

CHALMERS
UNIVERSITY OF TECHNOLOGY

Meter-Resolution Canopy Height Data Alone Can Be Used for Predicting the Conservation Value of Boreal Forests

Downloaded from: https://research.chalmers.se, 2026-11-14 20:19 UTC

Citation for the original published paper (version of record):
Della Vedova, M., Wahde, M. (2026). Meter-Resolution Canopy Height Data Alone Can Be Used for Predicting the Conservation Value of
Boreal Forests. IEEE Geoscience and Remote Sensing Letters, 23.
http://dx.doi.org/10.1109/LGRS.2026.3741421

N.B. When citing this work, cite the original published paper.

© 2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, or reuse of any copyrighted component of this work in other works.

(article starts on next page)

Meter-Resolution Canopy Height Data Alone Can Be Used for Predicting the Conservation Value of Boreal Forests

Marco L. Della Vedova and Mattias Wahde

***Abstract*—We consider the problem of automated binary classification of forests as having either low or high conservation value. We demonstrate that the resolution of the data used in classification is a crucial factor. At resolutions of around 2 m or less, it is possible to distinguish individual trees, a situation that greatly improves classification accuracy: Using canopy height models with 1-m resolution our U-Net-based classifier obtains a classification accuracy of 0.956 over a test dataset covering southern Sweden. This result can be compared to earlier reported accuracies in the range 0.81–0.92 at lower resolution (10–100 m) over a similar region. We also show how the classification performance varies over a range of resolution values, and find a strong positive jump in accuracy at resolutions of 2 m or less, where individual trees can be discerned.**

***Index Terms*—Canopy height model (CHM), forests, geospatial artificial intelligence, machine learning, remote sensing.**

I. Introduction

IDENTIFYING and protecting forests with high conservation value is crucial for safeguarding biodiversity [1]. Those forests often play an important role as the last remnants of natural structures in otherwise intensely managed landscapes. However, despite their importance, reliable information on where such forests occur remains scarce. Traditional field inventories may provide such information, but they are costly and time-consuming to carry out, making them unsuitable for large-scale monitoring needs. The importance of rapidly developing high-quality biodiversity monitoring tools has been amplified by recent international agreements such as the Kunming–Montreal global biodiversity framework [2] and the EU biodiversity strategy for 2030 [3]. These policies set ambitious targets, such as protecting at least 30% of terrestrial areas, of which one third should be under strict protection. Reaching those goals will require global high-resolution mapping of biodiversity values. Current data gaps, regarding both resolution and accuracy of the mapped regions, are critical obstacles to reaching these goals.

Large-scale biodiversity assessments are now possible, as a result of recent advances in remote sensing and machine learning. Forest naturalness has been evaluated from a variety of perspectives in many studies pertaining to a large portion of the forested areas of the world, such as Northern Europe [4], Central Europe [5], [6], Canada [7], and Brazil [8]. Sweden's forests have also been the subject of several recent studies focused on the assessment of conservation value. Bubnicki et al. [9] used Random Forest classifiers with diverse landscape, structural, and socio-economic variables as input to map the relative likelihood of high conservation value forests (HCVFs) across Sweden at a spatial resolution of 1 ha (100×100 m) with an accuracy of 0.81. Their work also highlighted the importance of canopy structure and forest management intensity as key predictors. However, the coarse resolution of hectare-sized pixels makes it impossible to distinguish single trees or small habitat patches, which are often essential for biodiversity.

Andermann et al. [10] applied a deep learning segmentation model to a suite of remote sensing products, including LiDAR-derived metrics, Sentinel-2 imagery, and climatic and anthropogenic data. Their method generated a national-scale conservation value map of Sweden at 10-m resolution, reaching a classification accuracy of 0.92. Yet, this progress came at the cost of increased complexity, both regarding the model and the data that it uses: The model required input from 11 distinct environmental layers, each with its own preprocessing pipeline. Moreover, the 10-m resolution is still insufficient for identifying individual trees.

