remote sensing

MDPI

Article

From Wood–Foliar Semantics to 3D Fuel Characterization: Transferable Deep Learning for Terrestrial LiDAR Across Global Forest Ecosystems

Nadeem Fareed [1,*], Carlos Alberto Silva [1], Alexander J. Gaskins [1], Susan J. Prichard [2], Andrew T. Hudak [3], Jinyi Xia [1] and Cesar Ivan Alvites Diaz [1]

[1] School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL 32611, USA; c.silva@ufl.edu (C.A.S.)
[2] School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195, USA; sprich@uw.edu
[3] Rocky Mountain Research Station, USDA Forest Service, 1221 South Main St, Moscow, ID 83844, USA; andrew.hudak@usda.gov
* Correspondence: fareed.nadeem@ufl.edu or nadeem@geomatics.ncku.edu.tw; Tel.: +1-605-592-4878

Highlights

What are the main findings?

- Developed a sensor-agnostic deep learning framework for wood–foliar semantic segmentation of terrestrial LiDAR point clouds.
- Demonstrated robust transferability across globally distributed forest ecosystems, terrestrial LiDAR platforms, and seasonal leaf-on/leaf-off conditions.

What are the implications of the main findings?

- Point Transformer achieved the most consistent cross-dataset semantic segmentation without benchmark-specific optimization.
- Probability-based predictions revealed localized annotation uncertainty, particularly within fine branches and densely occluded canopies.
- Binary semantic segmentation enabled hierarchical geometric decomposition into ecologically relevant structural fuel classes.

Abstract

Terrestrial Light Detection and Ranging (LiDAR) provides detailed three-dimensional (3D) observations of forest structure, yet transferable wood–foliar semantic segmentation remains challenging because of forest structural heterogeneity, occlusion, and variability across forest ecosystems and terrestrial LiDAR platforms. Existing Deep Learning (DL) approaches are commonly developed for localized forest conditions or complex multi-class semantic taxonomies, often limiting their transferability to structurally distinct forests. Here, we introduce points2SBL, a geometry-driven, terrestrial LiDAR platform-agnostic DL framework that reduces heterogeneous forest vegetation to a transferable wood–foliar representation and subsequently uses these semantics as the foundation for 3D forest fuel characterization. The framework was developed using globally harmonized benchmark datasets acquired across multiple terrestrial LiDAR systems and broadleaf, coniferous, mixed, regenerating, and structurally complex forests. Three point-based architectures (PointNet++, PointNeXt, and Point Transformer) were benchmarked under identical training and evaluation protocols. Point Transformer provided the most consistent cross-dataset performance, achieving overall accuracy (OA) of 92–94%, mean

Academic Editors: Antonio Luis Montealegre Gracia, Alberto García-Martín and Raúl Hoffrén

Received: 31 July 2026
Revised: 10 September 2026
Accepted: 11 September 2026
Published: 14 September 2026

Intersection-over-Union (mIoU) of 0.78–0.87, macro F1-score of 0.87–0.93, and Matthews correlation coefficient (MCC) of 0.74–0.86 across five independent benchmark datasets. Species-level, vertical-profile, and qualitative cross-ecosystem assessments further demonstrated consistent preservation of wood–foliar semantics throughout the 3D forest fuel continuum, including previously unseen forest point clouds, while revealing that some apparent semantic disagreements originated from manual annotation of benchmarks. Finally, the points2SBL wood–foliar semantics were geometrically decomposed using a geometric algorithm named "Points2woodyseg" into fuel-related components: stem, branches, leaf, surface wood, and surface foliage. Collectively, these results demonstrate that a simplified wood–foliar abstraction can provide a transferable semantic foundation for characterizing structurally heterogeneous terrestrial LiDAR point clouds and deriving essential 3D forest fuel information across diverse forest ecosystems.

Keywords: deep learning; terrestrial laser scanning; forest; wildland fuel mapping; semantic segmentation; global ecosystems

1. Introduction

Since the release of the first commercial Terrestrial Laser Scanner (TLS) in 1998 [1], Light Detection and Ranging (LiDAR) systems have substantially improved the ability to capture detailed understory and near-surface forest structure through high-density point clouds [2]. Modern LiDAR systems encompass a broad range of sensor platforms, including TLS, handheld, backpack, mobile laser scanning (MLS), personal laser scanning (PLS), and Simultaneous Localization and Mapping (SLAM)-based LiDAR systems, enabling flexible and efficient data acquisition across structurally complex forest environments [1,3–5].

Terrestrial LiDAR can resolve detailed 3D structural information associated with stems, branches, foliage, understory, shrubs, coarse woody debris (CWD), and fine woody debris (FWD). This information supports a wide range of ecological and operational applications, including forest inventory, aboveground biomass (AGB) estimation, habitat assessment, branch characterization [1], canopy cover estimation, gap fraction analysis, Leaf Area Index (LAI), Leaf Area Density (LAD), Canopy Base Height (CBH), and Canopy Bulk Density (CBD). These are essential metrics for currently available wildfire behavior models used for operational decision support [6–9]. TLS-based point clouds are highly heterogeneous in forested sites and can require automated segmentation frameworks to classify points into ecologically meaningful structural components [10–12]. In terrestrial LiDAR applications, semantic segmentation assigns a class label to each point according to its structural category (e.g., stem, branch, and foliage), whereas instance segmentation further distinguishes individual object instances within the same semantic classes, such as individual trees [13–15]. Of these two tasks, accurate semantic segmentation is particularly important for characterizing complete forest strata and near-surface fuel complexes [16,17].

To achieve semantic segmentation of terrestrial LiDAR point clouds, parametric algorithms, e.g., Density-Based Spatial Clustering of Applications with Noise (DBSCAN), machine learning (ML), and DL have been extensively investigated in past studies [10,18–20]. Parametric algorithms and ML frameworks are generally easier to implement and computationally efficient, yet their performance often remains sensitive to forest structural variability, sensor characteristics, and manually tuned parametric thresholds, limiting their transferability [3,10,21,22]. Recent studies comprehensively evaluated a large number of widely used wood–leaf separation algorithms and ML/DL frameworks to assess their performance on wood–leaf segmentation [23,24]. The results demonstrated substantial variability in individual tree-AGB (IAGB) among methods, primarily due to difficulties in accurately separating branches from leaf clouds. Inaccurate branch delineation resulted in notable underestimation of IAGB, reaching up to 14% in broadleaf and 17% in coniferous species [23,24]. The study further reported that 3DSegFormer [14], a DL framework, achieved better IAGB compared with other approaches, highlighting the potential of DL frameworks to better capture complex branch architecture, whereas trunk biomass predictions remained comparatively consistent across algorithms [22,23,25].

In particular, supervised point-wise DL frameworks directly operate on 3D irregular point clouds and can learn complex geometric representations from the data, thereby improving segmentation quality under structurally heterogeneous forest conditions [14,26–28]. Recent advances in point-wise supervised DL architectures, including PointNet, PointNet++ [24], PointNeXt, Point-based convolutional neural networks (e.g., PointCNN), KPConv, RandLA-Net [29], Point Transformer [27], LWSNet, and Forestformer3D, among others, have substantially advanced semantic segmentation of terrestrial LiDAR point clouds by enabling hierarchical feature extraction and local neighborhood learning [3,16,30]. These frameworks have consistently demonstrated improved segmentation compared with conventional parametric approaches, particularly in complex forest environments where occlusion and fine-scale structural variability remain challenging [14,23,24].

1.1. Related Work

LiDAR semantic segmentation generally falls into two categories: tree-centric and plot-level methods. Tree-centric approaches operate on individual trees isolated from surrounding vegetation and are primarily designed for applications such as branch characterization, tree morphology analysis, IAGB estimation, and species classification [1,18,23,25,31,32]. Though these methods provide detailed information on individual tree structure, they are not intended to represent the full semantic complexity of forest plots containing canopy, understory, and surface fuels [23,33]. Plot-level semantic segmentation remains comparatively less investigated and is typically formulated as a multi-class semantic problem involving classes such as stem, branch, leaf, understory, ground, and CWD [18,34–37]. Although many frameworks report high semantic accuracies, model development and evaluation are often conducted within localized forest environments, where training and testing data originate from similar ecological and structural conditions [38,39]. Consequently, performance may decline when applied to forests with different species compositions, structural characteristics, fuel complexes, disturbance histories, or sensor acquisition geometries, frequently requiring additional labeled data and model retraining to maintain performance [8,26,28,36,37,40].

Furthermore, many existing segmentation frameworks rely on supplementary attributes such as intensity, optical imagery red–green–blue (RGB)-based colorization, multispectral imagery, multispectral LiDAR, geometric descriptors, or combinations thereof to improve classification performance [18,20,41]. Radiometric or spectral features can enhance wood–leaf discrimination [13,41]; however, their consistency is often affected by sensor characteristics, acquisition geometry, radiometric responses, and variability in wood and leaf reflectance, limiting transferability across sensor platforms and forest conditions [16,37,42]. In addition, most frameworks have been evaluated using relatively few forest plots representing localized environmental conditions, leaving their broader transferability capability largely unexplored [34,40,43].

1.2. Terrestrial LiDAR on Wildfire Fuel Mapping

In fire behavior modeling, forest structure is broadly classified into three interconnected fuel strata—surface fuels (e.g., litter, grasses, and low shrubs), ladder fuels (vegetation and wood materials that create vertical fuel continuity between surface and canopy layers, facilitating the upward spread of fire into tree crowns), and canopy fuels (live and dead foliage, twigs, and branches within the tree crowns). Fuel strata vertical and horizontal continuity strongly influences fire ignition, crown fire transition, spread dynamics, and fire intensity [9,44,45]. Among forest fuel strata, canopy fuels have been the most extensively studied because the overstory is comparatively easier to characterize using remote sensing, including optical imagery, spaceborne LiDAR (e.g., Global Ecosystem Dynamics Investigations (GEDI)), airborne laser scanning (ALS), and Synthetic Aperture Radar (SAR), which effectively characterize canopy height, cover, structure, and AGB across large spatial scales [46–48]. In contrast, ladder and surface fuels are substantially more difficult to quantify because of canopy occlusion and fine-scale structural heterogeneity [16,35,49,50].

However, forest lower strata contain herbaceous (grasses and forbs) and highly heterogeneous woody fuel complexes, including shrubs, regenerating seedlings and saplings, downed woody materials, e.g., dead twigs, branches, stems, and boles broadly classified as CWD (diameter ≥ 7.6 cm) and fine woody debris (FWD; diameter < 7.6 cm), which collectively govern horizontal and vertical fuel continuity, surface fire propagation, fire intensity, and crown-fire transition dynamics [10,51,52]. CWD and FWD are commonly quantified using field-based fuel inventory methods that measure fuel diameter, length, volume, spatial distribution, and fuel loading across surface and near-surface strata [52]. Fine fuels strongly influence fire behavior (including grasses, shrubs, and FWD). Fine and coarse woody fuels collectively influence fuel continuity, fuel moisture dynamics, fire intensity, rate of spread, and crown-fire initiation. Fine woody fuels primarily govern ignition and rapid fire spread, whereas CWD sustains higher heat release and prolonged smouldering combustion after flaming has ceased, contributing to extended wildfire smoke emissions [53–55]. For that/those reason/s, 3D fuel characterization is essential yet remains challenging due to their highly heterogeneous spatial organization, variable decomposition states, and complex 3D structure across forest ecosystems [55–57].

Forest ecosystems exhibit substantial variability in species composition, canopy architecture, understory complexity, shrub abundance, regeneration dynamics, and the spatial distribution of CWD and FWD [58–60]. Consequently, semantic classes commonly used in existing multi-class segmentation frameworks—such as ground, understory, lower objects, stem, and foliage—may exhibit substantial variability across diverse forest ecosystems, with some classes becoming sparse, structurally ambiguous, or absent under certain environmental conditions [1,36,60]. Recent studies have similarly highlighted difficulties in separating understory and lower-object classes because of their strong geometric overlap with terrain and fine-scale mixing of wood–leaf components [35,36]. In contrast, although forest fuels comprise multiple structural strata and fuel types, their vegetation components can be represented for semantic segmentation by two primary combustible materials: wood–foliar biomasses (Figure 1).

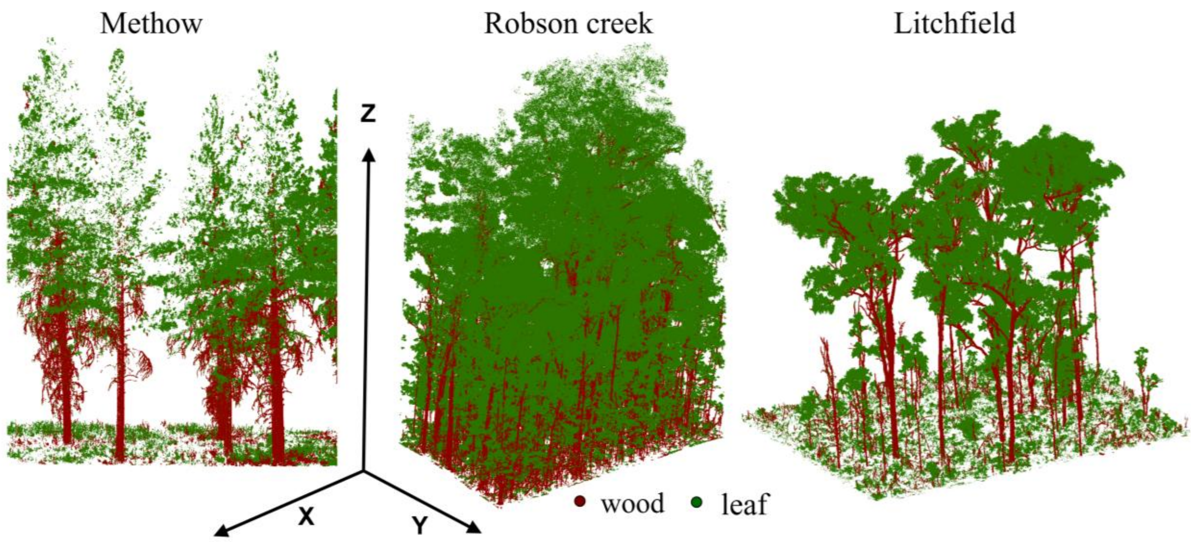

Figure 1. Representative wood–leaf organization across globally distributed forest ecosystems—Methow from the United States of America (USA), Robson Creek, and Litchfield from Australia. Wood is shown in red and foliage in green. Ground points were removed to highlight the 3D distribution of vegetation across canopy, ladder, understory, and near-surface fuel strata using FAST-GC (https://pypi.org/project/fastgc/ (accessed on 15 July 2026))—developed by the corresponding author.

Wood fuels—including stems, branches, CWD, and FWD—form relatively persistent structural frameworks, whereas foliar fuels such as foliage, grasses, and fine litter are more dynamic and strongly influence ignition and fire spread [44]. Wood–leaf abstraction aligns with established fire behavior concepts while reducing semantic ambiguity in terrestrial LiDAR point clouds [9]. Existing segmentation frameworks primarily emphasize structural classes rather than explicitly representing fuel heterogeneity across canopy, ladder, and surface fuel strata [29]. Following robust ground filtering, the remaining above-ground point cloud consistently consists of wood and foliar components distributed throughout the vertical 3D fuel continuum, regardless of forest type or structural complexity [61,62]. Since wood–leaf classes represent fundamental structural elements common to virtually all forest ecosystems, they can provide a more stable and transferable target representation than ecosystem-specific semantic classes (Figure 1). Moreover, the millimeter-scale geometric detail captured by terrestrial LiDAR systems offers an opportunity to develop sensor-agnostic segmentation frameworks that not only separate wood and foliar components but also support subsequent decomposition into ecologically meaningful fuel classes relevant to wildfire behavior modeling [36].

1.3. Research Objectives

To the best of our knowledge, no previous study has comprehensively evaluated wood–leaf semantic segmentation transferability across complete global forest plots while explicitly representing the full 3D fuel continuum, including canopy, ladder, and surface fuel strata. We focus exclusively on terrestrial LiDAR systems because their proximal sensing capabilities enable ultra-high-density point clouds with substantially greater structural detail than airborne LiDAR platforms, thereby providing improved characterization of surface, understory, and vertically connected fuel complexes that strongly influence wildfire behavior [49,63]. Furthermore, the datasets used in this study were acquired using several widely used terrestrial LiDAR platforms, e.g., RIEGL, GeoSLAM, and Leica [27,35]. Collectively, the independent terrestrial LiDAR datasets encompass a wide spectrum of forest ecosystems, vegetation structures, climatic conditions, phenological states, terrestrial LiDAR platforms, and acquisition strategies, providing a globally representative evaluation framework for assessing the robustness, transferability, and operational applicability of DL models beyond conventional benchmark datasets [36]. Consequently, this study aims to develop a terrestrial LiDAR platform-agnostic DL framework for semantic segmentation. Specifically, the study addresses the following objectives:

1. Evaluate whether point-based supervised DL frameworks can achieve robust and transferable wood–leaf semantic segmentation across diverse forest ecosystems, sensor platforms, point densities, acquisition conditions, and seasonal leaf-on and leaf-off conditions [36].

2. Benchmark three widely used supervised point-based architectures—PointNet++, PointNeXt, and Point Transformer [23,27,35,64] —using harmonized, manually annotated terrestrial LiDAR point clouds and identical training, validation, and evaluation protocols, thereby isolating the influence of network architecture on segmentation accuracy and cross-ecosystem transferability.

3. Investigate whether wood–leaf segmentation can serve as a foundation for automated fuel characterization by developing a geometric post-processing framework that further partitions wood–foliar binary abstractions into ecologically and operationally relevant fuel classes (i.e., stems, branches, canopy foliage, wood debris, and surface leafy fuels). We hypothesize that the wood component contains sufficient geometric information to support subsequent structural decomposition while reducing computational requirements and improving transferability relative to existing multi-class DL segmentation frameworks.

2. Materials and Methods

2.1. Datasets

To address the research objectives (Section 1.3), a comprehensive set of manually annotated point clouds representative of globally distributed forest ecosystems is required to train DL frameworks [27,36,65]. Given the limited availability of manually annotated forest point clouds, a targeted search of the published literature and publicly available data repositories was conducted to identify suitable datasets for model development. Search terms combined forest point cloud descriptors (e.g., "forest," "point cloud," "LiDAR," and "labelled/annotated") with acquisition-platform terms, including "TLS", "MLS", and "PLS". Additionally, terms such as "benchmark," "semantic segmentation," and "wood–foliage/leaf classification" are also used to cover the wide range of keywords used for annotated datasets [66,67].

Candidate datasets were considered for inclusion when (i) the point clouds and corresponding point-level semantic annotations were publicly accessible and open-source, (ii) the annotations contained classes that could be explicitly mapped to wood–foliage taxonomy—a desired input for DL framework training and evaluation, (iii) metadata were available to interpret the LiDAR sensor and forest ecosystem context, and (iv) the annotation quality was considered adequate following the quality-control (QC) procedures described below (Section 2.2).

For the datasets satisfying these criteria, we assembled a multi-source benchmark collection spanning different geographic regions, forest structural conditions, acquisition platforms, and spatial representations (Table 1). The resulting collection was intended to maximize the diversity of forest and sensor conditions represented in the currently accessible annotated data rather than to constitute a statistically representative sample of global forests [3]. The integration of these complementary datasets enables model development across multiple hierarchical levels of forest structure, ranging from individual trees and plot-level forest stands (understory and surface removed) to complete forest scenes containing understory vegetation, surface fuels, and terrain, thereby supporting robust learning across increasingly complex structural environments.

Table 1. Summary of the main characteristics of open-source terrestrial LiDAR point clouds used in the DL framework's training and validation.

Dataset	Country	Data Type	Total	Sensor	Labels
Lin3D [35]	China	Plot	1	TLS	Ground, understory, wood, leaf
SYSSIFOSS [68]	Germany	Individual trees	11	SLAM	Wood, leaf
SegmentedForest [60]	Austria	Plot	5	MLS	* Multi-class
	USA	Plot	3	TLS	* Multi-class
Plot-semantics [69]	Cameroon	plot	1	TLS	Wood, leaf
	Finland	plot	4	TLS	Wood, leaf
	Spain	plot	3	TLS	Wood, leaf
	Poland	plot	4	TLS	Wood, leaf

* Multi-class: Semantic labels comprising 16 manually annotated vegetation and non-vegetation categories, including ground and ground vegetation, shrubs, understory and lower-canopy trees, stems, branches and foliage, down wood, stumps, ivy, and ancillary non-vegetation objects (e.g., rocks, stakes, and people), following the SegmentedForests annotation scheme [60].

2.1.1. Wood–Foliar Plot-Level Datasets

Plot-level semantically labeled TLS point clouds from diverse forest ecosystems (Figure 2a,b) were incorporated to strengthen the representation of complete forest structural conditions. These datasets were acquired using multi-scan TLS protocols with extensive upright and tilted scan positions to minimize occlusion and preserve fine wood structures throughout the canopy profile [20]. Unlike isolated individual-tree datasets, plot-scale acquisitions capture complex interactions among stems, branches, foliage, understory, and surface fuels under natural stand conditions. The manually curated binary wood–leaf annotations provide high-quality benchmarks for learning wood–leaf structure across complete forest scenes [69].

The Lin3D dataset is a forest-scene point cloud benchmark developed for semantic segmentation across forest types and sensor platforms. It contains forest point clouds annotated into four semantic categories: foliage, wood (stems and branches), ground, and lower objects (e.g., grass and shrubs) [35]. In this study, we used TLS-annotated Lin3D (*v0.1)*, which comprises six annotated forest plots acquired in Guangxi Zhuang Autonomous Region, southern China. The TLS plots originate from the same regional Lin3D acquisition subset and therefore represent geographically related forest conditions; a single plot (plot # 1) with sufficient representation of both woody and foliar components was selected to incorporate the regional subset into the global forest ecosystem training and validation data pool. The remaining five plots were intentionally excluded from model development and retained as independent data for assessing model generalization using a spatially held-out sample [35].

The Plot-Semantics dataset is a high-density TLS dataset developed for wood–leaf semantic segmentation across diverse temperate forest ecosystems, comprising plots collected from Spain, Austria, Germany, Finland, and the United Kingdom (UK). Data were acquired using a RIEGL VZ-400i TLS (RIEGL Laser Measurement Systems GmbH, Horn, Austria) following a systematic multi-scan protocol within 30 × 30 m forest plots to minimize occlusion and capture detailed canopy architecture. The final dataset consists of twelve plots spatially partitioned into 10 × 10 m blocks at 1 cm resolution, containing xyz coordinates, reflectance values, and wood–leaf semantics of canopy and understory [69].

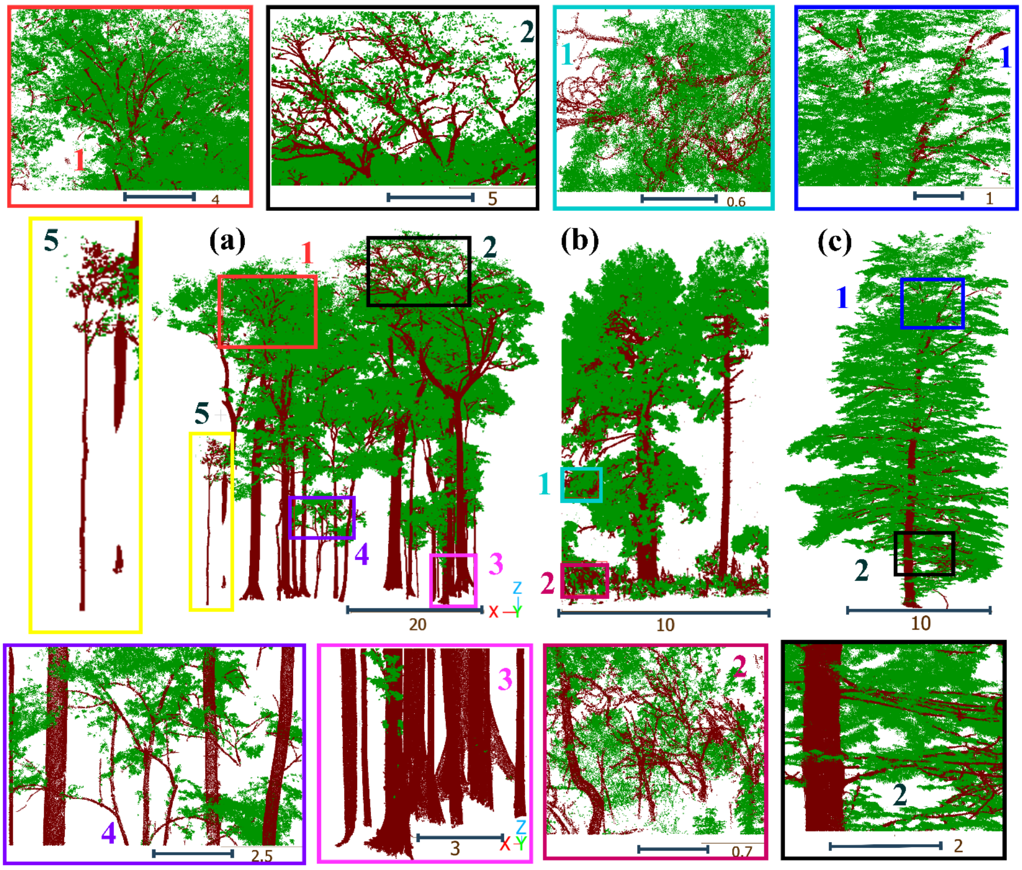

Figure 2. Representative examples of the labeled terrestrial LiDAR datasets used for training the wood–leaf segmentation framework. (**a**) Plot-level tree datasets containing only tree structures after removal of ground, understory, and surface fuel components. (**b**) Complete forest-plot datasets representing the full 3D fuel continuum from surface fuels to canopy foliage. (**c**) Individual-tree datasets with detailed wood–leaf annotations. Insets (1–5) show representative examples from the corresponding datasets and demonstrate variability in branching architecture, crown complexity, vegetation density, understory structure, and fuel organization. Red points represent wood, and green points represent foliar-labeled point clouds.

2.1.2. Individual Tree Point Clouds

Second, manually annotated individual-tree TLS point clouds with wood–leaf annotations (Figure 2c) were incorporated to complement the plot-level benchmarks. Unlike plot-level acquisitions, these datasets provide detailed representations of individual tree architecture, preserving fine branches, crown morphology, and wood–foliar interactions while substantially reducing occlusion through multi-scan acquisitions [68]. This enables the training framework to learn detailed structural characteristics across diverse branching patterns and crown forms, thereby improving its ability to generalize across heterogeneous forest conditions.

