

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

# Combination of lidar intensity and texture features enable accurate prediction of common boreal tree species with single sensor UAS data

M. Kukkonen, T. Lähivaara and Petteri Packalen

**Abstract**— We evaluated the performance of unmanned aerial systems (UAS) airborne light detection and ranging (lidar) data in the species classification of pine, spruce, and broadleaf trees. Classifications were conducted with three machine learning (ML) approaches (multinomial logistic regression, random forest, and multilayer perceptron) using features computed from automatically segmented point clouds that represent individual trees. Trees were segmented from the point cloud using a marker-controlled watershed algorithm, and two types of features were computed for each segment: intensity and texture. Textural features were computed from gray-level co-occurrence matrices built from horizontal cross-sections of the point cloud. Intensity features were computed as the average intensity values within voxels. The classification accuracies were validated on 39 rectangular 30 m x 30 m field plots using leave-one-plot out cross-validation. The results showed only very small differences in the classification performance between the different ML approaches. Intensity features provided greater classification accuracy ( $\kappa$  0.73-0.77) than textural features ( $\kappa$  0.60-0.64). However, the best classification results ( $\kappa$  0.81) were achieved when both intensity and textural features were used. Feature importance in the different ML approaches was also similar. We conclude that the accurate classification of the three tree species considered in this study is possible using single sensor UAS lidar data.

**Index Terms**—

## I. INTRODUCTION

TREE species is a crucial component of forest inventories and essential for forest management planning. Optical remote sensing techniques, such as satellite images, aerial images, and images from unmanned aerial systems (UAS), have been shown to be effective in capturing tree species information at different spatial scales. The cost, spatial resolution, and extent of the data are crucial factors in

determining their applicability. Satellite images provide wide coverage with lower spatial resolution, making them ideal for large-scale projects such as national-level forest inventories [1], [2]. UAS data, on the other hand, provides high spatial resolution but covers smaller areas, making it suitable for precision forestry [3], [4]. Aerial imagery offers high to moderate spatial resolution over relatively large areas, making it a preferred option for management-oriented forest inventories [5].

Classification of tree species using optical information relies on the assumption that the amount of radiation reflected at distinct wavelengths is different between tree species. This assumption is true in the sense that the reflectance of radiation is affected by chemical properties of plants, such as water content and photosynthetic pigments, leaf morphology, and canopy structure [6]. However, these discriminating features not only vary between tree species but also within species due to factors such as age and health [7] and for suppressed, co-dominant and dominant trees due to competitive status and shadows [8]. In addition, these features can also be affected by properties that are not directly related to tree species or habitat such as sensor configuration, weather condition, and solar zenith angle. Thus, the operational use of optical data for tree species classification can be problematic.

Airborne light detection and ranging (lidar) data can be used to describe the geometrical properties of the trees. In the context of tree species classification, this information is primarily related to the geometry of crown structure, which is based on, for example, the shape of the crown [9] or crown base height [10]. Moreover, leaf orientation, size, clumping, and foliage density are associated with lidar intensity and have been found useful in the classification of tree species [11]. Tree species classification based solely on geometric, or intensity information can be prone to errors caused by several factors,

This paragraph of the first footnote will contain the date on which you submitted your paper for review, which is populated by IEEE. It is IEEE style to display support information, including sponsor and financial support acknowledgment, here and not in an acknowledgment section at the end of the article. For example, "This work was supported in part by the U.S. Department of Commerce under Grant 123456." The name of the corresponding author appears after the financial information, e.g. (*Corresponding author: Second B. Author*). Here you may also indicate if authors contributed equally or if there are co-first authors.

The next few paragraphs should contain the authors' current affiliations, including current address and e-mail. For example, First A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@boulder.nist.gov).

Second B. Author Jr. was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

Third C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba 305-0047, Japan (e-mail: author@nrim.go.jp).

Mentions of supplemental materials and animal/human rights statements can be included here.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

including variations in crown structure and morphology within the same species, as well as the absence of distinct features that differentiate between different tree species. Perhaps due to these challenges, tree species classification using lidar data has remained a challenging task. Note that in mixed forests tree species can be classified only at the level of individual trees, but lidar data also describes geometrical properties of tree groups and this information can be used in the area-based lidar inventory [12]. However, in that case tree species must be predicted for stands instead of classifying individual trees.

Although geometrical information might have limited benefits for tree species classification, in some experiments, tree species have been classified exclusively using aircraft-borne lidar data. However, these single sensor approaches often require specific conditions, such as seasonal changes in leaf phenology [13], multispectral lidar data [14], or exceptionally high point densities [15], [16]. Despite these challenges, single-sensor approaches offer several advantages over data fusion from different sensors. One significant advantage is that they simplify the data collection, storage, and processing workflows, reducing the potential for errors and inconsistencies that can arise during data integration. It is also worth noting that the use of lidar data in forest inventories expands beyond the classification of tree species e.g., to the prediction of biomass or extraction of ground level. As a result, a single-sensor solution to tree species classification based on lidar would be highly advantageous for conducting tree species-specific forest inventories.

