



Article

Tree Species Classification from UAV Canopy Images with Deep Learning Models

Yunmei Huang ¹, Botong Ou ², Kexin Meng ², Baijian Yang ², Joshua Carpenter ³, Jinha Jung ³ and Songlin Fei ^{1,*}

¹ Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907, USA; huan1643@purdue.edu

² Department of Computer and Information Technology, Purdue University, West Lafayette, IN 47907, USA; bou@purdue.edu (B.O.); meng147@purdue.edu (K.M.); byang@purdue.edu (B.Y.)

³ Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA; jcarpene@purdue.edu (J.C.); jinha@purdue.edu (J.J.)

* Correspondence: sfei@purdue.edu

Abstract: Forests play a critical role in the provision of ecosystem services, and understanding their compositions, especially tree species, is essential for effective ecosystem management and conservation. However, identifying tree species is challenging and time-consuming. Recently, unmanned aerial vehicles (UAVs) equipped with various sensors have emerged as a promising technology for species identification due to their relatively low cost and high spatial and temporal resolutions. Moreover, the advancement of various deep learning models makes remote sensing based species identification more a reality. However, three questions remain to be answered: first, which of the state-of-the-art models performs best for this task; second, which is the optimal season for tree species classification in a temperate forest; and third, whether a model trained in one season can be effectively transferred to another season. To address these questions, we focus on tree species classification by using five state-of-the-art deep learning models on UAV-based RGB images, and we explored the model transferability between seasons. Utilizing UAV images taken in the summer and fall, we captured 8799 crown images of eight species. We trained five models using summer and fall images and compared their performance on the same dataset. All models achieved high performances in species classification, with the best performance on summer images, with an average F1-score was 0.96. For the fall images, Vision Transformer (ViT), EfficientNetB0, and YOLOv5 achieved F1-scores greater than 0.9, outperforming both ResNet18 and DenseNet. On average, across the two seasons, ViT achieved the best accuracy. This study demonstrates the capability of deep learning models in forest inventory, particularly for tree species classification. While the choice of certain models may not significantly affect performance when using summer images, the advanced models prove to be a better choice for fall images. Given the limited transferability from one season to another, further research is required to overcome the challenge associated with transferability across seasons.

Keywords: species classification; UAV; RGB; deep learning; temperate forest; supervised learning



Citation: Huang, Y.; Ou, B.; Meng, K.; Yang, B.; Carpenter, J.; Jung, J.; Fei, S. Tree Species Classification from UAV Canopy Images with Deep Learning Models. *Remote Sens.* **2024**, *16*, 3836. <https://doi.org/10.3390/rs16203836>

Academic Editor: Jie Shao

Received: 30 August 2024

Revised: 2 October 2024

Accepted: 11 October 2024

Published: 15 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forest, as a significant component of terrestrial ecosystems, provides critical ecosystem services such as food, water, and wood products. Understanding forest compositions, especially regarding tree species, is essential for forest management and conservation. Recently, machine learning algorithms have advanced rapidly. In light of this, increasing attention is being paid to utilizing them with various data to identify species [1]. Due to its effectiveness in cost and labor, remote sensing data such as hyperspectral, multispectral, and RGB images are commonly used in tree species classification. Using hyperspectral and multispectral images to identify tree species can achieve accuracies of about 70–90%, depending on the number of species and type of forests studied [2–10]. Despite their

effectiveness, the collection and processing of high resolution hyperspectral and multi-spectral images are costly and require spectral selection. In contrast, the cost of acquiring RGB images using unmanned aerial vehicles (UAVs) has significantly decreased in recent years due to improvements in consumer-grade UAV technology. UAV-based RGB sensors have gained increasing attention and applying these low-cost RGB sensors can address the practical challenges of data collection in resource-constrained situations, where access to hyperspectral and multispectral sensors may not always be feasible. High spatial resolution RGB images allow for a detailed analysis of individual trees and for the identification of subtle morphological features, such as shapes of the canopy and leaves, or colors of the flowers. As a result, UAV-based RGB images have become a more attractive option for identifying tree species, quantifying tree populations, and other forestry applications [11].

To identify tree species, many algorithms have been developed. Applications of deep learning models on UAV-based RGB images have become the most promising and practical methods to automate this procedure [12]. Four types of tasks have been explored for tree species recognition with deep learning models including classification, detection, semantic segmentation, and instance segmentation. Tree species classification is used to determine what species are present in an image. Tree detection focuses on localizing trees or determining what species are detected. Semantic segmentation aims to detect trees or species at the pixel level and instance segmentation of forests can achieve segmenting each individual tree with species information. For tree species classification applications, the Residual network (ResNet) and its variants are most commonly used [13–16]. Other CNN-based models, such as DenseNet, MaskRCNN, DeepLab++, GoogLeNet, and EfficientNets, are also popularly applied [11,14,16–21]. For tree detection, several models that detect or localize individual canopies with bounding boxes are applied, including YOLOs, FasterRCNN, and RetainNet [22–25]. To conduct semantic segmentation of tree canopy images, U-net, MaskRCNN, and DeepLab+ have been applied to detect live trees and windthrows at the pixel level [26,27]. Some studies use self-crafted models for their needs, which can offer more flexibility but require an extensive dataset to train these models from scratch [28,29].

Previous studies have demonstrated the potential of applying deep learning algorithms and UAVs for tree species identification. We reviewed 11 recent studies that focused on the use of UAV-based RGB images for tree species recognition with deep learning models. In Table 1, we included the type of task, location, number of species, time of data acquisition, model(s) applied, performance of model(s), and ground sample distance(s) (GSD). These studies investigated different deep learning models (all CNN-based models) on images collected from different seasons and regions including tropical, temperate, and boreal forests. Data were collected at different altitudes (varying from 30 to 150 m), resulting in GSDs ranging from 0.82 to 15.00 cm. These studies reported overall accuracy varying from 40% to 92% depending on their images GSD, spatial coverage, number of species included, and complexity of the forests. However, there is still a lack of comparisons of the accuracy of species classification among different models, and model accuracy in temperate forests is relatively low. Notably, the deep neural networks employed have been limited to CNN-based models [30], with few investigations of transformer-based models [30]. Transformer-based models such as the Vision Transformer (ViT) [31] have demonstrated a powerful and promising performance in solving computer vision tasks; however, there is limited discussion on selecting a particular deep learning model over others in species classification. CNN-based models, which rely on convolutional layers to extract hierarchical features, and process features layer by layer. Transformer-based models leverage the multi-head attention mechanism to learn globally and understand spatial relationships between all parts of the images without relying on filters or localized kernels [32].