The SYSSIFOSS-Trees benchmark consists of 11 manually annotated individual-tree TLS point clouds representing seven temperate and boreal tree species, namely Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), Douglas fir (*Pseudotsuga menziesii*), European beech (*Fagus sylvatica*), sycamore maple (*Acer pseudoplatanus*), silver birch (*Betula pendula*), and European aspen (*Populus tremula*), acquired under leaf-on conditions in managed mixed forests of southwest Germany [70]. Individual trees were scanned from five to eight terrestrial laser-scanning positions using a RIEGL VZ-400 to maximize canopy coverage and minimize occlusion. Each point cloud was manually annotated into binary wood and foliar classes, providing a high-quality benchmark for evaluating fine-scale wood–foliar semantic segmentation across diverse crown architectures and branching organizations [26].

2.1.3. Multi-Class Plot Labels

Existing plot-level forest semantic segmentation benchmarks provide detailed multi-class semantic annotation comprising 16 manually annotated vegetation and non-vegetation categories, including ground and ground vegetation, shrubs, understory and stems, branches and foliage, down wood, stumps, ivy, and ancillary non-vegetation objects (e.g., rocks, stakes, and people), following the SegmentedForests annotation scheme [60]. However, a major limitation of SegmentedForest is the absence of explicit wood–foliar semantics; e.g., leaves and fine branches are often grouped within the same vegetation class, obscuring the structural distinction between fine wood and foliar components (Figure 3a). Consequently, fine branches are frequently grouped with foliage within the same canopy class, obscuring the structural distinction between wood–foliar components (Figure 3a).

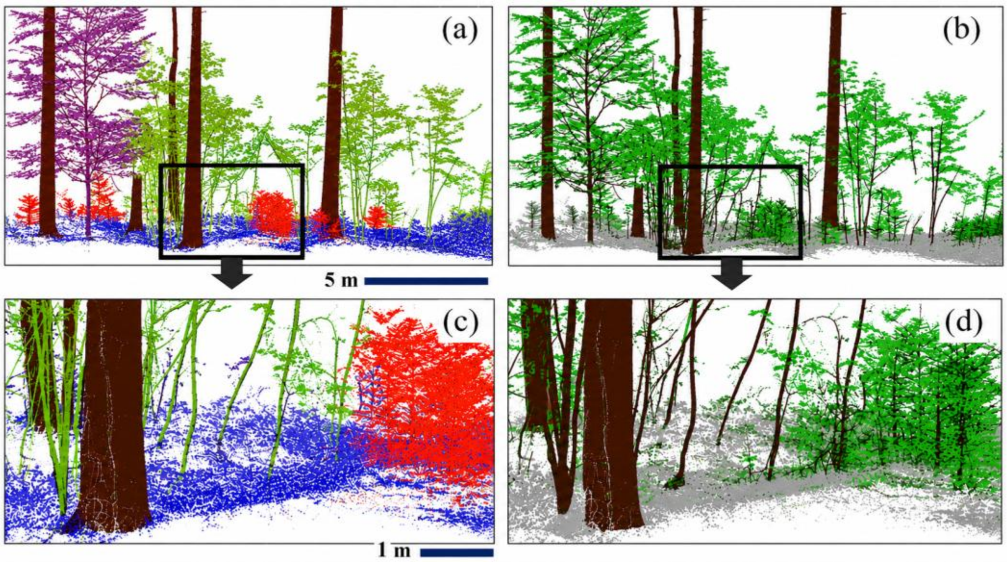

Figure 3. Harmonization of the SegmentedForests [60] multi-class benchmark into wood–foliar annotation used in this study. Panels (**a**) and (**c**) present the original semantic labels comprising ground and surface vegetation (blue), stems (brown), shrubs (red), understory vegetation (green), and fine branches and foliage (pink). Panels (**b**) and (**d**) show the harmonized annotations after semantic aggregation and manual refinement, where all woody structures were merged into a single wood class (brown), all foliage-bearing vegetation into a foliar class (green), and ground into a grey class.

To overcome this limitation, we incorporated a manually refined subset of the SegmentedForests benchmark [68], which contains forest point clouds acquired using TLS and MLS across multiple forest ecosystems. The original plots range from approximately 1300 to 3590 m^2. Rather than annotating the complete forest plots, representative 5 m × 10 m transects were extracted from each plot (plots #6–10 and 12–14) to capture these contrasting structural conditions while substantially reducing the manual annotation effort required for the much larger original scenes. Manual refinement was performed in CloudCompare (*Version:2.14 beta*) following the same slice-based workflow described by the original authors, ensuring methodological consistency with the published benchmark [60]. To this end, branches were manually separated from foliage throughout tree crowns, understory, shrubs, and wood debris to produce high-quality binary semantic labels (Figure 3c,d). Finally, the refined annotations were harmonized into a unified binary representation by merging stems, branches, CWD/FWD, and woody shrubs into a single wood class, while foliage and all foliar vegetation were assigned to the foliar class (Figure 3b–d).

https://doi.org/10.3390/rs18183158

ForestSegmented represents diverse natural and managed temperate forest ecosystems across Austria and the USA, spanning broadleaf, coniferous, and mixed stands dominated by European beech (*Fagus sylvatica*), European ash (*Fraxinus excelsior*), Norway spruce (*Picea abies*), silver fir (*Abies alba*), ponderosa pine (*Pinus ponderosa*), and longleaf pine (*Pinus palustris*) [60]. Structurally, these plots encompass multi-layer canopies, dense understory, natural regeneration, open woodlands, and stands with strong vertical fuel continuity (Figure 3a–c). The processed subset contributed approximately 51.2 million labeled points, comprising 10.9 million wood points (21.29%) and 40.3 million leaf points (78.71%).

2.2. Quality Control and Data Inclusion Criteria

All labeled datasets were subjected to rigorous QC to ensure consistency in wood–foliar annotations across datasets (Table 1). Plot-level binary labels were assessed through statistical sampling to identify potential labeling errors. Because visual assessment of entire forest plots is difficult in high-density terrestrial LiDAR point clouds, a transect-based QC approach was implemented in which total (n = 100) non-overlapping (10 × 5) m transects were automatically extracted from each plot for detailed visual qualitative assessment. This procedure enabled efficient assessment of canopy, understory, and wood–foliar labels across structurally heterogeneous forest environments. Datasets exhibiting poor labeling quality or combined wood–leaf classes were excluded. Although manually annotated forest point clouds inherently contain some degree of labeling uncertainty. Point clouds were retained when residual labeling inconsistencies were visually assessed to be minor (≈5%), with particular attention given to ambiguous wood–foliar boundaries, occluded branches, and dense canopy regions. The 5% value therefore represents a conservative visual quality-control criterion rather than a statistically estimated annotation-error rate [71]. This approach is consistent with the inherent uncertainty reported for manually annotated forest point cloud benchmarks, particularly in regions affected by occlusion and complex crown structure. Recent independent validation of manually corrected TLS wood–foliar annotations has similarly reported 96.34% mean agreement between final labels and independent manual re-classification, corresponding to approximately 3.66% disagreement [66].

While recent studies have increasingly relied on synthetic forest point clouds to augment training data, simulated environments cannot fully reproduce the structural complexity, ecological variability, sensor-specific artifacts, and labeling uncertainty present in real forests [36]. Furthermore, training on narrowly distributed datasets may increase susceptibility to domain shift and negative transfer when models are applied to structurally distinct forest conditions [36,72]. Negative transfer occurs when a model trained on one or more source forest ecosystems experiences a measurable decline in semantic accuracy after transfer to previously unseen target forests exhibiting different structural characteristics, species compositions, or acquisition conditions than those represented in the training data. To the best of our knowledge, this represents the first effort to systematically combine tree-assembly (plots without understory and ground; Figure 2a), individual-tree (Figure 2c), and complete plot-scale terrestrial LiDAR datasets (Figures 2b and 3) for the development and evaluation of a generalized wood–leaf semantic segmentation framework.

2.3. Structural Variability and Class Composition of the Training Datasets

Across all benchmark datasets, foliar points accounted for approximately 74% of all labeled observations, whereas wood points represented 26% (Figure 4a), reflecting the natural class imbalance commonly observed in LiDAR acquisitions under leaf-on canopy conditions [18]. At the individual dataset level, however, class composition varied substantially, with wood fractions ranging from approximately 3% to 62% and foliar fractions from 38% to 97%, likely indicating pronounced differences in canopy density, branching

https://doi.org/10.3390/rs18183158

architecture, and structural organization among forest ecosystems. Within-dataset distribution of class proportions further highlights this heterogeneity (Figure 4b). The median foliar fraction was 0.68, compared with 0.32 for wood points, while the broad interquartile ranges reflect considerable variability in vegetation composition across benchmark datasets. Such variability is advantageous for model development because it exposes DL frameworks to a wide range of structural conditions rather than a narrow, species-specific domain. Point densities spanned more than one order of magnitude, ranging from approximately 5×10^3 to over 1.5×10^5 points m^{-2} (Figure 4c). Both wood and leaf classes were represented across the full density spectrum, indicating substantial diversity in sensor configurations, acquisition geometries, and scanning protocols. Similarly, vertical structural complexity varied markedly among datasets, with vertical extents ranging from 12 to 57 m (Figure 4d), encompassing low-stature vegetation, intermediate-height stands, and tall dominant canopy forests.

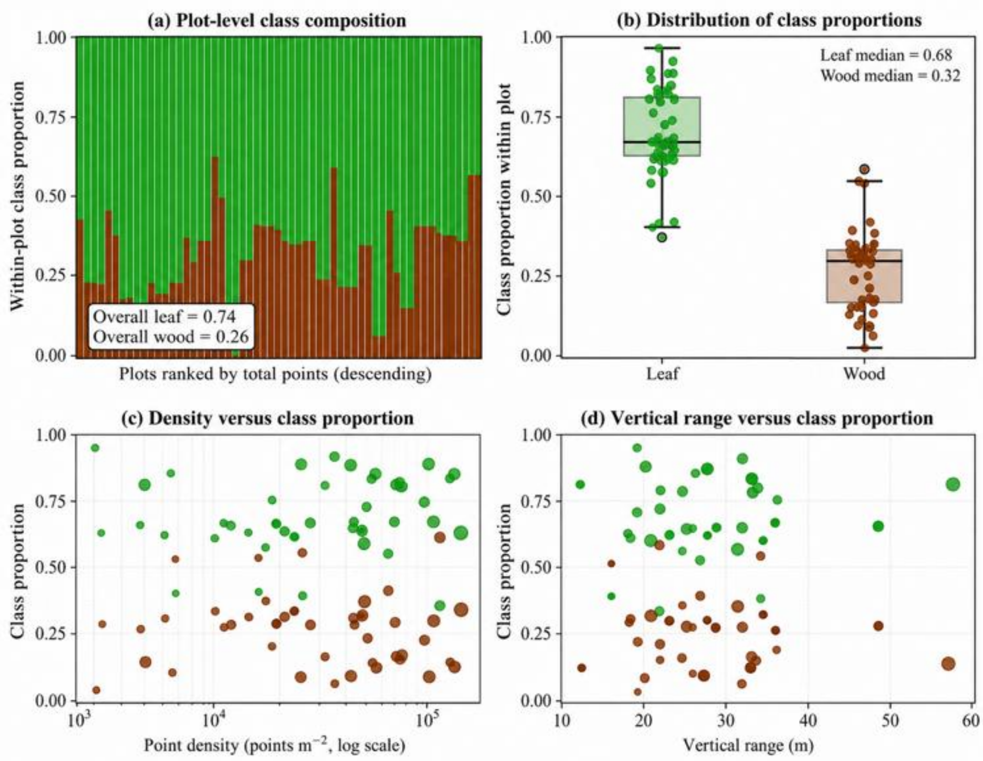

Figure 4. Structural and acquisition diversity of the training datasets used for wood–foliar semantics. (**a**) Within-scan wood and foliar class composition for all plot-level and individual-tree point clouds ranked by total labeled points. (**b**) Distribution of wood and foliar class proportions across datasets. (**c**) Variation in class composition across a broad gradient of point densities, representing different terrestrial LiDAR sensors. Point size is proportional to the total number of labeled points within each scan. (**d**) Variation in class composition across forest stands spanning contrasting vertical structural complexity, illustrating the broad structural domain. Brown represents wood points, whereas green represents foliar points throughout the figure.

2.4. Benchmark Datasets for DL Frameworks Evaluation

To evaluate model transferability and transferability, five independent benchmark datasets representing contrasting forest structures, fuel conditions, and spatial scales were used for model assessment (Table 2). Importantly, these datasets were not represented during training and validation (except Lin3D single plot (plot # 1) to understand how in-domain training validation influences the DL frameworks); they therefore provide an independent evaluation of model robustness across different forest regimes.

https://doi.org/10.3390/rs18183158

The TLS benchmark dataset consists of plot-level point clouds collected across Finland and Canada using Leica HDS6100 and Optech ILRIS terrestrial laser scanners operating under different acquisition configurations and wavelengths (Table 2). The subset used in this study includes lodgepole pine (*Pinus contorta Douglas*; LPine), red pine (*Pinus resinosa Aiton*; RPine), Scots pine (*Pinus sylvestris*; SPine), Norway spruce (*Picea abies*; NSpruce), sugar maple (*Acer saccharum Marshall*; SMaple), and trembling aspen (*Populus tremuloides*; TAspen), representing a broad range of forest structures, species compositions, and canopy architectures [73,74]. Representative examples of the plot-level benchmark datasets are presented in Figure 5a–c, illustrating wood–leaf annotations for NSpruce, RPine, and TAspen genera.

The Tropical Tree benchmark dataset consists of 147 individual tropical tree TLS point clouds acquired from three tropical rainforest sites in north-eastern Australia, namely Daintree Rainforest Observatory (DRO), Oliver Creek (OC), and Robson Creek (RC), using a RIEGL VZ-400i operating at 1550 nm, 300 kHz pulse repetition rate, 0.35 mrad beam divergence, and 0.04° angular resolution. The benchmark manually annotated trees representing 41 tropical species, selected from structurally diverse rainforest plots spanning tree heights from approximately 10 to 37 m and encompassing a wide range of crown architectures, stem diameters, and branching complexities. Trees were extracted from registered plot-level TLS acquisitions and subsequently labelled into wood and foliar classes (Figure 5e) through a semi-automated workflow followed by rigorous manual refinement [27].

The LeWoS benchmark dataset consists of 61 individual tropical tree TLS point clouds acquired from tropical forests in eastern Cameroon using a Leica C10 ScanStation under multi-scan acquisition protocols. The benchmark encompasses structurally diverse tropical trees spanning a broad range of heights (8.7–53.6 m) and diameters at breast height (10.8–186.6 cm), providing high-quality binary wood–foliar annotations (Figure 5f) for independent semantic segmentation evaluation [75].

ForestSemantic is a TLS benchmark dataset collected from six 32 × 32 m boreal forest plots in Evo, Finland, using a Leica HDS6100 TLS (Leica Geosystems AG, Heerbrugg, Switzerland) [76]. The dataset contains manually annotated plot-level point clouds representing the major boreal tree species, including Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), and silver birch (*Betula pendula*), with semantic labels for ground, trunk, first-order branch, higher-order branch, foliage, and miscellaneous objects [76]. It was used to evaluate the capability of the points2woodyseg framework (research objective 3) to further separate wood components into stem and branch classes. Surface wood and surface leaf annotations are not available; quantitative assessment was limited to stem, branch, and foliage classes within the annotated single plot data [76].

Table 2. Summary of main characteristics of open-source benchmark terrestrial LiDAR point clouds used for generalization and transferability assessment of the development of a DL framework.

Dataset	Country	Type	Total	Labels
TLS [73,74]	Finland and Canada	plot	12	Leaf–wood
Lin3D [35]	China	plot	5	Ground, understory, wood, leaf
Tropical [27]	Australia	Individual tree	147	Wood-leaf
LeWOS [75]	Cameroon	Individual tree	61	Wood–leaf
ForestSemantic [76]	Finland	plot	1	Ground, understory, stem, branches, leaf

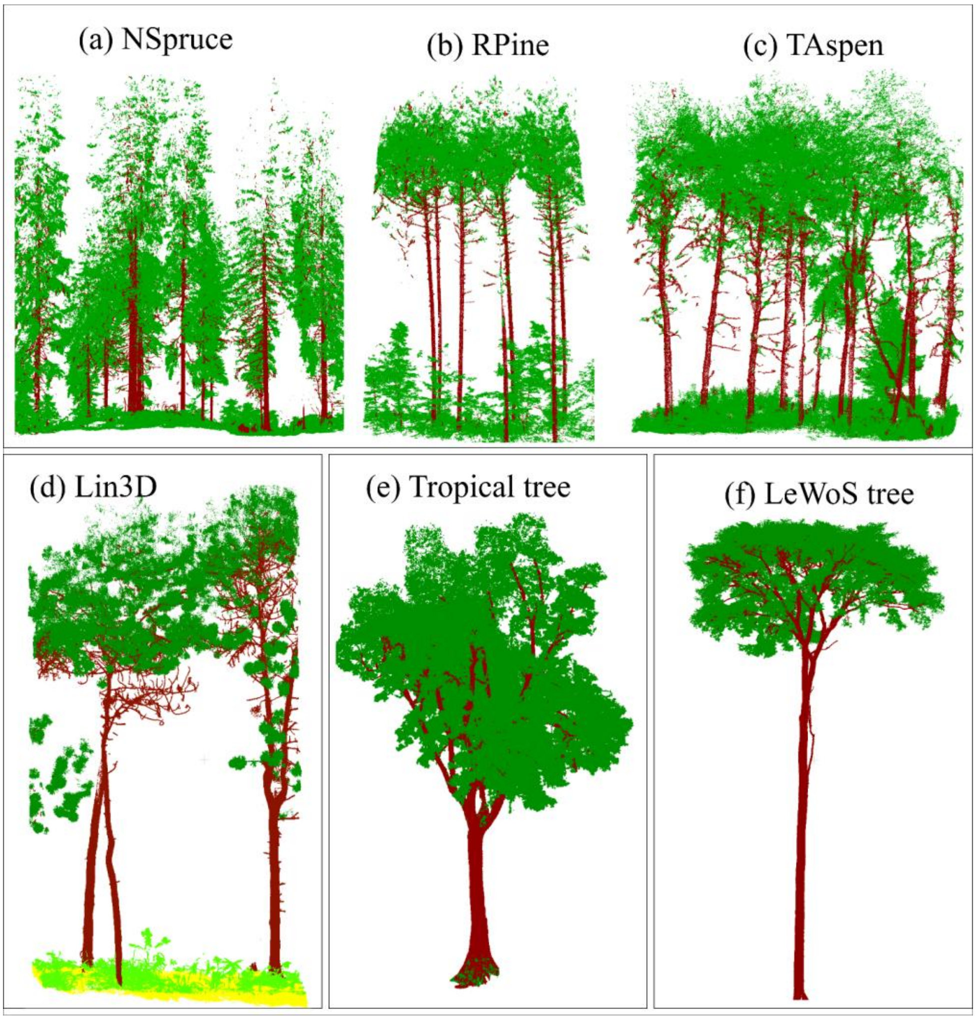

Figure 5. Independent benchmark datasets used for quantitative evaluation of wood–foliar semantic segmentation. The benchmarks comprise (**a**–**c**) complete forest plots, (**d**) plot-level tree datasets without understory (light green) and ground (yellow), and (**e**–**f**) individual-tree datasets. Red points denote wood components and green points denote foliar vegetation.

2.5. Global Datasets for Qualitative Evaluation

In addition to the benchmark (Table 2), a diverse collection of publicly available TLS and MLS datasets was assembled for qualitative assessment (Table 3). These datasets, including Wytham Woods, BlueCat, Ofental, Robson Creek, Litchfield, and other independent forest sites, have been widely adopted for DL frameworks for tree instance segmentation [12,39,77,78]. In contrast, their application to wood–foliar semantic segmentation has remained comparatively limited owing to the lack of consistent semantic annotations and standardized evaluation protocols. Collectively, these datasets encompass diverse forest ecosystems, acquisition geometries, sensor platforms, and structural conditions, thereby providing a rigorous assessment of DL frameworks' transferability under realistic operational scenarios.

Datasets (Table 3) were intentionally excluded from quantitative evaluation because they do not satisfy the QC criteria established in the present study (Section 2.2). For example, Wytham Woods, which is a leaf-off TLS dataset [36], and several recently released benchmarks provide high-quality TLS point clouds but do not include point-wise wood–foliar semantics required for quantitative evaluation. Conversely, datasets such as BlueCat provide wood–foliar semantics and have been adopted in previous semantic segmentation studies [30]; however, visual inspection revealed substantial annotation inconsistencies, particularly within fine branches, branch–foliage interfaces, and partially occluded woody structures. Consequently, these datasets were retained exclusively for qualitative assessment, which is a visual quality assessment of the processed dataset [35]. This evaluation strategy ensures that quantitative benchmark metrics are derived only from datasets satisfying the proposed QC protocol, while simultaneously demonstrating the transferability and operational applicability of the trained framework across a broad range of independently acquired forest LiDAR datasets.

Table 3. Summary of main characteristics of open-source global terrestrial LiDAR point clouds used for transferability assessment across diverse global forest ecosystems.

Dataset	Country	Type	Total	Sensor
SegmentedForest [60]	Spain, Austria, UK, USA	Plot	6	TLS, MLS
Wytham Woods [30,39]	UK	Plot	1	TLS
BlueCat [30]	Czech Republic	Plot	1	TLS
Robson Creek [39]	Australia	Plot	1	TLS
Ofental [39]	Germany	Plot	1	TLS
Australia	Australia	Plot	1	TLS
Methow [79]	USA	Plot	1	TLS
Litchfield [39]	Australia	Plot	1	TLS
TUWIEN [80]	Austria	Plot	1	ULS
NIBIO_MLS	Austria	Plot	1	MLS

2.6. Methodology Overview

The proposed framework, points2SBL, which performs segmentation of LiDAR point clouds into wood–foliar semantics, uses the diverse training and validation datasets presented in Section 2.1. The framework was specifically designed to generalize across heterogeneous forest ecosystems, structural conditions, and close-range LiDAR acquisition systems, including TLS, MLS, backpack laser scanning (BLS), and other near-ground platforms. Unlike conventional multi-class semantic segmentation approaches that rely heavily on radiometric information, spectral attributes, or sensor-specific features [80], points2SBL emphasizes geometry-driven learning using spatial structure and local neighborhood characteristics derived directly from 3D point clouds [18]. points2SBL is implemented in Python (*Version 3.1*) within the Anaconda environment with PyTorch (*Version 2.5.1*) and is publicly available at https://github.com/nadeemfareed/points2SBL (accessed on 15 July 2026) . Processing and benchmarking were performed on a high-performance workstation running Windows 11 Enterprise (*64-bit*), equipped with an Intel Xeon W5-2465X processor (*3.10 GHz*), 384 GB DDR5 RAM (4400 MT/s), and an NVIDIA RTX 5000 Ada Generation GPU (31 GB VRAM).

points2SBL supports CUDA-accelerated inference, configurable block batching, geometric-feature caching, and multi-layout voting to accommodate different computational and prediction-quality requirements. The maximum-quality configuration uses 10 confidence-weighted spatial voting layouts (HYBRID8-V10), whereas a four-layout configuration (GRID4-V4) provides a lower-cost alternative for large-scale processing. Processing time depends on point cloud size and density, GPU resources, and the selected voting configuration.

Figure 6 demonstrates the schematic design of points2SBL, which comprises six automated modules: (a) point cloud preparation and label harmonization, (b) spatial block generation and sampling, (c) geometric feature extraction from point clouds, (d) DL network training, testing, and validation, (e) tiled multi-vote inference, and (f) prediction refinement and post-processing—an optional step. Together, these modules produce (g) point clouds with wood–foliar semantics along with a continuous leaf-probability map. Each module is described in detail in the following subsections.

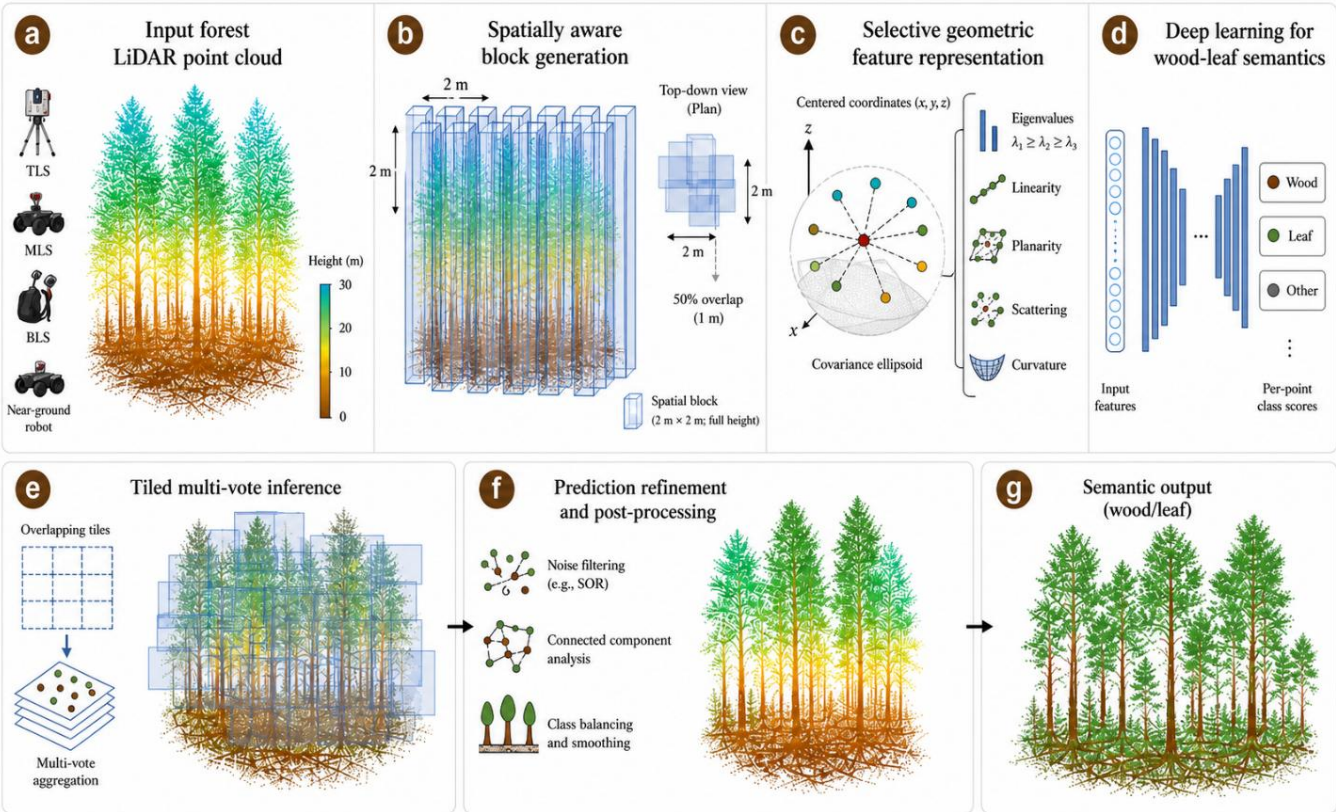

Figure 6. Conceptual workflow of points2SBL for wood–leaf semantics of terrestrial LiDAR. (**a**) Multi-platform near-ground LiDAR point clouds are partitioned into overlapping (**b**) spatial blocks represented using local (**c**) geometric descriptors derived from neighborhood covariance structure, to train (**d**) point-wise supervised DL frameworks. Classification outputs are aggregated through (**e**) tiled multi-vote inference and refined using (**f**) geometric post-processing operations to produce (**g**) final wood and foliar semantics.

