RESNET-BASED TREE SPECIES CLASSIFICATION USING UAV IMAGES

S. Natesan [1, *], C. Armenakis [1], U. Vepakomma [2]

[1] Geomatics Engineering, Department of Earth and Space Science and Engineering, Lassonde School of Engineering, York University, 4700 Keele St., Toronto, ON, M3J 1P3 Canada - (sowmy, armenc)@yorku.ca
[2] FPInnovations, 570 Blvd. Saint-Jean, Pointe-Claire, QC, H9R 3J9 Canada - udayalakshmi.vepakomma@fpinnovations.ca

Commission I, ICWG I/II

KEY WORDS: Tree Species, Classification, UAV, RGB Images, CNN, Deep Learning Networks, ResNet

ABSTRACT:

Tree species classification at individual tree level is a challenging problem in forest management. Deep learning, a cutting-edge technology evolved from Artificial Intelligence, was seen to outperform other techniques when it comes to complex problems such as image classification. In this work, we present a novel method to classify forest tree species through high resolution RGB images acquired with a simple consumer grade camera mounted on a UAV platform using Residual Neural Networks. We used UAV RGB images acquired over three years that varied in numerous acquisition parameters such as season, time, illumination and angle to train the neural network. To begin with, we have experimented with limited data towards the identification of two pine species namely red pine and white pine from the rest of the species. We performed two experiments, first with the images from all three acquisition years and the second with images from only one acquisition year. In the first experiment, we obtained 80% classification accuracy when the trained network was tested on a distinct set of images and in the second experiment, we obtained 51% classification accuracy. As a part of this work, a novel dataset of high-resolution labelled tree species is generated that can be used to conduct further studies involving deep neural networks in forestry.

1. INTRODUCTION

Tree species diversity is an important aspect in the study of forest ecosystems. Applications in conservation and sustainable management of forests such as forest inventories, monitoring of biodiversity, wildlife habitat modelling, hazard management and climate change studies are largely based on tree species classification. Most of the existing methods for tree species classification are constrained by portability, restricted to specific species or requiring large datasets for adaptation to a new site, which limits their applicability and makes them cost-intensive. (Fassnacht et al., 2016) Recently, given the flexibility of acquiring data anytime, anywhere with limited logistics, unmanned aerial vehicles (UAV) are becoming an essential tool in gathering ultra-high resolution imagery on forests for detailed characterization of canopies in contrast to any other higher platform. These have the potential to acquire large datasets at close range needed to train algorithms. Thus, many researches have focused on the use of UAV imagery for tree species classification. For example, Gini et al. (2018) investigated if the use of texture features, derived from UAV multispectral imagery, can improve the accuracy of tree species classification. Franklin and Ahmed (2018) have used UAV multispectral images acquired over forests to successfully classify few tree species by means of a machine-learning classifier which was found effective in separating individual crowns with spectral response, textural, and crown shape variables. UAV images have also been used to separate forest species and dead trees in temperate forest stands (Brovkina et al., 2018).

Recent advancement in deep learning has gained attention for several image classification tasks (Dyrmann et al., 2016; Ji et al., 2018; Li et al., 2017; Scott et al., 2017). One of the main advantages of deep learning approaches is that, they do not need manual feature extraction, unlike other machine learning algorithms. This drastically reduces data preparation and reduces observer bias. In regards to forest mapping, few studies have explored the ability of Convolutional neural networks (CNN) for tree species classification from images collected by different methods. CNNs have been demonstrated to classify tree species using pictures of cross-section surfaces of the trees captured by a regular digital camera (Hafemann et al., 2014), terrestrial lidar data (Mizoguchi et al., 2017), RGB images of bark (Carpentier et al., 2018), airborne Light Detection and Ranging (LiDAR) data (Ko et al., 2018) and RGB images from UAVs to classify broad species types (Onishi and Ise, 2018). Trier et al. (2018) performed tree species classification using deep learning with a combination of three selected bands from airborne hyperspectral images and canopy height from Airborne Laser Scanning. Although RGB images have been used for classification it is unclear if the models are robust on images acquired from different seasons and angles of acquisition.