Although efforts have been made to build more robust deep learning models for tree species classification by using images from multiple seasons, there are only a few studies that trained and tested their models with images from different times and sites with RGB

data; and they found that model accuracy significantly decreased and the transferability of models was low [15,20]. The transferability of a machine learning model reflects its ability to generalize knowledge from existing data and reuse such knowledge in unseen data [33]. By exploring models' transferability among seasons, we can evaluate the adaptability and generalizability of a model on images acquired from different seasons, and improve model performance by choosing the optimal time windows for data acquisition [34]. However, few studies have examined the impacts of seasonal variations on species classification using multi-season UAV RGB images for temperate forests and the model's transferability across seasons. Despite the progress in applying RGB images and deep learning models in tree species classification, several issues remain. First, model accuracy for temperate forests is relatively low, and deep learning methods employed thus far have been limited to CNN-based models, lacking investigation of transformer-based models. Second, the impact of image acquisition time on species classification and model transferability are underexplored.

To address these issues, this study focuses on the following main research questions: (1) What is the comparative performance of various deep learning models on the same dataset for tree species classification? (2) Does the seasonal variation in tree canopy features impact the accuracy of species classification in temperate forests? (3) Can deep learning models trained on a one season be generalized and transferable to another season?

Through these research inquiries, this study aims to advance the current understanding of the applicability and effectiveness of deep learning models for tree species classification. We thoroughly compared the state-of-the-art AI models, built a dataset across two seasons, compared the accuracy of five state-of-the-art deep learning models, and tested transferability across seasons. Thus, our findings can provide insights into the application of state-of-the-art deep learning models on species classifications across different seasons.

Table 1. Related work using UAV RGB images and deep learning algorithms for tree species identification. This study's results are included at the bottom.

Author	Task Types *	Location	#Species	Acquisition Time	Models	Results	GSD (cm/pixel)
Karrenborn et al. (2019) [35]	SS	Chile	2	Fall	U-net	84%	3
Natesan et al. (2019) [15]	CL	Canada	2	Summer	ResNet50	80% on two pines	1, 2, 4
Santos et al. (2019) [24]	DT	Brazil	1	Winter, Spring, Summer	Faster RCNN, YOLOv3, RetinaNet	Urban tree 92%	0.82
Natesan et al. (2020) [19]	CL	Canada	5	Summer, Fall	DenseNet	5 coniferous trees 84%	2.5
Ferreira et al. (2020) [36]	SS	Brazil	3	Summer	ResNet18 in DeepLabv3+	3 species of Palm trees 78.6–96.6%	4
Schiefer et al. (2020) [11]	SS	Germany	9	Fall, Winter	U-net	Average F1-score 0.73	2
Osco et al. (2021) [29]	DT	Brazil	2	Summer, Fall	CNN	Plantation detection and counting 87.6%	1.55–2.28
Martins et al. (2021) [37]	SS	Brazil	9	Spring	DeepLabv3+, ResNet	F1-score of 0.79 on urban trees	15
Veras et al. (2022) [16]	SS	Brazil	8	Summer, Fall, Winter, Spring	ResNet, DeepLab	90.50%	4
Onishi et al. (2022) [20]	SS	Japan	58	Summer, Fall	EfficientNet B7	Kappa: 0.97 and 0.72, species 0.47 on 26 species	2.74
Wang et al. (2023) [38]	CL	China	5	Summer	DenseNetBL	Overall accuracy 0.90	5
This study	CL	United States	8	Summer, Fall	ResNet18, DenseNet, EfficientNetB0, Vision Transformer, YOLOv5	Average F1-scores 0.93	1.51–2.01, 0.77

* SS = Semantic segmentation; CL = Classification; DT = Detection. #Species: Number of species.

2. Methods

The pipeline for this study consists of five main steps: data acquisition, label generation with ground information and visual interpretation, model exploration, model training, and model's transferability experiments (Figure 1).

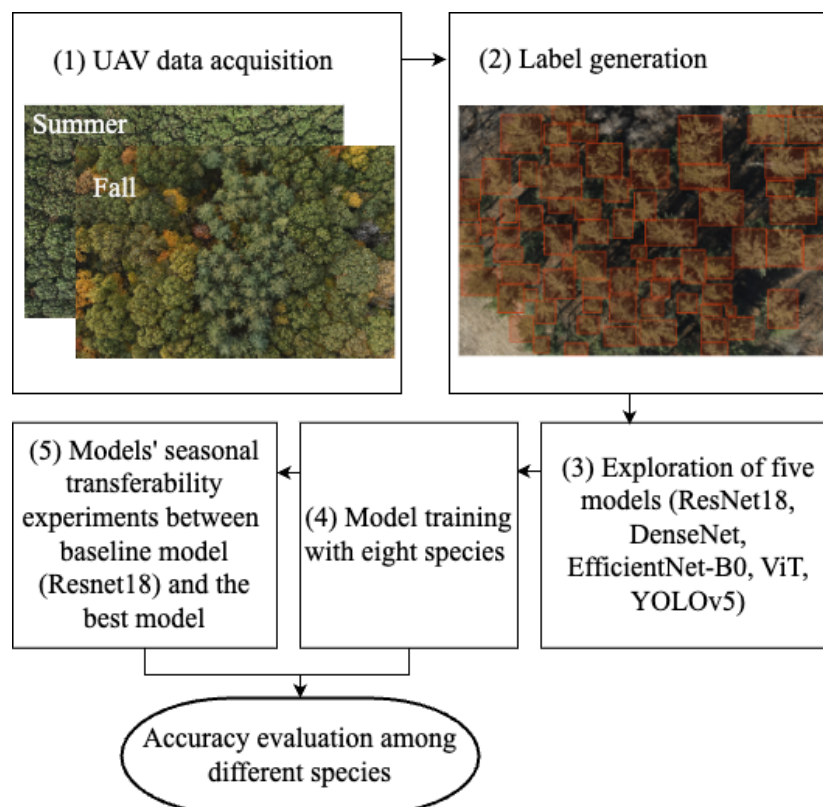


Figure 1. Work pipeline for tree species classification with UAV images and deep learning models.

2.1. Data Acquisition

The study area, Martell Forest, located in Tippecanoe County, Indiana, USA, is a research forest owned by Purdue University that contains monoculture plantations, mixed plantations, and natural forests (Figure 2a). We collected UAV images using the DJI ZENMUSE P1 (35 mm) camera, flying the drone with 85% overlap and 85% sidelap for two flights in August and November 2021 (Table 2). Our flights were programmed at a height of 120 m above the ground. The plots we covered had about 30 m differences in altitude due to the sloped terrain, resulting in a GSD of approximately 1.5 cm to 2.01 cm. We focused on eight ecologically and economically important tree species from plantations (Table 3). Black cherry (*Prunus serotina*), northern red oak (*Quercus rubra*), red pine (*Pinus resinosa*), black walnut (*Juglans nigra*), and white oak (*Q. alba*) were primarily located at the top of the slope, while butternut (*J. cinerea*), American chestnut (*Castanea dentata*), and white pine (*P. strobus*), were located near the valley at the Martell Forest. The flights were carried out during summer and fall to capture seasonal variations in the appearance of the tree species, particularly in terms of foliage color change. These two seasons were selected as they represent critical phenological stages. The summer images included full leaf development and the fall images captured senescence and leaf color change.

Table 2. Flight information, including date, flight settings, and brief description for images.

