Multi-branch Deep Learning Architecture for bathymetric LiDAR Point Cloud Classification

Lida Asgharian Pournodrati [1], Gottfried Mandlburger[2], Uwe Soergel [1]

[1] Institute for Photogrammetry and Geoinformatics, University of Stuttgart, Germany – {lida.asgharian-pournodrati, uwe.soergel @ifp.uni-stuttgart.de}
[2] Department of Geodesy and Geoinformation, TU Wien, Vienna, Austria– {gottfried.mandlburger@geo.tuwien.ac.at}

Keywords: Feature Fusion, Laser Bathymetry, Semantic Labelling, Coastal Zone.

Abstract

Accurate classification of topo-bathymetric LiDAR data remains challenging due to the heterogeneous nature of land-water transitional environments, where terrestrial, water surface, and submerged features must be distinguished simultaneously. This study presents a multi-branch deep learning architecture for classifying bathymetric LiDAR data into different classes: soil ground, trees and vegetation, water surface, seabed, aquatic plants and other underwater objects (dead wood, coral reef). The proposed framework employs three parallel feature extraction branches, while the first branch captures spatial structure by focusing on three-dimensional geometric coordinates (XYZ), the other two branches use two independent 1D U-Net architectures to extract signal-based features from RGB spectral reflectance and waveform-derived attributes (intensity, return number, number of returns). The discrete LiDAR attributes, though represented as point-wise numerical values, preserve signal characteristics derived from full-waveform analysis. The encoder-decoder of 1D U-Net architecture with skip connections effectively captures sequential patterns and multi-return patterns in different classes especially in vegetation canopies. The three feature streams are fused through fully-connected layers before final classification. Evaluation using different metrics demonstrates the capability of the framework to simultaneously classify diverse coastal zone and inland waters contexts spanning terrestrial and submerged domains within a unified processing pipeline, eliminating the need for separate terrestrial and bathymetric classification workflows.

1. Introduction

Airborne green wavelength LiDAR (532 nm) (Philpot, 2019) offers a critical advantage over near-infrared systems: it can simultaneously map terrestrial and underwater surfaces in a single flight mission. This dual capability makes it invaluable for coastal zones, river corridors, and wetlands where understanding both above-water and below-water features is essential. However, this versatility creates significant classification challenges. A single survey captures diverse targets across different physical contexts such as terrestrial surfaces, water surfaces, submerged vegetation, underwater terrain, land vegetation, and spurious returns from suspended particles. Successfully processing this data requires methods that can distinguish not only between different materials but also between different environmental contexts: aerial, at the surface, and submerged. To address this complexity, all green LiDAR systems record the complete temporal profile of backscattered laser pulses as full waveforms, providing significantly more information than peak-detection-only systems (Wang et al., 2015).

Although the delivered LiDAR 3d point cloud consists of discrete numerical attributes, these discrete values inherently preserve characteristics of the underlying continuous signals captured during full-waveform acquisition. The intensity values are amplitude measurements sampled from the returned waveform, while return sequencing parameters encode the temporal structure of waveforms with multiple peaks. Most modern bathymetric LiDAR sensors are additionally equipped with an RGB camera for the purpose of point cloud colorization. However, the color information is rarely used for classification. LiDAR attributes exhibit sequential and correlated characteristics when examined spatially. Within local neighborhoods, intensity values vary smoothly based on surface properties, with terrestrial features showing higher values than attenuated aquatic returns. Return characteristics (Tseng et al., 2015) show spatial autocorrelation: dense tree canopies exhibit multiple returns (3-5 per pulse) through vertical layering, soil ground produces single returns, and bathymetric points show structured patterns (surface, water column noise, seabed).

This signal-like behavior motivates applying signal processing architectures for feature extraction. One-dimensional U-Net architectures (Stoller et al., 2019), with their encoder-decoder structure and skip connections, effectively capture multi-scale patterns in data. Although 2D U-Net originally developed for image segmentation (Ronneberger et al., 2015), the U-Net’s hierarchical encoder–decoder architecture has proven adaptable to one-dimensional signal analysis. By treating LiDAR attributes as one-dimensional signals, 1D U-Net can extract features capturing both local patterns and broader context. The 1D convolutions learn filters detecting diagnostic patterns abrupt intensity changes at land-water boundaries, return sequences distinguishing vegetation types, and spectral signatures identifying materials through hierarchical feature learning and multi-scale information preservation.

Despite the rich information available from full-waveform green laser LiDAR, most existing classification approaches either process attributes independently (Diab et al., 2022; Tarsha Kurdi et al., 2023) or rely primarily on geometric features while underutilizing radiometric and return characteristics (Azadbakht et al., 2016; Li et al., 2016; Mandlburger et al., 2015; Yan and Shaker, 2012). The challenge of simultaneously classifying terrestrial and aquatic features, soil ground, terrestrial vegetation (trees and other plants), water surfaces, underwater features (seabed), submerged aquatic vegetation, and water column noise within a unified framework remains largely unaddressed in the literature. Multi-branch architectures in deep learning employ multiple independent network streams, each processing a different input modality, whose outputs are combined into a unified classification decision. Building on this principle, this

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-81-2026

study addresses these gaps by proposing a multi-branch deep learning architecture that: (1) separately processes geometric, color, and waveform-derived attributes through specialized feature extraction networks tailored to their respective information content, (2) employs point cloud neural networks PointMLP (Ma et al., 2022) for geometry learning and 1D U-Net architectures for signal-based attribute processing that recognizes the sequential and correlated nature of intensity and return characteristics, and (3) fuses these complementary feature representations for robust topo-bathymetric classification spanning terrestrial vegetation, water surface, seabed, and aquatic plants. The paper is structured as follows: Section 2 reviews related work in point cloud classification and bathymetric LiDAR processing. Section 3 describes the proposed multi-branch architecture incorporating PointMLP for geometry and dual 1D U-Net branches for radiometric and waveform-derived attributes. Section 4 presents experimental results and performance evaluation. Section 5 provides concluding remarks.

