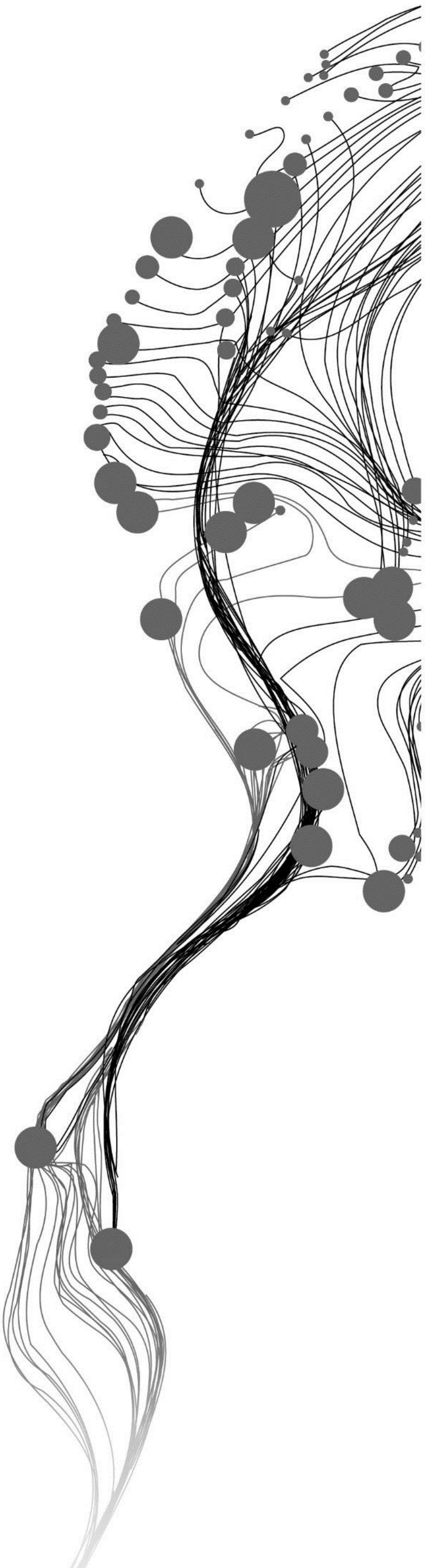
Mapping Dominant Tree Species from Remotely Sensed Image using Machine Learning Algorithms

DEVARA PRAWIRA ADININGRAT
June, 2017

SUPERVISORS:
Prof. Dr. Andrew K. Skidmore
Dr. Tiejun Wang

ADVISOR:
Yifang Shi



Mapping Dominant Tree Species from Remotely Sensed Image using Machine Learning Algorithms

DEVARA PRAWIRA ADININGRAT

Enschede, The Netherlands, June, 2017

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Environmental Modelling and Management

SUPERVISORS:
Prof. Dr. Andrew K. Skidmore (First Supervisor)
Dr. Tiejun Wang (Second Supervisor)

ADVISOR:
Yifang Shi

THESIS ASSESSMENT BOARD:
Prof. Dr. Andy Nelson (Chair)
Dr. Claudio Persello (External Examiner, ITC-EOS, University of Twente)

DISCLAIMER
This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

In forest management, dominant tree species information plays an important role in many applications ranging from biodiversity to the economic assessment of forest for managing logging. Remote sensing data can be used effectively and efficiently to provide information of tree species distribution. Moreover, recent technological developments allow the user to access remote sensing data freely, including the higher resolution (both spatial and spectral) from Sentinel-2 imagery. The use of machine learning algorithms for remote sensing image classification has increased in the last two decades since the accuracy results typically outperform traditional parametric classifiers. The most popular algorithms are Artificial Neural Network (ANN), Random Forest (RF) and Support Vector Machine (SVM). However, none of these algorithms has been shown to consistently outperform, particularly for tree species classification. The other issue is that the use of red-edge bands in different seasonal imageries is still limited for tree species classification, and since Sentinel-2 allows the three red-edge bands to be approximated, it is desirable to investigate this advantage.

This study aims to compare the performance of ANN, RF and SVM in tree species classification, particularly for the dominant species. The experiments were carried out using four images of different season (i.e., summer, autumn, winter and spring) using Sentinel-2 acquired over the Bavarian Forest National Park, Germany. The accuracy of ANN, RF and SVM in tree species classification were assessed through two main experiments: 1) assessing the accuracies of images with and without red-edge bands in a single season (summer) and 2) assessing the accuracies of different season images and multi-season image. The sensitivity of map accuracy to the algorithms' parameter settings was investigated to arrive at the optimum value settings for model parameters. The performance of ANN, RF and SVM were evaluated using four approaches, i.e., overall accuracy, statistical difference in accuracy, ease of use regarding the parameters optimization and computational time. Besides the performance of the classifiers, the quality of training samples was investigated through class separability, sensitivity analysis of training data size and class definition threshold.