Dataset	Date	Flight Information	Description
Summer	18 August 2021	Altitude at 120 m, sidelap 85%, overlap 85%, DJI ZENMUSE P1	Fully green canopies
Fall	2 November 2021	Same as above	Mixed canopies with different coloration, leaf-on and off

Table 3. Labeled number of images for eight tree species over two seasons.

Common Name	Species Name	Summer	Fall
Black cherry	<i>Prunus serotina</i>	585	548
Butternut	<i>Juglans cinerea</i>	358	1002
American chestnut	<i>Castanea dentata</i>	44	121
Northern red oak	<i>Quercus rubra</i>	739	1954
Red pine	<i>Pinus resinosa</i>	192	102
Black walnut	<i>J. nigra</i>	585	1840
White oak	<i>Q. alba</i>	119	96
White pine	<i>P. strobus</i>	200	317
Total		2819	5980

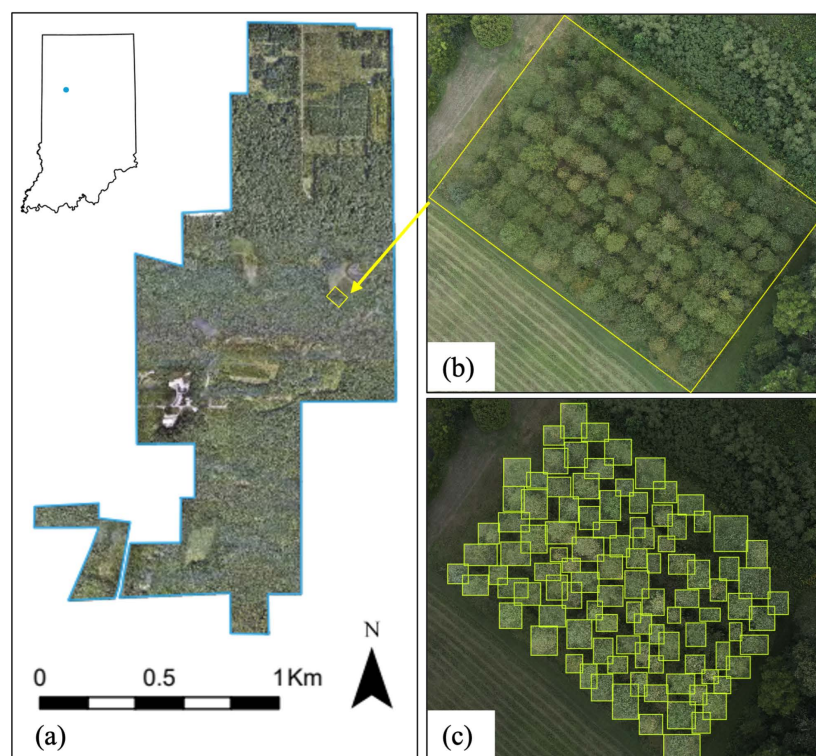


Figure 2. Study area and label examples. (a) Martell Forest in Indiana, USA; (b) Canopy image of a black cherry (*Prunus serotina*) plantation; (c) Label examples of the black cherry plantation (all crowns were identified with bounding boxes).

2.2. Label Generation

With the field record about species information, we used an open-source program, Label Studio [39], to label individual canopies from raw images with a bounding box on each crown. We used the raw images, which have a better quality than orthophotos, to generate training and validation datasets. In Figure 2b,c, we show examples of raw images and label examples in bounding boxes. The reason for using original UAV images is that trees have complicated geometry appearance and structures, making it challenging to overcome the limitations of orthophoto generation without blurring (Figure 3). Due to the different sizes of trees, all of the crown images had various numbers of pixels ranging from 100 to 300. Before training, all of the images were resized to a size of 224×224 .

Visible differences could be observed in the photos taken from the same trees collected in different seasons (Figure 3). Specifically, images that were captured on 18 August 2021 showed fully green canopies. In contrast, images taken on 2 November 2021 showed a range of colors, reflecting the trees' different phenological stages. Among the deciduous species in fall images, both black walnut and black cherry had already shed most of their

leaves, while other species demonstrated varying degrees of leaf coloration and defoliation. The two coniferous species, white pine and red pine, exhibited yellowish coloration on certain leaves in their green crowns in November.

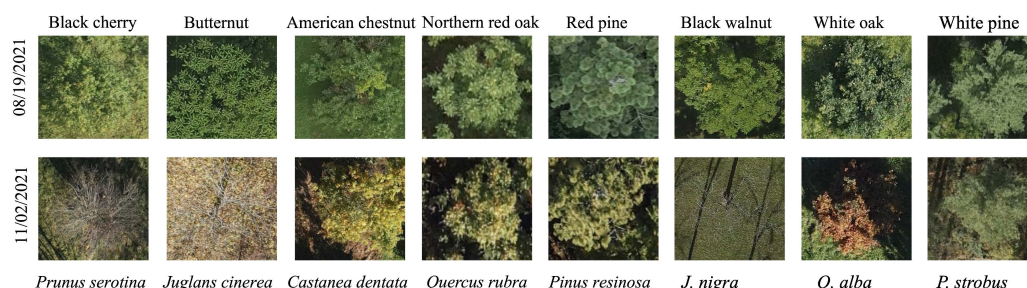


Figure 3. Examples of seasonal differences among eight species. These crown images are cropped from orthophotos of our study area and show the crown variation of the same trees.

We generated 8799 labeled crown images for these eight tree species from the two seasons (Table 3). Different species have different numbers of images depending on the size of the plantations. The dataset represents two seasons that have different numbers of images because the area of different flights covered is diverse and different images are picked. With raw images, we observed the same plantations from different angles so that a small portion of images covered the same plantations from different perspectives, especially for small plantations. Northern red oak, black walnut, and black cherry had more images than others. American chestnut and white oak had the least images, with an average of 82 and 107, respectively.

2.3. Model Explored

To use the best available models in this study, we carefully considered the previous applications of deep learning models in species classification and chose five state-of-the-art models. Two of the models we selected, ResNet18 and Densenet, have been frequently employed in earlier investigations (Table 1). We used ResNet18 as a baseline model. Another model we selected, YOLOv5 [40], was designed for real-time object detection, but can also be adapted for classification tasks. While more advanced versions like YOLOv8 [41] offer improvements for detection tasks, YOLOv5 strikes an effective balance among speed, accuracy, and ease of implementation, making it an ideal candidate for tree classification tasks and future tree detection [42,43]. In addition to these models, we also included EfficientNet-B0, a highly effective CNN model that is less sophisticated than ResNet18, yet still achieves similar accuracy levels. EfficientNet-B0 is known to be effective for applications with limited computational resources. Finally, we also included Vision Transformer (ViT) in our selection. This transformer-based model can learn rich, hierarchical representations of the input data, making it more flexible and adaptable to various tasks. It has been shown to outperform ResNet18 on a sizable dataset and such transformer-based models are the state-of-the-art model for solving image-based problems. This makes it an ideal choice for applications where accuracy is of the utmost importance, as it has been shown to achieve a state-of-the-art performance on a range of computer vision tasks [44,45].

