Development and Application of Tree Species Identification System Using UAV and Deep Learning

Masanori Onishi

2022

Table of contents

CHAPTER 1
General Introduction and Overview

1.1 Background

1.1.1 Tree species identification from the air

Identifying individual tree species in the canopy of a forest is beneficial for a wide variety of activities in forest management and conservations. These activities include management and protection of native vegetation (Shang and Chisholm, 2014), monitoring of invasive species (Boschetti et al., 2007), wildlife habitat mapping (Jansson and Angelstam, 1999), sustainable forest management (European Environmental Agency, 2006), and biodiversity monitoring. In recent years, monitoring of biodiversity has been emphasized to meet goals such as Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+), and forest certification (e.g. FSC), both of which programs are financial mechanisms deployed to prevent deforestation and degradation. Species information of each tree has been used for biodiversity assessment from the aspect of tree-community composition (Aoyagi et al., 2017; Fujiki et al. 2016; Imai et al., 2014; Kitayama et al., 2018), but surveying in the field takes much labor and cost. Thus, the method for mapping the distribution of tree species in the canopy in large areas has been desired and expected to help and reduce the cost and labors in biodiversity surveys, as well as other various activities in forest managements and conservations.

The number of studies focusing on tree species classification using remote sensing data has constantly increased since the 1990s. In the beginning, Imageries from satellite such as Landsat with multispectral sensors were used, and hyperspectral sensors and airborne LiDAR has been used frequently in the 2010s (Fassnacht et al., 2016). Multispectral sensors can collect four to eight wavelengths including RGB and near-infrared (NIR), and hyperspectral sensors can collect one to two hundred wavelengths including RGB, NIR, and short-wavelength infrared (SWIR). From the progressing toward the high performance of those sensors, tree species

identification from airborne has been improved. Especially, a combination of hyperspectral sensors and LiDAR achieved the high performance of more than 80% accuracy in identifying 4-7 tree species (Dalponte et al., 2012; Heinzel and Koch, 2012; Shen and Cao, 2017). Using these spectra is valid for tree species identification because various spectra reflecting tree crowns have a relationship with 1) water contained in the woody tissue, photosynthetic pigment, and characteristics of structural carbohydrates (Asner, 1998; Clark and Roberts, 2012; Gao and Goetz, 1990; Knipling, 1970; Ustin et all., 2009), 2) morphology of leaves (Asner, 1998; Clark et al., 2005; Grant, 1987) and 3) tree crown structures (Leckie et al., 2005). However, these approaches utilizing multi/hyperspectral data have often experienced limitations. One of these involves the spectral features, which can differ not only between species, but also across densities of leaves, health conditions, and background noises such as understory vegetation or bare soil (Shang and Chisholm, 2014; Treuhaft et al., 2002; Waser et al., 2014). When there is a shadow, the spectrum of the shadow differs from that of the no-shadow area, resulting in a lower accuracy (Shen and Cao, 2017). In terms of identifying species in mixed forest, the performance might be lower because multiple species are included in one pixel (Immitzer et al., 2012). These challenges can be attributed to the dependence of these approaches on the use of spectral information. Furthermore, the cost of one flight and specialized hardware equipment is highly expensive, thus local people or forest managers cannot use the system regularly.

1.1.2 UAV

Recently, Unmanned Aerial Vehicles (UAVs) have been developed and general people can purchase them at a low price due to miniaturization and cost reduction of multiple sensors such as gyro, acceleration, and GPS. UAVs can be categorized into multi-copter UAVs and fixed-wing UAVs. Multi-copter UAVs can hover and fly stably for around 30 minutes, while fixed-wing UAVs can fly 45 minutes but needs a large area for taking-off and landing. Due to the limitation of the duration of the flight, UAVs can only take data of limited areas, especially multi-copter UAVs can obtain less than 10 hectares in one flight when taking a photo with overlapping from

100m altitude. Meanwhile, the characteristics of UAVs are cost-effective, and they can obtain images in a few centi-meter resolutions, in which tree shapes or individual leaves of some tree species are obvious. Furthermore, three-dimensional data can be produced by using the images and Structure from Motion (SfM). Using these data, the number of trees, tree height, and diameter at breast height (DBH) can be estimated (Goodbody et al., 2017; Iizuka et al., 2017; Mlambo et al., 2017; Tang and Shao, 2015). For those advantages, UAVs are expected to be a community-based monitoring tool that reduce cost and labor in periodic surveys of forestry, biodiversity conservation, and REDD+ (Paneque-Gálvez et al. 2014). UAVs can be a low-cost, community-based monitoring tool and used in a wide range of applications in forest management if they aid in identifying tree species. However, although tree species are essential for forest management, tree species identification using UAVs has not been sufficiently studied and its application is therefore unforeseeable (Csillik et al., 2018; dos Santos et al., 2019; Natesan et al., 2019; Safonova et al., 2019).

1.1.3 Deep Learning

Deep learning has become an effective tool for object detection and classification (He et al., 2016; Krizhevsky et al., 2012; Tan et al., 2019). It is a kind of machine learning and means learning of deep neural networks. The mechanism of neural networks imitates the human brain, in which neurons activate when they receive strong electronic signals from another neuron. Using combinations of the tens of billions of such connectivity of neurons, the brain process large amounts of information. Among deep neural networks, convolutional neural networks (CNNs) are used for object detection and classification. In CNN, convolution layers conduct filtering processes such as edge detection, and these processes are adjusted from the error of prediction of the CNN and true label. From this mechanism, the CNN can automatically detect various detailed features which other general filtering methods cannot detect from images.

In the field of tree species identification from the air, previous research studies had used other machine learning method such as random forest or support vector machines (Fassnacht et al.

2017). With the appearance of deep learning, studies combining deep learning and remote sensing data such as airborne LiDAR or aerial image, have exhibited high potential for individual tree detection, dead forest cover mapping, and forest damage assessment (Hamdi et al., 2019; Khan and Gupta, 2018; Sylvain et al., 2019; Weinstein et al., 2019).

1.2 The aims of this study and research questions

From these backgrounds, it is expected that the combination of high-resolution digital images taken by UAVs and deep learning has the potential of identifying various tree species, and it can be a low-cost and community-based monitoring tool of forest management including biodiversity assessment.

The objective of this thesis is to develop a system for tree species identification from UAV imagery with the help of deep learning and examine its feasibility in terms of robustness and its application to biodiversity monitoring. For achieving the objective, this study is composed of three main chapters (Figure 1.1) and answered the following research questions in each chapter;

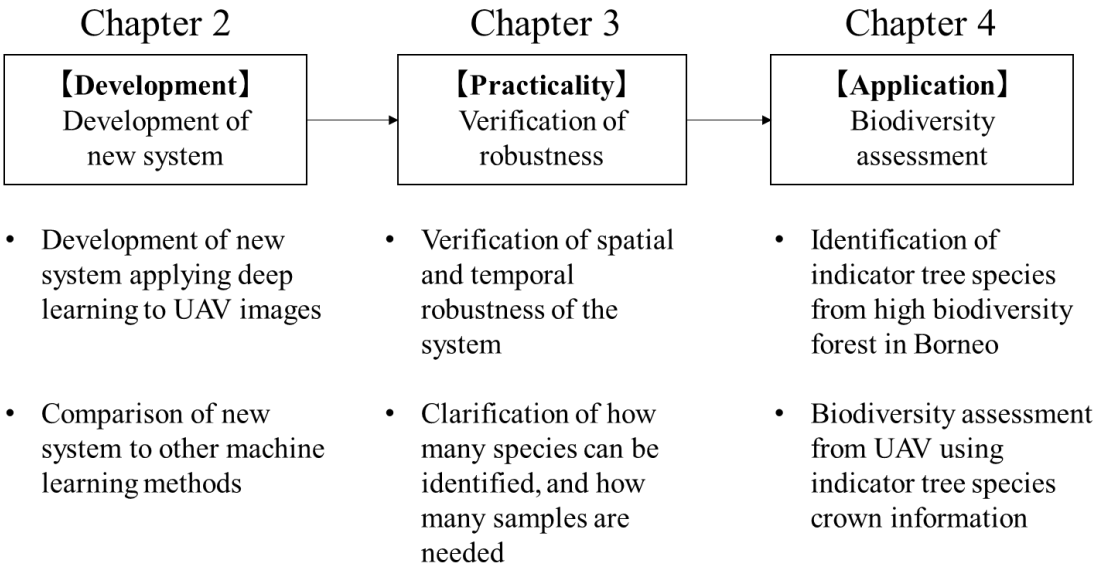


Figure 1.1 Structure of this study

(1) How can tree species be identified from UAV images combining deep learning?

The combination of high-resolution digital images taken by UAVs and deep learning has the

potential of identifying tree species interpreting from the features of texture information, which is different from previous studies that use hyperspectral sensors. In Chapter 2, I developed and proposed a chain of tree species identification and mapping system using CNN and UAV digital images and compared it to other machine learning methods. In addition, I visualized the features which CNN used as a basis of the performance of classification.

(2) To what extent can the identification system be used practically?

The performance of tree species identification can be affected by some factors such as shadow, weather conditions, individual variations of tree crowns. Moreover, how many tree species and what species can be identified is required for practical use. In Chapter 3, I verified the spatial and temporal robustness of classification performances using datasets obtained from various times and sites in Japan. The study site was six temperate forests, and I prepared 56 tree species, dead trees and gaps for training and validation for CNN.

(3) How can it be possible to apply the identification system for biodiversity monitoring?

Biodiversity can be assessed based on multiple aspects, such as the existence of endangered animals, species richness, and tree composition (Barlow et al., 2007; Lawton et al., 1998; Schulze et al., 2004; Struebig, 2013; Uehara-prado et al., 2009). Among them, tree-community composition monitoring could be a cost-effective manner because changes in canopy tree-composition can be detected from remotely sensed data (Fujiki et al. 2016, Kitayama et al. 2018). In Chapter 4, we identified *Macaranga* and *Neolamarckia* which are indicators of disturbed-forest (Aoyagi et al. 2017), and we revealed the relationship between UAV data including tree crown areas of those indicator tree species, and field data including tree-community composition as a representative of biodiversity. The study was conducted in Borneo which is one of the most biologically diverse islands (Myers et al., 2000) and was subjected to deforestation and forest degradation,

1.3 References

Aoyagi, R. *et al.* The mixing ratio of tree functional groups as a new index for biodiversity monitoring in Bornean production forests. *For. Ecol. Manage.* **403**, 27–43 (2017).

Asner, G. P. Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sens. Environ.* **64**, 234–253 (1998).

Barlow, J. *et al.* Quantifying the biodiversity value of tropical primary, secondary, and plantation forests. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18555–18560 (2007).

Bo-Cai Gao & Goetz, A. F. H. Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. *J. Geophys. Res.* **95**, 3549–3564 (1990).

Boschetti, M., Boschetti, L., Oliveri, S., Casati, L. & Canova, I. Tree species mapping with airborne hyper-spectral MIVIS data: The Ticino Park study case. *Int. J. Remote Sens.* **28**, 1251–1261 (2007).

Clark, M. L. & Roberts, D. A. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sens.* **4**, 1820–1855 (2012).

Clark, M. L., Roberts, D. A. & Clark, D. B. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens. Environ.* **96**, 375–398 (2005).

Csillik, O., Cherbini, J., Johnson, R., Lyons, A. & Kelly, M. Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. *Drones* **2**, (2018).

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.* **123**, 258–270 (2012).

dos Santos, A. A. *et al.* Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. *Sensors (Switzerland)* **19**, (2019).

European Environmental Agency. *European Forest Types. Categories and types for sustainable forest management reporting and policy*. *EEA Technical report No9/2006* https://www.eea.europa.eu/publications/technical_report_2006_9 (2006) doi:10.3832/efor0425-003.

Fassnacht, F. E. *et al.* Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **186**, 64–87 (2016).

Fujiki, S. *et al.* Large-scale mapping of tree-community composition as a surrogate of forest degradation in Bornean tropical rain forests. *Land* **5**, (2016).

Goodbody, T. R. H., Coops, N. C., Marshall, P. L., Tompalski, P. & Crawford, P. Unmanned aerial systems for precision forest inventory purposes: A review and case study. *For. Chron.* **93**, 71–81 (2017).

Grant, L. Diffuse and specular characteristics of leaf reflectance. *Remote Sens. Environ.* **22**, 309–322 (1987).

Hamdi, Z. M., Brandmeier, M. & Straub, C. Forest damage assessment using deep learning on high resolution remote sensing data. *Remote Sens.* **11**, 1–14 (2019).

He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 770–778 (2016) doi:10.1109/CVPR.2016.90.

Heinzel, J. & Koch, B. Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. *Int. J. Appl. Earth Obs. Geoinf.* **18**, 101–110 (2012).

Iizuka, K., Yonehara, T., Itoh, M. & Kosugi, Y. Estimating tree height and diameter at breast height (DBH) from digital surface models and orthophotos obtained with an unmanned aerial system for a Japanese Cypress (Chamaecyparis obtusa) Forest. *Remote Sens.* **10**, (2018).

Imai, N. *et al.* Tree community composition as an indicator in biodiversity monitoring of REDD+. *For. Ecol. Manage.* **313**, 169–179 (2014).

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.* **4**, 2661–2693 (2012).

Jansson, G. & Angelstam, P. Threshold levels of habitat composition for the presence of the long-tailed tit (Aegithalos caudatus) in a boreal landscape. *Landsc. Ecol.* **14**, 283–290 (1999).

Khan, S. & Gupta, P. K. Comparitive study of tree counting algorithms in dense and sparse vegetative regions. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. Vol. XLII-5, 2018 ISPRS TC V Mid-term Symp. "Geospatial Technol. – Pixel to People"* **XLII**, 801–808 (2018).

Kitayama, K. *et al.* Biodiversity observation for land and ecosystem health (BOLEH): A robust method to evaluate the management impacts on the bundle of carbon and biodiversity ecosystem services in tropical production forests. *Sustain.* **10**, (2018).

Knipling, E. B. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sens. Environ.* **1**, 155–159 (1970).

Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 1097–1105 (2012) doi:http://dx.doi.org/10.1016/j.protcy.2014.09.007.

Lawton, J. H. *et al.* Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391**, 72–76 (1998).

Leckie, D. G. *et al.* Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sens. Environ.* **94**, 311–326 (2005).

Mlambo, R., Woodhouse, I. H., Gerard, F. & Anderson, K. Structure from motion (SfM) photogrammetry with drone data: A low cost method for monitoring greenhouse gas emissions from forests in developing countries. *Forests* **8**, (2017).

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, G. A. B. da & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).

Natesan, S., Armenakis, C. & Vepakomma, U. Resnet-based tree species classification using uav images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLII**, 475–481 (2019).

Paneque-Gálvez, J., McCall, M. K., Napoletano, B. M., Wich, S. A. & Koh, L. P. Small drones for community-based forest monitoring: An assessment of their feasibility and potential in tropical areas. *Forests* **5**, 1481–1507 (2014).

Safonova, A. *et al.* Detection of fir trees (Abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sens.* **11**, (2019).

Schulze, C. H. *et al.* Biodiversity indicator groups of tropical land-use systems: Comparing plants, birds, and insects. *Ecol. Appl.* **14**, 1321–1333 (2004).

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.* **7**, 2481–2489 (2014).

Shen, X. & Cao, L. Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. *Remote Sens.* **9**, (2017).

Struebig, M. J. *et al. Quantifying the Biodiversity Value of Repeatedly Logged Rainforests. Gradient and Comparative Approaches from Borneo*. *Advances in Ecological Research* vol. 48 (Elsevier Ltd., 2013).

Sylvain, J. D., Drolet, G. & Brown, N. Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS J. Photogramm. Remote Sens.* **156**, 14–26 (2019).

Tan, M. & Le, Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. *36th Int. Conf. Mach. Learn. ICML 2019* **2019-June**, 10691–10700 (2019).

Tang, L. & Shao, G. Drone remote sensing for forestry research and practices. *J. For. Res.* **26**, 791–797 (2015).

Treuhaft, R. N., Asner, G. P., Law, B. E. & Van Tuyl, S. Forest leaf area density profiles from the quantitative fusion of radar and hyperspectral data. *J. Geophys. Res.* **107**, (2002).

Uehara-Prado, M. *et al.* Selecting terrestrial arthropods as indicators of small-scale disturbance: A first approach in the Brazilian Atlantic Forest. *Biol. Conserv.* **142**, 1220–1228 (2009).

Ustin, S. L. *et al.* Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sens. Environ.* **113**, S67–S77 (2009).

Waser, L. T., Küchler, M., Jütte, K. & Stampfer, T. Evaluating the potential of worldview-2 data to classify tree species and different levels of ash mortality. *Remote Sens.* **6**, 4515–4545 (2014).

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A. & White, E. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. *Remote Sens.* **11**, (2019).

CHAPTER 2
Explainable Identification and Mapping of Trees Using UAV RGB Image and Deep Learning

2.1 Introduction

The accurate characterisation of tree species distribution in forest areas is an important task for forest management and forest research. In particular, management and protection of native vegetation (Shang and Chisholm, 2014), monitoring of invasive species (Boschetti et al., 2007), wildlife habitat mapping (Jansson and Angelstam, 1999), and sustainable forest management (European Environmental Agency, 2006) are some of the objectives of studies that require tree species distribution characterisation on a wide scale. To this end, many studies have been conducted using remote sensing data (Fassnacht et al., 2016). Thus far, remote sensing research with these objectives has mainly employed satellites or aircraft. In the past, much attention has been devoted to multispectral Landsat satellites, which, because of their low cost, facilitate the mapping of forest types (Salovaara et al., 2005). The advantage of Landsat satellite is that it enables the coverage of vast areas, i.e. on a country scale. However, its resolution is 30 m, which does not allow easy identification of tree species. Since 2000, many studies have used very high-resolution data for tree species classification from commercial satellites, e.g. from World-view2 and QuickBird, with resolutions of 0.5 m/2.0 m and 0.6 m/2.4 m for panchromatic/multispectral data, respectively (Immitzer et al., 2012; van Lier et al., 2009; Wang et al., 2004; Waser et al., 2014). Further, in recent years, studies using aircraft have succeeded in identifying several tree species. The spatial resolution of images used therein is also very high: approximately 0.2–3.0 m. Most of these studies used specialised hardware such as multispectral, hyperspectral, and LiDAR sensors, and achieved high performance in identifying tree species. For example, Shen and Cao (Shen and Cao, 2017) succeeded in identifying five tree species with more than 85% accuracy. Dalponte et al. (2012) identified seven tree species and no-forest

class at approximately 80% accuracy. These two studies revealed that combining hyperspectral sensor and LiDAR data acquired by an airplane is superior to the use of only multispectral or hyperspectral sensor. However, although this combination of data acquisition methods performs superiorly, the equipment involved therein is highly expensive. These approaches utilising multi/hyperspectral data have often experienced problems. These involves the spectral features differ across densities of leaves, and health and forest conditions, and affected by background noises and shadows (Immitzer et al., 2012; Shang and Chisholm, 2014; Shen and Cao, 2017; Treuhaft et al., 2002; Waser et al., 2014).

In recent decades, unmanned aerial vehicles (UAVs) have been used experimentally in forestry applications (Goodbody et al., 2017; Iizuka et al., 2018; Mlambo et al., 2017; Paneque-Gálvez et al., 2014; Tang and Shao, 2015). Regarding tree identification, the most important difference between manned aircraft and UAVs is that UAVs can fly near canopies and acquire extremely high-resolution images; the images from UAVs have a spatial resolution of a few centimetres. In some cases, even the tree features at the leaf level can be seen. If I can use these features in tree identification systems, it is possible to map various tree species with the use of simple red–green–blue (RGB) digital images; such a method would facilitate cost-effective monitoring with broad application potential.

Meanwhile, deep learning has become an effective tool for object detection. Among deep learning techniques, convolutional neural networks (CNN) have demonstrated high classification performance for digital images in the computer vision field (Ise et al., 2018; Krizhevsky et al., 2012; Szegedy et al., 2014). Thus far, other machine learning approaches such as support vector machines (SVMs) or the random forest have been used in this field. Especially SVMs, which enable us to conduct supervised non-parametric prediction, are one of the conventional machine learning methods and have been used for tree species identification (Dalponte et al., 2012, 2013; Dian et al., 2017; Fassnacht et al., 2016; Louarn et al., 2017). One of the notable advantages of deep learning compared to such machine learning methods is that deep learning does not require manual feature extraction. Other machine learning methods

required researchers to manually extract features: in this field, texture features and vegetation indices as well as raw band values are extracted in the process of feature extraction. As for texture features, in addition to the mean and variation values of each band, the grey-level co-occurrence matrix (GLCM) has been widely used to provide texture features for improving performance (Dian et al., 2017; Lim et al., 2019; Louarn et al., 2017; Tsai and Chou, 2006). After the feature extractions, past studies utilising hyperspectral data sometimes conducted feature selection for reducing dimension in order to avoid the curse of dimensionality and high computational cost (Dalponte et al., 2013; Shen and Cao, 2017). Alternatively, deep learning can utilise full feature information, especially that pertaining to the spatial relationship of pixels, which provides information regarding the textures and shapes of trees. Therefore, even when simple digital images are used, deep learning is expected to succeed in identifying trees with high detail and accuracy.

For interpreting the performance of deep learning classification, the feature which deep learning focus on should be clear. However, deep learning was called as black box due to its multilayer nonlinear structure, comparing to other machine learning. Recently algorithms which visualize some features of CNN's models have been developed. Guided gradient-weighted class activation mapping (Grad-CAM) is one of such algorithms, and it can highlight particular image regions which provide meaningful information for model prediction (Selvaraju et al., 2016). Using this algorithm helps us to know the features that deep learning used, which means I can know whether deep learning really used detailed features such as textures of leaves and tree shapes, and it can provide understandable visual information about model performance.

In this work, I applied deep learning to RGB images taken by a UAV. This combination is expected to have a high potential for identifying tree types and tree species, even if the RGB image is acquired with a consumer-grade digital camera. Some recent studies have shown the potential for individual tree detection and classification of one or a few specific tree species (Csillik et al., 2018; dos Santos et al., 2019; Natesan et al., 2019; Safonova et al., 2019). The object detection method enables us to know the existence of the object tree species; thus, it may

be suitable for finding specific tree species such as invasive tree species. On the other hand, forest mapping, which enables us to know the crown ratio to the area of each tree class, as well as existence of trees, helps me to monitor the change of the forest dynamics and predict biomass. A forest mapping system using UAVs and digital images would be a particularly cost-effective and useful tool for forest management applications.

The contributions of this study are threefold: 1) I propose a forest mapping system using UAVs and CNN classification. 2) I reveal its classification potential for several tree classes including tree types (e.g. deciduous, evergreen, coniferous and braoedleaved) and species in two seasons. To evaluate the performance of the CNN, I compared it with that of another machine learning method. In this study, I employed SVM as a machine learning platform, and the pixel values of each band and GLCM texture values as the features for machine learning. 3) I reveal what kind of tree features a CNN uses and show its importance (contribution) for classification by comparing it to other machine learning methods.

2.2 Material and Methods

2.2.1 Study site

The study site was the Kamigamo Experimental Station of Kyoto University, located in a suburban area of Kyoto, Japan (Figure 2.1). This area is located in a warm and humid climate zone, with an elevation of 109–225 m above sea level. The mean annual precipitation and temperature are 1,582 mm and 14.6 ℃, respectively. The overall area is 46.8 ha. 65% of the area is naturally generated forest, primarily consisting of Japanese cypress (*Chamaecyparis obtusa)* and some broad-leaved trees such as oak (*Quercus serrata* or *Quercus glauca).* Within this area, 28% is planted forest, mainly consisting of foreign coniferous species. 7% consists of sample gardens, nurseries, or buildings.

In this work, I focused on the northern part (an area of 11 ha) of the Kamigamo Experimental Station, containing a naturally regenerated forest of Japanese cypress, and a managed forest of Metasequoia (*Metasequoia glyptostroboides*), strobe pine (*Pinus strobus)*, slash pine (*Pinus elliottii),* and taeda pine *(Pinus taeda)*.

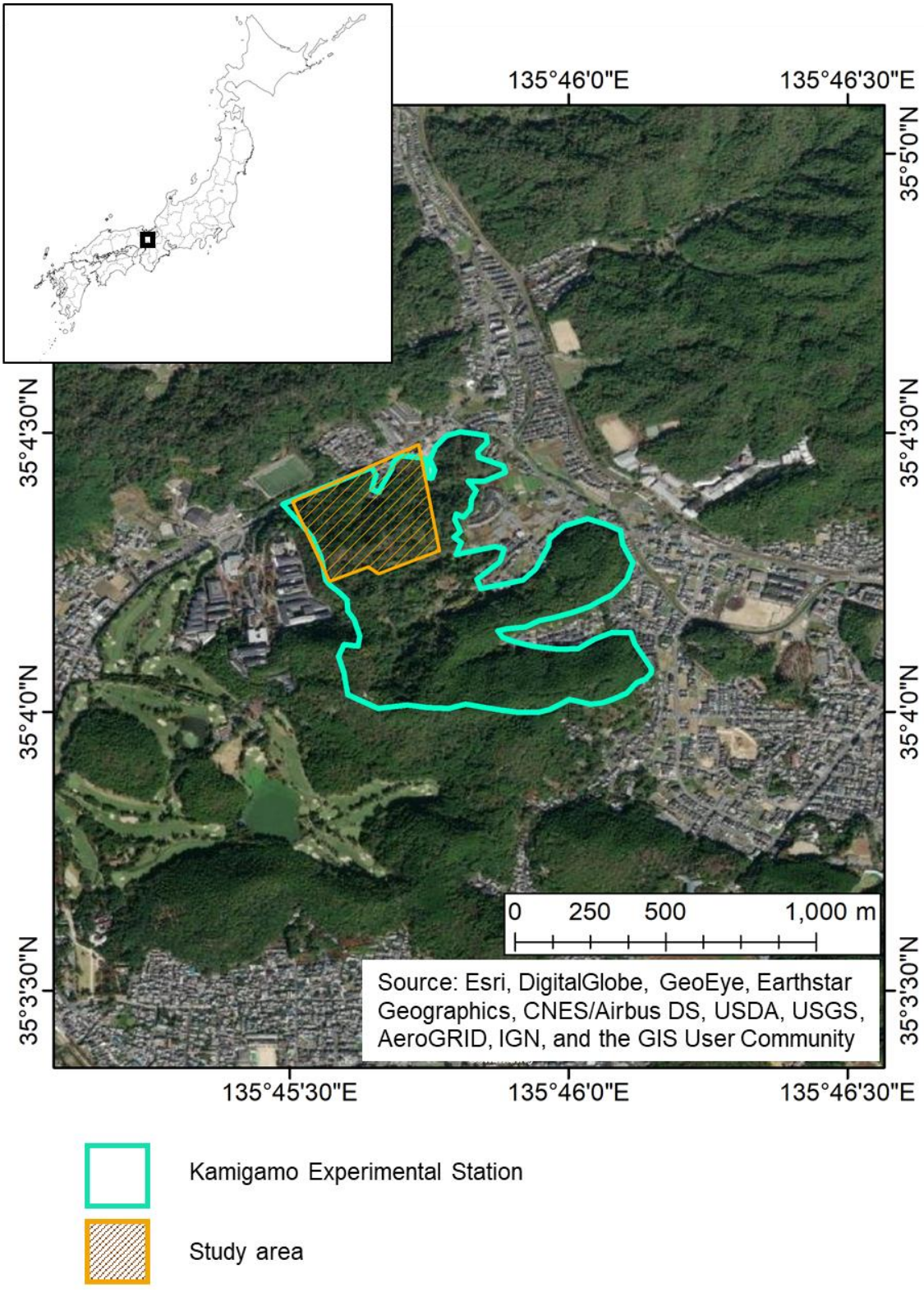


Figure 2.1 Location map of Kamigamo Experimental Station and study area.

2.2.2 Remote sensing data

UAV flights were conducted around noon in two seasons: on October 2, 2016, which is the end of the leaf season, and November 20, 2016, the peak of the fall leaf offset season. I used UAV DJI Phantom 4 (DJI, Shenzhen China). The UAV had an onboard camera with a 1/2.3 CMOS sensor that can capture RGB spectral information. The UAV was operated automatically using the DroneDeploy v2.66 application (www.dronedeploy.com, Infatics Inc., San Francisco, United States). On October 2, I set flight parameters as follows: both the overlap and sidelap were set to 75%, and the flight height was set to 80 m from the take-off ground level. However, I failed to align some parts of the images; thus, I changed the overlap and height parameters to 80% and 100 m on November 20. I used 10 ground-control points (GCPs) for reducing the error of the GPS with the images. From the images taken by the UAV, I produced an orthomosaic photo and a digital surface model (DSM) using the Agisoft PhotoScan Professional v1.3.4 software (www.agisoft.com, Agisoft LLC, St. Petersburg, Russia). An orthomosaic photo is an image that is composed of multiple overhead images corrected for perspective and scale. The parameter settings used in generating the orthomosaic photo are shown in Table 2.1. These parameters are for November 20. The parameters for October 2 differ only in that the resolution were approximately one centimetre. resolution of the orthomosaic photo and DSM was approximately 5 cm and 9 cm, respectively.

Table 2.1 Parameter settings for the workflow used in generating orthomosaic photo by photoscan software.

Workflow	Parameter	Settings
Align Photos Build	Accuracy	Highest
	Preselection	Reference
	Key Point Limit	80,000
	Tie Point Limit	8,000
	Adaptive Camera Model Fitting	Yes
Build Dense Cloud	Quality	High
	Depth Filtering	Aggressive
Build Digital Elevation Model	Projection	JGD2011
	Source Data	Dense Cloud
	Interpolation	Enabled
	Resolution	0.093 m
Build Orthomosaic	Surface	DEM
	Enable Hole Filling	Yes
	Resolution	0.047 m

2.2.3 UAV data processing

2.2.3.1 Segmentation and preparation of supervised data

The technological workflow of the individual tree image segmentation and extraction method I used is summarised in Figure 2.2. First, I segmented each tree crown using UAV image (orthomosaic photo), a DSM, and a slope model. Second, I visually constructed the ground truth map. Third, I extracted each tree image with a ground truth label which is a class name identified visually from UAV photo. Further details are discussed in Sections from "Object-based tree crown segmentation" to "Tree image extraction with ground truth label".

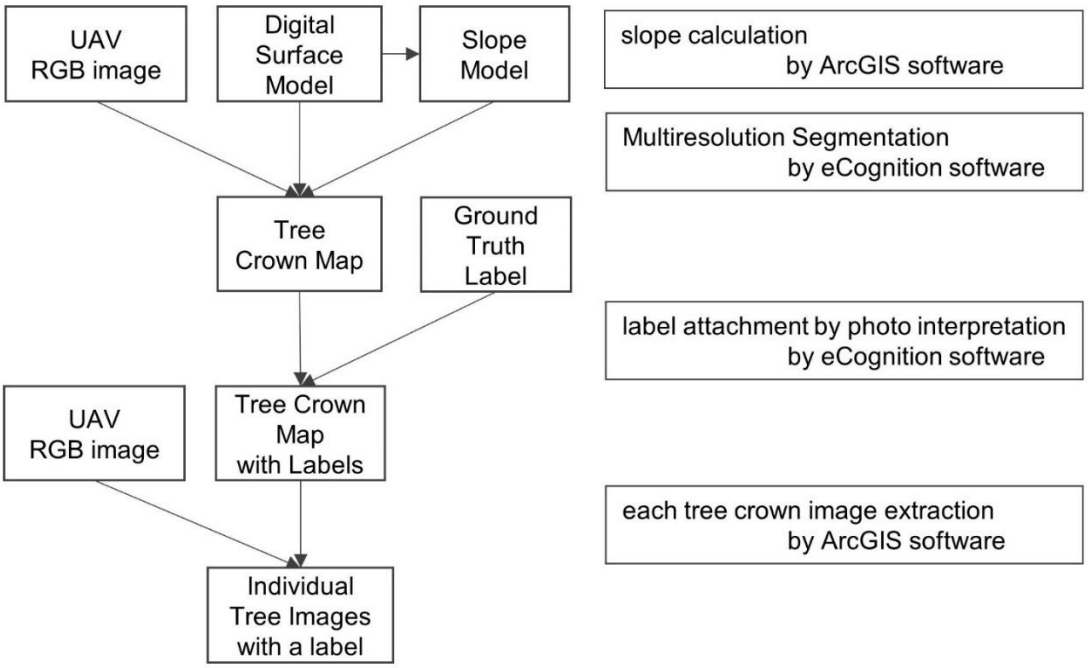


Figure 2.2 Workflow for supervised images extraction.