A complementary approach was recently proposed by Della Vedova and Wahde [11], who demonstrated that features derived solely from the canopy height model (CHM) at 1-m resolution could serve as interpretable indicators of forest naturalness. Their method achieved accuracies of up to 0.90 in southern Sweden and showed that canopy structure alone contains sufficient information to distinguish between highly natural and managed forests. However, their framework required prior delineation of polygons of interest, restricting its applicability to predefined forest stands rather than open, continuous landscapes. Note that in this context, the concepts of *forest naturalness* and *conservation value* are closely aligned: forests with a higher degree of naturalness are regarded as having higher conservation value.

The main contributions of this study are threefold: 1) we introduce a U-Net-based segmentation framework for predicting conservation value directly from CHM tiles, without requiring any predefined polygon boundaries; 2) we provide a controlled analysis of how spatial resolution affects

Received 24 September 2025; revised 23 February 2026 and 10 August 2026; accepted 29 September 2026. Date of publication 7 October 2026; date of current version 20 October 2026. This work was supported in part by the Sweden's Innovation Agency (VINNOVA) Project under Grant 2022-01702 and in part by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) through Swedish Research Council under Agreement 2022-06725. *(Corresponding author: Marco L. Della Vedova.)*

The authors are with the Department of Mechanical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden (e-mail: marco.dellavedova@chalmers.se).

Digital Object Identifier 10.1109/LGRS.2026.3741421

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

performance, isolating resolution as the primary variable; and 3) we show that tree-resolving CHM data alone are sufficient to achieve high predictive accuracy in Boreal forests, reaching 0.956 accuracy at 1-m resolution and exceeding the state-of-the-art performance despite relying on a single input layer.

The source code supporting this work is publicly available.[1]

II. Material and Methods

A. Problem Definition

The core objective of this study is to develop and evaluate automated methods for estimating the degree of forest naturalness from remote sensing data, specifically using the CHM. The data utilized here provide a detailed raster representation of tree canopy heights, captured with a certain spatial resolution (e.g., 1×1 m), covering the geographical area of interest. In detail, the input to the method consists of a raster grid, where each pixel encodes the height of the canopy at that location relative to the ground level. The desired output is a conservation value map, where each pixel is associated with a continuous naturalness score ranging from 0 (low naturalness) to 1 (high naturalness). Formally

$$\mathcal{N}(A) = f(\mathcal{C}(A)) \tag{1}$$

where A denotes the geographical area of interest, defined by a bounding box, $\mathcal{N}(A)$ is the predicted naturalness score for every point in A, taking values in the range $[0, 1]$, $\mathcal{C}(A)$ denotes the CHM input, representing the canopy height at every point in A, and f is the function, approximated by the proposed U-Net model (see Section II-D), that maps CHM values to naturalness scores.

B. Data

The primary data source used in this study is the CHM, a geo-referenced raster dataset representing the height of the forest canopy relative to ground level. Each pixel in the CHM used in this work corresponds to a 1×1 m square on the ground, with values expressed in decimeters, offering a vertical precision of 0.1 m. This high spatial resolution makes it generally possible to identify individual trees and to carry out a detailed structural analysis of forest stands.

The CHM data originate from national airborne laser scanning campaigns conducted by the Swedish Forest Agency as part of their national forest mapping program.[2] In this study, we consider a subset of the data covering southern Sweden.

The dataset is partitioned into training, validation, and test sets using the same geographical stratification strategy described in [11], ensuring spatial independence between the subsets. The geographical areas covered by the three subsets are 24 559, 6177, and 7680 km^2, respectively.

Ground-truth labels are provided as raster layers at the same spatial resolution as the CHM, with three discrete classes: background, low conservation value forests (LCVFs), and HCVFs. We derived these labels from a polygon-based classification of forest management intensity defined by environmental experts in field campaigns. The background class includes nonforested areas, regions outside the labeled polygons where naturalness is undefined, and areas with mixed labels (i.e., marked both high and low conservation value), which account for approximately 0.2% of the labeled area. Note that around 92.3% of the area of study is labeled as background. HCVFs, which cover around 3.5% of the total area, include habitat-classed areas within Natura 2000[3] as well as an inventory of key habitats conducted by forestry companies and by the Swedish Forest Agency. LCVFs, which cover around 4.2% of the total area, include forests harvested between 1990 and 2000, forest stands between 30 and 80 years old, and older forest stands that have been assessed as having low natural value by forestry companies. More details about these labels can be found in [11].

