

Tree Species Detection and Enhancing Semantic Segmentation Using Machine Learning Models with Integrated Multispectral Channels from PlanetScope and Digital Aerial Photogrammetry in Young Boreal Forest

[Arun Gyawali](#)*, [Mika Aalto](#), [Tapio Ranta](#)

Posted Date: 11 March 2025

doi: 10.20944/preprints202503.0722.v1

Keywords: Tree species detection; Semantic Segmentation; Digital Aerial Photogrammetry; PlanetScope; Vegetation Indices; YOLO; random forest; CatBoost; CNN



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Tree Species Detection and Enhancing Semantic Segmentation Using Machine Learning Models with Integrated Multispectral Channels from PlanetScope and Digital Aerial Photogrammetry in Young Boreal Forest

Arun Gyawali *, Mika Aalto and Tapio Ranta

Laboratory of Bioenergy, Lappeenranta-Lahti University of Technology, LUT, Lönnrotinkatu 7, 50100 Mikkeli, Finland

* Correspondence: arun.gyawali@lut.fi

Abstract: The precise identification and classification of tree species in young forests during their early development stages are vital for forest management and silvicultural efforts that support their growth and renewal. Yet, achieving accurate geolocation and species classification is often a labor-intensive and complex task through field-based surveys. Remote sensing technologies combined with machine learning techniques present an encouraging solution, offering a more efficient alternative to conventional field-based methods. This study aimed to detect and classify young forest tree species by employing remote sensing imagery and machine learning techniques. The study involved mainly two different objectives: first, the tree species detection using the latest version of You only Look Once (YOLOv12) and second, semantic segmentation (classification) using random forest, Categorical Boosting (CatBoost) and Convolutional Neural Network (CNN). To the best of our knowledge, this marks the first exploration utilizing YOLOv12 for tree species identification, along with the groundbreaking study that integrates digital aerial photogrammetry with Planet imagery to achieve semantic segmentation in young forest. The study utilized two remote sensing datasets: RGB imagery from UAV ortho photography and RGB-NIR from PlanetScope. For YOLOv12-based tree species detection, only RGB from ortho photography was used, while semantic segmentation was performed with three sets of data: (1) Ortho RGB, (2) Ortho RGB + canopy height model (CHM) + Planet RGB-NIR (8 Bands), and (3) ortho RGB + CHM + Planet RGB-NIR + 12 vegetation indices (20 Bands). With three models applied to these datasets, a total of nine machine learning models were trained and tested using total of 57 images (1024×1024 pixels) and their corresponding mask tiles. The YOLOv12 model achieved 79% overall accuracy, with Scots pine performing best (precision: 97%, recall: 92%, mAP50: 97%, mAP75: 80%) and Norway spruce showing slightly lower accuracy (precision: 94%, recall: 82%, mAP50: 90%, mAP75: 71%). For semantic segmentation, the CatBoost model with 20 Bands outperformed other models, achieving 85% accuracy, 80% Kappa, and 81% MCC, with CHM, EVI, NIRPlanet, GreenPlanet, NDGI, GNDVI, and NDVI being the most influential variables. These results indicate that a simple boosting model like CatBoost can outperform more complex CNNs for semantic segmentation in young forests.

Keywords: tree species detection; semantic segmentation; digital aerial photogrammetry; PlanetScope; vegetation indices; YOLO; random forest; CatBoost; CNN

1. Introduction

Tree species identification and classification using remote sensing and machine learning (ML) has become vital in forestry. The detection and classification is essential for ecosystem evaluation,

biomass and biodiversity monitoring, and effective utilization of forest resources [1,2]. This provides essential perspectives into environmental health, supports conservation efforts, and promotes sustainable forestry management practices. Tree species (object) detection is a powerful tool that enables the geolocation and identification of individual trees while utilizing data from diverse tree species, making these models highly adaptable for use in various regions [3]. There are many object detection methods depending on architecture, computational speed, efficiency, and accuracy; for example, You Only Look Once (YOLO) [4], RetinaNet [5], region-based convolutional neural networks (R-CNN) [6], faster R-CNN [7], Mask R-CNN [8], etc. Previous studies have demonstrated the success of various machine learning methods in detecting forest tree species using remote sensing data, such as Light Detection and Ranging (LiDAR) and Digital Aerial Photogrammetry (DAP) [9–16]. The YOLO framework is a popular type of real-time object detection model that's been carefully built to focus on accuracy, speed, and efficiency. Introduced by Joseph Redmon in 2016 [4], the YOLO framework has notably altered the field of computer vision by reconceptualizing object detection as a unified regression task, allowing the model to predict bounding boxes and class probabilities concurrently from a single input image in one evaluation. This technique illustrates considerable distinction from established object detection methods, such as R-CNN or Faster R-CNN, which require a series of processing steps. Previous studies also found that the YOLO model has higher detection accuracy than Mask-R CNN and U-Net [13], or Faster R-CNN and Retina Net [15] in tree species segmentation and detection. In recent years, various versions of the YOLO model have been successfully applied to forest species detection, consistently demonstrating high levels of accuracy [13,15,17–22]. With these substantial improvements, the majority of earlier YOLO applications has primarily targeted the detection of tree crowns, rather than examining the species-level classification. Relatively few studies [13,15,17,19] have explored the more specific task of tree species identification, including one on the classification of species such as *Pinus koraiensis* and *Abies holophylla* [13] and another identifying *Populus bolleana*, *Ulmus pumila*, *Elaeagnus angustifolia*, etc. [19]. Additionally, only a few studies have been dedicated to the detection of young tree species, stressing the need for further research in this area. These highlight the capabilities of deep learning model like YOLO for species-level identification, while also emphasizing the need for further research in distinguishing individual tree species within complex forest environments, including young forest, mixed stand age, and different biogeographic environments. Addressing this gap is also essential for improving the applicability of YOLO models to these diverse forest conditions for tree species detection. The latest version, YOLOv12 [23] (released in February 2025), has not yet been employed for the purposes of detecting and classifying forest tree species. Clearly, considering the recent launch of this specific version, we assert that, at the very least, no empirical investigations have been conducted employing it within field of forestry. Therefore, one of the objectives of this study is to utilize the YOLOv12 algorithm for the identification of tree species within young boreal forest, significantly enhancing precision and operational efficacy compared to its previous versions [23].

Semantic segmentation is a classification and computer vision technique, entailing the organized classification of each individual pixel within an image into established classes or labels. Unlike conventional methodologies for image classification, which assign a singular label to an entire image, or object detection methods that outline and confine objects using bounding boxes, semantic segmentation provides a detailed classification of the scene at the pixel level [24]. Advancements in ML – specifically, deep learning (DL) – have created new possibilities for automating the processing of remote sensing data, including sources such as PlanetScope (hereafter referred to as Planet) imagery, LiDAR, and DAP [9–14,25,26], and have expanded the possibilities for effectively classifying the forest species at pixel level [27–32]. Previous studies have conducted forest species classification at the pixel level (semantic segmentation) using ML models like random forest (RF), support vector machines (SVM), and convolutional neural networks (CNNs) [13,25,33–36]. The most frequent machine learning models for tree species classification are RF, SVM and CNNs. Only a limited number of studies have explored tree species classification using boosting methods such as LightGBM, XGBoost, and CatBoost, despite their superior or competitive performance compared to

decision trees, support vectors, k-nearest neighbors (KNN), and other ensemble methods [37–39]. Although boosting algorithms have proven their superior accuracy and efficiency, we have yet to fully unlock their potential in the classification of tree species. Furthermore, previous research on semantic segmentation largely centered around specific remote sensing data types, such as LiDAR, DAP, Sentinel, etc. Few studies have combined different data sources such as, high-resolution RGB imagery with satellite multispectral data [38,40–43]. Nonetheless, the combination of different data sources is proven to significantly enhance classification accuracy, elevating results from 73% with RGB alone to 92% when utilizing RGB + Hyperspectral + LiDAR [40]. An extensive review of the literature on tree species classification reveals that most studies have focused on three primary areas: (1) individual tree detection (ITD) using LiDAR and/or DAP data in combination with ML and DL models, (2) forest species classification using RGB DAP, where each image tile corresponds to a specific class, and (3) classification of tree species using high resolution unmanned aerial vehicle (UAV) images at pixel level. Apart from LiDAR and DAP, commonly used satellite remote sensing data for forest species classification include multi/hyperspectral imageries, the Landsat and Sentinel. It is evident that the current literature lacks adequate coverage on the utilization of other remote sensing data, like Planet, for classifying forest species [44]. Although Planet lab operates as a commercial service and its multispectral data is not freely accessible to the public, it offers the Planet Education and Research Program (ERP), which grants limited free access to researchers, students, and institutions for non-commercial scientific research. The multispectral data provided by Planet, covering red, green, blue, and near-infrared bands, offer an optimal balance of high temporal (daily) and spatial (3 m) and temporal resolution, making it highly suitable for tree species classification studies. A previous study has employed Planet imagery and high-resolution aerial photographs for forest species classification [45], and the fusion of these different data sources could further enhance accuracy by combining their spectral and structural attributes. This indicates that coarse-resolution satellite data can be integrated with high-resolution DAP data by resampling the coarse data and applying a classification model [42]. Moreover, many studies have mainly focused on mature trees, where fully developed branches and leaves create distinct spectral characteristics for different tree species. This clear differentiation makes it easier for models to learn during training, ultimately leading to higher classification accuracy. On the contrary, at the early growth stages, tree species exhibit similar spectral characteristics due to their comparable crown structure, height, and diameter at breast height (dbh). Different young boreal tree species, such as *Picea*, *Pinus*, and *Abies*, exhibit nearly the same reflectance, transmittance, and albedo characteristics at wavelengths below 700 nm [46]. This similarity makes it more challenging for models to distinguish between species accurately. Hence, our second objective of this research was to classify (semantic segmentation) young forest tree species by integrating RGB-NIR Planet lab data, RGB DAP, and their derived spectral indices (vegetation indices) using ML models. Furthermore, incorporating the canopy height model (CHM) into the dataset is anticipated to improve the accuracy of the models by offering valuable structural insights about tree height, which complements spectral and spatial data. The overall approach of this study was, first, to carry out tree species detection using the YOLOv12, second, to compare model classification accuracy at different levels of data complexity: (1) DAP RGB, (2) 8-band data (DAP RGB + CHM + Planet RGB-NIR), and (3) 20-band data (DAP RGB + CHM + Planet RGB-NIR + 12 spectral indices), and third, to prepare a forest species classification map from the best-performing model.

2. Materials and Methods

2.1. Study Area

The study was conducted in a young private forest located in Pieksämäki, Finland (62°21'0" N, 27°6'32" E) (**Figure 1**). The site comprises two forest stands, part of which underwent their first commercial thinning in 2021. The study site covered a total of 16 hectares of young boreal forest, predominantly composed of Scotch pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), and Silver birch (*Betula pendula*). The area also included other Deciduous tree species, such as European aspen

and Downy birch. In 2024, the region recorded a mean annual precipitation of 608 mm and an average temperature of 5.4°C [47]. The two stands were specified within the young forest developmental category and were categorized to the drained heath fertility and mesic heath classifications in both stands. The average height and dbh of the trees in the forest are 13 m and 14 cm, respectively [48].

2.2. Digital Aerial Photogrammetry

The DAP scan was conducted on June 16, 2020, utilizing a 42 MP Sony RX1R II camera (Sony, Tokyo, Japan) mounted to the unmanned aerial vehicle (UAV) GEODRONE X4L (Geotrim Oy, Vantaa, Finland) (**Figure 1**). The drone, engineered to endure wind speeds of up to 18 m/s and capable of traversing distances of up to 2500 m, performed a singular flight for the purpose of image acquisition. The flight trajectory of the UAV for the DAP was directed by a pre-established ground control system, while its IMU and GNSS components recorded the spatial coordinates (longitude and latitude), altitude, and angle of each image in real time. The flying height was 140 meters with an average flying speed of 7 meters per second. The distance between the flight lines was 27 meters. We used Trimble real time kinetic (RTK) technology for establishing accurately positioned five ground control point within the study area. The RGB images achieved a spatial resolution of 4 cm per pixel, with overlaps of 80% longitudinally and 65% laterally. A total of 812 images in JPEG were obtained, covering a broader area that includes the specific study site, all with a photo resolution of 7952 × 5304 pixels.