2.6.1. Spatially Aware Block Generation

A large fraction of points in terrestrial LiDAR scans correspond to ground, which is not relevant for wood–leaf segmentation [62]. Ground points were first classified using the Fully Adaptive Self-Tuning Ground Classification (FAST-GC) algorithm (https://pypi.org/project/fastgc/, accessed on 25 July 2026). After removing the classified ground points, the remaining 3D point clouds consisted solely of wood and leaf components [18]. This pre-processing step also reduces the computational burden of geometric feature calculation and DL framework training. Wood–foliar point clouds were then partitioned into spatial blocks to further improve scalability and mitigate boundary artifacts during DL training and inference. Spatial partitioning was performed in the horizontal plane while preserving the complete vertical structure of vegetation and fuel strata within each block (Figure 6b). The spatial support of each block was defined as:

$$B_m = \{p_i \in P \mid x_m \le x_i < x_m + w_x,\ y_m \le y_i < y_m + w_y\} \quad (1)$$

where B_m represents the m-th spatial block, P denotes the input point cloud, and w_x and w_y represent the horizontal window dimensions. A set of candidate tile sizes (0.5, 1, 2, 4, and 6 m) was evaluated to select an appropriate horizontal block extent. Because of the strong wood–leaf class imbalance (Figure 4), very small tiles (<2 m) produced a large number of leaf-dominated blocks, whereas large tiles (e.g., 5 m) substantially reduced the number of training tiles for individual-tree point clouds (Figure 2). Based on these experiments, we adopted 2 × 2 m spatial windows with 50% overlap in both horizontal directions (x, y), which provided a well-balanced total number of training tiles across wood- and leaf-dominated regions. For each valid block, a fixed number of 8192 points was randomly sampled to maintain consistent tensor dimensions during optimization. Blocks containing fewer than 64 valid points were discarded, whereas partially sparse but still valid blocks were sampled with replacement to preserve fixed-size network inputs. To improve geometric invariance and reduce directional bias, training blocks were randomly rotated around the vertical axis and spatially centered prior to model optimization. Instead of splitting the input datasets into three distinct blocks (e.g., 70% training, 15% inference, and 15% testing), blocks from entire labelled point clouds were pooled and randomly shuffled before being partitioned into 80% training and 20% validation subsets to minimize site-specific and acquisition-specific bias during training and validation. The 50% spatial overlap was applied during training block generation, whereas overlapping windows were disabled for validation. Because blocks were pooled before partitioning, strict spatial independence between training and validation samples was not enforced, and spatially neighboring observations could therefore occur across the two subsets. Accordingly, the validation subset was used primarily to monitor model convergence and select the best-performing checkpoint, while model generalization was assessed separately using spatially held-out plots Lin3D and independent benchmark datasets.

2.6.2. Selective Geometric Feature Representation

The proposed framework was designed to operate exclusively on geometric features derived from local 3D neighborhood structure without relying on radiometric attributes and RGB [80]. This limitation is generally associated with labelled point clouds in that a large amount of labeled data is missing radiometric/intensity and RGB information [80,81]. Geometric features-based design improves transferability across heterogeneous terrestrial LiDAR systems where radiometric responses and acquisition geometries vary substantially among sensors and forest environments [82]. For each point, local geometric structure was estimated by k-nearest-neighbors (KNN) neighborhoods using covariance-based eigenvalue decomposition (Figure 6c). The covariance matrix was computed as:

$$\mathbf{C}_i = \frac{1}{k-1}\sum_{j=1}^{k}\left(\mathbf{p}_{ij} - \bar{\mathbf{p}}_i\right)\left(\mathbf{p}_{ij} - \bar{\mathbf{p}}_i\right)^T \tag{2}$$

where $\mathbf{C}_i$ denotes the covariance matrix associated with the point p_i, $\mathbf{p}_{ij}$ represents the j-th neighboring point, and $\bar{\mathbf{p}}_i$ denotes the centroid of the local neighborhood (Figure 6c). Eigenvalue decomposition was subsequently applied as:

$$\mathbf{C}_i\mathbf{v}_m = \lambda_m\mathbf{v}_m \tag{3}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ are the ordered eigenvalues and $\mathbf{v}_m$ are the corresponding eigenvectors. Four covariance-derived geometric descriptors were computed from the eigenvalues:

$$L_i = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \varepsilon} \tag{4}$$

$$P_i = \frac{\lambda_2 - \lambda_3}{\lambda_1 + \varepsilon} \tag{5}$$

$$S_i = \frac{\lambda_3}{\lambda_1 + \varepsilon} \tag{6}$$

$$C_i = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3 + \varepsilon} \tag{7}$$

where L_i, P_i, S_i, and C_i represent linearity, planarity, scattering, and curvature, respectively. The final point-wise feature vector comprises centered spatial coordinates and a compact set of covariance-derived geometric descriptors. Although many additional geometric metrics can be computed from TLS point clouds [24,65], all of them would substantially increase computational costs for feature calculation, model training, and inference on large scenes [11,81]. Moreover, several higher-order descriptors are sensitive to forest structure, sensor modality, and local point density, which can reduce robustness when transferring models across sites or LiDAR systems. We therefore focused on a small subset of descriptors that complement x, y, z and are most informative on wood–leaf separation. In particular, linearity, planarity, and curvature provide a stable foundation for distinguishing wood components from foliage, while scattering is highly indicative of leaf clusters. Using the Classification and Regression Tree (CART), a recent study concluded that curvature, planarity, and linearity are important distinguishing features for woody surfaces [24]. Furthermore, the training data contains a very large number of leaf samples but relatively fewer wood samples (Figure 4); selective geometric features characterize wood structure and potentially compensate for class imbalance for DL frameworks. All feature channels were independently standardized prior to model training using z-score normalization. DL frameworks were input with the same sets of seven-dimensional training, validation, and inference datasets (Equation (8)).

$$\mathbf{x}_i = [x_i, y_i, z_i, L_i, P_i, S_i, C_i] \tag{8}$$

2.6.3. DL Frameworks for Wood–Foliar Semantics

Three supervised point-based DL architectures were evaluated: PointNet++, PointNeXt, and Point Transformer. The primary configuration used seven input channels (Equation (8)) and binary semantic outputs corresponding to wood and foliar classes. The optimization objective used in all reported experiments combined weighted cross-entropy and Dice losses (Equation (9)), with weighting coefficients $\alpha = 1.0$ and $\beta = 0.2$, respectively. Class-specific weights of 3.0 and 1.0 were assigned to the wood and foliar classes, respectively, to account for class imbalance in the training data (Figure 4). Focal loss was implemented as an optional component of the framework but was disabled in the experiments reported here ($\gamma = 0.0$) [35,80].

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{CE}} + \beta\mathcal{L}_{\text{Dice}} + \gamma\mathcal{L}_{\text{Focal}} \tag{9}$$

Herein, α, β, and γ are weighting coefficients controlling the contribution of each loss component. Weighted cross-entropy loss was estimated as:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} w_c\, y_{ic}\log(p_{ic}) \tag{10}$$

Herein w_c denotes the class-specific weighting coefficient, y_i represents the reference class label, and $p_{i,c}$ denotes the predicted posterior probability for the class. Dice loss was computed as:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i\, p_i + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i + \varepsilon} \tag{11}$$

Focal loss was additionally implemented as an optional optimization component to emphasize difficult samples during training:

$$\mathcal{L}_{\text{Focal}} = -\frac{1}{N}\sum_{i=1}^{N}(1-p_i)^{\gamma} y_i \log(p_i) \tag{12}$$

https://doi.org/10.3390/rs18183158

Here, the Focal-loss focusing parameter was set to $\gamma = 2.0$; however, because the Focal-loss contribution weight was set to $\delta = 0.0$, this component did not contribute to the optimization objective.

Model optimization was performed using the *AdamW optimizer* with an initial learning rate of 1×10^{-4} weight decay of 0.01, a batch size of 8, and 40 training epochs with early stopping enabled. Learning-rate adaptation used a *OneCycleLR* schedule with a maximum learning rate of 3×10^{-4}, a warm-up fraction of 0.1, and cosine annealing. The best-performing validation checkpoint was retained for inference.

2.6.4. Multi-Vote Inference

To support operational-scale prediction of large terrestrial LiDAR point clouds, tiled multi-vote inference was implemented using overlapping spatial offsets (Figure 6e). During inference, each point cloud was partitioned into multiple overlapping tiles, and semantic predictions were independently generated for each tile configuration. Individual points, therefore, received multiple semantic predictions from neighboring inference windows. Final posterior probabilities were estimated using weighted ensemble averaging:

$$\bar{p}_i = \frac{\sum_{m=1}^{M_i} \omega_i^m p_i^m}{\sum_{m=1}^{M_i} \omega_i^m} \tag{13}$$

where p_i^m represents the predicted probability from the m-th vote, ω_i^m denotes the corresponding vote weight, and M_i is the total number of valid predictions received for the point p_i. The final semantic label was assigned using maximum posterior probability:

$$\hat{y}_i = \arg\max_c \bar{p}_{ic} \tag{14}$$

The multi-vote inference strategy substantially reduced tile-boundary artifacts and improved spatial consistency across neighboring prediction regions while preserving fine-scale wood and foliage structures under strong canopy occlusion conditions.

2.6.5. Prediction Refinement and Post-Processing

To improve the spatial consistency of wood–foliar semantics and reduce isolated segmentation artifacts, particularly when offset voting contributes less to per-point predictions at the scene edges of plots or individual trees, post-processing operations were applied following neural network inference. These operations were designed to preserve continuous wood structures while suppressing noisy leaf-to-wood and wood-to-leaf misclassifications commonly observed in dense canopy environments.

Following semantic prediction, neighborhood-based spatial refinement was applied to improve the local consistency of semantic labels. For each point, neighboring semantic probabilities were aggregated using a k-nearest-neighbor (KNN) neighborhood defined in 3D Euclidean space, with $K = 16$, to reduce isolated point-wise prediction noise:

$$\tilde{p}_i = \frac{1}{K} \sum_{j=1}^{K} p_j \tag{15}$$

where $\tilde{p}_i$ denotes the refined posterior probability for point i, p_j represents neighboring probabilities, and K is the number of local neighbors. Additional structural refinement operations were implemented to preserve the spatial continuity of wood structures and suppress isolated false-positive wood predictions. A subsequent structural-consistency refinement, also using $\mathrm{K} = 16$, protected high-confidence woody-core points based on $\mathrm{p}_{\mathrm{leaf}} \leq 0.18$, local woody support ≥ 0.55, geometric woody support ≥ 0.52, vote variance ≤ 0.030, and a minimum vote count of 2. Weak non-core woody predictions were retained when supported by a woody-core neighborhood fraction of at least 0.20, strong geometric woody evidence (≥ 0.68), or strong local woody support (≥ 0.72). These thresholds were ascertained with subsequent pooling of all possible combinations to obtain the threshold

https://doi.org/10.3390/rs18183158

of maximum accuracies over benchmarks. Nevertheless, continuous posterior probabilities were retained in the output point clouds for downstream analysis and uncertainty assessment. All quantitative performance metrics reported in this study were calculated from the final semantic classifications after prediction aggregation and the post-processing operations described above.

2.6.6. Performance Evaluation of DL Frameworks

For wood–foliar semantics, the DL frameworks were evaluated using standard point-level metrics derived from the confusion matrix, including overall accuracy (OA), mean class accuracy (mACC), mean Intersection over Union (mIoU), precision (P), recall (R), F1-score, macro F1-score (MF1), and the Matthews correlation coefficient (MCC) (Equations (16)–(23)). Class-specific omission error and commission error were additionally calculated to characterize semantic errors (Equations (24) and (25)). These metrics were computed for wood–leaf semantics across all terrestrial LiDAR point cloud datasets (Section 2.2). Global semantic performance was quantified using OA, mACC, mIoU, MF1, and MCC, whereas class-specific performance was assessed using precision, recall, and F1-score. Mean class accuracy represents the average class-wise recall and is therefore equivalent to balanced accuracy (BA) under the formulation used here; consequently, BA was not reported separately. Similarly, omission error is the complement of recall ($OE_i = 1 - R_i$), whereas commission error is the complement of precision ($CE_i = 1 - P_i$). Herein, TP_i, TN_i, FP_i, and FN_i represent the true positives, true negatives, false positives, and false negatives for class i, respectively; n denotes the total number of semantic classes, and Ndenotes the total number of evaluated points.

$$mIoU = \frac{1}{n}\sum_{i=1}^{n}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{16}$$

$$(mACC = BA) = \frac{1}{n}\sum_{i=1}^{n}\frac{TP_i}{TP_i + FN_i} \tag{17}$$

$$OA = \frac{\sum_{i=1}^{n}(TP_i + TN_i)}{\sum_{i=1}^{n}(TP_i + TN_i + FP_i + FN_i)} \tag{18}$$

$$P_i = \frac{TP_i}{TP_i + FP_i} \tag{19}$$

$$R_i = \frac{TP_i}{TP_i + FN_i} \tag{20}$$

$$F1_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \tag{21}$$

$$MF1 = (1/\mathrm{n})\ \Sigma_{\mathrm{i}=1}{}^{\mathrm{n}}\ \mathrm{F1_i} \tag{22}$$

$$MCC = \frac{(TP_\mathrm{i}TN_\mathrm{i} - FP_\mathrm{i}FN_\mathrm{i})}{\sqrt{[(TP_\mathrm{i} + FP_\mathrm{i})(TP_\mathrm{i} + FN_\mathrm{i})(TN_\mathrm{i} + FP_\mathrm{i})(TN_\mathrm{i} + FN_\mathrm{i})]}} \tag{23}$$

$$Omission\ Error = 1 - R_i \tag{24}$$

$$Commission\ Error = 1 - P_i \tag{25}$$

$$\mathrm{h_{r,\,i}} = (\mathrm{z_i} - \mathrm{z_{min}})\ /\ (\mathrm{z_{max}} - \mathrm{z_{min}}) \tag{26}$$

To quantify semantic performance along the vertical forest strata, mIoU, wood F1-score (WF1), wood precision (WP), and wood recall (WR) were assessed across normalized relative height (h_r) from the surface to the canopy. This analysis was designed specifically to characterize the vertical stratification of wood semantics and identify height-dependent patterns in wood classification performance within the forest profile. For each dataset, the normalized relative height of point i, $h_{r,i}$ was calculated using min–max normalization (Equation (26)), where z_i is the elevation of point i, and $z_{\min}$ and $z_{\max}$ are the minimum and maximum elevations, respectively, within the corresponding benchmark dataset. This transformation scaled the vertical extent of each dataset from 0 to 1, with the minimum and maximum elevations corresponding to relative heights of 0 and 1, respectively.

2.6.7. Evaluation Scenarios

DL frameworks were evaluated under three complementary scenarios: (1) complete forest plots containing surface, understory, and overstory vegetation (Table 2, Figure 5a–c); (2) individually segmented trees (Figure 5e,f); and (3) plot-level tree point clouds with surface and understory vegetation removed to isolate overstory tree structures and assess improvements in segmentation performance following the exclusion of lower vegetation strata, where wood–leaf discrimination is often more challenging [27]. This third scenario was particularly significant for the Lin3D benchmark, where understory and surface fuel components are not explicitly annotated for wood–leaf semantics (Figure 5d). The three-tier evaluation framework, applied across five benchmark datasets, i.e., three plot-level and two tree-level datasets (Section 2.3), enables a more comprehensive investigation of terrestrial LiDAR semantic segmentation across varying structural complexities and operational scales. To the best of our understanding of existing investigations of similar scopes, such a unified multi-scenario 3D forest vertical stratum evaluation remains largely unexplored, whereas a large number of studies have documented overall accuracy assessments for individual trees or forest plots [35,83] using a combination of equations (Equations (16)–(26)). Furthermore, the proposed framework was designed to establish a fully automated DL framework capable of operating robustly across these complementary scenarios, thereby extending its practical applicability and utility for broader forest vegetation, AGB, and fuel characterization using emerging terrestrial LiDAR systems [40,84].

3. Results

3.1. Comparative Assessment of DL Frameworks Performance

DL frameworks showed stable convergence, with rapid loss reduction during the first 10–15 epochs followed by gradual stabilization in both training and validation phases (Figure 7a,b). Point Transformer consistently achieved the lowest training and validation losses, indicating more effective optimization than PointNeXt (PNx) and PointNet++ (PN++). This trend was consistently reflected across all validation metrics, where PT maintained the highest and most stable performance throughout training, reaching peak values of approximately OA = 92.00, mIoU = 78.00, MF1 = 86.00, WF1 = 78.00, LF1 = 95.00, and mACC = 87.00 (Figure 7c–h). PointNeXt and PointNet++ followed similar optimization trajectories and converged to comparable performance, generally remaining 1–3 percentage points lower across most metrics (Figure 7c–h).

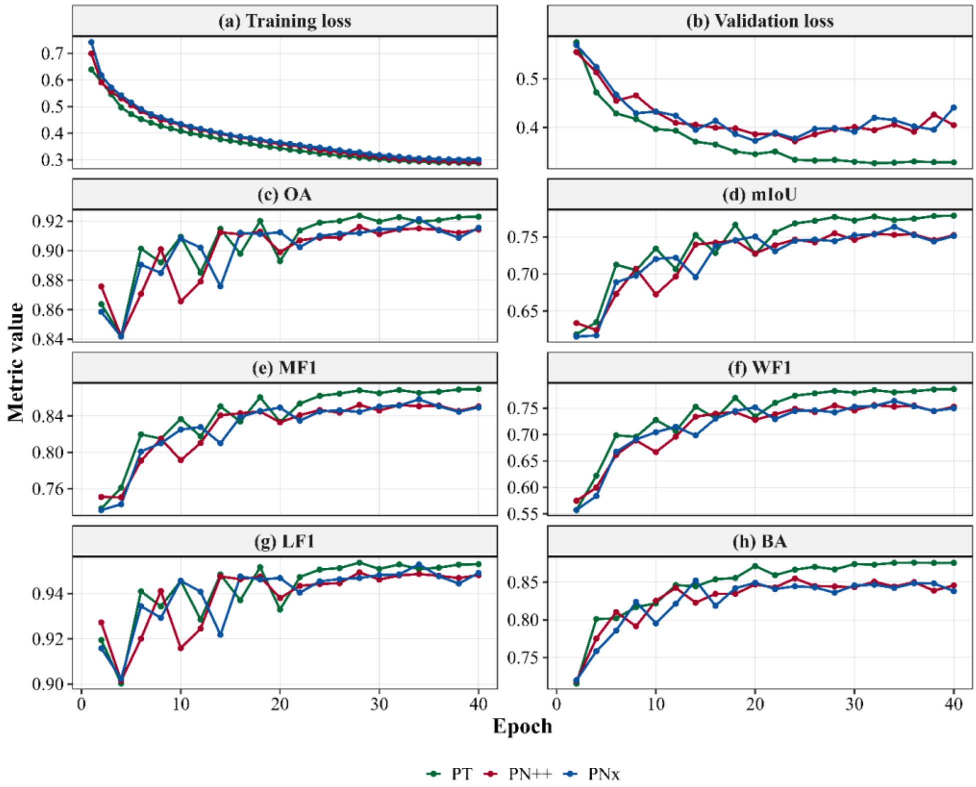

Figure 7. Comparison of training convergence and validation performance for Point Transformer, PointNet++ (PN++), and PointNeXt (PNx). Panels show (**a**) training loss, (**b**) validation loss, (**c**) overall accuracy (OA), (**d**) mean intersection over union (mIoU), (**e**) macro F1-score (MF1), (**f**) weighted F1-score (WF1), (**g**) leaf F1-score (LF1), and (**h**) (BA = mACC) over 40 training epochs. All models reached stable convergence within 40 epochs, after which no meaningful improvement in training or validation performance was observed.

Overall, these results indicate that all three architectures successfully learned the generalized wood–leaf semantics under identical training conditions, with Point Transformer showing superior optimization stability and stronger within-distribution validation performance [27]. However, because the validation subsets represent sparse samples drawn from the broader harmonized training pool (Figure 4), these results primarily reflect within-distribution model behavior. Therefore, the independent benchmark evaluations presented in the following sections provide a more rigorous assessment of model robustness and true cross-ecosystem transferability.

3.2. Post-Processing

An optional post-processing refinement step was applied after DL inference to improve the spatial consistency of wood–foliar semantics. This is an optional yet important step to suppress isolated artefacts in wood–foliar semantics. This refinement combined neighborhood-based posterior smoothing and local wood-support filtering to preserve continuous stem and branch structures under dense canopy occlusion. The effect of refinement was quantified as the change in performance metrics between raw and refined predictions, expressed in percentage points (pp). The response to refinement was strongly model-dependent (Figure 8). Point Transformer showed consistent improvement across nearly all metrics, including OA (+0.26 pp), mIoU (+0.36 pp), MF1 (+0.21 pp), MCC (+0.58 pp), WF1 (+0.12 pp), WIoU (+0.20 pp), and notably wood precision (+2.17 pp), indicating that its predictions already exhibited stronger spatial coherence and clearer structural boundaries between wood and leaf components (see Figure 8).

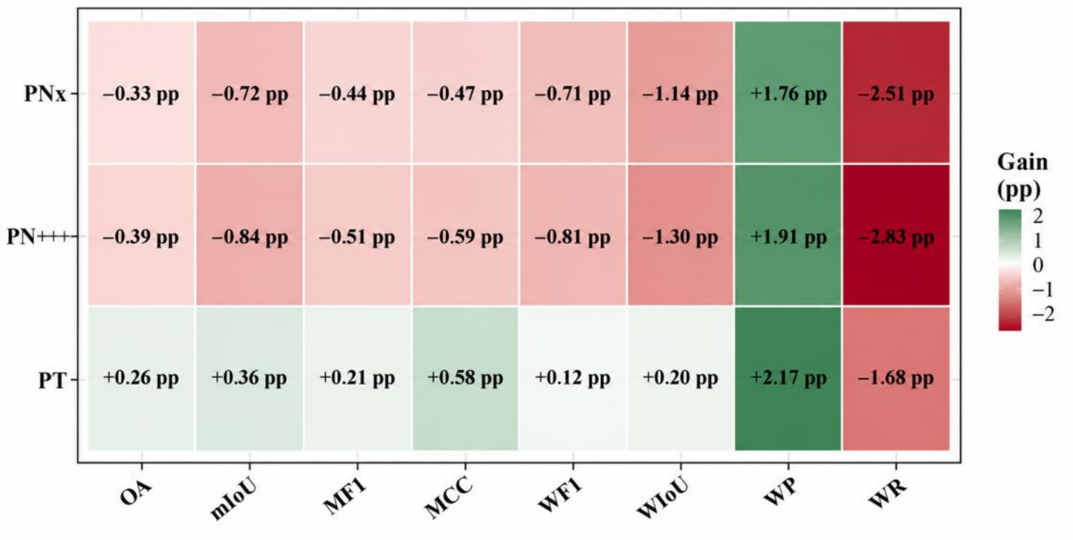

Figure 8. Performance change after optional post-processing refinement for PNx, PN++, and Point Transformer. Heatmap values show the percentage-point (pp) difference between refined and raw predictions across all evaluation metrics. Positive values indicate performance gains, while negative values indicate reductions. Refinement yielded the most consistent improvements for Point Transformer, whereas PNx and PN++ showed increased WP but reduced WR, reflecting greater sensitivity to over-smoothing of ambiguous local predictions.

Results indicate that neighborhood-based refinement was able to reinforce existing local consistency while suppressing localized salt-and-pepper inference artifacts, resulting in cleaner and structurally more coherent semantic boundaries even where overall metric gains remained modest. In contrast, PN++ and PNx showed slight declines across most metrics, despite modest gains in wood precision (+1.91 and +1.76 pp, respectively), accompanied by modest reductions in wood recall (−2.83 and −2.51 pp). This pattern shows that these architectures produced noisier and less distinct class boundaries, where refinement tended to over-smooth ambiguous regions and remove valid wood predictions. Overall, the results indicate that post-processing refinement is most effective when the underlying model already provides spatially stable and structurally coherent predictions (Point Transformer). It should be noted that post-processing refinement remains an optional user-defined step. Based on the performance trends observed (See Figure 8), its application is not recommended for PN++ and PNx, as these frameworks already provide results comparable to PT and may experience slight reductions in overall performance following refinement.

3.3. Models Generalization Evaluation on Lin3D

To provide an intermediate assessment between within-distribution validation and fully independent external benchmarking, Lin3D was used for a spatially held-out cross-plot evaluation. One comparatively small Lin3D plot, with ground and understory removed, was included in the harmonized training and validation pool. The remaining five plots (plots 2–6) were excluded entirely from model development and used only for final evaluation. The experiment therefore assesses generalization to unseen forest plots from the same benchmark, while retaining consistency in annotation taxonomy and acquisition domain.

Across all metrics, Point Transformer achieved the highest overall performance (Table 4), with an OA of 92.35, mIoU of 85.12, MF1 of 91.93, MCC of 84.08, WF1 of 90.09, and WIoU of 81.97, consistently outperforming PN++ and PNx. Although PN++ and PNx showed comparable overall performance, both remained lower than Point Transformer across all major segmentation metrics, consistent with previous benchmark studies [27]. The largest improvements were observed in wood-specific agreement, where Point Transformer improved WF1 by 1.78–2.10% and WIoU by 2.90–3.41% relative to the other architectures.