2. Related Work

2.1 Bathymetric LiDAR Classification

Bathymetric LiDAR point cloud classification addresses the fundamental challenge of separating water surface, column, bottom, and noise returns in shallow aquatic environments through various computational approaches. Kim et al. (2023) developed a pseudo-waveform technique that reconstructs waveform-like patterns from discrete point clouds by organizing points spatially and applying Gaussian decomposition to identify distinct water layers, offering an advantage of working without full waveform data but requiring extensive parameter tuning and potentially failing when surface and bottom returns are temporally close in very shallow conditions. Janowski et al. (2022) employed a geographic object-based image analysis framework combined with traditional machine learning classifiers that extract geomorphometric features like terrain curvature and ridge characteristics to automatically map seabed morphology and structures, demonstrating robust performance for geomorphological classification but depending heavily on handcrafted feature engineering and potentially struggling with novel seabed types not represented in training data. Song and Kim (2024) applied PointNet architecture (Qi et al., 2017) directly to bathymetric point clouds for end-to-end classification, eliminating manual feature design through learned representations but encountering difficulties distinguishing between closely related classes like surface and water column points, and requiring substantial computational resources for training. An emerging approach (Paul et al., 2025) using PointCNN (Li et al., 2018) for depth bias correction leverages semantic segmentation of water levels as a preprocessing step for bathymetric accuracy enhancement, showing promise for purely point-based deep learning but remaining relatively unexplored compared to traditional waveform processing methods and facing data scarcity challenges for training robust models. Asgharian et al. (2025) introduced a multi-scale transformer for classifying bathymetric LiDAR point clouds, using voxel-based embeddings and windowed self-attention with HDBSCAN clustering to reduce computational load. The model captures both local and global spatial relations. Rhomberg-Kauert et al. (2024) used UMAP dimensionality reduction combined with clustering algorithms to detect submerged macrophyte vegetation based on structural characteristics extracted from point clouds, representing a departure from traditional supervised learning by leveraging unsupervised methods commonly used in other remote sensing domains, though the approach requires careful selection of feature combinations and clustering parameters.

2.2 Multi-Branch Deep Learning Architectures

Multi-branch neural network architectures represent a promising approach for point cloud classification and segmentation tasks by enabling parallel processing of heterogeneous data sources or multi-view representations. Schilling et al. (2018) proposed a multi-branch fully convolutional network that fuses elevation and optical data through a pseudo-Siamese architecture, with separate branches for object recognition and high-resolution segmentation, demonstrating effective multi-sensor integration for vehicle detection in remote sensing applications. Vaquero et al. (2020) developed a dual-branch CNN for LiDAR-only vehicle detection that processes two different view representations independently before fusion, with their dual-view approach consistently outperforming single-branch methods on KITTI benchmarks and effectively handling occlusions through probabilistic bounding box growing. Zhao et al. (2018) introduced a multi-scale convolutional neural network that transforms point clouds into multi-scale contextual images for hierarchical feature learning, eliminating manual feature engineering while capturing both fine-grained local details and broader contextual information across multiple scales for airborne LiDAR applications. Feng et al. (2019) extended multi-branch fusion to hyperspectral and LiDAR data integration through a modified two-branch CNN with an adaptive fusion module that re-weights features based on importance rather than simple concatenation, achieving significant accuracy improvements in urban land-use mapping. Yu et al. (2023) addressed the computational efficiency challenge by proposing a dynamic multi-branch module utilizing structural re-parameterization to maintain rich multi-branch feature learning during training while collapsing to an efficient single-branch structure during inference, avoiding the accuracy loss inherent in traditional pruning or quantization approaches. Yuan et al. (2025) designed a multi-branch feature extractor combining neighborhood embedding and point transformer blocks for transfer learning on airborne LiDAR data with limited training samples, addressing point density variations and domain transfer challenges while significantly reducing annotation requirements.

2.3 U-Net and Encoder-Decoder Architectures

U-Net and encoder-decoder architectures have evolved significantly for point cloud semantic segmentation, balancing local detail preservation with global context understanding. Ronneberger et al. (2015) introduced the foundational U-Net with its symmetric encoder-decoder design featuring skip connections that enable precise localization while capturing hierarchical context, proving particularly effective for limited training data scenarios despite memory overhead from feature concatenation. Building upon this foundation, Biasutti et al. (2019) proposed LU-Net, which adapts the U-Net framework for 3D LiDAR data by projecting point clouds into 2D representations and learning 3D features end-to-end, achieving computational efficiency at the cost of some geometric information loss during projection. Deng et al. (2025) proposed 1D-CloudNet, a nested 1D U-Net for pixel-level cloud classification from multi-channel satellite sequences, demonstrating that encoder-decoder architectures effectively capture spatial correlations in one-dimensional remote sensing signals. Russo et al. (2021) applied a 1D encoder-decoder with skip connections to denoise underwater inertial sensor sequences, confirming U-Net effectiveness beyond image domains. Yan et al. (2021) used 1D U-Net on sonar backscatter sequences for seabed detection, demonstrating robust sequential signal processing in underwater acoustic sensing.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-81-2026 | © Author(s) 2026. CC BY 4.0 License.

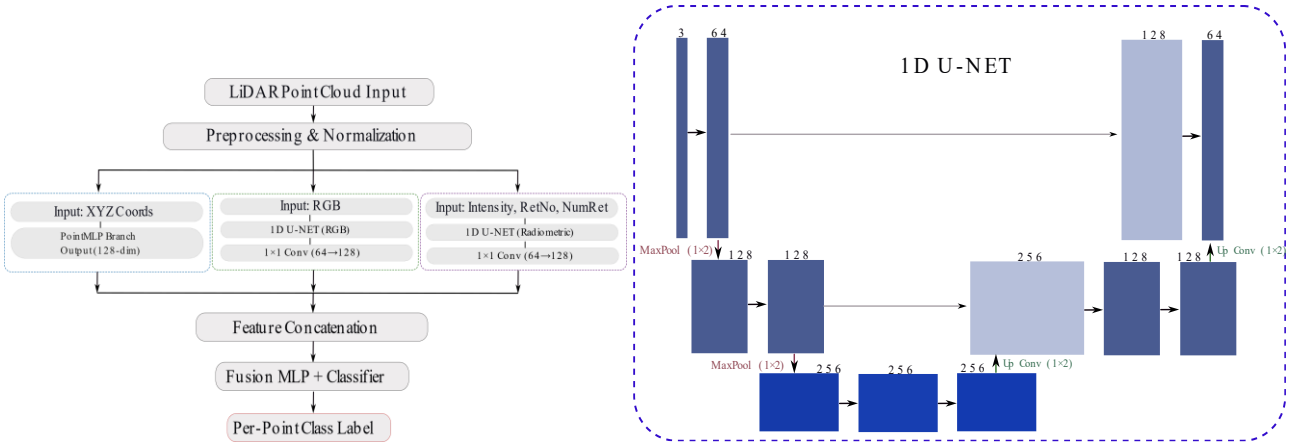

Figure 1. Proposed multi-branch architecture. (a) Overall pipeline showing three parallel branches and fusion module. (b) Detailed 1D U-Net architecture shared by RGB and radiometric branches.