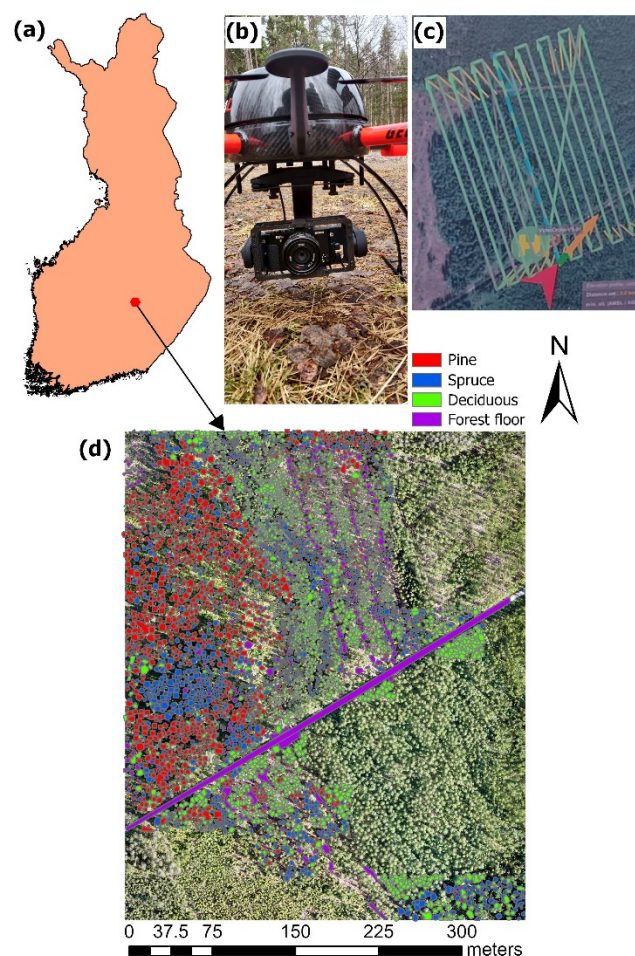


Figure 1. the location of the study area.

The JPEG images were subsequently transformed into digital maps and three-dimensional point clouds utilizing Pix4D software [49], which, through its structure-from-motion algorithm, produced dense point clouds from aerial imagery with an average density of 732 points per square

meter. To assess tree heights, a digital surface model (DSM) was initially created from the point cloud to depict the highest surface elevations, encompassing tree canopies and other objects. A digital terrain model (DTM) was derived by classifying and filtering the ground points within the point cloud, and the CHM was calculated by deducting the DTM from the DSM, thereby accurately illustrating the vertical extent of vegetation above the ground surface. The complete RGB orthomosaic and CHM of the study area were produced with the pixel sizes of 4 cm and 20 cm, respectively. The CHM has been resampled into 4 cm using a bilinear method to make both the RGB and CHM resolutions uniform in ArcGIS Pro, version 3.3.1.

2.3. Preparation of Reference Data

The training data was prepared in ArcGIS Pro, version 3.3.1. The process started with the manual delineation of different tree crowns and Forest floor boundary polygons in an orthomosaic (**Figure 1d**). This necessitated extensive working hours, as a total of 6011 polygons were delineated. Approximately 25% of the study area has been allocated for training and testing data in the models. Given the distinctiveness of the Forest floor and the assumption of effective training across all classes in any ML model, we polygonised the Forest floor with minimal representation—only 13% of the total delineated area. Nonetheless, class balance techniques were implemented during the training phase in all machine learning models. Further details of the delineated polygons are presented in **Table 1**. Additionally, an orthomosaic RGB image of the study area was also available from spring 2020 (**Figure A1**), as two different drone scanning campaigns were conducted during the spring and summer of 2020. The orthoimage of spring 2020 was used to differentiate Deciduous tree species from evergreen ones in a crown delineation process in ArGIS pro.

Table 1. The details of manually prepared reference datasets.

| Species | Number of polygons | Total area (m ²) | % of total delineated area | % of total study area |
|----------------|--------------------|------------------------------|----------------------------|-----------------------|
| Scots pine | 1261 | 11650 | 29.45 | 7.33 |
| Norway spruce | 1622 | 8850 | 22.37 | 5.57 |
| Deciduous spp. | 2755 | 13816 | 34.93 | 8.69 |
| Forest floor | 373 | 5240 | 13.25 | 3.30 |
| Total | 6011 | 39556 | 100 | 24.89 |

2.4. Planet Data

The RGB-NIR planet data [44] were collected for the date of 16 June 2020, aligning with the acquisition period of DAP RGB imagery for the study area. The cloud free four bands with 3-meter pixel resolution images were downloaded from the Planet lab database. An orthorectified, atmospherically corrected, and radiometrically calibrated surface reflectance (Level 3B) image was employed from Dove Classic PS2. As a result of differing pixel resolutions, images of Planet data were resampled to 4 cm using bilinear interpolation in ArcGIS Pro.

2.5. Vegetation Indices

Twelve vegetation indices were derived from three and four spectral bands using DAP and Planet data, respectively. Six different indices for each DAP and planet were calculated for input data. The vegetation indices calculated were DAP: ExG= Excess of Green, GLI= Green Leaf Index, MGRVI= Modified Green Red Vegetation Index, NGRDI= Normalized green–red difference index, RGBVI= Red Green Blue Vegetation Index, VARI= Visible atmospherically resistant index, Planet: ARVI= Atmospherically Resistant Vegetation Index, EVI= Enhanced Vegetation Index, GARI= Green Atmospherically Resistant Index, GNDVI= Green Normalized Difference Vegetation Index, NDGI=

Normalized Difference Greenness Index, and NDVI= Normalized Difference Vegetation Index. The formula and references are provided in **Table 2**.

Table 2. The vegetation indices and their formula used in this study.

| Data | Vegetation index | Formula | Reference |
|--------|------------------|------------------------------------------------------------------------|-----------|
| DAP | ExG | $2 * G - R - B$ | [50] |
| | GLI | $\frac{((2 * G) - R - B)}{((2 * G) + R + B)}$ | [51] |
| | MGRVI | $\frac{G^2 - R^2}{G^2 + R^2}$ | [52] |
| | NGRDI | $\frac{(G - R)}{(G + R)}$ | [53] |
| | RGBVI | $\frac{G^2 - (B * R)}{G^2 + (B * R)}$ | [52] |
| | VARI | $\frac{(G - R)}{((G + R) - B)}$ | [54] |
| Planet | ARVI | $\frac{NIR - (R - 1 * (B - R))}{NIR + (R - 1 * (B - R))}$ | [55] |
| | EVI | $2.5 * \left(\frac{NIR - R}{(NIR + (6 * R) - (7.5 * B)) + 1} \right)$ | [56] |
| | GARI | $\frac{NIR - (G - (B - R))}{NIR + (G - (B - R))}$ | [57] |
| | GNDVI | $\frac{NIR - G}{NIR + G}$ | [57] |
| | NDGI | $\frac{G - R}{G + R}$ | [58] |
| | NDVI | $\frac{NIR - R}{NIR + R}$ | [59] |

2.6. Preparation of the Input Data

In total, 20 raster channels were created by combining all individual bands (DAP: 3 channels + Planet: 4 channels), the CHM (DAP:1 channel), and vegetation indices (12 channels) from both DAP and Planet. For each channel, 57 raster tiffs were created, each having 1024*1024 rows and columns. In order to conduct species identification through the YOLO framework, dataset containing 57 RGB raster images, paired with their relevant annotated data, was applied in the analytical evaluation. For semantic segmentation, every pixel in corresponding masks should be classified with the predefined labels Pine, Spruce, Deciduous and Forest floor. Pixels that were excluded from the manual classification labels were classified utilizing the integrated U-Net model available in ArcGIS Pro 3.3.1

[60]. About 50% of unlabeled pixels were classified with the U-net model, in which most were assigned to the Forest floor. Each set of tiles has image tiles and their corresponding mask layer of 1024*1024 rows and columns with four classes of information, i.e., Pine, Spruce, Deciduous, and Forest floor. Finally, we prepared the complete dataset and categorized into three sets for modelling purposes: 1. 3 channels: RGB only (DAP), 2. 8 channels: RGB+CHM (DAP) + RGB+NIR (Planet) and 3. 20 channels: RGB+ CHM+ 6 VIs (DAP) + RGB+NIR+ 6 VIs (Planet). The input and labeled mask raster including DAP and Planet RGB-NIR, the canopy height model, and some of their vegetation index maps (merged 57 raster tiles) are presented in **Figure 2**.

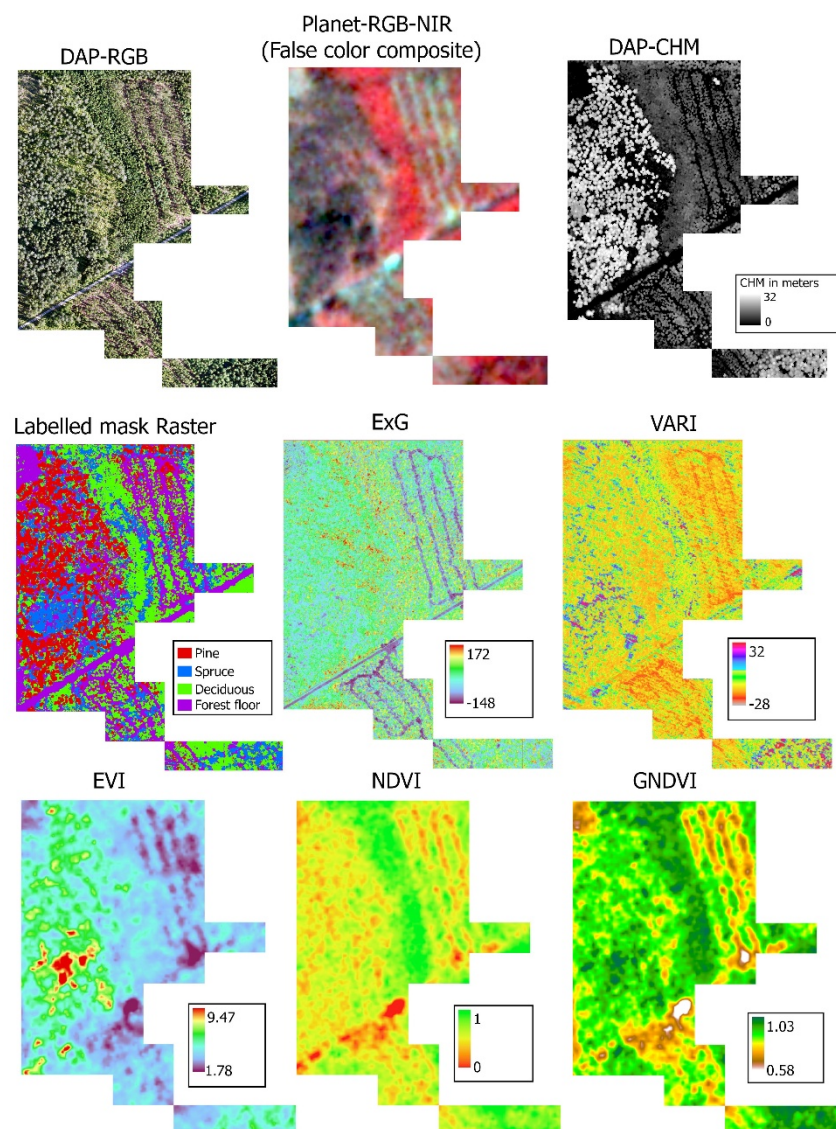


Figure 2. The input rasters and labeled mask of the study area. From ExG to GNDVI are the example of vegetation indices calculated for DAP (first two) and Planet (last three) for the study area. There were 12 vegetation indices (6 each) calculated for DAP and Planet.

2.7. Tree Species Detection and Classification

The species detection was carried out using YOLOv12. The primary objective of the species detection using YOLO was to assess the capability of the latest version of YOLO for individual tree detection in young forest. Since the delineation of tree crowns was carried out manually, we believe that the YOLO detection accuracy will meet expectations, thereby confirming that the training and validation datasets created are appropriately annotated and representative of all the species. The

classifications were then carried out using a random forest, categorical boosting, and convolutional neural networks.

2.7.1. YOLOv12

The YOLOv12 model is the newest addition to Ultralytics' YOLO series of real-time object detection tools [23]. This version is considered better among other versions in terms of accuracy for object detection (**Figure 3**). The manually delineated DAP-RGB region (total crown delineation area) was systematically cropped to dimensions of 1024x1024 pixels to establish a standardized dataset for all semantic segmentation algorithms; however, these images were subsequently resized to 640x640 pixels to comply with the requisite input specifications for the YOLO framework. A distinct format is required for the YOLO model, necessitating the partitioning of the dataset into training, validation, and testing subsets, adhering to an allocation ratio of 80/10/10 for both the images and their corresponding labels. The image and label data were prepared in .tif and .txt formats respectively. The training of the model was performed with a batch size of 25, covering a total of 300 epochs. The model was trained using Google Colab Pro T4 GPU with high ram and Ultralytics 8.3.78, Python-3.11.11, torch-2.5.1+cu124 CUDA:0 (NVIDIA A100-SXM4-40GB, 40507MiB), model=yolo12s.pt.