2.2.3.2 Object-based tree crown segmentation

At the segmentation stage, I segmented at the tree level. First, I constructed a slope model by calculating the slope from the DSM using the ArcGIS Desktop v10.4 software (www.esri.com, Environmental Systems Research Institute, Inc., Redlands, United States). The slope model showed the maximum rate of elevation change between each cell and its neighbours, such that the borders of trees were emphasised. From the orthomosaic photo, the DSM, and the slope model, tree crown segmentation was performed in the eCognition Developer v9.0.0 software (www.trimble.com, Trimble, Inc., Sunnyvale, United States) using the 'Multiresolution Segmentation' algorithm (Baatz and Schäpe, 2000). The parameter values were adjusted by trial and error. The tree crown map made by this segmentation process is shown in Figure 2.3 with enlarged images for visual confirmation of the result, and the best parameters are presented in Table 2.2.

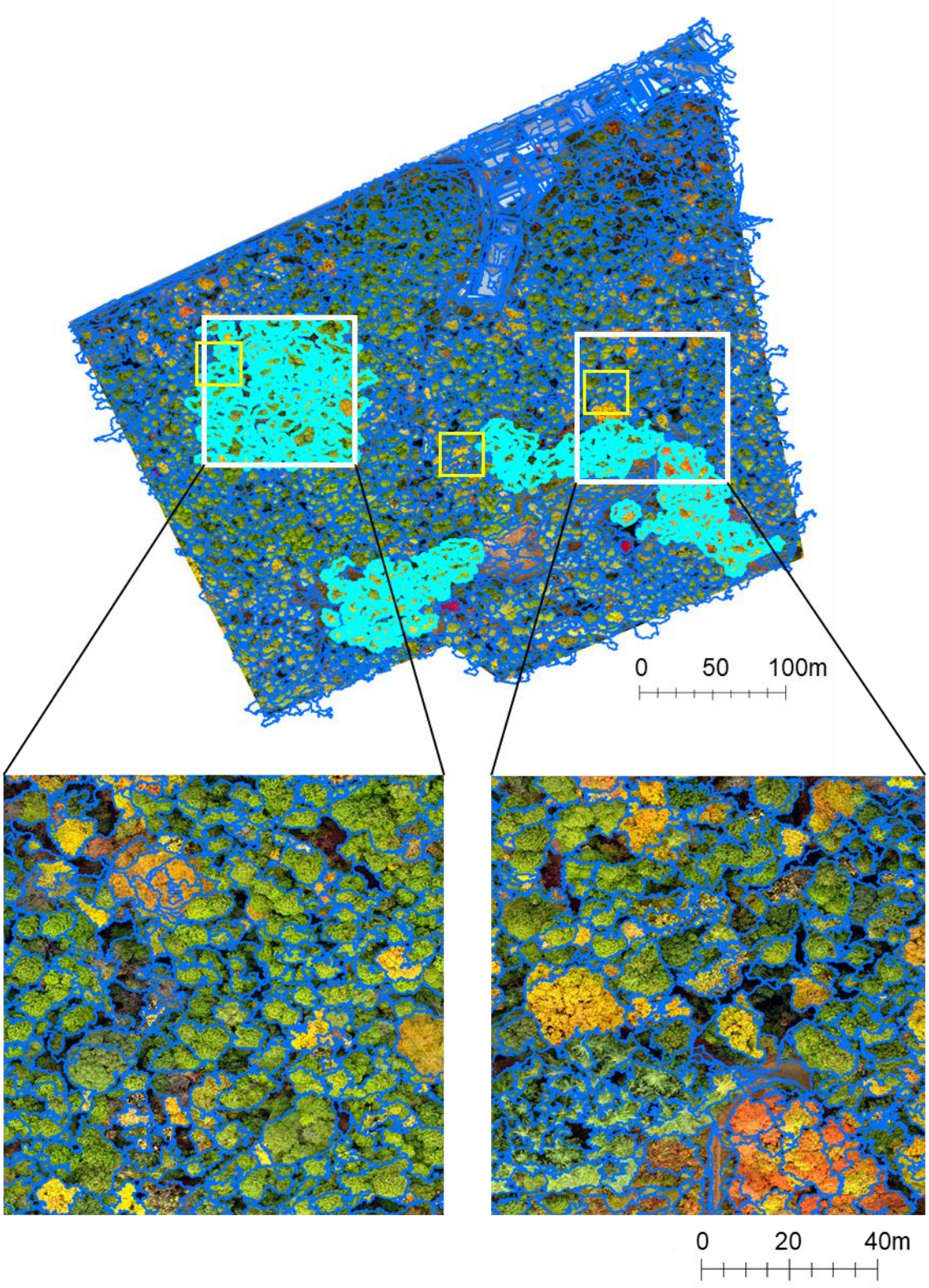


Figure 2.3 Whole area and representative enlarged tree crown map. The blue line show border-line of segmented polygons. The white rectangle shows the location of enlarged area,

and light blue polygons in the upper figure are used for evaluating the accuracy of tree segmentation. The yellow rectangle shows the location of field survey. This map was constructed via multiresolution segmentation using colour, DSM, and Slope model.

Table 2.2 Parameters for multiresolution segmentation in eCognition software.

Setting	Selected option
Weight of R, G, B, DSM, Slope	1, 1, 1, 2, 3
Scale	200
Compactness	0.5
Shape	0.2

Herein, I evaluated the accuracy of the segmentation. The segmented crowns were placed into the following five categories according to their spatial relationships with the visually confirmed reference crown. The five categories, set based on a previous study (Jing et al., 2012), and illustrated in Figure 2.4, are as follows.

(a) Matched: If the overlap of the segmented polygon and the reference crown was more than 80%, the segmented polygon was categorized as “Matched”

(b) Nearly matched: If the overlap of the segmented polygon and the reference crown was 60-80%, the segmented polygon was categorized as “Nearly matched”.

(c) Split: If the overlap of the segmented polygon and the reference crown was 20-60%, the segmented polygon was categorized as “Split”

(d) Merged: If multiple reference crowns covered by the segmented polygon, and even one overlap between each reference crown and the segmented polygons was more than 20%, the segmented polygon was categorized as “Merged”. If the segmented polygon had only one class reference crowns, the polygon was categorized as “one class merged”. If the segmented polygon had multiple class reference crowns, the polygon was categorized as “multiple class merged”.

(e) Fragmented: If one or multiple reference crowns covered by the segmented polygon, and the all respective overlaps between each reference crown and the segmented polygons were less than 20%, the segmented polygon was considered as a “fragmented polygon”.

I calculated the segmentation accuracy of trees at four areas: Areas 1–4. Area 1 was a deciduous coniferous forest and Area 2 was a strobe pine forest, for which I calculated the entire area. Area 3 was a slash pine and taeda pine forest, for which I calculated part of the areas. Area 4 was a naturally regenerated forest, for which I calculated 1 ha in area. As a result, some segmented images had multiple tree crowns, but this method almost succeeded in separating each tree class (Table 2.3).

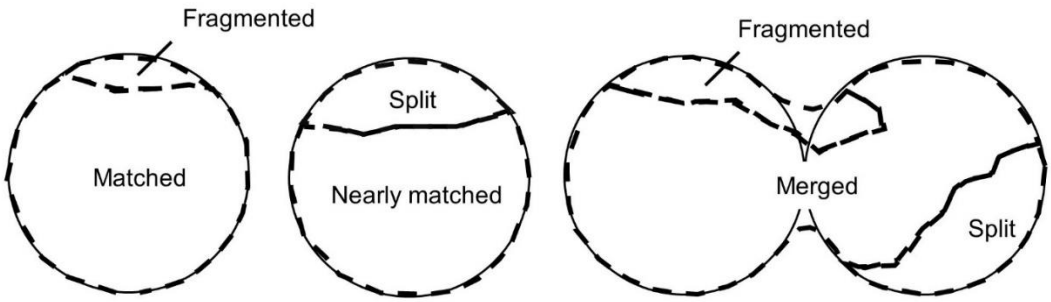


Figure 2.4 Diagram of the five categories in which segmented tree crowns were placed.

Table 2.3 Accuracy statistics of the tree crown maps. (Area 1: deciduous coniferous tree; Area 2: strobe pine forest; Area 3: slash pine and taeda pine forest; Area 4: naturally regenerated forest)

Forest area	Matched	Nearly matched	Split	Merged		Fragmented
				one class	multiple class	
Area 1	7	7	13	31	2	17
Area 2	1	0	5	41	2	17
Area 3	19	6	7	27	2	7
Area 4	38	11	23	48	24	16

2.2.3.3 Ground truth label attachment to tree crown map

After segmentation, I classified segmented images into the following seven classes: deciduous broad-leaved tree, deciduous coniferous tree, evergreen broad-leaved tree, *Chamaecyparis obtusa*, *Pinus elliottii* or *Pinus taeda*, *Pinus strobus,* and non-forest. The ‘non-

forest' class included understory vegetation and bare land, as well as artificial structures. For deciding these classes, I conducted field research. I set three rectangular plots sized 30 m × 30 m (Figure 2.3) and checked the tree species, regarding the classes I decided could be identified from the November 20 UAV images. The *Pinus elliottii* or *Pinus taeda* class consisted of two *Pinus* species, because these two species are difficult to identify from UAV images. At the ground truth map-making phase, I visually attached the class label to each tree crown, using nearest neighbour classification in the eCognition software to improve operational efficiency, which was then used for forest mapping (Machala and Zejdová, 2014) (Figure 2.5). More specifically, I chose some image objects as training samples and applied that algorithm to the overall tree crowns. In subsequent steps, by adding wrongly classified objects to correct classes of the training samples, I improved the accuracy of the ground truth map.

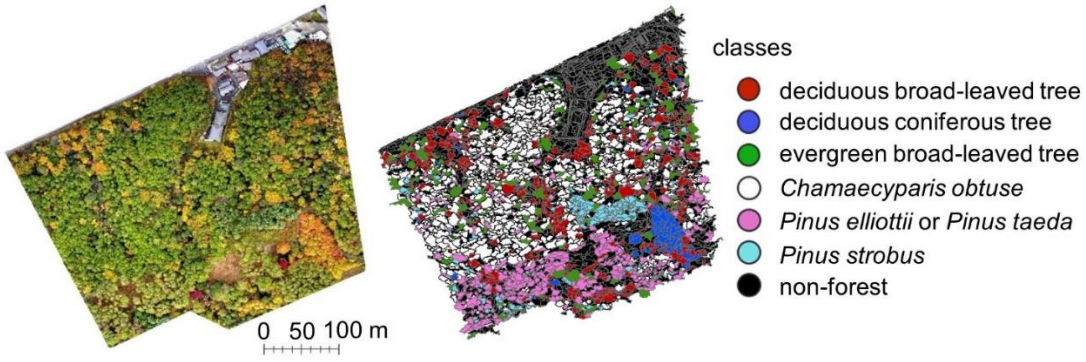


Figure 2.5 Segmentation and ground truth map. The tree classes found in the image on the left are represented by the colours explained in the legend in the figure on the right.

2.2.3.3 Tree image extraction with ground truth label

From the orthomosaic photos of the two season and the ground truth map, I extracted each tree image with a class label using the 'Extract by Mask' function in ArcGIS. There were some inappropriate images, such as fragments of trees, those difficult to be interpreted or classified visually, and those including multiple classes; thus, I manually deleted inappropriate images and

placed wrongly classified images into the correct class by group consensus. Representative images of the tree classes are shown in Figures 2.6 and 2.7. The number of extracted images and that of arranged images are shown in Table 2.4. After arrangement, the number of each class ranged from 37 to 416. The images had a wide range of sizes, but the length of one side of the largest image was approximately 400 pixels

After extraction, I resized the images from October 2 to the size of images from November 20 in order to align the two season conditions. Thus, all images were adjusted to the size of images taken from a height of approximately 100 m.

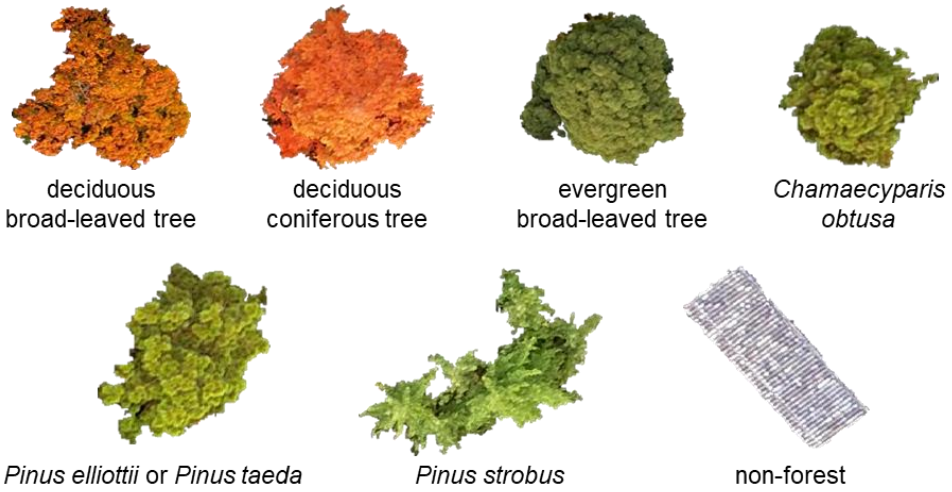


Figure 2.6. Representative extracted images from each class in the November 20 images. These images were segmented well at each tree crown level. However, the image of *Pinus strobus* includes several tree images. The image of the non-forest class shows the roof of a house.

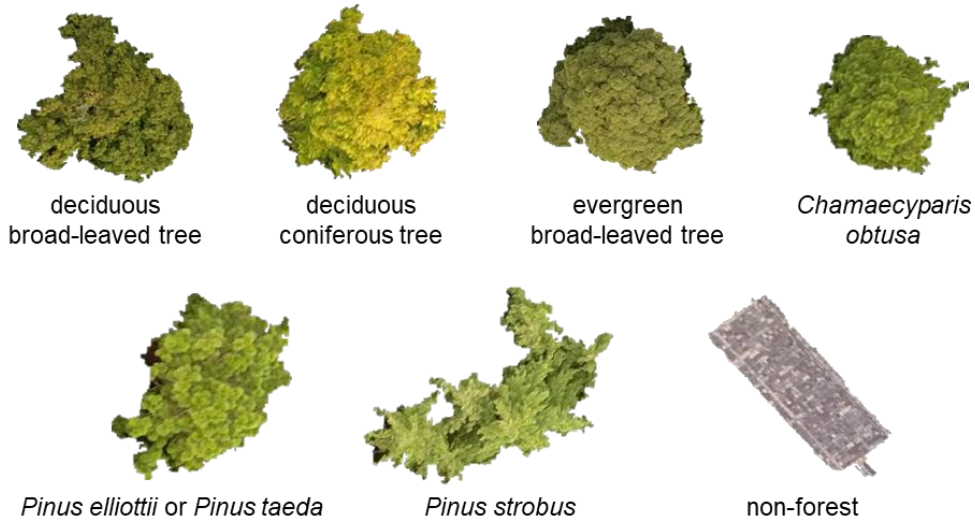


Figure 2.7 Representative extracted images from each class in the October 2 images. These images were extracted from the same tree crown map polygon as the November 20 images.

Table 2.4 Number of images in each class (Class 1: deciduous broad-leaved tree; Class 2: deciduous coniferous tree; Class 3: evergreen broad-leaved tree; Class 4: *Chamaecyparis obtusa*; Class 5: *Pinus elliottii* or *Pinus taeda*; Class 6: *Pinus strobus*; Class 7: non-forest). These are the values for one dataset.

Class	Extracted	Arranged	Training	Augmented training	Validation	Test
1	418	315	174	1392	70	71
2	87	60	33	264	11	16
3	218	73	28	224	24	21
4	758	333	184	1472	65	84
5	276	166	94	752	38	34
6	129	37	18	144	6	13
7	1439	416	207	1656	110	99

2.2.4 Machine learning

To construct a model for object identification, I used the publicly available package PyTorch v0.4.1 (Paszke et al., 2017) as a deep learning framework and four standard neural network models—specifically, AlexNet (Krizhevsky et al., 2012), VGG16 (Simonyan and Zisserman, 2014), Resnet18, and Resnet152 (He et al., 2016) —for fine-tuning. Fine-tuning is an effective method to improve the learning performance, especially when the amount of data is insufficient for training (Girshick et al., 2014). I used each neural network model, which had been learned with the ImageNet dataset (Deng et al., 2009), and trained all neural network layers using my data. At the CNN training phase, I augmented the training images eight times by flipping and rotating them. Further augmentation did not improve accuracy. For the input to the CNN, I applied 'random resized crop' at a scale of 224 × 224 pixel size for training, which crops the given image to a random size and aspect ratio. For validation and training, I resized the images into 256 × 256 pixel sizes and used 'centre crop' at a scale of 224 × 224 pixel size. These cropping algorithms extracted only one resized image (patch) from each cropped image. The ranges of the other learning settings are outlined in Table 2.5.

To evaluate the performance of the CNN, I used SVM as a machine learning platform. I used the average and standard deviation of each band and GLCM texture values as features. GLCM is a spatial co-occurrence matrix that computes the relationships of pixel values, and uses these relationships to compute the texture statistics (Haralick et al., 1973). For calculating GLCM, images with a large number of data bits result in huge computational complexity. In this case, the images that were converted to grey scale were 8-bit data. It is known that reduction of bit size causes only minor decrease in classification accuracy; hence, I rescaled from 8-bit to 5-bit (Narayanan et al., 2000; Xu et al., 2003). After calculation of GLCM, I extracted five GLCM texture features (angular second moment (ASM), contrast, dissimilarity, entropy, and homogeneity). Their algorithms are defined in Equations (1)–(5):

$$\text{ASM} = \sum_{i,j}^{N} (P_{i,j})^2 \tag{1}$$

$$\text{Contrast} = \sum_{i,j}^{N} P_{i,j}(i-j)^2 \tag{2}$$

$$\text{Dissimilarity} = \sum_{i,j}^{N} P_{i,j}|i-j| \tag{3}$$

$$\text{Entropy} = \sum_{i,j}^{N} P_{i,j} \log\left(P_{i,j}\right) \tag{4}$$

$$\text{Homogeneity} = \sum_{i,j}^{N} P_{i,j}/(1+(i-j)^2) \tag{5}$$

where $P_{i,j}$ is the GLCM at the pixel which is located in row number *i* and column number *j*. I obtained these GLCM texture features centering on each pixel, excluding pixels close to the image margin, and then calculated their mean and standard deviation for each image. Another important parameter that affects classification performance is the kernel size (Franklin et al., 1996; Marceau et al., 1990). To determine the most suitable kernel size for GLCM operation, I calculated GLCM texture features with various kernel sizes of 3, 11, 19, 27, 35, 43, 51, and 59.

For SVM validation, I used radial basis function (rbf) kernel and conducted a parameter grid search in the range of gamma from 10^{-1} to 10^{-5} and cost from 1 to 10^{5}. As a result of the grid search, I obtained the best validation accuracy and the best parameters at each GLCM kernel size (Figure 2.8). The validation accuracy slightly increased along with the increase in kernel size, and the accuracy stopped increasing at the 51 × 51 kernel size. Considering this result, I adopted the 51 × 51 kernel size and the best parameters as follows: gamma and cost were 10^{-2} and 10^{3} in the fall peak season, and 10^{-3} and 10^{4} in the green leaf season, respectively. I then used these parameters for SVM learning and the comparative evaluation.

Table 2.5 Settings of the PyTorch image classification model.

Setting	Selected option
Training epochs	50–100
Batch size	16–64
Solver type	SGD
Base learning rate	0.1–0.001
Momentum	0.90

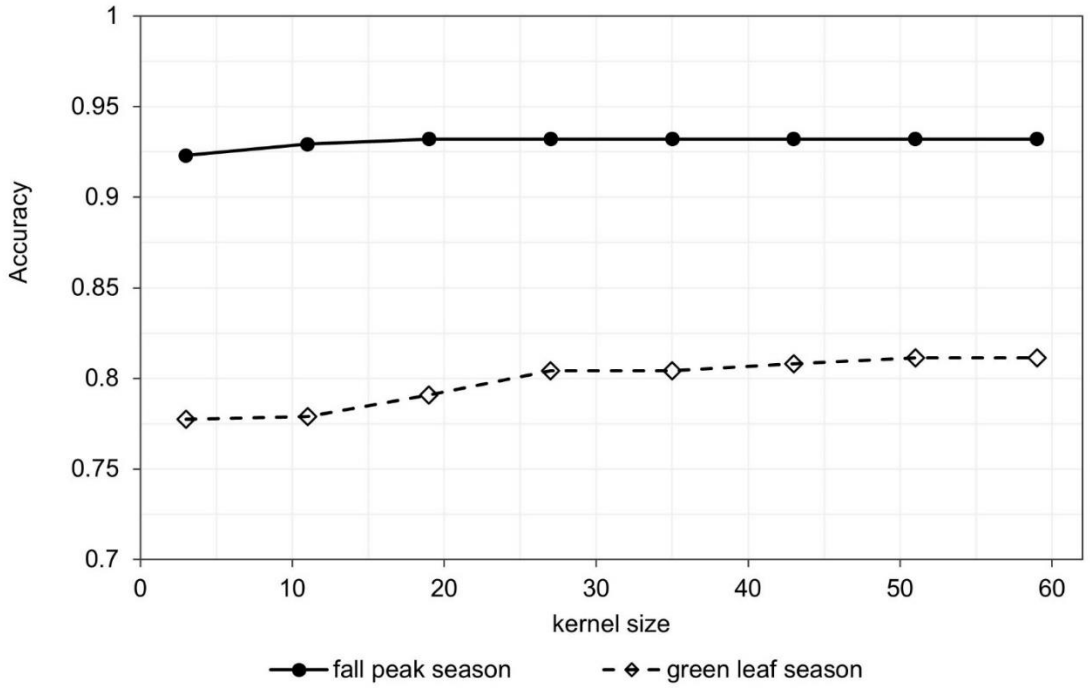


Figure 2.8 Relationship between GLCM kernel size and SVM validation accuracy in each season.

For machine learning, I divided the data into training, validation, and testing sets. The validation dataset was used for hyperparameters tuning such as learning rate, batch size for deep learning, and kernel size, cost, and gamma values for SVM. In the testing phase, I used the data which had not been used for training and parameter tuning. Validation accuracy is not suitable for comparing performance as a final result because validation accuracy can be higher than testing accuracy; I tuned the hyperparameters to get higher accuracy using the validation data. Using testing data, I can exclude the bias of parameter tuning. I also used a kind of cross-validation because I had a limited amount of data and decreased the contingency of accuracy. In this case, I randomly divided all the images evenly into four datasets and used two of them for training, one for validation, and one for testing. Subsequently, I interchanged successively the datasets used for training, validation, and testing. This process was repeated four times. For the accuracy evaluation and confusion matrix, I used total accuracy and all the images.

For this calculation, I used a built to order (BTO) desktop computer with a Xeon E5-2640 CPU, 32 GB RAM, and a Geforce GTX 1080 graphics card; the OS was Ubuntu 16.04.

2.2.5 Evaluation

For evaluation, I used the overall accuracy, Cohen's Kappa coefficient (Cohen, 1960), and the macro average F1 score. F1 score is the harmonic mean of Recall and Precision. In this study, the number of images acquired for each class varied significantly. The overall accuracy, which is typically utilised for evaluating the machine learning performance, is subject to the difference in the amount of data available to each class. Therefore, I used the Kappa and F1 score, which is suitable for evaluating imbalanced dataset accuracy, as well as overall accuracy to obtain an objective evaluation. Additionally, for evaluating the per-class accuracy, I used the F1 score of each class.

2.3 Results

2.3.1 Fall peak season

First, I made four standard neural network models and trained them on the data of the November 20 images. Although Resnet152 exhibited the highest performance, there was little difference among the accuracies of the neural network models (Table 2.6). I therefore adopted Resnet152 as a representative CNN model and denoted it as CNN Model 1.

The classification results of CNN Model 1 are shown in Figure 2.9; the overall accuracy was 0.976, Kappa 0.970, and F1 score 0.962. Model 1 succeeded in identifying almost all classes with more than 90% accuracy; one class identified with less than 90% accuracy was still identified at a relatively high accuracy greater than 85.0%. The confusion matrix shows the detailed results of model prediction. The vertical axis is the ground truth that I identified visually; the horizontal axis is the class that the model predicted. The number in each cell represents the number of classified images; each cell is coloured by the ratio of number of images per ground truth class. For example, a ratio of 0.0 (light blue) indicates that no image was classified to that cell, whereas a ratio of 1.0 (dark blue) indicates that all images of the ground truth were classified to that cell. From this confusion matrix, it can be seen that almost all classes were classified correctly.

I also tested the SVM method with the same images. The accuracy of the SVM was found to be lower than that of the CNN; however, it also exhibited a high performance (Figure 2.10). For the SVM, the overall accuracy, Kappa, and F1 score were 0.918, 0.896, and 0.857, respectively. These values are lower than the corresponding results of the CNN by approximately 6–10%; however, they are still relatively high. Most per-class accuracies are lower than those of the CNN results, especially Class 2, deciduous coniferous tree, which was difficult to identify. As can be seen in the comparison of the confusion matrix of the CNN results to that of the SVM results, misclassifications of the deciduous broad-leaved tree and deciduous coniferous tree, in particular, increased in the latter.

Table 2.6. Classification results of neural networks in fall peak season and green leaf season.

CNN model	Fall peak season			Green leaf season		
	Overall	Kappa	F1 score	Overall	Kappa	F1 score
AlexNet	0.960	0.949	0.930	0.902	0.875	0.836
VGG 16	0.973	0.965	0.958	0.925	0.903	0.878
Resnet 18	0.972	0.964	0.954	0.919	0.897	0.876
Resnet 152	0.976	0.970	0.962	0.933	0.914	0.901

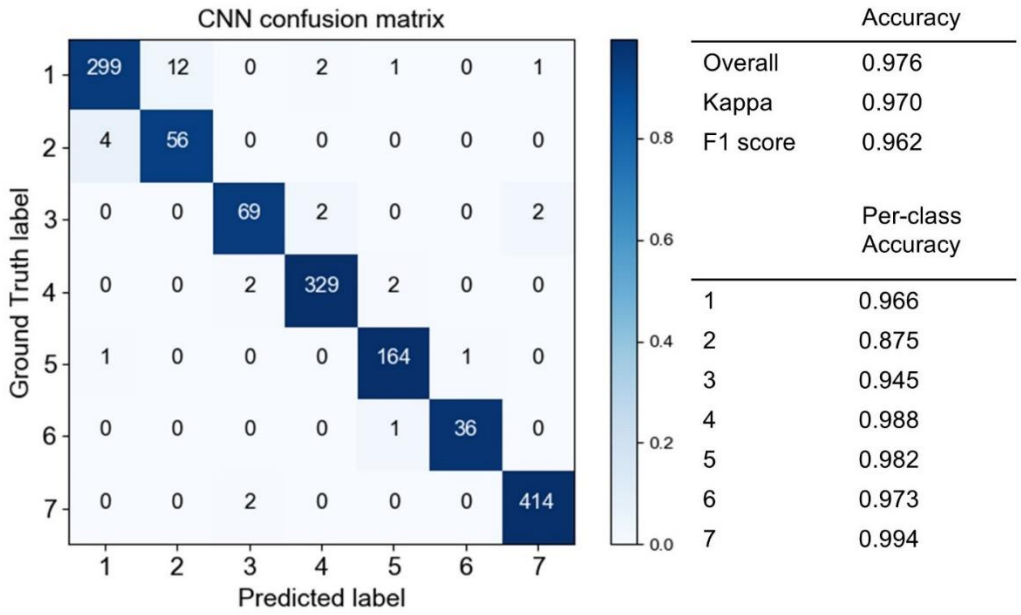


Figure 2.9. Confusion matrix of CNN in the fall season (Model 1). The vertical axis is the ground truth and the horizontal axis the model prediction. The number in each cell indicates the number of classified images; each cell is coloured according to the percentage of the number of images in each class. (Class 1: deciduous broad-leaved tree; 2: deciduous coniferous tree; 3: evergreen broad-leaved tree; 4: *Chamaecyparis obtusa*; 5: *Pinus elliottii* or *Pinus taeda*; 6: *Pinus strobus*; 7: non-forest.)

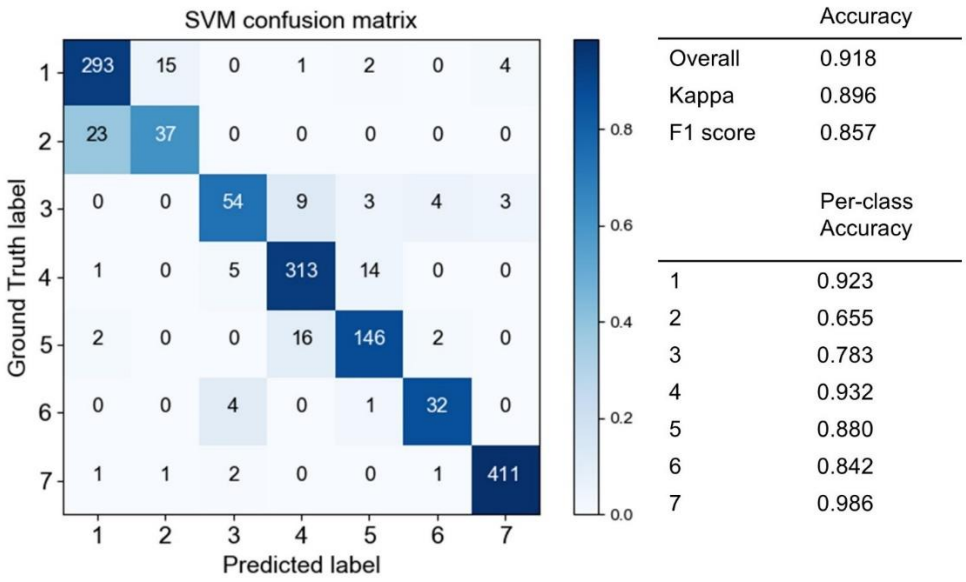


Figure 2.10. Confusion matrix of SVM in the fall season. (Class 1: deciduous broad-leaved tree; 2: deciduous coniferous tree; 3: evergreen broad-leaved tree; 4: *Chamaecyparis obtusa*; 5: *Pinus elliottii* or *Pinus taeda*; 6: *Pinus strobus*; 7: non-forest)

2.3.2 Green leaf season

Next, I constructed four standard neural network models and trained them on the data from the end of the green leaf season. As in the fall peak season, Resnet152 exhibited the highest performance, but the differences in the performance of the models were greater than for that season (Table 2.6). I therefore adopted Resnet152 as a representative CNN model and denoted it as CNN Model 2. I expected that in this season it would be more difficult to identify deciduous classes because the colours of each class are similar. However, Model 2showed slight decrease in performance . The performance of Model 2 is shown in Figure 11. The overall accuracy, Kappa, and F1 score were 0.933, 0.914, and 0.901, respectively. Per-class accuracies are also high, at approximately 90% in most classes. Compared to misclassifications in the fall season, the misclassification of deciduous broad-leaved trees and evergreen broad-leaved trees in the green leaf season increased.