In contrast to prior work, which performed classification at the polygon level using aggregated features, in this work we consider raster tiles as input to the classification task. Each tile has a fixed size of 1280×1280 m (i.e., 1280×1280 pixels at 1-m resolution) and two layers: One containing canopy height values and the other containing the corresponding label values.

C. Data Preprocessing

Prior to model training, several preprocessing steps were applied to the CHM and the corresponding label rasters. First, to reduce the influence of outliers in canopy height values (e.g., artifacts from misclassified LiDAR returns), we clipped all CHM values to a maximum threshold $H_{\max}$. All pixel values above $H_{\max}$ were set to $H_{\max}$, after which the entire raster was normalized linearly to the range $[0, 1]$

$$c' = \frac{\min(c, H_{\max})}{H_{\max}} \tag{2}$$

where c is the original CHM pixel value and c' is the normalized value used as model input. In this work, the value of $H_{\max}$ has been set to 30 m. This value was chosen after a visual inspection of the height distribution over the training set, which indicated that only a very small fraction of trees (0.09% of the pixels) exceed 30 m in height. The normalization ensures numerical stability during training and provides consistent scaling across different tiles.

To study the effect of spatial resolution, we also generated downsampled versions of the dataset. For the CHM rasters, downsampling was performed by block averaging, such that each downsampled pixel represents the mean canopy height of the corresponding block of higher-resolution pixels. For the label rasters, downsampling was carried out by majority voting within each block, assigning the class (high naturalness, low naturalness, or background) that occurred most frequently among the original pixels. This procedure ensured alignment between CHM inputs and labels across all considered resolutions.

The final dataset therefore consists of multiple versions of the same geographical tiles: one at the native resolution of 1 m, and additional versions at coarser resolutions obtained through

[1] See https://github.com/mldv/1mCHM-forest-conservation

[2] Available at https://www.skogsstyrelsen.se/sjalvservice/karttjanster/skogliga-grunddata, *trädhöjd* layer.

[3] Natura 2000 is a network of protected areas throughout the EU.

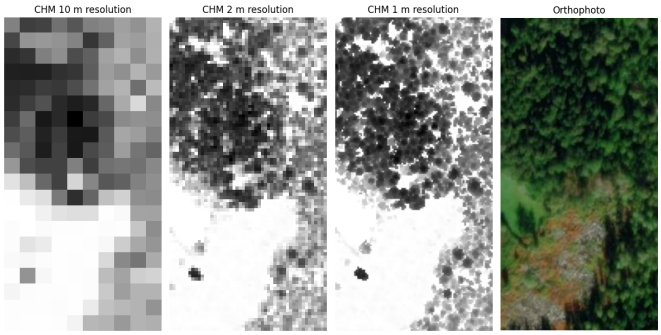


Fig. 1. Comparison of different CHM resolutions and corresponding orthophoto. The region in the figures is 100 × 200 m in size.

downsampling (see Fig. 1). These data allow us to evaluate the role of spatial resolution in predicting forest naturalness and to benchmark our 1-m model against coarser representations of the canopy structure.

D. Deep Learning Classifier

The proposed method is based on a fully convolutional encoder–decoder architecture. We adopt a U-Net model [12] to approximate the mapping f from CHM raster tiles $\mathcal{C}(A)$ to pixelwise naturalness predictions $\mathcal{N}(A)$. The network receives a single-channel input representing normalized canopy height values and outputs a raster of the same dimensions, where each pixel contains a score in $[0, 1]$ corresponding to the predicted probability of high naturalness.

The U-Net follows a standard architecture composed of a downsampling path, a bottleneck, and an upsampling path with skip connections. The downsampling path uses repeated blocks of two 3 × 3 convolutions (with batch normalization and ReLU activation), followed by max pooling. The bottleneck doubles the number of feature channels and preserves spatial resolution. The upsampling path consists of transposed convolutions followed by concatenation with the corresponding feature maps from the encoder, enabling precise localization.

We use four levels of resolution, halving feature sizes in the encoder and decoder. In the full resolution model, i.e., with 1280 × 1280 pixel inputs, it corresponds to {640, 320, 160, 80} pixels. A final 1 × 1 convolution followed by a sigmoid activation produces the output probability map.