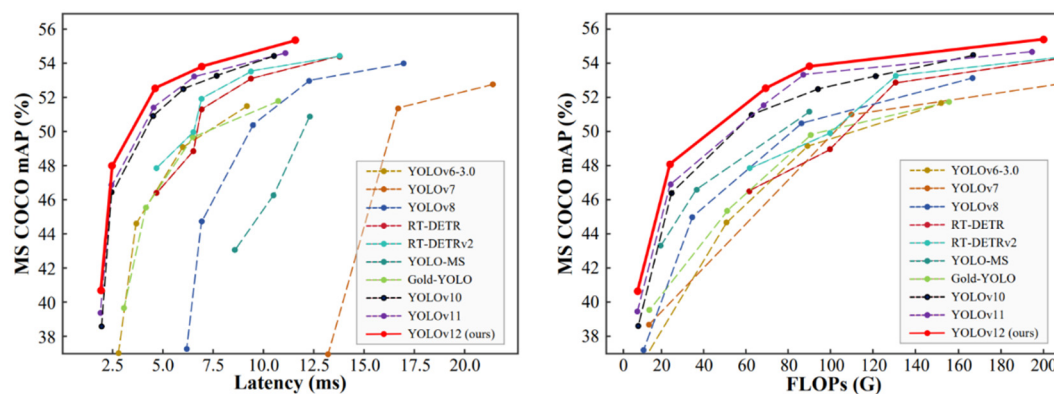


Figure 3. The comparisons of different YOLO versions for latency (left) and FLOPs (right) trade-offs [23].

2.7.2. Random Forest

A random forest constitutes an ensemble method for classification, which is comprised of numerous weak decision trees, each of which is constructed utilizing randomly chosen predictors derived from various subsets of the training dataset. The conclusive prediction or classification outcome is determined through majority voting among the trees [61]. In this approach, decision trees are built independently, with no pruning applied during their construction. The splitting at each node is guided by a randomly selected subset of features, determined by the user-defined `min_sample_split` and `max_depth`. This method creates trees characterized by high variance but low bias, resulting in a collection, or "forest," of multiple trees specified by the `Ntree` parameter (also called `n_estimators`). To classify new data points, the model combines the predictions from all the trees using a majority voting system, ensuring robust and accurate classification. We used the sklearn Python library to run a random forest classifier with optimized parameters: `n_estimators=500`, `max_depth=10`, `min_sample_split=5`, `minsamples_leaf=2`, `max_features= sqrt`, `class_weight= class_weight_dict`. The 5-fold cross validation (CV) grid search hyper parameters tuning was performed using GridSearchCV (`param_grid = {'n_estimators': [50, 100, 500], 'max_depth': [5, 10, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]}`) from sklearn library.

2.7.3. Categorical Boosting

Categorical boosting is a powerful algorithm in the field of machine learning, fundamentally based on the concepts of gradient boosting over decision trees [62]. The algorithm uses ordered target encoding to prevent data leakage and improve generalization during training. Its symmetric tree

structure accelerates the training process and ensures consistent splits across all trees. This design also enhances regularization, reducing the risk of overfitting. The CatBoost ordered boosting technique processes data sequentially, making it robust and accurate for diverse datasets. It automatically handles missing values, further simplifying the data preparation process. Overall, it delivers competitive accuracy with faster inference compared to other gradient boosting models, such as XGBoost and LightGBM [63]. We used the Python library “catboost” with iterations=3000, depth=10, learning rate= 0.1, and loss function= multiclass for this classification. The 5-fold cross-validation (CV) grid search hyperparameter tuning was performed using GridSearchCV (param_grid = {'iterations': [1000, 2000, 3000], 'depth': [5,10,20] 'learning_rate': [0.01, 0.1]}) from the sklearn library. We used GPU acceleration with a CUDA backend through the PyTorch framework to train the CatBoost model.

2.7.4. Convolutional Neural Networks

Convolutional neural networks (CNNs) are distinct category of neural networks designed with care for the analysis of structured grid-like datasets, such as images [64]. These networks skillfully utilize convolutional layers to extract spatial features from the input data, significantly reducing the need for manual feature engineering. CNNs achieve this by applying convolutional filters across an input image, detecting patterns such as textures, edges, and more complex shapes. Pooling layers are subsequently employed to diminish the dimensions of the feature maps, thereby decreasing their size whilst preserving critical information, which enhances computational efficiency. The network generally comprises numerous convolutional and pooling layers, facilitating the acquisition of hierarchical feature representations. Fully connected layers positioned towards the conclusion of the architecture integrate the features to facilitate predictions.

In the current study, we utilized a deep convolutional neural network (CNN) to enable pixel-wise classification in semantic segmentation tasks, whereby each pixel of an input image is assigned a corresponding class label. A 2D CNN architecture with ResNet50V2 [65] based encoder-decoder backbone was implemented and run in a Python environment using Keras with a TensorFlow backend for model execution. The dataset consists of either RGB or 8-band or 20-band raster images, all exhibiting a resolution of 1024×1024 pixels. The ResNet50V2 model was used as the feature extractor with pre-trained weight (“imagenet”) for RGB and without pre-trained weights for 8-band and 20-band, ensuring compatibility with our input.

The encoder extracts multi-scale hierarchical features from the input, progressively reducing spatial dimensions while increasing feature complexity. The decoder consists of multiple upsampling layers to restore the original image resolution. The upsampling process is performed in five stages, progressively increasing the spatial dimensions from 32×32 to 1024×1024. Each upsampling step is succeeded by a convolutional layer incorporating ReLU activation and a kernel dimension of 3×3, with filter sizes gradually decreasing from 512 to 32 as the spatial resolution is restored. The final output layer applies a 1×1 convolution with the number of filters aligned to the total count of classes, subsequently utilizing a softmax activation function to generate probability distributions at the pixel level. The model was constructed using the Adam optimization algorithm and the categorical cross-entropy loss function, which is suitable for multi-class segmentation tasks. Sample weights were reshaped to align with the batch size and averaged across the dataset to address class imbalances. The training of the model spanned 300 epochs, using a batch size of 4 to ensure optimal learning given the high-resolution input images. The model's performance was assessed using a test data, with test loss and accuracy metrics recorded to assess semantic segmentation performance.

2.8. Model Evaluation

The evaluation of the models was conducted using 20% of the testing data, with the exclusion of YOLO species detection. Five metrics from the confusion matrix were employed to measure the effectiveness of the model: overall accuracy, precision (user's accuracy), recall (producer's accuracy), Kappa [66], and the Matthews correlation coefficient (MCC) [67]. The performance of the models was

assessed using 20% data from the test set except for YOLO species detection. Five evaluation metrics derived from the confusion matrix were used to measure model effectiveness: overall accuracy (OA), precision (also referred to as user’s accuracy), recall (known as producer’s accuracy), and the Matthews correlation coefficient. The complete workflow from creating reference and training datasets using Planet and DAP to the final classification map is presented as a chart in **Figure 4**.

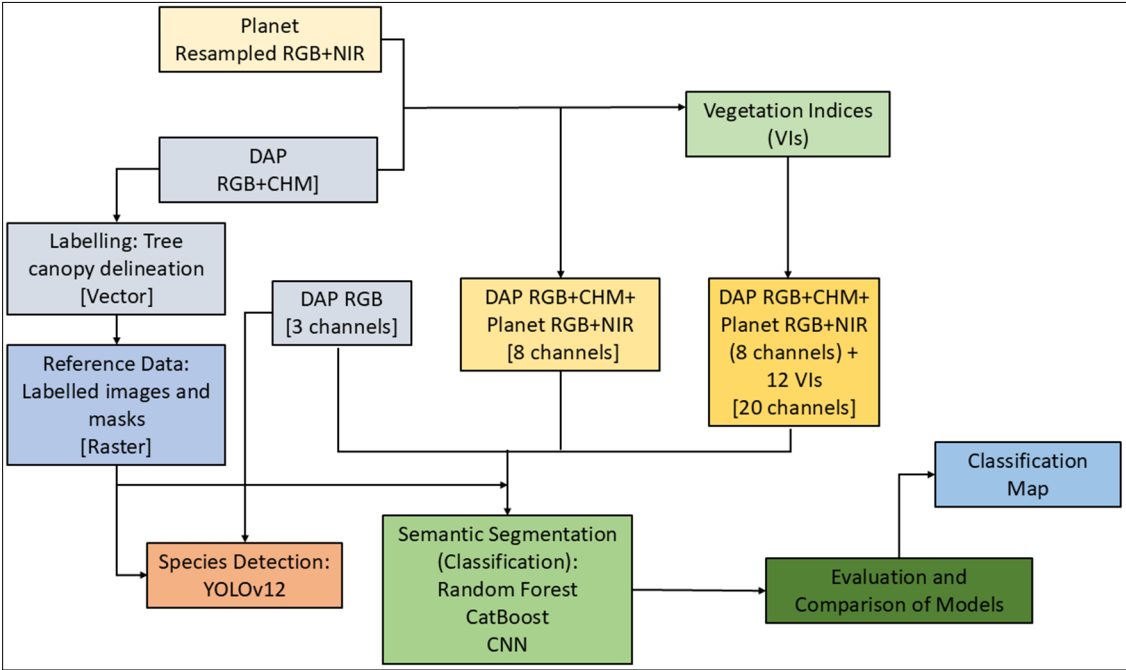


Figure 4. The workflow chart of forest species classification in this study.

The training environment for YOLO is presented in section 2.7.1. For all classification models, the training platform was Intel Xeon Gold 6230. The random forest model was run in CPU using 2x20 cores @2.1 GHz. The CatBoost and CNNs were run using GPU partitioning, NVIDIA Volta V100 GPU. The Jupyter Notebook with Python version 3.10-24.04 was utilized for all the classification training and output.

3. Results

3.1. Tree Species Detection

The confusion matrix (**Figure 5**) shows three determined categories: Pine, Spruce, and Deciduous and their classification by the YOLOv12 model. Pine achieves a high level of accuracy, with 512 instances correctly identified out of 543, demonstrating minimal misclassification into other categories. In a similar fashion, Spruce attained considerable classification performance, with 407 accurate predictions out of 477 instances. Deciduous achieved substantial classification performance, with 1518 accurate predictions out of 1,646 instances. In CNN models such as YOLO, the presence of the background class as a non-detected object is common. In this context, the total instances detected as background are distributed among the different classes in the reference dataset. Notably, 80% of these non-detected objects were classified as Deciduous, with the remaining instances divided between Pine and Spruce. An important observation in this model is that misclassifications into the Background class are less critical than misclassifications into other class labels, such as Pine being predicted as Spruce or Deciduous [68]. Fortunately, our model demonstrates strong performance in this regard, with very few instances incorrectly classified as another class. To provide a deeper assessment of the model’s performance, additional metrics are presented in the accompanying **Table 3**.

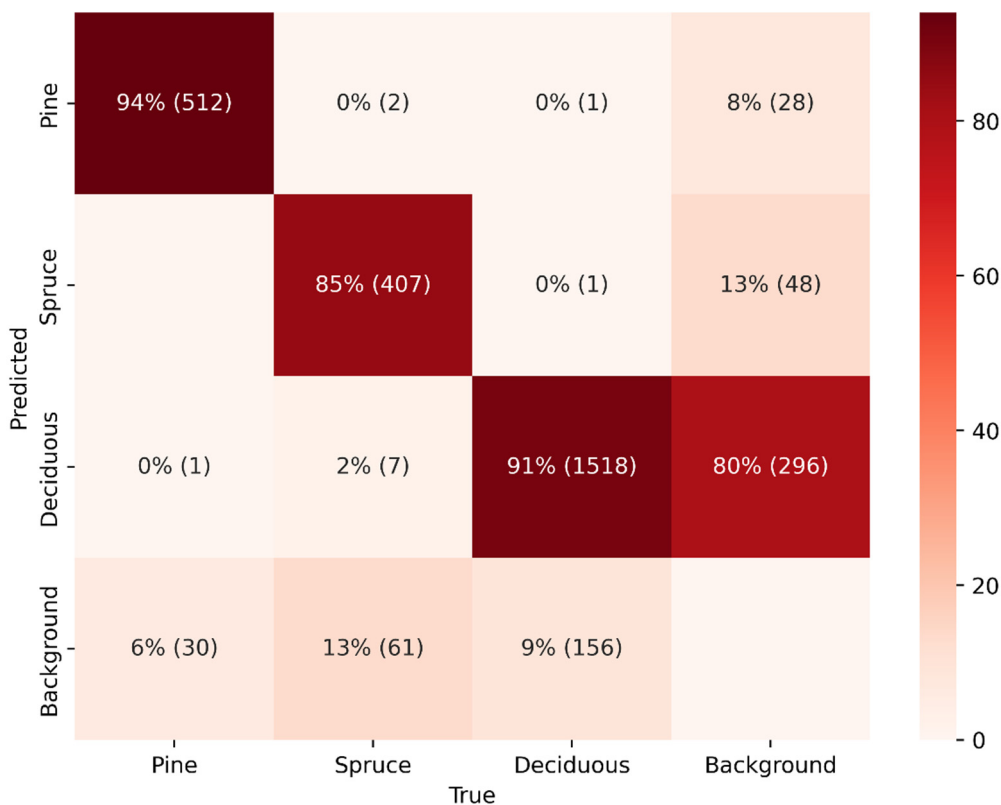


Figure 5. The confusion matrix of YOLOv12 model.