2.4. Model Training Setting

Considering the relatively limited amount of images in the dataset, we chose to fine-tune the pre-trained open-sourced models [46]. ResNet18, DenseNet, EfficientNet, and ViT were pre-trained with ImageNet [47] dataset. YOLOv5 was pre-trained by another benchmark dataset (COCO) [48]. In all of the experiments, the tree images dataset was divided into a training set comprising 80% of the data and a testing set comprising the remaining 20%. The training was performed on a Linux machine with an NVIDIA Tesla T4 GPU (TSMC's fabs, Taichung, Taiwan) and Intel(R) Xeon(R) CPU @ 2.00 GHz with 12 G memory (Intel, Hillsboro, OR, USA). The ResNet18, DenseNet, and YOLOv5 models were

trained for 100 epochs using a batch size of 8. The EfficientNet-B0 model was trained for 100 epochs with a batch size of 16. The ViT model has more hyperparameters and is more computationally intensive. For the training process, ViT was trained for 30 epochs as it converged at five epochs. During training, data augmentation techniques were employed, including random horizontal flip and random-sized crop on the training dataset, and resize and center crop on the test dataset.

2.5. Transferability Experiments

To explore the model's performance across different seasons, we conducted transferability testing experiments. Our key approach was to train a model using one seasonal dataset and then utilize the model to predict species on another seasonal dataset. To evaluate the transferability accuracy, the best models' weights were loaded to predict species and to assess accuracy. We tested the transferability of models using ResNet18 (as the baseline model).

2.6. Model Accuracy Assessment

The performance of each model was evaluated using three accuracy metrics—precision, recall, and F1-score. Precision refers to the proportion of correctly classified positive examples over the total number of images the model classifies as positive. Recall refers to the proportion of correctly classified positive images over the dataset's total number of reference images. In other words, it measures how well the model can identify positive examples. F1-score combines precision and recall, making it a useful metric to evaluate the overall performance of a classification model.

3. Results

3.1. Model Performance by Seasons and Species

All five models achieved F1-scores ranging from 0.84 to 0.99 across seasons, with an average F1-score of 0.93, which demonstrates that the state-of-the-art models have an exceptional performance for tree species classification. ResNet18 and Densenet have similar results (with average F1-scores of 0.87 and 0.89, respectively). The other two CNN-based models, EfficientB0 and YOLOv5, had good results based on the average F1-score (0.98 and 0.92, respectively). ResNet18, as our baseline model, achieved the lowest average F1-scores of 0.87, and, in contrast, ViT model achieved the highest average F1-score of 0.986 (Table 4). Among the data collected in different seasons, all models achieved better performance on the summer dataset than the fall dataset (average F1-score of 0.96 vs. 0.91) (Figure 4).

The results of classification accuracy varied slightly among species. Butternut, northern red oak, red pine, black walnut, and white pine had F1-scores above 0.9 across all models and seasons. Black cherry, American chestnut, and white oak had F1-scores ranging from 0.78 to 0.9. Specifically, northern red oak and white pine had the highest F1-scores across two seasons (Table 4). Black walnut and red pine also attained similar results with high F1-scores in the two seasons. Models on American chestnut had the lowest performances and the least images. The number of images for each class might have affected the accuracy, but the unique features of different species also affected the classification accuracy. These results also matched the observations for the baseline model (see the Figures A1 and A2 in Appendix A).

To sum up, our results demonstrate that summer is the optimal season for species classification in temperate forests. If only fall season images are available, EfficientNet-B0, ViT, and YOLOv5 are better choices than ResNet18 and DenseNet. Within the same training, different species have slightly different results depending on the crown features and the number of images.

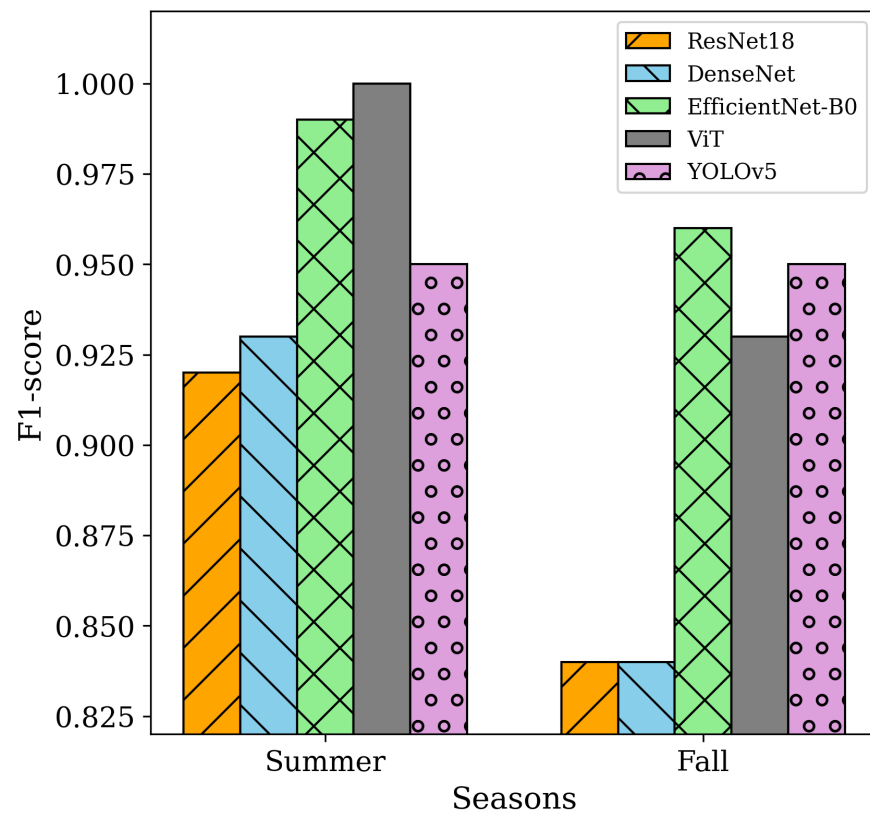


Figure 4. F1-scores of five models for summer and seasons.

Table 4. Five models test accuracy on two datasets. F1-scores under 0.8 are highlighted with background color.

