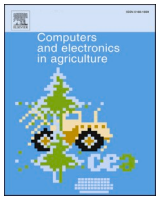
Computers and Electronics in Agriculture 237 (2025) 110576

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Enhancing forest inventory Accuracy: Comparing 3D-CNN and k-NN with genetic algorithm Approaches using ALS data across boreal bioregions

Andras Balazs *, Sakari Tuominen, Annika Kangas

Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00790 Helsinki, Finland

ARTICLE INFO

Keywords:
CNN
Forest inventory
Deep learning
Remote sensing
ALS

ABSTRACT

In 2020, the National Land Survey of Finland (NLS) started to acquire airborne laser scanning (ALS) data with a nominal pulse density of 5 pulse/m^2, representing a significant increase compared to the ALS data collected in earlier scanning campaigns. In our previous study utilizing lower density ALS data we have found that a three-dimensional convolutional neural network (3D-CNN) yielded results comparable to our benchmark genetic algorithm supported k-nearest neighbors (k-NN) method. In this paper, we compared the performance of our 3D-CNN model using voxelized ALS data with an increased point density to the benchmark k-NN method in estimating forest stand variables. The emphasis of our study was on the generalizability of tested models trained on independent, spatially uncorrelated datasets. For this reason, we utilized three datasets from different parts of Finland, of which two (Mikkeli and Äänekoski) are situated in the southern and one (Kolari) in the northern boreal bioregion. Models were trained with each of the datasets and cross-validated using the two other datasets and results reported separately.

The 3D-CNN outperformed our benchmark model by more than 9 % in terms of relative RMSE over all training sets, validation sets and forest variables. On the other hand, k-NN performed slightly better than CNN regarding average relative bias by 4.9 % over all datasets and variables. It Is also important to mention, that over 94 % of RMSE results were proven to be statistically significant, whereas over 40 % of bias results showed no statistically significant differences between the methods. Most remarkable differences are produced by models trained and validated by datasets from the same bioregion (Mikkeli and Äänekoski). Relative RMSE scores calculated from CNN predictions were lower compared to the benchmark method by 5.7 %, 25.3 %, 19.6 %, and 63.3 % for volume of total growing stock, pine, spruce, and broad-leaved, respectively. Based on these results we can state that models trained with the 3D-CNN used in our study are better generalizable at least within the same bioregion compared to our benchmark method. Increased prediction accuracy especially in terms of tree species-specific volumes could bring benefits to forest management practices, planning of harvest, or biodiversity research.

1. Introduction

Airborne laser scanning (ALS) of entire Finland is currently underway as part of a national laser scanning program by the National Land Survey of Finland (NLS). The current laser scanning is the second rotation of the entire country, aiming at significantly higher pulse density (nominal density of 5 pulse/m^2) than the previous rotation carried out in 2008–2019, which had a pulse density of 0.5 pulse/m^2.

ALS-based point clouds are currently seen as the most accurate remote sensing (RS) method in forest inventory (Fassnacht et al., 2024; Kangas et al., 2018; Maltamo et al., 2021). ALS, as a "true" 3-dimensional (3D) RS method, can provide information of multiple canopy layers as well as the terrain surface. Even very low density ALS data can significantly improve prediction of forest stand variables as demonstrated for instance by Tuominen et al. (2017). Typically, operational ALS-based forest inventories utilize area-based approach (ABA) as opposed to individual tree detection. In ABA stand variables are predicted e.g. for square shaped elements of a systematic inventory grid using field plots of comparable size as a reference data. Statistically it is a case of model-based estimation, where the value in each pixel (grid cell) is predicted with a model calculated from a sample (Maltamo et al., 2021).

* Corresponding author.
E-mail addresses: andras.balazs@luke.fi (A. Balazs), sakari.tuominen@luke.fi (S. Tuominen), annika.kangas@luke.fi (A. Kangas).

https://doi.org/10.1016/j.compag.2025.110576
Received 24 January 2025; Received in revised form 13 May 2025; Accepted 16 May 2025
Available online 20 May 2025

In ABA quantitative features representing the structure of the point cloud and the distribution of the points (i.e., returned LiDAR echoes) are extracted for predicting the actual forest inventory variables of interest (Næsset, 2002). The amount of features that can be extracted is often very high which results in a high-dimensional feature space (e.g. Tuominen and Haapanen, 2013).

Utilization of the data extracted from the point clouds in an optimal way usually benefits from a suitable machine learning algorithm. The algorithm selects an appropriate subset of features, develops, and trains a model that predicts the variables of interest. Typical machine learning methods used for this task are, for example, decision tree methods such as random forest or the k-nearest neighbor method (k-NN) combined with genetic algorithms that mimic natural selection (e.g. Corte et al., 2020; Pohjankukka et al., 2018).

A number of studies have indicated that even with high density ALS point cloud data, it is difficult to achieve accurate predictions of tree species-specific volumes or proportions with geometric point cloud features (based on single photon lidar scanning). This challenge arises when using traditional machine learning algorithms such as decision tree based methods, or support vector machines in combination with user defined point cloud features (Deng et al., 2016; Prieur et al., 2021; Yu et al., 2017).

Instead of extracting user-defined features from point clouds, it is possible to utilize deep learning methods for extracting self-directed point cloud features. Deep learning methods are a sub-class of machine learning. Due to their capability of learning abstract features from the input data, deep learning techniques are able to extract higher level features independently from the input data without user-defined algorithms (Ayrey et al., 2021; Seely et al., 2023). Convolutional neural networks (CNN) are a category of deep learning methods. CNNs are one type of artificial neural networks commonly applied for analyzing remote sensing data (Kattenborn et al., 2021). 2D-CNNs use two-dimensional input data, like aerial imagery or canopy height models (e.g. Schiefer et al., 2020), whereas 3D-CNNs three-dimensional data in form of point clouds or voxels derived from point clouds. 3D-CNNs have been used for various tasks in forestry related applications, for example, tree species recognition (e.g. Briechle et al., 2020) and estimation of growing stock (Ayrey et al., 2021; Oehmcke et al., 2024).

Although 3D-CNNs have considerable potential in extracting information, e.g. for predicting forest characteristics, they also have certain limitations. ALS-based point clouds typically have an unordered and unstructured format. While there are some recent examples of 3D-CNNs accepting point clouds as input (Oehmcke et al., 2024; Wielgosz et al., 2023), most 3D-CNNs require input data that is regularized to a standard format. For that reason, point cloud data are converted to, e.g. voxel-based structure which is an ordered data format, and the reformatted data is then used as an input for 3D-CNNs (Bello et al., 2020). When using relatively low-density ALS data, large part of the voxels in the voxel space will remain empty, depending on the number and size of voxels per sample unit that is used for training the 3D-CNN model. The results of our earlier study with lower density point cloud data indicate that a 3D-CNN can perform better in an area-based estimation of forest attributes, but the difference between the 3D-CNN and non-deep learning machine learning method was not significant (Balazs et al., 2022). Our hypothesis is that when using higher density ALS data, the difference between 3D-CNN and non-deep learning algorithms might be more prominent. Denser point cloud data should enable the 3D-CNN to make better use of its capability in composing features that are beneficial for the estimation task.

The objective of this study was to examine how a three-dimensional deep neural network performs in predicting total and species-specific forest attributes compared to the k-NN method utilizing ALS data at 5 pulse/m^2 nominal pulse density. K-NN is widely used in forest inventory applications, including the Finnish national forest inventory (Tomppo, 1991) making it a suitable benchmark method. The validation approach applied in our study allowed us to also assess the generalizability of these methods. Specifically, we investigated the accuracy of model predictions when the predictor and training datasets originated from different geographic locations having different type of forest structure. This evaluation procedure is in accordance with the findings of Kattenborn et al. (2022), who observed that CNNs' performance can be considerably overestimated by using spatially auto-correlated training and validation data.

2. Materials

2.1. Study area and reference data

The study material was collected from three different study areas, resulting in spatially uncorrelated datasets. The Mikkeli study area is located in Eastern Finland, approximately 61° 41′ N – 62° 11′ N and 26° 25′ E – 27° 27′ E. The study area of Äänekoski is situated in Central Finland, approximately 62° 34′ N – 63° 02′ N and 25° 20′ E – 26° 27′ E. In both study areas, the majority of land area is covered by managed forests mainly in private ownership, and the dominant land use is forestry for wood production. The third study area is located in Kolari in Northern Finland, approx. 66° 58′ N – 67° 26′ N and 23° 30′ E – 24° 40′ E. Major part of the forests covered by the Kolari study area are owned by the state of Finland and managed by the state-owned enterprise Metsähallitus.

Geo-botanically, the study areas in Mikkeli and Äänekoski belong to the southern boreal bioregion, and Kolari to the northern boreal bioregion. Forests in the Mikkeli and Äänekoski study areas mainly grow on eutrophic to sub-xeric forest site types, and the main tree species in those areas are Scots pine (Pinus sylvestris), Norway spruce (Picea abies), Silver birch (Betula pendula) and Downy birch (Betula pubescens). In the Kolari study area the site fertility classes are generally similar, but due to less favourable climatic conditions and larger proportion of peatland soils, the forest growth is low, and the forests are mainly dominated by pine. The terrain on all three study areas is relatively flat with highest elevation above sea level not exceeding 400 m. Each study area covers approx. 300 000 ha.

The map of the study areas is shown in Fig. 1.

Field reference data was collected by the Finnish Forest Centre (SMK) in 2020. Measurements were carried out on irregularly shaped and sized test plots between 1000 and 1500 m^2, aiming at 120–150 tally trees/plot. In the Mikkeli area 271 plots (average size: 1200 m^2), in the Äänekoski area 188 plots (average size: 1100 m^2) and in the Kolari area 216 plots (average size: 900 m^2) were measured.

In the test plots, all trees with a minimum diameter of 50 mm at breast height (DBH) were measured. Additionally, trees with a DBH of 30——49 mm were measured if their height was more than half of the average stand height. For all tally trees DBH, tree species and location within the plot were measured. From each sample plot, every fifth tally tree was sampled for height measurement. Furthermore, additional trees were sampled for height measurement for covering all tree species strata and diameter classes present at the study plot, if necessary. For the rest of the tally trees the height was predicted using ALS data (if the tree was detected from ALS point cloud) or using generic tree height models (if not detected from ALS).

The precise locations of tally trees were measured using a combination of GNSS and pseudolite-based positioning system by the Finnish Forest Centre. Pseudolites act as local base stations which are able to determine their relative locations in a local coordinate system. The locations of individual trees are determined using digital calipers that communicate with the pseudolites. The local coordinates are linked to geographic ETRS-TM35FIN coordinate system by GNSS positioning using a number of control points measured from the proximity of the sample plots (Luopa, 2022). Based on the system developer's specifications, the position accuracy (RMSE) is 0.13 m with 99.5 % of the trees having a position error less than 0.30 m (Kostensalo et al., 2023).