I also tested the SVM's performance with Model 2, wherein the SVM had difficulty identifying these classes (Figure 2.12). The overall accuracy decreased to 0.700–0.803, with per-class accuracy of approximately 70–95% for some classes, 48% for the *Pinus strobus* class, and 22% for the evergreen broad-leaved tree class. According to the confusion matrix, the evergreen broad-leaved tree class is mostly misclassified as a deciduous broad-leaved tree, whereas the *Pinus strobus* is mostly misclassified as the other *Pinus* class and the deciduous broad-leaved tree class.

Finally, I visualised the classification results of CNN Model 1 in GIS (Figure 2.13). Most of the object images in this figure were used for training; taking this into account, I can see that the CNN succeeded in correctly identifying the trees.

Consequently, I applied Guided Grad-CAM for generating an attention map to visualise the features that the CNN used for classification. Guided Grad-CAM can visualise locations and detailed features which are related to the judged class (Selvaraju et al., 2016). I applied the algorithms for the classes except for the non-forest class, to layer 4, the last layer of Model 2. Figure 2.14 shows the original image, the attention map visualised by Guided Grad-CAM, and the grayscale original image overlaid on the highlighted attention map of each class. For deciduous broad-leaved tree, it is difficult to explain clearly what region CNN used. For deciduous coniferous tree, the features are extracted on the whole area, and their locations look to be matched to contrast of edge of the small foliage. For the evergreen broad-leaved tree, though the colour of features is pale, I can see the highlighted features along the edge of the bush of the branch compared to deciduous broad-leaved tree. For *Chamaecyparis obtusa,* I can see the features along the edge of the hierarchical branch clearly. For *Pinus elliottii* or *Pinus taeda* pictures, the features match the edge of the branch. For *Pinus strobus*, the outlines of the trees are highlighted.

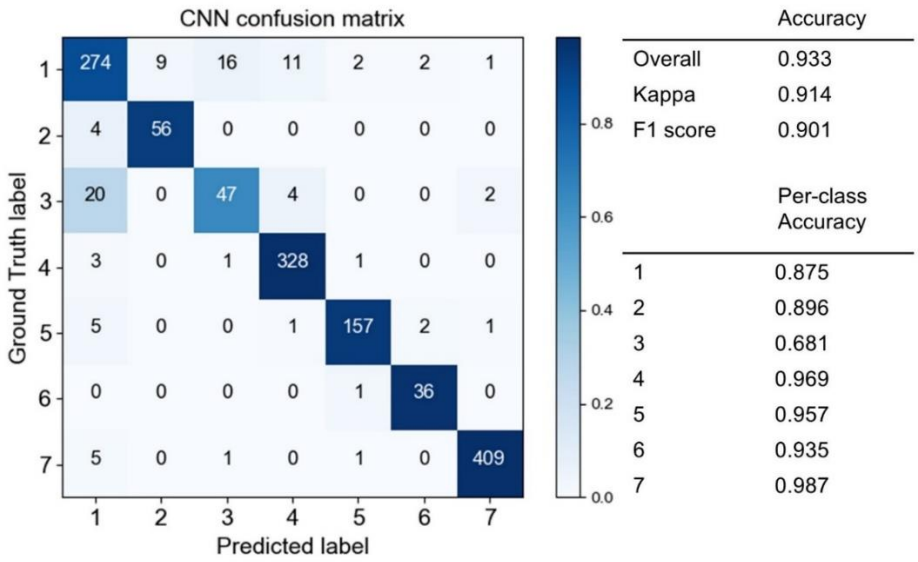


Figure 2.11. Confusion matrix of the CNN in the leaf season (Model 2). (Class 1: deciduous broad-leaved tree; 2: deciduous coniferous tree; 3: evergreen broad-leaved tree; 4: *Chamaecyparis obtusa*; 5: *Pinus elliottii* or *Pinus taeda*; 6: *Pinus strobus*; 7: non-forest.)

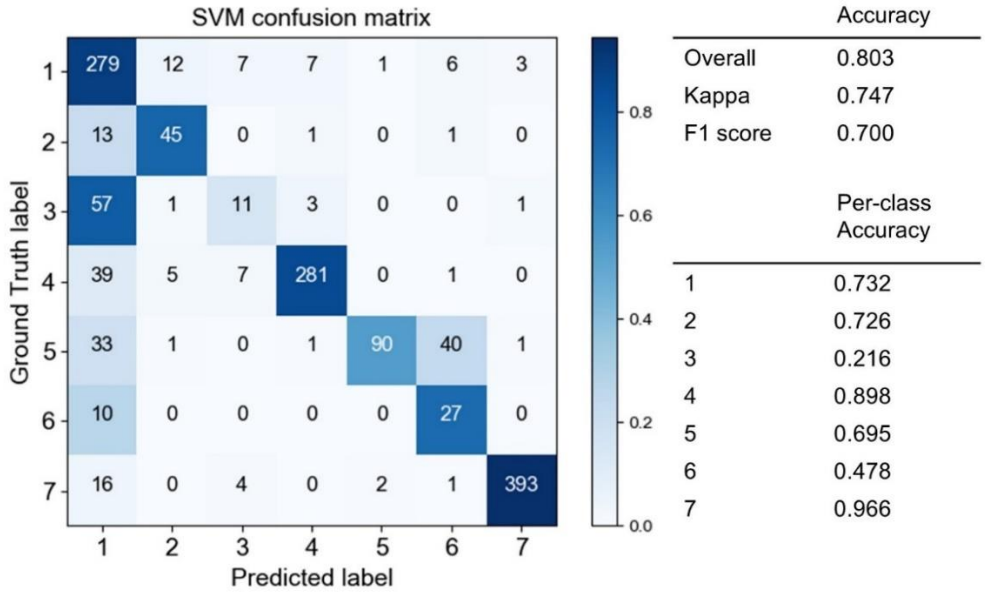


Figure 2.12. Confusion matrix of SVM in the green leaf season. (Class 1: deciduous broad-leaved tree; 2: deciduous coniferous tree; 3: evergreen broad-leaved tree; 4: *Chamaecyparis obtusa*; 5: *Pinus elliottii* or *Pinus taeda*; 6: *Pinus strobus*; 7: non-forest)

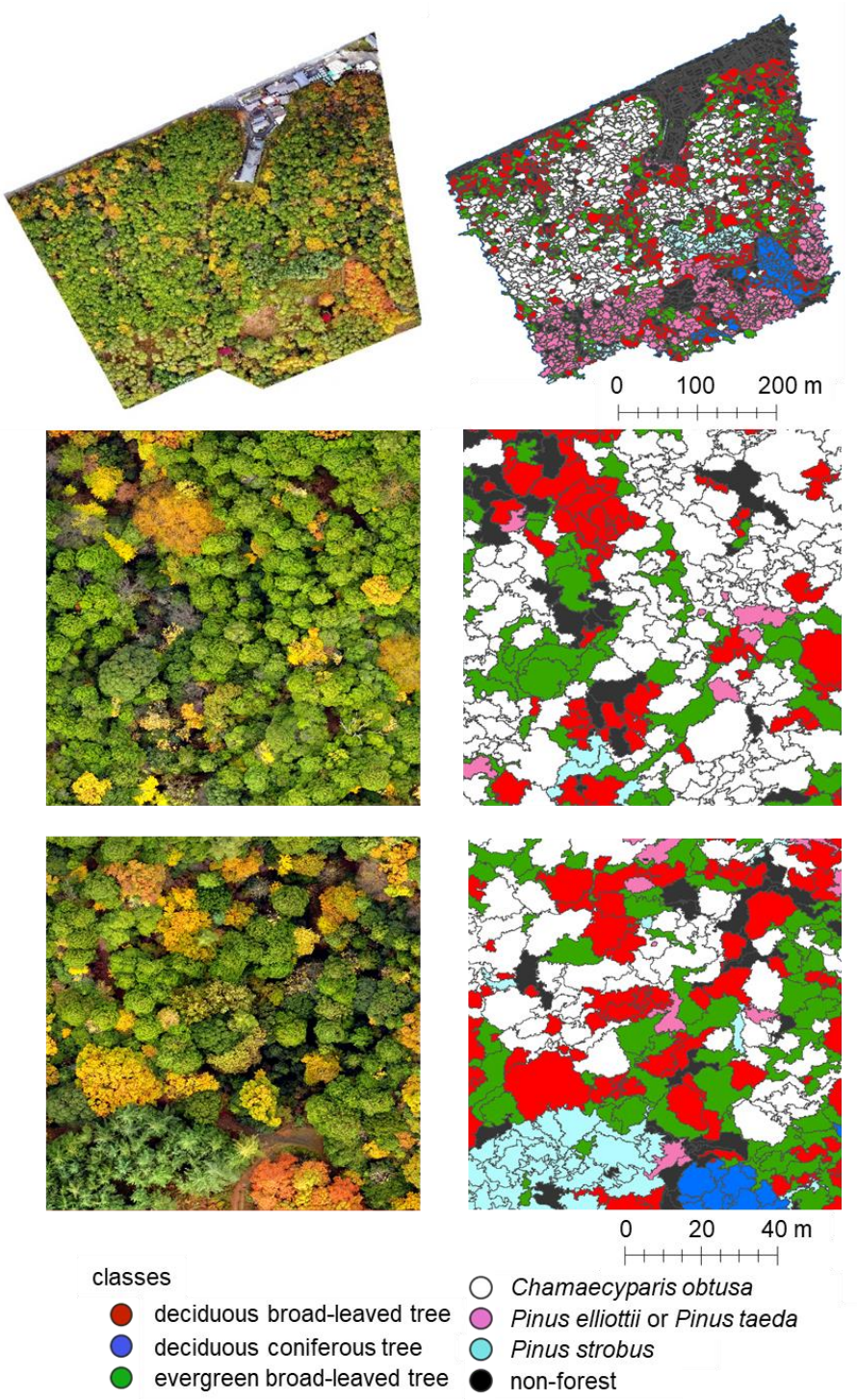


Figure 2.13. Orthomosaic photos and classification maps obtained with the CNN classifier. The topmost images each gives an overall view of one area, and the lower images in each column show the respective enlarged area. The tree classes found in the images are identified in the legend at the bottom of the figure.

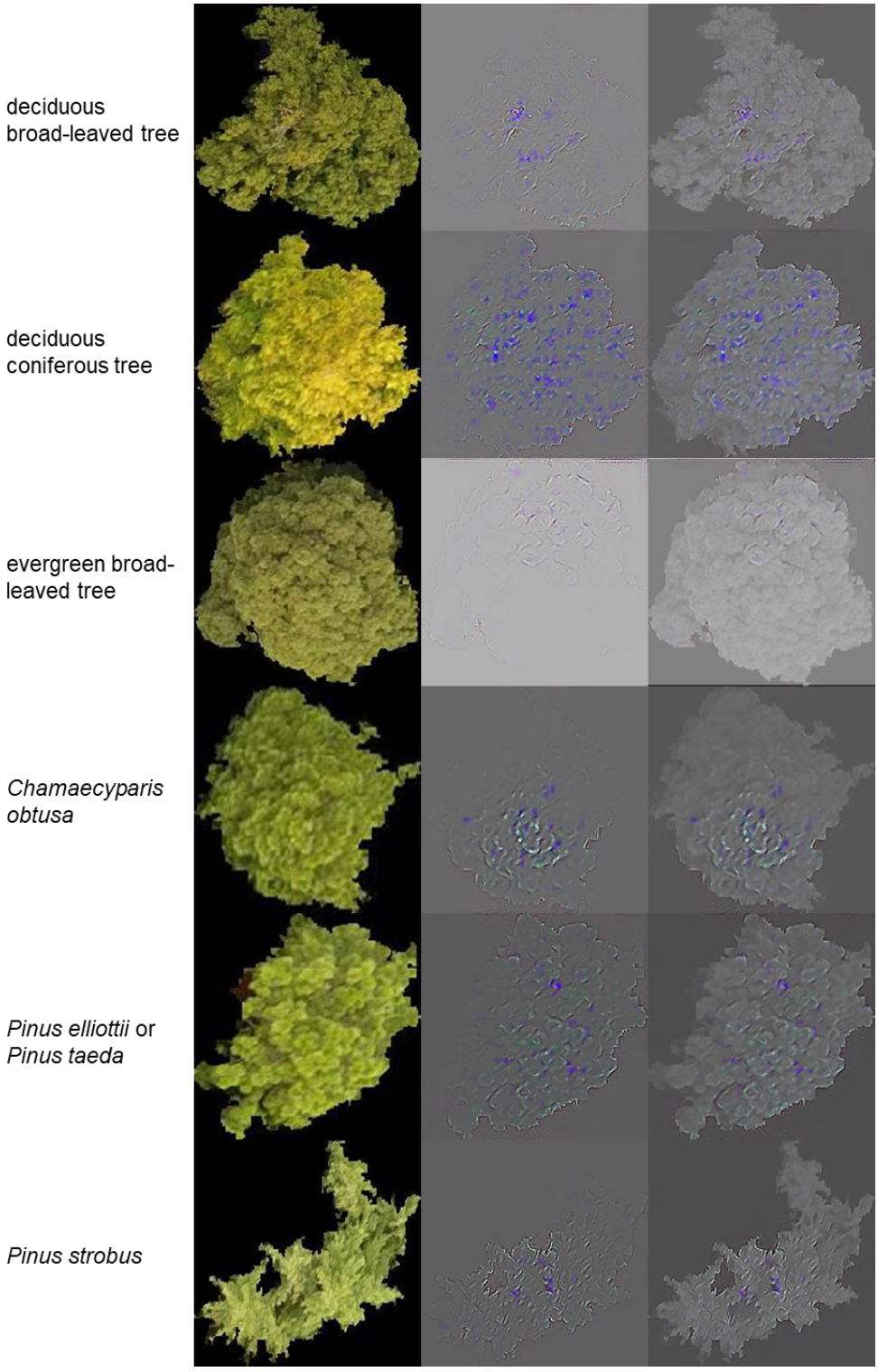


Figure 2.14. Attention maps of Guided Grad-CAM. (Left to right) class name, original image, result of Guided Grad-CAM, result of Guided Grad-CAM overlaid with the grayscale original image.

2.4 Discussion

This study was conducted to investigate the ability of my machine vision system, which combines UAV RGB image and deep learning, to classify individual trees into several tree types and identify specific tree species. Regarding the classification of tree types, Model 1 (pertaining to the fall season) classified four tree types and one non-forest class with Kappa 0.971 and F1 score 0.955, whereas Model 2 classified them with Kappa 0.911 and F1 score 0.882. At the species level, Models 1 and 2 were able to classify *Chamaecyparis obtusa*, *Pinus strobus* from *Pinus elliottii* and *Pinus taeda* with more than 90 % accuracy. Thus, my system was able to classify tree types and has the potential to classify several tree species in some seasons. This performance is notable because I used easily available digital RGB images only. Contrastingly, most previous studies used expensive hardware, such as multispectral imagers, to improve performance. Regarding the spatial scale, my method, using a UAV, may be more limited than previous methods using airborne sensors. However, the low-cost and easy-to-use feature of UAVs can enable periodic monitoring. Meanwhile, other researches using UAVs identified only a few tree species or applied rectangle object detection (Csillik et al., 2018; dos Santos et al., 2019; Natesan et al., 2019; Safonova et al., 2019). Regarding these points, my system successfully identified several tree classes, and enabled forest mapping. Thus, my machine vision system can be used as a cost-effective and handy tool for application in forest mapping and management.

Comparing the CNN performance to that of the SVM, in the fall peak season, both the SVM and the CNN demonstrated a high performance. However, in the green leaf season, the CNN exhibited a much higher performance than that of the SVM. I used average and standard deviation values of RGB bands and GLCM texture features for SVM learning; thus, this difference in features may have contributed to the performance difference between the SVM and the CNN. This result also indicates that the CNN used more features other than the average and standard deviation values of RGB bands and GLCM texture features.

Applying Guided Grad-CAM, I succeeded in visualising features that CNN used. The results of Guided Grad-CAM suggest that the CNN model used the difference in edge shape of foliage and bush of branch, hierarchical branching pattern and outlines of tree shapes for classification. From these considerations, it is supported that CNN identified trees based on the features of biological structures that represent physiological and ecological characteristics of the target tree classes. This result also indicates that the CNN with UAV image can successfully identify trees according to the differences in structural features, even if trees have similar colours. Considering this fact, high spatial and temporal robustness of this system is expected because if the leaf colour changes owing to factors such as weather or season, it can still be classified correctly using the structural features.

In my method, there are two reasons for the extremely successful result. First, I conducted object-based classification. Previous studies have shown that object-based classification can obtain higher accuracy than pixel-based classification (Immitzer et al., 2012; Tarabalka et al., 2010). One advantage of applying object segmentation is that each image can have more common features. For example, judging from the image of *Chamaecyparis obtusa* in Figure 2.14, CNN used the features around the treetop. Using object segmentation, most images can include this feature. However, if I use pixel-based segmentation, each image would be cut by a rectangular image such as 64×64 pixel. Therefore, the possibility that the image contains those common features will be low and the accuracy of classification may be lower. My method could not segment every tree crown perfectly; thus, improving the segmentation method would lead to higher classification accuracy and also enable us to count the number of trees. It is difficult to segment tree crowns using a specific parameter because the tree size and clarity of the treetop are highly diversified, especially in a mixed forest. In future work, applying a deep learning method such as Instance Segmentation (Dai et al., 2016) can have high potential for tree crown segmentation.

Second, I selected training and testing images from the same area and the same time. Tree shapes have some variation in different environments, and leaf colours and illuminations are

different across periods, weather types, and seasons. Utilisation of tree shapes (or DSM) and the seasonality of leaf colours would improve classification accuracy; however, generally, these properties may have a negative influence on simple machine learning. Considering practicability, a versatile model which involves trained images from various sites and times is desired in further study.

Although I achieved high accuracy classification by using the CNN, I still had some misclassification. Notably, Model 2 struggled to identify deciduous broad-leaved trees and evergreen broad-leaved trees. This misclassification may be attributed to the class inclusion of several tree species, i.e. the branch pattern and colour may not be common across the same tree type. Therefore, separating the classes into more fine-grained branching types or tree species levels may improve the classification accuracy. Thus, the way in which I separated each class may be one of the key parameters affecting classification performance.

Ultimately, I constructed a chain of machine vision systems that can segment and identify trees automatically. Although the classes are limited, this method can be utilised as a base system for tree mapping systems using UAV images.

2.5 Conclusion

In this study, I proposed a chain of low-cost machine vision system for identifying trees using UAV RGB image and deep learning. My system achieved an accuracy of more than 90% for classifying tree types and specific tree species. Additionally, the results suggest that the CNN classified trees according to the features of their biological structures such as foliage shapes, and branching pattern; thus, this system may have the potential for identifying several tree species even when the colours of the trees are similar. In a follow-up study, tree crown segmentation using deep learning needs to be conducted, and I are planning to identify more detailed species and also evaluate the spatial and temporal robustness of the developed system.

2.6 References

Baatz, M. & Schäpe, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. *XII Angew. Geogr. Informationsverarbeitung, Wichmann-Verlag, Heidelb.* (2000).

Boschetti, M., Boschetti, L., Oliveri, S., Casati, L. & Canova, I. Tree species mapping with airborne hyper-spectral MIVIS data: The Ticino Park study case. *Int. J. Remote Sens.* **28**, 1251–1261 (2007).

Cohen, J. A coeffient of agreement for nominal scales. *Educ. Psychol. Meas.* **XX**, 37–46 (1960).

Dai, J., He, K. & Sun, J. Instance-aware semantic segmentation via multi-task network cascades. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, Las Vegas, NV, USA, 27–30 June* 3150–3158 (2016) doi:10.1109/CVPR.2016.343.

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.* **123**, 258–270 (2012).

Dalponte, M., Ørka, H. O., Gobakken, T., Gianelle, D. & Næsset, E. Tree species classification in boreal forests with hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **51**, 2632–2645 (2013).

Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. *IEEE Conf. Comput. Vis. pattern Recognit.* 248–255 (2009).

Dian, Y., Li, Z. & Pang, Y. Spectral and texture features combined for rorest tree species classification with airborne hyperspectral imagery. *J. Indian Soc. Remote Sens.* **43**, 101–107 (2015).

dos Santos, A. A. *et al.* Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. *Sensors (Switzerland)* **19**, (2019).

European Environmental Agency. *European Forest Types. Categories and types for sustainable forest management reporting and policy*. *EEA Technical report No9/2006* https://www.eea.europa.eu/publications/technical_report_2006_9 (2006) doi:10.3832/efor0425-003.

Fassnacht, F. E. *et al.* Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **186**, 64–87 (2016).

Franklin, S. E., Wulder, M. A. & Lavigne, M. B. Automated derivation of geographic window sizes for use in remote sensing digital image texture analysis. *Comput. Geosci.* **22**, 665–673 (1996).

Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 580–587 (2014) doi:10.1109/CVPR.2014.81.

Goodbody, T. R. H., Coops, N. C., Marshall, P. L., Tompalski, P. & Crawford, P. Unmanned aerial systems for precision forest inventory purposes: A review and case study. *For. Chron.* **93**, 71–81 (2017).

Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 610–621 (1973).

He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 770–778 (2016) doi:10.1109/CVPR.2016.90.

Iizuka, K., Yonehara, T., Itoh, M. & Kosugi, Y. Estimating tree height and diameter at breast height (DBH) from digital surface models and orthophotos obtained with an unmanned aerial system for a Japanese Cypress (Chamaecyparis obtusa) Forest. *Remote Sens.* **10**, (2018).

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.* **4**, 2661–2693 (2012).

Ise, T., Minagawa, M. & Onishi, M. Classifying 3 moss species by deep learning, using the "chopped picture" method. *Open J. Ecol.* **08**, 166–173 (2018).

Jansson, G. & Angelstam, P. Threshold levels of habitat composition for the presence of the long-tailed tit (Aegithalos caudatus) in a boreal landscape. *Landsc. Ecol.* **14**, 283–290 (1999).

Jing, L., Hu, B., Noland, T. & Li, J. An individual tree crown delineation method based on multi-scale segmentation of imagery. *J. Photogramm. Remote Sens.* **70**, 88–98 (2012).

Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 1097–1105 (2012) doi:http://dx.doi.org/10.1016/j.protcy.2014.09.007.

Lim, J., Kim, K. M. & Jin, R. Tree species classification using hyperion and sentinel-2 data with machine learning in South Korea and China. *Int. J. Geo-Information* **8**, 1–23 (2019).

Louarn, M. Le, Clergeau, P., Briche, E. & Deschamps-Cottin, M. 'Kill two birds with one stone': Urban tree species classification using bi-temporal pléiades images to study nesting preferences of an invasive bird. *Remote Sens.* **9**, (2017).

Machala, M. & Zejdová, L. Forest mapping through object-based image analysis of multispectral and LiDAR aerial data. *Eur. J. Remote Sens.* **47**, 117–131 (2014).

Marceau, D. J., Howarth, P. J., Dubois, J.-M. M. & Gratton, D. J. Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery. *IEEE Trans. Geosci. Remote Sens.* **28**, 513–519 (1990).

Mlambo, R., Woodhouse, I. H., Gerard, F. & Anderson, K. Structure from motion (SfM) photogrammetry with drone data: A low cost method for monitoring greenhouse gas emissions from forests in developing countries. *Forests* **8**, (2017).

Narayanan, R. M., Sankaravadivelu, T. & Reichenbach, S. E. Dependence of image information content on gray-scale resolution. *Proc. IGARSS* **1**, 153–155 (2000).

Natesan, S., Armenakis, C. & Vepakomma, U. Resnet-based tree species classification using uav images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLII**, 475–481 (2019).

Paneque-Gálvez, J., McCall, M. K., Napoletano, B. M., Wich, S. A. & Koh, L. P. Small drones for community-based forest monitoring: An assessment of their feasibility and potential in tropical areas. *Forests* **5**, 1481–1507 (2014).

Paszke, A. *et al.* Automatic differentiation in PyTorch. *31st Conf. Neural Inf. Process. Syst. Long Beach, CA, USA* (2017) doi:10.1145/24680.24681.

Safonova, A. *et al.* Detection of fir trees (Abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sens.* **11**, (2019).

Salovaara, K. J., Thessler, S., Malik, R. N. & Tuomisto, H. Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sens. Environ.* **97**, 39–51 (2005).

Selvaraju, R. R. *et al.* Grad-CAM: visual explanations from deep networks via gradient-based localization. *arXiv:1610.02391* (2016) doi:10.1007/s11263-019-01228-7.

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.* **7**, 2481–2489 (2014).

Shen, X. & Cao, L. Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. *Remote Sens.* **9**, (2017).

Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. (2014).

Szegedy, C. *et al.* Going deeper with convolutions. in *Preprint at http://arxiv.org/abs/1409.4842.* (2014). doi:10.1109/CVPR.2015.7298594.

Tang, L. & Shao, G. Drone remote sensing for forestry research and practices. *J. For. Res.* **26**, 791–797 (2015).

Tarabalka, Y., Fauvel, M., Chanussot, J. & Benediktsson, J. A. SVM- and MRF-based method for accurate classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **7**, 736–740 (2010).

Treuhaft, R. N., Asner, G. P., Law, B. E. & Van Tuyl, S. Forest leaf area density profiles from the quantitative fusion of radar and hyperspectral data. *J. Geophys. Res.* **107**, (2002).

Tsai, F. & Chou, M. J. Texture augmented analysis of high resolution satellite imagery in detecting invasive plant species. *J. Chinese Inst. Eng.* **29**, 581–592 (2006).

van Lier, O. R., Fournier, R. A., Bradley, R. L. & Thiffault, N. A multi-resolution satellite imagery approach for large area mapping of ericaceous shrubs in Northern Quebec, Canada. *Int. J. Appl. Earth Obs. Geoinf.* **11**, 334–343 (2009).

Wang, L., Sousa, W. P., Gong, P. & Biging, G. S. Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama. *Remote Sens. Environ.* **91**, 432–440 (2004).

Waser, L. T., Küchler, M., Jütte, K. & Stampfer, T. Evaluating the potential of worldview-2 data to classify tree species and different levels of ash mortality. *Remote Sens.* **6**, 4515–4545 (2014).

Xu, B., Gong, P., Seto, E. & Spear, R. Comparison of gray-level reduction and different texture spectrum encoding methods for land-use classification using a panchromatic Ikonos image. *Photogramm. Eng. Remote Sensing* **69**, 529–536 (2003).

CHAPTER 3
Practicality and Robustness of Tree Species Identification Using UAV RGB Image and Deep Learning in Temperate Forest in Japan

3.1 Introduction

Recently, tree species identification from aircraft has drawn attention because of its success in high-performance classification (Fassnacht et al., 2018). From the aircraft, especially the use of hyperspectral sensors exhibited superior performance (Dalponte et al., 2012; Fassnacht et al., 2018; Shen and Cao, 2017). Although a general digital camera can capture the spectra of three bands of red, green, and blue within visible areas, multispectral cameras can capture 4 to 8 bands, including near infrared (NIR) spectra, and hyperspectral cameras can catch from 100 to 200 bands, including NIR and short-wave infrared (SWIR). The reason for using these spectra is valid for identification because various spectra reflecting tree crowns have a relationship with ① water contained in the woody tissue, photosynthetic pigment, and characteristics of structural carbohydrates (Asner, 1998; Clark and Roberts, 2012; Gao and Hoetz, 1990; Knipling, 1970; Ustin et all., 2009), ② morphology of leaves (Asner, 1998; Clark et al., 2005; Grant, 1987) and ③ tree crown structures (Leckie et al., 2005). To apply machine learning, hundreds or thousands of spectra sensed using hyperspectral sensors were compressed dimensionally by PCA, and by applying machine learning such as random forest or SVM, several tree species can be identified with high accuracy of approximately 80%–90% (Dalponte et al., 2012; Shen and Cao, 2017). Another approach is LiDAR. LiDAR is sometimes used for tree species identification. Even by itself, certain characteristic tree species, for example, *Cryptomeria japonica* and *Chamaecyparis obtusa* are detected using features such as reflectance strength and crown shapes (Nakatake et al., 2018). The combination of hyperspectral sensors and LiDAR showed the highest accuracy for the identification of several species (Dalponte et al., 2012; Shen and Cao, 2017). The advantage of using airplanes is that they can scan large areas, such as prefecture scales. The disadvantages are 1) they are costly, and 2) the system is easily affected by several factors. With respect to the first

disadvantage, it takes dozens of millions of yen per scan; therefore, only huge projects or administrations can apply the method on rare occasions. With respect to the second disadvantage, because the system relies on spectral information, it can be affected by shadows owing to the weather (Asner and Warner, 2003; Dalponte et al., 2012; Shen and Cao, 2017). Furthermore, background signals attributed to the conditions of vegetation, soil and litter, density of leaves, and health of trees can affect performance (Clark and Roberts, 2012; Roberts et al., 1997; Treuhaft et al., 2002; Waser et al. 2014).

For practical use, the temporal and spatial robustness should be considered. However, these studies were conducted using the data that were obtained simultaneously, and most of them were obtained from one site (Ballanti et al., 2016; Boschetti et al., 2005; Cao et al., 2016; Clark et al., 2012; Dalponte et al., 2009; Dalponte et al., 2012; Dalponte et al., 2013; Féret and Asner, 2013; Machala and Zejdová, 2014; Shen and Cao, 2017; Zhang et al., 2016;). The accuracy obtained by separating the data into training and test may be highly biased because the light condition and the state of vegetation, such as tree sizes or understory vegetation, is almost the same between datasets. Therefore, the performance of these models will be lower when they are applied on another day or place.

Recently, UAVs have been used as a framework for tree identification. Certain studies load hyperspectral sensors for identifying trees (Navelalainen et al., 2017; Sothe et al., 2019). On the other hand, certain researchers have succeeded in identifying tree species from general digital cameras with the help of deep learning (Natesan et al., 2019; Onishi and Ise, 2021; Safonova et al., 2019). Owing to the high resolution of images and deep learning, specific tree species detection, classification, and mapping of several tree species or tree types have been reported. The advantages of this system are its low cost and the potential for robustness. First, Onishi and Ise (2021) successfully identified several tree classes using phantom 4 (DJI), which is a famous commercial UAV, and costs less than 3000 US dollars. Owing to their low cost, data can be obtained frequently. Second, the system may have the potential for robustness. This means that the model that was trained at one place has the potential to be used on another day or place. This

is because this system does not rely on the spectral reflectance information. With the help of high-resolution images and deep learning, trees are identified using the texture images that are generated by leaf shapes and branching patterns (Onishi and Ise, 2021). Therefore, if the location or timing of the data is different from that of the training dataset, the model is expected to perform well. The disadvantage is that UAVs can only cover limited areas compared to the airplane. General multi-copter UAVs such as phantoms can fly for less than 40 min, and only approximately 10 ha can be scanned in one flight.

For the practical use of tree species identification from UAVs, I should explore their potential and robustness. Potential refers to the limit of the amount or kinds of tree species that can be identified. Although the UAV and deep learning can identify trees based on texture information, the texture may not be different for every tree species. In addition, the spatial and temporal robustness of the model should be assessed for general usage.

In this study, I used general UAVs and deep learning, collected large amounts of tree species data from temperate forests in Japan, and tested the robustness of this identification system to the difference in shooting date, individual tree and site.

3.2 Materials and Methods

3.2.1 Study sites

This study was conducted at six sites in a temperate forest in Japan. I used three sites for training and validation, and three sites for testing; the sites are outlined below. (Figure 3.1, Table 3.1).