Following the aforementioned research, this work presents the use of simple RGB images from a consumer grade camera mounted on a UAV to gather large datasets of individual trees in multi-season and Convolutional Neural Networks (CNN) to classify different tree species at the individual tree level.

2. METHODOLOGY

In this work, we aim to classify tree species based on ultra-high resolution RGB images of tree canopies acquired by UAVs. We propose to use Convolutional Neural networks for classification which can learn highly descriptive features from the tree canopies. A convolutional neural network is a structured stack of

---

convolutional layers, spatial pooling layers and fully connected layers. The convolutional layers are comprised of a series of filters which are used to extract deeper features from the input, and each filter is used to calculate a feature map. The pooling layers are responsible for reducing the dimensionality of the images so as to reduce the computational load. The fully connected layer is used to classify the data by imparting probability to each class.

For implementing the CNN in our research, a suitable architecture had to be chosen based on models that achieved high levels of accuracy on object classification tasks and also work on high resolution images. In tree species dataset, several complex features need to be learned by the network. Therefore, owing to the nature of the data, choosing a deeper network with more layers was crucial to extract high level features. Also, compared to shallow networks, deep network architectures are better at generalizing because they learn all the intermediate features between the input data and the high-level classification. We did a detailed literature review on the performance of most successful and relevant deep CNN architectures (Huang et al., 2017; Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Szegedy et al., 2015). We chose to experiment with Residual Neural Network (ResNet) (He et al., 2015) architecture since it is efficient and simple to develop a much deeper network (hundreds of layers), and has performed well on several classification problems (Heredia, 2017; Šulc et al., 2016).

2.1. Network Architecture

The ResNet architecture is composed of stacked entities referred to as residual blocks. It works with identity shortcut connection that skips one or more layers while training using skip-connections or residual connections. The intermediate layers can learn to gradually adjust their weights toward zero such that the residual block represents an identity function. A building block in residual learning is shown in Figure 1. ResNet overcomes the problem of vanishing and exploding gradients problem which is encountered by typical deep neural networks. We chose our model to have 50 convolutional layers i.e. ResNet50. The architecture of ResNet50 is shown in Table 1.
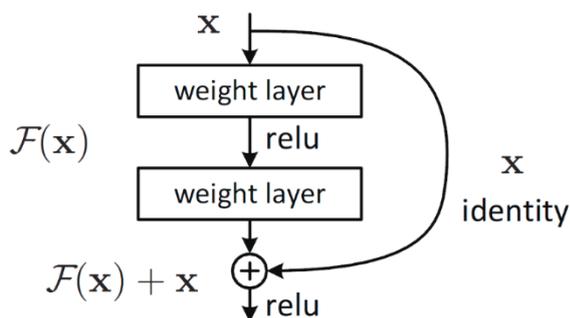

Figure 1. A building block in residual learning (He et al., 2015)

2.2 Structure of the proposed CNN model

Due to the limited availability of labelled data for training at this initial phase, we speeded up the learning process by applying transfer learning from ResNet-50 model which was pre-trained on ImageNet database. In typical deep CNN architectures including ResNET50, the early layers learn and extract general low level features and the last layers learn task specific features. Therefore, in our work, the first half of the ResNet-50 model was frozen so that the weights of the convolution filters are not modified and is used to generate low level features. We unfroze

the second half of the model so as to allow the weights to modify during training so that the prior extractor weights could be fine-tuned for our data. We replaced the average pooling layer at the end of the original ResNet architecture by a max-pooling layer because in respect to the nature of our data, average pooling can sometimes over- smoothens the image and fail to extract important features. We then added four extra fully- connected layers at the end of the network. This allowed us to fine tune the higher order feature representations along with our final classifier so as to make them more appropriate for our data. The structure of the proposed model is shown in Figure 2.