An increasing number of recent studies have experimented with machine learning (ML) methods to predict tree species of individually segmented trees using very high-point density lidar data acquired from terrestrial laser scanning (TLS) [17]–[20], mobile platforms [21] or UAS platforms [20], [22], [23]. For instance, [22] and [20] investigated the classification of birch and larch trees at Saihanba National Forest Park in Hebei province, China. Chen et al. [22] proposed a neural network architecture that operated directly on point clouds, which were acquired using terrestrial laser scanning (TLS) and unmanned aerial system (UAS) LiDAR data sources. The authors reported an increase in classification accuracy when the number of points at the individual tree-level increased from 512 to 2048 and found that TLS data yielded higher classification accuracy than UAS LiDAR under identical experimental conditions. Similarly, [20] proposed a deep neural network architecture that also operated directly on point clouds. Their approach achieved a higher classification accuracy with TLS data (kappa 0.85) than with UAS LiDAR data (kappa 0.76) in their experimental setup. In another related study, [24] investigated the simultaneous classification of tree species and standing dead trees using UAS LiDAR and helicopter-borne LiDAR data at two test sites. The authors created multiple side-view silhouette images of LiDAR intensity from different viewing directions as geometrical features for each tree. Using these geometrical features as the sole inputs to a neural network classifier, the authors reported overall accuracies (OA) ranging from 0.72 to 0.87, depending on the test site, LiDAR data characteristics, and

classifier used. However, when the geometrical features were used as complementary information to multispectral (MS) data, the OA values improved and ranged from 0.82 to 0.96, whereas the use of MS data alone resulted in OA values ranging from 0.86 to 0.94. The study underscores the challenges associated with classifying tree species using geometrical data alone in comparison to the advantages of incorporating auxiliary MS data. The objectives of this study were to explore the feasibility of using UAS lidar data to classify tree species in a boreal forest environment and to assess the effectiveness of features derived solely from point cloud data in achieving this classification. Specifically, the study aimed to:

1. Evaluate the accuracy of different machine learning models in classifying tree species using UAS lidar data.
2. Investigate the use of intensity and textural features derived from point cloud data in improving the accuracy of machine learning models for tree species classification.
3. Explore the importance of these descriptive features in achieving accurate tree species classification.

Overall, the study aimed to contribute to the development of more accurate and efficient methods for tree species classification in boreal forest environments, which can support forest management efforts.

## II. MATERIALS

### A. Field data

Field data were collected from 39 square 30 m 30 m forest plots located in eastern Finland (Fig. 1, Table I). The forests in the area are predominantly privately owned and are thus primarily used for timber production. Tree species distribution in the area is typical of Finnish inland areas with the forests dominated by Scots pine (*Pinus sylvestris*) and Norway spruce (*Picea abies*). Broadleaved tree species, mainly Silver birch (*Betula pendula*) and Downy birch (*Betula pubescens*), are found abundantly in more fertile habitats but are typically in the minority.

TABLE I  
AVERAGE GROWING STOCK VOLUME, STEM COUNT, HEIGHT OF THE BASAL AREA MEDIAN TREE (HGM), AND DIAMETER OF THE BASAL AREA MEDIAN TREE (DGM) IN THE FIELD PLOTS.

	Pine	Spruce	Broadleaved	Total
Volume (m <sup>3</sup> /ha)	92.9	99.1	48.5	240.5
Stem count (n/ha)	282.9	467.5	244.4	994.8
HGM (m)	20.9	16.8	19.3	21.0
DGM (cm)	24.8	20.4	20.5	23.7

Each tree with a diameter at breast height (DBH) >5 cm was

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

recorded. For these trees, tree species was determined, DBH was calipered, and height was measured using an electronic Vertex instrument. In addition, the location of each measured tree was determined with sub-meter accuracy using an adaptation of the triangulation method described in [25].

### B. Lidar data

The UAS lidar data were scanned using a Riegl VUX-1 UAS with an AP20 inertial measurement unit. The data were acquired in July 2020 using an Avartek Boxer hybrid drone. The scanner had a scanning angle of 120 degrees and pulse repetition frequency of 380 kHz. Each plot was scanned from 10 flight lines: five in an east-west direction and five in a south-north direction, at a cruising speed of about 4 m/s from an altitude of 50 - 60 m above ground level. This setup resulted in a nominal density of approximately 3700 pulses per m<sup>2</sup>. After the 10 flight lines were merged, the lidar echoes were classified as ground and non-ground with the method proposed by [26]. A triangulated irregular network (TIN) was created from ground lidar echoes. By subtracting the interpolated TIN value from the height of the echoes, all lidar echoes were normalized to ground level.

### A. Tree crown segmentation

The delineation of individual tree crowns from lidar data was performed using marker-controlled watershed segmentation (reference). Initially, a 0.5m resolution canopy height model (CHM) was generated from the lidar echoes by assigning the highest echo within each 0.5m resolution raster cell. The resulting CHM was Gaussian filtered with a sigma value of 0.8 (in pixel coordinate system). Subsequently, the markers for the watershed segmentation were identified from the Gaussian-filtered CHM using a modified variable window filter method (as detailed in [27]).