E. Model Training and Validation

We trained the U-Net following a standard supervised learning procedure implemented in PyTorch. The model parameters were optimized by minimizing a loss function computed over the training set. The performance on the validation set was monitored after each epoch to guide early stopping and to select the best-performing model, which was then used in final testing as reported below. Several loss functions were evaluated, each emphasizing different aspects of segmentation quality.

1) *Binary focal loss*, focusing learning on difficult-to-classify pixels.
2) *Dice loss*, promoting spatial coherence by maximizing overlap between predicted and ground-truth regions.
3) *Pixel accuracy loss*, directly optimizing classification accuracy, using a differentiable approximation: instead of hard thresholding, we use a sharp sigmoid around 0.5.
4) *Combined loss*, a weighted average of the three above.

All losses support masking of background regions and operate at the pixel level. The combined loss function is defined as

$$\mathcal{L}_{\text{combined}} = \alpha \cdot \mathcal{L}_{\text{focal}} + \beta \cdot \mathcal{L}_{\text{pixel}} + (1 - \alpha - \beta) \cdot \mathcal{L}_{\text{dice}} \quad (3)$$

where α and β are weighting parameters ($\alpha = 0.3$ and $\beta = 0.4$ in the best-performing setup, identified by grid search).

Optimization was carried out with the Adam optimizer (learning rate 10^{-3}), a batch size of 8, and a maximum of 50 epochs. Note that for downsampled versions of the input, we adapt the batch size and the learning rate for speedup. Early stopping was applied with a patience of ten epochs and a minimum validation loss improvement of 10^{-4}. At each epoch, training and validation losses and accuracies were computed, and the model with the lowest validation loss was stored as the selected model for the configuration under study.

F. Inference

After training, the selected model can be applied to unseen CHM tiles to produce pixelwise predictions of forest naturalness. Inference proceeds tile by tile: Each 1280 × 1280 m CHM raster is fed into the trained U-Net, which outputs a probability map of the same size. Each pixel value represents the estimated probability of belonging to the *high naturalness* class, with values close to 0 corresponding to low naturalness and values close to 1 corresponding to high naturalness.

For model evaluation, probability maps are thresholded at 0.5 to obtain binary predictions, which are then used to compute accuracy and other classification metrics (precision, recall, IoU, and Dice). All metrics are calculated only over areas with available ground-truth labels, i.e., not on background. Fig. 2 illustrates typical inputs and outputs of the method, together with labeled regions and the corresponding accuracy of the predictions.

III. Results

A. Effect of Spatial Resolution

Spatial resolution is a key factor in determining the amount of structural information that can be extracted from the CHM. To quantify its impact, we trained otherwise identical models on downsampled versions of the CHM at resolutions ranging from 40 to 1 m. In all cases, the ground area covered by a tile was kept constant (1280 × 1280 m), which implies that the number of pixels and thus the model size increased with resolution.

Table I and Fig. 3 show the results. As expected, accuracy improved monotonically with finer resolution, from 0.82 at 40 m to 0.96 at 1 m. The largest relative gains occur when moving from very coarse scales (40 and 20 m) to intermediate ones (10, 8, and 5 m). Between 40 and 5 m, accuracy increases by more than 0.10 (absolute), but the curve starts to flatten, suggesting diminishing returns as more pixels are