Layer Name	Layers
Conv1	3×3 max pool, stride 2
Conv2_x	1×1, 64 3×3, 64 x 3 1×1, 256
Conv3_x	1×1, 128 3×3, 128 x 4 1×1, 512
Conv4_x	1×1, 256 3×3, 256 x 6 1×1, 1024
Conv5_x	1×1, 512 3×3, 512 x 3 1×1, 2048
	average pool, 1000-d fc, softmax

Table 1. The architecture of ResNet50 model (He et al., 2015)
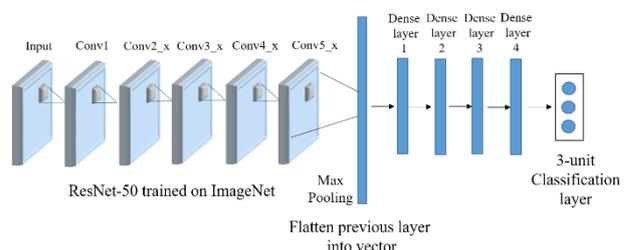

Figure 2. Structure of the proposed CNN model

3. EXPERIMENTS

3.1 Test site and data collection

For our experiments, the selected field test site is part of the Petawawa Reserved Forest (Ontario, Canada) which is dominated by pine species (White Pine and Red Pine) intermixed with Balsam Fir, White Spruce, Maple, Birch and Beech. The test area was about 10.5 ha. In this work, we used UAV RGB images acquired through three years (2015, 2016 and 2018). The UAV platform flown in 2018 flight campaign and the RGB camera mounted are shown in Figure 3 and 4 respectively. The images were collected with the leaf on (summer in 2016 and 2018) and off (fall in 2015) conditions to capture varying seasonal conditions, foliage density and the greenness, as well as in

different times of the day to bring in illumination variability. The UAV flying height in all campaign was about 150 m above ground and camera looking nadir. Every year, approximately 1600 images were collected.

Figure 3. UAV platform flown in 2018 flight campaign

Figure 4. RGB camera mounted on UAV in 2018 flight campaign

3.2 Preparation of Training data

From the UAV acquired RGB images, the training data was prepared following a workflow shown in Figure 5.

The RGB images from every year were processed using the Agisoft Photoscan v1.4.3 (Agisoft, 2017) photogrammetric software generating a digital surface model (DSM) and an orthomosaic image. The spatial resolution of the orthomosaic images produced from data in 2018, 2016 and 2015 were 2cm, 4 cm and 1 cm respectively. The orthomosaic image from year 2018 is shown in Figure 6.

To perform tree crown delineation, an iterative local maxima filtering of varying moving window size based on the tree size measured in the field was used on a Gaussian smoothed DSM reconstructed from RGB images for identifying probable tree tops. Using these as markers, a marker controlled watershed segmentation (Vepakomma et al., 2018) was then performed on the complement of the DSM for segmenting the crowns. The tree crown segmented polygons for a part of the data is shown in Figure 7.

Following the delineation, the tree species of each individual tree crown was identified and labelled. The labels were reviewed and approved by a forestry specialist. The labelled tree crowns were extracted as individual tree images and were used to train the CNN. Few examples of the individual tree crown images used for training are shown in Figure 8. Since the orthomosaic images from three years were georectified, we could overlay all three orthomosaic images and the delineation polygons. In this way, we were able to obtain three different images of the same tree. Thus, by spatializing and labelling trees in one orthoimage, we