The results show that the map accuracy when adding red-edge bands or using a multi-seasonal image is improved, yet it is comparable regardless of the algorithm used. Although in all experiments, SVM outperformed RF and ANN, the differences were not statistically significant. Furthermore, this study finds that the most efficient and the easiest algorithm to use is the RF algorithm as it needs the least number of parameters to be set, and it processes faster than ANN and SVM. The other finding in this study is that the threshold 75% of canopy coverage for species plot sample definition yielded an unsatisfactory result in class separability, and further yielded a classification accuracy which is considerably low.

This study concludes that the Sentinel-2 has the potential for monitoring dominant tree species in mixed temperate European forest. It would be interesting to further investigate how soft-classification/fuzzy approaches and ancillary data such as DEM and LiDAR may further increase the mapping accuracy of tree species using Sentinel-2.

ACKNOWLEDGEMENTS

It has been an excellent opportunity for 22 months of being a GEM student. Studying in two countries (Sweden and Netherlands) broadened and enriched my knowledge and life experience. At this time of finishing my thesis, I would like to express my gratitude to all people who helped and supported me during this course.

First and foremost, I would like to express my deepest gratitude to my first supervisor Prof. Dr. Andrew K. Skidmore for his patience, constructive critics, and advice in guiding the entire process of this M. Sc. thesis. Thank you for directing me to stay on the right path of my research. Also, thank you for sharing your experience, I learned a lot from it.

Great appreciation goes to my second supervisor, Dr. Tiejun Wang, for his assistance during this research. Having a discussion with him is always enjoyable and many fruitful things that I can gain.

Special thanks to Yifang Shi as my advisor. Thank you for your willingness to supporting and discussing my thesis.

Not to forget, thanks also go to Petter, Raymond, Laura and all GEM teachers and staffs. Thank you for facilitating all matters during the study.

To all my GEM's fellow, thank you for spending a wonderful time together. I will not forget and always cherish this friendship moment and hope we still connected no matter where we are. Also thanks to my friends in Lund and ITC which I cannot mention here one by one for your support and friendship.

Last but not least, a great appreciation for my family and friends in Indonesia. I believe I could not finish this entire M.Sc. process without your infinite support. I am going to miss you all!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

1.1.1. The importance of dominant tree species information

Climate change and the increase in demand for wood products may cause a harmful impact on the forest biodiversity and conservation management. Unfortunately, there is no good understanding on the historical and present drivers of the forest biodiversity (Schulze et al., 2016). Therefore, several methods have been proposed to tackle this issue, and one of them is classification of tree species (Franklin, 2001). Knowledge about tree species is central to describe forest ecosystems as a parameter for assessing forest biodiversity (Innes & Koch, 1998; Hill et al., 2010; Immitzer et al., 2012), particularly for dominant tree species information.

Information about the dominant tree species is essential to assist the management of plant community protection (Abdollahnejad et al., 2017). Through dominant tree species assessment, information regarding forest resilience and vulnerability such as drought, pathogens and climate change adaptation can be gathered regarding to prevent and mitigate the forest ecosystem from disturbances (Guyot et al., 2015; Périé & de Blois, 2016), the sustainability of forests, as well as the commercial value of forests (Franklin, 2001). Why does the dominant tree species have such a critical impact on the forest ecosystem? One may say that it is because of the relationship either between inter or intra-species within the forest. In the forest, all tree species live either individually or in association with each other, where some species may dominate over other species based on biotic and abiotic factors (Abdollahnejad et al., 2017). Thus, the domination of some species may influence the ecosystem functions since characteristics of each tree species characterize the ecosystem (Lohbeck et al., 2016).

1.1.2. Remote sensing of tree species

The ground-based survey was the traditional method for recording tree species distribution. However, for assessing large areas, ground-based survey is no longer effective in terms of cost and time since it is labour-intensive and only limited area can be covered due to limited accessibility. Therefore, remote sensing data are commonly in forest mapping (Immitzer et al., 2012). A remote sensing synoptic overview provides integrated and high-level detailed information at various levels of resolution, allowing the retrieval of information of forest characteristics, including tree species composition (Franklin, 2001; Immitzer et al., 2012; Barrett et al., 2016).

Remotely sensed data such as MODIS and Landsat can be used freely by the public and optimized for many remote sensing applications. However, regarding limited spectral and spatial resolution, this imagery is not sufficient to capture the ecological characteristics of trees and resulting coarser scale of monitoring (Wulder et al., 2009; Lisein et al., 2015). A solution of using (airborne) hyperspectral sensor seems promising since from spectral domain the sensor uses hundreds of continuous narrow bands allowing finer result in distinguishing spectral characteristics among tree species (Dalponte et al., 2012; Raczko & Zagajewski, 2017). Moreover, the airborne hyperspectral has a finer spatial resolution (up to 1m) which may be ideal for classifying the individual tree species (Shang & Chisholm, 2013; Ballanti et al., 2016; Vaglio Laurin et al., 2016). However, the hyperspectral system is costly and their availability is limited (Immitzer et al., 2012) .