In order to generate regular input data for our models, we created

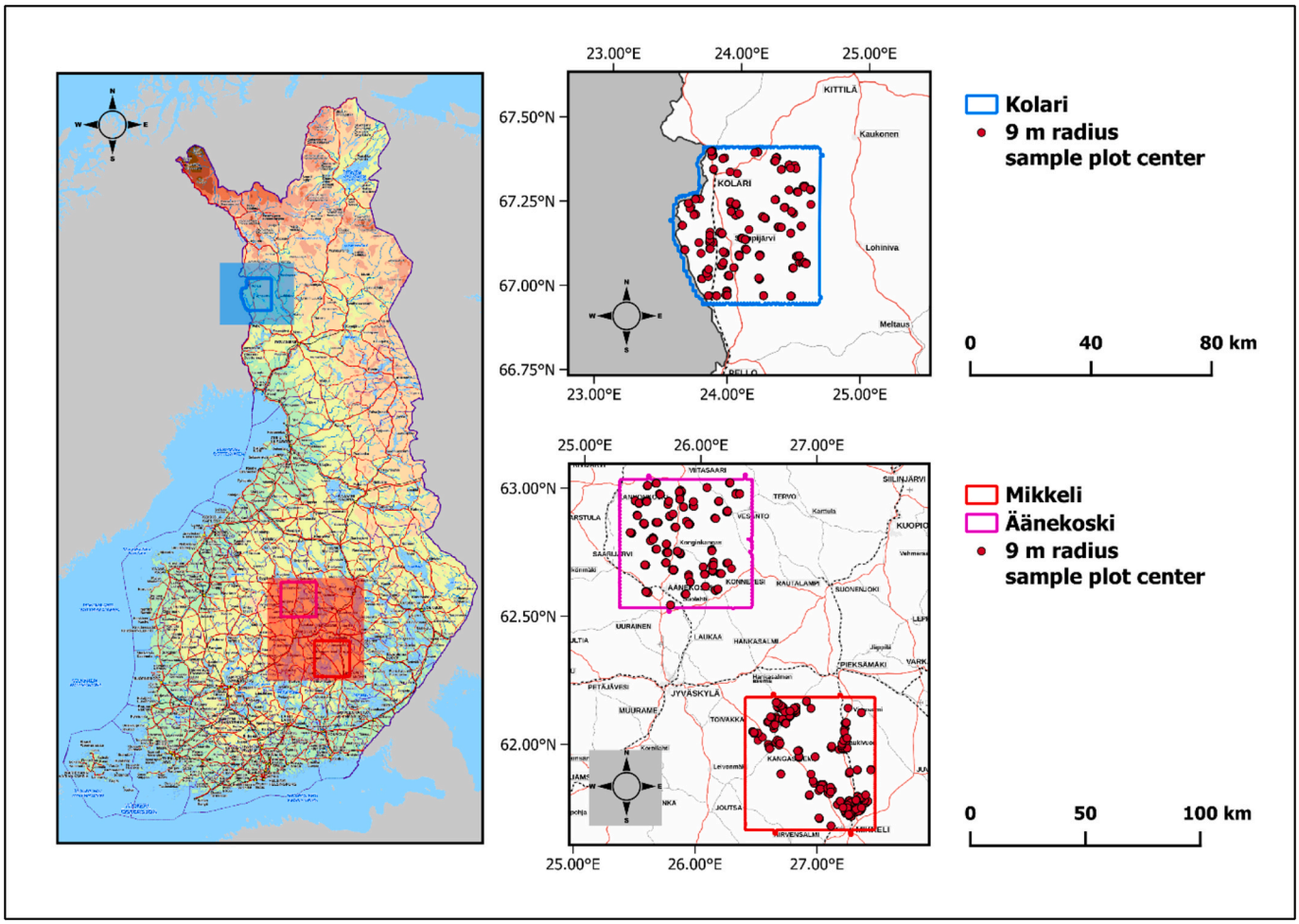


Fig. 1. Map of the study areas. Laser scanning area is marked by red, magenta and blue outlines; sample plots are indicated by red dots. Background map raster 1:800 000 by the NLS was used as background.

circular sample plots of 9-meter radius within the original test plots. In the Mikkeli, Äänekoski and Kolari study areas, 1483, 940 and 760 circular plots were placed, respectively. Based on our previous experience, these amounts of circular plots were found to be sufficient to cover forest attributes' variation in the respective test areas. The circular plots were placed so that they are completely within the limits of the test plots. On the other hand, overlap between the circular plots was allowed. Thus, an individual tally tree often belongs to more than one circular plot, and the tally trees form various combinations of tree sets in the circular plots. Overlapping circular plots are a way of data augmentation as deep learning methods usually require a substantial amount of training data due to their numerous trainable parameters (Shorten and Khoshgoftaar, 2019). An example of 9 m radius circular plot placements within an SMK test plot is depicted in Fig. 2.

Forest stand variables used in our study were: volume (V), mean height (H) and mean diameter (D) of the total growing stock as well as volume of growing stock of pine (V_p), spruce (V_s) and broad-leaved (V_b).

Properties of forest stand variables in each test area based on the circular plots are shown in Table 1.

3D-CNN and k-NN models were trained with each study area's sample plot set separately. Following the recommendations for training neural networks (Brownlee, 2017), during training of 3D-CNN models 15 % of the training set was used for validation to avoid overfitting which is a common issue with neural networks (e.g. Sarle, 1995). The performance of trained models was evaluated with the two other datasets not used in the training process. Evaluation was carried out with both test sets independently and results reported separately. Validation sets were not used in the training of k-NN models. Distributions of sample plot sets in volume classes of total growing stock are shown in Fig. 3.

According to Brownlee (2019a), different scales between estimands may add an unwanted complexity to the problem. Mean height and diameter values were multiplied by 10, to approximately match the value range of growing stock attributes.

2.2. ALS data

Laser scanning was carried out in June 2020 on all study areas. In Mikkeli RIEGL VQ-1560i laser scanner was used at 1525 m above ground with 20 % side overlap and 7.2 pulse/m^2 pulse density. In Äänekoski and Kolari RIEGL VQ-780i at a flight altitude of 1280 m was used with 20 % side overlap and pulse density was 6.4 pulse/m^2. The average point density of the resulting ALS point clouds covering the sample plots was 13.8 point/m^2.

Laser returns were extracted from the ALS data within SMK's plots. In the next step returns were clipped by the 9 m radius sample plot boundaries allocated into the SMK plots. Finally, heights of laser returns above sea level were normalized using the NLS's digital terrain model to obtain heights above ground.

The 3D-CNN used in this study accepts fixed-dimension input data. To fulfil this requirement laser data covering each sample plot were transformed to a voxel space. The original CNN architecture used by Ayrey and Hayes (2018) was edited to be able to experiment with different voxel sizes. Voxel sizes of 40x40x40, 45x45x45 and 50x50x50

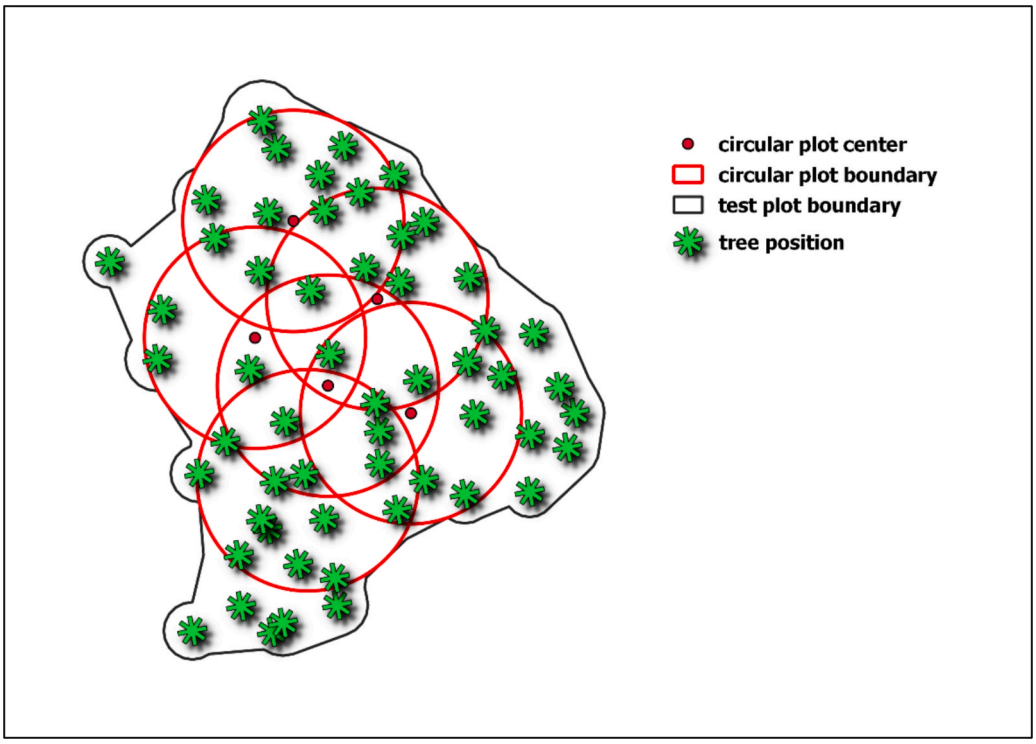


Fig. 2. Allocation of 9 m circular plots within test plot.

Table 1
Summary statistics of sample plots.

Äänekoski	V (m^3/ha)	H (m)	D (cm)	V_p (m^3/ha)	V_s (m^3/ha)	V_b (m^3/ha)
Min	27.1	8.6	9.5	0.0	0.0	0.0
Max	667.7	31.6	43.7	555.7	648.8	359.9
Mean	241.1	20.0	22.9	109.1	101.3	30.7
Standard deviation	129.6	4.5	6.1	108.9	134.5	52.0
Mikkeli	**V (m^3/ha)**	**H (m)**	**D (cm)**	**V_p (m^3/ha)**	**V_s (m^3/ha)**	**V_b (m^3/ha)**
Min	38.3	8.4	9.6	0.0	0.0	0.0
Max	1016.6	32.7	46.6	939.6	828.1	414.3
Mean	266.4	21.4	24.9	96.6	131.1	38.7
Standard deviation	139.3	4.7	6.8	120.5	151.4	68.3
Kolari	**V (m^3/ha)**	**H (m)**	**D (cm)**	**V_p (m^3/ha)**	**V_s (m^3/ha)**	**V_b (m^3/ha)**
Min	11.7	5.3	8.0	0.0	0.0	0.0
Max	351.6	21.7	38.0	351.6	270.4	170.3
Mean	114.4	13.9	19.3	85.6	15.0	13.8
Standard deviation	54.3	3.2	5.0	56.3	31.1	20.2

cm were tested with total growing stock variables. 10 independent training runs were carried out for each voxel size. The best performing voxel size was selected based on the mean relative RMSE of the training runs. The voxel size of 40x40x40 cm was chosen which amounted to 45x45x90 voxels covering a space of 18x18x36 m around the 9 m radius sample plots in x, y, and z directions, respectively. Each voxel was assigned the number of laser returns within the voxel. Voxel values were later converted to binary values, 0 = no return in the voxel, 1 = at least one return in the voxel, as binary voxel type was performing better in preliminary tests. Voxels outside the circular sample plots had a value of 0. The voxel data was subsequently split into training, validation, and test sets, corresponding to the ground truth data division.