The evaluation metrics for the YOLOv12 model are summarized in **Table 3**. The validation dataset comprises 30 images with a total of 2,696 annotated instances across three classes: Pine, Spruce, and Deciduous. The model revealed a precision of 95% and a recall of 87%, alongside a mean average precision (mAP) of 93% at a 50% intersection over union (IoU) threshold and 75% across various IoU thresholds from 50% to 95%. Among the three classes, Pine exhibited the highest detection performance, with a precision of 97%, recall of 92%, mAP50 of 97%, and mAP50-95 of 80%. In contrast, Spruce showed the lowest recall (82%) and mAP50-95 (71%), indicating potential challenges in detecting all instances accurately, especially at stricter IoU thresholds. These findings emphasize the model’s effectiveness in identifying Pine and Deciduous species, while also indicating areas where Spruce tree detection can be enhanced. The overall accuracy calculated from the confusion matrix (total number of correct prediction/ total number of prediction) was 79%. Additionally, the Kappa and MCC matrices were 66% and 67%. More information on precision and recall is provided in the form of a precision-recall curve in **Figure A3**.

Table 3. The accuracy assessment of tree species detection from YOLOv12.

| Class | Images | Instances | Precision | Recall | mAP50 | mAP50-95 |
|-----------|--------|-----------|-----------|--------|-------|----------|
| All | 30 | 2696 | 0.95 | 0.87 | 0.93 | 0.75 |
| Pine | 20 | 543 | 0.97 | 0.92 | 0.97 | 0.80 |
| Spruce | 25 | 477 | 0.94 | 0.82 | 0.90 | 0.71 |
| Deciduous | 30 | 1676 | 0.93 | 0.87 | 0.93 | 0.73 |

Figure 6 provides a detailed visualization of the spatial distribution of three tree species—Pine, Spruce, and Deciduous—within the study area, overlaid with reference polygons and YOLOv12-detected bounding boxes. This visual approach facilitates a detailed evaluation of the model’s effectiveness in detecting and classifying tree species. The insets reveal cases of correct detections, where bounding boxes align with reference polygons, alongside instances of misclassification and

missed detections. The **Figure A4** provides an elaborate illustration of a small segment within the study area. The findings indicate that YOLOv12 effectively recognizes tree species in isolated or conspicuous regions. However, the model is challenged by areas with dense seedling vegetation (upper-right and lower-left corner in the **Figure 6**), where visual complexity obstructs the identification of individual trees.

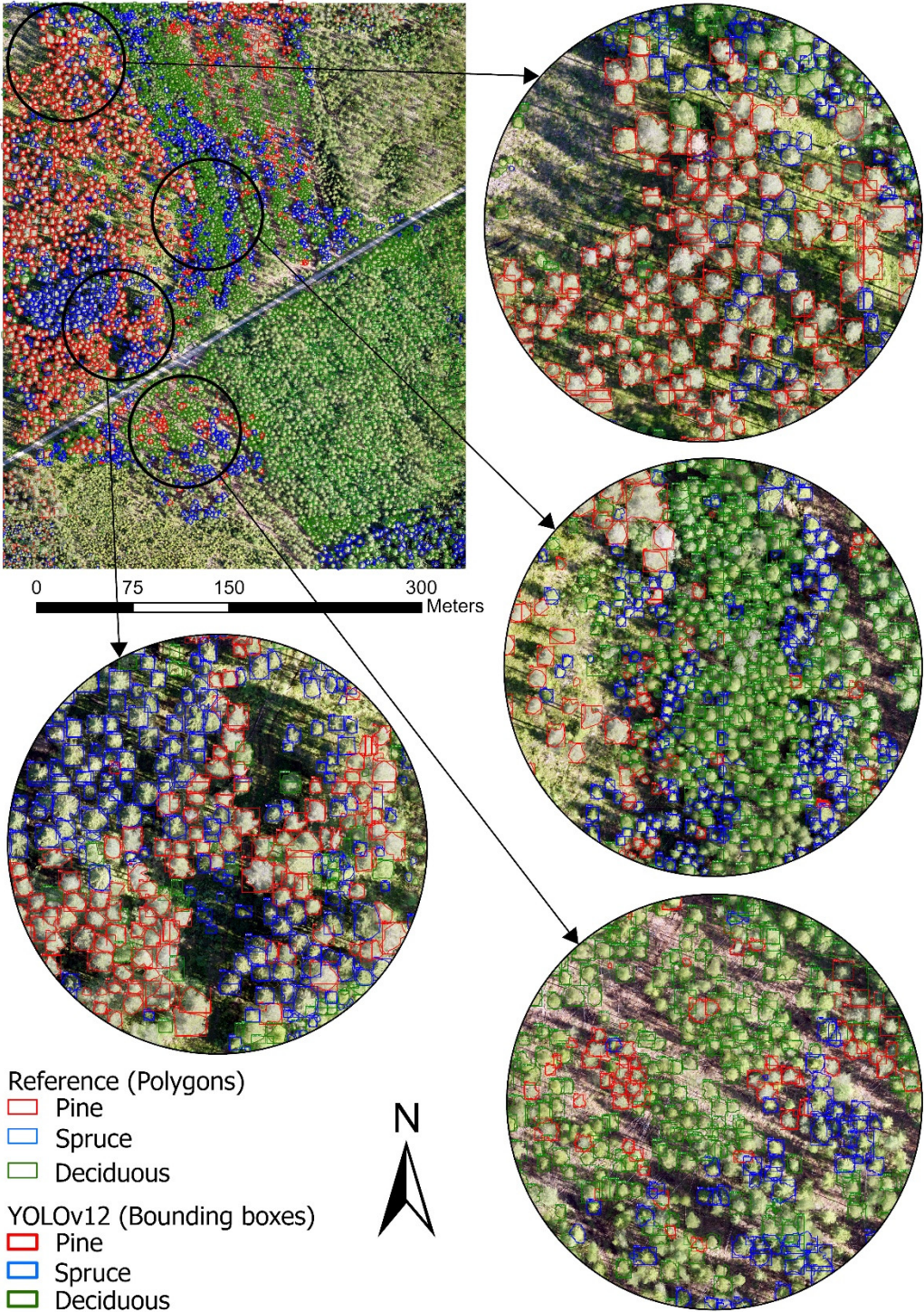


Figure 6. The species detection from YOLOv11 for the study site.

3.2. Classification

The classification process employed three types of models: random forest, CatBoost, and convolutional neural network. Each of these models was further tested in three configurations,

utilizing raster tiles with 3 channels (RGB), 8 channels, and 20 channels. Our analysis focused on evaluating the impact of incorporating additional channels on model accuracy, comparing their performance in various configurations. The classification task extended to include the Forest floor, enabling a comprehensive categorization of the study area. This approach resulted in four distinct classes: Pine, Spruce, Deciduous, and Forest floor, ensuring that the entire study area was represented within the classification framework.

3.2.1. Performance of the Models

Figure 7 shows confusion matrices (calculated from 20% test data) of three models: random forests, CatBoost, and CNNs when varying input features are applied: RGB, 8 bands, and 20 bands. Both random forests and CatBoost demonstrate a clear trend of improved classification accuracy as the number of spectral bands increases. For example, CatBoost’s accuracy for Deciduous rises from 47% with RGB to 91% with 20 bands, while the random forest shows a marked improvement for Forest floor, moving from 60% with RGB to 83% with 20 bands. However, CNNs exhibit inconsistent patterns; significant performance gains are noted for all classes except Pine. Pine’s performance slightly declines from 75% (RGB) to 72% (8 bands) and 71% (20 bands), whereas Spruce improves, oscillating between 60% (RGB), 65% (8 bands), and 60% (20 bands). Deciduous demonstrates a steady improvement from 73% (RGB) to 73% (8 bands) and 75% (20 bands). Similarly, the forest floor class declines from 83% (RGB) to 77% (8 bands) but rises to 85% (20 bands), indicating a significant benefit from enhanced spectral information. These findings suggest that the improvement of CNN performance attributable to additional spectral bands is dependent on the class. These findings highlight that random forest and CatBoost effectively utilize the supplementary spectral data, while CNNs demonstrate limited adaptability to complexities beyond RGB bands.

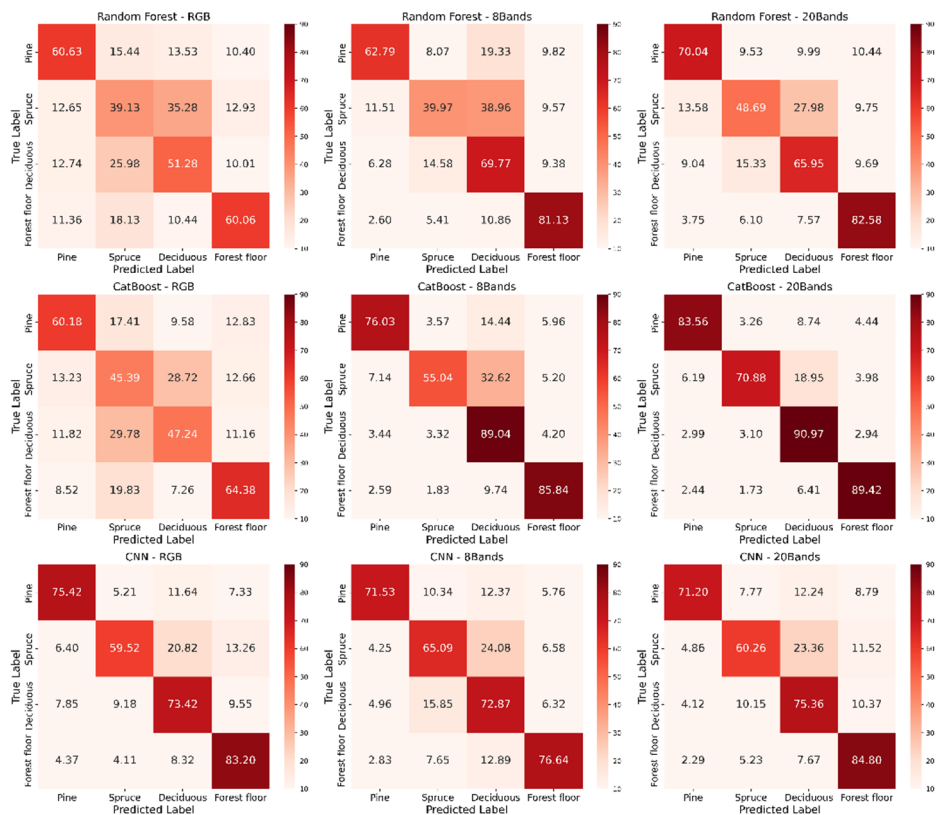


Figure 7. The confusion matrices (%) using RGB to 20 channels in all three models. Each row contains the same model, and each column has the same channel, i.e., RGB or 8 bands or 20 bands.

Likewise, **Table 4** shows the different performance matrices calculated using a confusion matrix derived from 20% test datasets. Both random forests and CatBoost demonstrate a consistent

improvement in all classification performance with increasing input channels, as reflected by higher overall accuracy, precision and recall, F1 scores, MCC, and Kappa values. For example, in random forest, the precision for Spruce improves from 0.29 with RGB to 0.49 with 20 bands, and the overall score increases from 0.54 to 0.70. Similarly, CatBoost shows a substantial increase in performance, with the recall for Deciduous rising from 0.47 (RGB) to 0.91 (20 bands), and the overall score increasing from 0.55 to 0.85. In contrast, CNN demonstrates less consistent improvement in performance when increasing from RGB channels to 20 channels. There is a precision improvement from 81% (RGB) to 88%(20 bands) for Pine. Other than this, there is no any significant improvement in any matrices while adding the spectral channels into RGB in CNN model. The overall score remains constant at 0.74 for RGB and 20 bands for CNN, suggesting reduced advantages from supplementary spectral channels in comparison to random forest and CatBoost.