3. Methods

This study proposes a multi-branch deep learning architecture that integrates geometric, spectral, and waveform-derived features from airborne LiDAR data for bathymetric data classification. Figure 1 illustrates pipeline of the proposed method. The architecture comprises three parallel feature extraction branches that independently process different attribute modalities before fusion for final classification.

3.1.1 Geometric Feature Extraction

The geometric branch employs PointMLP (Ma et al., 2022), a point-based deep learning architecture specifically designed for 3D point cloud processing. PointMLP operates directly on unordered point sets, extracting hierarchical spatial features from raw XYZ coordinates through progressive geometric abstraction layers.

This branch processes the three-dimensional spatial coordinates XYZ of the point cloud. Prior to network input, the coordinates undergo normalization. The PointMLP architecture consists of several key components. The embedding layer transforms the 3-channel input (XYZ) into a 64-dimensional feature space through 1D convolution, batch normalization, and ReLU activation. This is succeeded by successive point-wise convolutional layers that expand the feature dimensionality to 128 and then 256 channels, enabling the extraction of higher-level geometric features.

The architecture preserves per-point information throughout the forward pass. A final projection layer reduces the 256-dimensional features to 128 dimensions, producing point-level descriptors that encode fine-grained geometric relationships. These 128-dimensional features are subsequently used in the fusion module together with features from the other modality branches.

3.1.2 1D U-Net for Signal Processing

While PointMLP excels at geometric feature extraction, spectral and radiometric attributes from LiDAR systems exhibit signal-like characteristics that benefit from different architectural approaches. This study employs two parallel 1D U-Net architectures to process these complementary information sources.

U-Net, originally developed for biomedical image segmentation (Ronneberger et al., 2015), features a symmetric encoder-decoder architecture with skip connections that preserve spatial information across resolution scales. The 1D variant adapts this architecture for sequential signal processing, making it particularly suitable for LiDAR full-waveform attributes.

The input to each 1D U-Net branch is a sequence of points sampled from a contiguous spatial region of the point cloud, where each point contributes its modality-specific attributes, either RGB spectral values or waveform-derived properties, as one element in the sequence. The channel dimension carries the attribute values while the sequence dimension spans the sampled points within the region. This design allows the 1D convolution to learn attribute co-occurrence patterns across spatially proximate points and effectively decouples the attribute processing from the three-dimensional geometry, which is handled separately by the PointMLP branch.

The 1D U-Net architecture consists of three main components: *Encoder Path*: The encoder progressively extracts hierarchical features through convolutional downsampling. The first encoding block applies a one-dimensional convolution with kernel size 3 and 64 filters, followed by batch normalization and ReLU activation, transforming the input channels into 64-dimensional feature maps. A second encoding block applies a one-dimensional convolution with kernel size 3, doubling the channel dimension to 128. Max pooling with stride 2 reduces information between encoding stages, creating a multi-scale feature hierarchy.

Bottleneck: At the deepest level, the bottleneck layer expands features to 256 dimensions, capturing the most abstract representations of the input signal characteristics.

Decoder Path: The decoder reconstructs features through transposed convolutions that progressively upsample the feature maps. At each decoding stage, skip connections concatenate corresponding encoder features with upsampled decoder features, preserving fine-grained information lost during downsampling. Two decoding blocks mirror the encoder structure, reducing dimensions from 256 → 128 → 64 while recovering the original resolution.

In our implementation, the 1D U-Net outputs per-point features, maintaining the spatial correspondence of each sample. These 64-dimensional features are then projected to 128 dimensions

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-81-2026

through a 1×1 convolution before being passed to the multi-modal fusion stage.

3.1.3 Radiometric and Spectral Branches

Color information provides critical discriminative power for vegetation classification, as different plant species exhibit distinct spectral signatures due to variations in chlorophyll content, leaf structure. The 1D U-Net processes these spectral channels as correlated signals, learning hierarchical color patterns and waveform-derived features that can efficiently distinguish different classes.

RGB Feature Branch: The first branch extracts feature from the RGB information of the point cloud and employs 64 base channels, producing 64-dimensional feature maps through the encoder-decoder process. These features are then globally aggregated and projected to 128 dimensions through a 1×1 convolution (linear layer) to ensure dimensional consistency with other modality branches before fusion.

Waveform-Derived Property Branch: The second branch processes features derived from LiDAR full-waveform attributes, using three input channels: *Intensity*: Represents the amplitude of the returned laser pulse, influenced by surface reflectivity, range, incidence angle, and atmospheric conditions. *Return Number*: Denotes the sequence of a return within a multi-return pulse. Early returns typically correspond to canopy tops or water surfaces, whereas later returns arise from deeper vegetation or ground reflections. *Number of Returns*: Indicates the total number of discrete echoes detected from a single emitted pulse, reflecting structural complexity. Dense vegetation canopies and water areas often yield multiple returns, while bare ground typically produce a single return.

This branch uses the same 1D U-Net architecture as the RGB branch (with 64 base channels) and similarly projects its 64-dimensional output to 128 dimensions to maintain uniform feature dimensionality for the multi-modal fusion stage.

3.1.4 Multi-Branch Fusion and Classification

The three feature branches operate independently during forward propagation, with each specializing in its respective data modality. The 128-dimensional feature vectors from all three branches are concatenated into a 384-dimensional combined representation, which is passed to a fusion network designed to learn optimal feature integration.

The fusion module consists of two fully connected layers, each followed by batch normalization, ReLU activation, and dropout regularization with rates of 0.5 and 0.3, respectively. The first layer compresses the 384-dimensional input to 256 dimensions, and the second layer further reduces it to 128 dimensions, forming a compact unified representation that integrates geometric, color and waveform information.

A final linear classification layer maps the 128-dimensional fused features to class logits for the target categories. This parallel fusion strategy allows each branch to develop specialized feature extractors independently, avoiding premature feature entanglement and preserving modality-specific information.

The waveform-derived and color branches both share the same 1D U-Net configuration with 64 base channels, and their 64-dimensional outputs are projected to 128 dimensions for compatibility in the fusion stage.