Table 4. Performance of the evaluated DL architectures on five spatially held-out Lin3D plots excluded from model training and validation. Metrics include overall accuracy (OA), mean Intersection over Union (mIoU), macro F1-score (MF1), Matthews correlation coefficient (MCC), wood F1-score (WF1), wood Intersection over Union (WIoU), wood precision (WP), wood recall (WR), and leaf F1-score (LF1).

Model	**OA**	**mIoU**	**MF1**	**MCC**	**WF1**	**WIoU**	**WP**	**WR**	**LF1**
Point Transformer	**92.35**	**85.12**	**91.93**	**84.08**	**90.09**	**81.97**	94.11	**86.41**	**93.77**
PointNet++	91.33	83.09	90.71	82.27	88.31	79.07	96.62	81.32	93.11
PointNeXt	91.13	82.71	90.48	81.89	87.99	78.56	**96.73**	80.70	92.97

Note: All values are reported as percentages. Boldface indicates the highest-performing value for each evaluation metric.

The metric distributions further support this pattern (Figure 9), with PT showing consistently higher median values and narrower interquartile ranges across mIoU, MF1, WF1, WIoU, MCC, and OA, indicating stronger and more stable generalization and transferability. In contrast, PN++ and PNx exhibited broader variability and lower lower-bound performance, particularly in wood-sensitive metrics (WF1, WIoU, MCC), reflecting higher uncertainty in structural discrimination.

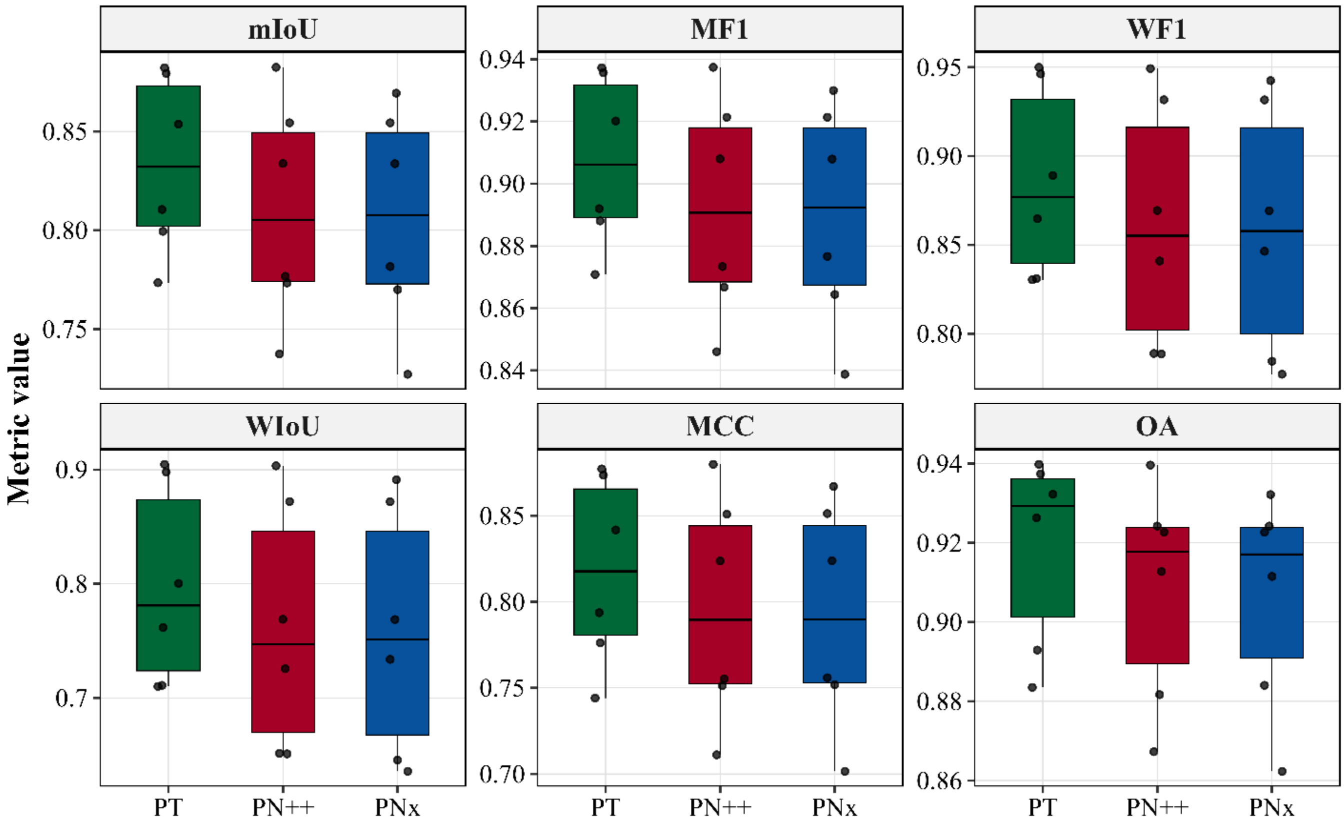

Figure 9. Distribution of segmentation metrics across five spatially held-out Lin3D plots for Point Transformer (PT), PointNet++ (PN++), and PointNeXt (PNx). Boxplots summarize mean Intersection over Union (mIoU), macro F1-score (MF1), wood F1-score (WF1), wood Intersection over Union (WIoU), Matthews correlation coefficient (MCC), and overall accuracy (OA). Black points represent individual plot-level results. Point Transformer achieved the highest median performance across most metrics, indicating stronger cross-plot generalization within the Lin3D benchmark.

https://doi.org/10.3390/rs18183158

These results confirm that the performance metrics observed during training and validation remained consistent under a spatially held-out complete Lin3D benchmark, with Point Transformer maintaining its generalization capabilities and higher accuracy [27,85].

3.4. Point Transformer Transferability Assessment of Benchmark Datasets

Table 4 summarizes the overall wood–foliar semantic performance achieved by the Point Transformer model across benchmark datasets (Section 2.3). Point Transformer was selected for detailed evaluation because it consistently outperformed PN++ and PNx during model generalization and transferability assessment (Sections 3.1 and 3.3). Consequently, the following analyses focus on evaluating Point Transformer across benchmark datasets comprised of forest plots, LeWoS Trees, and Tropical Trees (Table 2).

Table 5 shows that the Point Transformer demonstrated consistently high wood–foliar semantic segmentation performance across benchmarks despite substantial differences in forest structure, species composition, point density, acquisition geometry, and sensor characteristics (Section 2.3; Figure 5). OA remained remarkably stable, ranging from 91.63 to 94.09, while mACC and BA varied between 87.60 and 93.57. Similarly, mIoU ranged from 77.57 to 86.57, MF1 from 86.80 to 92.73, MCC from 73.65 to 85.57, WF1 from 78.82 to 90.01, WP from 76.72 to 87.11, and WR from 81.04 to 93.11, indicating a balanced trade-off between wood precision and recall across all evaluation scenarios. The close correspondence between mACC and BA indicates that wood–foliar semantic segmentation remained relatively insensitive to class imbalance despite the predominance of foliar points within the training data (Figure 4b).

Table 5. Overall dataset-level semantic segmentation performance across the evaluated benchmark datasets. Performance is summarized using Overall Accuracy (oACC), mean Accuracy (mACC), mean Intersection over Union (mIoU), Macro F1-score (MF1), and Matthews Correlation Coefficient (MCC). Higher values indicate better classification performance for all evaluation metrics.

Dataset	Total Points (M)	OA	mACC	mIoU	MF1	MCC	WF1	WP	WR
TLS plots	133.71	91.63	87.60	77.57	86.80	73.65	78.82	76.72	81.04
Tropical Tree	43.37	**94.09**	92.66	84.41	91.33	82.78	86.44	82.97	90.21
LeWoS Trees	183.89	93.75	**93.57**	**86.57**	**92.73**	**85.57**	**90.01**	**87.11**	**93.11**

Note: All values are reported as percentages. Boldface indicates the highest-performing value for each evaluation metric.

The TLS plots dataset represented the most challenging evaluation scenario because it contained the complete 3D fuel continuum, including canopy fuels, ladder fuels, understory vegetation, shrubs, regenerating seedlings and saplings, CWD, and FWD. Consequently, this dataset produced the lowest yet comparable metrics, with OA = 91.63, mACC = 87.60, mIoU = 77.57, MF1 = 86.80, MCC = 73.65, WF1 = 78.82, WP = 76.72, and WR = 81.04. In contrast, the highest performance was observed for the individual-tree datasets. LeWoS Trees achieved the strongest overall agreement across nearly all metrics (OA = 93.75, mACC = 93.57, mIoU = 86.57, MF1 = 92.73, MCC = 85.57, WF1 = 90.01, WP = 87.11, WR = 93.11), followed closely by the Tropical Tree benchmark (OA = 94.09, mACC = 92.66, mIoU = 84.41, MF1 = 91.33, MCC = 82.78, WF1 = 86.44, WP = 82.97, WR = 90.21) (Table 5).

3.5. Confusion Matrix Analysis

To further investigate the sources of classification uncertainty, the distributions of TP, TN, FP, and FN together with class-specific omission and commission errors were examined across all benchmark datasets (Figure S1). Although overall wood–foliar semantics remained consistently high (Table 5), the confusion analysis revealed important differences in how classification errors were distributed among benchmark datasets and structural conditions.

Across all datasets, TN represented the largest proportion of classified points, ranging from 62.0% in Lin3D to 76.0% in TLS plots (Figure S1). TN reflects the large proportion of foliar points present within benchmarks and indicates consistently reliable segmentation. Likewise, FP fractions remained low across all benchmarks, varying between 2.4% and 4.7%. The lowest FP fraction was observed in Lin3D, whereas TLS plots exhibited an FP fraction of 4.7%. Lin3D and LeWoS Trees exhibited the largest TP fractions, 29.8% and 28.2%, respectively, whereas TLS plots contained the smallest TP fraction (15.6%). However, Lin3D also produced the largest FN fraction (5.8%), indicating a tendency toward conservative wood classification in which uncertain points were preferentially assigned to the leaf class. In contrast, the Tropical tree and LeWoS Trees exhibited substantially lower FN fractions (2%), indicating improved recovery of wood structures.

The class-specific error metrics are shown in Figure 10 to further clarify metric trends. Wood omission decreased progressively from 0.19 in TLS plots to 0.07 in the LeWoS Trees, while wood commission declined from 0.23 to 0.13 across the same datasets. The relatively large wood-related errors observed in TLS plots indicate that most residual uncertainty originated from the classification of wood components rather than foliage. By comparison, LeWoS Trees and Tropical Trees exhibited substantially lower wood-related errors, reflecting improved visibility of wood elements. Foliar-related errors remained consistently small across all benchmark datasets (Figure 10). Leaf omission varied between 0.04 and 0.06, while leaf commission remained below 0.09 in all benchmarks. The comparatively low magnitude and limited variability of these errors indicate that foliage was identified reliably regardless of forest type, species composition, or acquisition platform.

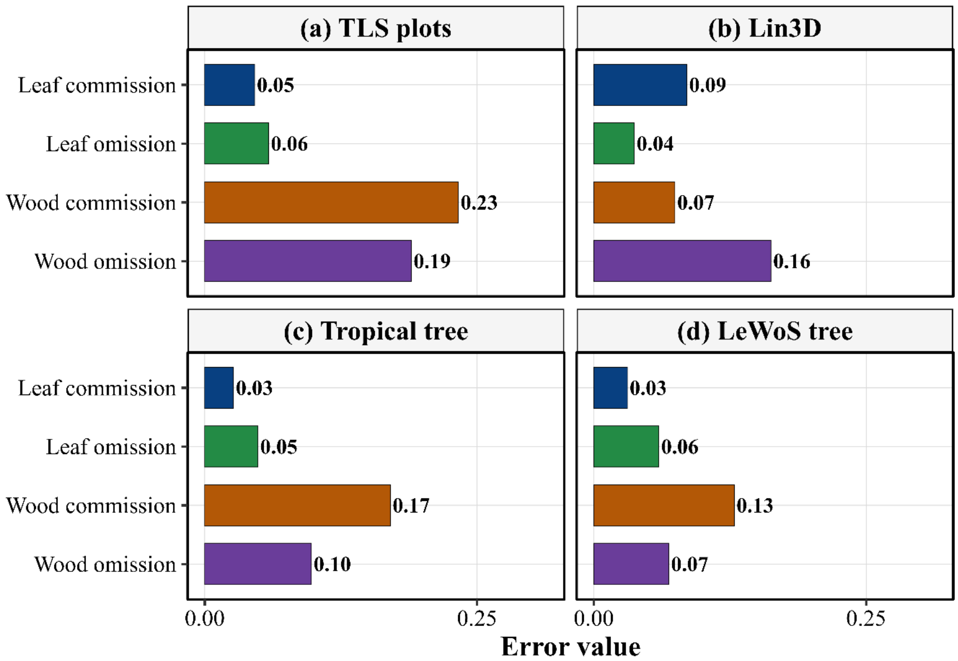

Figure 10. Class-specific error decomposition for the four benchmark datasets. Wood and leaf omission and commission errors are shown for (**a**) TLS plots; (**b**) Lin3D; (**c**) Tropical tree; and (**d**) LeWoS Trees. The profiles summarize the distribution of segmentation errors and facilitate comparison of class-specific uncertainty among benchmark datasets.

3.6. Quantitative Analysis Across Forest Vertical Strata

Section 3.4 demonstrated overall Point Transformer wood–foliar semantic segmentation performance, which is a widely used technique in DL framework evaluations. Nevertheless, aggregate statistics do not reveal where within the canopy profile these errors occurred. Given that wood–foliar semantics exhibit complex vertical stratification in forest ecosystems, segmentation performance was further evaluated as a function of relative canopy height of plots and individual tree benchmarks.

Vertical performance profiles revealed consistent height-dependent trends across all benchmark datasets (Figure 11). Segmentation accuracy generally increased from lower canopy layers toward the mid-canopy region, where maximum performance was observed, before progressively declining toward the overstory. The Tropical tree and LeWoS Trees datasets exhibited the most stable vertical responses, maintaining mIoU, WF1, WP, and WR values above approximately 0.80 throughout much of the canopy profile (Figure 11c,d). TLS plots displayed a similar pattern but with a more pronounced reduction in performance above approximately 0.80 relative height (Figure 11a), whereas Lin3D maintained high WP throughout most canopy layers but exhibited a marked decline in WR within overstory regions (Figure 11b).

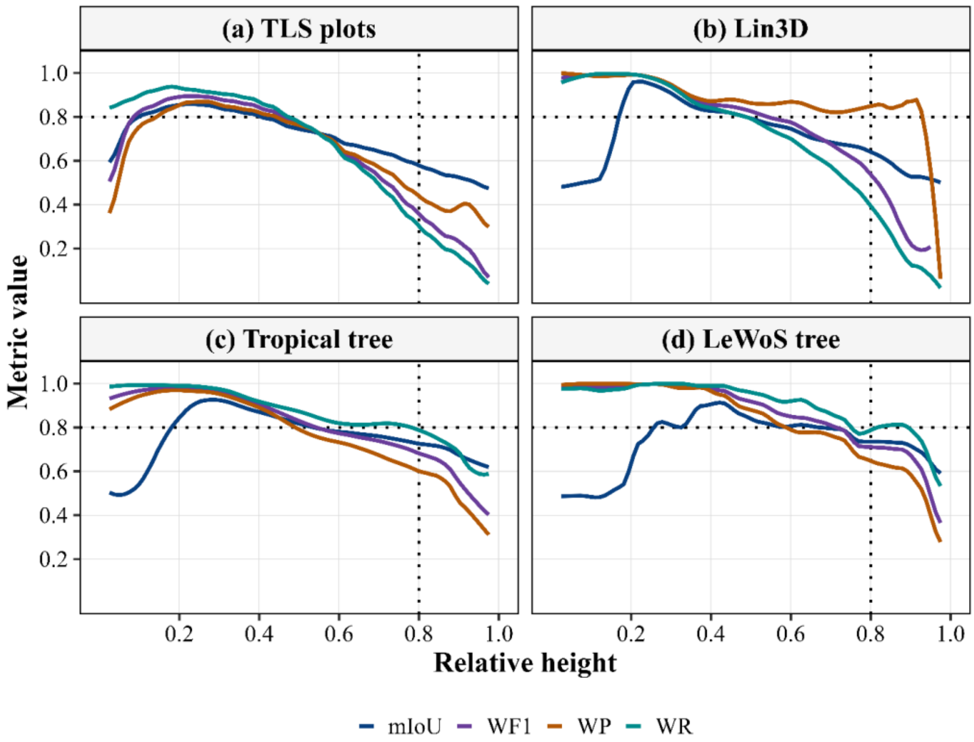

Figure 11. Vertical performance profiles of Point Transformer across the four benchmark datasets. mIoU, WF1, WP, and WR are plotted as a function of relative height for (**a**) TLS plots; (**b**) Lin3D; (**c**) Tropical tree; and (**d**) LeWoS Trees. Dashed reference lines indicate a metric value of 0.80 and a relative height of 0.80. The profiles illustrate height-dependent variation in wood segmentation performance throughout the canopy.

Point Transformer preserved the vertical distribution of wood throughout the 3D fuel continuum (Figure 12). The background shading represents the relative abundance of wood (brown) and leaf (green) points within each benchmark dataset, providing ecological context for interpreting prediction behavior across contrasting forest structures. The reference woody profile represents the benchmark-derived vertical distribution of woody points, whereas the predicted woody profile denotes the corresponding distribution predicted by Point Transformer. The woody agreement profile quantifies the proportion of points simultaneously classified as woody by both the benchmark and Point Transformer, while the predicted woody where the reference leaf identifies benchmark-labelled leaf points that were assigned woody labels by Point Transformer.

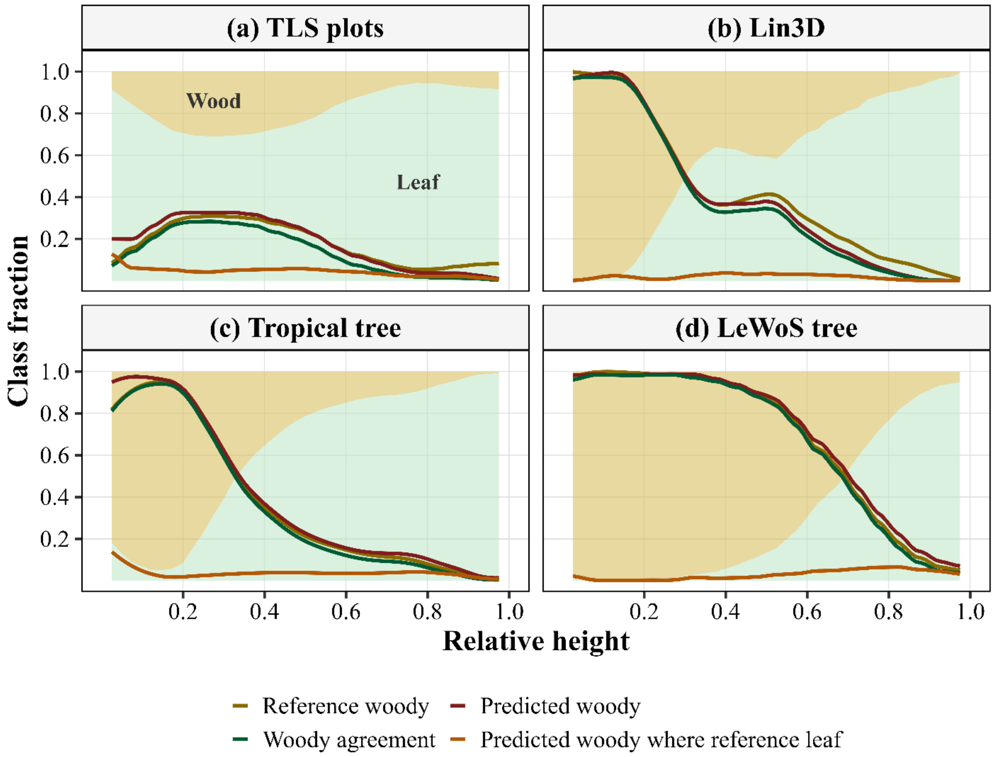

Figure 12. Vertical wood distribution and agreement profiles across the four benchmark datasets. The shaded background represents the relative abundance of wood and leaf points along the normalized canopy profile, while the curves depict the reference wood fraction (brown), predicted wood fraction (dark red), wood agreement (green), and predicted wood where reference leaf (orange) fraction. Results are shown for four benchmark datasets: (**a**) TLS plots; (**b**) Lin3D; (**c**) Tropical Tree; and (**d**) LeWoS Trees.

Substantial differences in vertical wood–foliar semantic composition were evident among the benchmark datasets (Figure 12). TLS plots exhibited the strongest leaf dominance, with foliage comprising most canopy layers (Figure 12a). In contrast, Lin3D, Tropical Tree, and LeWoS contained substantially larger woody fractions throughout much of their vertical domains, reflecting the predominance of stems and branches and the reduced influence of understory vegetation and surface fuels (Figure 12b–d). Across all benchmarks, Point Transformer closely reproduced the dominant vertical distribution of woody material. Predicted woody fractions closely followed the benchmark-derived profiles, while woody agreement remained high throughout most height intervals. The strongest correspondence was observed for the Tropical Tree and LeWoS benchmarks, where predicted and reference woody profiles were nearly coincident across extensive portions of the canopy (Figure 12c,d). Likewise, the major transitions between wood-dominated and leaf-dominated strata were consistently preserved across all datasets, demonstrating that Point Transformer effectively reconstructed the overall vertical organization of forest fuels despite substantial differences in canopy architecture and structural complexity.

For the predicted woody component, the reference leaf profile provides additional insight into the origin of prediction disagreement. Across all benchmarks, this component remained comparatively small throughout most height intervals, indicating that woody predictions assigned to benchmark-labelled leaf regions constituted only a minor fraction of the total point distribution. Localized departures from this near-zero baseline were primarily confined to canopy transition zones, particularly within TLS plots, the Tropical Tree, and parts of the Lin3D benchmarks, where predicted woody fractions exceeded the reference profile.

3.7. Species-Specific Quantitative Analysis

Species-specific evaluation revealed consistently balanced wood–foliar semantics across six benchmark tree species/genera, although measurable differences were observed among species (Figure 13). OA remained remarkably stable, ranging from 0.91 to 0.95, while MF1 varied between 0.84 and 0.91. Similarly, mIoU ranged from 0.74 to 0.84, indicating relatively consistent performance despite substantial variation in species composition, canopy structure, and wood–foliar distributions. The uniformly high OA and LF1 values across all species further demonstrate that the Point Transformer generalized effectively across diverse temperate forest conditions.

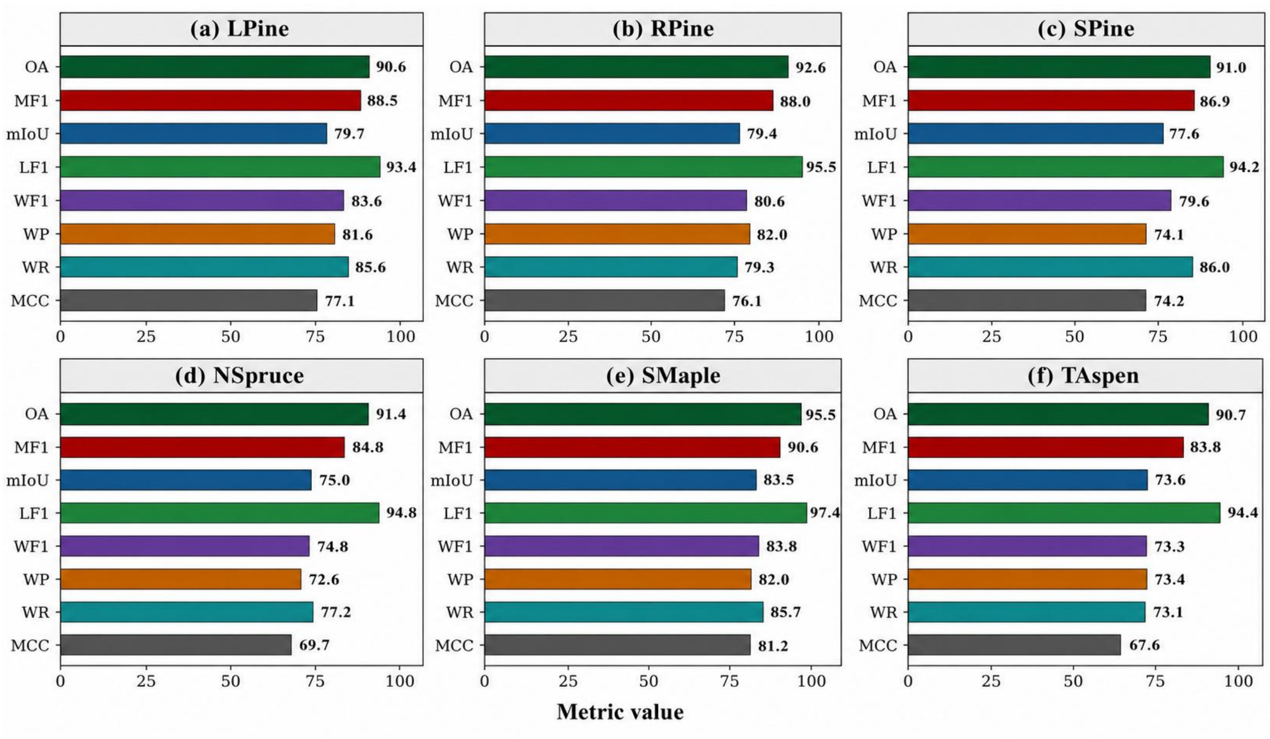

Figure 13. Species-level semantic segmentation performance of the Point Transformer across six TLS benchmark species. Overall Accuracy (OA), Macro F1-score (MF1), mean Intersection over Union (mIoU), Leaf F1-score (LF1), Wood F1-score (WF1), Wood Precision (WP), Wood Recall (WR), and Matthews Correlation Coefficient (MCC) are shown for (**a**) LPine; (**b**) RPine; (**c**) Spine; (**d**) NSpruce, (**e**) SMaple; and (**f**) TAspen, illustrating species-specific variation in semantic segmentation performance across contrasting canopy architectures, branching structures, and wood–foliar structural complexity.