3. Methods

Fig. 4 shows the workflow used in our study.

3.1. ALS features

ALS features were extracted from point clouds covering the 9 m circular sample plots. ALS features included height and intensity metrics as well as texture features. Texture features were obtained based on canopy surface models and intensity rasters, which were calculated from first and single laser returns at 0.5 m resolution.

Common height metrics (e.g. mean, max, skewness, kurtosis) were calculated using lidR package (Roussel, 2024) in R statistical software (R Core Team, 2023). In addition, the following features were calculated (Kirchhoefer et al., 2017; Næsset, 2002; St-Onge et al., 2008; Véga et al., 2016):

- L-moments (1–4) and L-moment skewness, kurtosis, and coefficient of variation,
- canopy relief ratio,
- percentage of canopy points having height greater than or equal to corresponding percentile of height distribution, at percentiles 20, 40, 60, 80, 95,
- average absolute deviation,
- quadratic and cubic mean,
- gap area,
- inner and outer volume,
- median of the absolute deviations from the overall median,
- rumple index.

Apart from rumple index, inner and outer volume, and gap area, where only single and first returns were used, metrics were calculated using first and last returns.

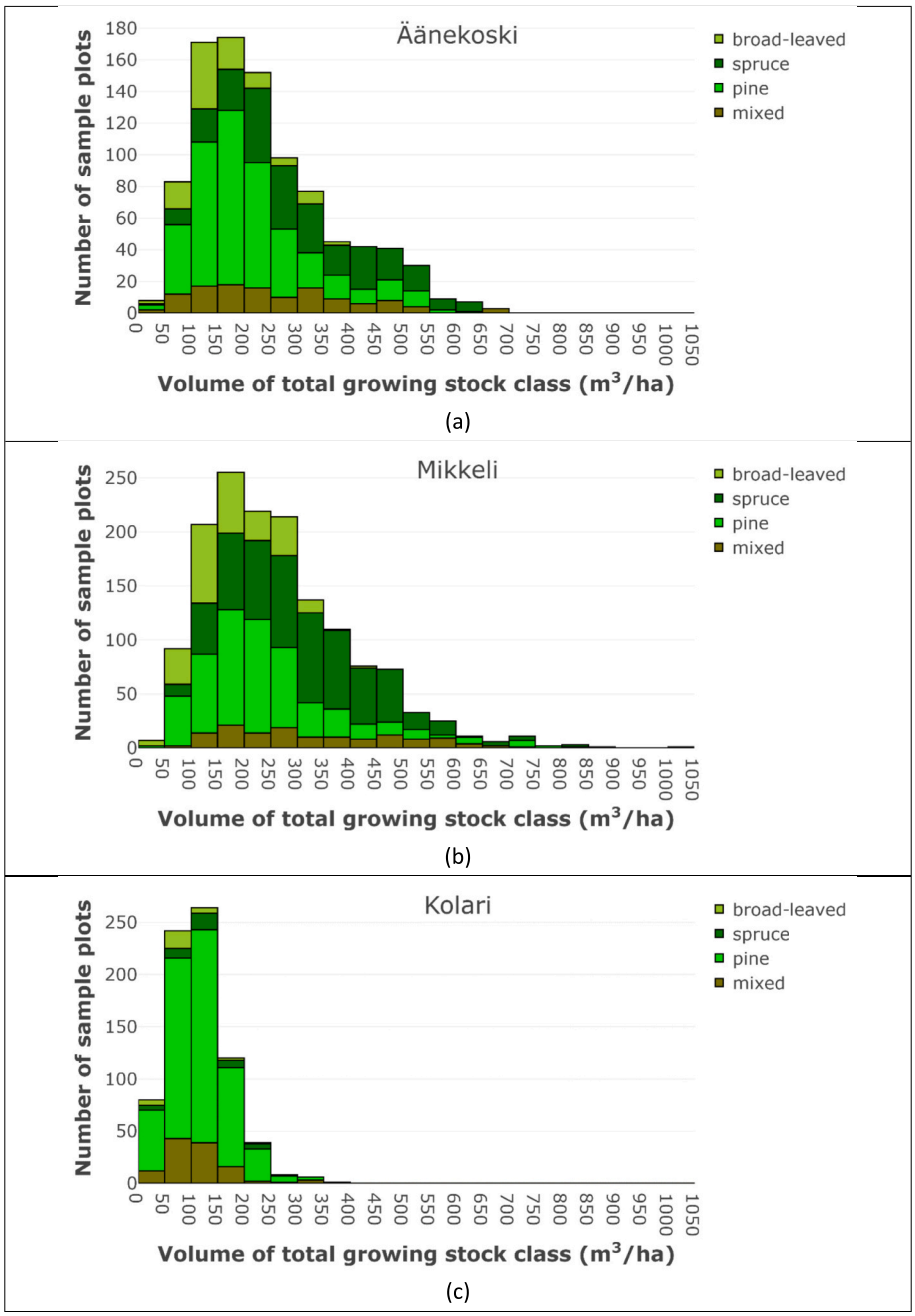


Fig. 3. Distributions of sample plots according to volume of total growing stock for each study area; color codes indicate dominating tree species of sample plots; the minimum share for a species to be dominating was 60% of the total volume, if none of the species reached 60%, the plot was mixed.

Texture features were calculated from grey-level co-occurrence matrices, derived from first and single returns using radiomics package (Hall-Beyer, 2017) in R statistical software.

We found that there is a high variation in the scale of intensity values between the test areas' ALS data. Due to this issue features based on intensity were excluded from the process except for those representing relative quantities and texture features based on rasterized intensity values. After removing problematic intensity features the total number of features amounted to 186.

k-nearest neighbors.

We compared the performance of the 3D-CNN to the k-NN method. Forest stand variables were estimated by weighted averages of the variables of k nearest neighbors. Nearest neighbors were chosen based on Euclidean distances between sample plots within an m-dimensional feature space, with m indicating the number of remote sensing features used for estimation. Weighing was carried out inversely proportional to the distances of the nearest neighbors as shown in Eq. (1) and Eq. (2) in order to reduce bias (Altman, 1992).

$$\hat{y}_j = \sum_{i=1}^{k} w_i y_{ji} \tag{1}$$

k = number of nearest neighbors,
y_{ji} = field observation of variable j of the i^{th} nearest neighbor,
$\hat{y}_j$ = estimate for variable j,
weight of the i^{th} nearest neighbor (Eq. (1):

$$w_i = \frac{1}{d_i^g} \Big/ \sum_{i=1}^{k} \frac{1}{d_i^g} \tag{2}$$

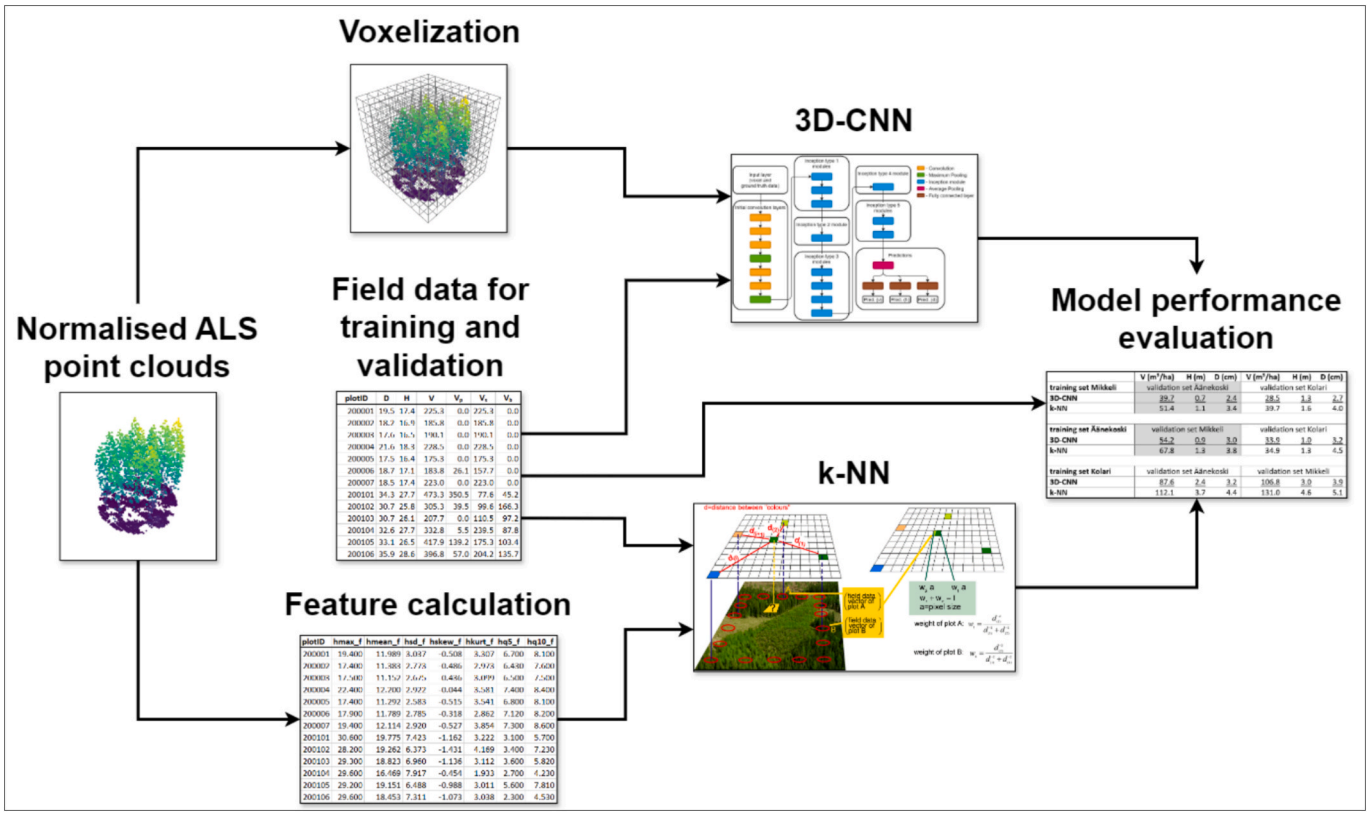


Fig. 4. Workflow.

g = distance weighting parameter.

d_i = Euclidean distance of the i^{th} nearest neighbor,

Multiple values for both parameters k and g were tested, finally 4 and 3.0 were selected for k and g, respectively.