3.1.5 Loss Function

The model employs a class-weighted focal loss (Lin et al., 2017) as the optimization objective for multi-class classification, extending the standard cross-entropy formulation to handle class imbalance and emphasize hard-to-classify examples. The focal loss is defined as

$$L_{focal} = -\alpha_c (1 - p_c)^{\gamma} \log(p_c) \quad (1)$$

Where p_c is the predicted probability of the true class $\boldsymbol{c}$, α_c is the class-specific weight, and γ is the focusing parameter that reduces the relative loss contribution of well-classified points. Although balanced spatial regions are used during dataset preparation to ensure equal class representation, the weighted focal loss further compensates for residual imbalance within sampled regions (Phan and Yamamoto, 2020). The loss is computed on the fused logits after integration of spatial, waveform-derived and color feature branches, ensuring end-to-end optimization of all network components. Model parameters are updated through backpropagation using the Adam optimizer with a learning rate of 0.001 and gradient clipping to maintain training stability.

4. Results

To evaluate cross-sensor transferability, this study applies the proposed method to two datasets representing different aquatic environments and acquired by different laser scanning platforms. The multi-branch architecture demonstrates inherent flexibility: while the PointMLP branch consistently processes XYZ geometry, the parallel 1D U-Net branches adapt their input channels to match the sensor-specific attribute suite whether conventional intensity and return characteristics or alternative parameters such as amplitude, reflectance, and deviation (Pfennigbauer and Ullrich, 2010). The specific attributes employed for each dataset are detailed in subsequent sections, demonstrating the capability of the framework to adapt to different sensor outputs.

4.1 Training Configuration

All models are trained on a single NVIDIA RTX 4090 GPU with 28 GB RAM. Due to severe class imbalance in topo-bathymetric LiDAR data where majority classes such as water surface and seabed constitute over 90% of points while minority classes such as aquatic vegetation and noise represent less than 3% training employs balanced spatial region sampling to ensure equal class representation per epoch and prevent the model from being dominated by majority classes. For the Stavanger dataset, the model converged at epoch 4 with early stopping triggered at epoch 19, requiring 2.3 minutes total training time. For the Loosdorf dataset, convergence occurred at epoch 6 with early stopping at epoch 21, requiring 0.2 minutes. Evaluation is performed on all sequential regions of the complete held-out test sets without any sampling, ensuring unbiased assessment over 16.1 million and 230,985 points for Stavanger and Loosdorf respectively, achieving inference speeds of approximately 2.3 million and 646,000 points per second. Overall accuracy is computed as the mean of per-class recall values, where each class contributes equally regardless of its size, and per-class accuracy represents the recall of each individual class.

The model is trained end-to-end using the Adam optimizer with learning rate 0.001 and L2 regularization (weight decay 10^{-4}). Additionally, gradient clipping with maximum norm 1.0 is applied to prevent gradient explosion, particularly important for this multi-branch architecture where gradients flow through

parallel networks before converging at the fusion layer. The learning rate follows a step decay schedule, multiplying by 0.7 every 10 epochs to enable fine-grained convergence in later training stages. To mitigate class imbalance, training is performed on balanced spatial regions, ensuring equal representation of all classes in each mini-batch. For every class, an equal number of 3D point samples are selected from distinct spatial regions, maintaining consistent class distribution across epochs. Each sample consists of 1024 points, obtained via sampling with replacement for classes containing fewer points.

4.2 Evaluation Metrics

Model performance is assessed using multiple metrics (Sokolova et al., 2006): overall accuracy (OA), recall (Rc), precision (Pr) and F1-score. Overall accuracy is computed as the mean of per-class recall values as defined in Equation 2, ensuring equal contribution from each class regardless of its frequency in the test set. Per-class accuracy represents the recall of each individual class. As evaluation regions maintain equal class representation, overall accuracy equals the mean of per-class accuracies. Recall indicates the ability to identify all relevant instances within each class. Precision assesses the proportion of correctly predicted instances among all positive predictions. F1-score balances recall and precision to provide a single performance metric calculated as:

$$Pr_i = \frac{TP_i}{TP_i + FP_i}, Rc_i = \frac{TP_i}{TP_i + FN_i}, F1_i = \frac{2.Pr_i.Rc_i}{Pr_i + Rc_i},$$

$$OA = \frac{1}{C}\sum_{i=1}^{C} \frac{TP_i}{TP_i + FN_i} \quad (2)$$

where TP_c, FP_c, and FN_c represent true positives, false positives, and false negatives for class i (i=1, …, C).

4.3 Dataset Structure and Discussion

The point clouds used in this study are refraction-corrected products delivered by the respective sensor processing pipelines. While refraction correction requires prior identification of water surface points, the unified classification framework remains advantageous as it simultaneously handles terrestrial, surface, and submerged classes without requiring separate preprocessing workflows for each domain.

4.3.1 Stavanger Dataset

The study utilizes airborne LiDAR point cloud data captured by Teledyne Optech CZMIL SuperNova laser scanner from an island called Fjøløy, in Stavanger municipality. The dataset is split into training (41.0 billion points) and testing (16.2 billion points) sets. The classification task focuses on five coastal zone classes: soil ground (5.1% of training data), Aquatic plants (3.3%), noise (0.9%), seabed (35.1%), and water surface (55.4%). The dataset exhibits substantial class imbalance, with water surface representing the majority class and noise being severely underrepresented at less than 1% of total training points. Figure 2 demonstrates the train and test parts of the Stavanger dataset. For the Stavanger dataset, the point cloud is spatially divided into 70% training and 30% test regions. From the training region, 15% is reserved as a validation set for early stopping and model selection with patience of 15 epochs. The test region remains completely unseen until final evaluation, ensuring unbiased performance reporting.

Table 1 represent the confusion matrix of the Stavanger dataset. The confusion matrix shows two main error patterns. The biggest problem is the noise class being extremely difficult to classify due to its omnipresent spatial distribution throughout the entire scene (underwater, at water surface level, above water, and mixed with seabed), resulting in only 30.4% recall and 8.6% precision. The largest confusion for noise is with water surface class (40.2% of noise misclassified as water surface), followed by seabed (14.7%), and soil ground (8.1%), showing that noise appears at all vertical levels and the model cannot learn consistent patterns. The second major error pattern is aquatic vegetation, which achieves only 55.1% recall, with 37.8% confused with seabed due to its close vertical proximity and similar geometric positioning at the bottom of the water column, and 5.4% confused with noise. The soil ground class achieves an excellent 96.7% recall, indicating that land features are well separated from underwater features. In contrast, underwater classes (vegetation and noise) and those near the water interface remain challenging due to extreme class imbalance (noise: 0.5%, vegetation: 2.3% in the test set versus water: 59.7%, seabed: 34.5%), the pervasive distribution of noise across spatial regions, and limited training examples to capture the high variability of these rare classes using the current features (XYZ, RGB, and single-return intensity). The overall accuracy and mean IoU for this dataset are OA=75.3% and mIoU=64.3%, respectively, with the high OA dominated by the abundant majority classes (water, seabed) and the lower mIoU revealing the severe struggles with minority classes.