Table 4. Performance metrics produced by all three models.

| Model | Channels | Species | Precision | Recall | F1 Score | Overall score | Kappa | MCC |
|---------------|----------|--------------|-----------|--------|----------|---------------|-------|------|
| Random forest | RGB | Pine | 0.59 | 0.61 | 0.60 | 0.54 | 0.38 | 0.38 |
| | | Spruce | 0.29 | 0.39 | 0.33 | | | |
| | | Deciduous | 0.56 | 0.51 | 0.53 | | | |
| | | Forest floor | 0.69 | 0.60 | 0.64 | | | |
| | 8 Bands | Pine | 0.76 | 0.63 | 0.69 | 0.66 | 0.54 | 0.54 |
| | | Spruce | 0.47 | 0.40 | 0.43 | | | |
| | | Deciduous | 0.60 | 0.70 | 0.64 | | | |
| | | Forest floor | 0.77 | 0.81 | 0.79 | | | |
| | 20 Bands | Pine | 0.72 | 0.70 | 0.71 | 0.70 | 0.60 | 0.60 |
| | | Spruce | 0.49 | 0.49 | 0.49 | | | |
| | | Deciduous | 0.68 | 0.66 | 0.67 | | | |
| | | Forest floor | 0.77 | 0.83 | 0.80 | | | |
| CatBoost | RGB | Pine | 0.62 | 0.60 | 0.61 | 0.55 | 0.40 | 0.40 |
| | | Spruce | 0.29 | 0.45 | 0.36 | | | |
| | | Deciduous | 0.61 | 0.47 | 0.53 | | | |
| | | Forest floor | 0.68 | 0.64 | 0.66 | | | |
| | 8 Bands | Pine | 0.85 | 0.76 | 0.80 | 0.79 | 0.71 | 0.72 |
| | | Spruce | 0.80 | 0.55 | 0.65 | | | |
| | | Deciduous | 0.70 | 0.89 | 0.78 | | | |
| | | Forest floor | 0.87 | 0.86 | 0.87 | | | |
| | 20 Bands | Pine | 0.88 | 0.84 | 0.86 | 0.85 | 0.80 | 0.81 |
| | | Spruce | 0.85 | 0.71 | 0.77 | | | |
| | | Deciduous | 0.80 | 0.91 | 0.85 | | | |
| | | Forest floor | 0.91 | 0.89 | 0.90 | | | |
| | RGB | Pine | 0.81 | 0.75 | 0.78 | | | |

| | | | | | | | | |
|-----|----------|--------------|------|------|------|------|------|------|
| CNN | | Spruce | 0.65 | 0.60 | 0.62 | | | |
| | | Deciduous | 0.74 | 0.73 | 0.74 | | | |
| | | Forest floor | 0.73 | 0.83 | 0.78 | | | |
| | | | | | | 0.74 | 0.65 | 0.65 |
| | 8 Bands | Pine | 0.86 | 0.72 | 0.78 | | | |
| | | Spruce | 0.52 | 0.65 | 0.58 | | | |
| | | Deciduous | 0.69 | 0.73 | 0.71 | | | |
| | | Forest floor | 0.80 | 0.77 | 0.78 | | | |
| | | | | | | 0.72 | 0.62 | 0.63 |
| | 20 Bands | Pine | 0.88 | 0.71 | 0.79 | | | |
| | | Spruce | 0.60 | 0.60 | 0.60 | | | |
| | | Deciduous | 0.73 | 0.75 | 0.74 | | | |
| | | Forest floor | 0.73 | 0.85 | 0.79 | | | |
| | | | | | | 0.74 | 0.65 | 0.65 |

Figure 8 summarizes the overall accuracy, Kappa and MCC for all nine models. The overall accuracy demonstrated a significant improvement of 30% for random forest and 55% for CatBoost when transitioning from the RGB model to the 20-band model. For Kappa scores, random forest achieved 58%, while CatBoost reached 100% with an MCC of 103%. However, the CNN exhibited no accuracy enhancement when shifting from RGB to 20 bands model . Based on the output from all models across different input channels, CatBoost with 20 bands emerged as the best-performing model, achieving an overall accuracy of 85%, a kappa value of 80%, and an MCC of 81% in this study.

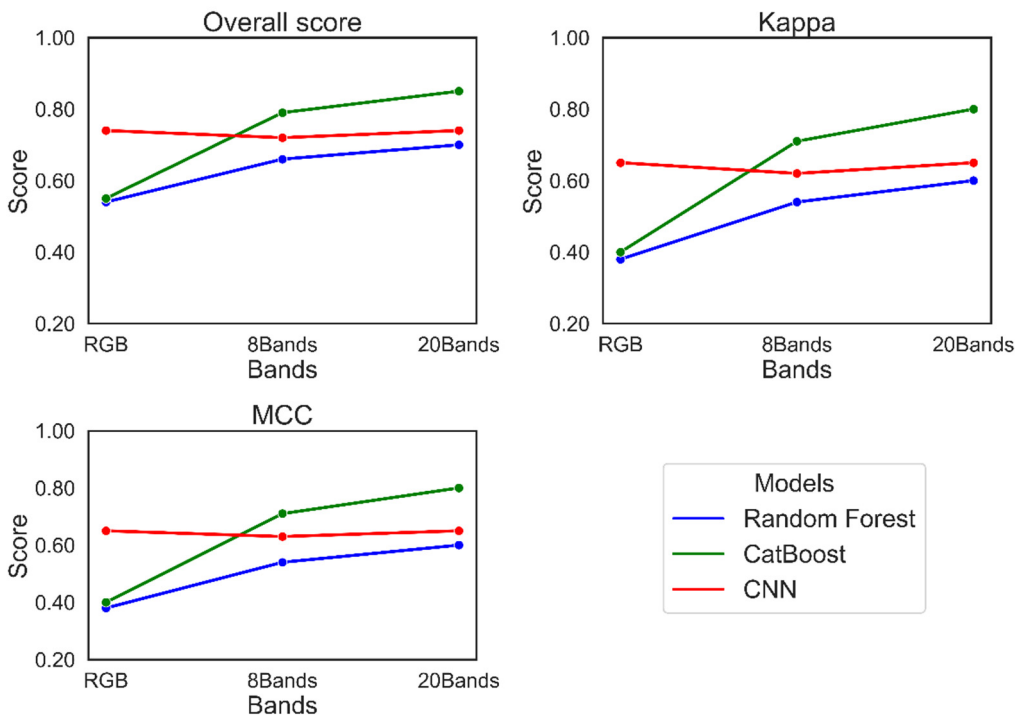


Figure 8. The Overall accuracy, Kappa and MCC matrices for all models.

3.2.2. Feature Importance

The variable importance score is presented as a percentage for the random forest and CatBoost models (**Figure 9**). Since RGB has only 3 channels, we calculated the feature importance score for 8 and 20 bands. In the 8-band random forest model, the blue band from DAP stood out as the leading feature, comprising 63% of the model's training. Moreover, the 8-band random forest model determined that blue and red bands from DAP and Planet datasets were significant variables, whereas CHM showed no contribution. Conversely, in the 20-band random forest model, CHM alone accounted for 43% of the total contribution. Other important indices, including EVI, GNDVI, VARI, and NDVI, each contributed approximately 8–10% to the model.

In the CatBoost model, the 8 bands again highlighted the blue from DAP as the most influential variable, contributing approximately 19%. However, unlike random forests, the contributions of other variables were more evenly distributed, ranging from about 5% to 19%. This balanced contribution of variables could be one of the key reasons for the higher accuracy observed in the CatBoost model. Red, NIR, and green bands from planet, along with CHM, were the second, third, fourth, and fifth most important spectral indices, each contributing around 14–17%. Similar to the random forest model, CHM (15%) and EVI (12%) were the top two contributing variables in the 20-band CatBoost model. Notably, spectral indices from Planet, such as NIR, green band, NDGI, GNDVI, NDVI, and blue band, played a major role in model training. Across all four models, it appears that aside from CHM and the blue band, the models primarily relied on spectral indices derived from Planet for classification.

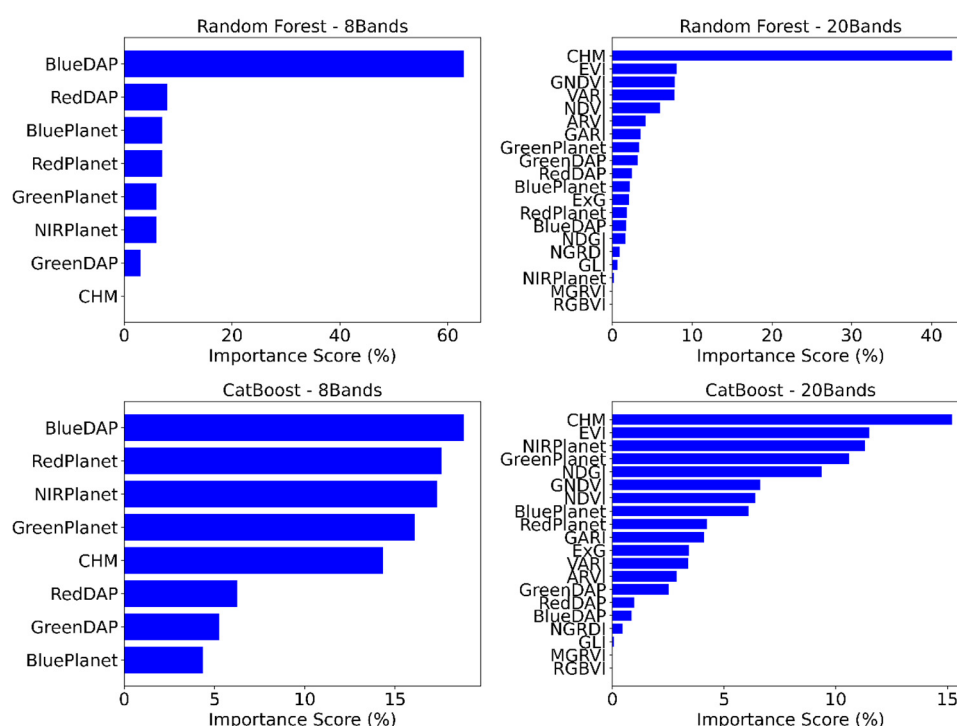


Figure 9. The feature importance score (%) for the random forest and CatBoost model.

3.3. Classification Map

Among nine model combinations, CatBoost with 20 bands proved to be the most accurate, clearly outpacing others in all evaluation metrics, including precision, recall, and MCC. Additionally, this model achieved the highest accuracy for all four classes. The classified raster map using the CatBoost 20 bands model is shown in **Figure 10**. It is important to highlight that the model exhibited remarkable competence in the classification of the Forest floor, while also demonstrating

praiseworthy performance in the categorization of Pine and Deciduous. However, some classification errors were observed for Spruce, where the model frequently confused it with Deciduous. This misclassification was also evident in other models, particularly in random forests with RGB and CNNs with 8-band inputs. Overall, the CatBoost 20-band model definitively demonstrated a strong ability for semantic segmentation of young boreal forest.