We used a genetic algorithm, that is available as the R statistical software package, genalg (Willighagen and Ballings, 2015). Initially, the algorithm randomly selected 100 feature combinations. During 30 iterations the feature combinations changed, working towards an optimum, which was defined by the sum of absolute values of biases and RMSEs of k-NN estimates (Eq. (3), Eq. (4) and Eq. (5). The feature combination that had the smallest evaluation value (EV) was selected. The number of initial feature combinations and iterations were chosen empirically.

$$EV = \sum_{j=1}^{J} \left(RMSE_j + |bias_j|\right) \tag{3}$$

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{n_t} \left(\hat{y}_{ji} - y_{ji}\right)^2}{n_t}} \tag{4}$$

$$bias_j = \frac{\sum_{i=1}^{n_t} \left(\hat{y}_{ji} - y_{ji}\right)}{n_t} \tag{5}$$

$RMSE_j$ = RMSE of variable j,

$bias_j$ = bias of variable j,

y_{ji} = field observation of variable j and sample plot i,

n_t = number of sample plots in the training set,

$\hat{y}_{ji}$ = estimate for variable j and sample plot i.

3D Convolutional Neural Network.

In this study we have used a 3D-CNN, that is loosely based on the Inception V3 network (Szegedy et al., 2016) by Google, and was developed by Ayrey and Hayes (2018). In a 3D-CNN three-dimensional convolutional kernels are scanning through the voxel space and trying to extract relevant information. Five initial convolutional layers are followed by eleven inception layers which are groups of various convolutional layers. Pooling and batch normalization layers are used between convolutional layers to reduce dimensionality and facilitate faster and stable training (Ioffe and Szegedy, 2015). The architecture of the CNN was described in detail by Balazs et al. (2022) and was slightly modified in order to be able to use voxels of various sizes.

The final output section of the CNN was modified to be able to estimate volumes of tree species that add up to the total volume. Tree species' volume estimates were calculated by passing the species' estimates through a Softmax activation function (Eq. (6) (Wikipedia contributors, 2024) obtaining ratios adding up to 1 (Bolyn et al., 2022). The estimate of total volume was then multiplied by the output of the Softmax function resulting in the absolute values of species' volumes.

$$Softmax(\boldsymbol{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} for\, i = 1, \ldots, N\, and\, \boldsymbol{z} = (z_1, \ldots, z_N) \in \mathbb{R}^N \tag{6}$$

e = exponential function

N = number of elements in vector z.

$Softmax(z)_i$ = the i^{th} element of input vector z after Softmax activation

Data was fed into the CNN in small batches with an initial batch size of 16, which was doubled through automatic augmentation before being input into the network. The augmentation was carried out by randomly rotating the voxel spaces by 90/180/270 degrees around the z axis. This additional augmentation step significantly improved the performance of the 3D-CNN. Learning rate defines how much the neural network changes between iterations. The learning rate was assigned a value of 0.0001, and the number of training iterations 20 000. Mean squared error (MSE) (Eq. (7) was used as loss function, which is recommended to be used with regression problems (Brownlee, 2019b). To avoid

overfitting, a common disadvantage of neural networks (Bejani and Ghatee, 2021), the model that produced the lowest sum of absolute root-mean-square errors (RMSE) (Eq. (4) on the validation dataset was saved to disk.

$$MSE_j = \frac{\sum_{i=1}^{n_v}\left(\hat{y}_{ji} - y_{ji}\right)^2}{n_v} \quad (7)$$

$\hat{y}_{ji}$ = estimate for variable j and sample plot i,
y_{ji} = field observation of variable j and sample plot i,
n_v = number of sample plots in the validation set.

The 3D-CNN was trained using NVIDIA Volta V100 GPUs with 32 GB GPU memory and each training run took approximately 4 h.

3.2. Validation

Both methods investigated in our study have a high amount of randomly initiated parameters. To account for the randomness each model training was carried out independently ten times. The model performance was assessed based on the two datasets not involved in the training process and results were reported separately for each set, resulting in 6 different training/validation combinations. Selected performance indicators were absolute and relative RMSE (Eq. (8) and bias (Eq. (9). As we were interested in the magnitude of bias, the absolute values of bias results were used. The mean of performance indicators over 10 runs was finally reported.

$$RMSE\% = 100 \cdot \frac{RMSE}{\bar{y}} \quad (8)$$

$$bias\% = 100 \cdot \frac{|bias|}{\bar{y}} \quad (9)$$

$\bar{y}$ = mean of ground observations of all circular plots of all study areas.

We also verified if the differences between the tested methods were statistically significant. To accomplish this, we carried out a *t*-test or Mann-Whitney *U* test based on RMSE and bias values of the 10 runs and for each validation dataset and response variable combination. If the tested values were not normally distributed according to the Shapiro-Wilk test, Mann-Whitney *U* test was used. If the RMSE/bias values of the methods were normally distributed and their variances were homogenous based on Levene's test, Student's *t*-test was applied, in case of inhomogeneous variances Welch's *t*-test was used. Conclusions were drawn based on tests' p-values with a threshold of 0.05.

Our study involved a high number of combinations of various factors: two tested methods, three training datasets, two validation sets for each training set, and six forest variables. To be able to demonstrate the impact of these factors, relative RMSE and bias scores of all training runs and tested methods were input into a regression model. Response variables were relative RMSE/bias, explanatory variables were method, forest variable, training and test dataset.

4. Results

Results of regression models for relative bias and RMSE are displayed in Table 2 and Table 3. Results of the RMSE regression indicate, that each term has a statistically significant impact on the prediction accuracy. According to Table 2, k-NN performs worse than CNN if we take all forest attributes, all training/validation dataset combinations, and results of all 10 training runs into account. Predictions of mean height are more accurate than mean diameter, but performance in terms of volume predictions, especially volumes of tree species are worse than mean diameter. Predictions are most accurate when models are trained by the Äänekoski dataset, and the best performing validation set is Kolari.

Results of the bias-regression indicate that on average k-NN produced less biased predictions. Unsurprisingly, volumes of tree species

Table 2
Regression analysis results of relative RMSE scores; the constant term includes the followings: method: 3D-CNN, forest variable: D, training set: Äänekoski, and validation set: Kolari.

Term	Coefficient	Standard Error	T-value	P-value
Constant	−16.9	2.9	−5.9	0.000
method: k-NN	9.2	1.7	5.3	0.000
forest variable: H	−5.8	3.0	−2.0	0.050
forest variable: V	13.7	3.0	4.6	0.000
forest variable: V_s	67.2	3.0	22.5	0.000
forest variable: V_b	118.4	3.0	39.7	0.000
forest variable: V_p	67.9	3.0	22.8	0.000
training set: Mikkeli	7.1	2.4	2.9	0.004
training set: Kolari	42.2	2.4	17.4	0.000
validation set: Mikkeli	27.5	2.4	11.3	0.000
validation set: Äänekoski	7.5	2.4	3.1	0.002

Table 3
Regression analysis results of relative bias scores; the constant term is as in the RMSE regression.

Term	Coefficient	Standard Error	T-value	P-value
Constant	2.4	1.7	1.4	0.157
method: k-NN	−4.9	1.0	−4.7	0.000
forest variable: H	0.8	1.8	0.5	0.638
forest variable: V	3.6	1.8	2.0	0.044
forest variable: V_s	29.6	1.8	16.7	0.000
forest variable: V_b	16.2	1.8	9.1	0.000
forest variable: V_p	23.3	1.8	13.1	0.000
training set: Mikkeli	6.5	1.5	4.5	0.000
training set: Kolari	26.1	1.5	18.0	0.000
validation set: Mikkeli	−5.4	1.5	−3.7	0.000
validation set: Äänekoski	−12.1	1.5	−8.3	0.000

are significantly more biased than mean diameter. Equally, estimates produced by Kolari trained models or models validated with the Kolari dataset are markedly more biased than the base level of factors.

RMSE and bias results of all training and validation set combinations and target variables along with the results of significance tests can be found in Appendix A. Fig. 5 shows the performance metrics of models trained with the Mikkeli dataset. Fig. 5 (a) demonstrates well that the 3D-CNN performed better than k-NN in terms of absolute RMSE of total growing stock variables. The same applies to all combinations of training and validation datasets (Table A1). Except for one case, the differences of absolute RMSE scores between the methods were statistically significant. The results of corresponding statistical tests are reported in Table A3. While differences in RMSE for mean height and diameter were negligible, in most cases CNN produced significantly better results predicting total growing stock volumes. The differences between RMSE scores were, for example 11.7 m^3/ha, 13.6 m^3/ha, and 24.5 m^3/ha when models were Mikkeli-trained and Äänekoski-validated, Äänekoski-trained and Mikkeli-validated, and Kolari-trained and Äänekoski-validated, respectively. For mean values of forest variables refer to Table 4.

According to the bias results shown in Fig. 5 (b) we can see that k-NN produced less biased results than the 3D-CNN for volume of total growing stock when models were validated with the Äänekoski dataset. In all other cases the neural network produced the same or slightly better bias results than k-NN. However, upon examining the summary of results in Table A5 and the corresponding significance values in Table A7, it becomes evident that approximately half of the differences of bias results were statistically insignificant. Out of the statistically significant cases k-NN performed better than 3D-CNN in only one case. Overall, volume predictions' biases for validation datasets from the same bioregion as the training set were considerably lower (2.3–7.2 m^3/ha) compared to validation data from a different bioregion (12.2–48.8 m^3/