Among the evaluated species, SMaple achieved the highest overall performance, with OA, MF1, and mIoU values of 95.5, 90.6, and 83.5, respectively (Figure 13e). SMaple also produced the highest LF1 value (97.4) while maintaining strong wood-specific performance (WF1 = 83.8, WP = 82.0, WR = 85.7) and the highest MCC (81.2) among all species. LPine exhibited comparable performance, achieving OA of 90.6, mIoU of 79.7, WF1 of 83.6, and MCC of 77.1 (Figure 13a). RPine and SPine similarly maintained high overall performance, with OA values of 92.6 and 91.0, mIoU values of 79.4 and 77.6, and MCC values of 76.1 and 74.2, respectively (Figure 13b,c). These results indicate that Point

Transformer remained highly effective across multiple conifer species despite differences in canopy organization and branching patterns [27].

Greater variability was observed within wood-specific metrics than in overall accuracy metrics. NSpruce and TAspen produced the lowest mIoU values (75.0 and 73.6, respectively) and correspondingly lower wood-specific performance (Figure 13d,f). For NSpruce, both WP (72.6) and WR (77.3) were reduced relative to the other species, resulting in a WF1 of 74.9 and an MCC of 69.8. TAspen exhibited the lowest overall wood performance, with WP, WR, and WF1 declining to 73.4, 73.1, and 73.3, respectively, together with the lowest MCC (67.6). In contrast, LPine, SPine, and SMaple maintained comparatively high WR values (85.6, 86.0, and 85.7, respectively), indicating strong recovery of wood structures. Notably, SPine exhibited relatively high WR (86.0) but lower WP (74.1), suggesting that wood structures were frequently detected but accompanied by increased commission errors. Conversely, RPine showed slightly higher WP (82.0) than WR (79.3), indicating a more conservative classification pattern.

3.8. Quantitative Analysis Across Species Vertical Stratum

To further assess framework robustness across contrasting canopy architectures, species-specific vertical performance profiles and wood agreement distributions were evaluated for all six TLS benchmark species (Figure 14). Across species, performance showed a consistent height-dependent pattern, increasing rapidly from lower canopy regions, peaking within lower- to mid-canopy strata, and declining toward canopy tops. WF1, WP, and WR generally exceeded 0.80 across most of the canopy profile, indicating robust wood–leaf separation across diverse crown architectures. Similar to the overall benchmark-level analysis (Figure 11), the largest performance reductions occurred in upper-canopy regions where fine branches and dense foliage increased structural ambiguity.

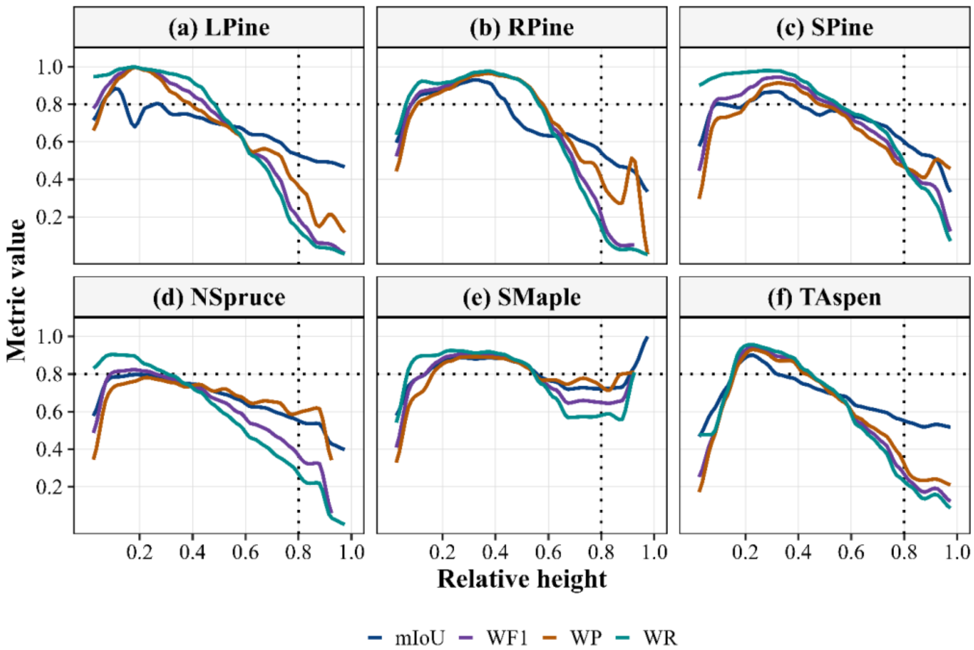

Figure 14. Species-specific vertical segmentation performance profiles for the six TLS benchmark species. Curves show mIoU, WF1, WP, and WR as a function of relative tree height for (**a**) LPine, (**b**) RPine, (**c**) SPine, (**d**) NSpruce, (**e**) SMaple, and (**f**) TAspen. The horizontal dashed line indicates a metric value of 0.80, and the vertical dashed line marks 0.80 relative height. The figure illustrates species-level variation in segmentation performance throughout the canopy profile and highlights the increasing difficulty of detecting fine wood structures within upper-crown regions.

Species-level differences were primarily expressed in the magnitude of upper-canopy degradation. SMaple maintained the most stable performance throughout the canopy profile (Figure 14e), whereas LPine, RPine, and SPine showed strong lower- and mid-canopy agreement followed by moderate declines above ~0.75 relative height (Figure 14a–c). In contrast, NSpruce and TAspen exhibited the strongest upper-canopy reductions, particularly in wood recall above ~0.70–0.80 relative height (Figure 14d,f). Across all species, the divergence between wood precision and recall increased toward canopy tops, indicating that performance reductions were primarily driven by reduced recovery of fine wood components rather than increased commission errors. Overall, these results demonstrate strong cross-species transferability, with most performance differences confined to structurally complex upper-canopy environments.

The genus-specific wood agreement profiles further evaluated the ability of Point Transformer to preserve the vertical distribution of wood material across contrasting canopy architectures (Figure 15). Consistent with the benchmark-level analysis (Figure 12), the predicted wood profiles closely followed the reference wood distributions across all species, indicating that the dominant vertical organization of wood material was largely preserved. The strongest agreement generally occurred within canopy regions containing the highest wood abundance, where the reference, predicted, and agreement profiles remained tightly aligned. Species differed primarily in the shape and vertical extent of their wood distributions. LPine and RPine showed pronounced wood peaks concentrated in lower- to mid-canopy strata, whereas SPine and TAspen exhibited broader wood distributions extending across larger portions of the canopy profile (Figure 15a–c,f). In contrast, NSpruce and SMaple maintained comparatively lower wood fractions throughout most height intervals, although the agreement between prediction and reference remained consistently high (Figure 15d,e).

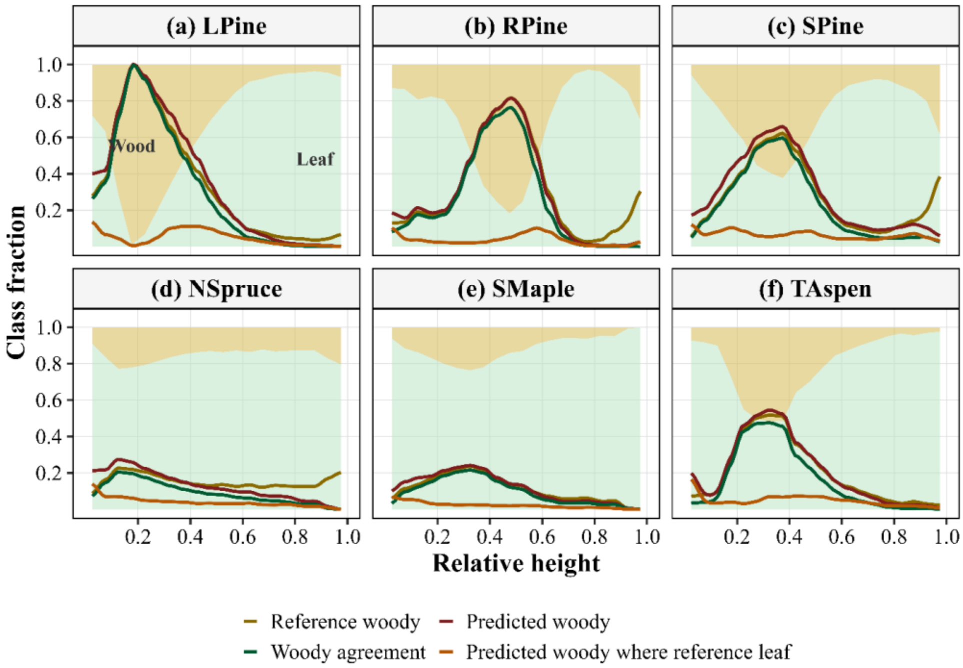

Figure 15. Species-specific vertical wood agreement profiles for the six TLS benchmark species. Shaded backgrounds indicate the benchmark-derived vertical distributions of wood and leaf points. Curves show the reference wood fraction, predicted wood fraction, wood agreement, and predicted wood where reference leaf as a function of relative tree height for (**a**) LPine, (**b**) RPine, (**c**) SPine, (**d**) NSpruce, (**e**) SMaple, and (**f**) TAspen. The figure illustrates the ability of Point Transformer to preserve species-specific vertical wood structure and identifies canopy regions where disagreement between predictions and benchmark annotations is concentrated.

Disagreement profiles remained comparatively small across all species and were largely confined to localized canopy regions rather than distributed systematically throughout the profile. These departures were most evident where predicted wood fractions exceeded reference wood fractions while maintaining strong overall agreement, suggesting that a portion of the mismatch may reflect localized benchmark labeling uncertainty or conservative annotation of fine wood components rather than persistent model overprediction.

3.9. Cross-Dataset Transferability and Structural Validation

The vertical performance profiles (Figure 11) are strongly supported by the qualitative assessment of benchmark vs. Point Transformer wood–foliar semantics (Figure 16), revealing how segmentation behavior changes along the canopy profile and across structural complexity gradients.

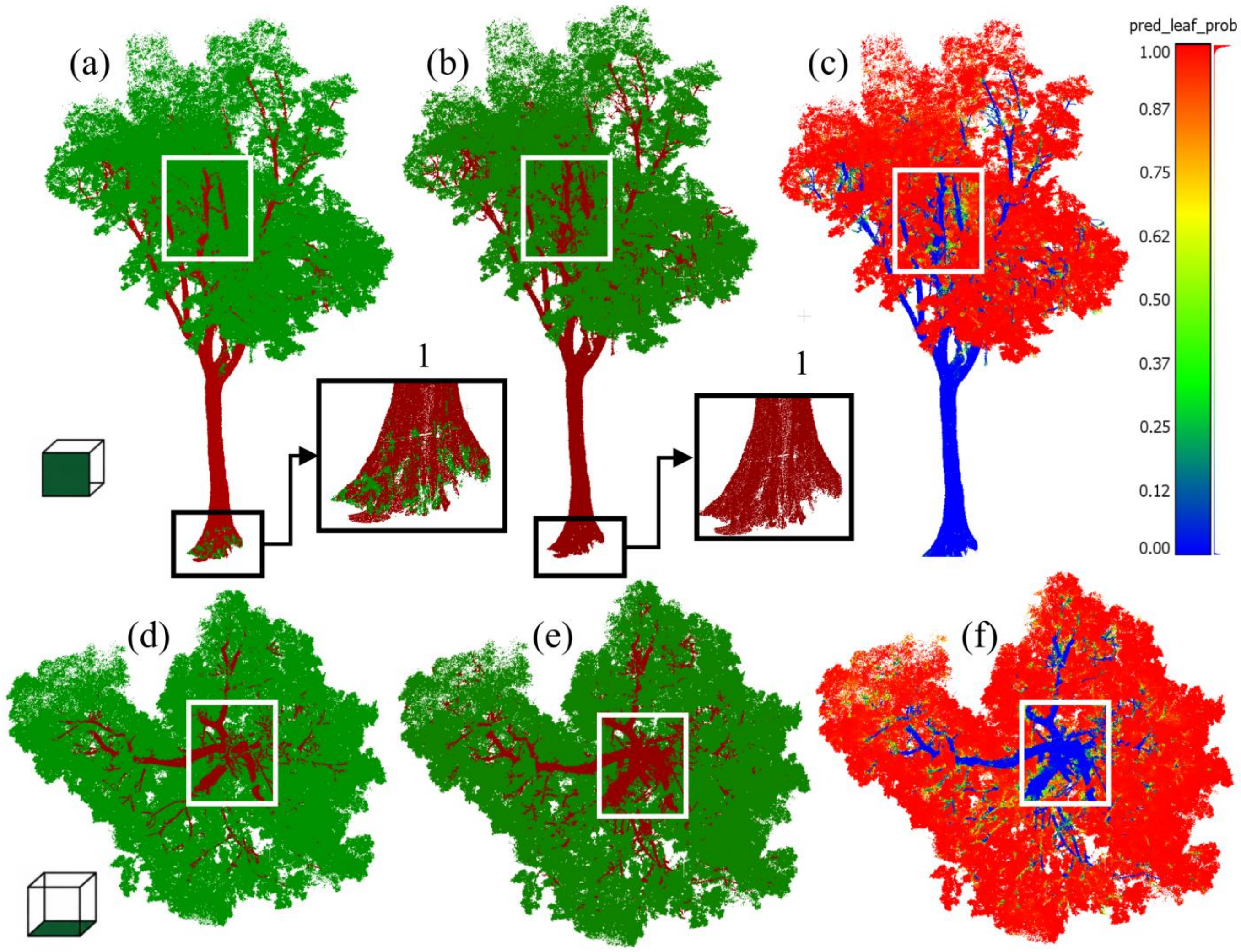

Figure 16. Qualitative assessment of Point Transformer predictions on a representative tree from the Tropical Tree benchmark dataset. Panels (**a**,**d**) show the benchmark wood–leaf annotations from lateral and bottom-up views, respectively; (**b**,**e**) present the corresponding Point Transformer predictions, and (**c**,**f**) show the predicted leaf probability maps (0–1). Insets highlight representative stem and crown regions for detailed visual comparison. Wood is shown in brown, foliage in green, and prediction confidence is represented by the probability color scale.

In the tropical benchmark (Figure 16), Point Transformer maintained consistently high wood recall (WR > 0.80) through most of the canopy (Figure 11c), which is visually evident from the strong preservation of internal branch architecture and continuous wood

https://doi.org/10.3390/rs18183158

skeletons in both lateral and bottom-up views (Figure 16b,e). The probability maps further show high confidence across crown interiors, indicating that most wood elements, including fine branches, were successfully retained (Figure 16c,f). However, the moderate decline in wood WP and WF1 in the upper crown corresponds to regions where the model recovered additional fine wood structures that appear absent or incompletely annotated in the benchmark reference, suggesting that part of the apparent error may reflect annotation incompleteness (see white square boxes in Figure 16a–f) rather than model inability to segment the wood and leaf in particular, performance differences with lower validation metrics in extreme lower i.e., surface wood and top of the canopy, are indicative of human labelling in these regions being found to be challenging and the model performed better than labelling efforts. A similar trend is observed in the LeWoS benchmark (Figure S2), where Point Transformer reproduced the dominant trunk and branch framework with strong structural continuity [27].

The vertical performance profiles of the Lin3D benchmark (Figure 11b) are strongly supported by the qualitative comparison between benchmark and Point Transformer semantics (Figure 17), revealing how segmentation behavior changes across canopy strata and structural complexity gradients. Point Transformer maintained consistently high wood recall (WR > 0.80) and wood precision (WP > 0.80) throughout most of the vertical profile, while wood F1 and mIoU remained above 0.80 over a large portion of the canopy (Figure 11b). This strong agreement is qualitatively evident, where the dominant stems, major branches, and canopy foliage are consistently preserved (Figure 17a,b).

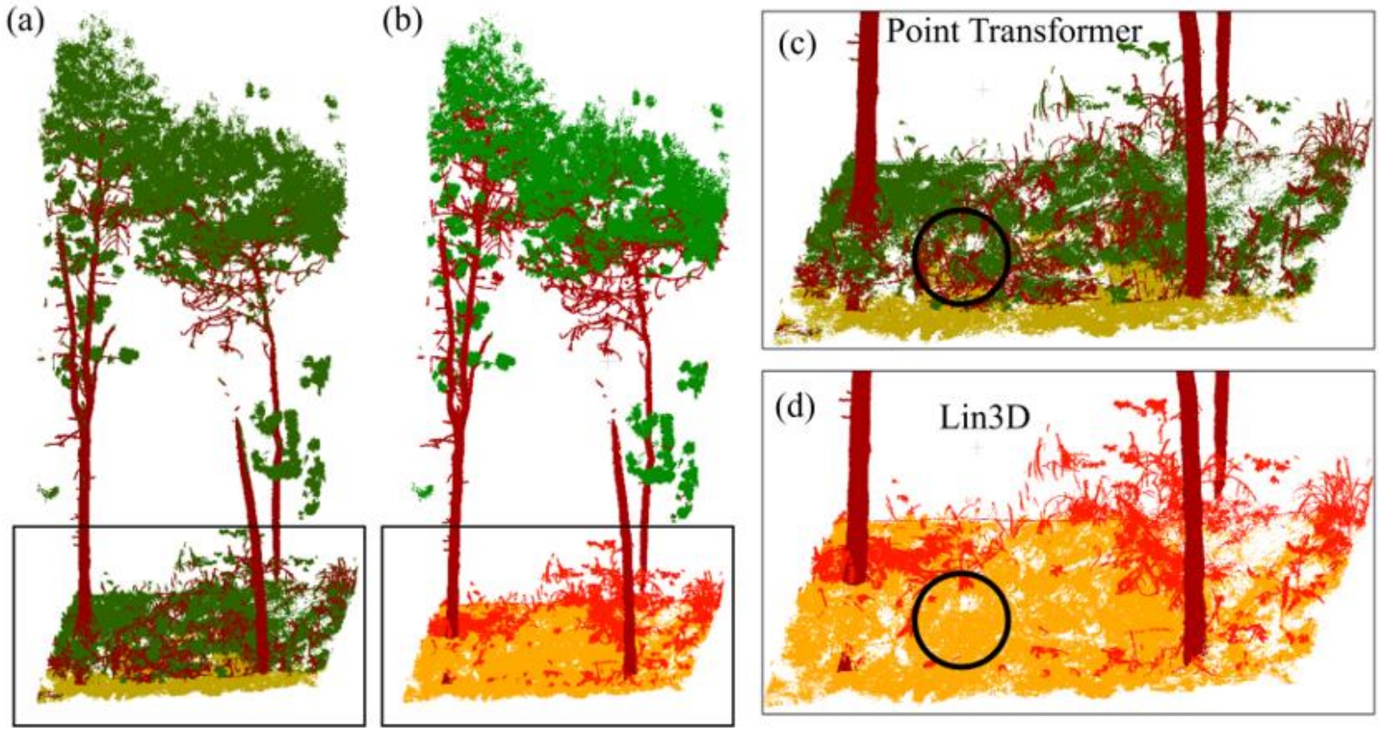

Figure 17. Representative examples of semantic segmentation discrepancies between reference labels and model predictions in complex forest scenes. Panels (**a**,**b**) show the same plot-level Lin3D point cloud from two perspectives, highlighting differences in the assignment of wood and foliar components, particularly surface and understory regions. Panels (**c**,**d**) illustrate plot-level examples where understory vegetation and near-ground wood structures exhibit substantial labeling ambiguity compared to Point Transformer (black circle).

In contrast, performance degradation was concentrated within the lower understory and upper-canopy transition zones. The reduction in mIoU, WF1, and WR above approximately 80% relative height (Figure 11b) coincides with highly fragmented branch networks and dense crown interfaces, where fine woody structures become increasingly sparse. Similarly, lower agreement within the near-ground region corresponds to understory vegetation, woody shrubs, and surface fuels, where semantic boundaries between wood, foliage, and lower objects become inherently ambiguous. These patterns are clearly illustrated in Figure 17c,d, where Point Transformer consistently identifies woody structures within regions annotated as lower objects or foliage in the benchmark dataset (black circles). Importantly, the localized disagreement observed in Figure 15 is not uniformly distributed throughout the canopy but is concentrated within structurally complex transition zones. The simultaneous decline in wood precision and recall in these regions suggests that part of the apparent segmentation error originates from semantic ambiguity and annotation uncertainty rather than systematic model failure.

3.10. Qualitative Assessment of Transferability Across Heterogeneous Forest Ecosystems

To further evaluate the qualitative performance of the proposed framework, the Point Transformer was applied to a large registered multi-scan TLS dataset acquired in an Australian eucalypt forest (Figure 18).

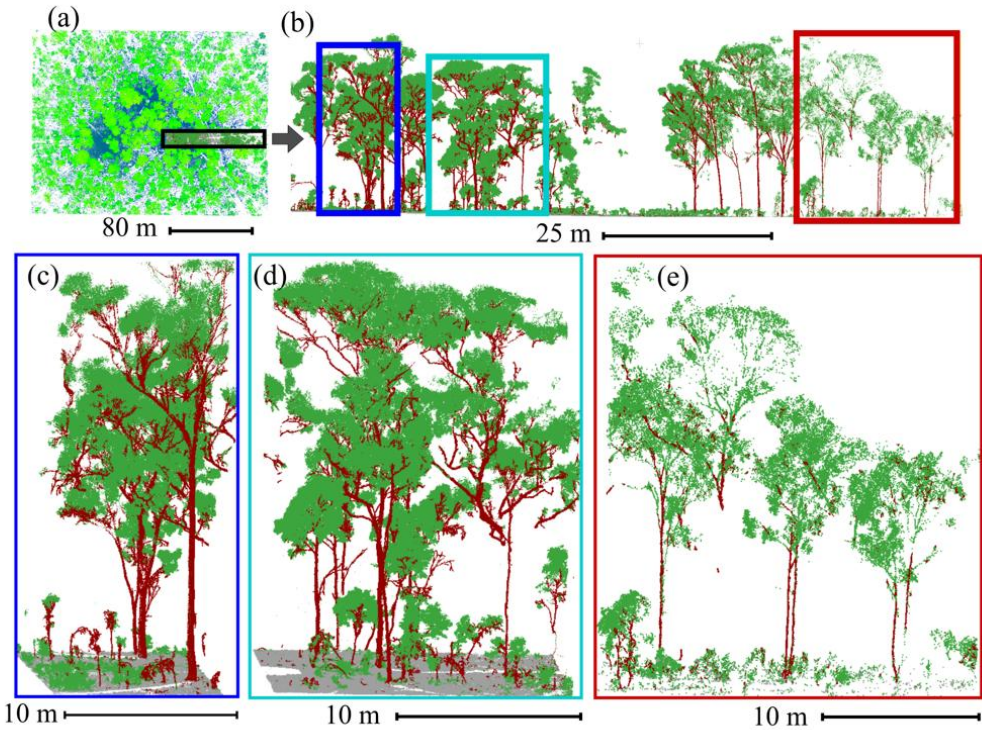

Figure 18. Point Transformer semantic segmentation of a large, registered multi-scan TLS dataset acquired in an Australian eucalypt forest. (**a**) Top-down view of the complete 180 × 160 m plot; the black rectangle identifies the transect displayed in (**b**). (**b**) Side view of the selected transect, with colored bounding boxes indicating the regions enlarged in (**c**–**e**). These regions extend from the relatively high-density plot center toward the lower-density outer edge, illustrating segmentation performance under spatially variable sampling density. Wood and foliage are shown in red and green, respectively.

The complete 180 × 160 m plot (Figure 18a) demonstrates successful semantic segmentation over a large forest scene, while the extracted transect (Figure 18b) provides a continuous side view extending from the plot center toward the outer boundary. Enlarged examples from the high-density central region (Figure 18c), intermediate region (Figure 18d), and lower-density outer edge (Figure 18e) show that the model consistently preserved the principal stems, branches, and foliar components throughout the plot.

Although fine-scale structural detail gradually decreased toward the plot boundary where point density was substantially reduced, the overall wood–leaf semantics remained spatially coherent across the registered multi-scan dataset, demonstrating stable performance under substantial spatial variations in sampling density.

Qualitative comparison with the BlueCat dataset further illustrates systematic differences between the manually annotated labels and the Point Transformer semantics (Figure 19).

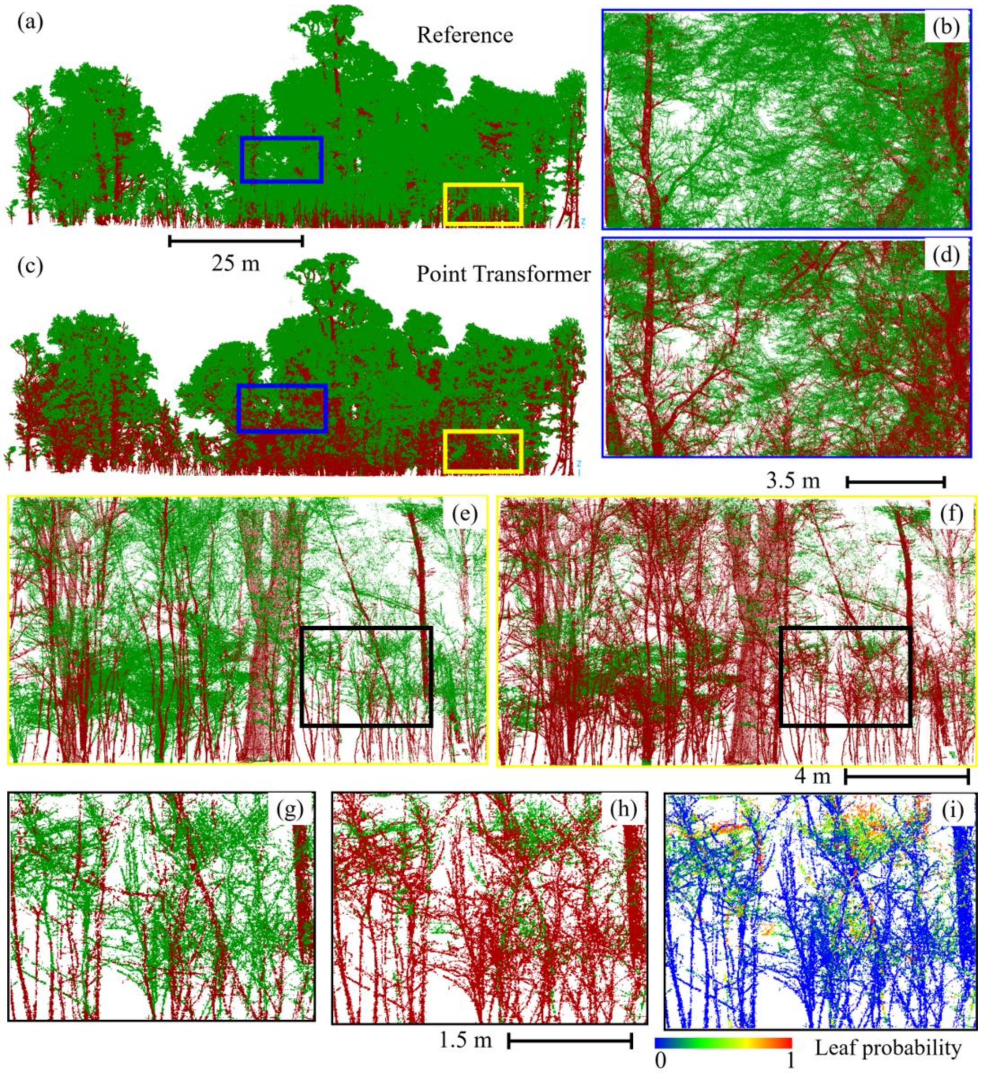

Figure 19. Comparison of manually annotated reference labels and Point Transformer semantic segmentation for the BlueCat dataset. Panels (**a**) and (**c**) show the reference labels and corresponding Point Transformer predictions, respectively. Blue and yellow boxes indicate representative regions enlarged in (**b**–**f**). The black boxes in (**e**,**f**) are further enlarged in (**g**–**i**), where (**g**,**h**) show the reference labels and Point Transformer predictions, respectively. Panel (**i**) presents the corresponding Point Transformer leaf-probability map, demonstrating confidence estimates that are consistent with the predicted semantic labels in (**h**).