Improving classification of minority classes such as noise and aquatic vegetation remains an open challenge. Several strategies could address this in future work: targeted data augmentation specifically for noise and vegetation points, such as synthetic point generation or geometric perturbation within minority class regions; more aggressive class-specific sampling strategies that further oversample rare classes relative to majority classes; and specialized loss formulations beyond focal loss, such as class-balanced softmax or contrastive losses that explicitly penalize confusion between geometrically similar but semantically distinct classes.

4.3.2 Loosdorf Dataset

In the second experimental scenario, we evaluate generalization capability of the model using two other pond datasets acquired by a *RIEGL* VQ-840-G laser scanner, which provides a different suite of LiDAR attributes compared to the primary dataset. Detailed specifications of these datasets, including acquisition parameters and site characteristics, are provided by Rhomberg-Kauert et al. (2024). These datasets lack RGB spectral information and employ alternative wave-derived parameters: *amplitude* (representing the peak intensity of the return signal), *reflectance* (quantifying the proportion of incident laser energy returned to the sensor), and *deviation* (indicating the temporal width or spread of the reflected laser pulse, describing surface roughness or scattering complexity), in place of the conventional intensity, return number, and number of returns attributes (Pfennigbauer and Ullrich, 2010). To assess transferability, one pond dataset serves as the training set with 15% of its points reserved for validation, and the second pond functions as the independent test set (Figure 3).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-81-2026 | © Author(s) 2026. CC BY 4.0 License.

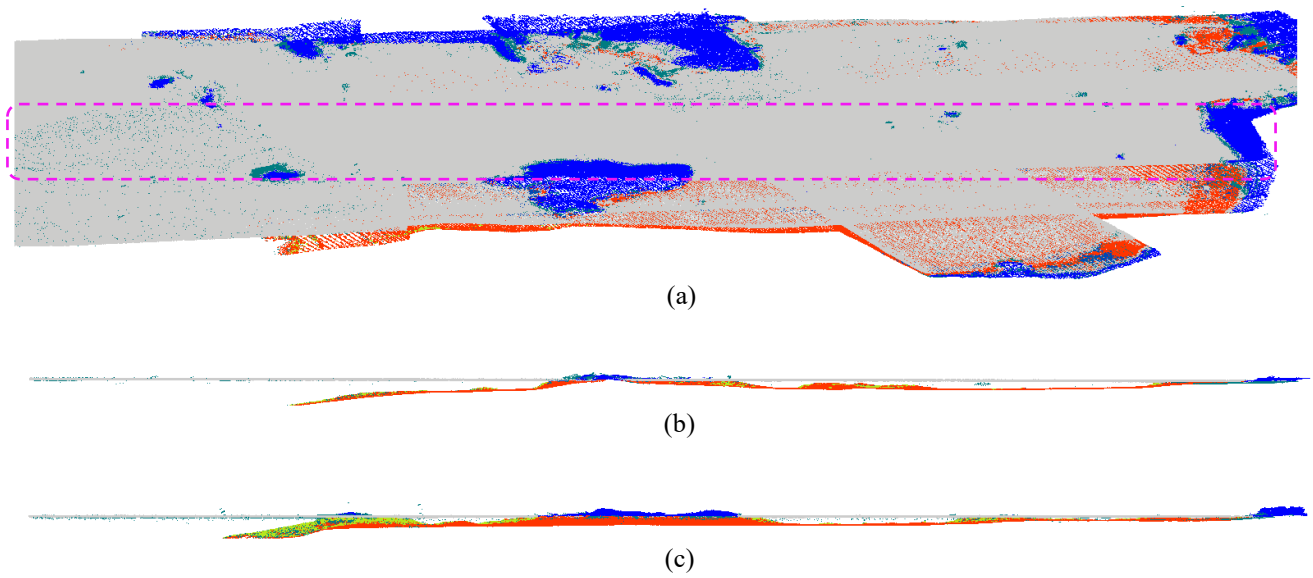

Figure 2. Train and Test sets of the Stavanger dataset that covers approximately 1.6 km × 1.8 km with vertical range of 145 m. (a) Perspective view of the whole dataset, purple dashed box displays the test set, and the other regions outside of the dashed box show the train set; (b) Cross-section view of the annotated train point clouds, water surface (gray), aquatic vegetation (light green), seabed (red), noise (dark green), ground (blue); (c) Labelled ground truth of the test set from cross-section view to better illustrate under water region.

Categories	Soil Ground	Aquatic Vegetation	Noise	Seabed	Water Surface
Soil Ground	**96.7**	0.1	0.4	0	2.8
Aquatic Vegetation	0.6	**55.1**	5.4	37.8	1.1
Noise	8.1	6.6	30.4	14.7	**40.2**
Seabed	0.3	2.3	0.2	**97.2**	0
Water Surface	0.4	0.3	2.2	0	**97.1**
Precision	87.4	56.5	8.6	97.2	99.4
Recall	96.7	55.1	30.4	97.2	97.1
F1	91.8	55.8	13	97.2	98.2

Table 1. Confusion matrix of Stavanger dataset. Precision, recall, F1 score are reported for each class.

This configuration evaluates whether features learned from one aquatic environment can generalize to classify region of an unseen pond. Despite their proximity in Loosdorf, Austria, the two ponds differ aquatic vegetation distribution (Rhomberg-Kauert et al., 2024). The balanced region sampling approach achieved an overall accuracy of 78.56% and mIoU of 59.6%, demonstrating reasonable performance on distinct classes with water achieving 97.1% recall and Trees achieving 98.1% recall (Table 2). However, the method revealed critical failures in separating geometrically mixed classes, most notably aquatic vegetation class which exhibited precision collapse at 20.3%. The primary confusion occurred because aquatic vegetation in the test pond is positioned closer to the seabed and appears in lower quantities with potentially different species composition compared to the training pond where vegetation samples were more abundant and positioned higher in the water column, causing 25.6% (14,753 points) of true seabed points to be misclassified as vegetation and reducing seabed recall to 62.4% despite high precision of 97.0%. In general, it is demonstrating effective cross-site transfer. The confusion matrix demonstrates overall good performance and reasonable discrimination for three classes and moderate challenges for two classes. It should be noted that this classification was performed using a dual-branch architecture where PointMLP processes XYZ spatial coordinates while a 1D U-Net processes waveform-derived features (amplitude, reflectance, and deviation). No RGB information was available in this dataset, meaning the classification relies solely on geometric and waveform properties without color information that could potentially improve vegetation discrimination.