Dataset	ResNet18			DenseNet			EfficientNet-B0			ViT			YOLOv5			All
	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Avg F1*
Summer																
Black cherry	0.97	0.96	0.97	0.99	0.97	0.98	0.99	0.99	0.99	1.00	1.00	1.00	0.98	0.99	0.98	0.98
Butternut	0.80	0.94	0.86	0.88	0.92	0.90	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.99	0.95	0.94
American chestnut	0.83	0.67	0.74	0.88	0.78	0.82	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.89	0.89	0.89
Northern red oak	0.98	0.97	0.98	0.98	0.97	0.97	1.00	0.99	0.996	1.00	1.00	1.00	0.99	1.00	1.00	0.99
Red pine	1.00	0.97	0.99	0.89	1.00	0.94	0.97	0.97	0.97	1.00	1.00	1.00	0.98	1.00	1.00	0.98
Black walnut	0.95	0.97	0.96	0.97	0.96	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98
White oak	0.92	0.92	0.92	0.84	0.88	0.86	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.96	0.96	0.95
White pine	0.97	0.98	0.97	1.00	0.95	0.97	0.97	1.00	0.98	1.00	1.00	1.00	1.00	0.75	0.86	0.96
Average	0.93	0.92	0.92	0.93	0.93	0.93	0.99	0.99	0.99	1.00	1.00	1.00	0.97	0.95	0.95	0.96
Fall																
Black cherry	0.78	0.64	0.70	0.69	0.88	0.77	0.91	0.99	0.95	0.98	0.95	0.97	0.94	0.99	0.96	0.87
Butternut	0.79	0.95	0.86	0.88	0.93	0.90	1.00	0.98	0.99	1.00	1.00	1.00	1.00	0.99	1.00	0.95
American chestnut	0.76	0.92	0.83	0.79	0.79	0.79	0.90	0.95	0.92	1.00	0.95	0.97	0.96	0.92	0.94	0.89
Northern red oak	0.89	0.99	0.94	0.92	0.96	0.94	0.97	0.99	0.98	0.98	0.98	0.98	0.99	1.00	1.00	0.97
Red pine	0.95	1.00	0.98	0.83	0.90	0.86	0.94	0.94	0.94	1.00	0.94	0.97	1.00	0.70	0.82	0.91
Black walnut	0.94	0.79	0.86	0.93	0.79	0.85	1.00	0.96	0.98	0.97	0.99	0.98	1.00	0.95	0.97	0.93
White oak	0.82	0.45	0.58	0.63	0.60	0.62	0.94	1.00	0.97	0.93	0.87	0.90	0.81	0.85	0.83	0.78
White pine	0.97	0.98	0.98	0.97	0.97	0.97	0.98	0.96	0.97	0.98	0.98	0.98	0.77	0.98	0.86	0.95
Average	0.86	0.84	0.84	0.83	0.85	0.84	0.95	0.97	0.96	0.98	0.96	0.97	0.93	0.92	0.92	0.91

* Prec = Precision, Rec = Recall, F1 = F1-score, Avg F1 = Average F1-score.

3.2. Transferability of Models across Two Seasons

As shown in Table 5, models trained on one season can hardly be transferred to another season, with an overall transferability under 0.5. Our results indicate that the models did not all transfer well to each other in the summer and fall seasons overall, but the models could learn some shared features from images and had huge differences among the two seasons and different species. Diving deeper to examine accuracy at the species level, we found that transferability among different species performed differently in two seasons. Taking the results of ResNet18 as an example (Table 6), the summer model could predict species well on fall images of northern red oak (F1-score: 0.84). Meanwhile, the fall images of white pine could be partially classified by the summer model (F1-score: 0.62). As for the transferability accuracy, many classes had large differences in precision and recall. Species with a higher recall and lower precision, such as northern red oak and white pine, tend to be over-classified compared with other species. For these classes with a higher precision and lower recall, models tended to misclassify these species as another species, including black walnut on summer and fall models. Interestingly, the fall model had a low precision and high recall on summer images of white pine. To summarize the results of the transferability experiments, we observed that models could not transfer well between these seasons because of their significant differences in canopy attributes, but the models learned the features of different species like leaves and colors. The models could partially transfer their ability to recognize canopies with continuous features such as white pine and red pine across seasons.

Table 5. Overall accuracy on transferability of models across seasons on ResNet18 and ViT.

Model	Dataset	Summer	Fall
ResNet18	Summer	-	0.40
	Fall	0.17	-
ViT	Summer	-	0.42
	Fall	0.39	-

Table 6. Results of model's transferability trained on ResNet18 across seasons for eight species.

Species	Precision	Recall	F1-Score
Summer model		Fall images	
Black cherry	0.106	0.182	0.134
Butternut	0.238	0.194	0.214
American chestnut	0.053	0.042	0.047
Northern red oak	0.763	0.936	0.840
Red pine	-	-	-
Black walnut	0.308	0.043	0.076
White oak	-	-	-
White pine	0.661	0.578	0.617
Fall model		Summer images	
Black cherry	0.333	0.009	0.017
Butternut	-	-	-
American chestnut	-	-	-
Northern red oak	0.873	0.324	0.473
Red pine	-	-	-
Black walnut	0.042	0.009	0.014
White oak	0.400	0.250	0.308
White pine	0.133	1.000	0.235

4. Discussion

With the rapid advancement of AI models, it is critical to understand and utilize them to solve challenges in ecology and forestry effectively. We conducted a comprehensive

evaluation of these deep learning models on species classification, along with corresponding performance and model transferability between summer and fall seasonal images. Our results outperformed most of the studies, especially working on classification tasks reported in Table 1, by comparing the number of species and accuracy metrics because advanced models and images with lower GSDs were applied. For instance, Mäyrä et al. (2021) [7] achieved an overall accuracy of 0.89 on four species in a Norway boreal forest with hyperspectral and Lidar data. Ferreira et al. (2020) [36] achieved an accuracy of 0.78 for Amazonian palm trees in Brazil. The data acquisition time plays a more important role than the selection of CNN models for classification tasks. Our study reveals that summer is better than fall to acquire data for species classification using the five deep learning models. In terms of model transferability across seasons, it is challenging due to the variations in seasonal features of temperate forests.

4.1. Performance among Different Models

The study findings demonstrate the effectiveness of applying deep learning models in species classification. The five models adopted in this study all achieved great accuracy, especially on the dataset collected during the growing season. EfficientNet-B0 [21] and YOLOv5 [40] had better accuracy than ResNet18 and DenseNet, with fall data. Among these models, the ViT consistently outperformed the others and achieved the highest F1-scores across two seasons due to its ability to capture both local and global features of tree canopies.

The differences in classification accuracy among models can stem from their underlying architectures. ResNet18 is known for its relatively shallow architecture with 18 layers. Its residual connections help mitigate vanishing gradients and make model training more effective. DenseNet achieved a comparable performance in our experiments to ResNet18 [49]. DenseNet with 121 layers utilizes dense connections between layers, through dense blocks, where each layer receives inputs from all previous layers. The dense connectivity helps retain more information across the network, leading to a slightly better performance than ResNet18 on certain species such as American chestnut and butternut. However, both models exhibited difficulty in maintaining a high classification accuracy for certain species in the fall dataset.

YOLOv5 and EfficientNet-B0 achieved a better performance on these two datasets, compared with DenseNet and ResNet18. YOLOv5 is primarily focused on object detection and uses real-time detection. Even though object detection is not the focus of this study, YOLOv5's balance between classification speed and accuracy makes it a strong candidate for integrating tree classification algorithms into handheld devices or drones, enabling efficient and large-scale forest monitoring. EfficientNet-B0, on the other hand, has outperformed other selected CNN-based models. Similarly, Dey et al. (2023) [50] found that EfficientNetV0 outperformed several other CNN models, including ResNets and DenseNets, in identifying swamp forest tree species. Its architecture leverages a compound scaling approach that balances the model's depth, width, and resolution to optimize accuracy while minimizing computational cost, and EfficientNet-B0 is particularly well-suited for future field applications where computational resources are limited [51].