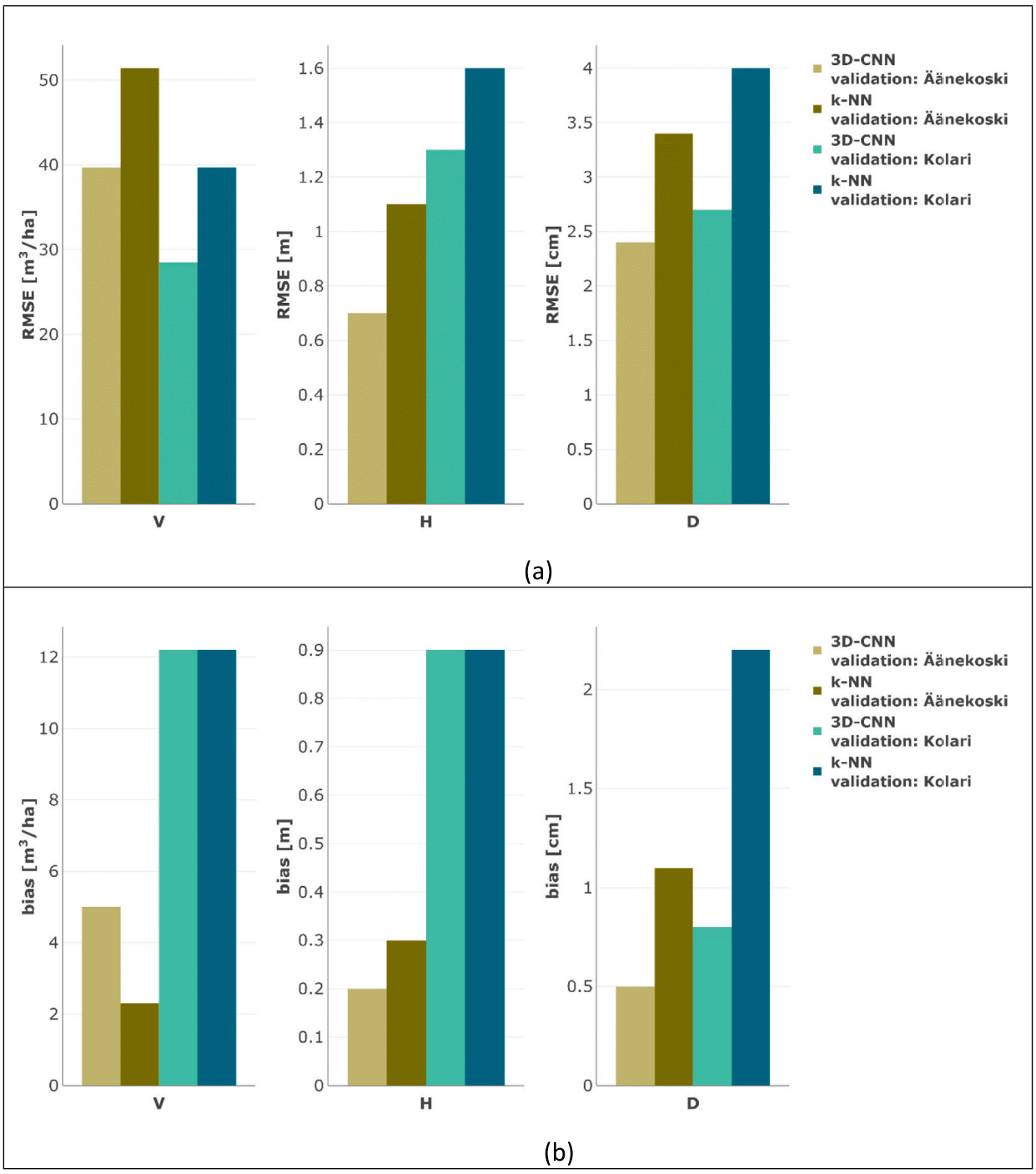


Fig. 5. Mean absolute RMSE (a) and bias (b) results of total growing stock predictions of models trained with the Mikkeli dataset.

Table 4
Mean values of forest attributes for each test area.

	V (m^3/ha)	H (m)	D (cm)	V_p (m^3/ha)	V_s (m^3/ha)	V_b (m^3/ha)
Mikkeli[1]	266.4	21.4	24.9	96.6	131.1	38.7
Äänekoski[1]	241.1	20.0	22.9	109.1	101.3	30.7
Kolari[2]	114.4	13.9	19.3	85.6	15.0	13.8

[1] Southern Boreal Region
[2] Northern Boreal Region

ha). Biases of mean height and diameter predictions were less affected by the geographic location of training and validation datasets, except for models trained by the Kolari dataset where biases of height predictions varied between 1.7 and 3.5 m.

The example in Fig. 6 shows RMSE and bias results regarding tree species-specific volume predictions of models trained by the Äänekoski dataset. RMSE scores of CNN models were lower than k-NN scores by 33.8, 22.8, and 23.7 m^3/ha for volume of pine, spruce and broad-leaved, respectively. In general, our CNN model performed significantly better in terms of RMSE when training and validation sets were from the same bioregion (Table A2 and Table A4). Differences in bias show no clear pattern, but apart from the Mikkeli/Äänekoski training/validation

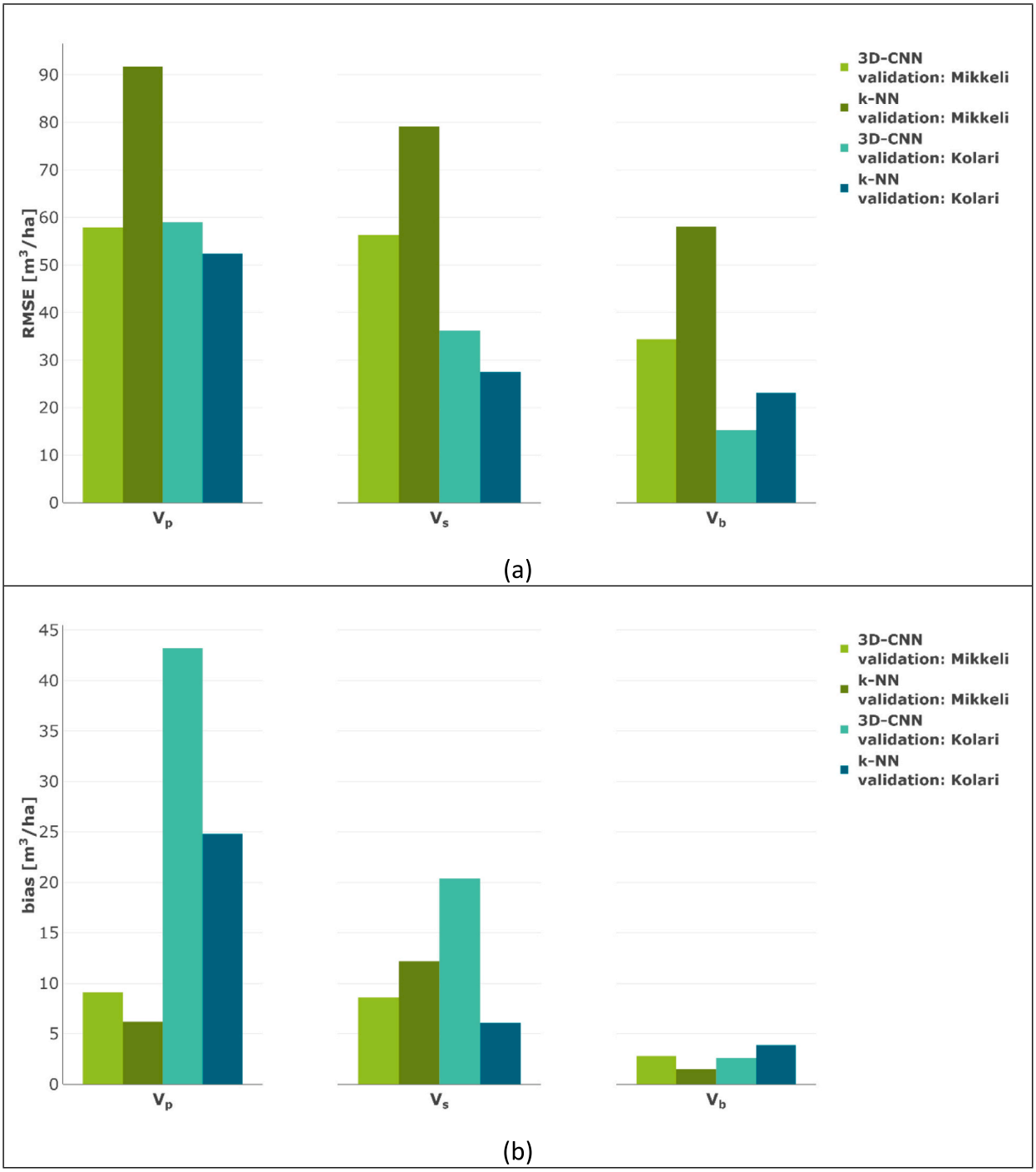


Fig. 6. Mean absolute RMSE (a) and bias (b) results of tree species' growing stock predictions of models trained with the Äänekoski dataset.

combination and volume of pine, the differences between methods within the same bioregion were statistically insignificant (Table A6 and Table A8). When training and validation datasets were from different bioregions (Kolari dataset used for validation or training), k-NN models produced markedly and statistically significantly less biased predictions in most of the cases.

In addition to RMSE and bias scores, we also compared the distribution of predicted forest variables to the ground truth data. For this purpose, predictions of the best training run were selected for both tested methods based on the lowest mean of absolute RMSE value of all forest variables. Below we are showing some examples how distributions of predictions relate to field observations.

Fig. 7 demonstrates how well CNN predictions follow the distribution of observed values, whereas k-NN is underestimating mean height and volume of growing stock. Training and validation datasets are located in the same bioregion.

Models trained with Äänekoski and validated with the Kolari dataset indicate that k-NN is overestimating total growing stock variables, while CNN predictions' distribution remains close to the ground observations (Fig. 8).

Fig. 9 shows the effect of the training data set's bioregion on the volume predictions of pine. In case of both datasets being from the same bioregion (in orange), predictions' distributions are closer to the distributions of field observations.

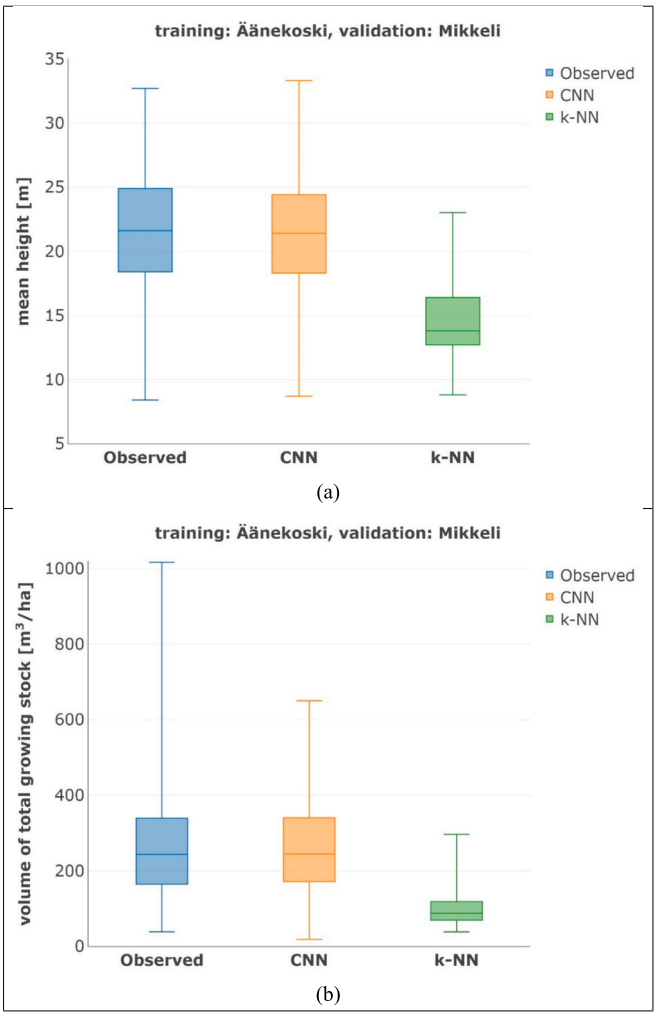


Fig. 7. Distributions of observed and predicted mean heights (a) and volumes of the total growing stock (b). Training and validation datasets are situated in the same bioregion. Boxes span from quartile 1 to 3, and the median is marked with a line inside the box; lower and upper fences indicate the minimum and maximum values, respectively.