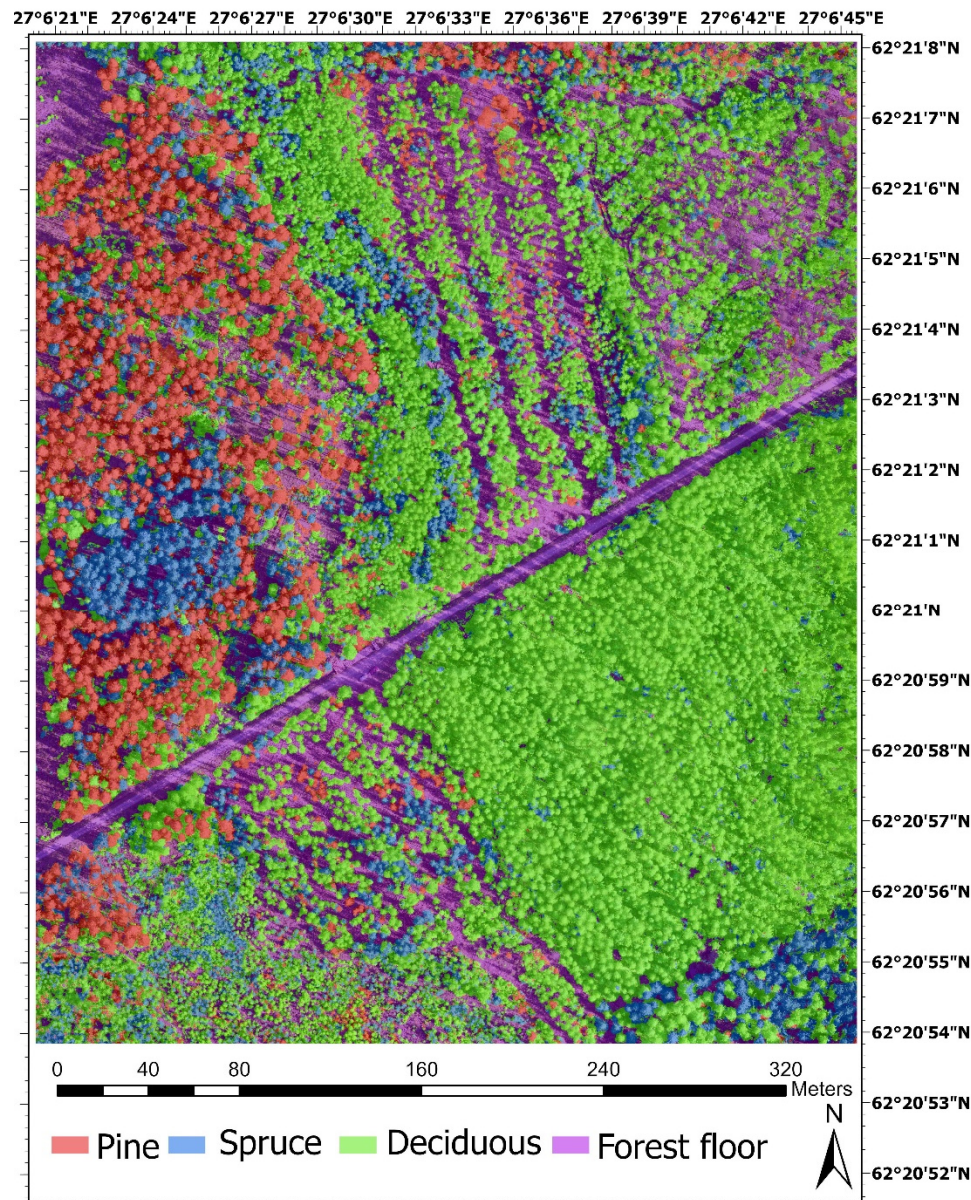


Figure 10. The classification map predicted from the CatBoost model for the whole study area. The predicted map with all classes of 50% transparency is placed right above the orthomosaic image of the study area.

4. Discussion

In this study, we successfully demonstrated that automated identification, classification and geolocation of three key tree species in young mixed boreal forests is possible using deep learning object detection and semantic segmentation, consistent with findings from previous studies [13,15,17,19,21,25,29,35,42]. However, our study adds the implication of integrating Planet and UAV for better classification of young boreal tree species. Moreover, to the best of our knowledge, this is the first study to apply the latest version of YOLO (version 12) for detection of young boreal tree species – namely, Scotch pine, Norway spruce and Deciduous species. This study also involved labeling and training models for over 6,000 annotated individual trees, including the Forest floor,

which were highly time-intensive processes. There are numerous well-established, publicly available labeled image datasets for deep learning applications in object detection, segmentation, and classification across various fields. Examples include MNIST for handwritten digits and fashion products, COCO, PASCAL VOC, OpenImages for general object detection, and medical datasets such as chest X-rays, CT scans, and MRIs. Additionally, datasets exist for plant disease detection in leaves and crop and weed classification. However, when it comes to tree species detection and classification, there is a significant lack of open access labeled datasets. To address this gap, it is crucial to gather and share as much labeled tree species data as possible, ensuring that future researchers do not have to manually label data from scratch. Our study contributes to expanding the tree species dataset pool, providing a valuable resource for future research in forest monitoring and classification. This large dataset and the trained models will also provide additional inputs to the database stated in a previous study [12] for future research in tree species detection and classification.

The tree species detection using YOLOv12 in our study is comparable to previous studies [13,15,19]. Our results show that young tree species – Scotch pine, Norway spruce and Deciduous trees (Silver birch, etc.) – can be detected with overall accuracy rates of 79% using the YOLOv12 model. The overall precision, recall and mAP are higher than those of [15] and vice versa with the prior studies [13,19]. The study by Zhong et al. [15] explored the possibility of different models, including Retina Net, Faster R-CNN and YOLO versions 5 and 8, for tree species identification and found YOLOv8 outperformed others (precision=0.74 and recall=0.72). One reason our study achieved better performance could be the use of the latest YOLOv12 model for tree detection. Previous studies also explored a wide range of spatial resolutions (from 2.7 cm to 80 cm) and found that tree species detection accuracy remained stable between 2.7 cm and 8 cm [15]. A similar trend is also found in a study conducted for tree crown detection using UAV RGB imagery and deep learning methods [14]. However, beyond 8 cm, accuracy began to decline significantly, with the sharpest drop occurring specifically between 15 cm and 80 cm. In contrast, another study [13] used 10 cm resolution image tiles (1024*1024) and found an overall precision of 99% in tree species detection using YOLOv8. While spatial resolution is always a contributing factor in detection accuracy, it is not the sole determinant. Other study for tree species detection in shelterbelts [19] used YOLOv7 with Kmeans++_CoordConv_CBAM (YOLOv7-KCC), meaning that the model was improved adding input data augmentation, K-means clustering, Coordinate Convolution and convolutional block attention module and found mAP50-95 of 0.78. Whereas our study found mAP50-95 of 0.75 without data augmentation and modification. In the same study [19], the YOLOv7 found only 0.67 of mAP50-95. The intricate background of the shelterbelts adversely affected the precision of identifying tree species [19].

This indicates the performance power of YOLOv12; however, it always depends on the characteristics of the input data, background of the images, architecture and parameters of the model. The key differences in accuracy among studies, including ours, could be attributed to background, forest type, UAV scanning period, the preparation of ground truth data and the type of model. Our study focused on a mixed coniferous-broadleaf young boreal forest, where the average dbh (14 cm) and height (13 m) made tree species detection particularly challenging in comparison to fully matured [15,19] or sparse and matured tree stands [13]. The scanning period also played a role in detection accuracy. Studies conducted around the fall season showed higher accuracy compared to our study, which was performed during the summer. Research on tree species classification across multiple seasons has also found that autumn is the ideal time for acquiring the UAV imagery, as it leads to the best performance of the model [69]. This might be due to seasonal changes in leaf color, canopy structure, and reduced foliage density, which may enhance spectral differentiation between species, specifically between evergreen and Deciduous species leading to improved classification accuracy.

Moreover, we compared the classification accuracy from three machine learning models for utilizing additional channels and spectral indices from DAP and planet. The results indicate that random forest and CatBoost made better use of the additional information from multispectral bands,

leading to improved performance. In this study, there was a significant increase in overall accuracy when the number of stacked input channels was expanded from RGB to 20 bands, with random forest improving from 54% to 70% and CatBoost rising from 55% to 85%. A similar trend was observed in the study by Immitzer et al. [35], where the overall accuracy increased from 87% to 96% when the number of input channels was expanded from four to eight using a random forest pixel-based classification model. However, increasing the number of features from 16 to 33 led to only a slight improvement in overall accuracy, rising from 82% to 83% when using the random forest classifier [31]. Additionally, limiting the input features to 20 was a practical choice in our study, as processing larger datasets would have required significantly more computing resources to run efficiently in any model. In our study, only structural (CHM) and spectral channels were used for classification, without incorporating textural matrices, which may have been a limiting factor, particularly for the random forest model. Since the random forest model typically performs better when combining spectral and structural indices [42], this omission may explain why it had the lowest accuracy among the models. Incorporating textural features, such as gray level co-occurrence matrix (GLCM) features, could potentially increase the complexity of the CNN model, requiring modifications to its architecture. Specifically, a dual-branch or two-parallel ResNet50V2 encoder setup (one for spectral and one for textural features) would need to be implemented in a 2D CNN. However, this would not only make the model more complex but also demand significantly higher computing resources. Additionally, previous studies have also found that adding structural features did not necessarily improve the performance of random forest, support vector machine, or CNN models [70–72]. Therefore, to maintain a balance between machine learning and deep learning input datasets, we decided to use only spectral input features across all our models.

In our study, CHM and EVI were the most influential variables in both the higher-performing (20-band) random forest and CatBoost models. Additionally, Green (Planet), GNDVI, NDGI, NDVI, and VARI also played a significant role, contributing to the improved performance of both models. In the best-performing model (CatBoost, 20-band), the most significant features were CHM, EVI, NIR (Planet), Green (Planet), NDGI, GNDVI, and NDVI, contributing between 6% and 15% to the performance of the model. A similar pattern was observed in the study of tree species classification using DAP and LiDAR [42], where height metrics related to CHM (H95, H50, etc.), GNDVI, and NDVI were identified as key contributing variables, aligning with the findings of our study. The significance of CHM in tree species classification must be highlighted, as it introduces a crucial additional dimension—tree height—which spectral indices alone cannot provide. Some tree species may exhibit similar spectral signatures but differ in height or canopy structure. For instance, Deciduous and coniferous trees may have overlapping spectral characteristics but can be distinguished by their height or crown shape. Incorporating CHM alongside spectral indices significantly improved semantic segmentation, particularly in the random forest and CatBoost models in our study. The similar results found in the study of tree species classification using random forest by integrating CHM with Sentinel-2 data [73], where they found up to 7% increase in the overall accuracy.

In contrast, CNN showed less sensitivity to the increase in input features, resulting in no accuracy gains. But there is an accuracy increasement in specific tree species such as Scots pine. A possible reason why CNN accuracy did not improve with the addition of vegetation indices could be its architecture with a backbone of ResNet50V2, which is originally pre-trained on ImageNet, a dataset containing millions of RGB images. In addition to this, lacking advanced backbones such as AlexNet, Resnet152, or EfficientNet that could typically enhance feature extraction and performance of the CNN model. Incorporating pre-trained advanced backbones can significantly enhance classification performance, boosting overall accuracy from 44% to 83% in tree species classification with random forest [36]. Alternatively, exploring different architecture, such as U-Net have provided outstanding overall accuracy of 89% in classifying forest tree species using UAV RGB of 2 cm spatial resolution with the inclusion of normalized digital surface model (nDSM) [25]. Therefore, incorporating advanced CNN models such as U-Net and DeepLabV3 could further enhance model

performance and is also recommended for future semantic segmentation for tree species studies. However, for semantic segmentation, preparing ground truth data was particularly challenging, as it required manual classification of every pixel to create mask data. In our case, manually labeling all 57 mask tiles (each with a 1024×1024-pixel size) was impractical due to the young forest, where many of the small tree species were difficult to distinguish with the human eye. To address this, we used a pretrained U-Net model from ArcGIS Pro to annotate pixels that were not manually classified. This was a key constraint in our study—had we been able to manually label all mask tiles, the performance of all the models would likely have been stronger in the classification.

In the recent development of deep learning models such as neural networks which are seemingly outperforming to all other models in most of the classification and regression tasks including tree species classification. However, the possibility of boosting methods such as CatBoost, XGBoost and LightGBM is still underrated/undermined in the forest species classification. We also suggest that future studies in tree species segmentation and classification using diverse models from simple machine learning such as SVM to advance deep neural networks. Deep learning does not always outperform machine learning methods, as performance depends on various factors. Deep neural networks typically require well-defined object boundaries for accurate classification, but in DAP images, tree crown edges often blend into the background, making it difficult for the model to extract clear spatial patterns. In contrast, boosting algorithms operate differently, as they do not rely on spatial structure but instead classify each pixel based on its spectral values, avoiding issues with blurry or mixed boundaries. Therefore, model performance is not solely determined by its architecture but also by factors such as the available input data, ground truth quality, and the morphological characteristics of forest tree species.

5. Conclusions

Our study demonstrates the successful utilization of the newest version of YOLO (version 12) model with an accuracy of 79 % for detecting young boreal tree species. The precision and recall values revealed that the 95% of the trees (true positives) were correctly detect and 13% (false negatives) were missed to detect by the model. The model achieved overall precision of 75 % on stricter IoU scale (0.50-0.95), which is a very good result. Moreover, the model performed exceptionally well for Pine, with precision of 97% and mAP50-95 of 80%. Whereas the model slightly underperformed for detecting Spruce trees with precision of 94% and mAP50 of 90%. For both Spruce and Deciduous species, the model performed well, however, mAP50-95 of 71% and 73% respectively suggest that bounding boxes may be slightly off from the detection of the trees.

Likewise for semantic segmentation, CatBoost model with input data of 20 channels performed the best with overall accuracy of 85 %. There was a maximum of 55% increase in the overall accuracy in CatBoost model with the addition of 17 spectral indices into RGB image. There is also a high improvement in accuracy up to 33% with the inclusion of additional spectral data from planet in DAP RGB using random forest model. However, the CNN performed well with the overall accuracy of 74% even in RGB data, feeding more features did not add value for semantic segmentation. The contributing features in random forest and CatBoost models were mainly CHM, EVI, NIR (Planet), Green (Planet), NDGI, GNDVI, and NDVI. Hence, from our study, we conclude that Catboost model with addition of canopy height model and spectral indices calculated from both DAP and planet lab would be best choice for sematic segmentation in young boreal forest.