could generate a good amount of training images from three years varying in several acquisition parameters such as season, time, illumination and angle. In few places, we encountered minor errors in georectification. In such cases, we manually adjusted the position of the delineation polygons to enclose the correct tree crown. In dense parts of the forest, extreme care was taken to ensure each tree crown was extracted with correct labels.
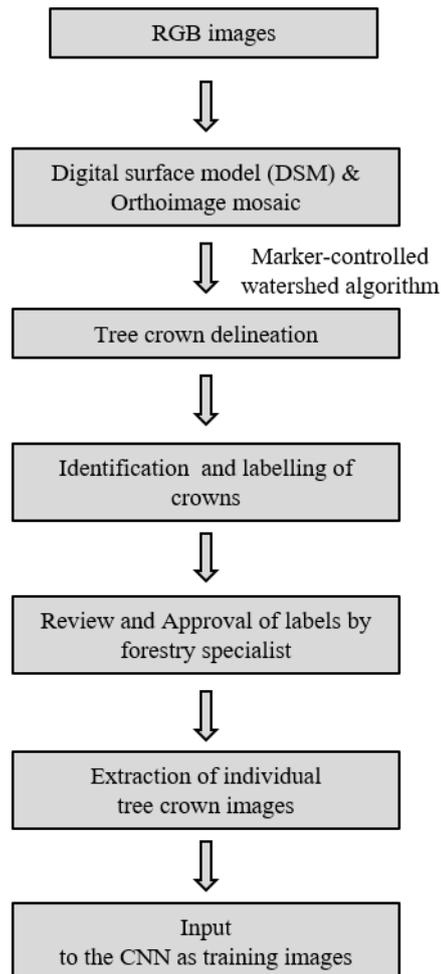

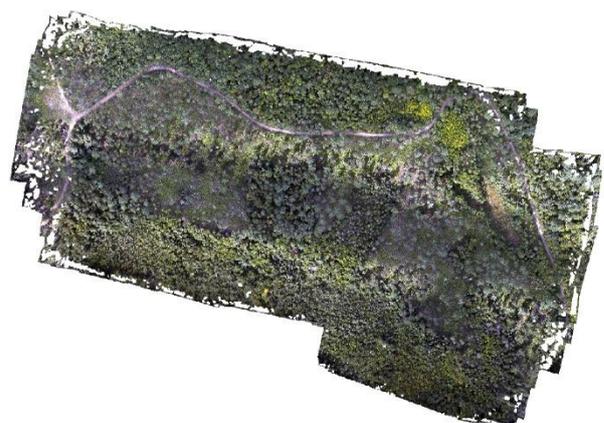Figure 5. Workflow for preparation of training data

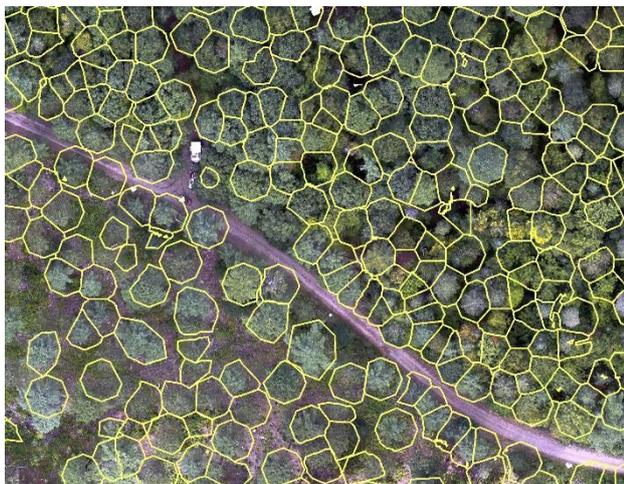Figure 6. Orthomosaic produced from RGB images in 2018 (630m X 300m)
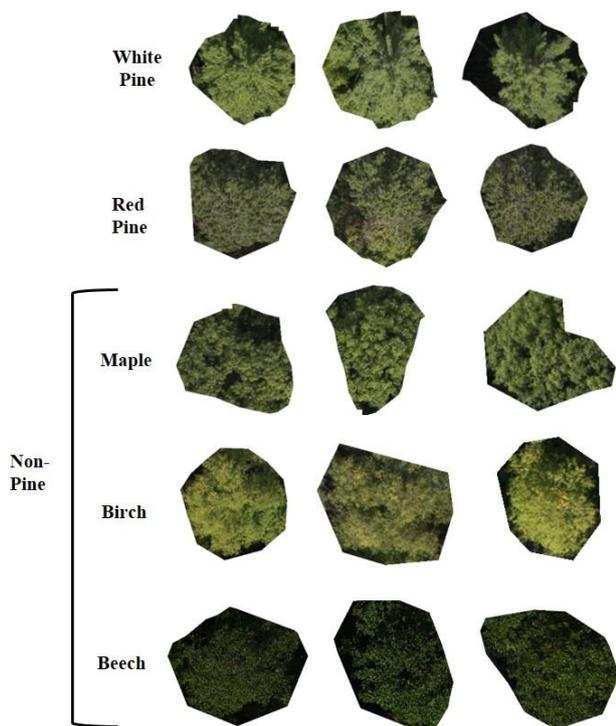
Figure 7. Tree crown delineation