Recently, a new generation of multispectral satellite imagery -Sentinel-2A & B- were launched and bringing a new dimension to the domain of freely accessible remote sensing data (ESA, 2015). Sentinel-2 has more bands with a wider range of spectral and finer resolution both spatial and temporal (13 bands and 10-20 m resolution, 10 days cycle, 3 red-edge bands) as compared to other free satellite data such as Landsat. These specifications open an opportunity to produce a more robust methodology for tree species mapping using freely accessible satellite imagery data, particularly for a group of species/stands (Immitzer et al., 2016).

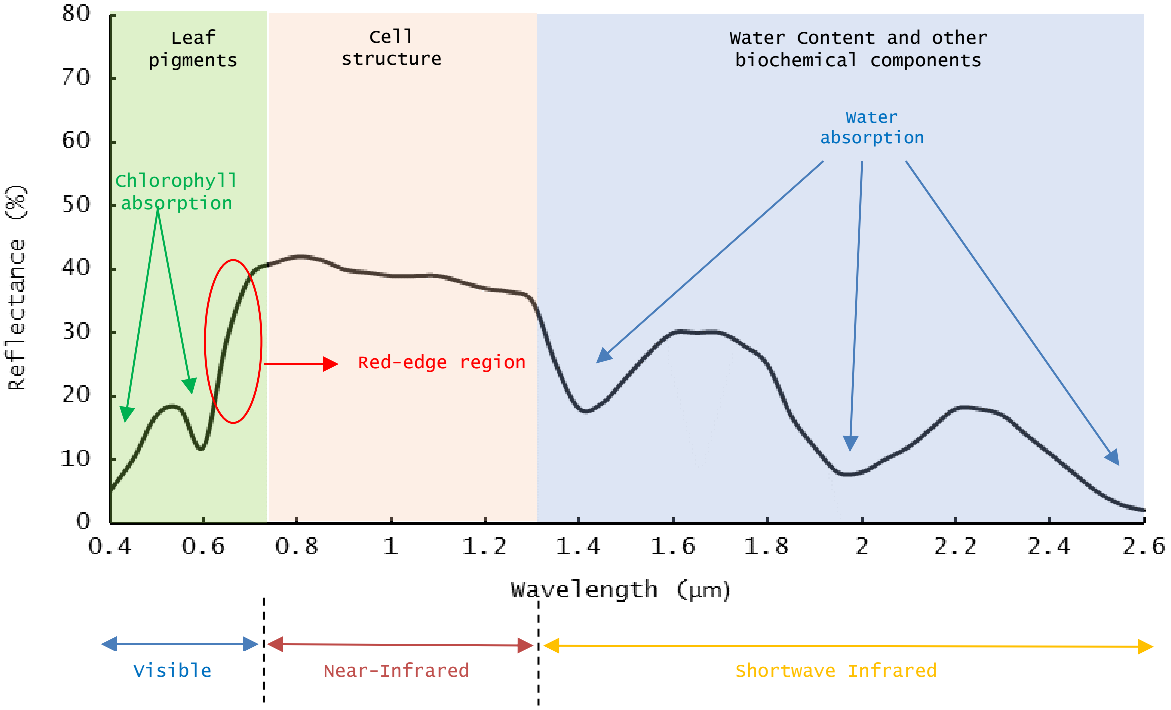


Figure 1.1. The spectral response characteristics of vegetation. The red-edge region lies on wavelength of 680-750 nm marked by red circle. (adapted and modified from Hoffer (1978)).

Fassnacht et al. (2016) reported that selection of the wavelength region is also important to obtain better discrimination among tree species. It was found that red and Near Infrared (NIR) wavelength are the most important regions where the intensive ratio between vegetation absorption and reflection occurred. Thus, these two regions are useful in detecting the leaf pigment which is the primary information for distinguishing tree species (Hoffer, 1978). Recently, many studies are trying to optimize the red-edge region for tree species classification and found it useful (Immitzer et al., 2012; Schuster et al., 2012). The red-edge denotes the maximum slope or transition region (Figure 1.1) of vegetation spectra (680nm -750nm) where the strong chlorophyll absorption in the red spectrum and the canopy reflectance in NIR occurs (Mutanga & Skidmore, 2007). This red-edge region has been successful in estimating vegetation information such as chlorophyll concentration, biomass and LAI. Since the chlorophyll concentration is related to the leaf pigment, several studies found the red-edge region to be useful in enriching spectral information to gaining higher accuracy in tree species classification (Immitzer et al., 2012; Adelabu et al., 2013; Li et al., 2015; Omer et al., 2015).