Lastly, compared with the CNN-based models, ViT achieved marginally better results on the two datasets. Our results agree with Bhojanapalli et al. (2021) [52], that using pre-trained ViT can outperform ResNets on a small dataset. Unlike CNNs, which primarily focus on learning local patterns through convolutional filters, ViT uses self-attention mechanisms to analyze entire images and incorporates more global features at lower layers [45]. This capability to learn from complex spatial relationships, both local and global features allowed ViT to outperform CNNs in tree species classification, where the variation in canopy structure and phenological features between seasons can significantly impact classification accuracy [53].

4.2. Seasonal Difference for Species Classification

Temperate forests have more complicated seasonal dynamics than tropical forests, including leaf-on, leaf-off, and different coloration stages. The impact of seasonal variations on temperate forests has mainly been studied through hyperspectral and multispectral images. Modzelewska et al. (2021) [54] found early summer hyperspectral images produced a slightly higher accuracy than late summer and fall for the temperate forest in Poland. Pu et al. (2018) [55] found that late spring images achieved a higher accuracy for tropical and subtropical urban forests than in other months using multispectral images. Hesketh and Sánchez-Azofeifa (2012) [56] analyzed tropical forests in Panama using seasonal spectral images and indicated that noticeable variations imposed by dry and wet seasons on leaf optical properties affected the classification accuracy.

From our results, we observed that summer images with fully green canopies achieved the highest accuracy across all models. Despite having more images in the fall dataset, the models achieved lower accuracies for black cherry and black walnut. For white oak, the model accuracy might have been influenced by the slightly lower number of images (119 in summer and 96 in fall). The phenological variations on canopies during fall might have had an impact on accuracy. When we collected the fall images, these forests were at their fall foliage peak, and different species displayed various phenological attributes. Notably, these three species tend to shed their leaves and transition to fall colors earlier than other deciduous species such as northern red oak. Consequently, their fall images present varying densities of remaining leaves or complete leaf-off. Furthermore, within other species that transition more slowly into fall phenological stages, the colors of their crowns can vary for the same species, and different species from the same genus can exhibit similar colorations, increasing the difficulty of classifying species. For example, northern red oak can display a range of colors from red to orange, even among trees of the same age and located in the same area during fall. Additionally, northern red oak and white oak, being from the same genus, can exhibit very similar canopy coloration. Hence, we can observe that seasonal phenological changes in tree canopies can impact species classification.

4.3. Model Transferability

Exploring a model's transferability from various perspectives, such as timing, location, and resolution, is important as it helps us understand the model's reliability. Our experiments observed that the model's transferability is time-sensitive. Among eight species, black cherry, butternut, and black walnut were almost leaf-offs in the fall dataset. Summer models were trained on images with green canopies and it was difficult to recognize leaf-off images, for which there are new features. Conifers, such as white pine and red pine, have better accuracy than other broadleaf trees for the accuracy of transferability. They also have a high accuracy for the model's transferability. White pine and red pine include more constant features from summer to fall. Our exploration of models' seasonal transferability demonstrates that AI models can learn the critical attributes of canopies, like leaf shapes and colors. Due to the massive changes in the appearances of canopies at various phenological stages for most temperate forest species, species classifications and models' transferability are time-sensitive. This demonstrates the unique and challenging features of trees in comparison with other object classification tasks. We need to explore more factors, particularly timing, that might impact the classification accuracy and models' transferability [34,57]. A deeper understanding of these factors can benefit us by identifying the optimal time for data collection and model application, and enhance our understanding of the models' compatibility and generality across various months or seasons. Additionally, further studies are necessary to explore the impacts of forest features under varying illumination, weather, phenological stages, seasons, and tree sizes. These factors could also have significant impacts on species classifications.

4.4. Limitations and Future Work

One of the major limitations of this study is the imbalanced number of images among different species, which could potentially impact classification accuracy. Future studies should aim to build a balanced dataset to mitigate this problem. Additionally, due to the absence of ground truth data for natural forests, this study focused on plantation images, which may limit generalizability. Furthermore, differences in ground sampling distance (GSD) across images could result in varying species features, introducing uncertainty in classification and transferability performance. Therefore, conducting more comprehensive comparative studies is necessary to determine the suitable range of image GSD for effective species classification. The lack of a general standard for effective data acquisition in this field is notable. For future studies, developing such a protocol would be invaluable in identifying general principles and best practices for UAV data acquisition. Furthermore, applying deep learning models requires substantial data for model training, yet there is no open-source benchmark dataset for this task. Given the time and cost associated with generating reference data for training models, potential strategies to tackle this challenge include developing open-source datasets through crowd-sourcing and standardizing label generation or even automating the procedures. To gain a deeper understanding of the impact of phenological variations from UAV images, further research is required to investigate the seasonal factors that influence tree species classification. Moreover, this study only focused on individual tree classification. For the automatic generation of tree canopy maps, there is a need to integrate tree detection and species classification procedures to achieve individual tree segmentation and accurate species classification.

5. Conclusions

Our study demonstrates the effectiveness of these state-of-the-art deep learning models in tree species classification tasks using UAV-based RGB images. We found summer to be the optimal season for species classification when applying models from our selection. Classification accuracy is further influenced by the attributes unique to each species; notably, the models tend to exhibit a higher performance with two coniferous species, white pine and red pine, compared with other deciduous species. The models trained on specific seasons are sensitive to timing and cannot reliably predict species of images from another season. Tree species with more consistent features across seasons tend to demonstrate better model transferability—coniferous trees, for instance. Deep learning models demonstrate promising results on tree species classification, and further studies are needed to investigate how to improve the transferability of models for broad applications.

Author Contributions: Methodology, Y.H., B.Y. and K.M.; Formal analysis, Y.H.; Investigation, Y.H.; Resources, S.F.; Data curation, B.O., J.C. and J.J.; writing—original draft preparation, Y.H.; Writing—review and editing, S.F.; supervision, S.F. and B.Y.; Funding acquisition, S.F.; Software, B.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the National Institute of Food and Agriculture [grant numbers 2023-68012-38992].