Figure 8. Examples of training images

3.3 Training procedure

Initially, we apply our approach with limited data towards the identification of two pine species from the rest of the trees. Hence, the three classes chosen were Red Pine, White pine and Non-Pine species. For all our experiments, we used Keras (Chollet and others, 2015) deep learning library running on TensorFlow v1.12.0 (Abadi, M et al., 2016) backend.

The number of labelled images from all three years used for each class is shown in Table 2. We made certain that the dataset was unbiased. We also implemented image augmentation to cater to a range of physical conditions in which an image can be captured during data acquisition and to increase the number of training images. Some of the image augmentation operations performed in this work were horizontal and vertical flips, rotations, height and width shift, zoom and brightness shift.

The images were resized to 224x224 pixels relative to the pretrained network's input image size. Before feeding the data into the network, the images were shuffled in unison with their corresponding labels since it is significant to randomize the input data in order to generalize the neural network. In this way, we could prevent the network from training on entire mini-batches of highly correlated training images. For training, we used learning rate of 0.0001 for a total of 100 epochs and used Adam as the optimization method.

Species	No. of images
Red Pine	602
White Pine	593
Non-Pine	591

Table 2. The composition of training dataset

4. PRELIMINARY RESULTS AND ANALYSIS

4.1 Test results from training on three years' data

The training loss and accuracy are plotted and shown in Figure 9. For prediction, we reserved a separate set of 90 images with 30 images in each class which were not used in training. We also made sure that this prediction set contains images from all three years. The labels obtained by prediction were evaluated by comparing them to the actual labels. The prediction results are tabulated in the form of a confusion matrix shown in Table 3. Since our classes are balanced, the performance of predictions was evaluated using the accuracy measure. In addition to accuracy, other measures such as Precision, Recall and F1 score are also calculated (Table 4) since these measures ignore the correct classification of negative examples, they instead reflect the importance of retrieval of positive examples (Sokolova and Lapalme, 2009). Precision gives a probability of number of images correctly identified as positive out of total images identified as positive. The precision average for multi-class classification is given by,

$$P_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l}$$

Recall gives a probability of number of images correctly identified as positive out of total true positives. The recall average for multi-class classification is given by,

$$R_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l}$$

F1 score is defined as the harmonic mean of precision and recall, given by,

$$F1 = 2 \frac{P_M \cdot R_M}{P_M + R_M}$$

where P_M = Macro Average of precision
R_M = Macro Average of recall
T_p = True positive
F_p = False positive
F_n = False negative
l = Number of classes