The temporal resolution of remote sensing imagery, particularly from space-borne/satellite platforms, allows the user to optimize different seasons to capture tree phenology throughout the year (Hill et al., 2010; Lisein et al., 2015). Considering the phenology, Hill et al. (2010) reported that combining time-series imagery in different growing seasons of the trees resulted in a robust species map. This report confirms the statement from Hoffer (1978) that considering temporal variations is necessary to understand the spectral characteristics of vegetation since the spectral is not static due to the vegetation biological temporal phenomenon or phenology. The term phenology is referred to leaf-on/leaf-off condition which covers the intense green colour of leaves as well as flower blossom events in spring, to the senescence events which may change the colour of the leaf under the autumn in deciduous forest (Fassnacht et al., 2016). Therefore, in spring and autumn, tree's leaves may have various spectral reflectance according to the alteration of biochemical and biophysical attributes (Hill et al., 2010). Thus, by combining different seasons/multitemporal imagery, the user can capture tree phenology to enrich the spectral variability between the tree species to gain higher accuracy in tree species classification (Mickelson et al., 1998; Hill et al., 2010; Lisein et al., 2015). However, it was suggested to have knowledge about species' phenological cycle in advance. Further, this knowledge can be used for determining the suitable season represented by a certain time of image acquisition to address the specific phenology of tree species (Fassnacht et al., 2016).

1.1.3. Overview of machine learning algorithms in tree species classification

Recently, advances in technology allowed the machine to adopt intelligent systems to process and solve problems. Such intelligent systems use learning process and trains its system through experience to solve the problem from the input data. Learning process imitates some facets of the human mind to solve highly complicated problems (Chaudhary et al., 2013). Thus, machine learning is about enabling the computers to modify and adapt their actions, so the actions may improve the efficiency and accuracy of the decisions which are drawn by the computer program (Chaudhary et al., 2013; Marsland, 2015). For the application, machine learning is widely used in classification, where data is separated into different segments to label the data object (Chaudhary et al., 2013). This classification can be applied in tabular/spreadsheet data to image recognition which is closely related to the most widely used remote sensing application (i.e., image classification).

In the domain of remote sensing, machine learning algorithms have become more popular compared to the traditional classifiers such as Maximum Likelihood Classifier. Machine learning algorithms mostly yield higher accuracy in image classification compared to the traditional parametric classifiers (Huang et al., 2002; Schuster et al., 2012; Shang & Chisholm, 2013; Khatami et al., 2016; Laurin et al., 2016). The higher accuracy obtained by machine learning algorithms is based on non-parametric data distribution (no prior probabilities). Therefore, this type of classifier does not depend on the performance of the pre-defined model and can handle complex data with respect to the nonlinear input data and features (Belgiu & Drăguţ, 2016; Ghosh et al., 2014). Another advantage of the machine learning algorithms is that several algorithms have capability to gain optimum classification accuracy by using small numbers of training/observation set and a high number of input features (Cracknell & Reading, 2014; Mountrakis et al., 2011). Fassnacht et al. (2016) and Khatami et al. (2016) reported from many types of machine learning algorithms, there are three most popular algorithms that are commonly used in remote sensing applications, i.e., Artificial Neural Network (ANN), Random Forest (RF) and Support Vector Machine (SVM). Generally, SVM outperforms the accuracy of ANN and RF classification in many applications (Kotsiantis, 2007; Khatami et al., 2016), but RF is used to be more efficient in the parameter handling and classification speediness (Kotsiantis, 2007). Since each algorithm has its own advantages, this study will apply and compare the performance of these three popular machine learning algorithms, specifically for mapping dominant tree species. In following paragraphs the characteristics of ANN, RF and SVM will be explained.

Artificial Neural Network (ANN)

The basic principal of ANN is that the algorithm imitates a biological brain through two tasks, i.e.: (i) the network gains knowledge from certain environment through a learning process and (ii) the knowledge is stored in synaptic weights or interneuron connection strengths (Mas & Flores, 2008). As well as the other machine learning algorithms, ANN is also based on non-parametric data distribution. Although there are several types of ANN, the most common type of ANN in remote sensing application is Multilayer

Perceptron (MLP). The applications of MLP can be found in many studies of remote sensing application (Cracknell & Reading, 2014; Omer et al., 2015; Li et al., 2016; Raczko & Zagajewski, 2017). MLP is a feed-forward network (Figure 1.2) where a single hidden layer of nodes exists (Cracknell & Reading, 2014) and contains several parameters i.e. size (number of input nodes and hidden layers); initial weight range; activation functions; learning rate and momentum; and stopping criterion (Mas & Flores, 2008).