Relative to the reference labels (Figure 19a), the Point Transformer identified substantially more woody points throughout the forest scene, particularly within tree crowns, fine branches, and understory vegetation (Figure 19c). Enlarged canopy regions (Figure 19b,d) show that many fine branches labeled as foliage in the reference were predicted as wood by the Point Transformer, resulting in a denser and more continuous woody representation. Similar differences are evident within the understory (Figure 19e,f), where a substantially greater number of fine-scale branches of seedlings and saplings were detected by Point Transformer (Figure 19e) compared with benchmark annotations (Figure 19f). Further enlargement (Figure 19g,h) highlights these differences at the individual branch level. The corresponding leaf-probability map (Figure 19i) demonstrates that the additional woody predictions are accompanied by consistently low leaf probabilities, indicating that these predictions were produced with high model confidence.

The Wytham Woods (UK) TLS dataset (Figure 20) provided an additional qualitative assessment under a leaf-off condition [36]. Leaf-off phenology exposed densely interconnected wood architecture while substantially reducing foliar returns, resulting in a strongly wood-dominated class distribution that has long been recognized as challenging for LiDAR-based forest segmentation because intertwined branches increase structural ambiguity and reduce separability [86].

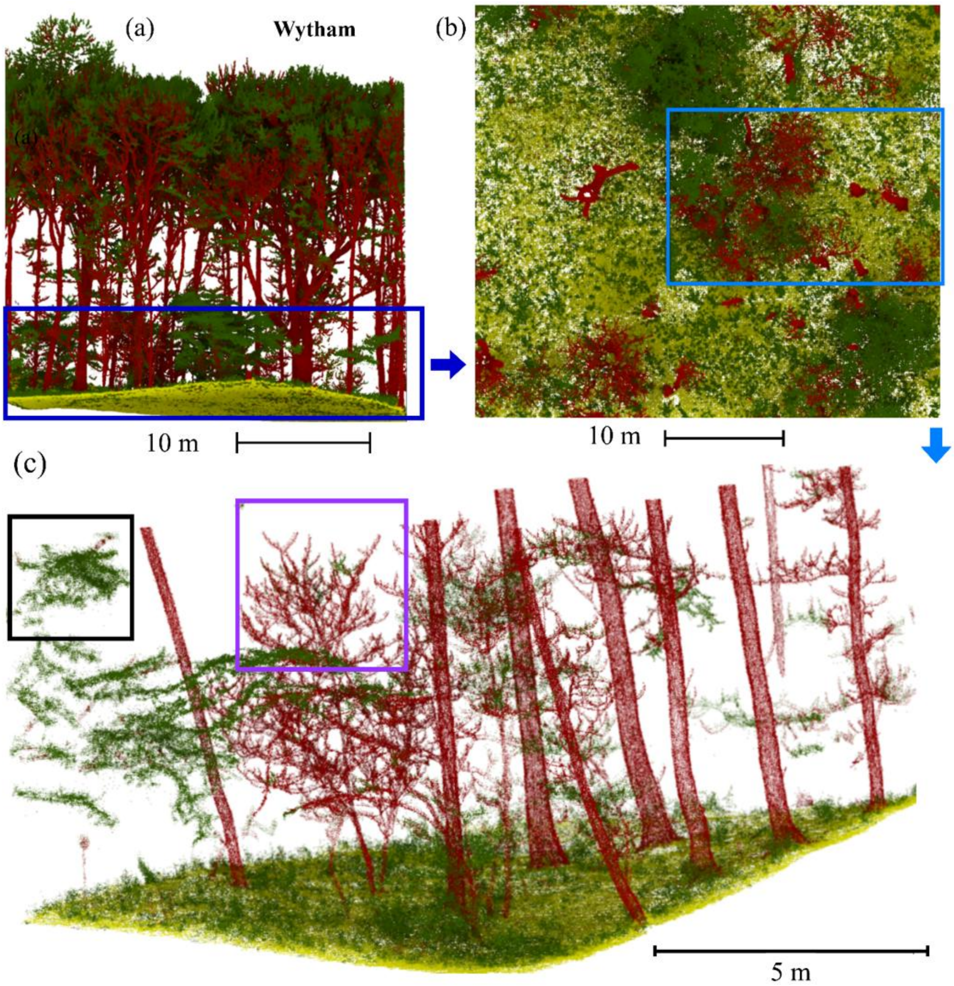

Figure 20. Qualitative assessment of the proposed Point Transformer framework on the independent leaf-off Wytham Woods (UK) terrestrial LiDAR dataset. (**a**) Oblique view of the predicted wood–foliar semantics. (**b**) Top-down view showing the region selected for detailed inspection. (**c**) Enlarged view illustrating semantic predictions under leaf-off conditions, where extensive leaf abscission resulted in a strongly woody-dominated forest structure. This independent dataset demonstrates the robustness of the proposed framework under contrasting seasonal class distributions.

Despite the substantial structural variation and pronounced class imbalance, with woody components dominating the Wytham Woods (Figure 20a) in contrast to the foliar-dominated training dataset (Figure 4a,b), Point Transformer consistently identified stems, major branches, and fine woody structures while simultaneously preserving the limited remaining foliar semantic class (Figure 20a,b). This result indicates that the learned geometric representations remained robust under contrasting seasonal and phenological conditions. The enlarged views further demonstrate coherent delineation of intricate branch networks and clear separation between wood–foliar semantics, with minimal evidence of semantic confusion despite the sparse leaf cover (Figure 20c). In particular, Point Transformer successfully recovered fine branches and crown connectivity that remained visually distinct from the residual foliage, as shown for understory leaf-on and leaf-off seedlings and saplings (see square boxes in Figure 20c). This shows that model transferability was not compromised by the substantial seasonal shift in canopy composition and class proportions on which the Point Transformer was trained (Figure 4). The Wytham Woods dataset was completely excluded from both training and validation, providing an independent qualitative assessment of model transferability under leaf-off conditions. The Wytham Woods benchmark has been extensively evaluated for instance segmentation and has largely remained unexplored for semantic segmentation [36,77,78].

Owing to space limitations, additional qualitative transferability examples are provided in the Supplementary Material (Figures S3 and S4). These include independent inference on the Methow, Robson Creek, Litchfield (Figure 1), wood–foliar predictions for Plot 2, a natural *Fagus sylvatica* forest in Asturias, Spain, acquired using a GeoSLAM ZEB Horizon MLS system (Figure S3), and the NIBIO MLS dataset (Figure S4), thereby extending the evaluation across contrasting forest ecosystems, stand structures, and terrestrial LiDAR systems (Table 3). Particularly, the SegmentedForests Plot 2 example (Figure S3) demonstrates that Point Transformer preserves the complex wood–foliar semantics while recovering visually continuous fine branches that are annotated as foliar components in the benchmark. Similarly, the NIBIO MLS (Figure S4) probability exhibits spatially coherent confidence patterns with disagreement largely confined to structurally ambiguous wood–foliar transition zones, further supporting the transferability of the proposed geometry-only framework across independent datasets and MLS acquisitions.

3.11. 3D Fuels Wildfire Fuel Mapping

The Point Transformer wood–foliar semantics were subsequently processed using a geometric post-processing framework, termed "Points2woodyseg", to derive wildland fuel-oriented classification (Figure 21). Points2woodyseg—the algorithm operates on the structurally simplified wood and ground point clouds (Figure 21a), where the wood points are first normalized relative to the ground and then used for stem detection. Once stem structures are identified, the algorithm uses the geometric relationships between ground and stems as reference constraints to further decompose the remaining wood points into ecologically relevant classes such as stem, branches, leaf, surface wood debris, and surface foliage (Figure 21b). This hierarchical strategy leverages the stable geometric properties of wood structures as a foundation for downstream structural decomposition, thereby reducing semantic complexity and providing an alternative pathway to conventional end-to-end multi-class DL segmentation (Figure 21b).

Points2WoodySeg is presented only to demonstrate the downstream applicability of the proposed DL wood–foliar semantic segmentation framework; its methodology is beyond the scope of the present study. However, to examine the applicability of Points2WoodySeg across different forest types and structural complexities, several plots from the benchmark datasets were processed, with a representative example from the SegmentedForest [60] plot 13, presented in Appendix A (Figure A1).

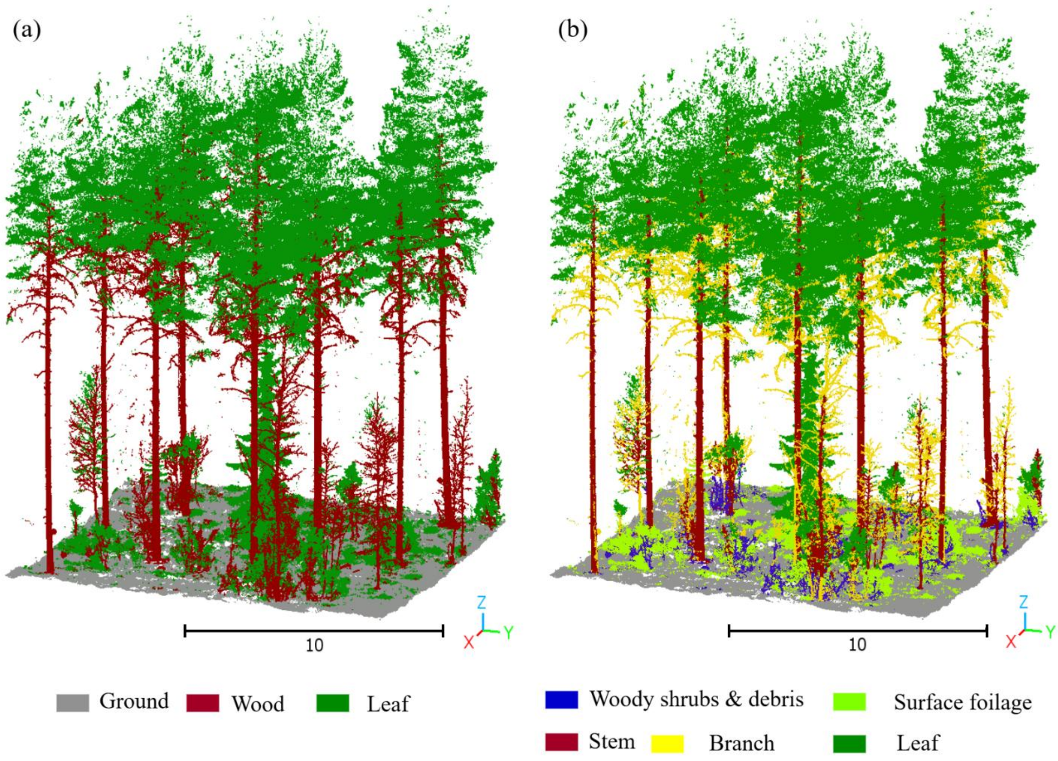

Figure 21. Example of the proposed fuel characterization workflow. (**a**) Wood–foliar semantics produced by Point Transformer, with ground (gray) classified by FAST-GC, wood (red), and foliage (green). (**b**) Geometric decomposition of the binary wood–foliar predictions using Points2WoodySeg into stems (red), branches (yellow), canopy foliage (green), surface foliage (light green), and woody shrubs and woody debris (blue). The multi-class decomposition presented in (**b**) is useful for next-generation fire behavior models [87].

Multi-Class Fuel Decomposition Benchmark Assessment

To further evaluate whether the wood–foliar semantics retained sufficient structural fidelity for ecologically meaningful fuel decomposition, the wood predictions generated by Point Transformer (Figure 21a) were subsequently processed using the Points2woodyseg geometric algorithm (Figure 21b) and evaluated against an independently annotated ForestSemantic [76] benchmark (Table 2) containing stem, branch, and leaf classes.

The vertical class profiles revealed strong overall correspondence between predicted and reference structural distributions throughout the canopy following Point Transformer-based wood–leaf segmentation and subsequent geometric decomposition (Figure 22a,b). Predicted class fractions closely followed the benchmark distributions across both wood and foliar semantics. The strongest agreement occurred within the lower and middle canopy, where stems and major branch structures dominate, whereas all class fractions progressively declined toward the upper crown where foliage became increasingly dominant. These decomposition patterns remain consistent with the wood–foliar semantics segmentation performance observed across the benchmark evaluations in Sections 3.5–3.7.

Stem segmentation achieved the highest overall performance, with P = 0.96, R = 0.99, F1 = 0.98, and IoU = 0.96, demonstrating near-complete detection of stems by the geometric framework (Points2woodyseg). Leaf segmentation also remained robust, with P = 0.91, R = 0.93, F1 = 0.92, and IoU = 0.85. In contrast, branch segmentation showed substantially lower agreement (P = 0.54, R = 0.51, F1 = 0.52, IoU = 0.36), indicating that branch-level lower accuracy is associated with the combined effect of uncertainty in benchmark annotations (Figure A2a) and Point Transformer wood–foliar semantics (Figure A2b).

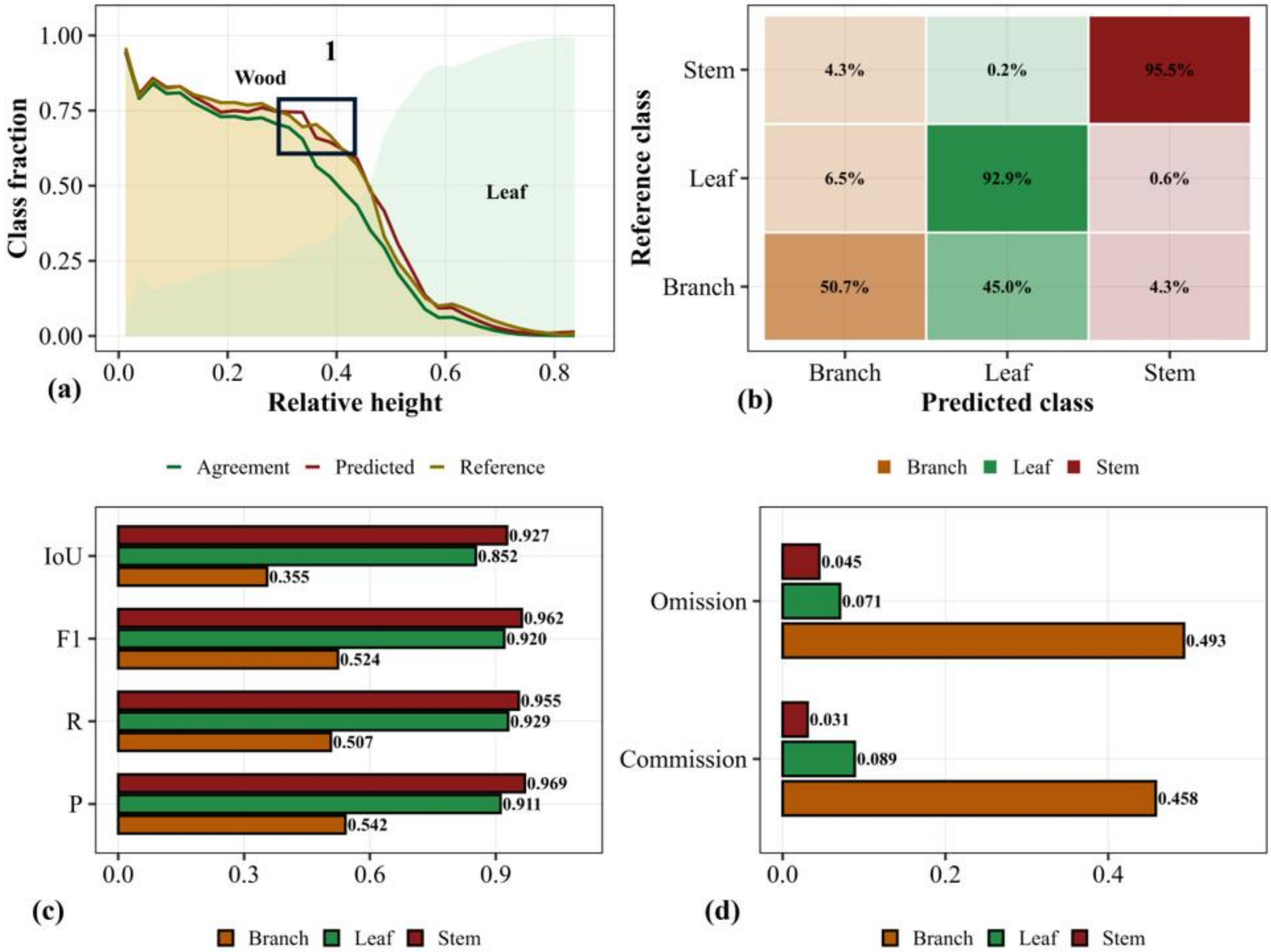

Figure 22. Performance assessment of the Points2WoodySeg structural decomposition framework for stem, branch, and leaf segmentation derived from Point Transformer wood–leaf predictions (Figure 21a). (**a**) Vertical distribution of reference, predicted, and agreeing class fractions along normalized canopy height of Point Transformer wood–foliar semantics (Figure 21a). The boxed region (1) highlights localized branch–leaf disagreement within the mid-canopy transition zone (see representative examples in Figure A2, Appendix A). (**b**) Normalized confusion matrix between predicted and reference structural classes. (**c**) Class-specific Precision (P), Recall (R), F1-score (F1), and Intersection over Union (IoU). (**d**) Omission and commission errors for each structural class decomposition using the Points2woodyseg algorithm.

The omission and commission analysis further supports this pattern (Figure 22c,d). Importantly, the near-symmetrical branch omission (0.49) and commission (0.46) errors, together with the localized vertical offsets between predicted and reference branch fractions in the mid-crown transition zone (black-box region in Figure 22a), indicate that the total branch point budget remained broadly comparable but differed in its precise spatial allocation. Collectively, these results indicate that branch disagreement largely reflects fine-scale spatial ambiguity in wood–leaf semantics rather than Points2woodyseg algorithm failure.

4. Discussion

4.1. Probabilistic Qualitative Evaluation of Wood–Foliar Semantics Error Propagation

We thoroughly investigated the disagreement between benchmark annotations and Point Transformer wood-leaf semantics using leaf continuous probabilistic maps generated by Point Transformer. Figure 23 provides additional insight into the nature of the remaining discrepancies between benchmark annotations and Point Transformer semantics. Unlike traditional segmentation, continuous leaf-probability predictions preserve semantic confidence information, allowing semantic uncertainty to be ascertained beyond the threshold-based binary semantics (Figure 23b,g). Across BlueCat and TUWIEN datasets, disagreement was highly localized rather than randomly distributed, occurring predominantly within densely occluded understories, canopy interiors, and fine-scale branch networks (Figure 23c–j). In these regions, the model consistently assigned low leaf probabilities to visually continuous fine branches that were annotated as foliage or remained unlabeled in the benchmark datasets (Figure 23d,e,i,j). On the contrary, disagreement was insignificant for stems and larger branches, where predictions and reference annotations showed almost complete agreement (Figure 23d,e). Similar relationships were observed in the BlueCat, where the wood–foliar semantics revealed additional fine branches, seedlings, and saplings that were absent from the benchmark annotations (Figure 23b,d,e–h). Comparable observations were evident for the Tropical Tree and LeWoS datasets, where Point Transformer consistently recovered more intertwined fine-scale branch architecture, particularly within densely obscured inner canopies (Figures 16 and S2).

These qualitative observations closely correspond with the quantitative height-dependent analyses presented in Figures 11 and 14. Across all benchmark datasets, segmentation performance declined primarily toward the lower understory and upper canopy extremes, whereas the mid-canopy remained comparatively stable for overall assessment (Figure 11) and genus-specific analyses (Figure 14). The height intervals coincide with regions exhibiting the greatest structural complexity, strongest obstruction, and highest abundance of fine-scale branches. Notably, the corresponding probability maps frequently retain coherent confidence patterns along visually continuous branch networks, suggesting that the remaining disagreement cannot be attributed exclusively to model failure (Figure 19g,h). Instead, the combined qualitative and quantitative evidence indicates that part of the measured reduction in benchmark performance originated from the practical limitations of manually annotating point clouds of complex forest environments. Exhaustive point-wise delineation of every small branch is inherently labor-intensive and requires subjective interpretation where wood–foliar semantics are densely intertwined, making some degree of annotation unreachable or significantly challenging (see Figures 16, 19, 20, S2 and S4).

Furthermore, a qualitative example demonstrating the transferability of "Points2WoodySeg" is provided in Appendix A (Figure A1). Building upon the multi-class decomposition. Figure A1 illustrates the application of the framework to Plot 13 of the SegmentedForests benchmark (Oregon, USA), a structurally complex *Pinus ponderosa* forest characterized by mature trees, dense understory, and complex branch morphologies [60]. Despite these challenging conditions, the framework consistently transformed Point Transformer wood–foliar semantics into comprehensible multi-class representations of stems, branches, leaves, surface foliage, surface woody debris, and woody shrubs, preserving tree architecture and structural continuity of the complete 3D fuel continuum (Figure A1, Appendix A).

These findings have broader implications for the development of future terrestrial LiDAR benchmarks. Continuous probability predictions provide substantially richer information than hard semantic labels alone by identifying regions of persistent semantic ambiguity while simultaneously highlighting high-confidence predictions (Figure 23g–i, Figure 23d,e,i,j). Nevertheless, relying exclusively on explicit manual annotation, probability-guided interactive labeling could concentrate expert review on uncertain regions while accepting high-confidence predictions elsewhere. Such a human-in-the-loop workflow has the potential to substantially reduce annotation effort, improve labeling consistency for fine intertwined branches (Figure 23d,e), and facilitate the development of larger, more reliable state-of-the-art datasets for transferable forest LiDAR semantics.

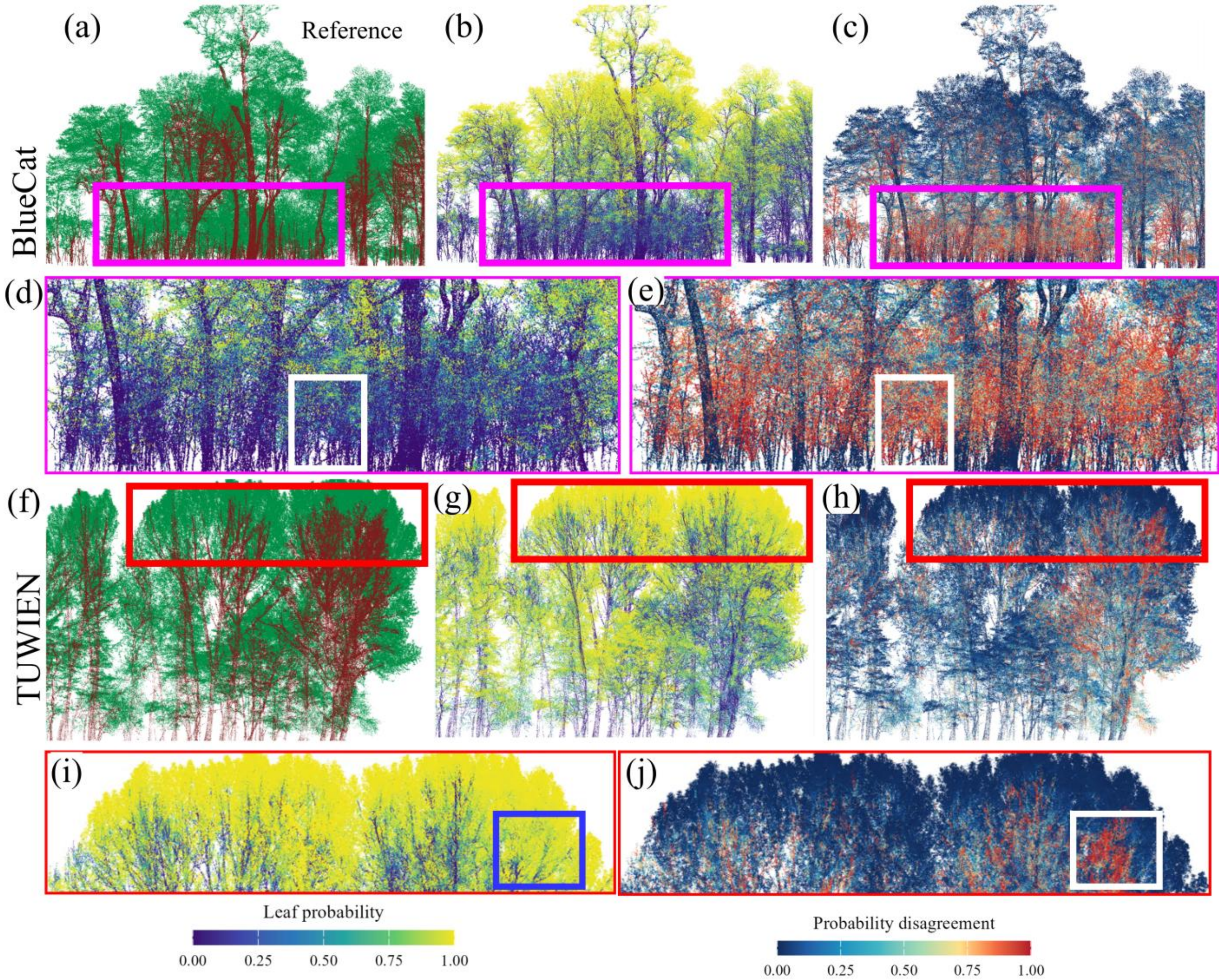

Figure 23. Qualitative comparison of benchmark annotations and Point Transformer predictions for the BlueCat and TUWIEN. (**a**,**f**) Reference wood–leaf annotations (wood: red; foliage: green). (**b**,**g**) Point Transformer continuous leaf probability confidence: 0 (lowest, blue) and 1 (highest, yellow). (**c**,**h**) Spatial probability disagreement between the reference annotations and model predictions (blue: lowest (0), and red: highest (1)). Enlarged views (**d**,**e**,**i**,**j**) highlight representative canopy regions where localized discrepancies occur. Although disagreement is concentrated within dense crowns (blue and white square boxes in (**i**,**j**)) and fine woody structures (pink and white rectangles in (**c**,**e**)), the probability maps indicate consistent geometric representations (**b**,**d**,**g**,**i**).