$$Accuracy = \frac{Correctly\ classified\ images}{Total\ number\ of\ images}$$

Model Loss
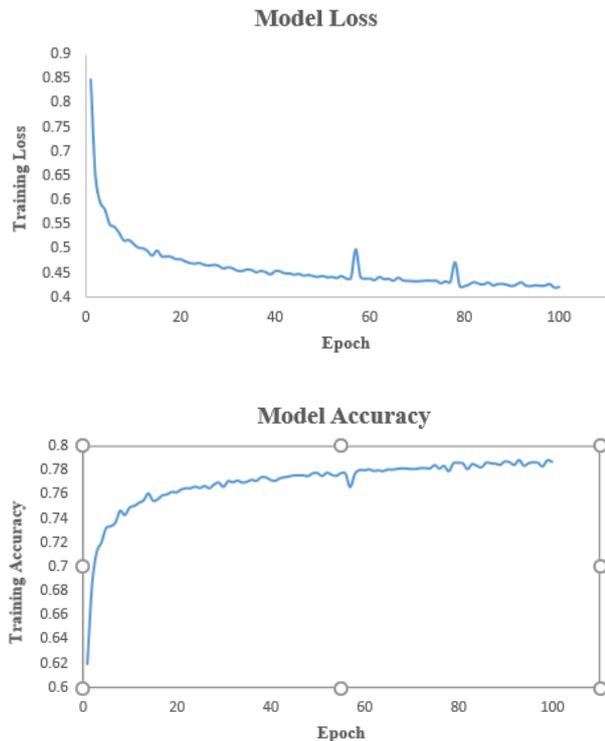

Model Accuracy

Figure 9. Plots showing training loss and accuracy

	Predicted Labels			
Species	**Red Pine**	**White Pine**	**Non-Pine**	**Total**
Red Pine	20	5	5	30
White Pine	3	23	4	30
Non-Pine	0	1	29	30
Total	23	29	38	**90**

Table 3. Confusion matrix for species classification using data from all acquisition years

Species	Precision	Recall
Red Pine	0.87	0.67
White Pine	0.79	0.77
Non-Pine	0.76	0.97
Macro Average	0.81	0.81
F1 score	0.8	
Accuracy	0.8	

Table 4. Performance measures for Species Classification using data from all acquisition years

The results indicate that the overall classification accuracy was 80%. We noticed that most of the misclassified images belong to the data acquired in the year 2015.

4.2 Test results from training on one-year data

We were interested to check if the size of dataset influenced the training results and also how the prediction is affected if we train the same network on images from only one-year flight campaign. For this purpose, we selected orthomosaic image generated from data of the year 2016 since it had the least fuzziness relative to the other orthomosaic images. The prediction results are shown in Table 5 and the accuracy estimation is shown in Table 6.

	Predicted Labels			
Species	**Red Pine**	**White Pine**	**Non-Pine**	**Total**
Red Pine	10	7	13	30
White Pine	10	17	3	30
Non-Pine	1	10	19	30
Total	21	34	35	**90**

Table 5. Confusion matrix for species classification using data from the acquisition year 2016

Species	Precision	Recall
Red Pine	0.48	0.33
White Pine	0.50	0.57
Non-Pine	0.54	0.63
Macro Average	0.51	0.51
F1 score	0.5	
Accuracy	0.5	

Table 6. Performance measures for Species Classification using data from the acquisition year 2016

We can see that the overall classification accuracy dropped from 80% to 51% by using the images from only one-year flight campaign. Therefore, using images of same trees from three different years has prevented the network to memorize and helped to generalise better. Also, by using the images from only one-year flight campaign, the number of images used for training reduced to one-third. Hence, the number of training images serve as a critical factor in classification accuracy.

5. CONCLUDING REMARKS

This work presents the ability of deeper CNN networks, such as ResNet, to classify individual trees into specific tree species from RGB images captured by UAV resulting in a cost effective and feasible approach. As a preliminary work, we performed our research towards the identification of two pine species from the rest of the trees. From the experiments on our dataset, we obtained an overall classification accuracy of 80%. We found that the classification accuracy significantly increases with increase in the number of training images. We also found that having a

temporal data of trees leads to better classification of species. As a part of this work, a large dataset of tree species is generated consisting of high resolution images of labelled tree crowns that can be used to conduct further studies involving deep neural networks in forestry. In future, we aim to experiment multispectral data integrated with RGB data to further classify the non-pine class into respective species. For this purpose, during the flight campaign in 2018, a separate flight was carried out with a Quadcopter UAV mounted with Micasense RedEdge-M multispectral sensor. In this manner, we intend to expand the approach to classify six or more species encompassing both hardwood and softwood such as Maple, Birch, Beech, Pines, Fir and Spruce. Besides ResNet, we also plan to experiment with several other deep network architectures such as VGGNet and DenseNet for tree species classification.

ACKNOWLEDGEMENTS

REFERENCES

Abadi, M. and TensorFlow, A.A.B.P., 2016, November. Large-scale machine learning on heterogeneous distributed systems. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)(Savannah, GA, USA (pp. 265-283). https://www.tensorflow.org/

Agisoft, 2017. Agisoft PhotoScan, http://www.agisoft.com/.