- Higashiyama site: located in (34° 59' 58" N, 135° 47' 17″ E), is a secondary forest of warm-temperate forest, dominated by deciduous broad-leaved trees such as *Quercus serrata*, and laurel evergreen broad-leaved trees *such as Castanopsis cuspidate*.
- Wakayama site: located in Wakayama Forest Research Station of Kyoto University (34° 03' 47″ N, 135° 31' 00″ E), is a natural forest of mid-temperate forest dominated by evergreen coniferous trees such as *Abies firma* and *Tsuga sieboldii.*
- Ashiu site: located at Yusen Valley in Ashiu Forest Research Station of Kyoto University (35° 18' 34″ N, 135° 43' 1″ E) is a natural forest of cold temperate forest dominated by deciduous broad-leaved trees such as *Fagus crenata*.
- Kamigamo site: located at Kamigamo Experimental Station of Kyoto University (35° 04' 00″ N, 135° 46' 01″ E) is a natural regeneration forest of warm-temperate forests dominated by evergreen coniferous trees such as *Chamaecyparis obtusa* and broad-leaved trees such as *Quercus serrata*.
- Kasugayama site: located in the Kasugayama Primeval Forest (34° 41' 14″ N, 135° 51' 24″ E) is a primeval forest of mid-temperate forest dominated by evergreen coniferous trees such as *Abies firma* and laurel evergreen broad-leaved trees *such as Castanopsis cuspidate*.
- Daisen site: located in (35°21' 32″ N, 133°33' 17″ E) is a naturally generated forest of cold-temperate forest dominated by deciduous broad-leaved trees such as *Fagus crenata*.

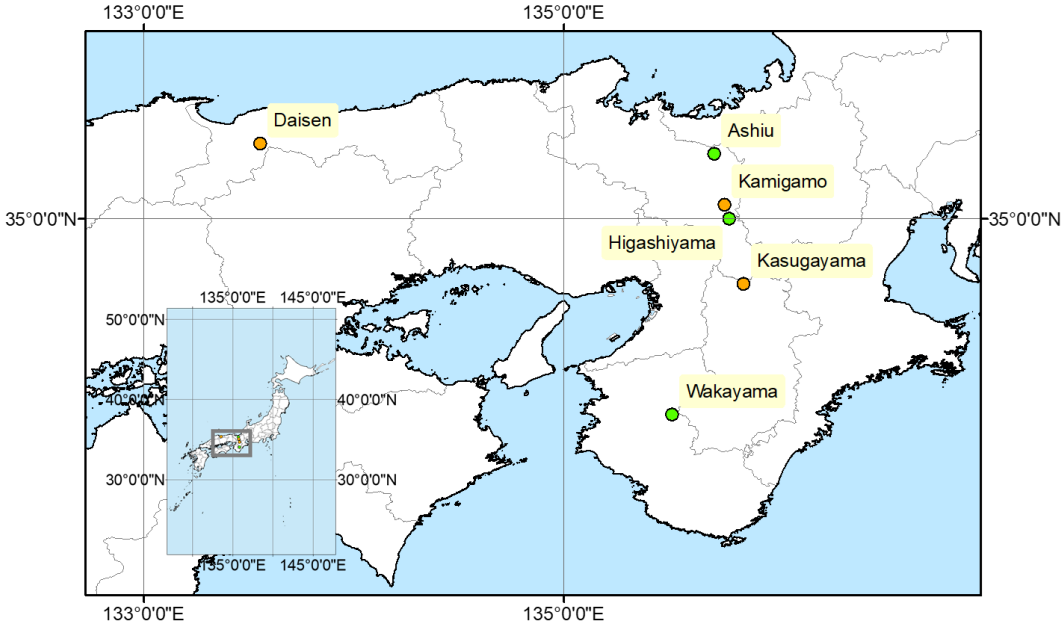


Figure 3.1 Location of each site. Green and orange points represent the training and validation sites, and test sites, respectively.

Table 3.1 Site information summary

Site name	Higashiyama	Wakayama	Ashiu
Data class	Training & Validation	Training & Validation	Training & Validation
Field research area	2ha	1ha	4ha
Flight date	July 4, 5, Sept. 14, & 15, 2019	July 15, 17, & Oct. 8, 2019	July 24, 25, 26 & Sep.18, 19, 20, 2019
dominant species	*Castanopsis cuspidate* *Quercus serrata*	*Abies firma*, *Tsuga sieboldii*	*Fagus crenata*

Site name	Kamigamo	Kasugayama	Daisen
Data class	Test	Test	Test
Field research area	1.5ha	1ha	1ha
Flight date	Oct. 10, 2019	Sept. 24, 2019	Oct. 1, 2019
dominant species	*Chamaecyparis obtusa* *Quercus serrata*	*Castanopsis cuspidate* *Abies firma*	*Fagus crenata*

3.2.2 UAV flight

I flew a UAV in the summer of 2019. At the training site, I flew the UAV several times to obtain the training data that looked different owing to the influence of light and weather conditions. At the test site, the UAV was flown once. At both sites, flight timing was set to 1 h before and after noon to avoid shadows, and I mainly used data obtained on a cloudy day. During the flight, I used Phantom 4 pro (DJI, Shenzhen, China) with a 20 million pixel camera, and UgCS v3.2 software (SPH engineering, Baložu pilsēta, Latvia) for automatic flight. The flight parameters were set as follows: the flight altitude was 100 m from the 10 m resolution DEM, which was provided by the Geospatial Information Authority of Japan. The front overlap was 90%, and the side overlap was 80%. The image format was set as RAW, exposure was set to 0.0, and other camera parameters were set automatically.

3.2.3 Field survey

I conducted a field survey in 2019, excluding the Kasugayama site in 2020. In the field survey, I identified tree species using leaves and attached the name as a label to all tree crown images in orthomosaic photos at the objective area obtained by the UAV. I identified 56 species and other two classes from the three training sites (Table 3.2).

3.2.4 UAV data processing

The UAV data processing method is illustrated in Figure 3.2. First, from UAV imagery, I created an orthomosaic photo and digital surface model (DSM) using Metashape software (Agisoft LLC, St. Petersburg, Russia). From the DSM, I created a slope model, calculated the slope using ArcGIS Desktop v10.6 software (Environmental Systems Research Institute, Inc., Redlands, United States), and orthomosaic photo, DSM, and slope, and applied multiresolution segmentation (Baatz and Schäpe, 2000) in eCognition Developer v9.0.0 software (Trimble, Inc., Sunnyvale, United States), and obtained tree crown polygons. Then, I attached species labels to each tree crown polygon using the field survey results. After that, I applied ExtractByMask tool

in ArcGIS to extract each tree crown image. For the training and validation dataset, I made orthomosaic photos from each flight, reduced the GPS error of each image by georeferencing, and used every orthomosaic photos for image extraction. From the UAV data processing above, I obtained the supervised data (Figure 3.3). For the training and validation dataset, the number of images was the same as the number of polygons multiplied by the number of flights.

Table 3.2 Tree species list of training data which identified in UAV imagery from field survey

class	Species name	class	Species name
1	*Chamaecyparis obtusa*	30	*Quercus serrata*
2	*Cryptomeria japonica*	31	*Pterocarya rhoifolia*
3	*Abies firma*	32	*Cinnamomum camphora*
4	*Pinus densiflora*	33	*Magnolia obovata*
5	*Tsuga sieboldii*	34	*Magnolia salicifolia*
6	*Ilex chinensis*	35	*Morella rubra*
7	*Ilex latifolia*	36	*Fraxinus lanuginosa f. serrata*
8	*Ilex macropoda*	37	*Ternstroemia gymnanthera*
9	*Ilex micrococca*	38	*Hovenia dulcis*
10	*Ilex pedunculosa*	39	*Hovenia tomentella*
11	*Chengiopanax sciadophylloides*	40	*Aria alnifolia*
12	*Evodiopanax innovans*	41	*Aria japonica*
13	*Kalopanax septemlobus*	42	*Malus tschonoskii*
14	*Betura grossa*	43	*Prunus grayana*
15	*Carpinus cordata*	44	*Prunus jamasakura*
16	*Carpinus japonica*	45	*Meliosma Myriantha*
17	*Carpinus laxiflora*	46	*Populus tremula var. sieboldii*
18	*Carpinus tschonoskii*	47	*Acer carpinifolium*
19	*Ostrya japonica*	48	*Acer mono Maxim*
20	*Cercidiphyllum japonicum*	49	*Acer nipponicum*
21	*Lyonia ovalifolia var.elliptica*	50	*Acer palmatum*
22	*Castanea crenata*	51	*Acer palmatum var. amoenum*
23	*Castanopsis cuspidata*	52	*Acer sieboldianum*
24	*Fagus crenata*	53	*Aesculus turbinata*
25	*Fagus japonica*	54	*Symplocos prunifolia*
26	*Quercus acuta*	55	*Stewartia monadelpha*
27	*Quercus crispula*	56	*Zelkova serrata*
28	*Quercus glauca*	57	dead_tree
29	*Quercus salicina*	58	Gap

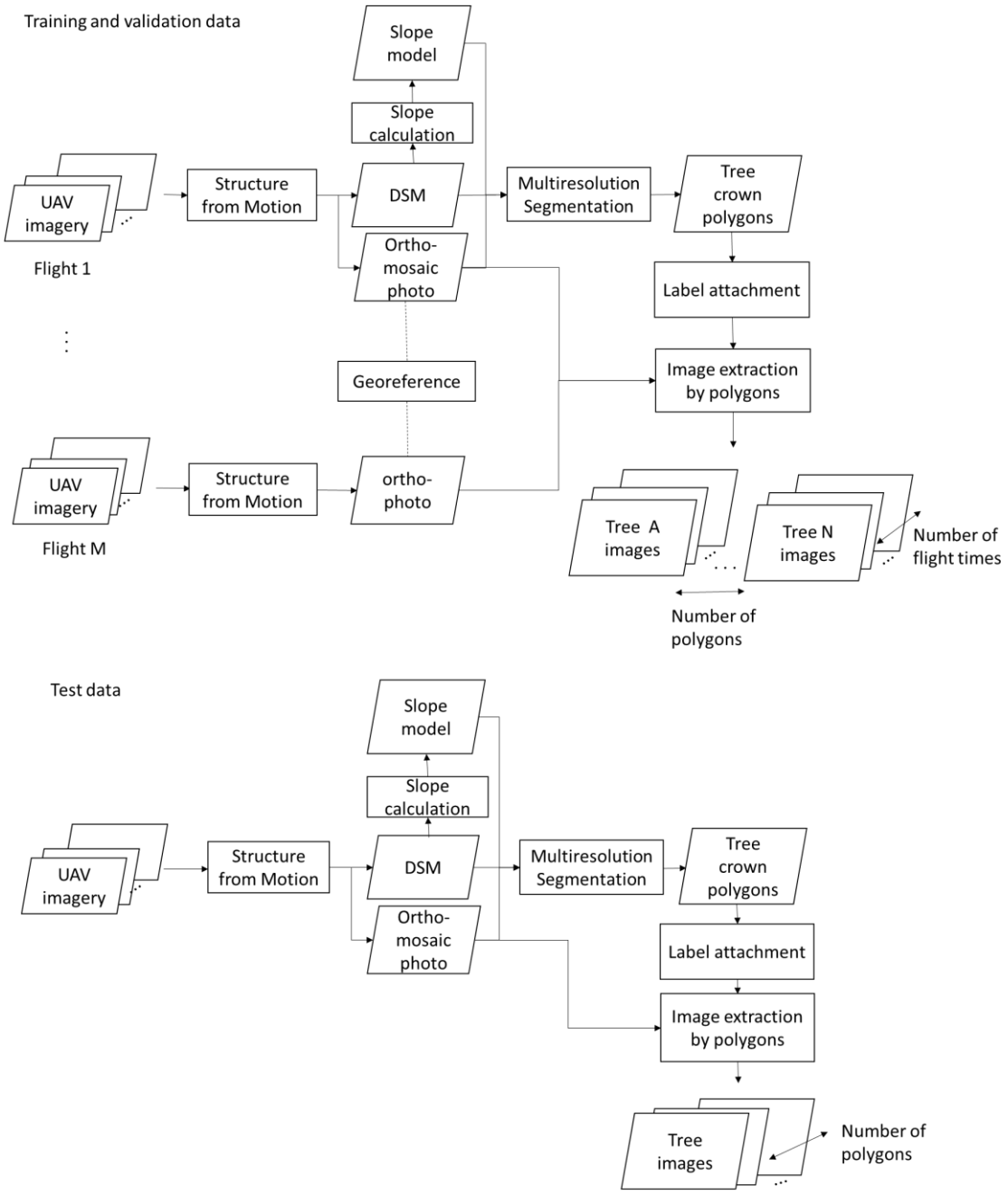
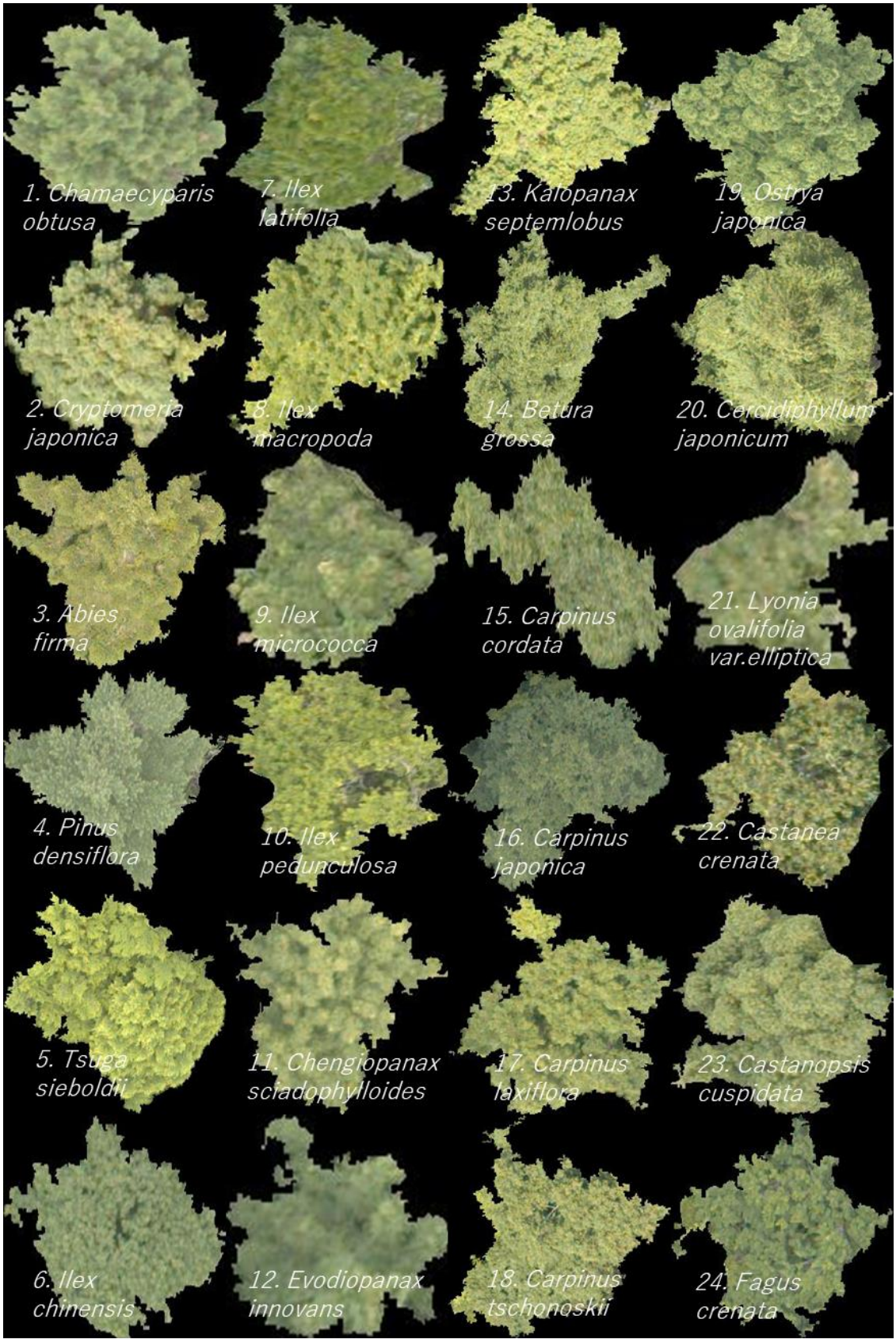
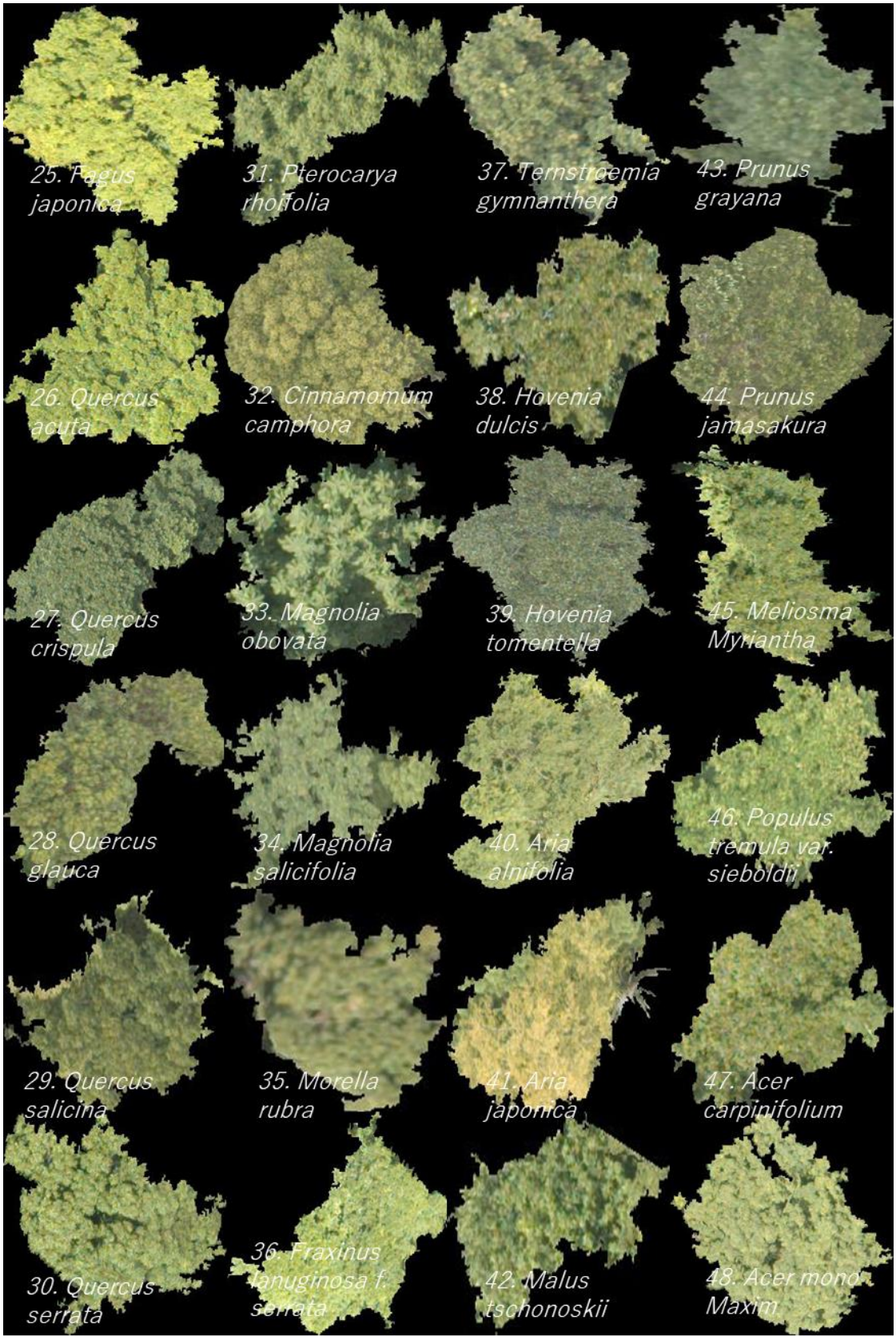


Figure 3.2 The method of UAV imagery processing. Upper process diagram shows the preparation method training and validation dataset from training site data, and lower process diagram shows that of test data from test site.


1. Chamaecyparis obtusa
7. Ilex latifolia
13. Kalopanax septemlobus
19. Ostrya japonica
2. Cryptomeria japonica
8. Ilex macropoda
14. Betura grossa
20. Cercidiphyllum japonicum
3. Abies firma
9. Ilex micrococca
15. Carpinus cordata
21. Lyonia ovalifolia var.elliptica
4. Pinus densiflora
10. Ilex pedunculosa
16. Carpinus japonica
22. Castanea crenata
5. Tsuga sieboldii
11. Chengiopanax sciadophylloides
17. Carpinus laxiflora
23. Castanopsis cuspidata
6. Ilex chinensis
12. Evodiopanax innovans
18. Carpinus tschonoskii
24. Fagus crenata


25. Fagus japonica
31. Pterocarya rhoifolia
37. Ternstroemia gymnanthera
43. Prunus grayana
26. Quercus acuta
32. Cinnamomum camphora
38. Hovenia dulcis
44. Prunus jamasakura
27. Quercus crispula
33. Magnolia obovata
39. Hovenia tomentella
45. Meliosma Myriantha
28. Quercus glauca
34. Magnolia salicifolia
40. Aria alnifolia
46. Populus tremula var. sieboldii
29. Quercus salicina
35. Morella rubra
41. Aria japonica
47. Acer carpinifolium
30. Quercus serrata
36. Fraxinus lanuginosa f. serrata
42. Malus tschonoskii
48. Acer mono Maxim

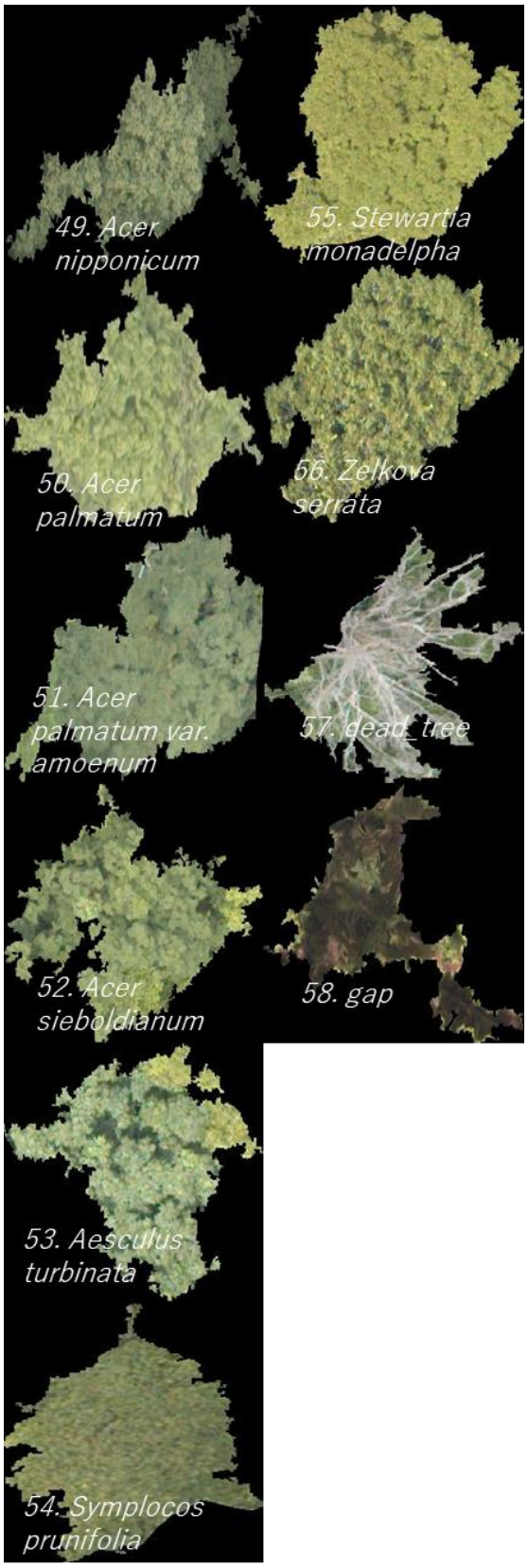


Figure 3.3 Tree crown images of each species taken by UAV from 100m altitude

3.2.5 Analysis

I checked the accuracy of the two types of validation and one test. The separation method of the validation dataset is shown in Fig. 3.4. In validation 1, I divided the dataset randomly into training and validation datasets at rates of 70% and 30%, respectively. Validation 1 can evaluate the performance of the model for the dataset obtained from the same time or same trees. In validation 2, the images were divided by polygons at the same rate as validation 1. This means tree A images are included only in training, and tree B images in validation dataset. Therefore, validation 2 showed the performance of the model at the same time but with different trees. In Table 3.3, I show the number of images for each dataset.

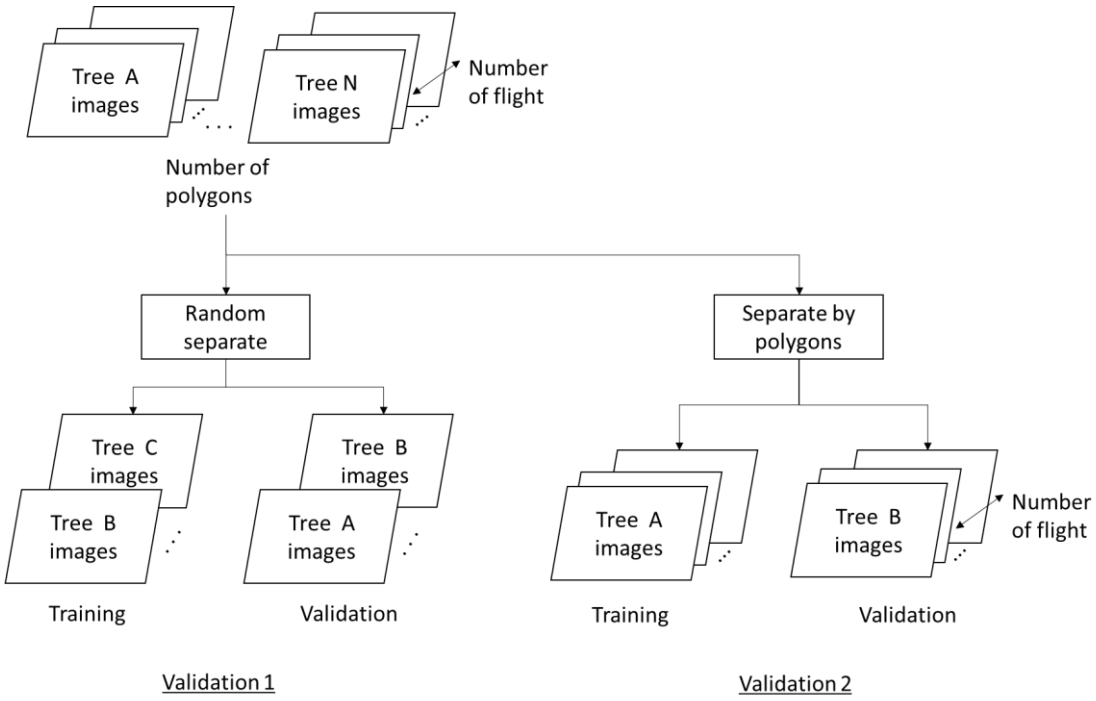


Figure 3.4 Method for validation dataset separation. The images were extracted from previous process shown in upper process diagram in Figure 3.2. In validation 1, training and validation datasets were divided randomly. In validation 2, training and validation datasets were divided by polygons.

Table 3.3 Number of images of each dataset

class	training and validation	validation 1		validation2		test
		training	validation	training	validation	
1	961	675	286	654	307	139
2	1619	1143	476	1213	406	17
3	1115	772	343	810	305	122
4	373	264	109	264	109	3
5	205	155	50	125	80	11
6	16	14	2	8	8	0
7	12	10	2	8	4	0
8	18	8	10	12	6	0
9	24	15	9	16	8	0
10	202	133	69	150	52	0
11	46	36	10	32	14	5
12	28	21	7	22	6	0
13	54	37	17	36	18	0
14	349	230	119	229	120	23
15	20	13	7	15	5	0
16	16	28	-12	11	5	0
17	168	123	45	112	59	31
18	236	169	67	157	79	0
19	70	55	15	35	35	0
20	138	103	35	120	18	0
21	12	9	3	8	4	0
22	10	7	3	5	5	0
23	1624	1142	482	1112	512	333
24	1297	895	402	988	309	157
25	182	137	45	106	76	0
26	200	140	60	160	40	0
27	165	121	44	103	62	0
28	87	61	26	62	25	25
29	330	252	78	246	84	15
30	650	455	195	444	206	97
31	248	166	82	178	70	21
32	8	7	1	4	4	1

33	30	18	12	12	18	3
34	12	9	3	6	6	0
35	260	182	78	192	68	0
36	30	19	11	24	6	0
37	28	25	3	24	4	0
38	18	14	4	12	6	0
39	36	23	13	24	12	0
40	6	5	1	6	0	3
41	6	6	0	6	0	0
42	6	5	1	6	0	0
43	24	15	9	16	8	4
44	42	29	13	34	8	9
45	12	9	3	6	6	0
46	20	15	5	20	0	0
47	18	11	7	12	6	0
48	177	118	59	131	46	25
49	6	3	3	6	0	0
50	10	4	6	5	5	2
51	17	11	6	11	6	1
52	18	8	10	6	12	0
53	102	65	37	90	12	3
54	144	100	44	88	56	6
55	65	44	21	45	20	0
56	60	47	13	48	12	0
57	73	49	24	64	9	18
58	759	520	239	495	264	401

3.2.6 Deep learning

I used PyTorch (Paszke et al. 2017) as the deep learning framework, and EfficientNet B7 (Tan et al. 2019) as the network model. I applied fine-tuning: the output layer was changed to 58 classes, and all parameters were trained. Other parameters were set to {batch: 8, learning rate: 0.005, optimizer: SGD, momentum: 0.9, epochs 200}. In the training phase, I augmented the

training image number, which was less than 100 to approximately 100, by random rotation and horizontal flip.

Furthermore, I developed a method for improving classification performance using the vegetation data of the target area. I called this method inventory tuning. Ideally, the model that trained only the tree species that existed in the target area showed optimal classification accuracy. However, making the local model for every target area is not realistic because the vegetation is constantly changing, and hundreds of models must be created. Conversely, for inventory tuning, I used one model that was trained by many classes (in this case, 58 classes). In the prediction phase, I limited the output to classes that were only listed in field-obtained inventory data. Using this system, I can create a local model from big model.

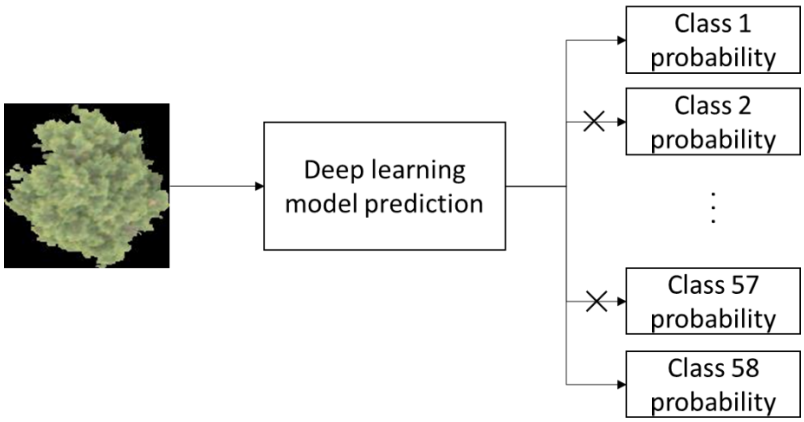


Figure 3.5 Image of inventory tuning.

3.2.7 Performance evaluation

For performance evaluation, I used Cohen's Kappa score (Cohen, 1960) to evaluate the overall performance, and precision and recall for each class classification accuracy. To visualize the details of the classification result, I used a confusion matrix where the vertical axis indicates the ground truth label and the horizontal axis shows the prediction label. The number of each cell represents the number of images. However, this method is not sufficient to confirm the misclassification between each class because each cell has to be small when the number of

classes is large. To visualize the misclassification between classes, which means similarities across classes, I developed a visualization method using a dendrogram.