4.4 Comparison with Baseline

To contextualize the proposed method against established point cloud classification approaches, PointNet++ (Qi et al., 2017) is evaluated on both datasets using identical XYZ geometric features as input, providing a geometry-only deep learning baseline. As shown in Tables 3 and 4, the proposed multi-branch architecture consistently outperforms PointNet++ across both datasets. On the Stavanger dataset, the proposed method achieves OA=75.3% and mIoU=64.3% compared to OA=38.34% and mIoU=20.1% for PointNet++, representing a gain of +36.96pp in overall accuracy. On the Loosdorf dataset, the proposed method achieves OA=78.56% and mIoU=59.6% compared to OA=53.68% and mIoU=29.5% for PointNet++, a gain of +24.88pp. The performance gap demonstrates the critical contribution of radiometric and spectral attributes, which provide discriminative information that geometry alone cannot capture in topo-bathymetric environments.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-81-2026

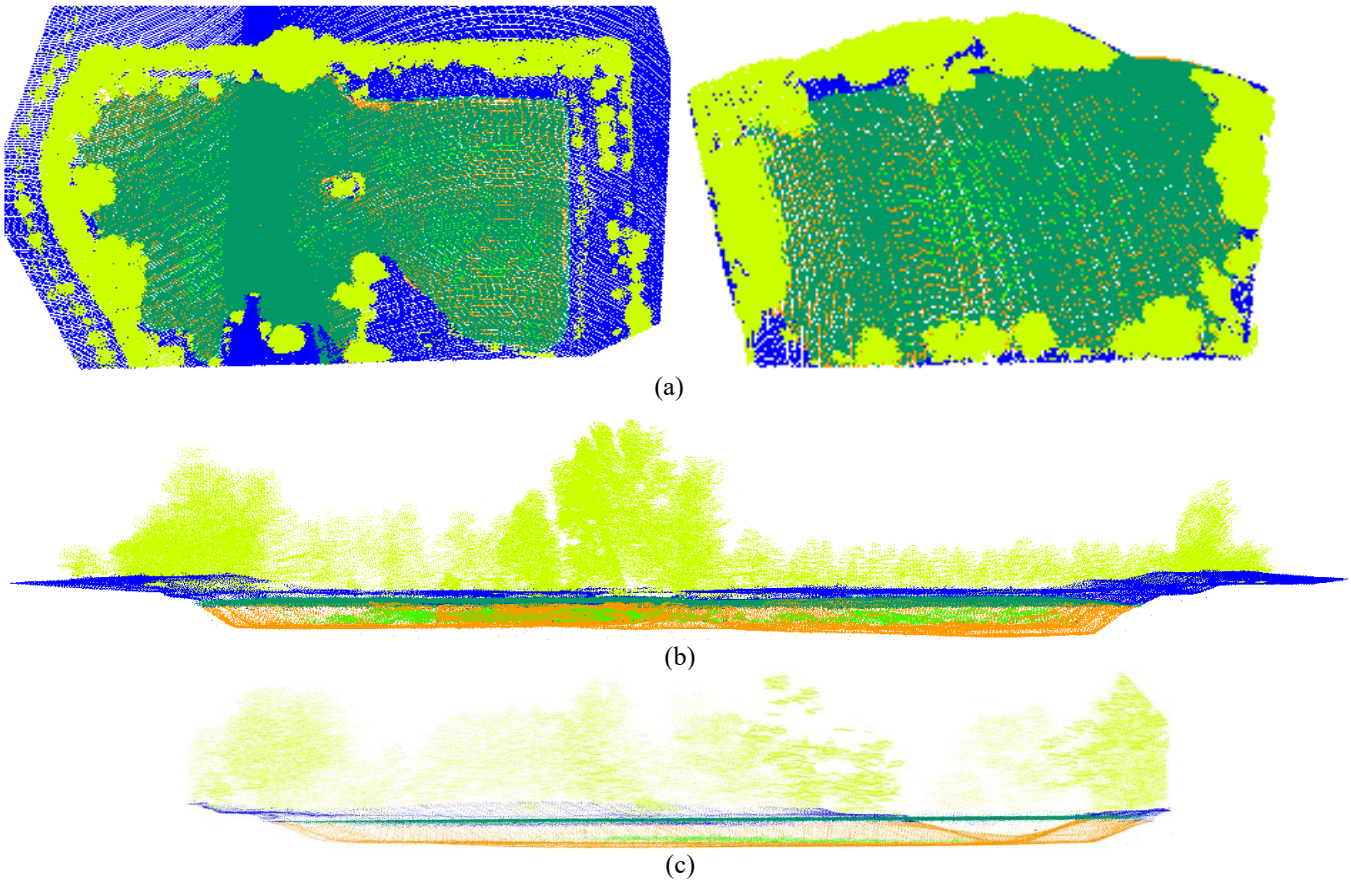

Figure 3. Train and Test sets of the Loosdorf dataset. (a) Perspective view of the train set (covers approximately 200 m × 216 m with vertical range of 30 m) and test set (covers approximately 95 m × 90 m with vertical range of 27 m); (b) Cross-section view of the annotated train point clouds, water surface (dark green), aquatic vegetation (light green), seabed (orange), tree (phosphorescent), ground (blue); (c) Labelled ground truth of the test set.

Categories	Soil Ground	Water Surface	Aquatic Vegetation	Tree	Seabed
Soil Ground	**62.1**	20.3	0.5	15.6	1.5
Water Surface	0.3	**97.1**	0	2.6	0
Aquatic Vegetation	0	9.9	**73.1**	0	17
Tree	1.9	0	0	**98.1**	0
Seabed	5.5	4.7	25.6	1.8	**62.4**
Precision	60.6	88.8	20.3	94	97
Recall	62.1	97.1	73.1	98.1	62.4
F1	56	92.8	31.8	96	75.9

Table 2. Confusion matrix of Loosdorf dataset. Precision, recall, F1 score are reported for each class.