Author Contributions: Conceptualization, A.G.; methodology, A.G.; software, A.G.; validation, A.G., M.A. and T.R.; formal analysis, A.G.; investigation, A.G., M.A. and T.R.; resources, A.G., M.A. and T.R.; data curation, A.G., and M.A.; writing—original draft preparation, A.G.; writing—review and editing, A.G., M.A. and T.R.; visualization, A.G., M.A. and T.R.; supervision, M.A. and T.R.; project administration, M.A. and T.R.; funding acquisition, M.A. and T.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research is conducted under the PUUSTI (Puun tarjonnan kestävä edistäminen digitalisoinnin ja verkostoitunisen keinoin) project, focusing on sustainable promotion of wood supply through digitalization and

networking. The initiative is funded by the Rural Development Program for Mainland Finland (Project code: 101242) and additional support from Suur-Savon Energiasäätiö.

Data Availability Statement: The study did not report any data. The raw data and python code will be made available upon request for a valid reason.

Acknowledgments: The authors wish to acknowledge CSC-IT Center for Science, Finland, for generous computational resources.

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

Declaration of use of AI: During the formulation of this manuscript, the author(s) employed Open AI ChatGPT and SCISPACE as a mechanism to ascertain grammatical errors and enhance the clarity of sentences. Subsequent to the utilization of these tools, the author(s) conducted a thorough review and refinement of the content as deemed necessary and accept(s) complete accountability for the material presented in this publication.

Appendix A



Figure A1. The RGB Orthomosaic of the study area captured from drone in spring 2020.

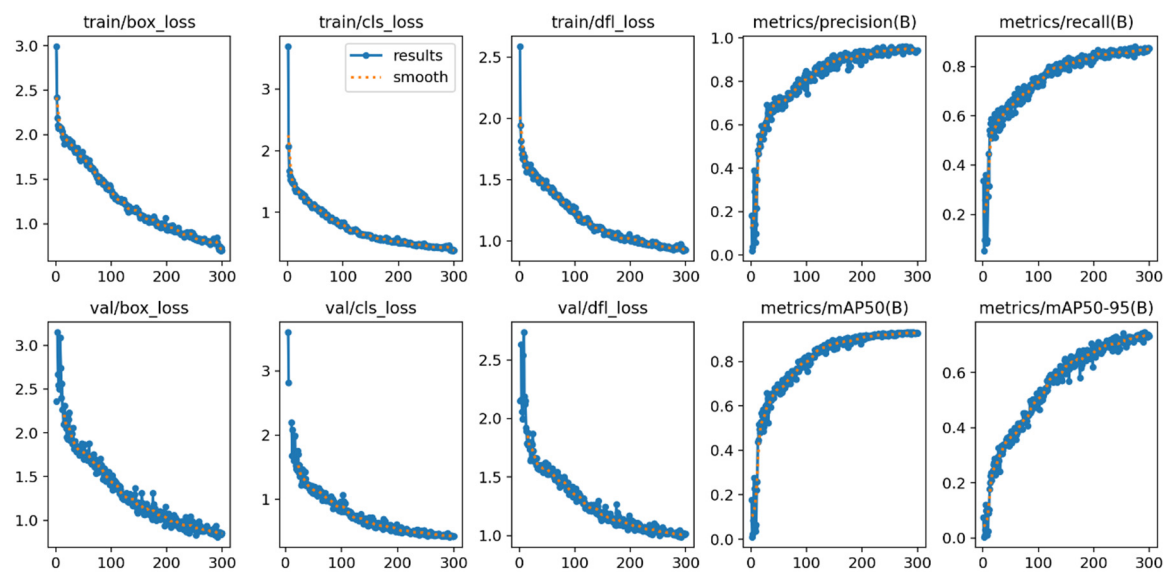


Figure A2. Visualization of YOLOv12 training and validation losses and performance metrics (Precision, Recall, and mAP) over 300 epochs.

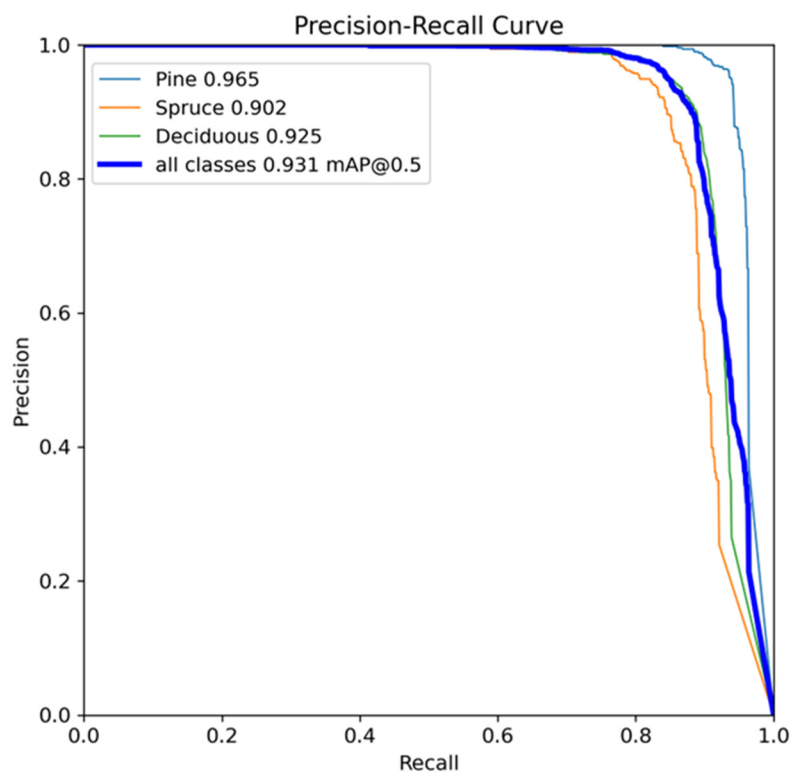


Figure A3. The precision-recall curve produced from YOLOv12.

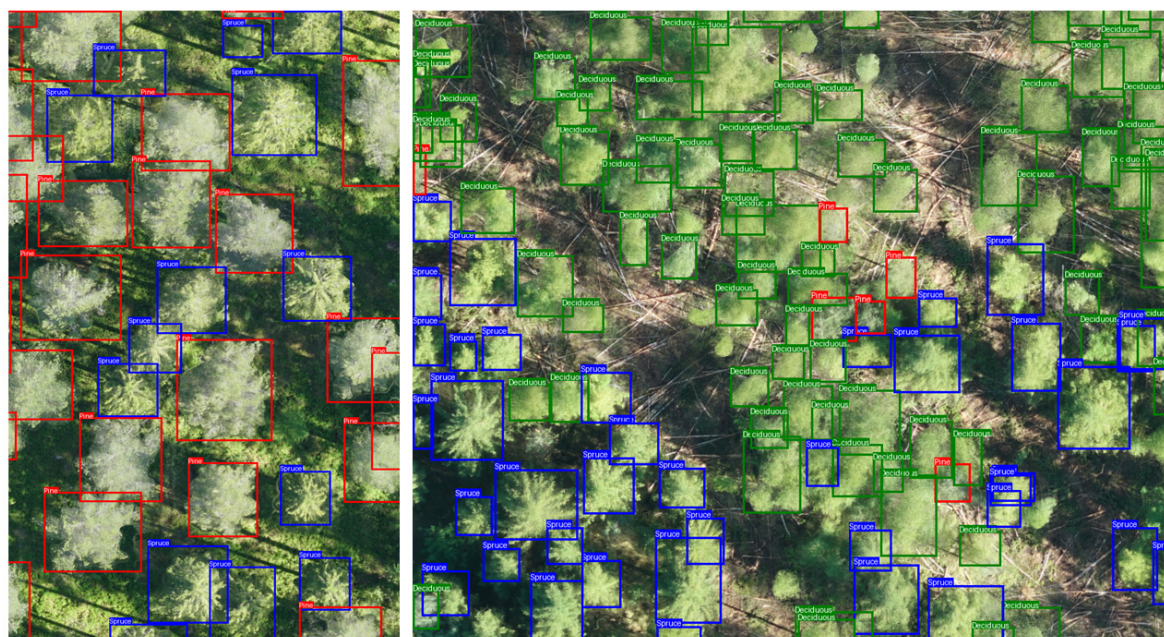


Figure A4. The bounding box visualization output from YOLOv12.

References

1. Liu, M.; Han, Z.; Chen, Y.; Liu, Z.; Han, Y. Tree species classification of airborne LiDAR data based on 3D deep learning. *Guofang Keji Daxue Xuebao/Journal Natl. Univ. Def. Technol.* **2022**, *44*, 123–130, doi:10.11887/j.cn.202202016.
2. Shang, X.; Chisholm, L.A. Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2481–2489, doi:10.1109/JSTARS.2013.2282166.
3. Weinstein, B.G.; Marconi, S.; Bohlman, S.A.; Zare, A.; White, E.P. Cross-site learning in deep learning RGB tree crown detection. *Ecol. Inform.* **2020**, *56*, 101061, doi:10.1016/j.ecoinf.2020.101061.
4. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2016; Vol. 2016-Decem, pp. 779–788.
5. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327, doi:10.1109/TPAMI.2018.2858826.
6. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2014; pp. 580–587.
7. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149, doi:10.1109/TPAMI.2016.2577031.
8. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the Proceedings of the IEEE international conference on computer vision; 2017; pp. 2961–2969.
9. Ferreira, M.P.; Almeida, D.R.A. de; Papa, D. de A.; Minervino, J.B.S.; Veras, H.F.P.; Formighieri, A.; Santos, C.A.N.; Ferreira, M.A.D.; Figueiredo, E.O.; Ferreira, E.J.L. Individual tree detection and species classification of Amazonian palms using UAV images and deep learning. *For. Ecol. Manage.* **2020**, *475*, 118397, doi:10.1016/j.foreco.2020.118397.
10. Hao, Z.; Lin, L.; Post, C.J.; Mikhailova, E.A.; Li, M.; Chen, Y.; Yu, K.; Liu, J. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 112–123, doi:10.1016/j.isprsjprs.2021.06.003.

11. Yang, M.; Mou, Y.; Liu, S.; Meng, Y.; Liu, Z.; Li, P.; Xiang, W.; Zhou, X.; Peng, C. Detecting and mapping tree crowns based on convolutional neural network and Google Earth images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, doi:10.1016/j.jag.2022.102764.
12. Beloiu, M.; Heinzmann, L.; Rehush, N.; Gessler, A.; Griess, V.C. Individual Tree-Crown Detection and Species Identification in Heterogeneous Forests Using Aerial RGB Imagery and Deep Learning. *Remote Sens.* **2023**, *15*, doi:10.3390/rs15051463.
13. Korznikov, K.; Kislov, D.; Petrenko, T.; Dzizyurova, V.; Doležal, J.; Krestov, P.; Altman, J. Unveiling the Potential of Drone-Borne Optical Imagery in Forest Ecology: A Study on the Recognition and Mapping of Two Evergreen Coniferous Species. *Remote Sens.* **2023**, *15*, doi:10.3390/rs15184394.
14. Gan, Y.; Wang, Q.; Iio, A. Tree Crown Detection and Delineation in a Temperate Deciduous Forest from UAV RGB Imagery Using Deep Learning Approaches: Effects of Spatial Resolution and Species Characteristics. *Remote Sens.* **2023**, *15*, doi:10.3390/rs15030778.
15. Zhong, H.; Zhang, Z.; Liu, H.; Wu, J.; Lin, W. Individual Tree Species Identification for Complex Coniferous and Broad-Leaved Mixed Forests Based on Deep Learning Combined with UAV LiDAR Data and RGB Images. *Forests* **2024**, *15*, 293.
16. Santos, A.A. dos; Marcato Junior, J.; Araújo, M.S.; Di Martini, D.R.; Tetila, E.C.; Siqueira, H.L.; Aoki, C.; Eltner, A.; Matsubara, E.T.; Pistori, H. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs. *Sensors* **2019**, *19*, 3595.
17. Itakura, K.; Hosoi, F. Automatic tree detection from three-dimensional images reconstructed from 360 spherical camera using YOLO v2. *Remote Sens.* **2020**, *12*, 988, doi:10.3390/rs12060988.
18. Safonova, A.; Hamad, Y.; Alekhina, A.; Kaplun, D. Detection of Norway Spruce Trees (*Picea Abies*) Infested by Bark Beetle in UAV Images Using YOLOs Architectures. *IEEE Access* **2022**, *10*, 10384–10392, doi:10.1109/ACCESS.2022.3144433.
19. Liu, Y.; Zhao, Q.; Wang, X.; Sheng, Y.; Tian, W.; Ren, Y. A tree species classification model based on improved YOLOv7 for shelterbelts. *Front. Plant Sci.* **2024**, *14*, 1265025, doi:10.3389/fpls.2023.1265025.
20. Dong, C.; Cai, C.; Chen, S.; Xu, H.; Yang, L.; Ji, J.; Huang, S.; Hung, I.K.; Weng, Y.; Lou, X. Crown Width Extraction of Metasequoia glyptostroboides Using Improved YOLOv7 Based on UAV Images. *Drones* **2023**, *7*, 336, doi:10.3390/drones7060336.
21. Xu, S.; Wang, R.; Shi, W.; Wang, X. Classification of Tree Species in Transmission Line Corridors Based on YOLO v7. *Forests* **2024**, *15*, 61, doi:10.3390/f15010061.
22. Jarahizadeh, S.; Salehi, B. Advancing tree detection in forest environments: A deep learning object detector approach with UAV LiDAR data. *Urban For. Urban Green.* **2025**, *105*, 128695, doi:10.1016/j.ufug.2025.128695.
23. Tian, Y.; Ye, Q.; Doermann, D. YOLOv12: Attention-Centric Real-Time Object Detectors. *arXiv:2502.12524* **2025**, doi:10.48550/arXiv.2502.12524.
24. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*; 2017; Vol. 39, pp. 640–651.
25. Schiefer, F.; Kattenborn, T.; Frick, A.; Frey, J.; Schall, P.; Koch, B.; Schmidtlein, S. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2020**, *170*, 205–215, doi:10.1016/j.isprsjprs.2020.10.015.
26. Gyawali, A.; Adhikari, H.; Aalto, M.; Ranta, T. From simple linear regression to machine learning methods: Canopy cover modelling of a young forest using planet data. *Ecol. Inform.* **2024**, *82*, 102706, doi:10.1016/j.ecoinf.2024.102706.
27. Morais, T.G.; Domingos, T.; Teixeira, R.F.M. Semantic Segmentation of Portuguese Agri-Forestry Using High-Resolution Orthophotos. *Agronomy* **2023**, *13*, 2741, doi:10.3390/agronomy13112741.
28. Kemker, R.; Salvaggio, C.; Kanan, C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 60–77, doi:10.1016/j.isprsjprs.2018.04.014.
29. Haq, M.A.; Rahaman, G.; Baral, P.; Ghosh, A. Deep Learning Based Supervised Image Classification Using UAV Images for Forest Areas Classification. *J. Indian Soc. Remote Sens.* **2021**, *49*, 601–606, doi:10.1007/s12524-020-01231-3.