Table 3.4 Sample matrix of classification result

		Prediction		
	class	**A**	**B**	**C**
ground truth	**A**	a	b	c
	B	d	e	f
	C	g	h	i

Here, I considered this matrix (Table 3.4). When I calculate the classification accuracy of class A, the F1 score is one of the major accuracy indices. The F1 score is the harmonic mean of the precision and recall. Precision, recall, and F1 score can be calculated using the formula below.

$$Precision_A = \frac{a}{a + d + g} \tag{1}$$

$$Recall_A = \frac{a}{a + b + c} \tag{2}$$

$$F1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{3}$$

When I focus on the classification accuracy between classes A and C, I can calculate the performance using equations (4) and (5) below.

$$F1\ score_A\ between\ A\ and\ C = \frac{2a}{2a + c + g} \tag{4}$$

$$F1\ score_C\ between\ A\ and\ C = \frac{2i}{2i + c + g} \tag{5}$$

From these two F1 scores, the identification accuracy between A and C can be calculated by averaging the scores.

$$F1\ score\ between\ A\ and\ C = \left(\frac{2a}{2a+c+g} + \frac{2i}{2i+c+g}\right) \times \frac{1}{2} = \frac{4ai + (a+1)(c+g)}{(2a+c+g)(2i+c+g)} \quad (6)$$

A low F1 score means low identification performance; therefore, the score means identification of these classes is difficult because of the similarity of the classes. Therefore, similarities in appearance can be calculated using equation (7) and the F1 scores.

$$similarity\ between\ A\ and\ C = F1\ score\ between\ A\ and\ C = \frac{4ai + (a+1)(c+g)}{(2a+c+g)(2i+c+g)} \quad (7)$$

Using equation (7), I calculated the similarities among all classes. and made similarities matrix which shows similarities value of each species Then, I applied Ward's clustering method (Ward, 1963) to the similarities matrix and visualized the similarities relationships between classes using the dendrogram in R v3.6.3 software.

3.3 Results

3.3.1 Validation 1: Performance for dataset of same time or same trees

At first, I evaluated a model using a dataset which obtained from same time or same trees as the training dataset. Though this evaluation is rarely reproduced in practical applications, I revealed the potential for identifying each tree species. As a result, the model showed a high validation accuracy: Kappa score was 0.979. The classification results are shown in Figure 3.6, and the details are presented in Table 3.5. Almost all tree species showed 90% accuracy in both precision and recall. This result means that when the data on light condition and trees are similar to those of the training data, almost all tree species can be identified. Meanwhile, classes 37, 41, and 50 showed a low classification accuracy. Class 41 had no recall and precision because there were no ground images. No ground images were obtained because of the program that separated each training data into a validation dataset randomly at 30%, and a small number of overall images of the class.

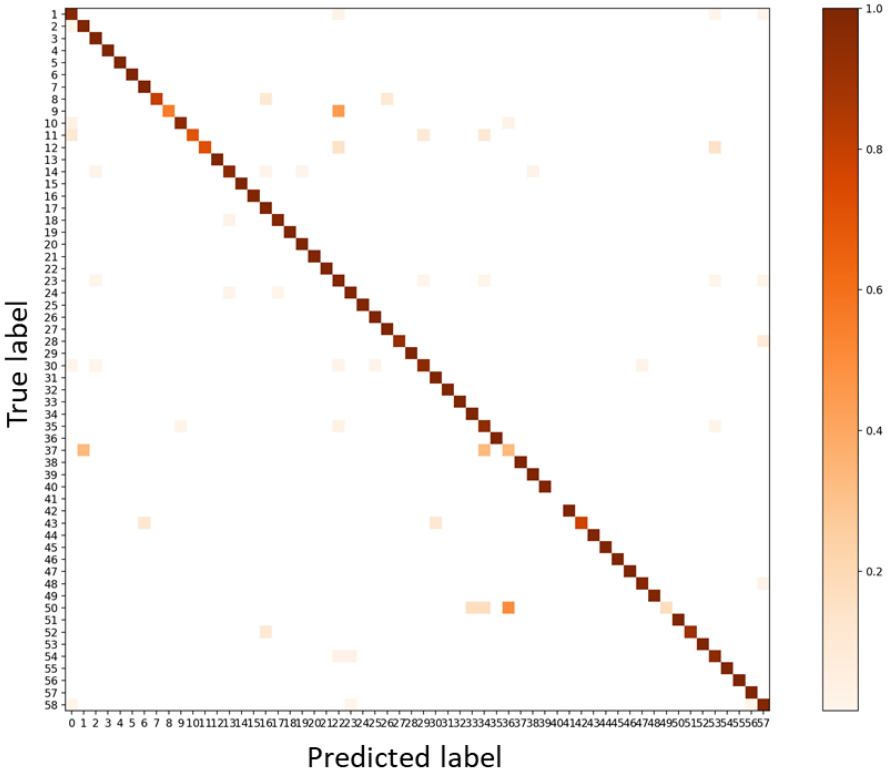


Figure 3.6 Confusion matrix of random separation dataset: same time or same trees. The vertical axis is the ground truth and the horizontal axis the model prediction. The number in each cell indicates the number of classified images; each cell is colored according to the percentage of the number of images in each class. The model showed 0.98 Kappa scores.

Table 3.5 Details on classification accuracy of the model that was trained from the same time and same trees. The slash means no denominator data.

class	precision	recall	class	precision	recall
1	0.98	0.97	30	0.98	0.96
2	1.00	1.00	31	0.99	1.00
3	0.99	1.00	32	1.00	1.00
4	1.00	1.00	33	1.00	1.00
5	1.00	1.00	34	0.75	1.00
6	1.00	1.00	35	0.95	0.95
7	0.67	1.00	36	1.00	1.00
8	1.00	0.80	37	0.20	0.33
9	1.00	0.56	38	1.00	1.00
10	0.99	0.96	39	0.87	1.00
11	1.00	0.70	40	1.00	1.00
12	1.00	0.71	41	/	/
13	1.00	1.00	42	1.00	1.00
14	0.97	0.96	43	1.00	0.78
15	1.00	1.00	44	1.00	1.00
16	1.00	1.00	45	1.00	1.00
17	0.94	1.00	46	1.00	1.00
18	0.99	0.99	47	1.00	1.00
19	1.00	1.00	48	0.97	0.98
20	0.97	1.00	49	1.00	1.00
21	1.00	1.00	50	1.00	0.17
22	1.00	1.00	51	1.00	1.00
23	0.97	0.98	52	1.00	0.90
24	1.00	0.99	53	1.00	1.00
25	1.00	1.00	54	0.88	0.95
26	0.97	1.00	55	1.00	1.00
27	0.98	1.00	56	1.00	1.00
28	1.00	0.92	57	0.96	1.00
29	1.00	1.00	58	0.98	0.99

3.3.2 Validation 2: Performance for dataset of same time and different trees

Second, I evaluated a model using a dataset which obtained from same time and different trees as the training dataset. From this evaluation, I revealed the spatial robustness of the model. As a result the model showed a relatively high accuracy, with a kappa of 0.715. This performance was approximately 25% lower than that of random separation. I show the details of the classification performance of each class (Figure 3.7, Table 3.6). The precision and recall of 21 classes exceeded 50%. In particular, coniferous trees (classes 1–5) showed stable performance, with more than 75% precision and recall.

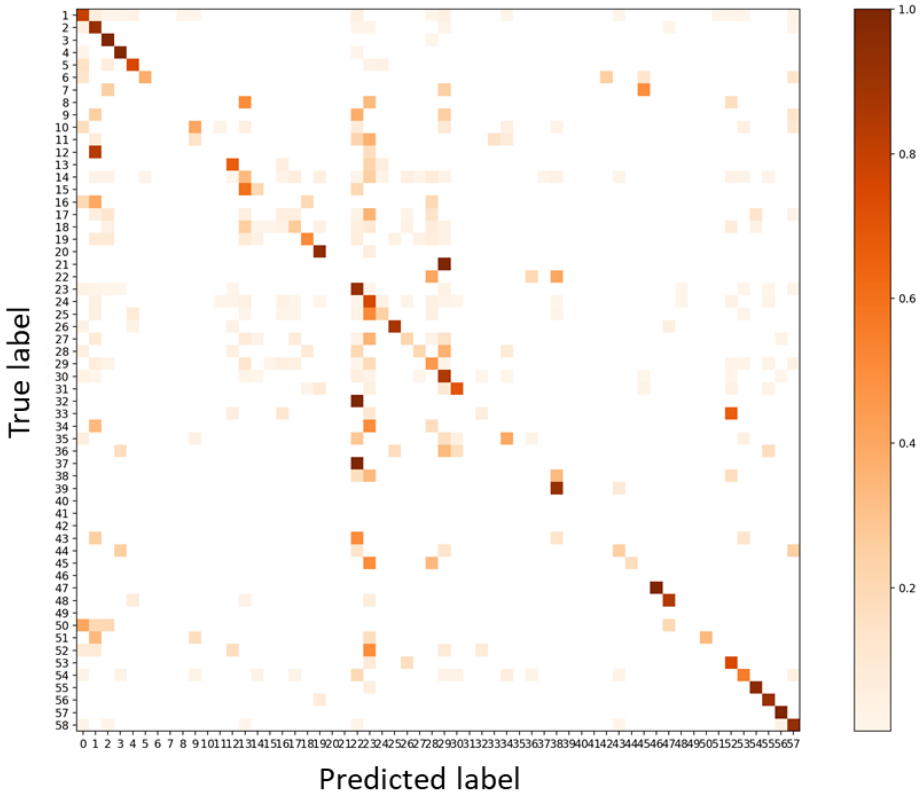


Figure 3.7 Confusion matrix of random separation dataset: same site and same time. The vertical axis is the ground truth and the horizontal axis the model prediction. The number in each cell indicates the number of classified images; each cell is coloredcoloured according to the percentage of the number of images in each class. The model yielded ashowed 0.72 Kappa score of 0.72scores.

Table 3.6 Details on classification accuracy of the model which trained from same time and different site. The slash means no denominator data.

class	precision	recall	class	precision	recall
1	0.76	0.79	30	0.67	0.85
2	0.84	0.92	31	0.91	0.70
3	0.91	1.00	32	/	0.00
4	0.93	0.98	33	0.33	0.06
5	0.79	0.75	34	/	0.00
6	0.75	0.38	35	0.73	0.40
7	/	0.00	36	0.00	0.00
8	/	0.00	37	0.00	0.00
9	0.00	0.00	38	0.00	0.00
10	0.75	0.40	39	0.50	0.92
11	0.00	0.00	40	/	/
12	0.00	0.00	41	/	/
13	0.57	0.67	42	/	/
14	0.37	0.33	43	0.00	0.00
15	0.17	0.20	44	0.29	0.25
16	0.00	0.00	45	1.00	0.17
17	0.13	0.05	46	0.00	0.00
18	0.49	0.27	47	1.00	1.00
19	0.82	0.51	48	0.91	0.85
20	0.53	0.94	49	/	/
21	/	0.00	50	/	0.00
22	/	0.00	51	1.00	0.33
23	0.82	0.92	52	0.00	0.00
24	0.56	0.76	53	0.20	0.75
25	0.60	0.24	54	0.74	0.55
26	0.95	0.88	55	0.66	0.95
27	0.54	0.21	56	0.58	0.92
28	0.63	0.20	57	0.50	1.00
29	0.46	0.45	58	0.90	0.93

3.3.3 Test: Performance for dataset of different time and different site

The model trained from different times and sites showed a even lower performance of a kappa score of 0.472. I show the details of the classification performance of each class (Figure 3.8, Table 3.7). Only three classes (1, 3, 58) showed good performance, with both precision and recall exceeding 50%. Classes 2, 4, 23, 24, 30, and 58 have the potential to be identified (precision or recall exceed 50%).

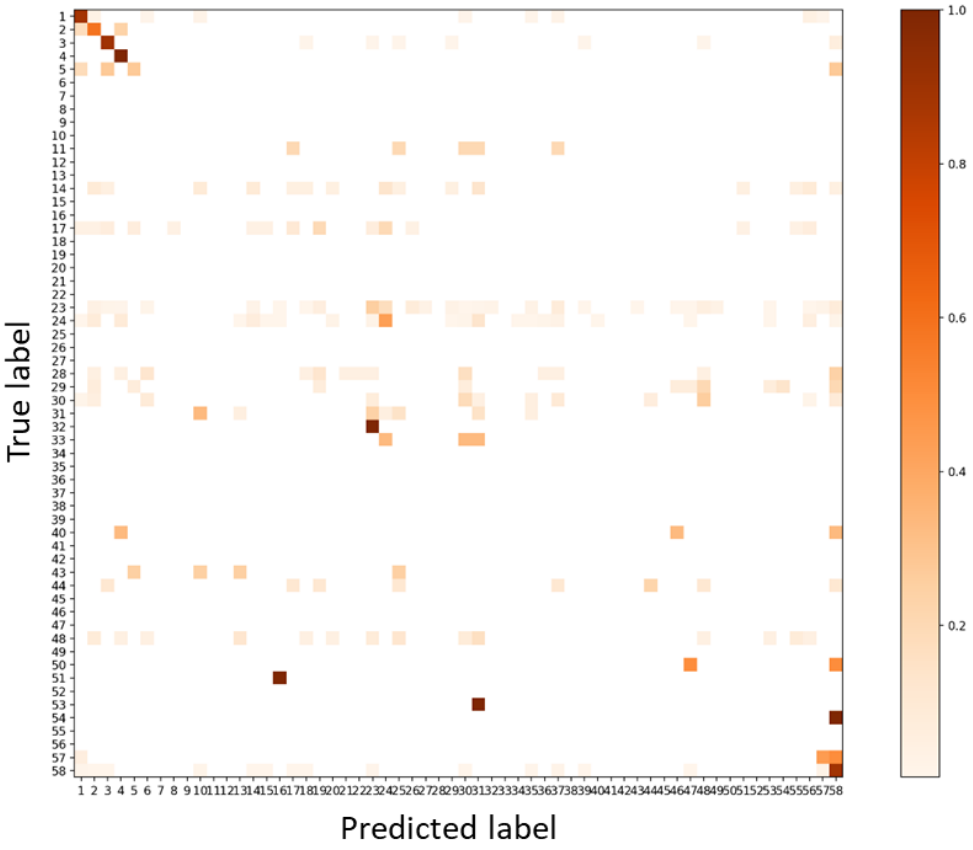


Figure 3.8 Classification result of the test. The vertical axis is the ground truth and the horizontal axis the model prediction. The number in each cell indicates the number of classified images; each cell is colored according to the percentage of the number of images in each class. The Kappa score is 0.47.

Table 3.7 Details on classification accuracy of the model which was trained from different times and different sites. The slash means no denominator data.

class	precision	recall	class	precision	recall
1	0.90	0.88	30	0.55	0.19
2	0.20	0.59	31	0.07	0.14
3	0.90	0.89	32	0.00	0.00
4	0.12	1.00	33	0.00	0.00
5	0.43	0.27	34	0.00	0.00
6	0.00	0.00	35	0.00	0.00
7	/	/	36	0.00	0.00
8	0.00	0.00	37	0.00	0.00
9	/	/	38	/	/
10	0.00	0.00	39	0.00	0.00
11	0.00	0.00	40	0.00	0.00
12	/	/	41	/	/
13	0.00	0.00	42	/	/
14	0.11	0.09	43	0.00	0.00
15	0.00	0.00	44	0.25	0.22
16	0.00	0.00	45	/	/
17	0.43	0.10	46	0.00	0.00
18	0.00	0.00	47	0.00	0.00
19	0.00	0.00	48	0.02	0.04
20	0.00	0.00	49	0.00	0.00
21	0.00	0.00	50	0.00	0.00
22	0.00	0.00	51	0.00	0.00
23	0.77	0.26	52	/	/
24	0.50	0.43	53	0.00	0.00
25	0.00	0.00	54	0.00	0.00
26	0.00	0.00	55	0.00	0.00
27	0.00	0.00	56	0.00	0.00
28	0.00	0.00	57	0.36	0.44
29	0.00	0.00	58	0.83	0.89

3.3.4 Test using inventory tuning

When I applied inventory tuning, the test result was improved from a Kappa score of 0.472 to 0.616. The classification details are presented in Figure 3.9 and Table 3.8. The precision and recall in 9 classes (1, 2, 3, 4, 23, 24, 30, 57, 58) were 50%. In particular, four of the five coniferous tree species (1–4) showed high classification performance.

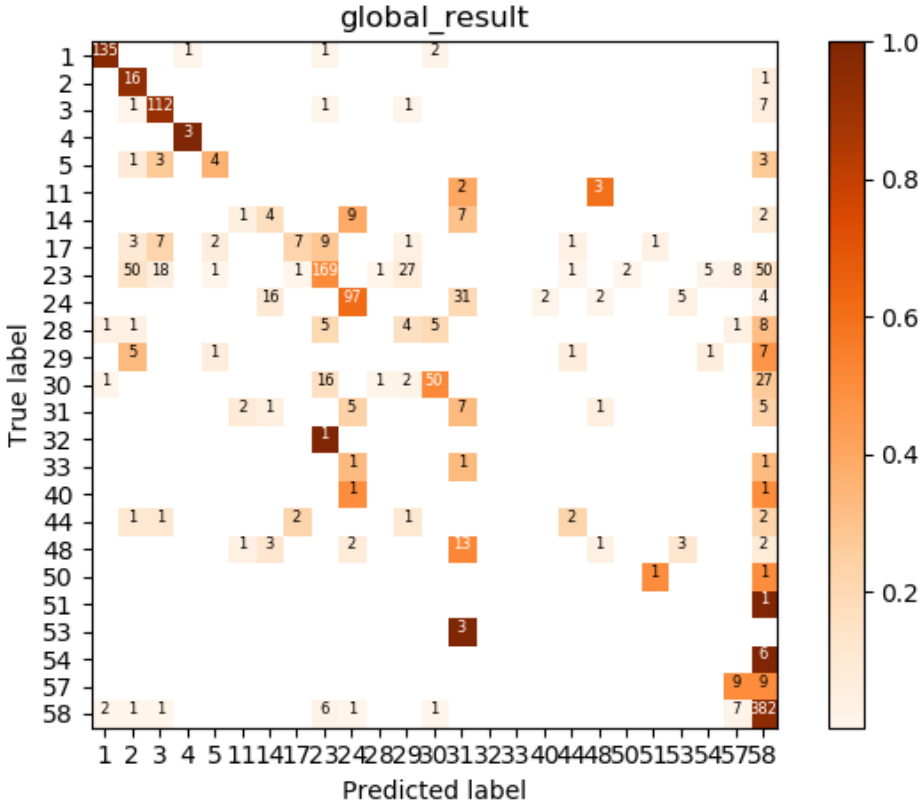


Figure 3.9 Classification result of the test using inventory tuning. The vertical axis is the ground truth and the horizontal axis the model prediction. The number in each cell indicates the number of classified images; each cell is colored according to the percentage of the number of images in each class. The Kappa score is 0.62.

I visualised the classification results of CNN with inventory tuning applied to Kasugayama site in GIS (Figure 3.10). We can know the Kasugayama which is a primeval forest is dominated by *Castanopsis cuspidate*, *Abies firma* and *Cryptomeria japonica*. Tree species map like this will help biodiversity assessment, monitoring of succession, and ecology researches.

Table 3.8 Details on classification accuracy of inventory tuning. The slash means no denominator data

class	precision	recall	class	precision	recall
1	0.97	0.97	30	0.86	0.52
2	0.20	0.94	31	0.11	0.33
3	0.79	0.92	32	0.00	0.00
4	0.75	1.00	33	0.00	0.00
5	0.50	0.36	34	/	/
6	/	/	35	/	/
7	/	/	36	/	/
8	/	/	37	/	/
9	/	/	38	/	/
10	/	/	39	/	/
11	0.00	0.00	40	0.00	0.00
12	/	/	41	/	/
13	/	/	42	/	/
14	0.17	0.17	43	0.00	0.00
15	/	/	44	0.40	0.22
16	/	/	45	/	/
17	0.70	0.23	46	/	/
18	/	/	47	/	/
19	/	/	48	0.13	0.04
20	/	/	49	/	/
21	/	/	50	0.00	0.00
22	/	/	51	0.00	0.00
23	0.81	0.51	52	/	/
24	0.82	0.62	53	0.00	0.00
25	/	/	54	0.00	0.00
26	/	/	55	/	/
27	/	/	56	/	/
28	0.00	0.00	57	0.36	0.50
29	0.00	0.00	58	0.73	0.95

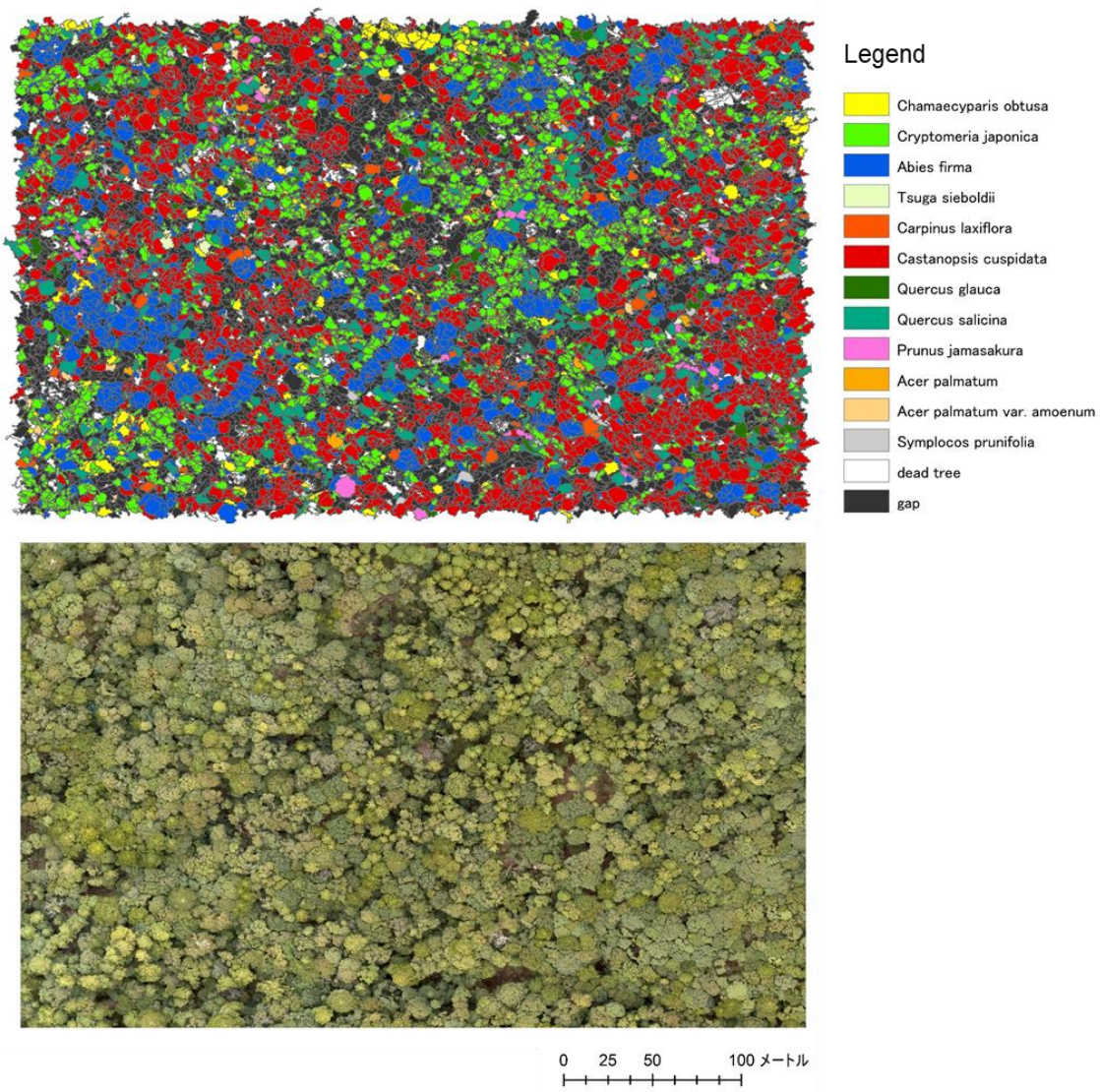


Figure 3.10 Orthomosaic photo and classification maps obtained with the inventory tuning CNN in Kasugayama site. The area is 12 ha.

3.4 Discussion

3.4.1 Classification performance and robustness

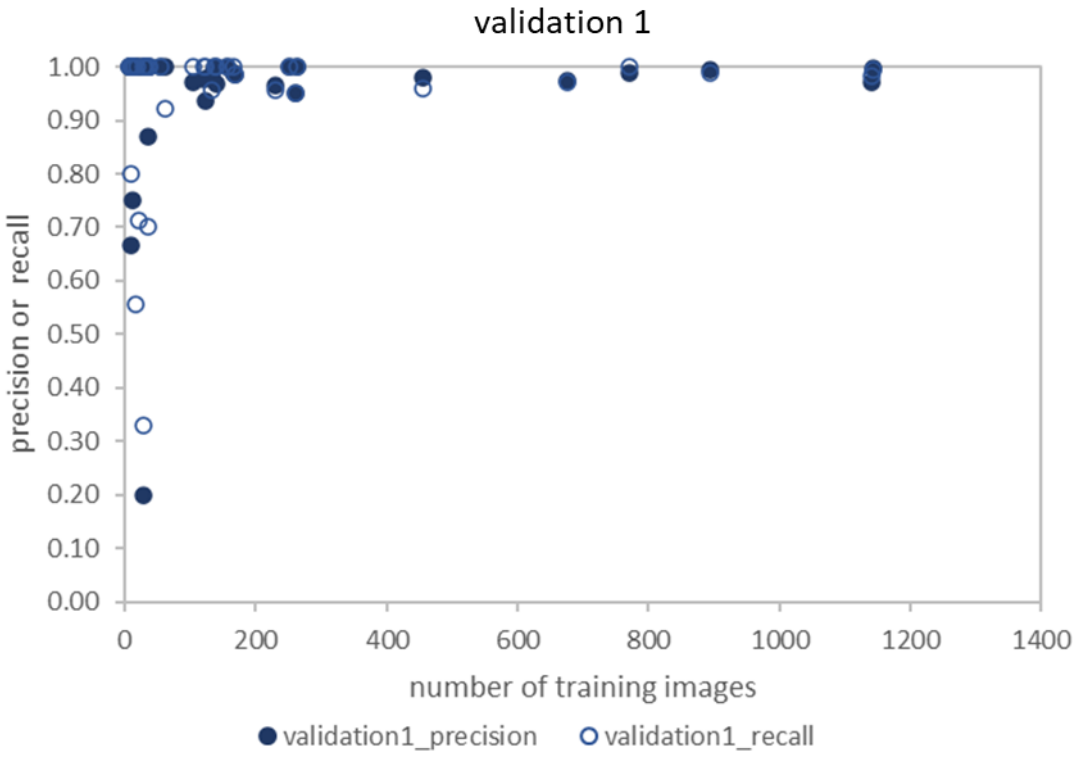
This study illustrates high potential and some stable robustness of the tree species identification systems using UAVs and deep learning. If the condition of the trees and shooting was optimal, almost all tree species could be identified (Kappa 0.979). This means that this method has the potential to identify a large number of tree species. However, this identification performance is rarely reproduced in practical applications. A similar classification accuracy is observed when this system is used for regular monitoring of the site.

When I used the dataset obtained at the same time but from different trees, the performance decreased from 0.979 to 0.715 Kappa score, and 36% (21 class) tree species could potentially be identified (more than 50% in both precision and recall). This classification accuracy is assumed when I take images from one flight and separate them into training and test datasets. The decrease was caused by the difference in features of the training images and test images. A previous study showed that CNN can use features made from the edge shape of foliage and bush of branches, hierarchical branching patterns, and outlines of tree shapes (Onishi and Ise, 2021). In this study, CNN may also use such features, but the limited number of training samples and large number of classes might cause a decrease in performance.

As for the test that shows the performance of general practical use, the accuracy was further decreased to Kappa 0.472. Conversely, some tree species, especially coniferous trees and representative species of stands, showed potential for identification. This result indicates that my system has temporal and spatial robustness to a certain extent. In general, previous studies have evaluated the identification performance using one dataset obtained at one flight of the UAV or airplane. Under these conditions, the shooting conditions, such as weather or season, and vegetation were the same. Under these conditions, the method using the multi-or hyperspectral sensors showed good identification performance, but if the condition was different, the method may not show good performance because spectral reflectance can change under different shooting conditions or vegetation. From this point of view, my method showed some robustness in these

different conditions. This robustness indicates that if the method were to be made a classification model, other researchers or general users would be able to identify tree species automatically. If the classification accuracy is increased by the increased number of training data, the model is feasible.

The relationship between precision or recall accuracy and the number of training images is shown in Figure 3.11. With respect to validation 1, the plot showed clear proportionality. In particular, classes that have less than 50 training samples sometimes show an accuracy ranging from 0.2 to 0.9, and classes with more than 50 training samples always show high precision and recall of more than 0.9. With respect to validation 2, I can confirm proportionality, but the plot was more scattered. The class that had fewer than 300 training images showed varied scores, but the class with more than 300 training images always showed the potential for being identified, where both precision and recall were more than 0.5. In addition, recall scores were higher than the precision in those classes. As for the test, most of the classes with less than 300 training samples showed zero precision and recall accuracy. The class with more than 300 training samples always showed more than 0.1 precision or recall. This means that at least 300 training samples should be gathered for practical use.


validation 1
precision or recall
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
200
400
600
800
1000
1200
1400
number of training images
validation1_precision
validation1_recall

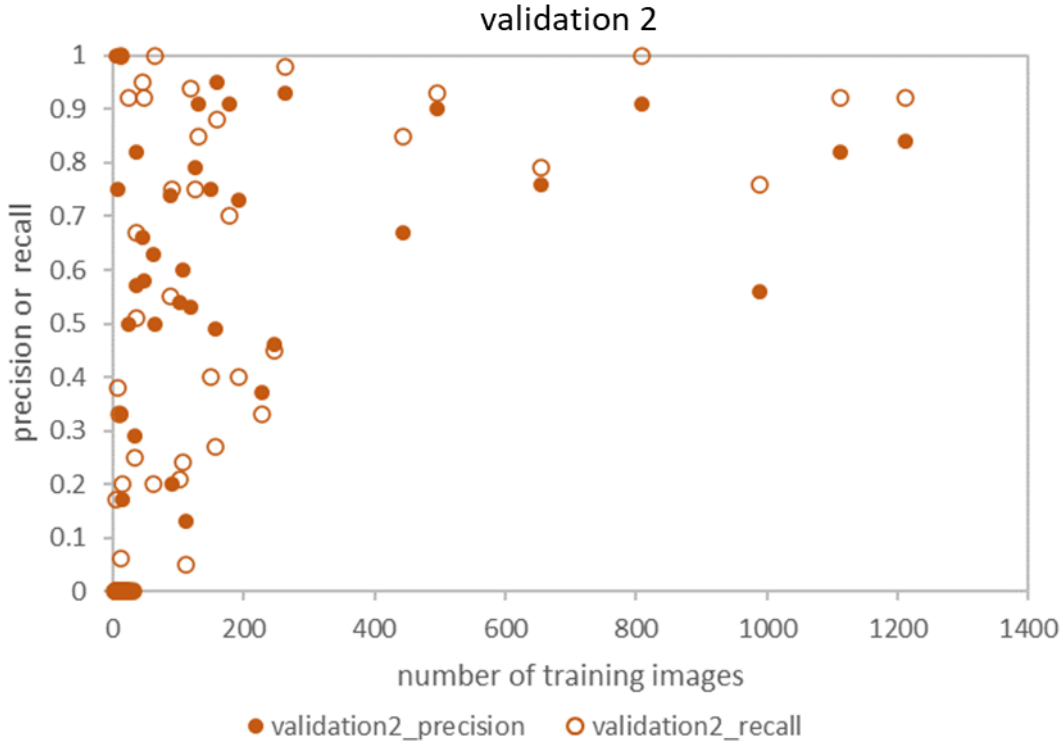

validation 2
precision or recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
200
400
600
800
1000
1200
1400
number of training images
validation2_precision
validation2_recall

Figure 3.11 Relationship between precision or recall accuracy and number of training images. From top to bottom, the accuracy of validation 1, validation 2 and the test.