The ANN-MLP works based on the backpropagation algorithm, where the error is propagated back to input data layer as the one iteration is completed (Ardö et al., 1997; Skidmore et al., 1997). In the neural network structure, the backpropagation encompasses two iterating phases, i.e., forward and backward phase. Forward phase is the first phase where the output nodes calculated from the input data. In the second phase, the calculated output node values are compared with the desired known targets value. The difference between the value of output node and the target node is treated as the error which is used to modify the weights in the previous layer. This process represents one epoch in backpropagation algorithm. The total error between the calculated value and the target value for each node is calculated as the Root-Mean-Square-error (RMSE) in the system. The algorithm's model is running until the total error decreased to a predetermined level, or the rate of decrease becomes asymptotic (Skidmore et al., 1997).

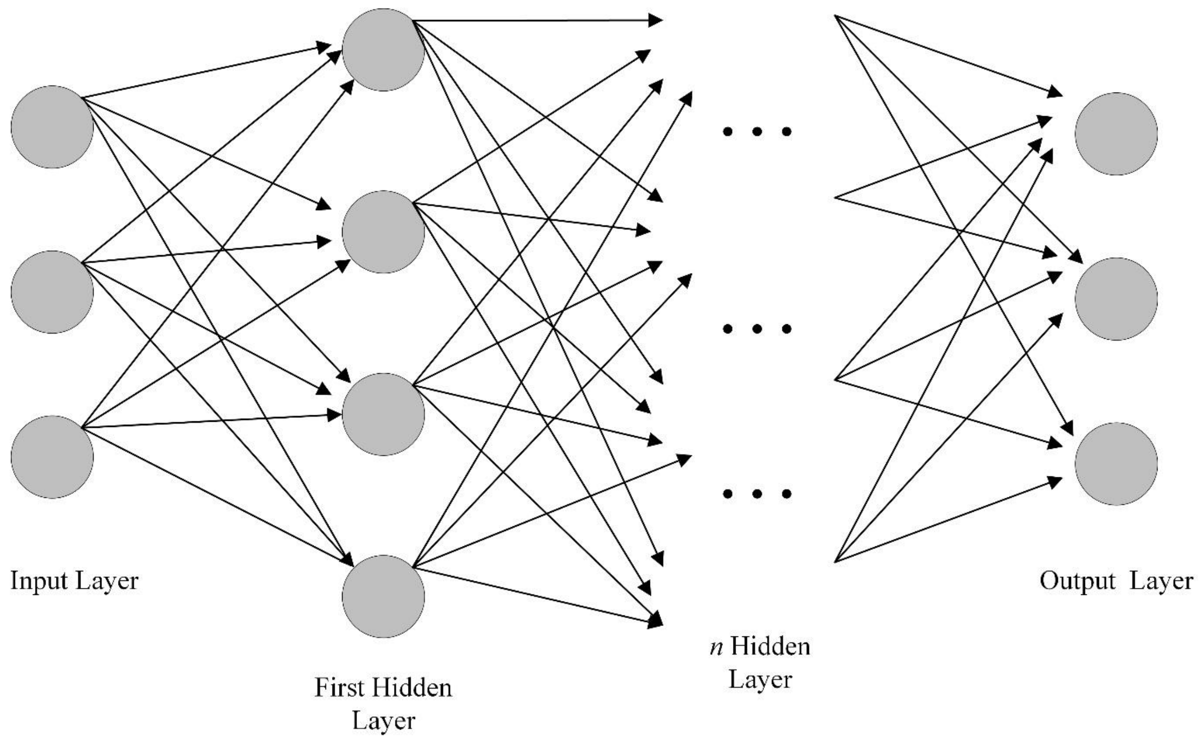


Figure 1.2. The architecture of ANN-Multilayer Perceptron (MLP) where forward propagation flows from left (input layer) to right (hidden layers and output layer). (adapted and modified from Mas & Flores, 2008)

Random Forest (RF)

The Random Forest (RF) algorithm is a combination of several 'branch' predictors where each tree depends on the values of the random vector which are independently sampled, and each tree has the same distributions within the forest (Breiman, 2001). RF is an advanced method of the decision tree algorithm which is designed as a hierarchical structure (binary structure) of certain classes (Breiman et al., 1984). Each class consisted of a root node and a number of interior nodes as non-terminal nodes and linked to the number of terminal nodes (decision stages) as a final classification (Swain & Hauska, 1977). RF works by assembling trees to draw a subset of training samples through replacement (a bagging approach) (Breiman, 1996). Only two-thirds of samples is used for training the trees while one-third are used for internal cross-validation. All samples were randomly chosen. After the samples are divided, the decision trees can be independently produced without any pruning and then the nodes are split by the user-defined number of features (Figure 1.3). Further, each decision trees will grow together formed as a forest depending on the user-defined number of trees (Belgiu & Drăguţ, 2016). The classification decision is taken by averaging the class assignment probabilities which is calculated by all produced trees to give a single vote to the most frequent class. In other words, a class with maximum votes is the final selected class (Rodriguez-Galiano et al., 2012a; Belgiu & Drăguţ, 2016).