Method	OA (%)	mIoU (%)	Class Accuracy (%)				
			Soil Ground	Water Surface	Noise	AquaticVegetation	Seabed
PointNet++	38.34	20.1	60.5	40	10.2	40.6	40.4
Proposed	75.3	64.3	96.7	55.1	30.4	97.2	97.1

Table 3. Comparison against PointNet++ method for Stavanger dataset.

Method	OA (%)	mIoU (%)	Class Accuracy (%)				
			Soil Ground	Water Surface	AquaticVegetation	Tree	Seabed
PointNet++	53.68	29.5	60.6	52.4	40.3	65.1	50
Proposed	78.56	59.6	62.1	97.1	73.1	98.1	62.4

Table 4. Comparison against PointNet++ method for Loosdorf dataset.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-81-2026

4.5 Ablation Study

In this section, we analyze the impact of each architectural component in the proposed method. To assess their individual contributions, four configurations are defined: Method A represents the baseline PointMLP architecture, Method B combines PointMLP with the waveform-derived U-Net, Method C combines PointMLP with RGB U-Net and Method D integrates all components PointMLP, U-Net (waveform), and U-Net (RGB).

The ablation study on Stavanger dataset (Table 5) demonstrates that spatial features alone (PointMLP) yield a low baseline accuracy of 38.98%, with severe misclassification of vegetation, noise, and water classes. Incorporating signal information through the Intensity U-Net significantly boosts performance by +24.86 percentage points, particularly improving aquatic vegetation detection (50.2% accuracy). Adding color features via the RGB U-Net provides an additional 11.46% gain, mainly enhancing water classification (50.8% accuracy). Although the combined model achieves a substantial overall improvement of 36.32% in accuracy, noise remains the most challenging class with only 30.4% accuracy. Overall, waveform-derived features contribute more to classification accuracy than color, underscoring their critical role in distinguishing complex terrain classes.

The ablation study on Loosdorf dataset (Table 6) reveals that incorporating waveform-derived features (amplitude, reflectance and deviation) into the PointMLP baseline substantially improves classification performance on *RIEGL* data. The baseline model relying solely on spatial features achieves 64.96% overall accuracy and 34% mIoU, performing particularly poorly on the seabed class (0%). After adding signal features, overall accuracy rises by +13.6 percentage points to 78.56%, and mIoU improves by +25.6 unit to 59.6%. The enhancement is most notable for the seabed class, which gains +62.4 unit, and for water, which improves by +28 unit, and trees by +15.5 unit. However, the addition of signal information reduces performance for aquatic vegetation (-25.2 unit), indicating sensitivity to feature-domain differences result by the type of underwater vegetation. Overall, radiometric information proves critical for accurate classification, especially for distinguishing seabed, water surfaces and trees, while remaining susceptible to domain shifts between training and testing environments.

Method	OA (%)	mIoU (%)	Class Accuracy (%)				
			Soil Ground	AquaticVegetation	Noise	Seabed	Water Surface
A	38.98	20.2	100	0	0	93.1	1.8
B	63.84	48.1	95.7	50.2	30.9	96.1	46.3
C	59.66	51.8	94.7	13.7	0.4	90.9	98.6
D	75.3	64.3	96.7	55.1	30.4	97.2	97.1

Table 5**.** Impact of successive method components on Stavanger dataset with RGB information.

Method	OA (%)	mIoU (%)	Class Accuracy (%)				
			Soil Ground	Water Surface	AquaticVegetation	Tree	Seabed
A	64.96	34	74.8	69.1	98.3	82.6	0
B	78.56	59.6	62.1	97.1	73.1	98.1	62.4

Table 6. Impact of successive method components on Loosdorf dataset without RGB information.

5. Conclusion

This study presented a multi-branch deep learning architecture for classifying topo-bathymetric LiDAR data. The framework addresses the fundamental challenge of simultaneously distinguishing terrestrial, surface, and submerged features including soil ground, trees and vegetation, water surface, seabed, aquatic plants, and water column noise within a unified processing pipeline. The proposed architecture leverages the complementary nature of different LiDAR attributes through parallel processing streams. PointMLP extracts hierarchical geometric features from three-dimensional spatial coordinates, capturing vertical structures and terrain morphology that distinguish vegetation from flat surfaces. Two independent 1D U-Net branches process RGB spectral reflectance and waveform-derived attributes (intensity, return number, number of returns) as one-dimensional signals, exploiting the sequential and correlated patterns preserved in discrete point cloud data despite their origin from continuous full-waveform recordings. Experimental results demonstrate that this multi-branch fusion approach effectively integrates geometric, spectral, and signal information for robust classification across coastal environments. Extending the framework to handle variable point densities and sensor configurations would enhance its applicability across different LiDAR systems and survey conditions. The evaluation covers spatial train/test splits and cross-site transfer within the same sensor. Generalization across different flight parameters or sensor types remains challenging due to varying waveform characteristics and attribute definitions, which will be addressed in future work.

Future work will focus on comprehensive benchmarking against state-of-the-art point cloud classification architectures. Specifically, evaluating the presented datasets using established methods such as KPConv and Point Transformer would provide critical comparative insights into the relative strengths of different architectural paradigms for bathymetric LiDAR classification. Comparison against traditional machine learning approaches such as Random Forest with handcrafted geometric and radiometric features will also be explored, addressing the performance gap between classical and deep learning methods for topo-bathymetric classification tasks.

Acknowledgements

The authors wish to thank Bernt Larsen, Andreas Mathisen Tanner and Charles de Jongh from Field Geospatial for generously providing access to the Stavanger dataset and for their valuable insights on data acquisition and processing protocols.

We are also grateful to Jan Rhomberg-Kauert from TU Wien for providing the Loosdorf pond datasets and for his technical support.

References

Asgharian Pournodrati, L., Baba, M.M., Gangelhoff, J., Sörgel, U., 2025: Multi-scale Transformer-based classification of bathymetric LiDAR data in shallow water environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48,1-6.

Azadbakht, M., Fraser, C.S., Khoshelham, K., 2016: Improved urban scene classification using full-waveform LiDAR. *Photogrammetric Engineering & Remote Sensing*, 82(12), 973-980.

Biasutti, P., Lepetit, V., Aujol, J.F., Brédif, M., Bugeau, A., 2019: Lu-net: An efficient network for 3d lidar point cloud semantic segmentation based on end-to-end-learned 3d features and u-net. *In Proceedings of the IEEE/CVF international conference on computer vision workshops*.