Data Availability Statement: The example dataset from the two seasons can be downloaded at https://github.com/YunmeiHuanghi/Tree_canopy_UAV_image_examples.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

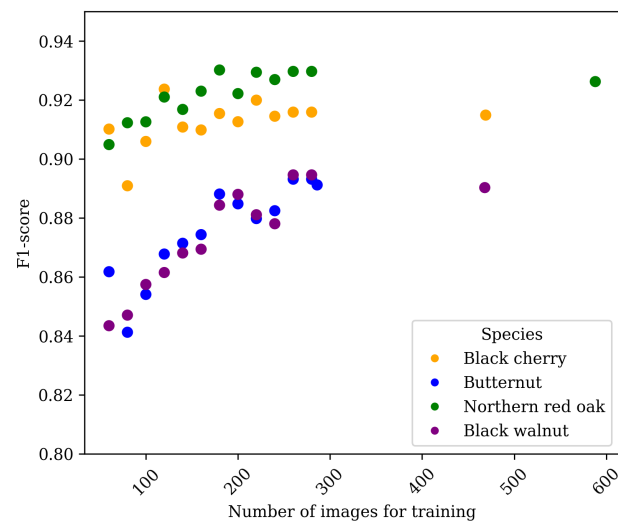


Figure A1. Number of images VS F1-score on summer datasets for four species with ResNet18. Due to the limits of numbers of images, we selected four species and the number of training images starts from 60 to 280, with 20 as an increment. In each training session, all four classes have an equal amount of training images and the same test dataset. Thus, we trained ResNet18 12 times with various numbers of images. When the number of images ranges from 60 to 180, the increment of accuracy is faster than the further part image numbers ranging from 200 to 280. For the experiments on images with the numbers 260 and 280, their change in accuracy was unremarkable. Hence, from our observation, the number of images impacts the model's classification accuracy, and after training images reach a certain amount, the influences decrease.

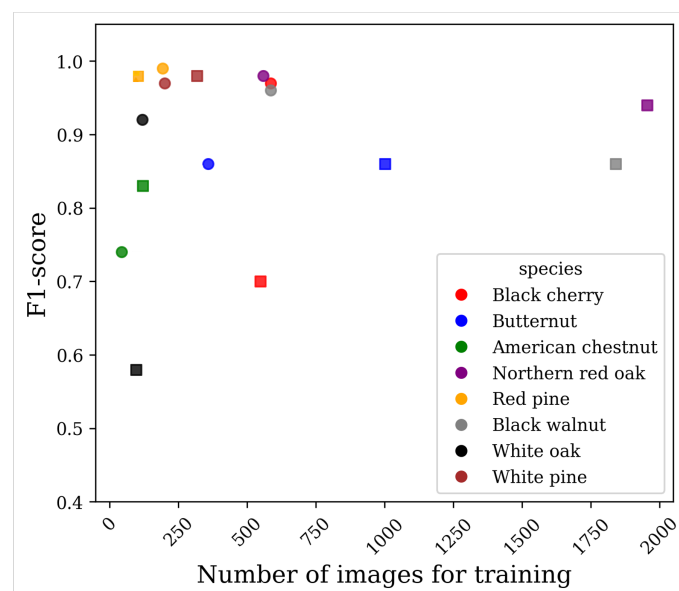


Figure A2. Number of images VS F1-scores with ResNet18 on two datasets for eight species. Different shapes of points stand for different seasons. The round shape points stand for the summer dataset, the squares belong to fall.

References

- Christin, S.; Hervet, É.; Lecomte, N. Applications for deep learning in ecology. *Methods Ecol. Evol.* **2019**, *10*, 1632–1644. [[CrossRef](#)]
- Martin, M.; Newman, S.D.; Aber, J.D.; Congalton, R.G. Determining forest species composition using high spectral resolution remote sensing data. *Remote Sens. Environ.* **1998**, *65*, 249–254. [[CrossRef](#)]

3. Alonzo, M.; Bookhagen, B.; Roberts, D.A. Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sens. Environ.* **2014**, *148*, 70–83. [[CrossRef](#)]
4. Dalponte, M.; Bruzzone, L.; Gianelle, D. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sens. Environ.* **2012**, *123*, 258–270. [[CrossRef](#)]
5. Lu, X.; Liu, G.; Ning, S.; Su, Z.; He, Z. Tree Species Classification based on Airborne Lidar and Hyperspectral Data. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 2787–2790. [[CrossRef](#)]
6. Roffey, M.; Wang, J. Evaluation of Features Derived from High-Resolution Multispectral Imagery and LiDAR Data for Object-Based Support Vector Machine Classification of Tree Species. *Can. J. Remote Sens.* **2020**, *46*, 473–488. [[CrossRef](#)]
7. Mäyrä, J.; Keski-Saari, S.; Kivinen, S.; Tanhuanpää, T.; Hurskainen, P.; Kullberg, P.; Poikolainen, L.; Viinikka, A.; Tuominen, S.; Kumpula, T.; et al. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens. Environ.* **2021**, *256*, 112322. [[CrossRef](#)]
8. 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. [[CrossRef](#)]
9. Liu, X.; Frey, J.; Munteanu, C.; Still, N.; Koch, B. Mapping tree species diversity in temperate montane forests using Sentinel-1 and Sentinel-2 imagery and topography data. *Remote Sens. Environ.* **2023**, *292*, 113576. [[CrossRef](#)]
10. Murray, B.A.; Coops, N.C.; Winiwarer, L.; White, J.C.; Dick, A.; Barbeito, I.; Ragab, A. Estimating tree species composition from airborne laser scanning data using point-based deep learning models. *ISPRS J. Photogramm. Remote Sens.* **2024**, *207*, 282–297. [[CrossRef](#)]
11. Schiefer, F.; Kattenborn, T.; Frick, A.; Frey, J.; Schall, P.; Koch, B.; Schmidlein, 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.
12. Hartling, S.; Sagan, V.; Sidike, P.; Maimaitijiang, M.; Carron, J. Urban tree species classification using a WorldView-2/3 and LiDAR data fusion approach and deep learning. *Sensors* **2019**, *19*, 1284. [[CrossRef](#)] [[PubMed](#)]
13. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
14. Ferreira, M.P.; Lotte, R.G.; D’Elia, F.V.; Stamatopoulos, C.; Kim, D.H.; Benjamin, A.R. Accurate mapping of Brazil nut trees (*Bertholletia excelsa*) in Amazonian forests using WorldView-3 satellite images and convolutional neural networks. *Ecol. Inform.* **2021**, *63*, 101302. [[CrossRef](#)]
15. Natesan, S.; Armenakis, C.; Vepakomma, U. Resnet-based Tree Species Classification using UAV Images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 475–481. [[CrossRef](#)]
16. Veras, H.F.P.; Ferreira, M.P.; da Cunha Neto, E.M.; Figueiredo, E.O.; Dalla Corte, A.P.; Sanquetta, C.R. Fusing multi-season UAS images with convolutional neural networks to map tree species in Amazonian forests. *Ecol. Inform.* **2022**, *71*, 101815. [[CrossRef](#)]
17. Guo, X.; Li, H.; Jing, L.; Wang, P. Individual Tree Species Classification Based on Convolutional Neural Networks and Multitemporal High-Resolution Remote Sensing Images. *Sensors* **2022**, *22*, 3157. [[CrossRef](#)]
18. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
19. Natesan, S.; Armenakis, C.; Vepakomma, U. Individual tree species identification using Dense Convolutional Network (DenseNet) on multitemporal RGB images from UAV. *J. Unmanned Veh. Syst.* **2020**, *8*, 310–333. [[CrossRef](#)]
20. Onishi, M.; Watanabe, S.; Nakashima, T.; Ise, T. Practicality and Robustness of Tree Species Identification Using UAV RGB Image and Deep Learning in Temperate Forest in Japan. *Remote Sens.* **2022**, *14*, 1710. [[CrossRef](#)]
21. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
22. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Montreal, QC, USA, 7–12 December 2015.
24. Santos, A.A.d.; 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.; et al. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs. *Sensors* **2019**, *19*, 3595. [[CrossRef](#)]
25. Beloiu, M.; Heinzmann, L.; Rehus, 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*, 1463. [[CrossRef](#)]
26. 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. [[CrossRef](#)]
27. Reder, S.; Mund, J.P.; Albert, N.; Waßermann, L.; Miranda, L. Detection of Windthrown Tree Stems on UAV-Orthomosaics Using U-Net Convolutional Networks. *Remote Sens.* **2021**, *14*, 75. [[CrossRef](#)]
28. Onishi, M.; Ise, T. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Sci. Rep.* **2021**, *11*, 903. [[CrossRef](#)] [[PubMed](#)]