30. Hizal, C.; Gülsu, G.; Akgün, H.Y.; Kulavuz, B.; Bakırman, T.; Aydın, A.; Bayram, B. Forest Semantic Segmentation Based on Deep Learning Using Sentinel-2 Images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* **2024**, *48*, 229–236, doi:10.5194/isprs-archives-XLVIII-4-W9-2024-229-2024.
31. You, H.; Huang, Y.; Qin, Z.; Chen, J.; Liu, Y. Forest Tree Species Classification Based on Sentinel-2 Images and Auxiliary Data. *Forests* **2022**, *13*, 1416, doi:10.3390/f13091416.
32. Ma, Y.; Zhao, Y.; Im, J.; Zhao, Y.; Zhen, Z. A deep-learning-based tree species classification for natural secondary forests using unmanned aerial vehicle hyperspectral images and LiDAR. *Ecol. Indic.* **2024**, *159*, 111608, doi:10.1016/j.ecolind.2024.111608.
33. Zagajewski, B.; Kluczek, M.; Raczko, E.; Njegovec, A.; Dabija, A.; Kycko, M. Comparison of random forest, support vector machines, and neural networks for post-disaster forest species mapping of the krkonoše/karkonosze transboundary biosphere reserve. *Remote Sens.* **2021**, *13*, 2581, doi:10.3390/rs13132581.
34. Raczko, E.; Zagajewski, B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *Eur. J. Remote Sens.* **2017**, *50*, 144–154, doi:10.1080/22797254.2017.1299557.
35. Immitzer, M.; Atzberger, C.; Koukal, T. Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. *Remote Sens.* **2012**, *4*, 2661–2693, doi:10.3390/rs4092661.
36. Yan, S.; Jing, L.; Wang, H. A new individual tree species recognition method based on a convolutional neural network and high-spatial resolution remote sensing imagery. *Remote Sens.* **2021**, *13*, 479.
37. Los, H.; Mendes, G.S.; Cordeiro, D.; Grosso, N.; Costa, H.; Benevides, P.; Caetano, M. Evaluation of Xgboost and Lgbm Performance in Tree Species Classification With Sentinel-2 Data. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS); IEEE, 2021; Vol. 2021-July, pp. 5803–5806.
38. Vanguri, R.; Laneve, G.; Hościło, A. Mapping forest tree species and its biodiversity using EnMAP hyperspectral data along with Sentinel-2 temporal data: An approach of tree species classification and diversity indices. *Ecol. Indic.* **2024**, *167*, 112671, doi:10.1016/j.ecolind.2024.112671.
39. Usman, M.; Ejaz, M.; Nichol, J.E.; Farid, M.S.; Abbas, S.; Khan, M.H. A Comparison of Machine Learning Models for Mapping Tree Species Using WorldView-2 Imagery in the Agroforestry Landscape of West Africa. *ISPRS Int. J. Geo-Information* **2023**, *12*, 142, doi:10.3390/ijgi12040142.
40. Qin, H.; Zhou, W.; Yao, Y.; Wang, W. Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data. *Remote Sens. Environ.* **2022**, *280*, 113143, doi:10.1016/j.rse.2022.113143.
41. Sun, Y.; Huang, J.; Ao, Z.; Lao, D.; Xin, Q. Deep learning approaches for the mapping of tree species diversity in a tropical wetland using airborne LiDAR and high-spatial-resolution remote sensing images. *Forests* **2019**, *10*, 1047.
42. Xu, Z.; Shen, X.; Cao, L.; Coops, N.C.; Goodbody, T.R.H.; Zhong, T.; Zhao, W.; Sun, Q.; Ba, S.; Zhang, Z.; et al. Tree species classification using UAS-based digital aerial photogrammetry point clouds and multispectral imageries in subtropical natural forests. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102173, doi:10.1016/j.jag.2020.102173.
43. Quan, Y.; Li, M.; Hao, Y.; Liu, J.; Wang, B. Tree species classification in a typical natural secondary forest using UAV-borne LiDAR and hyperspectral data. *GIScience Remote Sens.* **2023**, *60*, doi:10.1080/15481603.2023.2171706.
44. Planet Team Planet Application Program Interface; Space for Life on Earth: San Francisco, CA, USA Available online: <https://api.planet.com> (accessed on Sep 20, 2022).
45. Kluczek, M.; Zagajewski, B.; Zwijacz-Kozica, T. Mountain Tree Species Mapping Using Sentinel-2, PlanetScope, and Airborne HySpex Hyperspectral Imagery. *Remote Sens.* **2023**, *15*, 844, doi:10.3390/rs15030844.
46. Hovi, A.; Raitio, P.; Rautiainen, M. A spectral analysis of 25 boreal tree species. *Silva Fenn.* **2017**, *51*, 7753, doi:10.14214/sf.7753.
47. FMI Finnish Meteorological Institute (FMI) Available online: <https://en.ilmatieteenlaitos.fi/> (accessed on Jan 15, 2025).

48. Gyawali, A.; Aalto, M.; Peuhkurinen, J.; Villikka, M.; Ranta, T. Comparison of Individual Tree Height Estimated from LiDAR and Digital Aerial Photogrammetry in Young Forests. *Sustainability* **2022**, *14*, 3720, doi:10.3390/su14073720.
49. Pix4D S.A. Prilly, S. Pix4D Drone Mapping Software. Swiss Fed Inst Technol Lausanne, Route Cantonale, Switz 2014. Available online: <http://pix4d.com> (accessed on May 25, 2021).
50. Woebbecke, D.M.; Meyer, G.E.; Von Bargen, K.; Mortensen, D.A. Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. Am. Soc. Agric. Eng.* **1995**, *38*, 259–269, doi:10.13031/2013.27838.
51. Louhaichi, M.; Borman, M.M.; Johnson, D.E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto Int.* **2001**, *16*, 65–70, doi:10.1080/10106040108542184.
52. Bendig, J.; Yu, K.; Aasen, H.; Bolten, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.L.; Bareth, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 79–87, doi:10.1016/j.jag.2015.02.012.
53. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87, doi:10.1016/S0034-4257(01)00289-9.
54. Gitelson, A.A.; Vina, A.; Arkebauer, T.J.; Rundquist, D.C.; Keydan, G.; Leavitt, B. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* **2003**, *30*, doi:10.1029/2002gl016450.
55. Huete, A.R.; Liu, H.Q.; Batchily, K.; Van Leeuwen, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens. Environ.* **1997**, *59*, 440–451, doi:10.1016/S0034-4257(96)00112-5.
56. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213, doi:10.1016/S0034-4257(02)00096-2.
57. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298, doi:10.1016/S0034-4257(96)00072-7.
58. Yang, W.; Kobayashi, H.; Wang, C.; Shen, M.; Chen, J.; Matsushita, B.; Tang, Y.; Kim, Y.; Bret-Harte, M.S.; Zona, D.; et al. A semi-analytical snow-free vegetation index for improving estimation of plant phenology in tundra and grassland ecosystems. *Remote Sens. Environ.* **2019**, *228*, 31–44, doi:10.1016/j.rse.2019.03.028.
59. Rouse, R.W.H.; Haas, J.A.W.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS (Earth Resources Technology Satellites). *Goddard Sp. Flight Cent. 3d ERTS-1* **1974**, *1*, 309–317.
60. ESRI. Train Deep Learning Model (Image Analyst) Available online: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/train-deep-learning-model.htm> (accessed on Sep 10, 2024).
61. Breiman, L. Random forests. *Random Forests*, 1–122. *Mach. Learn.* **2001**, *45*, 5–32.
62. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. Catboost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *2018-Decem*, 6638–6648.
63. Odeh, A.; Al-Haija, Q.A.; Aref, A.; Taleb, A.A. Comparative Study of CatBoost, XGBoost, and LightGBM for Enhanced URL Phishing Detection: A Performance Assessment. *J. Internet Serv. Inf. Secur.* **2023**, *13*, 1–11, doi:10.58346/JISIS.2023.I4.001.
64. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 1–74, doi:10.1186/s40537-021-00444-8.
65. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In *Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer, 2016; Vol. 9908 LNCS, pp. 630–645.
66. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46, doi:10.1016/0034-4257(91)90048-B.
67. Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *BBA - Protein Struct.* **1975**, *405*, 442–451, doi:10.1016/0005-2795(75)90109-9.
68. Shen, L.; Lang, B.; Song, Z. DS-YOLOv8-Based Object Detection Method for Remote Sensing Images. *IEEE Access* **2023**, *11*, 125122–125137, doi:10.1109/ACCESS.2023.3330844.

69. Avtar, R.; Chen, X.; Fu, J.; Alsulamy, S.; Supe, H.; Pulpadan, Y.A.; Louw, A.S.; Tatsuro, N. Tree Species Classification by Multi-Season Collected UAV Imagery in a Mixed Cool-Temperate Mountain Forest. *Remote Sens.* **2024**, *16*, 4060, doi:10.3390/rs16214060.
70. Sothe, C.; De Almeida, C.M.; Schimalski, M.B.; La Rosa, L.E.C.; Castro, J.D.B.; Feitosa, R.Q.; Dalponte, M.; Lima, C.L.; Liesenberg, V.; Miyoshi, G.T.; et al. Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *GIScience Remote Sens.* **2020**, *57*, 369–394, doi:10.1080/15481603.2020.1712102.
71. Sothe, C.; La Rosa, L.E.C.; De Almeida, C.M.; Gonsamo, A.; Schimalski, M.B.; Castro, J.D.B.; Feitosa, R.Q.; Dalponte, M.; Lima, C.L.; Liesenberg, V.; et al. Evaluating a convolutional neural network for feature extraction and tree species classification using uav-hyperspectral images. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *5*, 193–199, doi:10.5194/isprs-Annals-V-3-2020-193-2020.
72. Maschler, J.; Atzberger, C.; Immitzer, M. Individual tree crown segmentation and classification of 13 tree species using Airborne hyperspectral data. *Remote Sens.* **2018**, *10*, 1218, doi:10.3390/rs10081218.
73. Yao, Y.; Wang, X.; Qin, H.; Wang, W.; Zhou, W. Mapping Urban Tree Species by Integrating Canopy Height Model with Multi-Temporal Sentinel-2 Data. *Remote Sens.* **2025**, *17*, 790, doi:10.3390/rs17050790.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.