3.4.2 Inventory tuning

In this study, I developed a new identification system called "inventory tuning." Using this method, I successfully created a local model from one large global model and improved the identification accuracy (0.472 to 0.616 Kappa score). The advantage of this method is that I can create a local model without retraining the initial model. With respect to the feasibility of the classification system, selecting the tree species and retraining the model is expected to be feasible, but it requires considerable computational load and time. From this perspective, inventory tuning does not require retraining and almost no computational load and time compared to the normal classification. Although in this study, I used the field-obtained inventory data for inventory tuning, I may use the vegetation inventory data predicted from the longitude, latitude, and altitude in the future. Therefore, automatic inventory tuning can be used.

3.4.3 Similarities in appearance of trees species

I visualized the similarities in appearance of each tree species from the classification results of the test in Figure 3.12. Although some species appear to be lined up chaotically, certain branches showed interesting tendencies. First, most of the branches on the trees in closest proximity to the center had large leaves (*Kalopanax septemlobus, Acer mono Maxim, Magnolia obovata, Pterocarya rhoifolia, and Aesculus turbinata*). *Kalopanax septemlobus* and Acer mono Maxim belong to different families, but both have large and palmate leaves. Therefore, the textures that can be seen from the UAV appear to be similar. The leaves of *Magnolia obovata*, *Pterocarya rhoifolia, Aesculus turbinata*, the leaves of *Magnolia obovata* and *Aesculus turbinata* are similar to those of verticillate big leaves. However, *Pterocarya rhoifolia* has a pinnately compound leaf; therefore, *Pterocarya rhoifolia* is thought to be different, but in the images, the pinnately compound leaf looks similar to *Aesculus turbinata* at 100 m altitude. Both *Aesculus turbinata* and *Pterocarya rhoifolia* prefer a wet environment along the stream and grow at the same site. Therefore, on a macro level, the trees of the same functional types and that thrive in the same environment may be hypothesized as having similar light acquisition performance and textures. In contrast, the phylogenetically close species were also located close to the dendrogram. For example, *Carpinus cordata* and *Carpinus laxiflora*, located in close proximity to *Acer carpinifolium* and *Acer palmatum* are also located close to the dendrogram. Furthermore, *Chamaecyparis obtusa*, *Cryptomeria japonica*, and *Pinus densiflora*, which are coniferous trees positioned close to each other. Meanwhile, some coniferous trees located close to broad-leaved trees in the dendrogram. This suggests that coniferous tree does not have unique textures from 100 m above even if the leaf shape differs from broad-leaved tree. In this time, I used some fragmented tree images for training and test giving priority to automatic segmentation. Using only whole tree images will improve the coniferous tree species identification performance and clearly locate them far from broad-leaved tree in the dendrogram judging from the existence of tree top and the difference of branching patterns. Further, some evergreen and deciduous tree located

disorderly, some evergreen broad-leaved trees located close such as *Ilex chinensis* and *Quercus glauca*, *Custanopsis cuspidate* and *Quercus acuta*. This means that though there may be some common features among some evergreen trees, those are not enough for identification in summer. To distinguish deciduous trees from evergreen trees, autumn leaf coloring season is best, but detailed deciduous tree species classification may be difficult due to the unstable number of leaves and colors. In summary, ①tree species that belong to close phylogenetically sometimes have similar textures, ②tree species that have similar leaf shapes sometimes look similar, ③tree species that prefer the same environment may sometimes show similar textures, and ④tree types such as coniferous and broad-leaved or evergreen and deciduous do not always promise common features among the tree type.

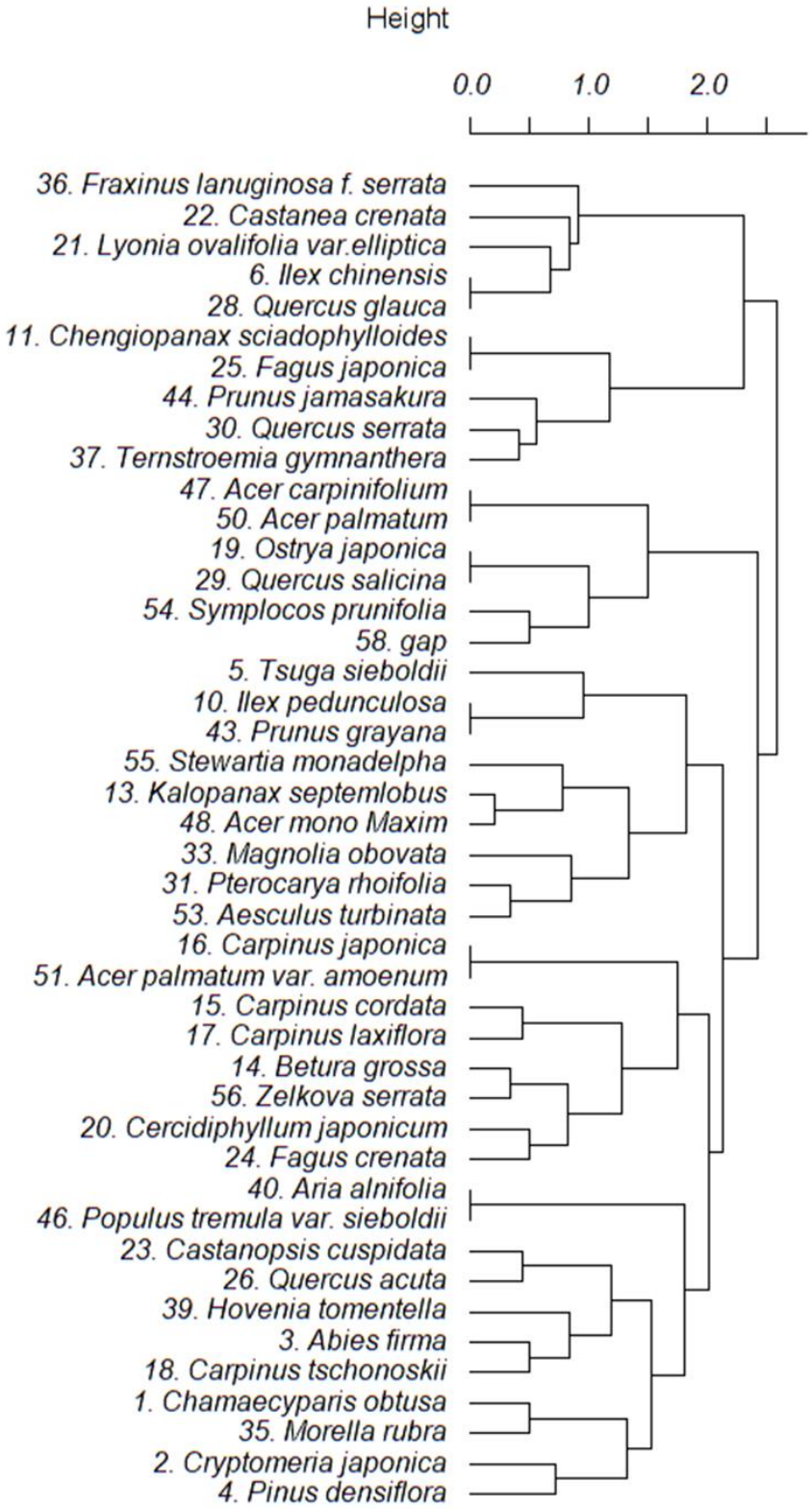


Figure 3.12 Similarities in appearance of trees species. This dendrogram was made from the distance calculated from F1 score of each class in test classification result.

3.5 Conclusion

This study demonstrated the stable robustness of the tree species identification system using UAV RGB imagery and deep learning for some tree species. In particular, coniferous tree species showed stable performance with respect to identification. To improve the performance, inventory tuning, which allows for the creation of a local model using inventory data, was proven to be an effective method. Furthermore, at least 300 images are needed to train the model for feasibility. In addition, misclassification occurred between ①tree species that belong to close phylogenetically, ②tree species that have similar leaf shapes, and ③tree species that prefer the same environment. These findings will promote the practicalization of identification systems using UAV RGB imagery and deep learning.

3.6 Reference

Asner, G. P. Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sens. Environ.* **64**, 234–253 (1998).

Asner, G. P. & Warner, A. S. Canopy shadow in IKONOS satellite observations of tropical forests and savannas. *Remote Sens. Environ.* **87**, 521–533 (2003).

Baatz, M. & Schäpe, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. *XII Angew. Geogr. Informationsverarbeitung, Wichmann-Verlag, Heidelb.* (2000).

Ballanti, L., Blesius, L., Hines, E. & Kruse, B. Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sens.* **8**, 1–18 (2016).

Bo-Cai Gao & Goetz, A. F. H. Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. *J. Geophys. Res.* **95**, 3549–3564 (1990).

Boschetti, M., Boschetti, L., Oliveri, S., Casati, L. & Canova, I. Tree species mapping with airborne hyper-spectral MIVIS data: The Ticino Park study case. *Int. J. Remote Sens.* **28**, 1251–1261 (2007).

Cao, L. *et al.* Tree species classification in subtropical forests using small-footprint full-waveform LiDAR data. *Int. J. Appl. Earth Obs. Geoinf.* (2016) doi:10.1016/j.jag.2016.01.007.

Clark, M. L. & Roberts, D. A. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sens.* **4**, 1820–1855 (2012).

Clark, M. L. & Roberts, D. A. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sens.* **4**, 1820–1855 (2012).

Clark, M. L., Roberts, D. A. & Clark, D. B. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens. Environ.* **96**, 375–398 (2005).

Cohen, J. A coeffient of agreement for nominal scales. *Educ. Psychol. Meas.* **XX**, 37–46 (1960).

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.* **123**, 258–270 (2012).

Dalponte, M., Bruzzone, L., Vescovo, L. & Gianelle, D. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sens. Environ.* **113**, 2345–2355 (2009).

Dalponte, M., Ørka, H. O., Gobakken, T., Gianelle, D. & Næsset, E. Tree species classification in boreal forests with hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **51**, 2632–2645 (2013).

Fassnacht, F. E. *et al.* Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **186**, 64–87 (2016).

Feret, J. B. & Asner, G. P. Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Trans. Geosci. Remote Sens.* **51**, 73–84 (2013).

Grant, L. Diffuse and specular characteristics of leaf reflectance. *Remote Sens. Environ.* **22**, 309–322 (1987).

Knipling, E. B. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sens. Environ.* **1**, 155–159 (1970).

Leckie, D. G. *et al.* Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sens. Environ.* **94**, 311–326 (2005).

Machala, M. & Zejdová, L. Forest mapping through object-based image analysis of multispectral and LiDAR aerial data. *Eur. J. Remote Sens.* **47**, 117–131 (2014).

Nakatake, S., Yamamoto, K., Yoshida, N., Yamaguchi, A. & Unom, S. Development of a single tree classification method using airborne LiDAR. *J. Japanese For. Soc.* **100**, 149–157 (2018).

Natesan, S., Armenakis, C. & Vepakomma, U. Resnet-based tree species classification using uav images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLII**, 475–481 (2019).

Nevalainen, O. *et al.* Individual tree detection and classification with UAV-Based photogrammetric point clouds and hyperspectral imaging. *Remote Sens.* **9**, (2017).

Onishi, M. & Ise, T. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Sci. Rep.* **11**, 1–15 (2021).

Paszke, A. *et al.* Automatic differentiation in PyTorch. *31st Conf. Neural Inf. Process. Syst. Long Beach, CA, USA* (2017) doi:10.1145/24680.24681.

Roberts, D. A., Green, R. O. & Adams, J. B. Temporal and spatial patterns in vegetation and atmospheric properties from AVIRIS. *Remote Sens. Environ.* **62**, 223–240 (1997).

Safonova, A. *et al.* Detection of fir trees (Abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sens.* **11**, (2019).

Shen, X. & Cao, L. Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. *Remote Sens.* **9**, (2017).

Sothe, C. *et al.* Tree species classification in a highly diverse subtropical forest integrating UAV-based photogrammetric point cloud and hyperspectral data. *Remote Sens.* **11**, (2019).

Tan, M. & Le, Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. *36th Int. Conf. Mach. Learn. ICML 2019* **2019-June**, 10691–10700 (2019).

Treuhaft, R. N., Asner, G. P., Law, B. E. & Van Tuyl, S. Forest leaf area density profiles from the quantitative fusion of radar and hyperspectral data. *J. Geophys. Res.* **107**, (2002).

Ustin, S. L. *et al.* Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sens. Environ.* **113**, S67–S77 (2009).

Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).

Waser, L. T., Küchler, M., Jütte, K. & Stampfer, T. Evaluating the potential of worldview-2 data to classify tree species and different levels of ash mortality. *Remote Sens.* **6**, 4515–4545 (2014).

CHAPTER 4
Biodiversity Assessment in Bornean Tropical Rain Forests Using UAVs to Identify Indicator Tree Species Combined with Deep Learning

4.1 Introduction

Tropical forests account for almost 50% of the global forest area (FAO, 2020) and are rich in biodiversity (Gardner et al., 2009). Currently, these tropical forests are undergoing deforestation and forest degradation owing to human activities such as land conversion and commercial logging (FAO, 2020; Hansen et al., 2009; Margono et al., 2012), thereby accounting for approximately 20% of anthropogenic carbon emissions (van der Wer et al., 2009) and posing a threat to global biodiversity (Giam, 2017). In particular, Borneo, one of the most biologically diverse island (Myers et al., 2000), was subjected to deforestation and forest degradation, wherein 34% of old-growth forests were cleared between 1973 and 2015 owing to commercial logging and industrial plantations of oil palm, pulpwood, and other species (Abood et al., 2015; Gaveau et al., 2016).

Two financial mechanisms have been deployed to conserve these forests. The first approach includes Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+). This framework was adopted as an agendum by the United Nations Framework Convention on Climate Change (UNFCCC) in 2007 (UNFCCC, 2008); in total, 50 developing countries submitted a REDD+ forest reference level or forest reference emission level for technical assessment to the UNFCCC in 2020. In this mechanism, the tropical countries can receive financial rewards by reducing greenhouse gas (GHG) emissions from deforestation and forest degradation. The second approach is forest certification, which represents certification by the Forest Steward Council (FSC). The timber produced from the forests, which fulfills the sustainability requirements and is certified by such third-party organizations, can gain improved market access. Thus, forest managers gain incentives in both the aforementioned approaches when they conserve forests, prevent excessive logging, and care for sustainability.

In these programs, monitoring and assessment of biodiversity is crucial. While the main objective of REDD+ is to reduce emissions, only focusing on the amount of carbon will have negative environmental impacts; for example, low-carbon forests with high biodiversity will be displaced by high-carbon forests with poor biodiversity. To avoid this situation, biodiversity assessment was introduced as a safeguard in 2010. In particular, various REDD+ projects adopted the Climate, Community, and Biodiversity (CCB) standard, in which biodiversity is not harmed, and positive environmental impacts are generated (Narasimhan et al., 2014). However, Panfil (2015) reported that the 80 REDD+ projects, which adopted the CCB standard, lacked quantitative targets for the biodiversity objectives and details on sampling design and methodology. Regular biodiversity monitoring, reporting, and verification systems are needed in REDD+, as well as in forest certification, which assesses the environmental aspects, including biodiversity, as one of the criteria (principles) of sustainability (FSC, 2021).

Biodiversity can be assessed based on multiple aspects, such as existence of endangered animals, species richness, and tree composition (Barlow et al., 2007; Lawton et al., 1998; Schulze et al., 2004; Struebig, 2013; Uehara-prado et al., 2009). Among these, tree-community composition monitoring could be a cost-effective manner because changes in canopy tree-composition can be detected from remotely sensed data (Fujiki et al. 2016, Kitayama et al. 2018). For example, Fujiki (2016) used Landsat satellite data for estimating overall tree-community composition based on all genera obtained by field survey. Even though resolution of the satellite is 30 m which is not enough for figuring each tree species out, multi-spectral reflectance values of each pixel and the texture information (relationship to its neighboring pixels) reflected the tree-community composition (adjusted R2 value ranged from 0.59 to 0.69). Further, the tree-community composition revealed a significant correlation with an index of forest degradation (remaining aboveground biomass, AGB) and has exhibited its appropriateness as a forest biodiversity indicator compared with other indicators such as tree species richness in logged-over Bornean forests (Imai et al., 2014).

This tree-community composition was calculated with a nonmetric multidimensional scaling

(nMDS) using the Chao index (Chao 2004) of all tree species (or genera) present in count plots. The greatest variation in tree-community composition occurs along the axis-1 of the nMDS coordinate and axis-1 values of the plots correlate with the magnitude of forest degradation derived from logging (or inversely intactness), provided that all plots have the same original vegetation. In this system, axis-1 values of nMDS are used as an index of forest intactness (i.e. forest biodiversity) (Imai et al., 2014, Fujiki et al. 2016, Kitayama et al. 2018).

This index can be extrapolated to a wider landscape at the forest management unit (FMU) scale as a digital map (the method "BOLEH, biodiversity observation for land and ecosystem health; Kitayama et al., 2018). This method combines count-plot sampling on the ground and an extrapolation using a satellite imagery. A total of 50 circle plots with a 20-m radius and the identification of all genera present are needed in the guideline. Landsat reflectance values (and texture matrixes) corresponding to the tree-community composition index (nMDS axis-1 values) of the 50 plots are extrapolated to the entire area of an FMU (Fujiki et al., 2016; Kitayama et al., 2018). This method is robust in elucidating forest degradation (or intactness) in terms of tree-community composition, which is not possible from the biomass aspect (Kitayama et al., 2018). However, this method is costly because a survey of a large number of count-plots is involved. Moreover, the same procedure must be repeated when updating the map.

Two guilds of tree genera are actually known to be involved for the variation of tree-species composition in these Bornean forests; one is a pioneer guild which increases in abundance with degradation, and the other is a climax guild which decreases with degradation. Therefore, mixing ratio of these two guilds can actually explain the magnitude of forest degradation (or intactness). Furthermore, the same genera are involved for a wider region spanning entire Borneo, i.e. the genera *Neolamarckia* and *Macaranga* for the pioneer guild and the dipterocarp genera (genera in Dipterocarpaceae) for the climax guild (Aoyagi et al., 2017). The aerial survey of the abundance of the selected indicators only may be able to provide the same information as BOLEH.

Recently, unmanned aerial vehicles (UAVs) have gained attention for helping with forest

surveys (Paneque-Galves et al., 2014). Specifically, tree height estimation, tree enumeration, and tree size evaluation are major factors that can be assessed using UAVs (Goodbody et al., 2017; Iizuka et al., 2018; Mlambo et al., 2017; Tang et al., 2015). Furthermore, aerial image identification of tree species is an important factor that may help in forest management and contribute to biodiversity monitoring.

To date, aerial image identification of tree species has been attempted using airborne and satellite imagery. Recently, airborne hyperspectral images have revealed high potential for identifying several tree species (Dalponte, 2012; Féret and Asner, 2013; Shen and Cao, 2017). Although airborne identification using airplanes is advantageous as it covers large areas, it is costly. In recent years, UAV imagery has revealed immense potential in identifying tree species. In particular, application of deep learning to UAV RGB imagery has succeeded in identifying several trees in an economic manner (Csillik et al., 2018; dos Santos, 2019; Natesan, 2019; Onishi and Ise, 2021; Safonova, 2019). Onishi and Ise (2021) reported that applying deep learning to RGB images captured by an UAV could identify tree types and tree species with high accuracy in Japan. In addition, the features that Convolutional Neural Network (CNN) used revealed that unique texture patterns created by structural differences such as tree shapes or branching patterns were extracted. Furthermore, an identification model using deep learning is highly robust because it relies on structural features, not spectral differences that can change and are subject to weather, shade, and solar altitude (Onishi and Ise, 2021).

For estimating tree-community composition from UAV, indicator tree species information detected from UAV imagery, especially tree crown areas of each tree species were expected to be significant clue for estimating the tree-community composition. However, overall tree species cannot be identified, and understory vegetation cannot be monitored from UAV. For more accurate estimation of the tree-community composition, other forest information might be helpful. First one is forest structure. Among many forest structure parameters, maximum and range of tree height can be indexes for tree community composition. Maximum tree height increases along with succession (Peña-Claros, 2003), and it can be useful index for species richness (Marks et al.,

2016). Marks et al. (2017) shown that maximum and range of tree height has correlation with alpha and beta diversity. Second is the condition of surrounding forest. Forest community can be affected by surrounding condition. For example, forming distinct edge by constructing roads can increase tree mortality (Prasad, 2009) via microclimatic change (Kapos,1989; Williams-Linera, 1990), mechanical damage (Chen et al., 1992), and high infestation rates by pathogens (Dickie & Reich 2005). Further, it facilitates exotic plant invasion (Gelbards and Belnap, 2003; Honnay et al., 2008; Lugo and Gucinski, 2000), and lead to invasive plants colonization (Fensham et al. 1994, Laurance, 1991).

To achieve the tree-community composition estimation from UAV, I set my study objectives as: 1) evaluating the potential of combining UAV and deep learning for identifying indicator tree species (genera) in Bornean tropical rain forests, and 2) clarifying relationship between overall tree-community composition based on count-plots on the ground, and the identified tree species crown area and forest structure information of UAV imagery. Considering the latter objective, I analyzed the appropriate size of projected area on the ground when analyzing crown aeras of the target genera and forest structure in order for reflecting the surrounding forest conditions. In addition to the tree-community composition, the aboveground biomass, which is simultaneously obtained from field survey was also evaluated using UAV imagery because biomass may covary with tree-community composition.

4.2 Materials and Methods

4.2.1 Study sites

This study was conducted mainly in two Bornean FMUs where legal commercial logging was being conducted. The mode of logging is selective in these FMUs. The studied FMUs were Deramakot (5°14'–28'N, 117°20'–38'E, 551 km^2), and Tangkulap (5°18'–31' N, 117°11'–22'E, 276 km^2) in Sabah, Malaysia. Supplementary inventory data were collected at Segaliud Lokan (5°20'–27' N, 117°23'–39' E, 576 km^2) (Figure 4.1).

In the Deramakot FMU, whole logging was suspended in 1989, and reduced-impact logging was introduced in 1995. In 1997, the FMU was certified by the FSC for the first time as a tropical timber forest. This certification was extended to 2019. In the Tangkulap FMU, conventional logging was conducted until 2003. Logging was suspended in 2003 and the entire FMU was certified by the FSC in 2011. In the Segaliud Lokan FMU, conventional logging was conducted until 2002. Since 2002, reduced-impact logging has been conducted and is currently certified by the Malaysian Timber Certification Council (MTCC).

I conducted a basic research such as flying UAVs, field surveys, tree identification, and analysis of biodiversity at Deramakot and Tangkulap. Biodiversity scores (i.e. community composition index as explained in the introduction) were derived based on the count-plots on the ground of Deramakot and Tangkulap, supplemented by field inventory data of Segaliud Lokan as will be explained later. 1).

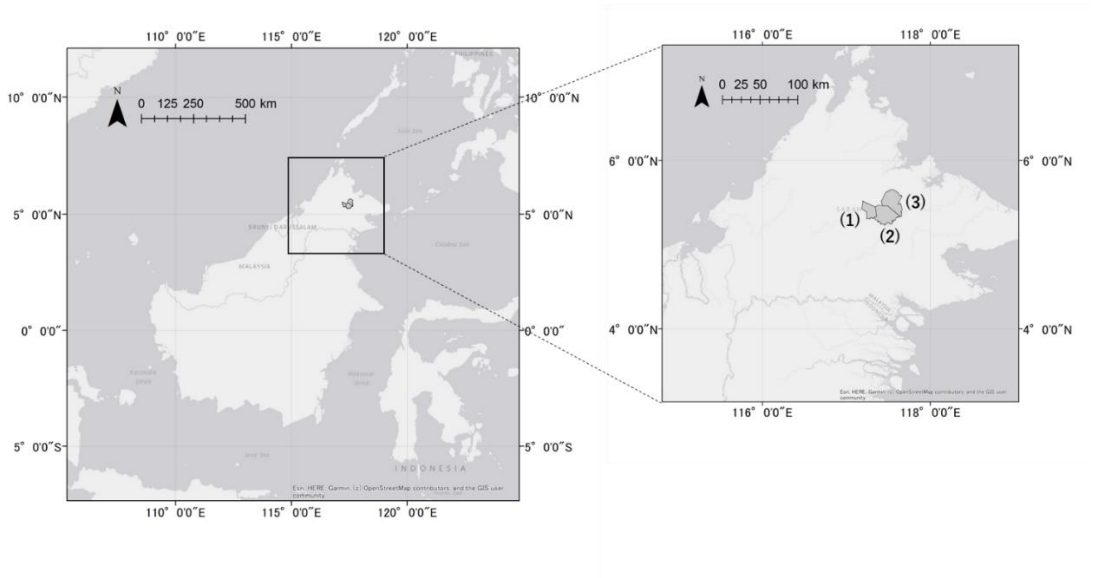


Figure 4.1 Location of study sites: (1) Tangkulap, (2) Deramakot, and (3) Segaliud Lokan FMU

4.2.2 Field survey

I conducted a field survey in the Deramakot and Tangkulap FMUs in 2019. Two types of plots were used. First, I researched three rectangular plots (R1-3) with dimensions of 100 * 200 m. These plots were established and used in a previous study (Imai et al., 2012); in the present study, I used them to obtain accurate supervised data and confirm that a researcher can visually identify the objective tree species from UAV imagery. R1 is a primary forest, R2 is a conventional logging forest, and R3 is a reduced-impact logged forest. At the plot and in the surrounding area, I checked the tree crown image obtained by the UAV and corresponding trees and tagged the tree species name to UAV imagery as a ground truth label. I used R3 as the test site for tree species identification.

The other plot was a circular plot with a 20 m radius. I researched 28 plots in total, 8 plots in Tangkulap and 20 plots in Deramakot. Any two plots have a distance of at least 100 m in-between and used in previous studies to assess the biodiversity (Aoyagi et al., 2017; Fujiki et al., 2016; Kitayama et al., 2018). In the present study, I used these small plots to analyze overall tree-community composition based on count-plots on the ground, and UAV imagery data. In the field

survey, I measured the girth at breast height (GBH) of all trees with ≥15 cm GBH. Woody vines were excluded from the study. The global positioning system location (GPS) of the center of a plot was recorded by averaging 2 h using portable global positioning system (GPS) in 2014. Trees with buttress were measured above (approximately 50 cm) the protrusions. While all trees were identified up to the species level, I analyzed the data only at the genus level.

4.2.3 Objective tree species

In this study, I focused on tree species of the genera *Macaranga* and *Neolamarckia*, which are indicator tree species of degraded forests with respect to tree community composition (Aoyagi et al., 2017). In my study area, the genus *Neolamarckia* comprised only *Neolamarckia cadamba.* In contrast, the genus *Macaranga* has several species. Thus, I classified the species into three classes based on their leaf size: *Macaranga gigantea*, *Macaranga pearsonii*, and *Macaranga conifera*. *The M. pearsonii* class includes other similar species, such as *Macaranga hypoleuca*, *Macaranga beccariana*, *and Macaranga triloba*, which present almost similar size and shape of leaves (Figure 4.2).

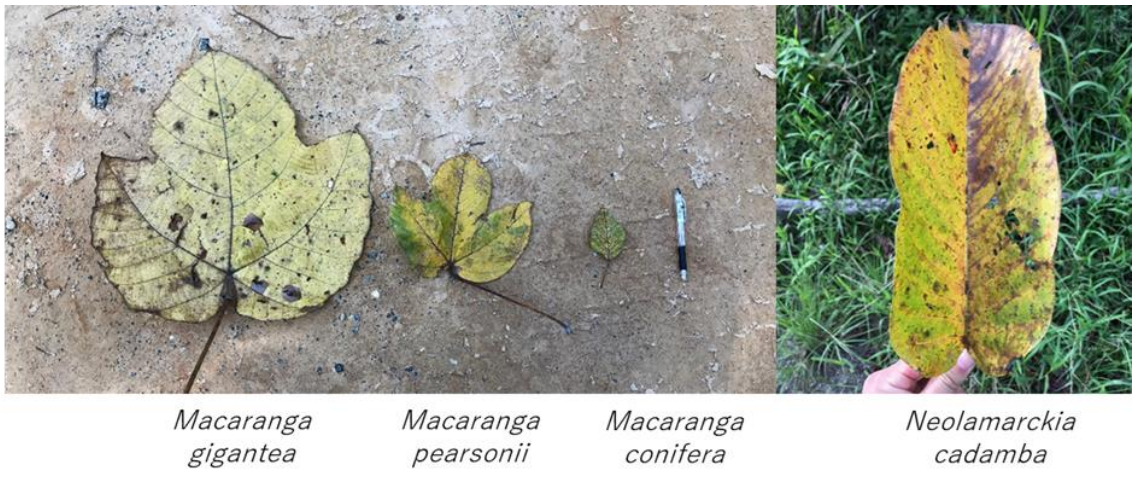


Figure 4.2 Objective *Macaranga* species leaf image.

4.2.4 UAV flight

In this study, I used two types of UAVs, Phantom 4 Pro and Mavic 2 Pro (DJI, China), as I obtained data simultaneously and checked the robustness of identification performance between the UAVs. In both Phantom 4 Pro and Mavic 2 Pro, the camera records approximately 20 million pixel images; however, the viewing angle slightly differs (84° and 77°, respectively).

I conducted automatic flight in 2019, using UgCS v3.2 software (SPH engineering, Baložu pilsēta, Latvia), which enabled us to set the relative flight altitude from the ground surface, overlaps, and others. Herein, the flight parameters were set as flight altitude: 100~120 m (GSD:2.5~3.3 cm) from the Shuttle Radar Topography Mission (SRTM), front overlap: 90%, and side overlap: 80%. I conducted the flight from 11:00 to 13:00 to avoid shade influence. The flight covered enough the whole area of rectangular plots, and I marked approximately 250 × 250 m area around the circular plots.

4.2.5 UAV data processing

For applying deep learning to UAV imagery and analyzing with field-obtained data, I processed UAV imagery. I present the technical workflow of the data processing in Figure 4.3. As a flow, I made orthophoto and 3-dimensional data from UAV imagery, and made polygons and canopy height model (CHM). I extracted tree crown image from orthomosaic photo and tree crown map which is higher than 5 m of CHM, and applied deep learning. As for analysis with field-obtained data, I made tree species map from tree crown map, orthomosaic photo and deep learning, gap map from polygons, and calculated forest structure information from CHM. I explain the details of the procedure below.