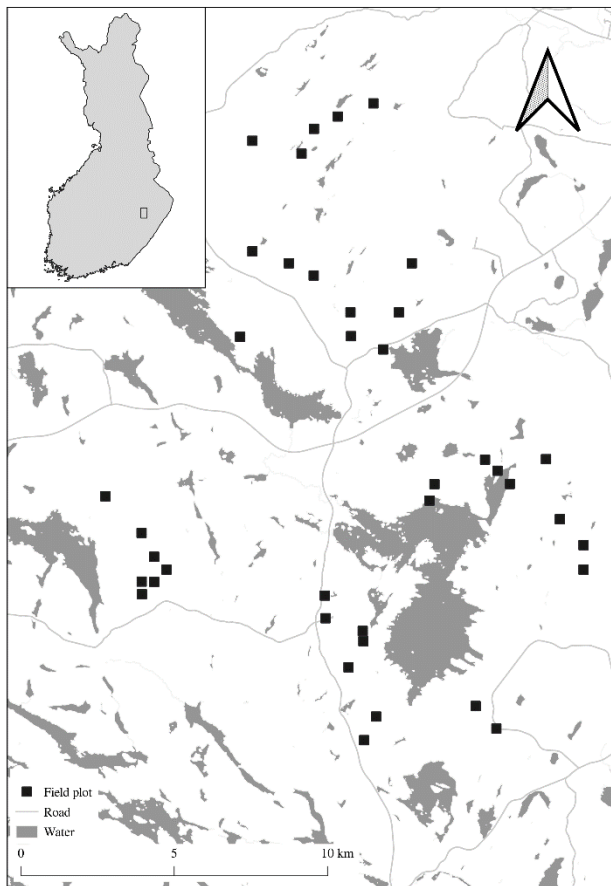
The crown segments were linked to field-measured trees using an automated linking procedure. Closest neighbors were computed from crown segments to field-measured trees and vice versa. A connection between a crown segment and a field-measured tree was created only if they were each other's closest neighbor and the distance between them was <2 m. The distances were computed in 3D using the local maxima from the CHM (i.e., detected treetop) and field-measured treetops.

### B. Predictor variables

Two categories of predictor variables were computed from the lidar echoes for each segmented tree: 1) intensity, and 2) texture (Fig. 2). The intensity features were computed at the voxel-level, whereas the texture variables were computed for multiple horizontal cross-sections of the lidar point cloud. Both categories of predictor variables were computed using first of many and only lidar echoes with height values >50% of the height of the tree. Height threshold was used to mitigate the impact of understory trees in the classification process. For the remainder of this subsection, when referring to lidar echoes, we refer to first of many and only lidar echoes with height values >50% of the detected treetop.

1) *Intensity features:* Each segmented tree was divided into 8 x 8 x 10 (x y z) voxels so that the detected treetop was located in the center of the highest layer of voxels. Fixing the number of voxels ensures that the size of the voxels can vary between different sized trees. The mean intensity of all lidar echoes within each voxel was computed. It should be noted that these intensity features indirectly contain information about the shape of the segmented tree. This is because an intensity value of zero for any given voxel means that the voxel contains no lidar echoes.

Looking downwards from the treetop, the crown of the tree can be considered to be rotationally symmetrical. Given this assumption, the original 8 x 8 x 10 voxel grid can be reduced to a 4 x 4 x 10 voxel grid (x y z). This computation can be explained in 2D by initially folding the 8 x 8 grid in half (4 x 8) and computing the mean of the voxels that are laid on top of each other. Thereafter, the 4 x 8 grid is folded in half in the other direction and the mean of the voxels that are on top of each other compose the final 4 x 4 grid, where the treetop is located at the lower right cell. The number of intensity features was 160



**Fig. 1.** Map of the study area. Black squares indicate the locations of the forest plots in the study area.

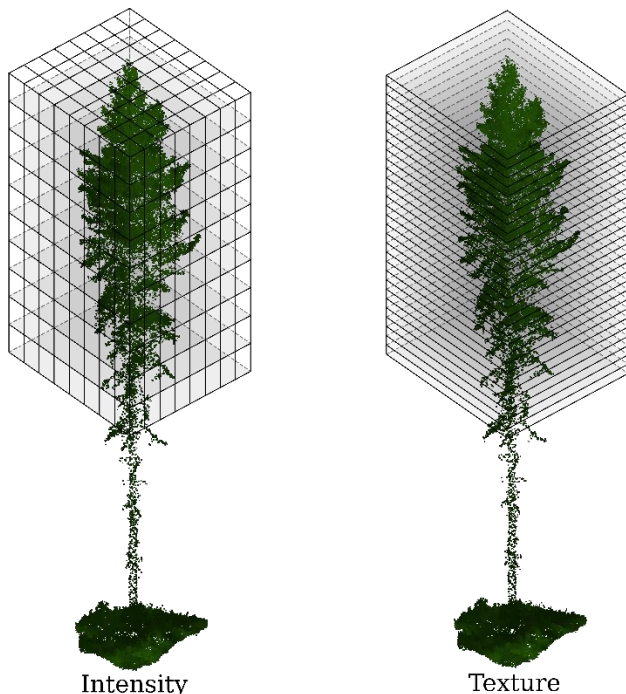
## III. METHODS

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

when this was repeated for all 10 layers.