Brovkina, O., Cienciala, E., Surový, P., Janata, P., 2018. Unmanned aerial vehicles (UAV) for assessment of qualitative classification of Norway spruce in temperate forest stands. Geo-Spat. Inf. Sci. 21, 12–20. https://doi.org/10.1080/10095020.2017.1416994

Carpentier, M., Giguère, P., Gaudreault, J., 2018. Tree Species Identification from Bark Images Using Convolutional Neural Networks, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Presented at the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1075–1081. https://doi.org/10.1109/IROS.2018.8593514

Chollet, F., others, 2015. Keras.

Dyrmann, M., Karstoft, H., Midtiby, H.S., 2016. Plant species classification using deep convolutional neural network. Biosyst. Eng. 151, 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87. https://doi.org/10.1016/j.rse.2016.08.013

Franklin, S.E., Ahmed, O.S., 2018. Deciduous tree species classification using object-based analysis and machine learning with unmanned aerial vehicle multispectral data. Int. J. Remote Sens. 39, 5236–5245. https://doi.org/10.1080/01431161.2017.1363442

Gini, R., Sona, G., Ronchetti, G., Passoni, D., Pinto, L., 2018. Improving Tree Species Classification Using UAS Multispectral Images and Texture Measures. ISPRS Int. J. Geo-Inf. 7, 315. https://doi.org/10.3390/ijgi7080315

Hafemann, L.G., Oliveira, L.S., Cavalin, P., 2014. Forest Species Recognition Using Deep Convolutional Neural Networks, in: 2014 22nd International Conference on Pattern Recognition. Presented at the 2014 22nd International Conference on Pattern Recognition, pp. 1103–1107. https://doi.org/10.1109/ICPR.2014.199

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Heredia, I., 2017, May. Large-scale plant classification with deep neural networks. In Proceedings of the Computing Frontiers Conference (pp. 259-262). ACM.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely Connected Convolutional Networks. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Ji, S., Zhang, C., Xu, A., Shi, Y. and Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sensing, 10(1), p.75.

Ko, C., Kang, J. and Sohn, G., 2018. Deep Multi-task Learning for Tree Genera Classification. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci, pp.153-159.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp. 1097–1105.

Li, Y., Zhang, H., Shen, Q., 2017. Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. Remote Sens. 9, 67. https://doi.org/10.3390/rs9010067

Mizoguchi, T., Ishii, A., Nakamura, H., Inoue, T., Takamatsu, H., 2017. Lidar-based individual tree species classification using convolutional neural network, in: Videometrics, Range Imaging, and Applications XIV. Presented at the Videometrics, Range Imaging, and Applications XIV, International Society for Optics and Photonics, p. 103320O. https://doi.org/10.1117/12.2270123

Onishi, M. and Ise, T., 2018. Automatic classification of trees using a UAV onboard camera and deep learning. arXiv preprint arXiv:1804.10390.

Šulc, M., Mishkin, D. and Matas, J., 2016. Very deep residual networks with maxout for plant identification in the wild. Working notes of CLEF.

Scott, G.J., England, M.R., Starms, W.A., Marcum, R.A., Davis, C.H., 2017. Training Deep Convolutional Neural Networks for Land–Cover Classification of High-Resolution Imagery. IEEE Geosci. Remote Sens. Lett. 14, 549–553. https://doi.org/10.1109/LGRS.2017.2657778

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going Deeper with Convolutions. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.

Trier, Ø.D., Salberg, A.-B., Kermit, M., Rudjord, Ø., Gobakken, T., Næsset, E., Aarsten, D., 2018. Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data. Eur. J. Remote Sens. 51, 336–351. https://doi.org/10.1080/22797254.2018.1434424

Vepakomma, U., Kneeshaw, D.D., De Grandpré, L., 2018. Influence of Natural and Anthropogenic Linear Canopy Openings on Forest Structural Patterns Investigated Using LiDAR. Forests 9, 540. https://doi.org/10.3390/f9090540