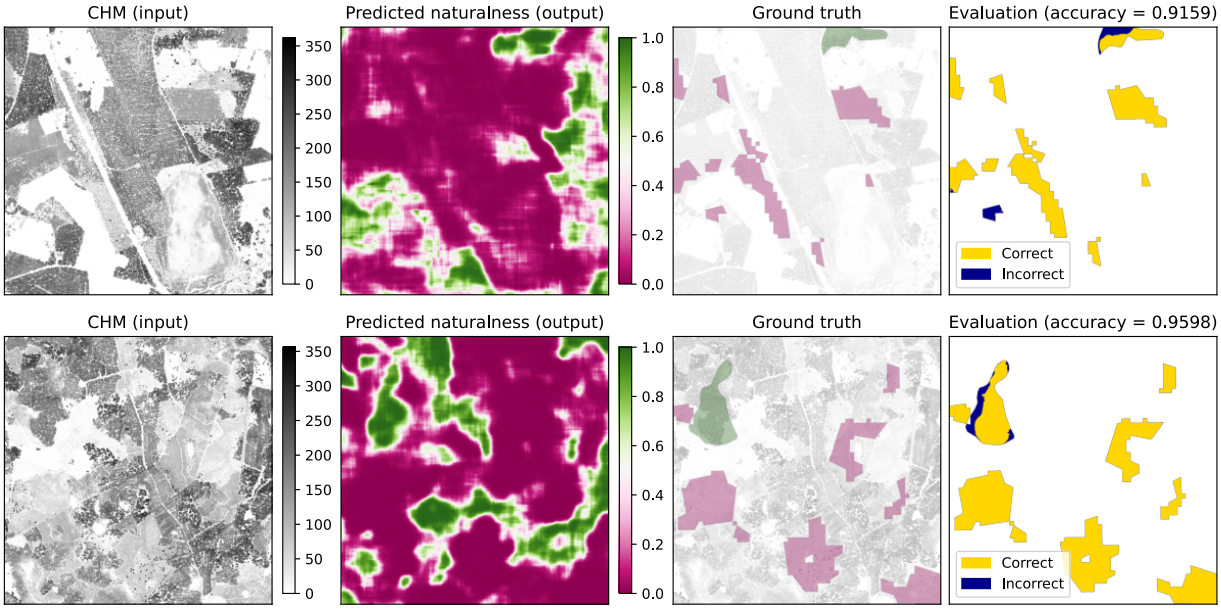


Fig. 2. Two illustrations of our method in action. In each row, from left to right, the first panel shows the input to the method, i.e., a CHM over a given region. The second panel shows the output of our method, where each pixel is assigned a value in the range $[0, 1]$, which corresponds to the predicted conservation value. The third panel shows the ground-truth assignments (with the CHM in the background for reference): green for high naturalness areas, violet for low naturalness areas. The fourth panel exemplifies the performance of our method (after thresholding): areas that are correctly predicted are marked in yellow, and incorrectly predicted areas are marked in blue (including examples of both false positives and false negatives). The centers of the regions are at (60.322692 N, 17.477156 E) and (56.563449 N, 15.122369 E).

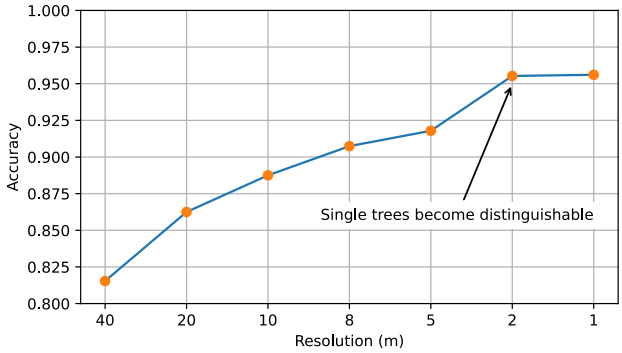


Fig. 3. Test accuracy plotted as a function of resolution. Note the jump in accuracy as the resolution reaches 2 m and below.

TABLE I

EFFECT OF THE SPATIAL RESOLUTION ON PERFORMANCE METRICS

Resol.	# pixels[a]	Accuracy	Precis.	Recall	IoU	Dice
40 m	32	0.8154	0.8015	0.8057	0.6717	0.8036
20 m	64	0.8625	0.8496	0.8570	0.7441	0.8533
10 m	128	0.8875	0.8833	0.8745	0.7840	0.8789
8 m	160	0.9074	0.8941	0.9092	0.8208	0.9016
5 m	256	0.9179	0.8907	0.9386	0.8417	0.9140
2 m	640	0.9553	0.9499	0.9546	0.9089	0.9523
1 m	1280	0.9561	0.9466	0.9597	0.9104	0.9531

[a]Number of pixels of the tile's side. Note that the size of the tile on the ground is 1280×1280 m in all cases.