Deng, M., Han, Y., Liu, Y., Dong, L., Zhou, Q., Zhang, Y., Deng, X., Lu, T., 2025. Development of a Novel One-Dimensional Nested U-Net Cloud-Classification Model (1D-CloudNet). Remote Sensing, 17(3), 519.

Diab, A., Kashef, R., Shaker, A., 2022: Deep learning for LiDAR point cloud classification in remote sensing. *Sensors*, 22(20), 7868.

Feng, Q., Zhu, D., Yang, J., Li, B., 2019: Multisource hyperspectral and LiDAR data fusion for urban land-use mapping based on a modified two-branch convolutional neural network. *ISPRS International Journal of Geo-Information*, 8(1), 28.

Janowski, L., Wroblewski, R., Rucinska, M., Kubowicz-Grajewska, A., Tysiac, P., 2022: Automatic classification and mapping of the seabed using airborne LiDAR bathymetry. *Engineering Geology*, 301, 106615.

Kim, H., Jung, J., Lee, J., Wie, G., 2023: Water Bottom and Surface Classification Algorithm for Bathymetric LiDAR Point Clouds of Very Shallow Waters. *Canadian Journal of Remote Sensing*, 49(1), 2172957.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X. Chen, B., 2018: Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31.

Li, Z., Jupp, D.L., Strahler, A.H., Schaaf, C.B., Howe, G., Hewawasam, K., Douglas, E.S., Chakrabarti, S., Cook, T.A., Paynter, I., Saenz, E.J., 2016: Radiometric calibration of a dual-wavelength, full-waveform terrestrial lidar. *Sensors*, 16(3), 313.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017: Focal loss for dense object detection. *In Proceedings of the IEEE international conference on computer vision*, 2980-2988.

Ma, X., Qin, C., You, H., Ran, H., Fu, Y., 2022: Rethinking network design and local geometry in point cloud: A simple residual MLP framework. *arXiv preprint arXiv*, 2202.07123.

Mandlburger, G., Hauer, C., Wieser, M., Pfeifer, N., 2015: Topo-bathymetric LiDAR for monitoring river morphodynamics and instream habitats—A case study at the pielach river. *Remote Sens (basel)* 7, 6160–6195.

Paul, O., Ekhtari, N., Glennie, C., 2025: Automated depth correction of bathymetric LiDAR point clouds using PointCNN semantic segmentation. *Frontiers in Remote Sensing*, 6, 1521446.

Pfennigbauer, M., Ullrich, A., 2010: Improving quality of laser scanning data acquisition through calibrated amplitude and pulse deviation measurement. In *Laser Radar Technology and Applications XV*, Vol. 7684, 463-472.

Phan, T.H., Yamamoto, K., 2020: Resolving class imbalance in object detection with weighted cross entropy losses. *arXiv preprint arXiv*, 2006.01413.

Philpot, W., 2019: Airborne Laser Hydrography II, Cornell University Library (eCommons), Coernell. https://ecommons.cornell.edu/handle/1813/66666

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017: Pointnet: Deep learning on point sets for 3d classification and segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 652-660.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Rhomberg-Kauert, J., Dammert, L., Grömer, M., Pfennigbauer, M., Mandlburger, G., 2024: Macrophyte detection with bathymetric LiDAR–Applications of high-dimensional data analysis for submerged ecosystems. *The International Hydrographic Review*, 30(2).

Ronneberger, O., Fischer, P., Brox, T., 2015: U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention*, 234-241.

Russo, P., Di Ciaccio, F., Troisi, S., 2021. DANAE++: A smart approach for denoising underwater attitude estimation. Sensors, 21(4), 1526.

Schilling, H., Bulatov, D., Niessner, R., Middelmann, W., Soergel, U., 2018: Detection of vehicles in multisensor data via multibranch convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11), 4299-4316.

Sokolova, M., Japkowicz, N., Szpakowicz, S., 2006: Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *In Australasian joint conference on artificial intelligence*, 1015-1021.

Song, A., Kim, H., 2024: Deep Learning Approach For Classification of Water Bottom and Surface from Bathymetric Lidar Point Clouds. *In IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, 6069-6071.

Stoller, D., Tian, M., Ewert, S., Dixon, S., 2019: Seq-u-net: A one-dimensional causal u-net for efficient sequence modelling. *arXiv preprint arXiv*,1911.06393.

Tarsha Kurdi, F., Amakhchan, W., Gharineiat, Z., Boulaassal, H., El Kharki, O., 2023: Contribution of geometric feature analysis for deep learning classification algorithms of urban LiDAR data. *Sensors*, *23*(17), 7360.

Tseng, Y.H., Wang, C.K., Chu, H.J., Hung, Y.C., 2015: Waveform-based point cloud classification in land-cover identification. *International journal of applied earth observation and geoinformation*, 34, 78-88.

Vaquero, V., del Pino, I., Moreno-Noguer, F., Solà, J., Sanfeliu, A., Andrade-Cetto, J., 2020: Dual-branch CNNs for vehicle detection and tracking on LiDAR data. *IEEE Transactions on Intelligent Transportation Systems*, 22(11), 6942-6953.

Wang, C., Li, Q., Liu, Y., Wu, G., Liu, P., Ding, X., 2015: A comparison of waveform processing algorithms for single-wavelength LiDAR bathymetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, *101*, 22-35.

Yan, J., Meng, J., Zhao, J., 2021. Bottom detection from backscatter data of conventional side scan sonars through 1D-UNet. Remote Sensing, 13(5), 1024.

Yan, W.Y., Shaker, A., Habib, A., Kersting, A.P., 2012: Improving classification accuracy of airborne LiDAR intensity data by geometric calibration and radiometric correction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 35-44.

Yu, S., Wang, M., Zhang, C., Li, J., Yan, K., Liang, Z. Wei, R., 2023: A Dynamic Multi-Branch Neural Network Module for 3D Point Cloud Classification and Segmentation Using Structural Re-parametertization. *In 2023 11th International Conference on Agro-Geoinformatics*, 1-6.

Yuan, J., Ma, H., Zhang, L., Deng, J., Luo, W., Liu, K., Cai, Z., 2025: Transfer Learning Based on Multi-Branch Architecture Feature Extractor for Airborne LiDAR Point Cloud Semantic Segmentation with Few Samples. *Remote Sensing*, 17(15), 2618.

Zhao, R., Pang, M., Wang, J., 2018: Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network. *International journal of geographical information science*, 32(5), 960-979.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-81-2026