Based on this process, RF has two important parameters that need to be set, i.e., number of features to be selected and tested for the best split when generating the tree (Mtry) and number of trees to be generated (Ntree), both are user-defined. The advantages of RF are that it is relatively insensitive to overfitting; it can generate an internal unbiased estimation of error; it gives estimates of important variables in the classification; and it can deal with thousands of input variables without any deletion (Rodriguez-Galiano et al., 2012a; Belgiu & Drăguţ, 2016).

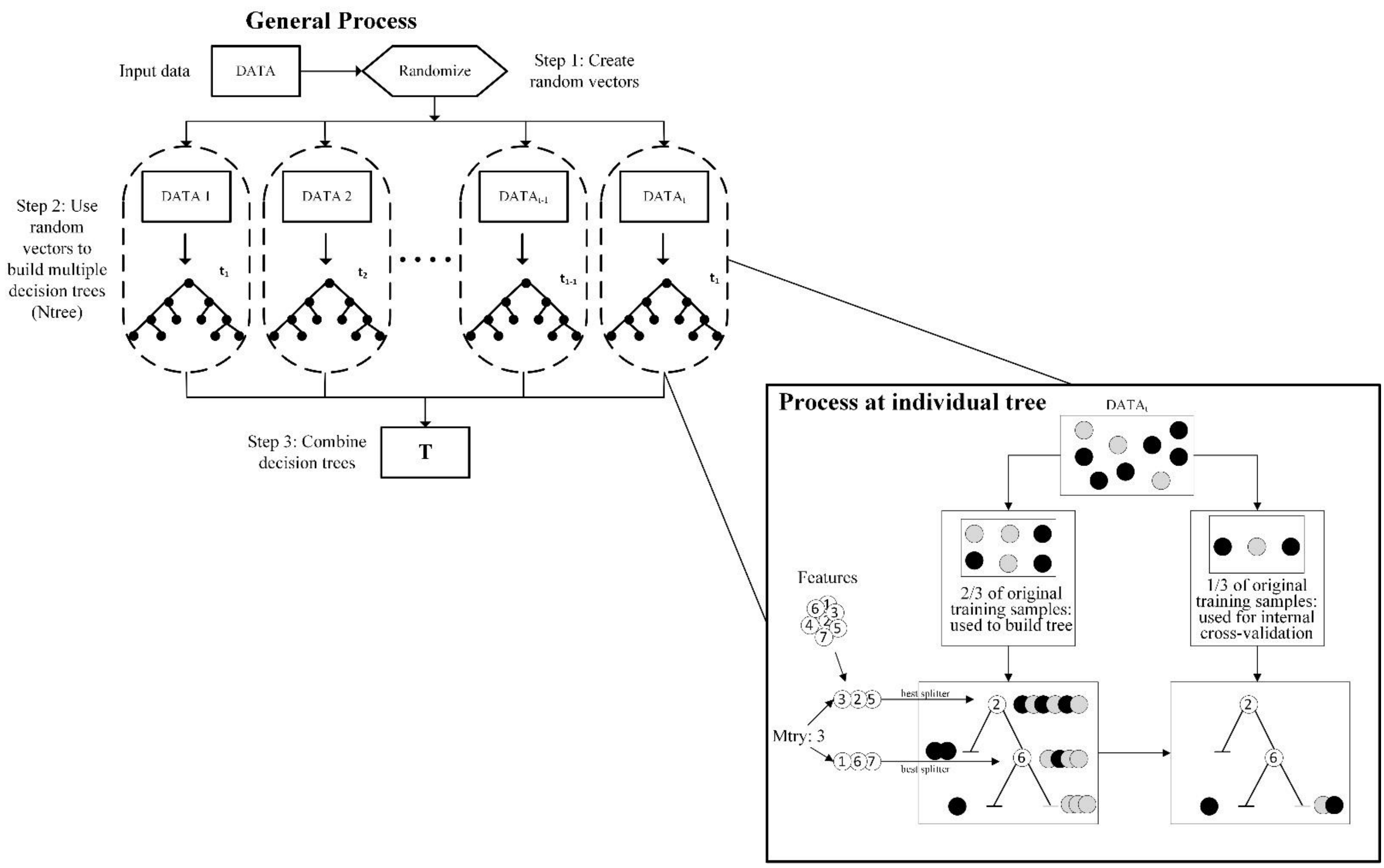


Figure 1.3. Random forest illustration with an example of using 7 features and Mtry-3 (adapted and modified from Thampi et al., 2013 and Kovanović et al., 2016).