TABLE II

BENCHMARKING RESULTS FOR THE SOUTH BOREAL REGION OF SWEDEN

Model	Resol.	Accuracy	Precision	Recall	Dice
Bubnicki [9]	100 m	0.810	–	–	–
Andermann [10]	10 m	0.917	0.798	0.809	0.804
Ours	10 m	0.888	0.883	0.875	0.879
Ours	1 m	**0.956**	**0.947**	**0.960**	**0.953**

added. However, when resolution is further improved from 5 to 2 m and 1 m, a distinct jump in performance is observed (see Fig. 3), with accuracy rising from about 0.92 to 0.96. This step change coincides with the scale at which individual trees and canopy gaps can be distinguished in the CHM, as illustrated in Fig. 1.

These results highlight that while moderate resolutions already provide useful information about stand-level structure, the ability to resolve individual trees is decisive for predicting naturalness with high confidence.

B. Benchmarking Against the State-of-the-Art

To assess the performance of our method relative to existing approaches, we compared it with the results reported by Bubnicki et al. [9] and Andermann et al. [10], both focusing on the South Boreal region of Sweden. While the former applied a tenfold cross-validation strategy on their dataset, Andermann's models and ours were evaluated using the same training–validation–test split. Table II summarizes the results in terms of accuracy, precision, recall, and Dice coefficient. Bubnicki et al. [9] reported an accuracy of 0.81 using a Random Forest model trained on a wide set of structural, environmental, and socio-economic predictors at 1-ha resolution. Andermann et al. [10] achieved substantially higher performance with an attention U-Net model based on 11 heterogeneous input bands at 10-m resolution (accuracy 0.917, Dice 0.804). Using the same training–validation–test split as them, our U-Net models trained solely on CHM data reached an accuracy of 0.888 at 10-m resolution, and further improved

to 0.956 at 1-m resolution. At this fine scale, both precision (0.947) and recall (0.960) were balanced, leading to a Dice coefficient of 0.953.

IV. Discussion and Conclusion

This study shows that conservation value in Boreal forests can be predicted accurately from CHMs alone when spatial resolution is sufficiently high. Using a U-Net segmentation model on 1-m CHM data, we achieved an accuracy of 0.956 in the South Boreal region of Sweden, exceeding previously reported results obtained at coarser scales. A clear performance transition was observed when moving from 5–10- to 1–2-m resolution, corresponding to the scale at which individual trees and canopy gaps become discernible. This indicates that tree-level structural heterogeneity, rather than perfect tree delineation, is the critical factor for predicting naturalness. Although dedicated individual tree detection methods exist [13], [14], they are not required to achieve the performance reported here.

The comparison against state-of-the-art approaches further emphasizes this point. Whereas Bubnicki et al. [9] relied on a wide set of predictors at 1 ha (100×100 m) resolution (accuracy 0.81), and Andermann et al. [10] used 11 heterogeneous input layers at 10-m resolution (accuracy 0.917), our method surpassed both using only a single data source at 1-m resolution. The fact that our model requires only CHM input strongly reduces the complexity of data preparation. High-quality LiDAR-based CHMs are already available in several countries, and emerging image-based [15] and radar-based [16] approaches suggest that such data may become increasingly accessible.

A key limitation lies in the scarcity of reliable labels for forest conservation value, particularly for low-value stands, rather than in the availability of canopy height data. Ground truth requires ecological expertise and is often tied to national inventories or protected area registers. Since conservation indicators vary across ecosystems, region-specific models may be necessary, though transfer learning could enable adaptation from data-rich to data-poor contexts. The results presented here are specific to Boreal forests, which are common in Canada, Alaska, Russia, and the Nordic countries, where canopy structure is typically characterized by relatively distinct crowns and moderate vertical stratification. The performance gain observed at 2-m resolution relies on the ability of the CHM to capture tree-level structural heterogeneity. In other biomes, such as dense tropical or multilayered temperate forests, crown overlap and complex vertical structure may reduce the separability between management regimes in canopy height data alone. Applying the proposed method to other forest types would therefore require: 1) region-specific ground-truth labels; 2) retraining or fine-tuning of the model on representative local data; and 3) verification that meter-scale CHM data provide sufficient structural contrast for the conservation indicators of interest. Without these conditions, predictive performance cannot be assumed to transfer directly across ecosystems.