4.2. Comparative Assessment

Recent advances in DL have substantially improved semantic segmentation of forest LiDAR point clouds through increasingly sophisticated network architectures and learning strategies [24,35,81,83]. However, direct comparison of wood–foliar semantics (our study) and published studies remains inherently challenging because benchmark datasets differ in semantic taxonomies, training, evaluation protocols, and annotation strategies [35,36]. Consequently, reported performance reflects not only network capability but also the degree of benchmark-specific optimization and characteristics of reference annotations. To provide a rigorous and unbiased comparison, the present assessment was restricted to overlapping benchmark datasets (Lin3D, Tropical Trees, LeWoS Trees, genus-specific investigations—TLS plots) for quantitative assessment. For that reason, extensive sets of evaluation metrics were derived (Section 2.6.6) and compared with metrics reported by the original authors.

The Tropical Tree benchmark [27] provides the most direct comparison with KPConv, RandLANet, Point Transformer, the leaf–wood separation (LeWoS), and graph-based leaf–wood separation (GBS), which adopted identical wood–leaf semantics of trees (n = 148) and were explicitly optimized for the target dataset (Table 6).

Table 6. Comparative assessment of wood–foliar semantics on the Tropical Tree benchmark. Performance is summarized using Overall Accuracy (OA), Leaf Intersection over Union (LeIoU), Wood Intersection over Union (WIoU), and mean Intersection over Union (mIoU). Higher values indicate better segmentation performance.

Model	Train/Test/Validate Split?	Evaluation Protocol	OA	LeIoU	WIoU	mIoU
GBS	Yes	Same-dataset held-out test	91.90	90.20	67.80	79.00
LeWoS	Yes	Same-dataset held-out test	93.70	92.30	73.60	82.90
RandLA-Net	Yes	Same-dataset held-out test	95.20	93.90	81.80	87.90
KPConv	Yes	Same-dataset held-out test	96.20	95.20	84.20	89.70
Point Transformer	Yes	Same-dataset held-out test	**96.50**	**95.60**	**85.80**	**90.70**
Point Transformer (ours)	No	cross-dataset inference	94.09	92.72	76.11	84.41

Note: All values are reported as percentages. Boldface indicates the highest-performing value for each evaluation metric.

Table 6 demonstrates the progressive improvement in wood–foliar semantics on the Tropical Tree benchmark from the traditional GBS (OA = 91.9, mIoU = 79.0), through the LeWoS parametric algorithms (OA = 93.7, mIoU = 82.9) and point-based DL architectures, including RandLA-Net (OA = 95.2, mIoU = 87.9), KPConv (OA = 96.2, mIoU = 89.7), and Point Transformer (OA = 96.5, mIoU = 90.7). Nevertheless, the Point Transformer (ours) was not trained, validated, or adapted using the Tropical Tree benchmark. Instead, it was evaluated exclusively through frozen cross-dataset inference and achieved OA = 94.09, LeIoU = 92.72, WIoU = 76.11, and mIoU = 84.41 on the unseen Tropical Tree benchmark. On the completely unseen Tropical Tree benchmark, Point Transformer outperformed the traditional GBS and LeWoS approaches while maintaining competitive performance relative to DL frameworks. These findings suggest that a substantial proportion of the remaining performance gap reflects differences in evaluation protocol and target-domain optimization rather than limitations of the Point Transformer, as target-domain optimized DL frameworks perform better in localized context with limited transferability to an entirely different forest ecosystem.

Lin3D represents a complementary validation scenario by evaluating generalization to independent spatial hold-out plots (Table 7). In comparison, multi-platform synergistic training (MST) [36] addressed the four-class Lin3D semantic taxonomy (ground, lower objects, wood, and foliage), whereas the present study adopted wood–foliar semantics specifically designed to improve semantic consistency across forest ecosystems [35]. Unlike Point Transformer, which used only Plot 1 during training (without ground and understory) and evaluated transferability on independent spatial hold-out plots (Plots 2–6), the MST framework followed the original Lin3D v0.1 benchmark split, which includes three training plots, one validation plot, and two test plots [36]. Therefore, the reported metrics should be interpreted descriptively rather than as direct performance rankings because both the semantic taxonomy and evaluation protocols differ substantially. In particular, the proposed Point Transformer performed wood–foliar segmentation on complete forest plots containing canopy, understory, and transitional vegetation layers, which constitutes a more demanding task than the simplified four-class segmentation framework illustrated in Figure 17c,d. Despite these differences, Point Transformer achieved a lower OA (91.82 vs. 94.76) but higher mean accuracy (90.01 vs. 85.59) and mean Intersection over Union (83.40 vs. 79.48) than MST (Table 7). It is important to note that the original study reported only a limited set of evaluation metrics for the TLS-based Lin3D v0.1 benchmark, with particular emphasis placed on Lin3D v0.2, which consists of unpiloted laser-scanning (ULS) data [36]. Nevertheless, the class-level confusion metric reported for Lin3D v0.1 reveals substantial variability among semantic categories, with IoU values of 95.91 for ground, 49.70 for lower objects, 84.28 for wood, and 88.02 for foliage. Furthermore, Sen-Net [35] employed an evaluation strategy similar to MST but focused primarily on the Lin3D v0.2, which consists of ULS data. As the present study evaluates transferable wood–foliar semantics on the TLS-based Lin3D v0.1 benchmark, a direct comparison is not feasible because of differences in acquisition platform and assessment protocol.

Table 7. Comparative assessment of semantic segmentation on the Lin3D v0.1 (TLS) benchmark. Performance of the original Multi-platform Synergistic Training (MST) framework and the Point Transformer under their respective evaluation protocols. Performance is summarized using Overall Accuracy (OA), mean Accuracy (mACC), mean Intersection over Union (mIoU), Leaf Intersection over Union (LeIoU), and Wood Intersection over Union (WIoU).

Model	Train/Test/Validate Split?	OA	mACC	mIoU
MST [36]	Yes	**94.76**	85.59	79.48
Point Transformer	spatial hold-out	91.82	**90.01**	**83.40**

Note: All values are reported as percentages. Boldface indicates the highest-performing value for each evaluation metric.

More importantly, the present study extends beyond benchmark-level aggregate metrics on entire datasets by quantifying metrics of vertical stratification (Figure 11b) and investigating semantic uncertainty using probabilistic wood–foliar distributions, thereby providing insights into the origin and spatial distribution of residual errors that were not examined by Sen-Net and MST [35,36].

The LeWoS benchmark provides an additional individual tree-level evaluation on highly complex crown architecture (Table 8). Despite the absence of benchmark-specific optimization, the Point Transformer achieved higher OA (93.75 vs. 91.0) together with balanced wood recall WR (93.11) and wood specificity (94.02), demonstrating effective transfer to structurally complex forest ecosystems.

Table 8. Comparative assessment of wood–foliar semantic segmentation performance on the LeWoS benchmark. The original LeWoS framework was evaluated using the published tree-level protocol, whereas the proposed Point Transformer (Point Transformer) was evaluated using aggregate pointwise inference. Performance is reported using Overall Accuracy (OA), Wood Recall (WR), and Wood Specificity (WS). Higher values indicate better segmentation performance.

Model	Train/Validation/Test Split?	Evaluation Protocol	OA	WR	Wood Specificity
LeWoS	Yes	Tree-level mean ± SD	91.0 ± 3.0	92.0 ± 4.0	89.0 ± 6.0
Point Transformer (ours)	No	Aggregate point-wise evaluation	**93.75**	**93.11**	**94.02**

Note: All values are reported as percentages. Boldface indicates the highest-performing value for each evaluation metric.

4.3. The Rationale of Wood–Foliar Semantic vs. Multi-Class Segmentation

Terrestrial LiDAR semantic segmentation has traditionally relied on multi-class taxonomies that partition forest scenes into categories such as ground, lower vegetation, shrubs, stems, branches, foliage, and CWD [64,83]. Although these taxonomies provide detailed ecological descriptions, they remain benchmark-dependent because class definitions, annotation protocols, and ecological objectives differ substantially among datasets [60]. Consequently, identical 3D structures may receive different semantic labels across benchmarks, limiting direct comparison and reducing transferability.

Evidence from recent DL studies consistently demonstrates that this limitation is concentrated within intermediate structural classes rather than the overall segmentation task [35]. This pattern is well evident in published multi-class semantic investigations in previous studies. For example, on the FOR-Instance benchmark, Xiang et al. (2024) reported that ForAINet achieved a confusion matrix between five semantic categories with 85.6 for low vegetation, 89.2 for ground, 55.3 for stems, 98.3 for live branches (a combined class comprising both woody branches and foliage), and 68.4 for dead branches, i.e., branches without foliage points [81]. Using the Lin3D and similar semantic taxonomy, Lu et al. (2025) reported confusion metrics of 82.1 for low vegetation, 96.2 for ground, 88.0 for wood (stem and branches collectively), and 99.6 for foliage [35]. In the same study, the confusion matrix of the five semantic categories on the FOR-Instance benchmark was ground (87.9), low vegetation (94.9), stem (53.2), live branches (99.5), and woody branches (85.2). The authors attributed this substantially lower accuracy to the strong geometric continuity and spatial proximity between lower objects and the ground surface, making their semantic boundaries inherently ambiguous [35]. The six-class HeliALS benchmark presented by Takhtkeshha et al. (2025) provides an even stronger illustration of this limitation [64]. Using geometric coordinates alone, KPConv achieved confusion matrices of 97.7 for ground and 99.7 for foliage, whereas low vegetation (95.1), trunks (77.7), branches (34.6), and woody debris (92.4). The authors attributed this poor performance to the increasing structural complexity, geometric similarity, and annotation uncertainty associated with progressively smaller branch hierarchies [83].

The proposed wood–foliar formulation therefore does not merely simplify a multi-class problem; it separates physically persistent foliage from application-specific structural interpretations. All woody points are assigned to wood, while photosynthetically active vegetation is assigned to foliage, avoiding the need to impose uncertain boundaries between stems, branch orders, standing deadwood, CWD, FWD, woody shrubs, and ladder fuels during DL inference. This rationale is supported by Hanousek et al. (2025), who removed CWD from a pretrained four-class taxonomy because lower branches were frequently confused with CWD or ground and subsequently achieved 96.3 overall accuracy and 97.05 wood-class accuracy using ground–wood–foliage labels [88]. Wood–foliar semantics can thus serve as a transferable intermediate representation, with ecologically specific stems, branches, and fuel classes subsequently derived through hierarchical geometric and topological decomposition.

Importantly, the proposed wood–foliar abstraction should not be interpreted as a simplification that sacrifices ecological information. Rather, it establishes a transferable intermediate representation from which higher-order structural classes can subsequently be recovered through geometry-driven post-processing (Objective 3; Figures 21 and A1). We showed that with Point Transformer wood–foliar semantics, the continuous woody skeleton can be hierarchically decomposed into stems, branches, foliage, woody shrubs, and coarse and fine woody classes without requiring the DL multi-class semantics to solve an increasingly ambiguous multi-class problem directly (Figures 21 and A1, Appendix A). This decoupling of semantic recognition from structural decomposition enables the network to learn transferable material classes while deterministic geometric algorithms exploit spatial topology to derive application-specific structural products. Consequently, the framework provides a wood–foliar semantic foundation for wildfire fuel characterization, quantitative structural modeling (QSM), AGB estimation, forest inventory, and leaf-off structural analysis across diverse forest ecosystems.

Recent studies have demonstrated that terrestrial TLS LiDAR point clouds can be voxelized and converted into 3D fuel semantics suitable for physics-based fire behavior models, including the Wildland–Urban Interface Fire Dynamics Simulator (WFDS), QUIC-Fire, and the Fire Dynamics Simulator (FDS), thereby enabling explicit characterization of fuel continuity and vertical fuel structure [87]. However, the semantic decomposition required to distinguish surface, ladder, canopy, woody, and foliar fuels remains a major bottleneck for operational fire simulations [60,89]. Previous studies have highlighted that semantic segmentation of forest point clouds is fundamentally constrained by structural overlap, occlusion, and semantic ambiguity among lower vegetation, woody debris, branches, and understory [89–91]. Consequently, most existing studies remain limited to laboratory settings [92], individual trees, or simplified fuel representations [93], restricting the applicability of TLS-based fire simulations to larger and more heterogeneous forest environments [89]. By preserving transferable wood–foliar semantics and geometric-based decomposition to fuel relevant classes across complete forest plots and diverse forest ecosystems (see Figures 21b and A1, Appendix A), the present study provides an important step toward scalable 3D fuel characterization for next-generation wildfire behavior models [87].

5. Conclusions

We investigated and developed a Python-based wood–foliar semantics deep learning (DL) framework known as “points2SBL” (https://pypi.org/project/points2sbl/, accessed on 9 September 2026). points2SBL was evaluated to address three major research questions concerning: the trained DL framework’s transferability across global forest ecosystems, the generalization and transferability assessment of widely used DL architectures (PointNet++, PointNeXt, Point Transformer), and subsequent wildland fuel characterization from wood–foliar semantics using the geometry-driven algorithm named “Points2woodyseg”.

The benchmark evaluations demonstrated that LiDAR point clouds, along with fewer geometric descriptors—linearity, planarity, curvature, and scattering-based DL frameworks—can provide transferable wood–foliar segmentation across global forest ecosystems and terrestrial LiDAR platforms. Over five benchmark forest point clouds, Point Transformer achieved overall accuracies (OA) of 91.63–94.09%, mean Intersection-over-Union (mIoU) of 77.57–86.57%, F1-scores of 86.80–92.73%, and Matthews correlation coefficients (MCC) of 73.65–85.57%. The consistency of these results across plot- and individual-tree LiDAR point clouds supports the transferability potential of simplified binary wood–foliar semantics under substantially different forest and acquisition conditions and LiDAR sensors. Secondly, the generalization and transferability potential of PointNet++, PointNeXt, and Point Transformer under identical training and evaluation protocols showed that Point Transformer provided substantially higher and consistent accuracy over benchmarks. Finally, from the wood–foliar predictions, we retained wood points to support subsequent decomposition into ecologically relevant structural components of forest surface fuel of wood–foliar classes, along with stem branches and leaves of dominant and co-dominant canopy.

Collectively, these findings demonstrated that binary wood–foliar semantics can reduce classification complexity while preserving essential three-dimensional (3D) forest structural and composition information, and that wood semantics can be decomposed into several other forest semantics through geometrical decomposition approaches rather than DL frameworks. points2SBL therefore provides a transferable semantic foundation for terrestrial LiDAR analysis and a practical basis for future development of 3D forest-fuel characterization.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18183158/s1. Figure S1: Distribution of point-wise classification outcomes for Point Transformer (PT) wood–foliar semantic segmentation across the four evaluation datasets: (a) TLS plots, (b) Lin3D, (c) Tropical tree, and (d) LeWoS tree. Stacked bars show the proportions (%) of true negatives (TN), true positives (TP), false positives (FP), and false negatives (FN), with the percentage contribution of each component indicated within the corresponding bar segment. Figure S2: Qualitative assessment of wood–leaf semantic segmentation using two representative trees from the LeWoS benchmark. Panels (a) and (c) show the benchmark wood–leaf annotations, whereas panels (b) and (d) present the corresponding Point Transformer (PT) semantics from lateral views. Panels (e) and (g) show the benchmark annotations from bottom-up views, whereas panels (f) and (h) present the corresponding PT semantics. Woody points are shown in brown and foliar points in green. Figure S3: Comparison of the original SegmentedForests annotations and Points2SBL wood–foliar predictions for Plot 2, a natural *Fagus sylvatica* forest in Asturias, Spain, acquired using a GeoSLAM ZEB Horizon mobile laser scanning (MLS) system (GeoSLAM Ltd., Nottingham, UK). The plot is characterized by mature, sparsely distributed trees, thick branching architecture, and minimal understory. Figure S4: Transferability example for the independent NIBIO MLS dataset. From left to right: manually generated reference annotations, Point Transformer (PT) binary wood–foliar semantic predictions, continuous leaf-probability predictions, and the absolute difference between the reference labels and predicted leaf probabilities, illustrating regions of prediction uncertainty.

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

Funding: This research was funded by the Strategic Environmental Research and Development Program (SERDP) and the Environmental Security Technology Certification Program (ESTCP)—FuelsCraft: An innovative wildland fuel mapping tool for prescribed fire decision support on Department of Defense (DoD) military installations (#RC23-7779).

Data Availability Statement: Data will be made available upon request from the corresponding author.

Acknowledgments: Generative AI: ChatGPT (https://chatgpt.com/, accessed on 15 July 2026), version (GPT-5.6 Sol) and Perplexity AI (https://www.perplexity.ai/, accessed on 15 July 2026), version (Sonar 2) were used to validate open-access datasets, acronyms, and referenced articles. Grammarly (https://app.grammarly.com/, accessed on 15 July 2026), version (v1.2.291.1949)was used to fix typographical and grammatical errors.

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

Appendix A

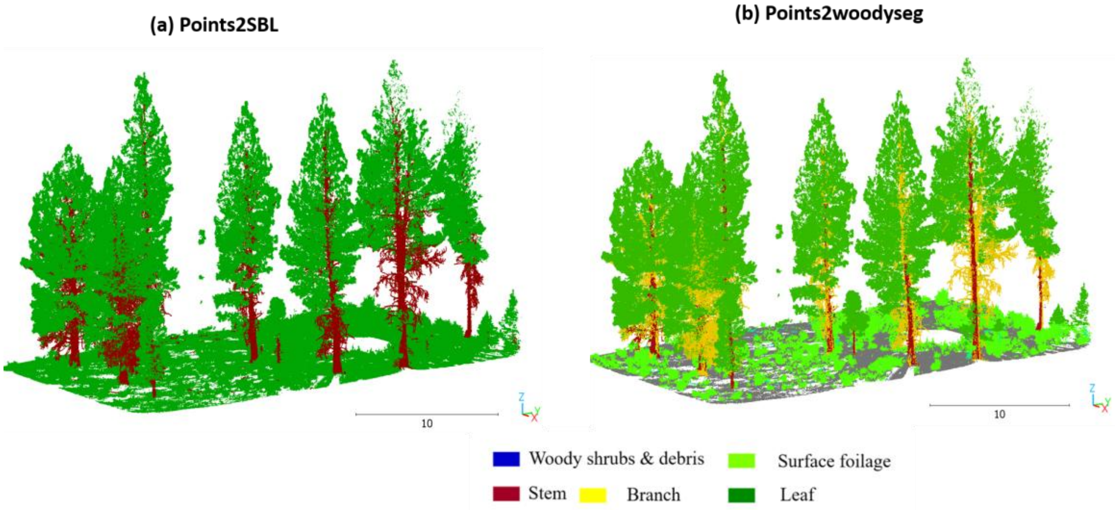

Figure A1. points2SBL wood–foliar semantics (**a**) to Points2WoodySeg (**b**) decomposition of Plot 13 from the SegmentedForests dataset [60], representing a structurally complex Pinus ponderosa forest in Oregon, USA. The output illustrates the multiclass decomposition of forest structure into stems, branches, foliage, ground, and surface vegetation under sparse mature-tree cover, dense understory, and strong vertical structural continuity.

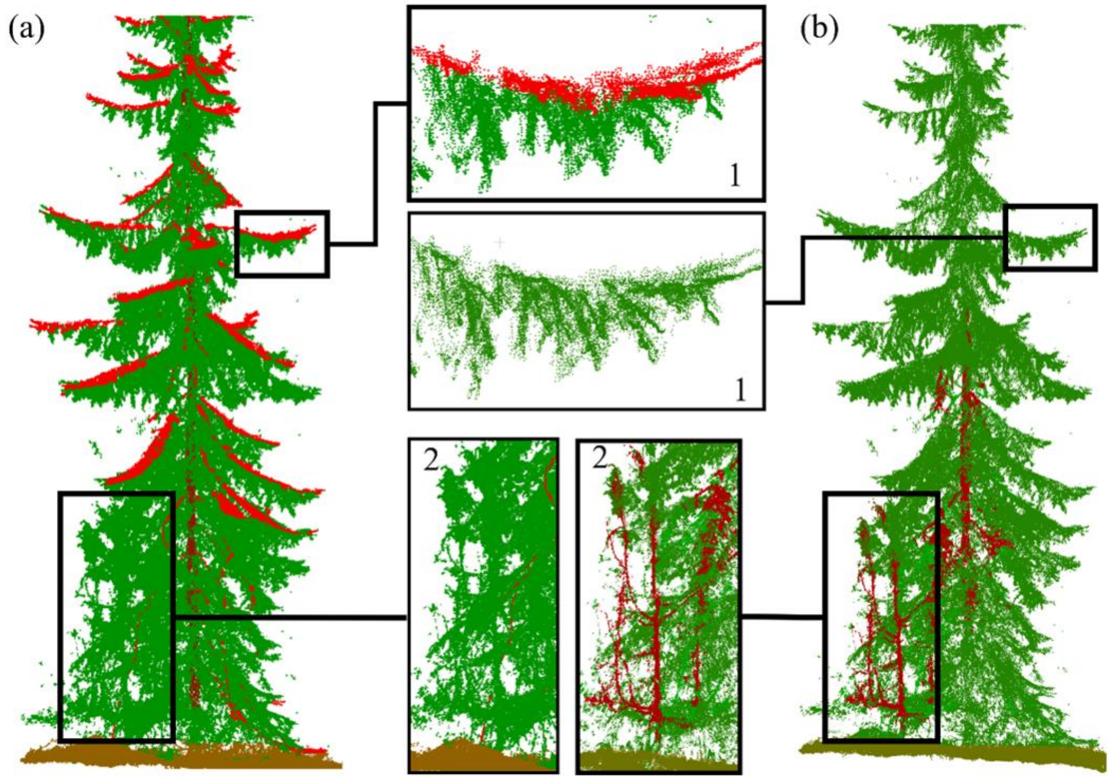

Figure A2. Visual comparison between (**a**) manually annotated ForestSemantic benchmark labels and (**b**) Point Transformer wood–foliar semantics. Wood points are shown in red and leaf points in green. Insets highlight representative regions of disagreement. Subset 1 illustrates differences at occluded outer crown margins, whereas subset 2 shows additional fine wood structures identified by Point Transformer within the understory that were not assigned as wood in the benchmark annotations. These examples illustrate the uncertainty associated with manual wood–leaf annotation in a structurally complex understory.

References

1. Hartley, R.J.L.; Jayathunga, S.; Morgenroth, J.; Pearse, G.D. Tree Branch Characterisation from Point Clouds: A Comprehensive Review. *Curr. For. Rep.* **2024**, *10*, 360–385. https://doi.org/10.1007/s40725-024-00225-5.
2. Alonso-Rego, C.; Arellano-Pérez, S.; Cabo, C.; Ordoñez, C.; Álvarez-González, J.G.; Díaz-Varela, R.A.; Ruiz-González, A.D. Estimating Fuel Loads and Structural Characteristics of Shrub Communities by Using Terrestrial Laser Scanning. *Remote Sens.* **2020**, *12*, 3704. https://doi.org/10.3390/rs12223704.
3. Kulicki, M.; Cabo, C.; Trzcinski, T.; Bedkowski, J.; Sterenczak, K. Artificial Intelligence and Terrestrial Point Clouds for Forest Monitoring. *Curr. For Rep.* **2025**, *11*, 5. https://doi.org/10.1007/s40725-024-00234-4.
4. Zhao, C.; Fei, S.; Habib, A. Integrated scan simultaneous trajectory enhancement and mapping (IS2-TEAM) for fine resolution forest inventory using backpack LiDAR. *Remote Sens. Environ.* **2026**, *334*, 115212. https://doi.org/10.1016/j.rse.2025.115212.
5. Shao, J.; Lin, Y.-C.; Wingren, C.; Shin, S.-Y.; Fei, W.; Carpenter, J.; Habib, A.; Fei, S. Large-scale inventory in natural forests with mobile LiDAR point clouds. *Sci. Remote Sens.* **2024**, *10*, 100168. https://doi.org/10.1016/j.srs.2024.100168.
6. Wilson, N.; Bradstock, R.; Bedward, M. Influence of fuel structure derived from terrestrial laser scanning (TLS) on wildfire severity in logged forests. *J. Environ. Manag.* **2022**, *302*, 114011. https://doi.org/10.1016/j.jenvman.2021.114011.
7. Rowell, E.; Loudermilk, E.L.; Hawley, C.; Pokswinski, S.; Seielstad, C.; Queen, L.; O'Brien, J.J.; Hudak, A.T.; Goodrick, S.; Hiers, J.K. Coupling terrestrial laser scanning with 3D fuel biomass sampling for advancing wildland fuels characterization. *For. Ecol. Manag.* **2020**, *462*, 117945. https://doi.org/10.1016/j.foreco.2020.117945.
8. Xi, Z.; Chasmer, L.; Hopkinson, C. Delineating and Reconstructing 3D Forest Fuel Components and Volumes with Terrestrial Laser Scanning. *Remote Sens.* **2023**, *15*, 4778. https://doi.org/10.3390/rs15194778.
9. Hall, S.A.; Burke, I.C. Considerations for characterizing fuels as inputs for fire behavior models. *For. Ecol. Manag.* **2006**, *227*, 102–114. https://doi.org/10.1016/j.foreco.2006.02.022.
10. Hui, Z.; Jin, S.; Xia, Y.; Wang, L.; Yevenyo Ziggah, Y.; Cheng, P. Wood and leaf separation from terrestrial LiDAR point clouds based on mode points evolution. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 219–239. https://doi.org/10.1016/j.isprsjprs.2021.06.012.
11. Ruoppa, L.; Oinonen, O.; Taher, J.; Lehtomäki, M.; Takhtkeshha, N.; Kukko, A.; Kaartinen, H.; Hyyppä, J. Unsupervised deep learning for semantic segmentation of multispectral LiDAR forest point clouds. *ISPRS J. Photogramm. Remote Sens.* **2025**, *228*, 694–722. https://doi.org/10.1016/j.isprsjprs.2025.07.038.
12. Xu, X.; Iuricich, F.; Calders, K.; Armston, J.; De Floriani, L. Topology-based individual tree segmentation for automated processing of terrestrial laser scanning point clouds. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103145. https://doi.org/10.1016/j.jag.2022.103145.
13. Wilkes, P.; Disney, M.; Armston, J.; Bartholomeus, H.; Bentley, L.; Brede, B.; Burt, A.; Calders, K.; Chavana-Bryant, C.; Clewley, D.; et al. TLS2trees: A scalable tree segmentation pipeline forTLSdata. *Methods Ecol. Evol.* **2023**, *14*, 3083–3099. https://doi.org/10.1111/2041-210x.14233.
14. Xi, Z.; Hopkinson, C.; Chasmer, L. Supervised terrestrial to airborne laser scanner model calibration for 3D individual-tree attribute mapping using deep neural networks. *ISPRS J. Photogramm. Remote Sens.* **2024**, *209*, 324–343. https://doi.org/10.1016/j.isprsjprs.2024.02.010.
15. Grilli, E.; Menna, F.; Remondino, F. A review of point clouds segmentation and classification algorithms. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 339–344.
16. Fareed, N.; Silva, C.A.; Numata, I.; Flores, J.P. Interdisciplinary Applications of LiDAR in Forest Studies: Advances in Sensors, Methods, and Cross-Domain Metrics. *Remote Sens.* **2026**, *18*, 219.
17. Viedma, O.; Moreno, J.M. Impact of LiDAR pulse density on forest fuels metrics derived using LadderFuelsR. *Ecol. Inform.* **2025**, *88*, 103135. https://doi.org/10.1016/j.ecoinf.2025.103135.
18. Wan, P.; Shao, J.; Jin, S.; Wang, T.; Yang, S.; Yan, G.; Zhang, W. A novel and efficient method for wood–leaf separation from terrestrial laser scanning point clouds at the forest plot level. *Methods Ecol. Evol.* **2021**, *12*, 2473–2486. https://doi.org/10.1111/2041-210x.13715.
19. Ferrara, R.; Virdis, S.G.P.; Ventura, A.; Ghisu, T.; Duce, P.; Pellizzaro, G. An automated approach for wood-leaf separation from terrestrial LIDAR point clouds using the density based clustering algorithm DBSCAN. *Agric. For. Meteorol.* **2018**, *262*, 434–444. https://doi.org/10.1016/j.agrformet.2018.04.008.
20. Vicari, M.B.; Disney, M.; Wilkes, P.; Burt, A.; Calders, K.; Woodgate, W.; Freckleton, R. Leaf and wood classification framework for terrestrial LiDAR point clouds. *Methods Ecol. Evol.* **2019**, *10*, 680–694. https://doi.org/10.1111/2041-210x.13144.