2) *Texture*: Haralick texture features [28] were computed from 40 horizontal cross-sections from the lidar echoes for each segmented tree. Haralick features which cannot always be computed, i.e., the two *Information Measures of Correlation* and *Maximal Correlation Coefficient*, were excluded. Each horizontal cross-section was divided into an array of 40 x 40 pixels. Therefore, similar to voxels of intensity features, pixel size, and the thickness of each horizontal cross-section varied in different sized trees. A pixel had a value of 1 if it contained lidar echoes, and a value of 0 if it contained no echoes (i.e., a binary image). The gray-level co-occurrence matrix was computed as a mean of four directions (0°, 45°, 90° and 135°) with a distance of 1 between neighboring pixels. Then, the gray-level co-occurrence matrix was normalized to the sum of 1.

Texture features with a strong correlation ( $<-0.99$  or  $>0.99$ ) and features that could have an invalid value were excluded. Tree species classification was performed by linear discriminant analysis (LDA) on one texture feature at a time. Features with very little explanatory potential (i.e., very poor classification accuracy) were excluded. Texture features with a strong correlation ( $<-0.99$  or  $>0.99$ ) were excluded, prioritizing the removal of the feature that displayed lower classification accuracy as indicated by the LDA classifier. The remaining texture features were angular second moment (ASM), contrast (CON), correlation (COR), and sum average (SAVE). Thus, the final number of textural features was 160 (four textural features in 40 cross-sections).



**Fig. 1.** Illustration of tree level intensity and texture features. Voxels are distributed in a regular 8 x 8 x 10 grid, and each voxel contains the average intensity of lidar echoes within it. Texture features are computed from 40 horizontal cross sections which are each a 40 x 40 binary image.

### C. Classification algorithms

Classification accuracy and feature importance were evaluated using three ML algorithms: random forest (RF), multinomial logistic regression (MLR) with LASSO, and multilayer perceptron (MLP). These algorithms were chosen as they are well-known, widely used, and partly because they all enable the computation of feature importance measures in a different fashion. For each classification algorithm, the two categories of predictor variables, i.e., intensities and texture, were evaluated separately and then in combination.

1) *Random Forest*: Random forests are an ensemble learning method used in a range of tasks [29]. The method is based on utilizing a multitude of decision trees that all vote on the outcome from a given set of inputs. In the context of classification, each individual decision tree produces a class prediction and the class with the greatest number of votes among all the decision trees becomes the models prediction. In this study, we used the random forest R implementation described by [30]. A permutation-based method was used to determine the importance of individual features to the output of the random forest model. The features were permuted (i.e., the input feature array was shuffled) one input feature at a time. The resulting prediction error was compared to the prediction error when that specific feature was not shuffled. The difference in the prediction error denotes the importance of that feature.

2) *Multinomial logistic regression with LASSO*: Least absolute shrinkage and selection operator (LASSO) is a regression and classification method that performs both variable selection and regularization [31]. It uses the L1 penalty for fitting and penalization of the model coefficients. We fitted a multinomial logistic regression model with LASSO, as implemented in the R-package glmnet [32]. The input features were standardized prior to fitting the model. Thus, the importance of each feature could be defined simply by the magnitude of the coefficient of the feature. The greater the coefficient, the more important the feature.

3) *Multilayer perceptron*: Multilayer perceptron (MLP) refers to a feedforward neural network, in which the layers are fully connected. The neural network architecture consists of input and output layers, and hidden layers between them. In addition, each of the hidden layers consists of neurons whose operation is controlled by activation functions. For more detailed discussion of MLP, see [33].

The hyperparameter tuning of the MLP model was done using a *BayesianOptimization* tuner in KerasTuner [34]. In our setup, the number of layers ranged from 1 to 3, and the number of neurons per layer from 3 to 300. In addition, each layer had an L2 regularization (weight decay) of 0.01, 0.001 or 0.0001, and the learning rate for the chosen Adam optimizer [35] of 0.01, 0.001, 0.0001, or 0.00001. The rectified linear unit was used as an activation function in all but the last layer where softmax activation was used. One must note that the parameters were optimized separately at each fold of the k-fold validation. Feature importance of the MLP model were calculated using SHAP (SHapley Additive exPlanations) values [36].



> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

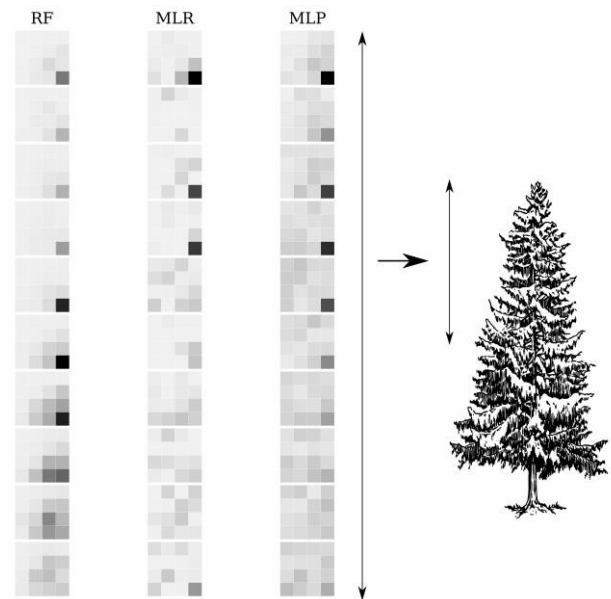
#### D. Validation

The results were validated using a leave-one-plot-out procedure. This validation scheme was chosen because the sample size was relatively small and to avoid the effect of spatial autocorrelation among trees. Training, prediction, and model evaluation were repeated 39 times, leaving the trees of a single plot out from the training data each time for validation. Thus, trees of a spatially excluded area (i.e., 30 m 30 m field plot) were always used for validation, thereby avoiding the case that training and validation trees exist in the same plot. For the MLP model, the trees from the 38 training plots at each k-fold were divided into training and test datasets (70% for training and 30% for test), which were used to regulate the early stopping criterion of the MLP model. Classification accuracies were evaluated using confusion matrices and Cohen's kappa [37], a statistical measure that accounts for chance agreement between raters. Additionally, class-specific accuracies were reported as F1-scores [38], which combine both precision and recall into a single metric. Feature importance measures were scaled between 0 and 1 for each classification method separately. This means that feature importance are comparable between feature categories for each classification method, although feature importance cannot be directly compared between classification methods.

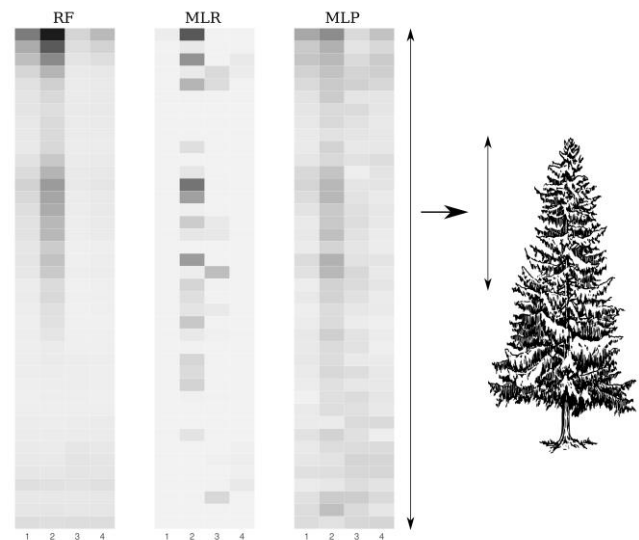
### IV. RESULTS

#### A. Feature importance

Scaled feature importance of the different classification methods is presented in Fig. 3 for intensities and in Fig. 4 for textural features. Similar features were ranked as important across the different classification methods except in the case of intensities, where RF appears to value voxels at a lower portion of crown than the other classifiers. This is an interesting observation because feature importance was computed with fundamentally different techniques. We see that MLR often sets unimportant features to 0, a result of how the coefficients are estimated in LASSO.



**Fig. 3.** Intensity feature importance with random forest (RF), multinomial logistic regression with LASSO (MLR), and multilayer perceptron (MLP). The upper 4 4 grid is located at the top of the tree, and the lowest 4 4 grid is located at half the height of the tree. The center of the treetop (and by extension the tree trunk) is located at the lower right corner of each 4 4 grid. The color of the cell indicates the importance of that feature, scaling from low (light) to high (dark).



**Fig. 4.** Texture feature importance with random forest (RF), multinomial logistic regression (MLR) with LASSO and multilayer perceptron (MLP). The upper row contains textural features from the top of the tree and the lowest row contains textural features from half the height of the tree. The color of the cell indicates the importance of that feature scaling from low (light) to high (dark). The columns correspond to (1) ASM, (2) CON, (3) COR, and (4) SAVG Haralick textural features.

All classification methods seemed to value the CON textural

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

feature at the top of the canopy and 1/3 of the vertical distance from the treetop to the middle of the tree. All classification methods also agreed that the intensity information attached to voxels is most important closest to the tree trunk and that the voxels furthest away from the trunk are less important. In RF the most important intensity features were computed from different heights of the tree than in the other classification methods. The MLR and MLP ranked the upper-most intensity features to be most important but in RF the most important intensity features were computed below the upper-most part of the tree (Figure. 2). Indirect information about tree shape was most apparent in Figure 2, where the intensity feature importance resembled the cone-like shape of a tree crown that expands towards the base of the tree.

### B. Classification accuracy

Kappa values of all classification methods, using the different input features, are presented in Table II. Corresponding confusion matrices can be found in Tables III-V. All classification methods performed similarly with the same input features. However, the input features noticeably influenced the classification outcome. Kappa values were least (0.60 - 0.64) when only textural features were used, whereas the use of intensity features resulted in clearly greater kappa values (0.73 - 0.77). All classification methods yielded the greatest kappa value (0.81) when both intensity and textural features were used.

TABLE II

KAPPA VALUES OF THE DIFFERENT CLASSIFICATION METHODS AND INPUT FEATURES. RF DENOTES RANDOM FOREST, MLR DENOTES MULTINOMIAL LOGISTIC REGRESSION WITH LASSO, AND MLP DENOTES MULTILAYER PERCEPTRON.

	RF	MLR	MLP
Texture	0.64	0.62	0.60
Intensity	0.75	0.77	0.73
Texture and intensity	0.81	0.81	0.81

TABLE III

TREE SPECIES CLASSIFICATION USING RANDOM FOREST (RF) WITH DIFFERENT INPUT FEATURES. CLASS NAMES ARE ABBREVIATED: P DENOTES PINE, S DENOTES SPRUCE, AND B DENOTES BROADLEAVED.