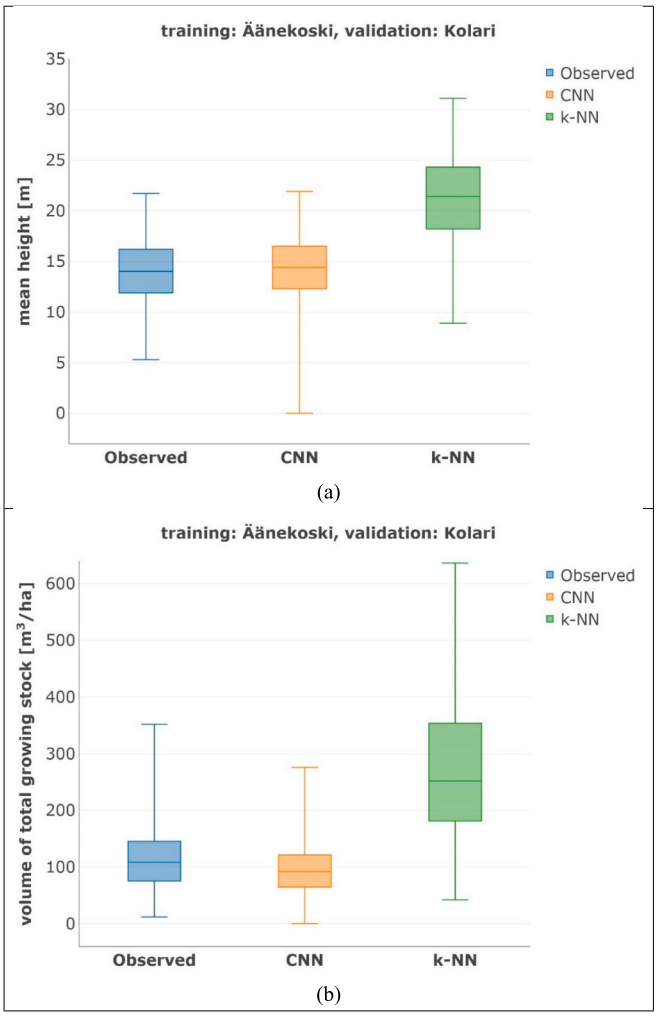


Fig. 8. Distributions of observed and predicted mean heights (a) and volumes of the total growing stock (b). Training and validation datasets are situated in different bioregions.

The impact of bioregions is even more obvious for tree species-specific variables, such as volume of pine shown in Fig. 10.

5. Discussion

In our study, we aimed to compare a 3-D convolutional neural network to the k-NN method in predicting forest stand variables using training and validation datasets from three different study areas in two different bioregions.

In terms of RMSE, the CNN consistently outperformed k-NN in predicting forest attributes such as mean diameter, height, and total volume, which can be easily derived from ALS data. In almost all cases, where differences were statistically significant, the CNN also produced smaller bias than k-NN. However, for tree species-specific volumes, where the relationship to the point cloud structure is more complex, the results were inconclusive. While the CNN provided better prediction accuracy for species-specific volumes in slightly more than half of the cases, it also tended to produce higher bias compared to k-NN.

When using training and validation data from the same bioregion, CNN predictions were more accurate in terms of RMSE for all forest stand variables. Differences between the tested methods were without exception statistically significant. Concerning bias, differences were either statistically insignificant or the CNN produced less bias with one exception.

When using training and validation data from different bioregions, the results were somewhat ambiguous. When predicting variables of northern boreal forest based on southern boreal training data, the RMSE of diameter and height did not change markedly. RMSEs of volume predictions decreased, apparently due to the lower target volumes in the Kolari study site. Predictions of forest stand variables of total growing stock produced by the CNN were more accurate compared to k-NN. However, for tree species-specific volumes, the prediction accuracy decreased markedly, and the results do not indicate any superiority of CNN based predictions. The cases, where the accuracy ranking of CNN and k-NN varied inconclusively, were primarily linked to tree species-specific volumes where training and validation data were from different bioregions. This indicates that the accuracy advantage of CNN may disappear when applying the model in areas with significantly different forests compared to the training data.

For the prediction of forest variables, the least workable arrangement was using training set from the northern boreal study site (Kolari) and validation data from the southern boreal region. Compared to other cases, this setup resulted in highest RMSEs and bias for all predicted forest variables. This is mainly because the variation of forest variables in the northern boreal site covers only part of the variation present in southern boreal areas, and the densely stocked or otherwise highly productive forests are missing from the training data. Thus k-NN is lacking suitable reference data, and the CNN model fails to adapt to

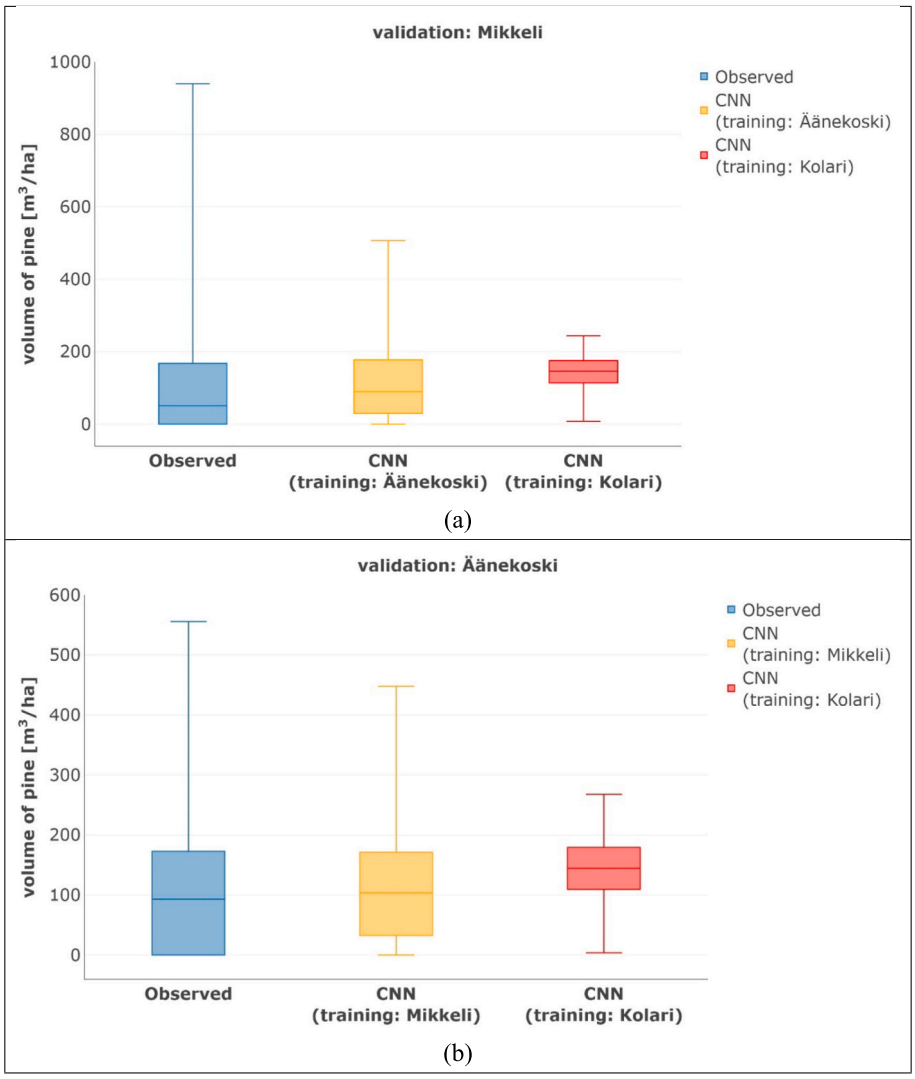


Fig. 9. Distributions of observed and predicted volumes of pine with Mikkeli (a) and Äänekoski (b) validation set.

significantly different validation data compared to the training dataset.

Our results indicate that the 3D neural network used in this study is performing significantly better than the benchmark k-NN method when it comes to generalizability of models predicting total growing stock variables. K-NN is restricted by the distribution of potential neighbors in the training data, while the CNN appears to be able to model the actual physical relationships between the laser scanning data and the measured variables. This applies to variables having clear physical dimensions.

Regarding tree species-specific variables, laser scanning data at 5 pulses/m^2 density does not provide physical information, which would help to accurately differentiate between tree species. As a result, k-NN's data-driven approach in some cases performed better than the CNN. However, RMSE values in Table A2 suggest, that CNN was consistently more accurate than k-NN, when both the training and validation data were from the same bioregion (Mikkeli/Äänekoski). This shows that the CNN is able to find useful patterns in the ALS data, which allow more accurate estimates than the manually extracted features used by k-NN. Table A2 also indicates, that these patterns don't seem to work as consistently when the training and validation datasets are significantly different (Kolari/Äänekoski or Kolari/Mikkeli). In these cases, the k-NN method mostly performed better than the neural network. The reason for this includes an element of randomness and may be specific to the particular data we used.

Ayrey et al. (2021) compared the same 3D-CNN to random forest (RF) models using ALS and satellite data to predict various forest stand variables. They found that while the CNN outperformed RF in predicting 12 out of the 14 target variables, RF was proven to be superior in estimating tree species proportions. The authors speculated that the reason for this was RF's ability to make better use of the spectral information than the CNN. It was also found that CNN did not outperform RF regarding bias, which is in accordance with our overall findings. Oehmcke et al. (2024) reported that their MSENet50 neural network outperformed a benchmark power regression model in predicting above ground biomass and volume of total growing stock in terms of RMSE and bias.

In our previous study (Balazs et al., 2022) utilizing sparse (0.5 pulse/m^2) ALS data, we found that the CNN improved prediction accuracy of for example total growing stock volume by a narrow margin of 1.3 % RMSE. In our current research, the CNN outperformed k-NN by 4.8 % and 5.2 % in terms of RMSE of total growing stock volume when training and validation datasets were from the same bioregion. This leads us to believe, that increasing pulse density of ALS data could improve the CNN's performance and make it a viable alternative to k-NN.