29. Osco, L.P.; de Arruda, M.d.S.; Gonçalves, D.N.; Dias, A.; Batistoti, J.; de Souza, M.; Gomes, F.D.G.; Ramos, A.P.M.; de Castro Jorge, L.A.; Liesenberg, V.; et al. A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 1–17. [[CrossRef](#)]
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.
31. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
32. Maurício, J.; Domingues, I.; Bernardino, J. Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Appl. Sci.* **2023**, *13*, 5521. [[CrossRef](#)]
33. Jiang, J.; Shu, Y.; Wang, J.; Long, M. Transferability in deep learning: A survey. *arXiv* **2022**, arXiv:2201.05867.
34. Verhulst, M.; Heremans, S.; Blaschko, M.B.; Somers, B. Temporal transferability of tree species classification in temperate forests with Sentinel-2 time series. *Remote Sens.* **2024**, *16*, 2653. [[CrossRef](#)]
35. Kattenborn, T.; Eichel, J.; Fassnacht, F.E. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Sci. Rep.* **2019**, *9*, 17656. [[CrossRef](#)]
36. Ferreira, M.P.; de Almeida, D.R.A.; de Almeida Papa, D.; 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. Manag.* **2020**, *475*, 118397. [[CrossRef](#)]
37. Martins, G.B.; La Rosa, L.E.C.; Happ, P.N.; Coelho Filho, L.C.T.; Santos, C.J.F.; Feitosa, R.Q.; Ferreira, M.P. Deep learning-based tree species mapping in a highly diverse tropical urban setting. *Urban For. Urban Green.* **2021**, *64*, 127241. [[CrossRef](#)]
38. Wang, N.; Pu, T.; Zhang, Y.; Liu, Y.; Zhang, Z. More appropriate DenseNetBL classifier for small sample tree species classification using UAV-based RGB imagery. *Heliyon* **2023**, *9*, e20467. [[CrossRef](#)] [[PubMed](#)]
39. Tkachenko, M.; Malyuk, M.; Holmanyuk, A.; Liubimov, N. Label Studio: Data Labeling Software, 2020–2022. Open Source Software. Available online: <https://github.com/heartexlabs/label-studio> (accessed on 2 October 2022).
40. Jocher, G.; Stoken, A.; Borovec, J.; NanoCode012; ChristopherSTAN; Changyu, L.; Laughing; tkianai; Hogan, A.; lorenzomama; et al. ultralytics/yolov5: V3.1—Bug Fixes and Performance Improvements. Available online: <https://zenodo.org/records/4154370> (accessed on 15 January 2023).
41. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8. 2023. Available online: <https://github.com/ultralytics/ultralytics/blob/main/docs/en/models/yolov8.md> (accessed on 15 May 2023).
42. Masum, M.I.; Sarwat, A.; Riggs, H.; Boymelgreen, A.; Dey, P. YOLOv5 vs. YOLOv8 in Marine Fisheries: Balancing Class Detection and Instance Count. *arXiv* **2024**, arXiv:2405.02312.
43. Hussain, M. YOLOv5, YOLOv8 and YOLOv10: The Go-To Detectors for Real-time Vision. *arXiv* **2024**, arXiv:2407.02988.
44. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [[CrossRef](#)]
45. Raghu, M.; Unterthiner, T.; Kornblith, S.; Zhang, C.; Dosovitskiy, A. Do vision transformers see like convolutional neural networks? *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12116–12128.
46. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [[CrossRef](#)]
47. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [[CrossRef](#)]
48. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:1405.0312.
49. Zhou, Q.; Zhu, W.; Li, F.; Yuan, M.; Zheng, L.; Liu, X. Transfer learning of the ResNet-18 and DenseNet-121 model used to diagnose intracranial hemorrhage in CT scanning. *Curr. Pharm. Des.* **2022**, *28*, 287–295. [[CrossRef](#)]
50. Dey, B.; Ahmed, R.; Ferdous, J.; Haque, M.M.U.; Khatun, R.; Hasan, F.E.; Uddin, S.N. Automated plant species identification from the stomata images using deep neural network: A study of selected mangrove and freshwater swamp forest tree species of Bangladesh. *Ecol. Inform.* **2023**, *75*, 102128. [[CrossRef](#)]
51. Kansal, K.; Chandra, T.B.; Singh, A. ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning. *Procedia Comput. Sci.* **2024**, *235*, 70–80. [[CrossRef](#)]
52. Bhojanapalli, S.; Chakrabarti, A.; Glasner, D.; Li, D.; Unterthiner, T.; Veit, A. Understanding robustness of transformers for image classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10231–10241.
53. Lu, K.; Xu, Y.; Yang, Y. Comparison of the potential between transformer and CNN in image classification. In Proceedings of the ICMLCA 2021—2nd International Conference on Machine Learning and Computer Application, Shenyang, China, 17–19 December 2021; pp. 1–6.
54. Modzelewska, A.; Kamińska, A.; Fassnacht, F.E.; Stereńczak, K. Multitemporal hyperspectral tree species classification in the Białowieża Forest World Heritage site. *For. Int. J. For. Res.* **2021**, *94*, 464–476. [[CrossRef](#)]
55. Pu, R.; Landry, S.; Yu, Q. Assessing the potential of multi-seasonal high resolution Pléiades satellite imagery for mapping urban tree species. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *71*, 144–158.

-
56. Hesketh, M.; Sánchez-Azofeifa, G.A. The effect of seasonal spectral variation on species classification in the Panamanian tropical forest. *Remote Sens. Environ.* **2012**, *118*, 73–82. [[CrossRef](#)]
 57. Hemmerling, J.; Pflugmacher, D.; Hostert, P. Mapping temperate forest tree species using dense Sentinel-2 time series. *Remote Sens. Environ.* **2021**, *267*, 112743. [[CrossRef](#)]

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.