		Observed											
		Texture			Intensity			Texture and intensity					
		P	S	B				P	S	B			
Predicted	P	527	87	67	P	574	50	64	P	606	18	46	
	S	94	453	56	S	4	513	31	S	33	555	37	
	B	57	60	233	B	63	37	261	B	39	27	273	
	F <sub>1</sub>	0.78	0.75	0.66	F <sub>1</sub>	0.84	0.87	0.73	F <sub>1</sub>	0.90	0.91	0.79	

TABLE IV

TREE SPECIES CLASSIFICATION USING MULTINOMIAL LOGISTIC REGRESSION (MLR) WITH DIFFERENT INPUT FEATURES. CLASS NAMES ARE ABBREVIATED: P DENOTES PINE, S DENOTES SPRUCE, AND B DENOTES BROADLEAVED.

		Observed											
		Texture			Intensity			Texture and intensity					
		P	S	B				P	S	B			
Predicted	P	539	83	82	P	600	32	73	P	620	24	60	
	S	107	487	61	S	32	546	38	S	25	558	36	
	B	32	30	213	B	46	22	245	B	33	18	260	
	F <sub>1</sub>	0.78	0.78	0.68	F <sub>1</sub>	0.87	0.90	0.73	F <sub>1</sub>	0.90	0.92	0.78	

TABLE V

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

TREE SPECIES CLASSIFICATION USING MULTILAYER PERCEPTRON (MLP) WITH DIFFERENT INPUT FEATURES. CLASS NAMES ARE ABBREVIATED: P DENOTES PINE, S DENOTES SPRUCE, AND B DENOTES BROADLEAVED.  
Observed

		Texture			Intensity			Texture and intensity						
		P	S	B			P	S	B			P	S	B
Predicted	P	527	87	67	P	574	50	64	P	606	18	46		
	S	94	453	56	S	4	513	31	S	33	555	37		
	B	57	60	233	B	63	37	261	B	39	27	273		
	F <sub>1</sub>	0.78	0.75	0.66	F <sub>1</sub>	0.84	0.87	0.73	F <sub>1</sub>	0.90	0.91	0.79		

## V. DISCUSSION AND CONCLUSIONS

We introduced two new approaches to compute tree-level features from UAS lidar data. One feature set uses lidar intensities and the second quantifies texture at different heights. These feature sets were used as inputs for three different ML approaches both individually and in combination. All three ML approaches delivered very similar results given identical input features. Thus, the focus of the discussion is related to the features and feature importance, rather than discussing the minute differences between the performance of the prediction methods.

The utilization of textural features alone resulted in the lowest classification accuracy (kappa 0.60-0.64) compared to using only intensity features, which yielded a notably higher accuracy (kappa 0.73-0.77). Intensity features contain information pertaining to crown characteristics that are also captured by texture. Intensity values of 0 indicate the absence of lidar echoes, while the mean intensity of a voxel can indicate foliage density, trunk, leaf orientation, and clumping. In this study, only the first echoes returns of each pulse were used, which ensures the validity of such interpretation of lidar intensity as for all other echoes of the pulse the energy loss from preceding echoes would be unknown. The differences in geometry and foliage among the tree species investigated may account for why intensity features were more effective than textural features for classification.

The study found that combining intensity and textural features resulted in the highest classification accuracies, regardless of the prediction method used. The addition of texture information improved the accuracy of identifying broadleaves the most, increasing the F1 score from 0.73 to 0.79 and the kappa value from 0.75 to 0.81. These improved accuracies are similar to those in other studies, which typically report overall accuracies around 90% depending on factors such as the number of tree species, biome, and data acquisition parameters. For example, [22] found that point density at the individual tree-level had a noticeable impact on tree species classification, with a kappa value of 0.77 for lower-point density data and 0.85 for higher-point density UAS LiDAR data when classifying pine and larch. Similarly, [23] achieved an overall accuracy of 86.6% when classifying four tree species with relatively low-point density UAS LiDAR data (40 pts/m<sup>2</sup>) using convex hull-based feature descriptors with the PointNet++ network.

In most cases different ML approaches ranked similar features important, although their methods of determining importance varied. Notably, the MLP approach assigned fewer features with zero importance, consistent with findings reported by [31][39]. Across all three approaches, intensity features closest to the tree trunk were considered most important. However, important features were more spread out in lower voxels due to the cone-like shape of trees, with increasing branches and leaves further away from the top. The central voxels of the trees were also important, even at lower cross-sections, suggesting the importance of not only the crown's shape but also the way branches and leaves are organized within it. Furthermore, the tree crown's central voxels were important in all prediction methods, indicating that differences in this area can distinguish the three tree species in the study. However, it should be noted that unlike the other classifiers, RF valued less intensity features computed from the upper-most part of the tree. The reason for this behavior of RF feature importance remains unclear. Spruce trees typically have conical tops, while pine and birch may have rounder tops and possibly multiple distinguishable treetops.