Point cloud voxelization might lead to loss of information. However, neural networks are also capable of handling point clouds in their original form, without the need for voxelization (e.g. Oehmcke et al., 2024). This method could further improve forest attribute estimates utilizing ALS data and CNNs.

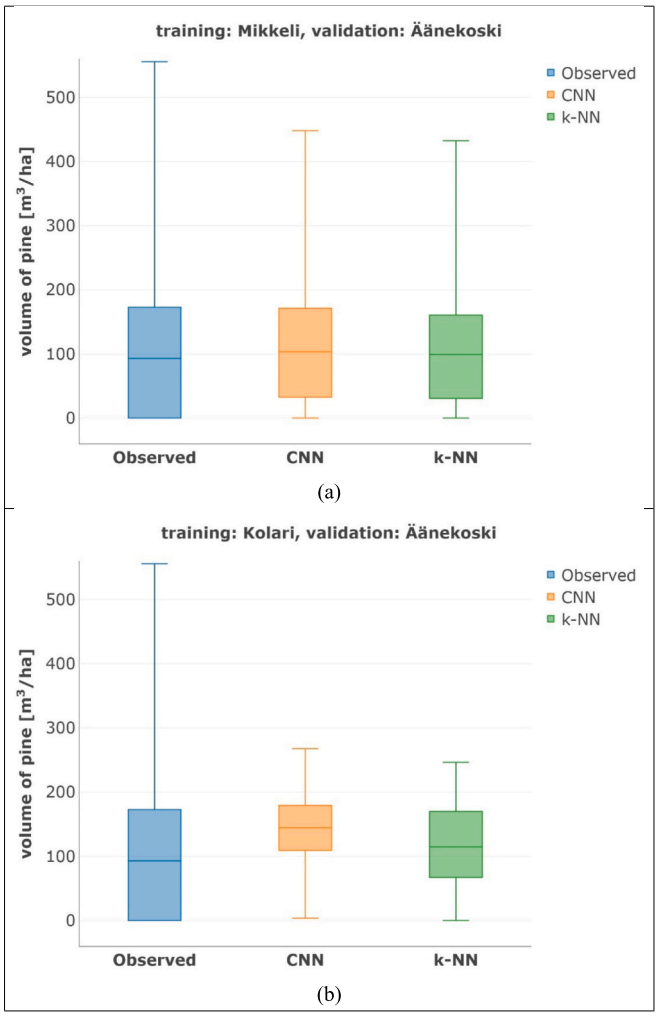


Fig. 10. Distributions of observed and predicted volumes of pine with training and validation sets in the same (a) and in different bioregions (b).

Furthermore, due to the sampling design of the ground truth data used in this study, it was not possible to create validated wall-to-wall forest attribute maps. In future studies it would be desirable to utilize suitable field data and carry out a similar comparison of such maps.

6. Conclusions

CNN generally performed markedly better than k-NN in predicting forest attributes, when training and validation datasets were from the same bioregion.

Using southern boreal training data for northern boreal study site (with sparsely stocked forest of low productivity) also achieved acceptable prediction accuracy with both CNN and k-NN. In contrast, applying northern boreal training data for southern boreal study sites resulted in low prediction accuracy and high amount of bias in the predictions with both methods.

When predicting forest variables that are directly connected to the dimension and volume of growing stock, CNN based predictions are more accurate and less biased than predictions based on k-NN. The same applies when predicting tree species-specific volumes within the same bioregion as the training data. On the other hand, k-NN often results in better accuracy and less bias of tree species-specific predictions, when training and validation data are from different bioregions.

Due to our model validation approach, we were also able to show that the 3D-CNN applied in our study is generalizable at least within the same bioregion and prediction accuracies outside the training dataset are at an acceptable level.

As a next step we are considering the integration of high-resolution aerial images into the training and prediction process of the 3D-CNN, which could help improve tree species-specific volume estimates. Aerial imagery containing information at least in the spectrum of green, red and near-infrared light is useful in differentiating between tree species. The higher the spectral resolution, the better results can be achieved (Briechle et al., 2020).

CRediT authorship contribution statement

Andras Balazs: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Sakari Tuominen:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Investigation, Data curation. **Annika Kangas:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Research Council of Finland's flagship project UNITE (decision numbers 357909 and 359174) and by the Natural Resources Institute Finland. The authors also wish to acknowledge CSC-IT Center for Science, Finland, for computational resources, as well as Dr. Mikko Kuronen (Natural Resources Institute Finland) for his valuable support in statistical testing of the results.

Appendix A

Model performances and corresponding significance test results.

Table A1
Mean RMSE results of total growing stock predictions (lower values with bold font).

	V (m^3/ha)	H (m)	D (cm)	V (m^3/ha)	H (m)	D (cm)
training set Mikkeli	validation set Äänekoski			validation set Kolari		
3D-CNN	**39.7**	**0.7**	**2.4**	**28.5**	**1.3**	**2.7**
k-NN	51.4	1.1	3.4	39.7	1.6	4.0
training set Äänekoski	validation set Mikkeli			validation set Kolari		

(*continued on next page*)

Table A1 (*continued*)

	V (m^3/ha)	H (m)	D (cm)	V (m^3/ha)	H (m)	D (cm)
3D-CNN	**54.2**	**0.9**	**3.0**	**33.9**^{si}	**1.0**	**3.2**
k-NN	67.8	1.3	3.8	34.9^{si}	1.3	4.5
training set Kolari	validation set Äänekoski			validation set Mikkeli		
3D-CNN	**87.6**	**2.4**	**3.2**	**106.8**	**3.0**	**3.9**
k-NN	112.1	3.7	4.4	131.0	4.6	5.1

^{si} statistically insignificant results ($p \geq 0.05$).

Table A2
Mean RMSE results of tree species' growing stock predictions (m^3/ha).

	V_p	V_s	V_b	V_p	V_s	V_b
training set Mikkeli	validation set Äänekoski			validation set Kolari		
3D-CNN	**55.6**	**45.7**	**29.3**	68.3	54.3	**18.5**
k-NN	71.2	60.0	44.1	**62.7**	**33.6**	26.6
training set Äänekoski	validation set Mikkeli			validation set Kolari		
3D-CNN	**57.9**	**56.3**	**34.4**	59.0	36.2	**15.3**
k-NN	91.7	79.1	58.1	**52.4**	**27.5**	23.1
training set Kolari	validation set Äänekoski			validation set Mikkeli		
3D-CNN	108.0	**112.8**	52.1	133.1	**141.2**	69.2^{si}
k-NN	**99.3**	133.3	**50.0**	**121.8**	162.0	**68.4**^{si}

Table A3
P-values of statistical significance tests for absolute RMSE results of total growing stock (p-values ≥ 0.05 with bold font, i.e. difference between methods statistically insignificant).

	V	H	D	V	H	D
training set Mikkeli	validation set Äänekoski			validation set Kolari		
	$< 0.001^M$	$< 0.001^S$	$< 0.001^S$	$< 0.001^M$	$< 0.001^M$	$< 0.001^M$
training set Äänekoski	validation set Mikkeli			validation set Kolari		
	$< 0.001^S$	$< 0.001^M$	$< 0.001^S$	**0.280**^{M}	$< 0.001^W$	$< 0.001^W$
training set Kolari	validation set Äänekoski			validation set Mikkeli		
	$< 0.001^S$	$< 0.001^S$	$< 0.001^S$	$< 0.001^S$	$< 0.001^S$	$< 0.001^S$

^{S} Student's *t*-test.
^{W} Welch's *t*-test.
^{M} Mann-Whitney U.

Table A4
P-values of statistical significance tests for absolute RMSE results of tree species' growing stock.

	V_p	V_s	V_b	V_p	V_s	V_p
training set Mikkeli	validation set Äänekoski			validation set Kolari		
	$< 0.001^S$	$< 0.001^S$	$< 0.001^S$	0.035^M	$< 0.001^S$	$< 0.001^W$
training set Äänekoski	validation set Mikkeli			validation set Kolari		
	$< 0.001^M$	$< 0.001^S$	$< 0.001^S$	0.008^W	$< 0.001^W$	$< 0.001^S$
training set Kolari	validation set Äänekoski			validation set Mikkeli		
	0.001^M	$< 0.001^S$	0.008^S	$< 0.001^S$	$< 0.001^S$	**0.290**^{W}

Table A5
Mean biases of total growing stock predictions (lower values with bold font).

	V (m^3/ha)	H (m)	D (cm)	V (m^3/ha)	H (m)	D (cm)
training set Mikkeli	validation set Äänekoski			validation set Kolari		
3D-CNN	5.0^{si}	**0.2**	**0.5**	**12.2**^{si}	**0.9**^{si}	**0.8**
k-NN	**2.3**^{si}	0.3	1.1	**12.2**^{si}	**0.9**^{si}	2.2
training set Äänekoski	validation set Mikkeli			validation set Kolari		
3D-CNN	**4.1**	**0.3**^{si}	**0.7**^{si}	23.4	0.5^{si}	**2.0**
k-NN	7.2	**0.3**^{si}	0.9^{si}	**14.8**	**0.3**^{si}	3.0
training set Kolari	validation set Äänekoski			validation set Mikkeli		
3D-CNN	**28.5**	**1.7**	**0.6**^{si}	**36.1**	**2.3**	0.7^{si}
k-NN	35.9	2.4	0.7^{si}	48.8	3.5	**0.6**^{si}

^{si} statistically insignificant results ($p \geq 0.05$).

Table A6
Mean biases of tree species' growing stock predictions (m^3/ha).

	V_p	V_s	V_b	V_p	V_s	V_b
training set Mikkeli	validation set Äänekoski			validation set Kolari		
3D-CNN	9.4	13.5^{si}	**5.2**^{si}	51.1	42.6	**3.7**^{si}
k-NN	**3.2**	**12.7**^{si}	7.6^{si}	**33.3**	**16.4**	4.6^{si}
training set Äänekoski	validation set Mikkeli			validation set Kolari		
3D-CNN	9.1^{si}	**8.6**^{si}	2.8^{si}	43.2	20.4	**2.6**
k-NN	**6.2**^{si}	12.2^{si}	**1.5**^{si}	**24.8**	**6.1**	3.9
training set Kolari	validation set Äänekoski			validation set Mikkeli		
3D-CNN	37.7	**48.3**	17.9	58.3	**73.1**^{si}	21.2
k-NN	**22.7**	58.8	**2.9**	**33.4**	80.9^{si}	**3.3**

Table A7
P-values of statistical significance tests for absolute bias results of total growing stock (p-values ≥ 0.05 with bold font, i.e. difference between methods statistically insignificant).