21. Xia, J.; Martin, T.A.; Peter, G.F.; Brock, K.M.; Atkins, J.W.; Gitzendanner, M.A.; Bueno, I.; Calders, K.; Corte, A.P.D.; Hudak, A.T.; et al. Combined impact of semantic segmentation and quantitative structure modelling of Southern pine trees using terrestrial laser scanning. *Sci. Rep.* **2025**, *15*, 24427. https://doi.org/10.1038/s41598-025-09681-w.
22. Wang, D. Unsupervised semantic and instance segmentation of forest point clouds. *ISPRS J. Photogramm. Remote Sens.* **2020**, *165*, 86–97. https://doi.org/10.1016/j.isprsjprs.2020.04.020.
23. Chen, S.; Verbeeck, H.; Terryn, L.; Van den Broeck, W.A.J.; Vicari, M.B.; Disney, M.; Origo, N.; Wang, D.; Xi, Z.; Hopkinson, C.; et al. The impact of leaf-wood separation algorithms on aboveground biomass estimation from terrestrial laser scanning. *Remote Sens. Environ.* **2025**, *318*, 114581. https://doi.org/10.1016/j.rse.2024.114581.
24. Xi, Z.; Hopkinson, C.; Rood, S.B.; Peddle, D.R. See the forest and the trees: Effective machine and deep learning algorithms for wood filtering and tree species classification from terrestrial laser scanning. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 1–16. https://doi.org/10.1016/j.isprsjprs.2020.08.001.
25. Dong, Y.; Ma, Z.; Xu, F.; Chen, F. Unsupervised Semantic Segmenting TLS Data of Individual Tree Based on Smoothness Constraint Using Open-Source Datasets. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. https://doi.org/10.1109/tgrs.2022.3218442.
26. Esmorís, A.M.; Weiser, H.; Winiwarter, L.; Cabaleiro, J.C.; Höfle, B. Deep learning with simulated laser scanning data for 3D point cloud classification. *ISPRS J. Photogramm. Remote Sens.* **2024**, *215*, 192–213. https://doi.org/10.1016/j.isprsjprs.2024.06.018.
27. Van den Broeck, W.A.J.; Terryn, L.; Chen, S.; Cherlet, W.; Cooper, Z.T.; Calders, K. Pointwise deep learning for leaf-wood segmentation of tropical tree point clouds from terrestrial laser scanning. *ISPRS J. Photogramm. Remote Sens.* **2025**, *227*, 366–382. https://doi.org/10.1016/j.isprsjprs.2025.06.023.
28. Wang, P.; Yao, W.; Shao, J.; He, Z. Test-time adaptation for geospatial point cloud semantic segmentation with distinct domain shifts. *ISPRS J. Photogramm. Remote Sens.* **2025**, *229*, 422–435. https://doi.org/10.1016/j.isprsjprs.2025.08.022.
29. Pehkonen, M.; Vastaranta, M.; Hyyppä, J.; Pyörälä, J. Segmentation of living and dead tree crowns using terrestrial laser scanning and deep learning. *Ecol. Inform.* **2026**, *95*, 103750. https://doi.org/10.1016/j.ecoinf.2026.103750.
30. Xiang, B.; Wielgosz, M.; Puliti, S.; Král, K.; Krůček, M.; Missarov, A.; Astrup, R. Forestformer3d: A unified framework for end-to-end segmentation of forest lidar 3d point clouds. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, Honolulu, HI, USA, 19–25 October 2025; pp. 24717–24727.
31. Terryn, L.; Calders, K.; Åkerblom, M.; Bartholomeus, H.; Disney, M.; Levick, S.; Origo, N.; Raumonen, P.; Verbeeck, H. Analysing individual 3D tree structure using the R package ITSMe. *Methods Ecol. Evol.* **2022**, *14*, 231–241. https://doi.org/10.1111/2041-210x.14026.
32. Luo, X.; Tian, X.; Liang, X.; Mokros, M.; Chai, G.; Li, Z.; Guo, Y.; Yang, Y.; Pang, Y.; Wang, Y.; et al. Cognition-inspired multimodal attention fusion of close-range laser scanning data for globally representative tree species classification. *Remote Sens. Environ.* **2026**, *335*, 115269. https://doi.org/10.1016/j.rse.2026.115269.
33. Van den Broeck, W.A.J.; Terryn, L.; Cherlet, W.; Cooper, Z.T.; Calders, K. Three-Dimensional Deep Learning for Leaf-Wood Segmentation of Tropical Tree Point Clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48*, 765–770. https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-765-2023.
34. Krisanski, S.; Taskhiri, M.S.; Gonzalez Aracil, S.; Herries, D.; Muneri, A.; Gurung, M.B.; Montgomery, J.; Turner, P. Forest Structural Complexity Tool—An Open Source, Fully-Automated Tool for Measuring Forest Point Clouds. *Remote Sens.* **2021**, *13*, 4677.
35. Lu, H.; Li, B.; Yang, G.; Fan, G.; Wang, H.; Pang, Y.; Wang, Z.; Lian, Y.; Xu, H.; Huang, H. Towards a point cloud understanding framework for forest scene semantic segmentation across forest types and sensor platforms. *Remote Sens. Environ.* **2025**, *318*, 114591. https://doi.org/10.1016/j.rse.2024.114591.
36. Jiang, J.; Shen, Y.; Wang, J.; Kissling, W.D.; Hollaus, M.; Su, H.; Wang, J.; Ferreira, V.; Pfeifer, N. Cross-platform forest understanding: A multi-platform synergistic training framework for generalized forest point cloud segmentation. *Remote Sens. Environ.* **2026**, *342*, 115467. https://doi.org/10.1016/j.rse.2026.115467.
37. Kaijaluoto, R.; Kukko, A.; El Issaoui, A.; Hyyppä, J.; Kaartinen, H. Semantic segmentation of point cloud data using raw laser scanner measurements and deep neural networks. *ISPRS Open J. Photogramm. Remote Sens.* **2022**, *3*, 100011. https://doi.org/10.1016/j.ophoto.2021.100011.
38. Oviedo de la Fuente, M.; Cabo, C.; Roca-Pardiñas, J.; Loudermilk, E.L.; Ordóñez, C. 3D Point Cloud Semantic Segmentation Through Functional Data Analysis. *J. Agric. Biol. Environ. Stat.* **2023**, *29*, 723–744. https://doi.org/10.1007/s13253-023-00567-w.
39. Cherlet, W.; Dayal, K.; Chen, S.; Cooper, Z.; Disney, M.; Hanzl, A.; Levick, S.; Nightingale, J.; Origo, N.; Senf, C.; et al. Benchmarking tree instance segmentation of terrestrial laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* **2026**, *231*, 230–247. https://doi.org/10.1016/j.isprsjprs.2025.10.033.

40. Arrizza, S.; Marras, S.; Ferrara, R.; Pellizzaro, G. Terrestrial Laser Scanning (TLS) for tree structure studies: A review of methods for wood-leaf classifications from 3D point clouds. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101364. https://doi.org/10.1016/j.rsase.2024.101364.
41. Comesaña-Cebral, L.; Martínez-Sánchez, J.; Suárez-Fernández, G.; Arias, P. Wildfire response of forest species from multispectral LiDAR data. A deep learning approach with synthetic data. *Ecol. Inform.* **2024**, *81*, 102612. https://doi.org/10.1016/j.ecoinf.2024.102612.
42. Brown, L.A.; Kadhim, I.; Danson, F.M. Leaf-wood classification of terrestrial laser scanning data with co-registered near-infrared photography. *Methods Ecol. Evol.* **2025**, *16*, 1425–1436. https://doi.org/10.1111/2041-210x.70060.
43. Dai, W.; Jiang, Y.; Zeng, W.; Chen, R.; Xu, Y.; Zhu, N.; Xiao, W.; Dong, Z.; Guan, Q. MDC-Net: A multi-directional constrained and prior assisted neural network for wood and leaf separation from terrestrial laser scanning. *Int. J. Digit. Earth* **2023**, *16*, 1224–1245. https://doi.org/10.1080/17538947.2023.2198261.
44. Ottmar, R.D.; Sandberg, D.V.; Prichard, S.J.; Riccardi, C.L. Fuel characteristic classification system. In *Proceedings of the Presentation at the 2nd International Wildland Fire Ecology and Fire Management Congress*; American Meteorological Society: Boston, MA, USA, 2003. Available online: http://ams.confex.com/ams/FIRE2003/techprogram/paper_67236.htm (accessed on 15 July 2026).
45. Scott, J.H. *Assessing Crown Fire Potential by Linking Models of Surface and Crown Fire Behavior*; US Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2001.
46. Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. Lidar Remote Sensing for Ecosystem Studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience* **2002**, *52*, 19–30. https://doi.org/10.1641/0006-3568(2002)052[0019:Lrsfes]2.0.Co;2.
47. Le Toan, T.; Quegan, S.; Davidson, M.W.J.; Balzter, H.; Paillou, P.; Papathanassiou, K.; Plummer, S.; Rocca, F.; Saatchi, S.; Shugart, H.; et al. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sens. Environ.* **2011**, *115*, 2850–2860. https://doi.org/10.1016/j.rse.2011.03.020.
48. Jucker, T.; Gosper, C.R.; Wiehl, G.; Yeoh, P.B.; Raisbeck-Brown, N.; Fischer, F.J.; Graham, J.; Langley, H.; Newchurch, W.; O'Donnell, A.J.; et al. Using multi-platform LiDAR to guide the conservation of the world's largest temperate woodland. *Remote Sens. Environ.* **2023**, *296*, 113745. https://doi.org/10.1016/j.rse.2023.113745.
49. Wilson, N.; Bradstock, R.; Bedward, M. Detecting the effects of logging and wildfire on forest fuel structure using terrestrial laser scanning (TLS). *For. Ecol. Manag.* **2021**, *488*, 119037. https://doi.org/10.1016/j.foreco.2021.119037.
50. Forbes, B.; Reilly, S.; Clark, M.; Ferrell, R.; Kelly, A.; Krause, P.; Matley, C.; O'Neil, M.; Villasenor, M.; Disney, M.; et al. Comparing Remote Sensing and Field-Based Approaches to Estimate Ladder Fuels and Predict Wildfire Burn Severity. *Front. For. Glob. Change* **2022**, *5*, 818713. https://doi.org/10.3389/ffgc.2022.818713.
51. Yépez-Rincón, F.D.; Luna-Mendoza, L.; Ramírez-Serrato, N.L.; Hinojosa-Corona, A.; Ferriño-Fierro, A.L. Assessing vertical structure of an endemic forest in succession using terrestrial laser scanning (TLS). Case study: Guadalupe Island. *Remote Sens. Environ.* **2021**, *263*, 112563. https://doi.org/10.1016/j.rse.2021.112563.
52. Brown, J.K. *Handbook for Inventorying Downed Woody Material*; Gen. Tech. Rep. INT-16; US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station: Ogden, UT, UDA, 1974; Volume 16, 24p.
53. Monsanto, P.G.; Agee, J.K. Long-term post-wildfire dynamics of coarse woody debris after salvage logging and implications for soil heating in dry forests of the eastern Cascades, Washington. *For. Ecol. Manag.* **2008**, *255*, 3952–3961. https://doi.org/10.1016/j.foreco.2008.03.048.
54. Lydersen, J.M.; Collins, B.M.; Coppoletta, M.; Jaffe, M.R.; Northrop, H.; Stephens, S.L. Fuel dynamics and reburn severity following high-severity fire in a Sierra Nevada, USA, mixed-conifer forest. *Fire Ecol.* **2019**, *15*, 43. https://doi.org/10.1186/s42408-019-0060-x.
55. Prichard, S.J.; O'Neill, S.M.; Eagle, P.; Andreu, A.G.; Drye, B.; Dubowy, J.; Urbanski, S.; Strand, T.M. Wildland fire emission factors in North America: Synthesis of existing data, measurement needs and management applications. *Int. J. Wildland Fire* **2020**, *29*, 132–147. https://doi.org/10.1071/wf19066.
56. Brown, J.K. *Coarse Woody Debris: Managing Benefits and Fire Hazard in the Recovering Forest*; US Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2003.
57. Loudermilk, E.L.; Pokswinski, S.; Hawley, C.M.; Maxwell, A.; Gallagher, M.R.; Skowronski, N.S.; Hudak, A.T.; Hoffman, C.; Hiers, J.K. Terrestrial Laser Scan Metrics Predict Surface Vegetation Biomass and Consumption in a Frequently Burned Southeastern U.S. Ecosystem. *Fire* **2023**, *6*, 151.

58. Parresol, B.R.; Blake, J.I.; Thompson, A.J. Effects of overstory composition and prescribed fire on fuel loading across a heterogeneous managed landscape in the southeastern USA. *For. Ecol. Manag.* **2012**, *273*, 29–42. https://doi.org/10.1016/j.foreco.2011.08.003.
59. Keane, R.E. *Wildland Fuel Fundamentals and Applications*; Springer: Berlin/Heidelberg, Germany, 2015.
60. Laino, D.; Cabo, C.; Ordóñez, C.; Bolanos, R.; Janvier, R.; Giulioni, F.; Herrmann, M.; Hudak, A.; Parsons, R.; Santin, C.; et al. SegmentedForests: A labelled dataset of terrestrial LiDAR point clouds for semantic segmentation of forests. *For. Int. J. For. Res.* **2026**, *99*, cpaf062. https://doi.org/10.1093/forestry/cpaf062.
61. Che, E.; Olsen, M.J. Fast ground filtering for TLS data via Scanline Density Analysis. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 226–240. https://doi.org/10.1016/j.isprsjprs.2017.05.006.
62. Qin, N.; Tan, W.; Guan, H.; Wang, L.; Ma, L.; Tao, P.; Fatholahi, S.; Hu, X.; Li, J. Towards intelligent ground filtering of large-scale topographic point clouds: A comprehensive survey. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *125*, 103566. https://doi.org/10.1016/j.jag.2023.103566.
63. Hillman, S.; Wallace, L.; Reinke, K.; Jones, S. A comparison between TLS and UAS LiDAR to represent eucalypt crown fuel characteristics. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 295–307. https://doi.org/10.1016/j.isprsjprs.2021.09.008.
64. Takhtkeshha, N.; Bocaux, L.; Ruoppa, L.; Remondino, F.; Mandlburger, G.; Kukko, A.; Hyyppä, J. 3D Forest Semantic Segmentation Using Multispectral LiDAR and 3D Deep Learning. *PFG–J. Photogramm. Remote Sens. Geoinf. Sci.* **2025**, *94*, 377–397. https://doi.org/10.1007/s41064-025-00369-4.
65. Zhang, F.; Chancia, R.; Clapp, J.; Hassanzadeh, A.; Dera, D.; MacKenzie, R.; van Aardt, J. Through the perspective of LiDAR: A feature-enriched and uncertainty-aware annotation pipeline for terrestrial point cloud segmentation. *ISPRS J. Photogramm. Remote Sens.* **2026**, *236*, 141–161. https://doi.org/10.1016/j.isprsjprs.2026.03.033.
66. Xie, Y.; Han, D.; Liu, J.; Yang, K.; Wang, X.; Li, L.; Ma, Q.; Yang, P. A TLS point cloud dataset with detailed annotations for leaf-on and leaf-off seasons across multiple tree species—HuashuTrees. *Sci. Data* **2026**. https://doi.org/10.1038/s41597-026-08199-8.
67. Abdollahi, A.; Yebra, M. Forest fuel type classification: Review of remote sensing techniques, constraints and future trends. *J. Environ. Manag.* **2023**, *342*, 118315. https://doi.org/10.1016/j.jenvman.2023.118315.
68. Weiser, H.; Ulrich, V.; Winiwarter, L.; Pena, A.M.E.; Höfle, B. *Manually Labeled Terrestrial Laser Scanning Point Clouds of Individual Trees for Leaf-Wood Separation*; heiDATA: Heidelberg, Germany, 2019.
69. Owen, H.J.F.; Lines, E.; Grieve, S. *Plot-Level Semantically Labelled Terrestrial Laser Scanning Point Clouds*, Version 1.0 [Dataset]; Zenodo: Geneva, Switzerland, 2024. https://doi.org/10.5281/zenodo.13268500.
70. Weiser, H.; Schäfer, J.; Winiwarter, L.; Krašovec, N.; Fassnacht, F.E.; Höfle, B. Individual tree point clouds and tree measurements from multi-platform laser scanning in German forests. *Earth Syst. Sci. Data* **2022**, *14*, 2989–3012. https://doi.org/10.5194/essd-14-2989-2022.
71. Maiti, A.; Oude Elberink, S.J.; Vosselman, G. Effect Of Label Noise in Semantic Segmentation of High Resolution Aerial Images and Height Data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *2*, 275–282. https://doi.org/10.5194/isprs-annals-V-2-2022-275-2022.
72. Liu, J.; Wang, D.; Gong, H.; Wang, C.; Zhu, J.; Wang, D. A synthetic data generation framework for deep learning-based LiDAR forest structure analysis. *Remote Sens. Environ.* **2026**, *341*, 115436. https://doi.org/10.1016/j.rse.2026.115436.
73. Xi, Z.; Hopkinson, C. *Terrestrial Laser Scanning (TLS) Plot Scans from Varying Natural Forest Environments*; Federated Research Data Repository: Toronto, ON, Canada, 2022.
74. Xi, Z.; Hopkinson, C. Detecting Individual-Tree Crown Regions from Terrestrial Laser Scans with an Anchor-Free Deep Learning Model. *Can. J. Remote Sens.* **2021**, *47*, 228–242. https://doi.org/10.1080/07038992.2020.1861541.
75. Wang, D.; Momo Takoudjou, S.; Casella, E. LeWoS: A universal leaf-wood classification method to facilitate the 3D modelling of large tropical trees using terrestrial LiDAR. *Methods Ecol. Evol.* **2020**, *11*, 376–389.
76. Liang, X.; Qi, H.; Deng, X.; Chen, J.; Cai, S.; Zhang, Q.; Wang, Y.; Kukko, A.; Hyyppä, J. ForestSemantic: A dataset for semantic learning of forest from close-range sensing. *Geo-Spat. Inf. Sci.* **2025**, *28*, 185–211. https://doi.org/10.1080/10095020.2024.2313325.
77. Wielgosz, M.; Puliti, S.; Xiang, B.; Schindler, K.; Astrup, R. SegmentAnyTree: A sensor and platform agnostic deep learning model for tree segmentation using laser scanning data. *Remote Sens. Environ.* **2024**, *313*, 114367. https://doi.org/10.1016/j.rse.2024.114367.
78. Henrich, J.; van Delden, J.; Seidel, D.; Kneib, T.; Ecker, A.S. TreeLearn: A deep learning method for segmenting individual trees from ground-based LiDAR forest point clouds. *Ecol. Inform.* **2024**, *84*, 102888. https://doi.org/10.1016/j.ecoinf.2024.102888.
79. Federal, P. *3D Fuels: Hierarchically Scaled Datasets for Three-Dimensional Wildland Fuel Characterization*; Springer: Berlin/Heidelberg, Germany, 2024.

80. Zhong, Y.; Liu, S.; Sun, H. A 3D point cloud instance segmentation network for extracting individual trees from complex forest scenes. *Comput. Electron. Agric.* **2026**, *242*, 111333. https://doi.org/10.1016/j.compag.2025.111333.
81. Xiang, B.; Wielgosz, M.; Kontogianni, T.; Peters, T.; Puliti, S.; Astrup, R.; Schindler, K. Automated forest inventory: Analysis of high-density airborne LiDAR point clouds with 3D deep learning. *Remote Sens. Environ.* **2024**, *305*, 114078. https://doi.org/10.1016/j.rse.2024.114078.
82. Zhang, Q.; Cai, S.; Liang, X. Individual tree segmentation in occluded complex forest stands through ellipsoid directional searching and point compensation. *For. Ecosyst.* **2024**, *11*, 100238. https://doi.org/10.1016/j.fecs.2024.100238.
83. Krisanski, S.; Taskhiri, M.S.; Gonzalez Aracil, S.; Herries, D.; Turner, P. Sensor Agnostic Semantic Segmentation of Structurally Diverse and Complex Forest Point Clouds Using Deep Learning. *Remote Sens.* **2021**, *13*, 1413. https://doi.org/10.3390/rs13081413.
84. Holvoet, J.; Eichhorn, M.P.; Giannetti, F.; Kükenbrink, D.; Liang, X.; Mokroš, M.; Novotný, J.; Pitkänen, T.P.; Puliti, S.; Skudnik, M.; et al. Terrestrial and mobile laser scanning for national forest inventories: From theory to implementation. *Remote Sens. Environ.* **2025**, *329*, 114947. https://doi.org/10.1016/j.rse.2025.114947.
85. Lu, X.; Wang, R.; Zhang, H.; Zhou, J.; Yun, T. PosE-Enhanced Point Transformer with Local Surface Features (LSF) for Wood–Leaf Separation. *Forests* **2024**, *15*, 2244.
86. Lu, X.; Guo, Q.; Li, W.; Flanagan, J. A bottom-up approach to segment individual deciduous trees using leaf-off lidar point cloud data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *94*, 1–12. https://doi.org/10.1016/j.isprsjprs.2014.03.014.
87. Tutland, N.J.; Wion, A.P.; May, C.J.; Hutchings, G.C.; Nowak, H.A.; Gattiker, J.R.; Hiers, J.K.; Linn, R.R.; Pokswinski, S.M.; Margolis, E.Q. Representing 3-dimensional fuels for physics-based fire behavior models: A general framework and case study in a type-converted post-fire shrubfield. *Fire Ecol.* **2025**, *21*, 43. https://doi.org/10.1186/s42408-025-00383-2.
88. Hanousek, T.; Novotný, J.; Navrátilová, B.; Švik, M.; Krejza, J.; Janoutová, R. Complete workflow for detailed 3D forest reconstruction: From terrestrial laser scanning to complex 3D radiative transfer modelling. *Silico Plants* **2025**, *7*, diaf019. https://doi.org/10.1093/insilicoplants/diaf019.
89. Bester, M.S. *The Burning Bush: Linking LiDAR-Derived Shrub Architecture to Flammability*; West Virginia University: Morgantown, WV, USA, 2022.
90. Cova, G.R.; Prichard, S.J.; Rowell, E.; Drye, B.; Eagle, P.; Kennedy, M.C.; Nemens, D.G. Evaluating Close-Range Photogrammetry for 3D Understory Fuel Characterization and Biomass Prediction in Pine Forests. *Remote Sens.* **2023**, *15*, 4837.
91. Hudak, A.T.; Kato, A.; Bright, B.C.; Loudermilk, E.L.; Hawley, C.; Restaino, J.C.; Ottmar, R.D.; Prata, G.A.; Cabo, C.; Prichard, S.J.; et al. Towards Spatially Explicit Quantification of Pre- and Postfire Fuels and Fuel Consumption from Traditional and Point Cloud Measurements. *For. Sci.* **2020**, *66*, 428–442. https://doi.org/10.1093/forsci/fxz085.
92. Marcozzi, A.A.; Johnson, J.V.; Parsons, R.A.; Flanary, S.J.; Seielstad, C.A.; Downs, J.Z. Application of LiDAR Derived Fuel Cells to Wildfire Modeling at Laboratory Scale. *Fire* **2023**, *6*, 394. https://doi.org/10.3390/fire6100394.
93. Rowell, E.; Loudermilk, E.L.; Seielstad, C.; O'Brien, J.J. Using Simulated 3D Surface Fuelbeds and Terrestrial Laser Scan Data to Develop Inputs to Fire Behavior Models. *Can. J. Remote Sens.* **2016**, *42*, 443–459. https://doi.org/10.1080/07038992.2016.1220827.