Support Vector Machine (SVM)

First suggested by Vapnik in 1979, SVM became widely used as a classifier for remote sensing image classification in past two decades (Mountrakis et al., 2011). The approach of SVM is to find the optimum separation of the hyperplane between the classes by using the support vectors (training samples that lie on the edge of class) and set aside other training samples (Figure 1.4). Through this approach, the optimal hyperplane will be fitted. SVM also allows small training datasets to gain higher accuracy, which is a very advantageous feature for remote sensing application (Foody & Mathur, 2004). As a matter of fact, SVM was initially designed for solving binary (two-class) problems. Therefore, when dealing with the multiple classes, which often occurs in remote sensing domain, an appropriate multi-class method was required (Pal & Mather, 2005). The multi-class method divides into two methods, i.e., One Against One (OAO) and One Against All (OAA). In OAO, several classifiers are combined and all possible combinations are evaluated from the training set of n classes. Meanwhile, in OAA method, a total of n SVMs are developed for each of the *n* classes, then each SVM trained to classify one class against all other classes (Tso & Mather, 2009). Based on Pal & Mather (2005), the OAO method gives higher accuracy as compared to OAA. Hence, OAO is the suitable SVM's method in image classification domain when dealing with small training sets against high dimension data.

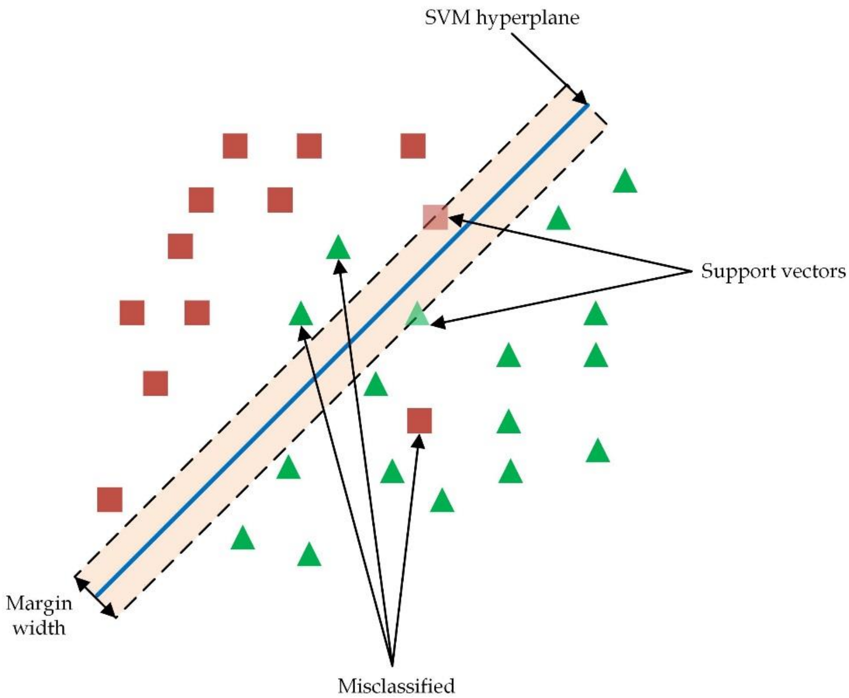


Figure 1.4. Support vector machine illustration (adapted from Burges, 1998 and Mountrakis et al., 2011)

Kavzoglu & Colkesen (2009) stated that SVM classification could also be divided into two approaches, i.e., linear and non-linear SVMs. Non-linear SVM is usually suitable for classification of remotely sensed image since the data are not linearly separable when the pixel is used as a sample. In this case, data sets cannot be classified into two classes with a linear function in the input space. The non-linear approach can be applied in remotely sensed image classification by using a kernel function, where the kernel function enables the data points to be spread along the hyperplane in such a way that a linear hyperplane can be fitted (Kavzoglu & Colkesen, 2009). There are four types of kernel functions that are used in SVM, i.e, linear, polynomial, Radial Basis Function (RBF) and sigmoid. However, only polynomial and RBF are commonly used in remote sensing image classification, and regarding the accuracy, RBF usually outperformed polynomial (Huang et al., 2002; Kavzoglu & Colkesen, 2009).

1.1.4. Sentinel-2

Sentinel-2 is a new generation of multispectral satellite imagery which was launched in June 2015 under the European Copernicus program. The launch mission brings a new dimension to the freely accessible data of remotely sensed imagery by providing an image with relatively higher resolution in both spectral, spatial and temporal (ESA, 2015). The satellite carries sensor of Multispectral Imager (MSI) with a capability to acquire data in 13 bands in various spatial resolution (4 bands 10 m; 6 bands 20 m; and 3 bands 60 m) and higher temporal resolution (10 days and 5 days with the twin satellites (Sentinel-2B) orbiting in 2017) with swath width of 290 km (Immitzer et al., 2016). The mission of Sentinel-2 is to deliver image in accordance with the land resources monitoring, emergency management, security and climate change (ESA, 2015).