	V	H	D	V	H	D
training set Mikkeli	validation set Äänekoski			validation set Kolari		
	0.064^{W}	0.010^{S}	0.001^{W}	**0.977**^{W}	**0.307**^{M}	$< 0.001^{M}$
training set Äänekoski	validation set Mikkeli			validation set Kolari		
	0.030^{S}	**0.197**^{W}	**0.157**^{W}	$< 0.001^{M}$	**0.082**^{W}	$< 0.001^{W}$
training set Kolari	validation set Äänekoski			validation set Mikkeli		
	0.011^{S}	$< 0.001^{S}$	**0.325**^{M}	0.002^{S}	$< 0.001^{S}$	**0.491**^{S}

Table A8
P-values of statistical significance tests for absolute bias results of tree species' growing stock.

	V_p	V_s	V_b	V_p	V_s	V_b
training set Mikkeli	validation set Äänekoski			validation set Kolari		
	0.022^{S}	**0.740**^{W}	**0.127**^{W}	$< 0.001^{W}$	$< 0.001^{S}$	**0.423**^{S}
training set Äänekoski	validation set Mikkeli			validation set Kolari		
	0.473^{M}	**0.169**^{W}	**0.108**^{W}	$< 0.001^{S}$	$< 0.001^{W}$	0.035^{S}
training set Kolari	validation set Äänekoski			validation set Mikkeli		
	0.003^{S}	0.013^{S}	$< 0.001^{W}$	$< 0.001^{S}$	**0.110**^{S}	$< 0.001^{W}$

Data availability

Data and code will be made available on GitHub when the paper is accepted.

References

Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. 46, 175–185. https://doi.org/10.1080/00031305.1992.10475879.

Ayrey, E., Hayes, D.J., 2018. The Use of Three-Dimensional Convolutional Neural Networks to Interpret LiDAR for Forest Inventory. Remote Sens. 10, 649. https://doi.org/10.3390/rs10040649.

Ayrey, E., Hayes, D.J., Kilbride, J.B., Fraver, S., Kershaw, J.A., Cook, B.D., Weiskittel, A. R., 2021. Synthesizing Disparate LiDAR and Satellite Datasets through Deep Learning to Generate Wall-to-Wall Regional Inventories for the Complex, Mixed-Species Forests of the Eastern United States. Remote Sens. 13, 5113. https://doi.org/10.3390/rs13245113.

Balazs, A., Liski, E., Tuominen, S., Kangas, A., 2022. Comparison of neural networks and k-nearest neighbors methods in forest stand variable estimation using airborne laser data. ISPRS Open J. Photogramm. Remote Sens. 4, 100012. https://doi.org/10.1016/j.ophoto.2022.100012.

Bejani, M.M., Ghatee, M., 2021. A systematic review on overfitting control in shallow and deep neural networks. Artif. Intell. Rev. 54, 6391–6438. https://doi.org/10.1007/s10462-021-09975-1.

Bello, S.A., Yu, S., Wang, C., Adam, J.M., Li, J., 2020. Review: Deep Learning on 3D Point Clouds. Remote Sens. 12, 1729. https://doi.org/10.3390/rs12111729.

Bolyn, C., Lejeune, P., Michez, A., Latte, N., 2022. Mapping tree species proportions from satellite imagery using spectral–spatial deep learning. Remote Sens. Environ. 280, 113205. https://doi.org/10.1016/j.rse.2022.113205.

Briechle, S., Krzystek, P., Vosselman, G., 2020. Classification of tree species and standing dead trees by fusing UAV-based LiDAR data and multispectral imagery in the 3D deep neural network Pointnet++. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. V-2–2020, 203–210. Doi: 10.5194/isprs-annals-V-2-2020-203-2020.

Brownlee, J., 2019a. How to use Data Scaling Improve Deep Learning Model Stability and Performance. accessed 4.27.23 MachineLearningMastery.com. https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/.

Brownlee, J., 2019b. How to Choose Loss Functions When Training Deep Learning Neural Networks. accessed 4.28.25 MachineLearningMastery.com. https://www.machinelearningmastery.com/how-to-choose-loss-functions-when-training-deep-learning-neural-networks/.

Brownlee, J., 2017. What is the Difference Between Test and Validation Datasets? MachineLearningMastery.com. URL https://machinelearningmastery.com/difference-test-validation-datasets/ (accessed 3.17.23).

Corte, A.P.D., Souza, D.V., Rex, F.E., Sanquetta, C.R., Mohan, M., Silva, C.A., Zambrano, A.M.A., Prata, G., Alves de Almeida, D.R., Trautenmüller, J.W., Klauberg, C., de Moraes, A., Sanquetta, M.N., Wilkinson, B., Broadbent, E.N., 2020. Forest inventory with high-density UAV-Lidar: Machine learning approaches for predicting individual tree attributes. Comput. Electron. Agric. 179, 105815. https://doi.org/10.1016/j.compag.2020.105815.

Deng, S., Katoh, M., Yu, X., Hyyppä, J., Gao, T., 2016. Comparison of Tree Species Classifications at the Individual Tree Level by Combining ALS Data and RGB Images Using Different Algorithms. Remote Sens. 8, 1034. https://doi.org/10.3390/rs8121034.

Fassnacht, F.E., White, J.C., Wulder, M.A., Næsset, E., 2024. Remote sensing in forestry: current challenges, considerations and directions. For. Int. J. for. Res. 97, 11–37. https://doi.org/10.1093/forestry/cpad024.

Hall-Beyer, M., 2017. GLCM Texture: A Tutorial v. 3.0 March 2017.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15. JMLR.org, Lille, France, pp. 448–456.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K.T., Maltamo, M., Nilsson, M., Nord-Larsen, T., Næsset, E., Olsson, H., 2018. Remote sensing and forest inventories in Nordic countries–roadmap for the future. Scand. J. for. Res. 33, 397–412. https://doi.org/10.1080/02827581.2017.1416666.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010.

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M.D., Dormann, C.F., 2022. Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. ISPRS Open J. Photogramm. Remote Sens. 5, 100018. https://doi.org/10.1016/j.ophoto.2022.100018.

Kirchhoefer, M., Schumacher, J., Adler, P., Kändler, G., 2017. Considerations towards a Novel Approach for Integrating Angle-Count Sampling Data in Remote Sensing Based Forest Inventories. Forests 8, 239. https://doi.org/10.3390/f8070239.

Kostensalo, J., Mehtätalo, L., Tuominen, S., Packalen, P., Myllymäki, M., 2023. Recreating structurally realistic tree maps with airborne laser scanning and ground measurements. Remote Sens. Environ. 298, 113782. https://doi.org/10.1016/j.rse.2023.113782.

Luopa, T., 2022. Puuston sijaintitarkkuus Metsäkeskuksen puukarttakoealoilla [WWW Document]. URL http://www.theseus.fi/handle/10024/781777 (accessed 4.11.25).

Maltamo, M., Packalen, P., Kangas, A., 2021. From comprehensive field inventories to remotely sensed wall-to-wall stand attribute data — a brief history of management inventories in the Nordic countries. Can. J. for. Res. 51, 257–266. https://doi.org/10.1139/cjfr-2020-0322.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5.

Oehmcke, S., Li, L., Trepekli, K., Revenga, J.C., Nord-Larsen, T., Gieseke, F., Igel, C., 2024. Deep point cloud regression for above-ground forest biomass estimation from airborne LiDAR. Remote Sens. Environ. 302, 113968. https://doi.org/10.1016/j.rse.2023.113968.

Pohjankukka, J., Tuominen, S., Pitkänen, J., Pahikkala, T., Heikkonen, J., 2018. Comparison of estimators and feature selection procedures in forest inventory based on airborne laser scanning and digital aerial imagery. Scand. J. for. Res. 33, 681–694. https://doi.org/10.1080/02827581.2018.1482955.

Prieur, J.-F., St-Onge, B., Fournier, R.A., Woods, M.E., Rana, P., Kneeshaw, D., 2021. A Comparison of Three Airborne Laser Scanner Types for Species Identification of Individual Trees. Sensors 22, 35. https://doi.org/10.3390/s22010035.

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Roussel, J.-R., 2024. stdmetrics: Predefined standard metrics functions [WWW Document]. URL https://rdrr.io/github/Jean-Romain/lidR/man/stdmetrics.html (accessed 10.24.24).

Sarle, W.S., 1995. Stopped training and other remedies for overfitting. Proc. 27th Symp. Interface Comput. Sci. Stat. 352–360.

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., Schmidtlein, S., 2020. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 170, 205–215. https://doi.org/10.1016/j.isprsjprs.2020.10.015.

Seely, H., Coops, N.C., White, J.C., Montwé, D., Winiwarter, L., Ragab, A., 2023. Modelling tree biomass using direct and additive methods with point cloud deep learning in a temperate mixed forest. Sci. Remote Sens. 8, 100110. https://doi.org/10.1016/j.srs.2023.100110.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0.

St-Onge, B., Vega, C., Fournier, R.A., Hu, Y., 2008. Mapping canopy height using a combination of digital stereo-photogrammetry and lidar. Int. J. Remote Sens. 29, 3343–3364. https://doi.org/10.1080/01431160701469040.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, pp. 2818–2826.

Tomppo, E., 1991. Satellite image-based national forest inventory of Finland. Int. Arch. Photogramm. Remote Sens. 28, 419–424.

Tuominen, S., Haapanen, R., 2013. Estimation of forest biomass by means of genetic algorithm-based optimization of airborne laser scanning and digital aerial photograph features. Silva Fenn. 47, 20. https://doi.org/10.14214/sf.902.

Tuominen, S., Pitkänen, T., Balazs, A., Kangas, A., 2017. Improving Finnish Multi-Source National Forest Inventory by 3D aerial imaging. Silva Fenn. 51. https://doi.org/10.14214/sf.7743.

Véga, C., Renaud, J.-P., Durrieu, S., Bouvier, M., 2016. On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters. Remote Sens. Environ. 175, 32–42. https://doi.org/10.1016/j.rse.2015.12.039.

Wielgosz, M., Puliti, S., Wilkes, P., Astrup, R., 2023. Point2Tree(P2T)—Framework for Parameter Tuning of Semantic and Instance Segmentation Used with Mobile Laser Scanning Data in Coniferous Forest. Remote Sens. 15, 3737. https://doi.org/10.3390/rs15153737.

Wikipedia contributors, 2024. Softmax function. Wikipedia.

Willighagen, E., Ballings, M., 2015. genalg: R based genetic algorithm [WWW Document]. URL https://cran.r-project.org/web/packages/genalg/index.html (accessed 3.2.17).

Yu, X., Hyyppä, J., Litkey, P., Kaartinen, H., Vastaranta, M., Holopainen, M., 2017. Single-Sensor Solution to Tree Species Classification Using Multispectral Airborne Laser Scanning. Remote Sens. 9, 108. https://doi.org/10.3390/rs9020108.