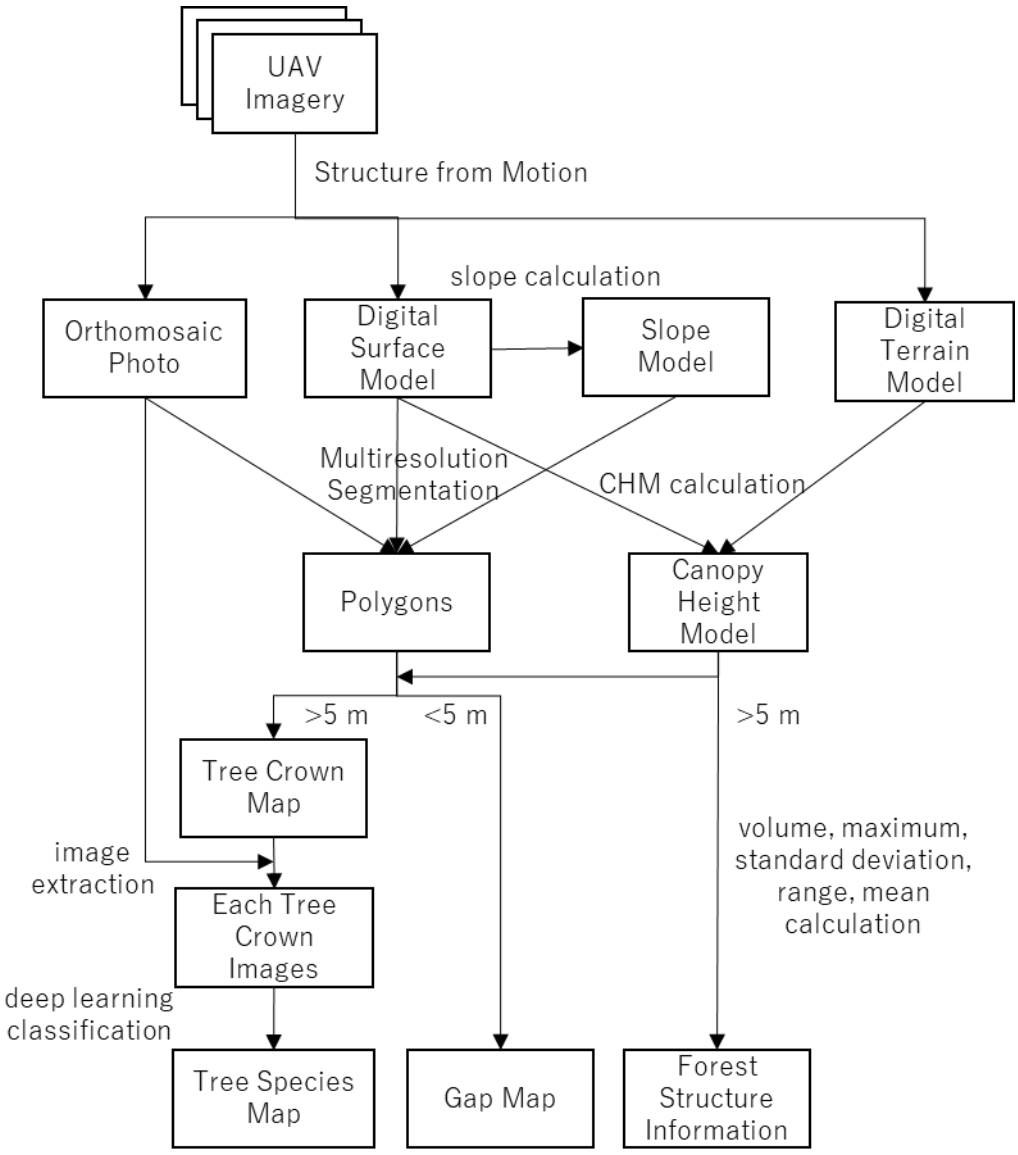


Figure 4.3 Technical workflow of data processing

4.2.5.1 UAV data preprocess

First, from the UAV data, I created an orthomosaic image, digital surface model (DSM), and digital terrain model (DTM) by using the structure from motion technique in photoscan software (Agisoft, Russia). The parameters required for detecting ground points are set at each study site by trial and error to create a DTM. For the R3 plot, I created flights using both Phantom 4 Pro and Mavic 2 Pro, and I adjusted the GPS location error via georeferencing.

For GIS processing, I used ArcGIS software (ESRI, USA). I calculated the slope from the DSM

to create the slope model. Moreover, I created a canopy height model (CHM) by subtracting DTM from DSM. I separated each tree crown using an orthomosaic photo, DSM, and slope model using a multiresolution segmentation algorithm (Baatz and Schäpe, 2000) in eCognition software (Trimble, USA). This method enables us to demarcate each tree species crown as a segment (Onishi and Ise, 2021). The parameters for multiresolution segmentation were set to {Weight of R, G, B, DSM, Slope:1, 1, 1, 2, 3, Scale: 200, Compactness: 0.5, Shape: 0.2}.

After segmentation, I divided these polygons into tree crown and gap polygons using the height information of CHM; gap areas are defined as the stature <5 m on CHM and tree areas as the stature ≧5 m on CHM.

Thereafter, I attached tree species labels to the tree polygons. For the rectangular plots, I attached each species label using field survey results; however, for the circle plots, I judged only the indicator species by visually inspecting photos. Next, I extracted each tree crown image using a tree crown map with ground truth labels and orthomosaic photos (Figure 4.4). Furthermore, for the tree areas, I calculated three-dimensional volume, maximum, range, standard deviation, and mean of CHM as representative variables of forest structure at five scales (20, 40, 60, 80, and 100 m radius) from the center of plots respectively.

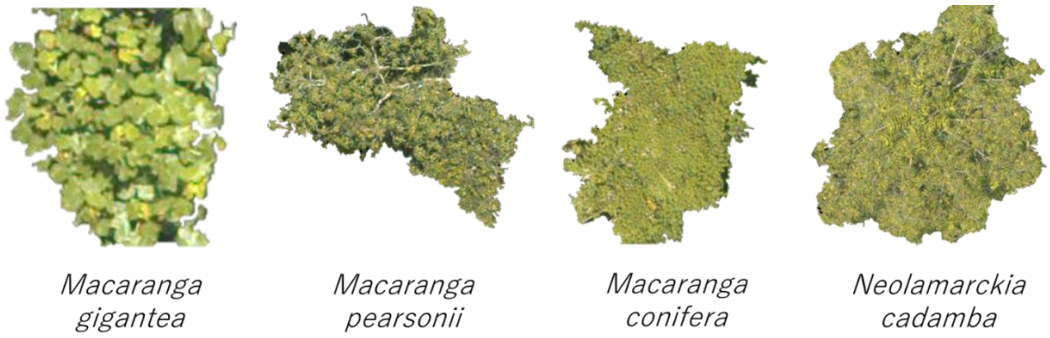


Figure 4.4 Representative images of indicator tree species from the left: *Macaranga gigantea*, *Macaranga pearsonii*, *Macaranga conifera*, and *Neolamarckia cadamba*.

4.2.5.2 Deep learning

For the training and validation datasets, I used photo of two rectangular plots (R1-R2), which interpreted from the field, and some photo of circle plots, which interpreted from UAV photo. Regarding to the latter dataset, I only used photo except for those present within a 100 m radius from the center of the plot, because for regression analysis within the radius, I should use the tree crown data that are identified initially via deep learning. Moreover, I previously augmented approximately 2000 indicator tree classes because if there is a bias in the dataset, the model would be adjusted to classify the images into a class with a large amount of data. For the classification test, I used the remaining 2 ha rectangular plots (R3) (Table 4.1).

I applied deep-learning identification using a convolutional neural network, wherein, I used PyTorch (Paszke et al., 2017) as the framework and pretrained EfficientNet B7 (Tan et al., 2020) as the neural network model. During the training phase, I conducted fine-tuning, that is, I changed the output of the fully connected layer to a value of 5, and then trained all parameters of each CNN layer. The parameter settings are summarized in Table 4.2. The input size was 224 × 224 pixels.

Furthermore, to improve the classification performance, I applied a threshold to each probability identification value. Thus, for each image of deep learning identification, I can obtain the probability value of individual class. The system that is commonly used adopts the top-class label with the highest probability; however, the other tree species that are absent in the training and identified as indicator tree species may exhibit low probabilities of the predicted class because this image does not reveal any features similar to the five classes. Therefore, I used the threshold value for probability, which indicates that the image with lower probability than the threshold is classified as another tree species class. The threshold value is set at each class and adjusted to obtain higher overall classification accuracy. This value is similar for Phantom 4 Pro and Mavic 2 Pro; however, it differs among classes. To evaluate the accuracy, I used the intersection over union (IoU) score for objective tree species segmentation using the ground truth polygons and predicted segmented polygons that are merged from the polygons; more than 50% of these belong

to the ground truth polygon, and precision, recall, and Kappa coefficient (Cohen, 1960) for deep learning classification.

For clarifying the relationship between overall tree-community composition obtained by field survey and the identified tree species from UAV, I used the deep learning model with threshold and made tree species map at twenty-eight areas. Then I calculated the tree crown areas of *M. gigantea, M. pearsonii, M. conifera, N. cadamba*, other tree species and gap at five scales (20, 40, 60, 80, and 100 m radius) from the center of plots respectively.

Table 4.1 Data summary

Class ID	Species	Training	Validation	Test
0	Other tree species	1874	475	453
1	*Macaranga conifera*	90	13	29
2	*Macaranga gigantea*	40	12	16
3	*Macaranga pearsonii*	116	29	38
4	*Neolamarckia cadamba*	199	64	7

Table 4.2 Parameter settings of deep learning

Parameter	Value
Network	EfficientNet B7
Epoch	100
Batch size	16
Learning rate	0.005
Momentum	0.9
Optimizer	SGD

4.2.6 Data analysis

For evaluating tree-community composition from field survey, I calculated Chao's distances (Chao et al., 2005) between any two plots using the number of trees of each genus recorded at field survey in twenty-eight circular plots, and applied nMDS analysis to plot data with the meta-MDS package in the vegan package of R software (Oksanen et al., 2013). I used nMDS axis-1 score as index of tree-community composition for subsequent analysis.

The AGB of each plot was calculated as a sum of the AGBs of all trees present based on the allometric equation obtained by Chave et al. (2014):

$$\text{AGB (kg)} = \rho \times \exp\left(-1.499 + 2.148\ln(\text{D}) - 0.207(\ln(\text{D}))^2\right) - 0.0281(\ln(\text{D}))^3)$$

where ρ represents wood-specific gravity (g cm^{-3}), and D represents DBH (diameter at breast height (cm)). Wood-specific gravity data were obtained from the Global Wood Density Database (Chave et al., 2009).

From the process above, I prepared AGB and nMDS axis-1 scores from field survey, and tree species map and CHM from UAV and deep learning in twenty-eight areas. Then I set field-obtained nMDS axis-1 score and log(AGB) as dependent variables, and UAV obtained indicator tree species information (tree crown area of *M. gigantea, M. pearsonii, M. conifera, N. cadamba*, other tree species and gap) and forest structure (three-dimensional volume, maximum, standard deviation, range and mean of tree part of CHM) as independent variables, and applied multiple regression analysis for discovering the relationship between the variables. From those independent variables, valuable parameter was chosen using forward–backward stepwise selection based on Akaike Information Criterion (AIC) (Akaike, 1974); thereafter, only variables with significant values ($p <$ 0.05) were selected. I applied this analysis to the independent variables derived at five scales (20, 40, 60, 80, and 100 m radius) from the center of plots respectively for reflecting surrounding forest conditions such as existence of indicator tree species or gap areas.

4.3 Results

4.3.1 Tree species identification

4.3.1.1 Tree crown segmentation

After calculating the IoU value of indicator tree species (Figure 4.5), I observed that *M. gigantea* exhibits a lower value (~0.6) than other species (~0.8); however, for *M. pearsonii*, the average value is high but it exhibits a higher range (<0.3 to >0.9) than that of the other species. *M. conifera* and *N. cadamba* exhibited high average values with a small range.

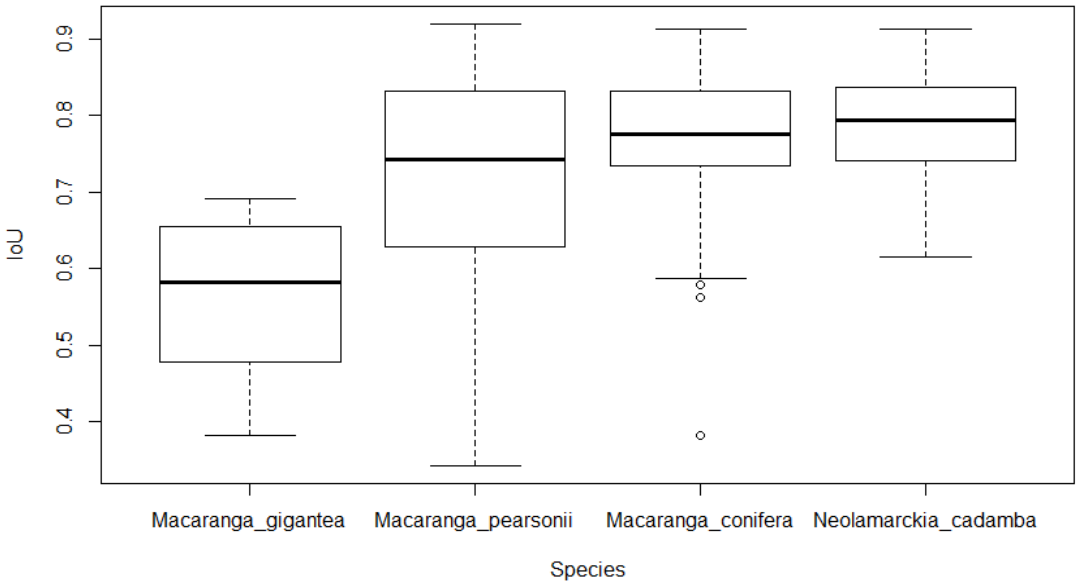


Figure 4.5 IoU value of each class.

4.3.1.2 Identification of indicator tree species

The CNN model succeeded in identifying these indicator tree species with high accuracy for Phantom 4 Pro data (Kappa 0.802) and a slightly lower accuracy for Mavic 2 Pro data (Kappa 0.591). I have summarized the details of the identification result as a confusion matrix in Table 4.3. Most classes were identified at higher than 90% precision in Phantom 4 Pro, except for *M. conifera* (63%). Moreover, most classes were classified at higher >85% recall, including *M. pearsonii* that exhibited relatively high accuracy (73%). However, the accuracy of each class in

Mavic 2 Pro was lower than that of Phantom 4 Pro. In particular, some other tree species were misclassified as *M. conifera* (56% precision), and most species of the *M. pearsonii* class were misclassified as another tree species (33% recall). In contrast, *M. gigantea* and *N. cadamba* were consistently identified in both the precision and recall (more than 75%).

Table 4.3 Result of indicator tree species identification using Phantom 4 Pro and Mavic 2 Pro.

phantom 4 pro

CNN predicted class	field reference class					Precision
	Macaranga gigantea	*Macaranga pearsonii*	*Macaranga conifera*	*Neolamarckia cadamba*	other tree species	
Macaranga gigantea	19	1	0	0	0	95.0%
Macaranga pearsonii	0	69	0	0	1	98.6%
Macaranga conifera	0	1	34	0	19	63.0%
Neolamarckia cadamba	0	0	0	10	0	100.0%
other tree species	1	23	6	0	433	93.5%
Recall	95.0%	73.4%	85.0%	100.0%	95.6%	Kappa: 0.802

mavic 2 pro

CNN predicted class	field reference class					Precision
	Macaranga gigantea	*Macaranga pearsonii*	*Macaranga conifera*	*Neolamarckia cadamba*	other tree species	
Macaranga gigantea	15	2	2	0	1	75.0%
Macaranga pearsonii	1	31	0	0	0	96.9%
Macaranga conifera	0	5	32	0	20	56.1%
Neolamarckia cadamba	0	0	0	10	2	83.3%
other tree species	4	56	7	0	430	86.5%
Recall	75.0%	33.0%	78.0%	100.0%	94.9%	Kappa: 0.591

I set the classification threshold: 0.5, 0.7, 0.4, and 0.8 for *M. conifera*, *M. gigantea*, *M. pearosnii*, and *N. cadamba,* respectively. These values were set to achieve better scores in each class for both UAVs. Although I expected this threshold to improve the performance, only a slight increase was observed in the Kappa score ranging from 0.002 to 0.005 (Table 4.4).

Table 4.4 Result of indicator tree species identification using Phantom 4 Pro and Mavic 2 Pro with threshold (0.5, 0.7, 0.4, and 0.8 for each tree species).

phantom 4 pro with threshold

CNN predicted class	field reference class					Precision
	Macaranga gigantea	*Macaranga pearsonii*	*Macaranga conifera*	*Neolamarckia cadamba*	other tree species	
Macaranga gigantea	16	0	0	0	0	100.0%
Macaranga pearsonii	0	69	0	0	0	100.0%
Macaranga conifera	0	1	34	0	19	63.0%
Neolamarckia cadamba	0	0	0	10	0	100.0%
other tree species	1	24	6	0	434	93.3%
Recall	94.1%	73.4%	85.0%	100.0%	95.8%	Kappa: 0.804

mavic 2 pro with threshold

CNN predicted class	field reference class					Precision
	Macaranga gigantea	*Macaranga pearsonii*	*Macaranga conifera*	*Neolamarckia cadamba*	other tree species	
Macaranga gigantea	15	1	0	0	0	93.8%
Macaranga pearsonii	1	31	0	0	0	96.9%
Macaranga conifera	0	5	32	0	20	56.1%
Neolamarckia cadamba	0	0	0	10	0	100.0%
other tree species	4	57	9	0	433	86.1%
Recall	75.0%	33.0%	78.0%	100.0%	95.6%	Kappa: 0.596

Furthermore, I visualized the identification result, which was classified by the CNN model with a threshold for the whole area and an enlarged area of one circle plot (Figure 4.6).

Using this segmentation system and CNN model, I succeeded in identifying and mapping the indicator tree species.

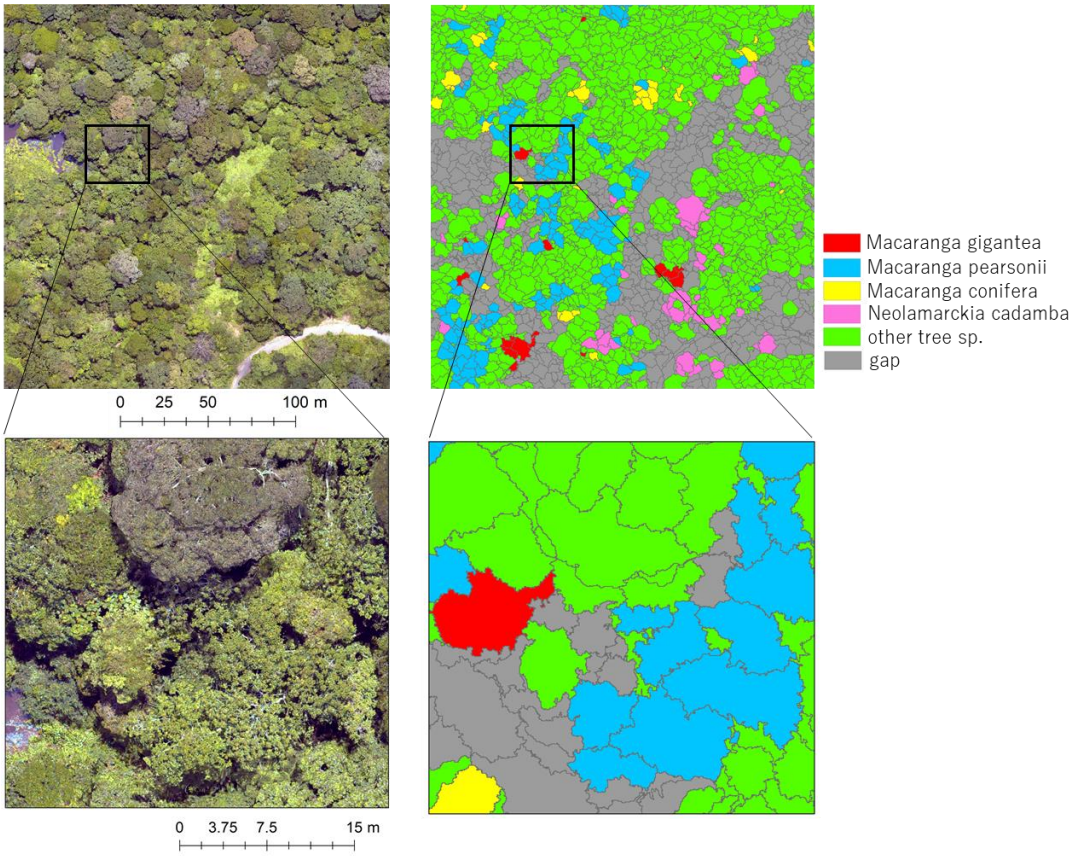


Figure 4.6 Visualization of the result of deep learning classification. The UAV image is depicted on the left and the segmented and classified image is on the right.

4.3.2 Biodiversity assessment

4.3.2.1 Evaluating the relationship between field-obtained nMDS axis-1 score (tree community composition based on all genera) and AGB

First, I evaluated the relationship between the nMDS axis-1 score and AGB, which were obtained from a field survey. In a previous study, they were positively correlated with each other (Aoyagi et al., 2017; Imai et al., 2014). My data revealed an obvious outlier plot, D4 (Figure 4.7). For further analysis, I excluded plot D4 for analysis with the nMDS axis-1 score. Except for plot D4, a significant correlation was observed between $p < 0.001$ and R2 = 0.55 (Figure 4.7).

In Figure 4.7, the plots in Deramakot reveal high log(AGB) and nMDS axis-1 scores. In contrast, the plot D3 exhibits a high log(AGB) and low nMDS axis-1 score

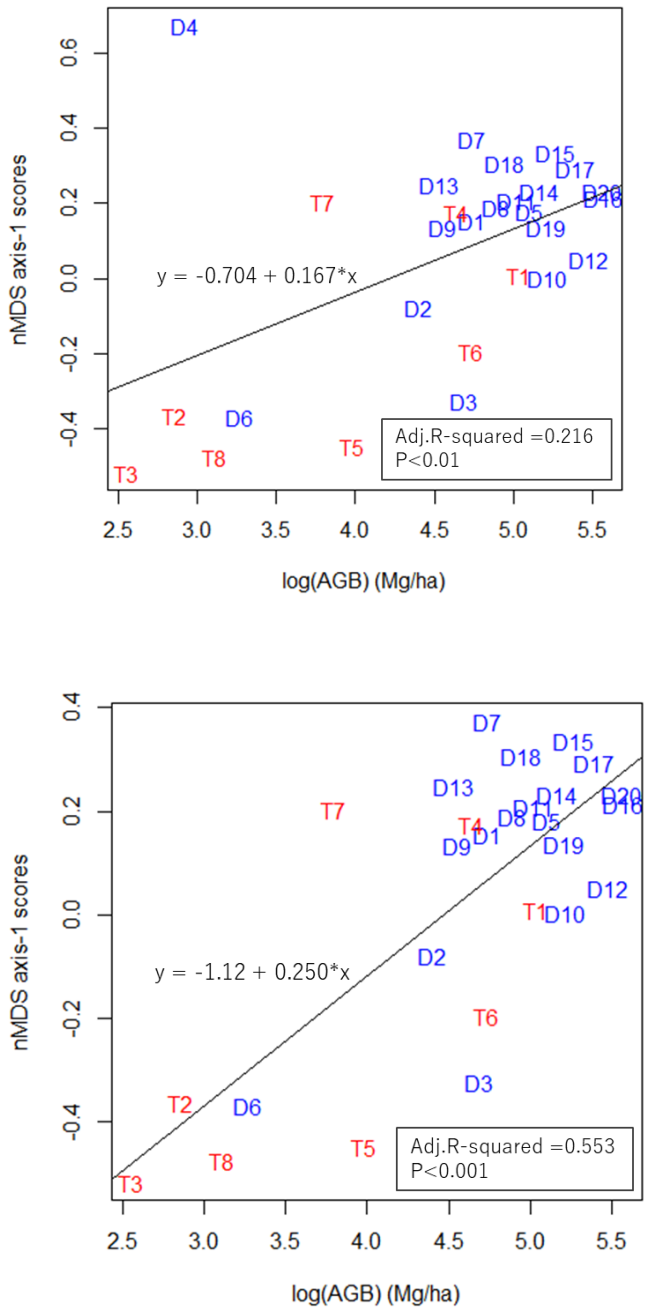


Figure 4.7 Result of field-obtained nMDS axis-1 score (overall tree community composition based on all genera) and aboveground biomass of each plot. D and T represents Deramakot and Tangkulap. As a result, the plot number D4 was detected as an outlier. Bottom image depicts the relationship without outlier D4 plot.

4.3.2.2 Relationship between field obtained index (AGB and overall tree-community composition based on all genera) and UAV obtained data (tree crown areas of indicator tree species and forest structure information)

I applied multiple regression analysis with stepwise selection for UAV obtained indicator tree species information (tree crown area of *M. gigantea*, *M. pearsonii*, *M. conifera*, *N. cadamba*, other tree species and gap) and forest structure (three-dimensional volume, maximum, standard deviation, range and mean of tree part of CHM) as independent variables , and field-obtained nMDS axis-1 and AGB score as dependent variables from 20 to 100 m radius (Figure 4.8).

The AIC score was lowest at the 40-m radius, considering the correlation with nMDS axis-1 score. In contrast, for the AGB, the AIC score was lowest at a radius of 20 m, which was the same size as the field plot, and this score increased with increasing radius. I summarized the details of the results and the parameters obtained by stepwise selection (Table 4.5).

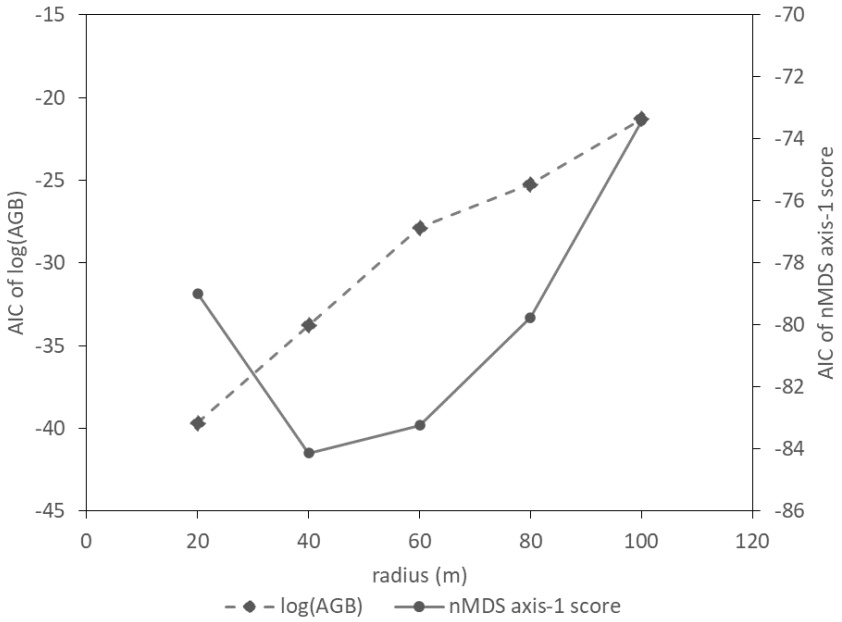


Figure 4.8 Result of relationship between radius of UAV data analysis and field-obtained nMDS axis-1 score (overall tree-community comsition) and log(AGB). For analysis, I applied multiple regression analysis with stepwise selection.

Table 4.5 Detailed result of multiple regression analysis

nMDS axis-1 score

Radius	AIC	Adj. R^2	*p*	Parameter 1	*p*1	Parameter 2	*p*2	Parameter 3	*p*3
20	−79.0	0.342	***	Other tree sp.(+)	***				
40	−84.1	0.474	***	Other tree sp.(+)	**	*M. gigantea*(-)	*		
60	−83.2	0.456	***	*M. gigantea*(-)	**	Gap(-)	*		
80	−79.8	0.382	**	*M. gigantea*(-)	**	Gap(-)	*		
100	−73.4	0.191	*	Other tree sp.(+)	*				

log(AGB)

Radius	AIC	Adj. R^2	*p*	Parameter 1	*p*1	Parameter 2	*p*2	Parameter 3	p3
20	−39.7	0.705	***	Gap(-)	***				
40	−33.8	0.636	***	Gap(-)	***				
60	−27.9	0.551	***	Gap(-)	***				
80	−25.2	0.537	***	Gap(-)	**	*M. gigantea*(-)	**	Other tree sp.(+)	*
100	−21.2	0.467	***	Gap(-)	**	*M. gigantea*(-)	*	Other tree sp.(+)	*

Significant Code *** 0.001, ** 0.01, * 0.05
+ or – means positive or negative of parameter estimate

As a result of stepwise selection, for the nMDS axis-1 score, other tree species or gap was chosen at all radii, and *M. gigantea* was selected at 20, 40, and 60 m radius. The adjusted R^2 value was the highest (0.474) at 40 m radius.

For the AGB, a gap was chosen at all radii. Additionally, *M. gigantea* and other tree species were also selected at 80 and 100 m radius. The adjusted R^2 value was the highest (0.705) at a 20

m radius.

I derived the best model for estimating nMDS axis-1 score (overall tree-community composition based on all genera) and AGB at the radius where AIC was lowest, as follows:

$$\text{nMDS axis} - 1 \text{ score} = -0.20066 + 0.48409 * \text{other tree species} - 21.45804 * \textit{Macaranga gigantea} \text{ (40 m radius)} \quad (1)$$

$$\text{Log(AGB)} = 5.2743 - 2.2382 * \text{gap (20 m radius)} \quad (2)$$

4.4 Discussion

4.4.1 Tree species identification

4.4.1.1 Tree crown segmentation

In tree crown segmentation, the IoU of *M. gigantea* is 0.6, and other indicator species is high (0.7~0.8). In general, the IoU scores 0.7~0.8 are considered high, even in deep learning segmentation (Wang et al., 2018). This result indicates that my method successfully segmented the objective trees with high accuracy; however, in reality, the polygon was separated into pieces, and the IoU value was calculated for the polygon that is merged with several polygons. In other studies, some researchers have applied deep learning methods such as instance segmentation of Mask RCNN (He et al., 2017), which can improve tree crown segmentation.

4.4.1.2 Indicator tree species classification

I succeeded to identify indicator tree species with Kappa accuracy values of 0.6~0.8. This performance is remarkable considering extremely rich tree species in my Bornean rain forests. Also, because this score was obtained at different sites and days from training data, my deep learning model was robust.

In contrast, certain differences exist between Phantom 4 Pro and Mavic 2 Pro. I used the sample of *M. pearsonii* of Phantom 4 Pro and Mavic 2 Pro (Figure 4.9). Here, I observed that the image of Phantom 4 Pro is clearer than that of Mavic 2 Pro. This may be due to difference in UAV, shooting conditions, and the influence of georeference. Considering the difference in UAV, Mavic 2 Pro is lighter than Phantom 4 Pro (907 g and 1388 g each); therefore, the image of Mavic 2 Pro is easily blurred. Another reason for this is the shooting conditions. Illumination and wind conditions can affect the image quality. Considering influence of georeference, I applied georeference for only the Mavic 2 Pro image to match the Phantom 4 Pro image; thus, the image might be blurred during the process.

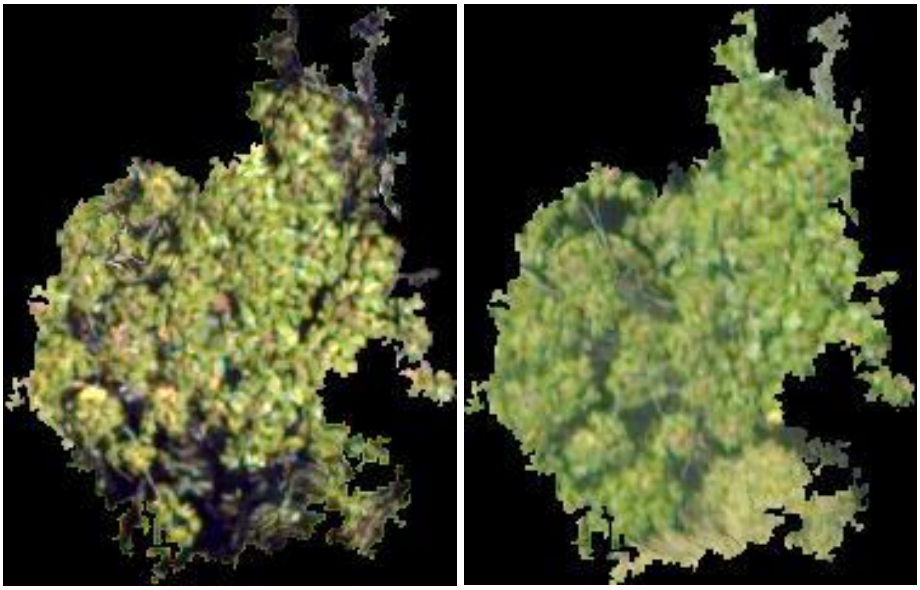


Figure 4.9 Representative image of *Macaranga pearsonii*. Left: Phantom 4 Pro; Right: Mavic 2 Pro.