Texture features were also ranked similarly with all approaches, with the CON feature appearing to be the most important at the top of the canopy and at about 75% of the height of the tree. The other three texture features were considered important only at the very top of the canopy. CON is a measure of local variation in the image and can be linked to the number of distinct features in a horizontal cross-section. The main contribution of textural features as complementary to intensity features was the increase in classification accuracy for broadleaved trees, which could indicate that the improvement is in some way related to distinct local variations in the point cloud due to leaves or multiple apparent treetops.

With regard to the operational applications of the proposed classification scheme, an obvious disadvantage of the intensity features, and to a certain extent the textural features as well, is that they are not sensor-agnostic. Lidar intensity is strongly dependent on the hardware and the associated software that decides where the discrete echoes are located in the waveform data. It is also affected by the flying parameters and scanning geometry that influences the size of the footprint of the laser pulse. This can be an issue as the collection of large amounts of tree-level georeferenced training data from a relatively small inventory area can be prohibitively expensive. The same

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

problem with sensor-specific features and exorbitant field measurement expenses applies to most tree-level inventory methods. Ideally, the training data from one inventory area could be utilized in another inventory area and different hardware, given that the tree species have similar characteristics between the different geographical locations.

This study demonstrated that the combination of two fundamentally different types of features from UAS lidar datasets can deliver very good classification accuracies for common boreal tree species. These features should also be tested with lower density airborne lidar data. However, it is reasonable to assume that the successful application of the features presented in this study requires point densities that a conventional airborne lidar is unable to achieve at a reasonably high altitude. Recently commercialized single photon lidar sensors could be used as an alternative to conventional airborne lidar sensors to enhance the point densities.

#### ACKNOWLEDGMENT

The study was supported by the Academy of Finland through the Finnish Flagship Programme for the Forest-Human-Machine Interplay Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) [337655], the Centre of Excellence of Inverse Modelling and Imaging, and projects [321761], and the project Unmanned Aerial Vehicles in Forest Remote Sensing [323484].