In order to apply the proposed method for generating large-area maps, tilewise inference requires appropriate border-handling strategies to avoid edge artifacts. Standard solutions include overlapping sliding windows or trimming border regions prior to tiling to ensure spatial continuity.

In summary, we demonstrate that, in Boreal forests, meter-resolution canopy height data alone provide sufficient structural information to accurately predict conservation value. By isolating spatial resolution as the key variable, we show that predictive performance improves markedly on fixed ground extent once the tree-level structure becomes visible in the CHM. These results highlight the potential of high-resolution canopy height data as a simple and scalable basis for conservation value mapping while emphasizing the need for region-specific validation beyond Boreal ecosystems.

References

[1] R. E. McRoberts, S. Winter, G. Chirici, and E. LaPoint, "Assessing forest naturalness," *Forest Sci.*, vol. 58, no. 3, pp. 294–309, Jun. 2012.

[2] *Convention on Biological Diversity (CBD), Kunming-Montreal Global Biodiversity Framework*, Secretariat of the Convention on Biological Diversity (CBD), Montreal, QC, Canada, 2022.

[3] European Commission, Directorate General for Environment, *EU Biodiversity Strategy for 2030: Bringing Nature Back into Our Lives*, LU: Publications Office, Brussels, Belgium, 2021.

[4] G. Brumelis, B. Jonsson, J. Kouki, T. Kuuluvainen, and E. Shorohova, "Forest naturalness in northern Europe: Perspectives on processes, structures and species diversity," *Silva Fennica*, vol. 45, no. 5, pp. 807–821, 2011.

[5] S. Winter, H. S. Fischer, and A. Fischer, "Relative quantitative reference approach for naturalness assessments of forests," *Forest Ecol. Manage.*, vol. 259, no. 8, pp. 1624–1632, Mar. 2010.

[6] A. Kusbach, M. Friedl, V. Zouhar, T. Mikita, and J. Šebesta, "Assessing forest classification in a landscape-level framework: An example from central European forests," *Forests*, vol. 8, no. 12, p. 461, Nov. 2017.

[7] M. Barrette, D. Dumais, I. Auger, Y. Boucher, M. Bouchard, and J. Bouliane, "Naturalness assessment performed using forestry maps to validate forest management sustainability," *Ecol. Indicators*, vol. 119, Dec. 2020, Art. no. 106832.

[8] A. Magalhaes Albuquerque et al., "Qualitative forest inventory in eucalyptus plantations using unmanned aerial vehicles, multispectral sensors, and deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[9] J. W. Bubnicki, P. Angelstam, G. Mikusiński, J. Svensson, and B. G. Jonsson, "The conservation value of forests can be predicted at the scale of 1 hectare," *Commun. Earth Environ.*, vol. 5, no. 1, p. 196, Apr. 2024.

[10] T. Andermann, J. Häggmark, P. Olsson, and A. Högström, "Deep learning segmentation model produces accurate biodiversity predictions at 10 m spatial resolution," *Res. Square, Durham, NC, USA, Tech. Rep. rs-4734879*, Sep. 2024, doi: 10.21203/rs.3.rs-4734879/v1.

[11] M. L. Della Vedova and M. Wahde, "Naturalness indicators of forests in southern Sweden derived from the canopy height model," *Eur. J. Remote Sens.*, vol. 58, no. 1, Dec. 2024, Art. no. 2441834.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–(MICCAI)*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham, Switzerland: Springer, 2015, vol. 9351, pp. 234–241.

[13] I. Harmon et al., "Injecting domain knowledge into deep neural networks for tree crown delineation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4415419.

[14] Z. Ding, H. Zhang, R. Wang, L. Zhang, H. Jiang, and T. Yun, "A dual-branch deep learning framework at the grid scale for individual tree segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 22, pp. 1–5, 2025.

[15] J. Tolan et al., "Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial LiDAR," *Remote Sens. Environ.*, vol. 300, Jan. 2024, Art. no. 113888.

[16] H. Luo, C. Yue, Y. Wu, X. Zhang, C. Lu, and G. Ou, "An efficient method for estimating tropical forest canopy height from airborne PolInSAR data," *Ecol. Indicators*, vol. 166, Sep. 2024, Art. no. 112566.