In addition to those indicator tree species, I tried to identify Dipterocarpaceae, which is an indicator of the climax guild and the inclusion of this indicator would enhance the accuracy and robustness of the representativeness of overall tree-community composition (Aoyagi et al. 2017). However, I could not identify the family due to the small number of images. To identify tree species, I need at least fifty image data at the species level preferably (Onishi and Ise, 2021) or apply other deep learning methods such as few-shot learning (Li et al., 2019; Wang et al., 2020).

4.4.2 Biodiversity assessment

4.4.2.1 Relationship between field-obtained nMDS axis-1 score (tree-community composition based on all genera) and AGB

As a result of the field data evaluation, an outlier plot D4 was detected. Its tree community composition was unique compared to other plots due to the disproportionately high abundance of the genus *Dendrocnide* in D4. Therefore, the value of the Chao distance is markedly different from other plots, and D4 was emitted as an outlier in nMDS axis-1 value.

4.4.2.2 Relationship between field-obtained index (AGB and overall tree-community composition based on all genera) and UAV obtained data (tree crown areas of indicator tree species and forest structure information)

First, for AGB, the 20 m radius exhibited a best model, and the gap area was selected as a significant parameter in this model; in this case, 20 m radius is reasonable because it is the same size as that of the field survey area. Moreover, as the gap area was selected, maximum tree height and volume, which refer to three-dimensional total volume of tree class was also considered (Kachamba et al., 2016); however, the tree height estimation from UAVs cannot be accurate in closed forests, because the DEM was created by connecting the ground point detected in UAV imagery and the ground could not be observed in such closed forests (Obeng-Manu, 2019). Therefore, the gap was selected rather than the volume.

Furthermore, when explaining the biodiversity scores based on overall tree-community composition (i.e. nMDS axis-1 scores), the best model was derived at the 40-m radius, and the crown areas of other tree species (tree areas except for tree species indicative forest degradation and lower biodiversity) and *Macaranga gigantea* were selected as significant parameters. *Macaranga gigantea* has biggest leaves in *Macaranga* genus, and emerges at earlier successional stages than the other *Macaranga* species (Davies 1998). Moreover, *Macaranga gigantea* is especially shade intolerant among *Macaranga* genus with most trees being exposed to light (Davies 1998). This life history traits suggests that *Macaranga gigantea* can be a more significant indicator species of forest degradation, and emerges at canopy. Thus UAV could detect the most *Macaranga gigantea* trees. For this reason, crown area of *Macaranga gigantea* was thought to be chosen as a significant independent variable of tree-community composition.

The result that highest relationships at 40-m radius and higher relationships at 60-m radius was revealed than at 20 m radius suggests that surrounding forest condition may affect to the tree community composition. The existence of gap around the plot may suggest high mortality rate and invasion of exotic plant (Gelbards and Belnap, 2003; Honnay et al., 2008; Lugo and Gucinski, 2000; Prasad, 2009), and the existence of indicator tree species may suggest the small trees of

those species which exists under crown and cannot be detected from UAV. In addition, other indicator tree species were not selected as significant variable. If identification performance of CNN model was improved, the result of parameter selection of regression analysis could be changed.

As a trial, I mapped the AGB and biodiversity (nMDS axis-1 score) based on the aforementioned formula (1) (2) (Figure 4.10). For making the heat map, I set points with a spacing of approximately 10 m and calculated the AGB and nMDS axis-1 scores in the area with 20 and 40 m buffer, respectively, at each point. To estimate nMDS axis-1 score, values lower than −0.20066 were modified to −0.20066. For mapping, I used interpolation by spline; for nMDS axis-1 score, I applied an averaging filter because a huge difference was observed among the areas that include *M. gigantea* in the D3 plot, which contains high biomass but biodiversity degraded forests with respect to tree community composition (Figure 4.7). As illustrated in Figure 4.10, the D3 plot is covered by trees; however, there are many *Macaranga* trees, and large gap areas and logging roads are close to the forest. Thus, UAV data succeeded to evaluate the forest as degraded forest with respect to tree community composition as same as field survey. Though there are some improvement point and I need to improve the performance, the day may come when UAV and deep learning will be used for visualizing and estimating biodiversity for the practical use in the future.

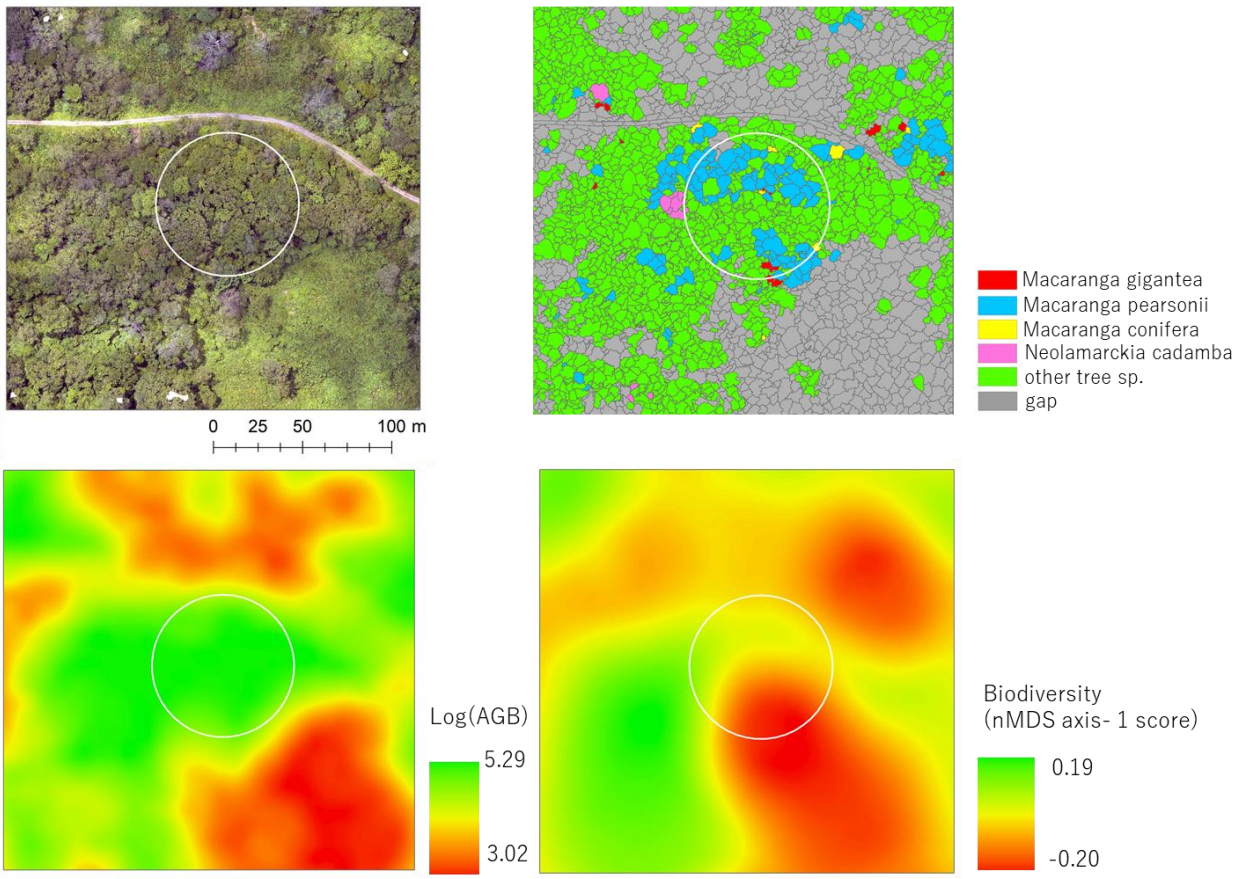


Figure 4.10 Estimation of log(AGB), and biodiversity (nMDS axis-1 score) in D3 plot. The white circle indicates 40 m buffer from the center of field plot.

4.5 Conclusion

In this study, I investigated the use of UAV remote sensing imagery and deep learning technology to identify and map the distribution of tree species indicative of forest degradation in logged-over tropical rain forests. My method and algorithm could successfully identify the target indicator tree species with a high accuracy. Furthermore, the tree crown areas of so identified indicator species could explain the tree-community composition based on all genera. Thus, the identification of key indicator species and mapping their crowns using UAV with my algorithm can be used as a cost-effective, biodiversity monitoring tool than the labor intensive survey that enumerates all tree genera on the ground.

There are still outstanding issues to be resolved for better assessment of forest canopy intactness at a larger scale. First, another important indicator group Dipterocarpaceae, which belongs to the climax guild, must be identified with UAVs. Second, the mapping by UAVs must be scaled up to wider areas such as FMU or lager. UAVs can cover at a scale of tens of hectare; however, forests need to be monitored at a larger scale when forest-management impacts and environmental safeguards are to be evaluated. Middle-resolution satellite imageries such as Landsat or Sentinel may be used for scaling up but appropriate algorithms must be developed.

4.6 References

Abood, S. A., Lee, J. S. H., Burivalova, Z., Garcia-Ulloa, J. & Koh, L. P. Relative contributions of the logging, fiber, oil palm, and mining industries to forest loss in Indonesia. *Conserv. Lett.* **8**, 58–67 (2015).

Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).

Baatz, M. & Schäpe, A. Multiresolution segmentation an optimization approach for high quality multi-scale image segmentation. in *Angewandte Geographische Informations-Verarbeitung, XII* 12–23 (2000).

Barlow, J. *et al.* Quantifying the biodiversity value of tropical primary, secondary, and plantation forests. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18555–18560 (2007).

Chao, A., Chazdon, R. L., Colwell, R. K. & Shen, T. J. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol. Lett.* **8**, 148–159 (2005).

Chave, J. *et al.* Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12**, 351–366 (2009).

Chave, J. *et al.* Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.* **20**, 3177–3190 (2014).

Chen, J., Franklin, J. F. & Spies, T. A. Vegetation responses to edge environments in old-growth Douglas-fir forests. *Ecol. Appl.* **2**, 387–396 (1992).

Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).

Csillik, O., Cherbini, J., Johnson, R., Lyons, A. & Kelly, M. Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. *Drones* **2**, 1–16 (2018).

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.* **123**, 258–270 (2012).

Davies, S. J. Photosynthesis of nine pioneer Macaranga species from Borneo in relation to life history. *Ecology* **79**, 2292–2308 (1998).

Dickie, I. A. & Reich, P. B. Ectomycorrhizal fungal communities at forest edges. *J. Ecol.* **93**, 244–255 (2005).

dos Santos, A. A. *et al.* Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs. *Sensors* **19**, 1–11 (2019).

FAO. *Global forest assessment resources 2020 main report*. *FAO* https://www.fao.org/3/ca9825en/ca9825en.pdf (2020).

Fensham, R. J., Fairfax, R. J. & Cannell, R. J. The invasion of Lantana camara L. in forty mile scrub National Park, north Queensland. *Aust. J. Ecol.* **19**, 297–305 (1994).

Feret, J.-B. & Asner, G. P. Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Trans. Geosci. Remote Sens.* **51**, 73–84 (2013).

FSC. Ecosystem services procedure: impact demonstration and market tools. *FCS, Germany.* 1–66 https://ic.fsc.org/en/document-center/id/328 (2021).

Fujiki, S. *et al.* Large-scale mapping of tree-community composition as a surrogate of forest degradation in Bornean tropical rain forests. *Land* **5**, 1–16 (2016).

Gardner, T. A. *et al.* Prospects for tropical forest biodiversity in a human-modified world. *Ecol. Lett.* **12**, 561–582 (2009).

Gaveau, D. L. A. *et al.* Rapid conversions and avoided deforestation: Examining four decades of industrial plantation expansion in Borneo. *Sci. Rep.* **6**, 1–13 (2016).

Gelbard, J. L. & Belnap, J. Roads as conduits for exotic plant invasions in a semiarid landscape. *Conserv. Biol.* **17**, 420–432 (2003).

Giam, X. Global biodiversity loss from tropical deforestation. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5775–5777 (2017).

Goodbody, T. R. H., Coops, N. C., Marshall, P. L., Tompalski, P. & Crawford, P. Unmanned aerial systems for precision forest inventory purposes: A review and case study. *For. Chron.* **93**, 71–81 (2017).

Hansen, M. C. *et al.* Quantifying changes in the rates of forest clearing in Indonesia from 1990 to 2005 using remotely sensed data sets. *Environ. Res. Lett.* **4**, (2009).

He, K., Gkioxari, G., Dollar, P. & Girshick, R. Mask R-CNN. in *Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017* 2980–2988 (2017). doi:10.1109/ICCV.2017.322.

Honnay, O., Verheyen, K. & Hermy, M. Permeability of ancient forest edges for weedy plant species invasion. *For. Ecol. Manage.* **161**, 109–122 (2002).

Iizuka, K., Yonehara, T., Itoh, M. & Kosugi, Y. Estimating tree height and diameter at breast height (DBH) from digital surface models and orthophotos obtained with an unmanned aerial system for a Japanese Cypress (Chamaecyparis obtusa) Forest. *Remote Sens.* **10**, 1–14 (2018).

Imai, N. *et al.* Effects of selective logging on tree species diversity and composition of Bornean tropical rain forests at different spatial scales. *Plant Ecol.* **213**, 1413–1424 (2012).

Imai, N. *et al.* Tree community composition as an indicator in biodiversity monitoring of REDD+. *For. Ecol. Manage.* **313**, 169–179 (2014).

Kachamba, D. J., Ørka, H. O., Gobakken, T., Eid, T. & Mwase, W. Biomass estimation using 3D data from unmanned aerial vehicle imagery in a tropical woodland. *Remote Sens.* **8**, 1–18 (2016).

Kapos, V. Effects of isolation on the water status of forest patches in the Brazilian Amazon. *J. Trop. Ecol.* **5**, 173–185 (1989).

Kitayama, K. *et al.* Biodiversity observation for land and ecosystem health (BOLEH): A robust method to evaluate the management impacts on the bundle of carbon and biodiversity ecosystem services in tropical production forests. *Sustainability* **10**, 1–15 (2018).

Laurance, W. F. Edge effects in tropical forest fragments: application of a model for the design of nature reserves. *Biol. Conserv.* **57**, 205–219 (1991).

Lawton, J. H. *et al.* Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391**, 72–76 (1998).

Li, A., Luo, T., Lu, Z., Xiang, T. & Wang, L. Large-scale few-shot learning: Knowledge transfer with class hierarchy. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 7212–7220 (2019). doi:10.1109/CVPR.2019.00738.

Lugo, A. E. & Gucinski, H. Function, effects, and management of forest roads. *For. Ecol. Manage.* **133**, 249–262 (2000).

Margono, B. A. *et al.* Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010. *Environ. Res. Lett.* **7**, 1–16 (2012).

Marks, C. O., Muller-Landau, H. C. & Tilman, D. Tree diversity in relation to maximum tree height: Evidence for the harshness hypothesis of species diversity gradients. *Ecol. Lett.* **20**, 398–399 (2017).

Marks, C. O., Muller-Landau, H. C. & Tilman, D. Tree diversity, tree height and environmental harshness in eastern and western North America. *Ecol. Lett.* **19**, 743–751 (2016).

Mlambo, R., Woodhouse, I. H., Gerard, F. & Anderson, K. Structure from motion (SfM) photogrammetry with drone data: A low cost method for monitoring greenhouse gas emissions from forests in developing countries. *Forests* **8**, 1–20 (2017).

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, G. A. B. da & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).

Narasimhan, P., Starr, I., Hayward, J., Noponen, M. & Durbin, J. CCB Standards. *Climate, Community and Biodiversity Alliance and the Rainforest Alliance, USA.* 1–119 (2014).

Natesan, S., Armenakis, C. & Vepakomma, U. Resnet-based tree species classification using uav images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLII-2/W13**, 475–481 (2019).

Obeng-Manu, C. Assessing the accuracy of UAV- DTM generated under different forest canopy density and its effect on estimation of aboveground carbon in Asubima forest, Ghana. (2019).

Oksanen, J. *et al.* Vegan: community ecology package. R package version. 2.0-10. *CRAN* 1–291 (2013).

Onishi, M. & Ise, T. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Sci. Rep.* **11**, 1–15 (2021).

Paneque-Gálvez, J., McCall, M. K., Napoletano, B. M., Wich, S. A. & Koh, L. P. Small drones for community-based forest monitoring: An assessment of their feasibility and potential in tropical areas. *Forests* **5**, 1481–1507 (2014).

Panfil, S. N. & Harvey, C. A. REDD+ and biodiversity conservation: A review of the biodiversity goals, monitoring methods, and impacts of 80 REDD+ projects. *Conserv. Lett.* **9**, 143–150 (2015).

Paszke, A. *et al.* Automatic differentiation in PyTorch. *NIPS* 1–4 (2017).

Peña-Claros, M. Changes in forest structure and species composition during secondary forest succession in the Bolivian Amazon. *Biotropica* **35**, 450–461 (2003).

Prasad, A. E. Tree community change in a tropical dry forest: The role of roads and exotic plant invasion. *Environ. Conserv.* **36**, 201–207 (2009).

Safonova, A. *et al.* Detection of fir trees (Abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sens.* **11**, 1–19 (2019).

Schulze, C. H. *et al.* Biodiversity indicator groups of tropical land-use systems: Comparing plants, birds, and insects. *Ecol. Appl.* **14**, 1321–1333 (2004).

Shen, X. & Cao, L. Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. *Remote Sens.* **9**, 1–24 (2017).

Struebig, M. J. *et al.* Quantifying the biodiversity value of repeatedly logged rainforests. gradient and comparative approaches from Borneo. in *Advances in Ecological Research* vol. 48 183–224 (Elsevier Ltd., 2013).

Tan, M. & Le, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. in *Proceedings of the 36th International Conference on Machine Learning, PMLR* vol. 97 6105-6114International Conference on Machine Learn (2019).

Tang, L. & Shao, G. Drone remote sensing for forestry research and practices. *J. For. Res.* **26**, 791–797 (2015).

Uehara-Prado, M. *et al.* Selecting terrestrial arthropods as indicators of small-scale disturbance: A first approach in the Brazilian Atlantic Forest. *Biol. Conserv.* **142**, 1220–1228 (2009).

UNFCCC. *Report of the conference of the parties serving as the meeting of the parties to the Kyoto Protocol on its fourth session , held in Poznan from 1 to 12 December 2008*. *UNFCCC* (2009).

van der Werf, G. R. *et al.* CO2 emissions from forest loss. *Nat. Geosci.* **2**, 737–738 (2009).

Wang, W., Shen, J., Porikli, F. & Yang, R. Semi-Supervised video object segmentation with super-trajectories. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 985–998 (2018).

Wang, Y., Yao, Q., Kwok, J. T. & Ni, L. M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.* **53**, 1–34 (2020).

Williams-Linera, G. Vegetation structure and environmental conditions of forest edges in Panama. *J. Ecol.* **78**, 356–373 (1990).

CHAPTER 5
General Discussion

5.1 Summary of each chapter

To develop a system for tree identification from UAV imagery and examine its practicality, I proposed a new automatic tree species identification system using deep learning (Chapter 2). In Chapter 3, I examine the potential and robustness of the system. In Chapter 4, I used the system to identify indicator tree species and evaluate the potential for biodiversity monitoring in Borneo. The following is a summary of each chapter.

In Chapter 2, I constructed a machine vision system for tree identification and mapping using a red –green–blue (RGB) image obtained using an unmanned aerial vehicle (UAV) and a convolutional neural network (CNN). In this system, I first calculated the slope from the three-dimensional model obtained by the UAV, and segmented the UAV RGB photograph of the forest into several tree crown objects automatically using color and three-dimensional information and the slope model, and finally applied object-based CNN classification for each crown image. This system managed to classify seven tree classes, including several tree species with more than 90% accuracy. Guided gradient-weighted class activation mapping (Guided Grad-CAM) showed that the CNN classified trees according to their shapes and leaf contrasts, thereby enhancing the potential of the system for classifying individual trees with similar colors in a cost-effective manner, a useful feature for forest management.

In Chapter 3, I set objectives to evaluate the practicality and robustness of the tree identification system using UAVs and deep learning. I sampled training and test data from three sites respectively in temperate forests in Japan. The objective tree species ranged from 56 species, dead trees, and gaps. When I evaluated the model performance on the dataset obtained from the same time and same trees as the training dataset, it yielded a Kappa score of 0.97. When I

evaluated the performance on the dataset obtained from the same time but with different trees, the kappa score was 0.72. When I evaluated the dataset obtained from different times and sites from the training dataset, which is the same condition as the practical one, the Kappa scores decreased to 0.47. To improve the performance in practical use, I developed "inventory tuning," which enables us to create a local model by limiting output classes from inventory data. Using inventory tuning, I increased the accuracy level to 0.62. Based on the classification result details, the classes that showed the potential for being identified were mainly coniferous trees. Further, by analyzing the relationship between the number of training images and the accuracy, I found that I need to prepare at least 300 samples for practical use. In addition, some misclassifications occurred between ①tree species that belong to close phylogenetically, ②tree species that have similar leaf shapes, ③ tree species that prefer the same environment, and ④ tree types such as coniferous and broad-leaved or evergreen and deciduous do not always promise common features among the tree type. In this study, I illustrated stable robustness for certain tree species in practical use, the amount of data required for stable performance, and the similarities between the appearances of tree species that cause misclassification. These findings will promote the practicalization of identification systems using UAV RGB images and deep learning.

In Chapter 4, I aimed to evaluate biodiversity cost-effectively from air using unmanned aerial vehicles (UAVs) and deep learning. First, I applied digital images obtained using UAVs and deep learning to identify tree species (genera *Macaranga* and *Neolamarckia*) that indicate degradation with respect to tree community composition in Bornean tropical rain forests. Then, using this identification system, I created an indicator species crown map and analyzed the relationship between UAV-derived information (tree crown areas of indicator species identified from UAV and forest structure information such as maximum tree height) and overall tree-community composition based on all genera obtained by field survey at 28 20-m radius plots, using a regression analysis and a stepwise selection. I successfully classified the indicator tree species and another tree species class at a Kappa score ranging from 0.6 to 0.8. Moreover, the regression analysis revealed that the total tree crown area, except for the indicator tree species and one

indicator tree species crown area, had a significant relationship with tree community composition. My study indicated that the system using UAVs digital imagery and deep learning can identify specific indicator tree species from a wide variety of species in the tropical rain forest, and using the identified tree species crown areas, I can reliably evaluate biodiversity aerially. This system is a cost-effective tool for monitoring biodiversity.

5.2 Comprehensive discussion about tree species identification from the air

This dissertation demonstrated that the proposed new tree identification system using UAV and deep learning. This system exhibited superior performance compared to other machine learning methods, and showed some spatial and temporal robustness, indicating the feasibility of the reliable identification of certain tree species. Regarding its application for biodiversity monitoring in Borneo, specific indicator tree species were successfully identified for various tree species, and the information of its existence is helpful for monitoring biodiversity from the air.

UAVs has an advantage in cost and easy-to-use comparing to airplane, while UAVs can only cover limited areas. In terms of low-costand easy-to-use, UAVs can be used for community-based regular monitoring in the case of forestry in developed countries, and REDD+ project (Paneque-Gálvez et al. 2014). In the point of the covering areas, fixed-wing UAVs or gasoline-powered UAVs can fly longer than multi-copter UAVs, but it cannot be an absolute substitute for airplane.

I developed a chain of automatic tree crown segmentation and tree species classification system using some software and deep learning for the first time in the world in April 2018 (Onishi and Ise, 2018). In the past few years, some studies have applied deep learning to detect some trees or classify some tree species. However, they applied general object detection method of deep learning, which detect objective trees by bounding boxes, instead of tree crown segmentation (dos Santos et al., 2019; Safonova et al., 2019). This method takes cost for annotation of regions and cannot map the tree crown areas. In that aspect, my method has advantages, unfortunately, my method did not succeed to segment various tree crown perfectly. In remote sensing field, the most common tree crown segmentation method is the combination of local maximum filter and region-

growing algorithms or watershed segmentation algorithms (Chen et al., 2006; Jakubowski et al., 2013; Ke et al., 2011; Koch et al., 2006; Solberg et al., 2006; Wulder et al., 2000). This method can be applied to height information of LiDAR or digital image from UAV, especially in coniferous forest. (Picos et al., 2020; Mohan et al., 2017). In broad-leaved forest or mixed forest, tree crown segmentation is still difficult task because the size of trees has various range value and the tree top is unclear, hence, the development of segmentation method, which do not need to set parameters such as pixel sizes, has been desired (Dalponte et al., 2019; Maschler et al., 2018). On this issue, masking region method using deep learning such as instance segmentation (He et al., 2017) will be expected to be a clue for tree crown segmentation in broad-leaved or mixed forest.

I used one of the most basic deep learning classification methods for identification in my study, however, one of the disadvantages of this method is it needs much training data. In order to overcome this disadvantage, one-shot learning (Koch et al. 2015) and few-shot learning (Wang et al. 2020) has been developed in the computer vision field. Representative one-shot learning uses siamese network, which is a kind of metric learnings, and the class was distinguished by comparing the test data to all training samples (Koch et al. 2015). It is opposed to basic method which trains CNN to identify the data without comparing to samples. These latest classification methods will help us to identify tree species with small amount of training samples in practical uses.

5.3 Future perspective

In future, tree species identification systems using UAVs and deep learning can be a standard tool for forest management. This will change the manner of logging planning, monitoring biodiversity, and estimating biomass, CO2 absorption, and emissions. The survey professional experts are doing with much labor and cost now may be addressed by local people cost-effectively.

To realize such a future, we need to do as follows.

1) Develop an accurate tree crown segmentation method for more accurate identification and

counting of trees, and estimation of DBH and biomass. Recently, segmentation technology using deep learning has been developed in the computer vision field, and instance segmentation is expected to pave way for the solution. A large amount of image data and annotation data obtained through intensive labor will enable the automatic and accurate segmentation of tree crowns.

2) Develop a software that enables us to segment tree crowns, identify tree species automatically, and estimate the number of trees, DBH, and biomass. Currently, only a few professionals can identify tree species using deep learning. For wide use, we have to develop a software that can be used by non-professionals and without a high-power computer. The number of trees, DBH, and biomass can be estimated from the values of tree crown size, tree height, and tree species. By developing a system that can estimate such information easily, various local forest managers can use the system for forest management.
3) Gather each tree species image for the reliable identification of tree species. In Chapter 3, I had to prepare at least 300 samples. To obtain samples, I need to develop a system that can obtain data collected by various people. The aforementioned method is an effective way of doing so and gathering the image data collected from the user.

If forest information data estimated from UAVs by forest managers are accumulated, I can study forests in various places at super high resolution and each tree species level around the world. Furthermore, by combining satellite data: correction of satellite data or expansion of UAV data, I can monitor the forest conditions and transitions of biomass and biodiversity more accurately on a global scale. This will promote appropriate and sustainable use and protection of forests.

5.4 References

Chen, Q., Baldocchi, D., Gong, P. & Kelly, M. Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogramm. Eng. Remote Sensing* **72**, 923–932 (2006).

Dalponte, M., Frizzera, L. & Gianelle, D. Individual tree crown delineation and tree species classification with hyperspectral and LiDAR data. *PeerJ* **2019**, (2019).

dos Santos, A. A. *et al.* Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. *Sensors (Switzerland)* **19**, (2019).

He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 770–778 (2016) doi:10.1109/CVPR.2016.90.

Jakubowski, M. K., Li, W., Guo, Q. & Kelly, M. Delineating individual trees from lidar data: A comparison of vector- and raster-based segmentation approaches. *Remote Sens.* **5**, 4163–4186 (2013).

Ke, Y. & Quackenbush, L. J. A comparison of three methods for automatic tree crown detection and delineation from high spatial resolution imagery. *Int. J. Remote Sens.* (2011) doi:10.1080/01431161003762355.

Koch, B., Heyder, U. & Welnacker, H. Detection of individual tree crowns in airborne lidar data. *Photogramm. Eng. Remote Sensing* **72**, 357–363 (2006).

Koch, G., Zemel, R. & Salakhutdinov, R. Siamese Neural Networks for One-shot Image Recognition. *Proc. 32nd Int. Conf. Mach. Learn.* **37**, 1–8 (2015).

Maschler, J., Atzberger, C. & Immitzer, M. Individual tree crown segmentation and classification of 13 tree species using Airborne hyperspectral data. *Remote Sens.* **10**, (2018).

Mohan, M. *et al.* Individual tree detection from unmanned aerial vehicle (UAV) derived canopy height model in an open canopy mixed conifer forest. *Forests* **8**, 1–17 (2017).

Onishi, M. & Ise, T. Automatic classification of trees using a UAV onboard camera and deep learning. (2018).

Paneque-Gálvez, J., McCall, M. K., Napoletano, B. M., Wich, S. A. & Koh, L. P. Small drones for community-based forest monitoring: An assessment of their feasibility and potential in tropical areas. *Forests* **5**, 1481–1507 (2014).

Picos, J., Bastos, G., Míguez, D., Alonso, L. & Armesto, J. Individual tree detection in a eucalyptus plantation using unmanned aerial vehicle (UAV)-LiDAR. *Remote Sens.* **12**, (2020).

Safonova, A. *et al.* Detection of fir trees (Abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sens.* **11**, (2019).

Solberg, S., Naesset, E. & Bollandsas, O. M. Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogramm. Eng. Remote Sensing* **72**, 1369–1378 (2006).

Wang, Y., Yao, Q., Kwok, J. T. & Ni, L. M. Generalizing from a few examples: A survey on few-shot Learning. *ACM Comput. Surv.* **53**, 1–34 (2020).

Wulder, M., Niemann, K. O. & Goodenough, D. G. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sens. Environ.* **73**, 103–114 (2000).

Acknowledgements

This dissertation was written with the support of many cooperators as follows. I appreciate all of them.

First, I would like to thank my supervisor, Associate Professor Takeshi Ise, for providing me with many opportunities to challenge the latest and interesting themes, as shown in this dissertation.

Next, I would like to thank Professor Kanehiro Kitayama for providing me with a spectacular theme for biodiversity assessment from UAVs in Borneo and financial support for the research.

I also appreciate Associate Professor Hisashi Hasegawa for supporting the research study conducted by Wakayama, Dr. Shuntaro Watanabe for helping with several field studies, Torimaru Takeshi for accepting my study and helping with Daisen, and staff of Kamigamo Experimental Forest, and Ashiu and Wakayama Forest Research Stations.

For the research in Borneo, I am deeply grateful to Ryuichi Takeshige who took and provided me with several UAV photos of Borneo, and Dr. Ryota Aoyagi and Dr. Yoshimi Sawada who helped with the research of Borneo. I thank Nobuo Imai for providing inventory data to create supervised data, the staff of the FMUs for helping with the field inventory, Sabah Forestry Department and Sabah Forest Research Centre for their generous support, and Miun Poster and Joel Bin Dawat for the tree species identification.

I would like to thank Dr. Syogoro Fujiki for teaching me the segmentation method using eCognition software, Dr. Soyoka Makino for helping me create figures and organize the manuscript, Associate Professor Masae Ishihara for teaching me about analysis, May Thet Su Kyaw Tint for English proofreading, and members of AI Research Center of AIST for teaching me about deep learning.

Finally, I appreciate my family for being remarkably understanding. I am deeply grateful to everyone that contributed to my research study.

This research was sponsored by JSPS KAKENHI Grant Number JP19J22591, JP18H03357, and supported by the United Nations University GGS Project Fund to Kanehiro Kitayama, and the Re:connect Program of the Nippon Foundation–Kyoto University Joint Project.

This doctoral dissertation was composed based on the following papers.

1. Onishi, M & Ise, T. Automatic classification of trees using a UAV onboard camera and deep learning. arXiv preprint arXiv:1804.10390 (2018).
   https://arxiv.org/ftp/arxiv/papers/1804/1804.10390.pdf

2. Onishi, M. & Ise, T. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Sci. Rep.* **11**, 1–15 (2021).
   https://doi.org/10.1038/s41598-020-79653-9