#### REFERENCES

- [1] J. Breidenbach, A. Granhus, G. Hylen, R. Eriksen, and R. Astrup, "A century of National Forest Inventory in Norway – informing past, present, and future decisions," *For. Ecosyst.*, vol. 7, no. 1, p. 46, Dec. 2020, doi: 10.1186/s40663-020-00261-0.
- [2] E. Tomppo, M. Haakana, M. Katila, and J. Peräsaari, *Multi-Source National Forest Inventory*, vol. 18. Springer, 2008, doi: 10.1007/978-1-4020-8713-4.
- [3] S. Puliti, H. Ørka, T. Gobakken, and E. Næsset, "Inventory of Small Forest Areas Using an Unmanned Aerial System," *Remote Sensing*, vol. 7, no. 8, pp. 9632–9654, Jul. 2015, doi: 10.3390/rs70809632.
- [4] E. Kotivuori, M. Kukkonen, L. Mehtätalo, M. Maltamo, L. Korhonen, and P. Packalen, "Forest inventories for small areas using drone imagery without in-situ field measurements," *Remote Sensing of Environment*, vol. 237, p. 111404, Feb. 2020, doi: 10.1016/j.rse.2019.111404.
- [5] M. Kirchhoefer, J. Schumacher, and P. Adler, "Potential of remote sensing-based forest attribute models for harmonising large-scale forest inventories on regional level: a case study in Southwest Germany," *Annals of Forest Science*, vol. 76, no. 2, p. 33, Jun. 2019, doi: 10.1007/s13595-019-0804-4.
- [6] E. B. Knipling, "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation," *Remote Sensing of Environment*, vol. 1, no. 3, pp. 155–159, 1970, doi: https://doi.org/10.1016/S0034-4257(70)80021-9.
- [7] L. Windrim, A. J. Carnegie, M. Webster, and M. Bryson, "Tree Detection and Health Monitoring in Multispectral Aerial Imagery and Photogrammetric Pointclouds Using Machine Learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2554–2572, 2020, doi: 10.1109/JSTARS.2020.2995391.
- [8] P. Kempeneers, K. Vandekerckhove, F. Devriendt, and F. Van Coillie, "Propagation of shadow effects on typical remote sensing applications in forestry," in *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2013, pp. 1–5, doi: 10.1109/WHISPERS.2013.8080613.
- [9] J. Vauhkonen, T. Tokola, P. Packalén, and M. Maltamo, "Identification of Scandinavian Commercial Species of Individual Trees from Airborne Laser Scanning Data Using Alpha Shape Metrics," *Forest Science*, vol. 55, no. 1, pp. 37–47, Feb. 2009, doi: 10.1093/forestscience/55.1.37.
- [10] J. Holmgren, Å. Persson, and U. Söderman, "Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images," *International Journal of Remote Sensing*, vol. 29, no. 5, pp. 1537–1552, 2008, doi: 10.1080/01431160701736471.
- [11] I. Korpela, H. O. Orka, M. Maltamo, T. Tokola, and J. Hyypä, "Tree Species Classification Using Airborne LiDAR - Effects of Stand and Tree Parameters, Downsizing of Training Set, Intensity Normalization, and Sensor Type," *Silva Fennica*, vol. 44, pp. 319–339, 2010.
- [12] J. Vauhkonen, A. Seppänen, P. Packalén, and T. Tokola, "Improving species-specific plot volume estimates based on airborne laser scanning and image data using alpha shape metrics and balanced field data," *Remote Sensing of Environment*, vol. 124, pp. 534–541, Sep. 2012, doi: 10.1016/j.rse.2012.06.002.
- [13] M. Villikka, P. Packalen, and M. Maltamo, "The suitability of leaf-off airborne laser scanning data in an area-based forest inventory of coniferous and deciduous trees," *Silva Fennica*, vol. 46, 2012.
- [14] A. Axelsson, E. Lindberg, and H. Olsson, "Exploring Multispectral ALS Data for Tree Species Classification," *Remote Sensing*, vol. 10, no. 2, 2018, doi: 10.3390/rs10020183.
- [15] H. Hamraz, N. B. Jacobs, M. A. Contreras, and C. H. Clark, "Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees," 2018.
- [16] S. Briechele, P. Krzystek, and G. Vosselman, "Semantic Labeling of ALS Point Clouds for Tree Species Mapping Using the Deep Neural Network POINTNET++," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4213, pp. 951–955, Jun. 2019, doi: 10.5194/isprs-archives-XLII-2-W13-951-2019.
- [17] T. Mizoguchi, A. Ishii, H. Nakamura, T. Inoue, and H. Takamatsu, "Lidar-based individual tree species classification using convolutional neural network," in *Videometrics, Range Imaging, and Applications XIV*, 2017, vol. 10332, pp. 193–199, doi: 10.1117/12.2270123.
- [18] X. Zou, M. Cheng, C. Wang, Y. Xia, and J. Li, "Tree Classification in Complex Forest Point Clouds Based on Deep Learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2360–2364, 2017, doi: 10.1109/LGRS.2017.2764938.
- [19] Z. Xi, C. Hopkinson, S. B. Rood, and D. R. Peddle, "See the forest and the trees: Effective machine and deep learning algorithms for wood filtering and tree species classification from terrestrial laser scanning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 168, pp. 1–16, 2020, doi: https://doi.org/10.1016/j.isprsjprs.2020.08.001.
- [20] M. Liu, Z. Han, Y. Chen, Z. Liu, and Y. Han, "Tree species classification of LiDAR data based on 3D deep learning," *Measurement*, vol. 177, p. 109301, 2021, doi: https://doi.org/10.1016/j.measurement.2021.109301.
- [21] H. Guan, Y. Yu, Z. Ji, J. Li, and Q. Zhang, "Deep learning-based tree classification using mobile LiDAR data," *Remote Sensing Letters*, vol. 6, no. 11, pp. 864–873, 2015, doi: 10.1080/2150704X.2015.1088668.
- [22] J. Chen, Y. Chen, and Z. Liu, "Classification of Typical Tree Species in Laser Point Cloud Based on Deep Learning," *Remote Sensing*, vol. 13, no. 23, 2021, doi: 10.3390/rs13234750.
- [23] Y. Lv et al., "A Convex Hull-Based Feature Descriptor for Learning Tree Species Classification From ALS Point Clouds," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3055773.
- [24] S. Briechele, P. Krzystek, and G. Vosselman, "Silvi-Net – A dual-CNN approach for combined classification of tree species and standing dead trees from remote sensing data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 98, p. 102292, 2021, doi: https://doi.org/10.1016/j.jag.2020.102292.
- [25] I. Korpela, T. Tuomola, and E. Välimäki, "Mapping forest plots: an efficient method combining photogrammetry and field triangulation," *Silva Fennica*, vol. 41, no. 3, pp. 457–469, 2007.
- [26] P. Axelsson, "DEM Generation from Laser Scanner Data Using Adaptive TIN Models," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 33, pp. 110–117, 2000.
- [27] S. C. Popescu and R. H. Wynne, "Seeing the Trees in the Forest: Using Lidar and Multispectral Data Fusion with Local Filtering and Variable



> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

Window Size for Estimating Tree Height,” *Photogrammetric Engineering and Remote Sensing*, vol. 70, pp. 589–604, 2004.

[28] R. Haralick, K. Shanmugam, and I. Dinstein, “Texture Features for Image Classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, 1973.

[29] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[30] A. Liaw and M. C. Wiener, “Classification and Regression by randomForest,” 2007.

[31] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” *Journal of the royal statistical society series b-methodological*, vol. 58, pp. 267–288, 1996.

[32] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization Paths for Generalized Linear Models via Coordinate Descent,” *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[33] “Dive into Deep Learning,” *CoRR*, vol. abs/2106.11342, 2021, [Online]. Available: <https://arxiv.org/abs/2106.11342>

[34] T. O’Malley *et al.*, “Keras Tuner.” 2019. [Online]. Available: <https://github.com/keras-team/keras-tuner>

[35] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization.” 2017.

[36] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *Advances in Neural Information Processing Systems* 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>

[37] J. Cohen, “A Coefficient of Agreement for Nominal Scales,” *Educational and Psychological Measurement*, vol. 20, no. 1, p. 37, 1960.

[38] C. Goutte and E. Gaussier, “A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation,” in *Advances in Information Retrieval*, Berlin, Heidelberg, 2005, pp. 345–359.

[39] G. Hooker, L. Mentch, and S. Zhou, “Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance,” *Stat. Comput.*, vol. 31, p. 82, 2021.