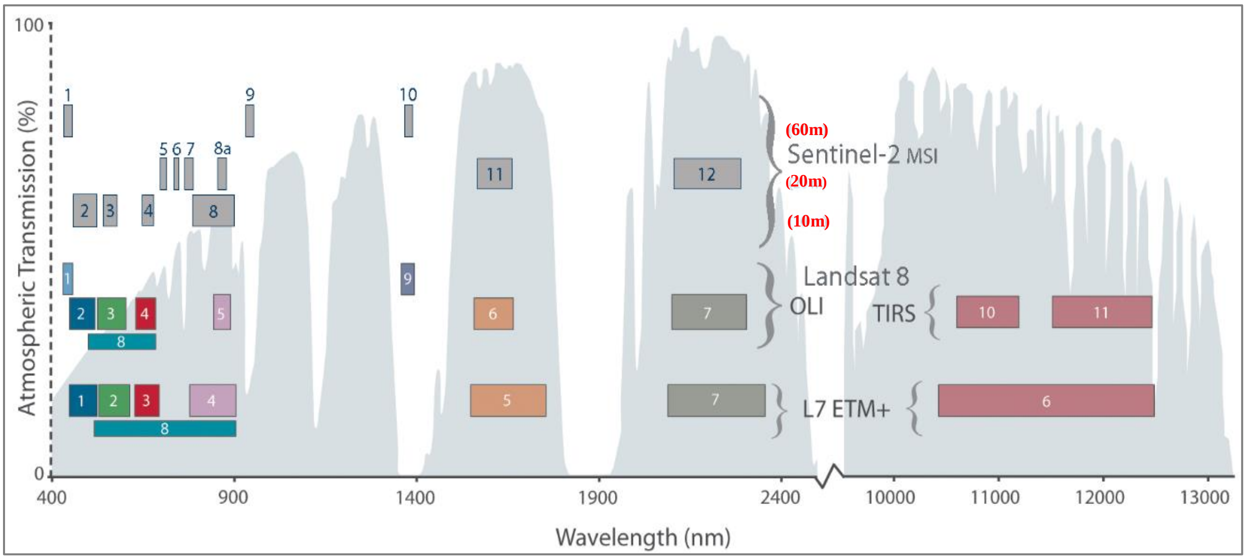


Figure 1.5. The spectral bands comparison between Sentinel-2, Landsat 8 and Landsat 7. Source: https://landsat.gsfc.nasa.gov/sentinel-2a-launches-our-compliments-our-complements/

Compared to the other free satellites imagery which has the closest specification such as Landsat 8 OLI or Landsat 7 ETM+, Sentinel-2 outperforms the Landsat specifications, particularly for land applications. In term of spectral resolution, the three red-edge bands (B5, B6, B7) of Sentinel-2 may give more advantages to monitoring and capturing vegetation properties such as biomass, chlorophyll concentration and LAI (Mutanga & Skidmore, 2007) as compared to Landsat which only has a single red band. The spectral configuration of Sentinel-2 is designed to match Landsat specifications to ease the integration of these satellite imageries (Mandanici & Bitelli, 2016). Regarding the spatial and temporal resolution, Sentinel-2 has again outperformed Landsat, with 10-20m resolution to Landsat 30m resolution and 5-10 days cycle to Landsat 16 days cycle. Again, Mandanici & Bitelli (2016) noted that the spatial and temporal resolutions of Sentinel-2 and Landsat permit easy integration. This study investigates the capability of advanced specifications of Sentinel-2 over Landsat for tree species classification.

1.2. Problem statement

From several prior studies of tree species classification, the accuracy yielded by SVM, RF and ANN algorithms showed that none of these algorithms consistently outperformed the others. Some results showed that SVM obtained higher accuracy than RF (Dalponte et al., 2012; Adelabu et al., 2013; Ballanti et al., 2016), while others reported that RF outperformed SVM (Pal, 2005; Sesnie et al., 2010; Li et al., 2016). Although there is a lack of study using ANN in tree species classification (Raczko & Zagajewski, 2017), in general, several studies indicate that ANN obtains the lowest accuracy as compared to RF and SVM

(Attarchi & Gloaguen, 2014; Omer et al., 2015). Therefore, further evaluation is needed to evaluate the performance of ANN, RF and SVM collectively to obtain a comprehensive reference of the performance of the algorithms in remote sensing applications, particularly for dominant tree species classification. Moreover, the studies that collectively compare the ANN, SVM and RF algorithms are still few and mostly were applied for land use/land cover classification or crop classification instead of tree species classification.

Previous studies have proven that the red-edge channel is useful for improving tree species classification since it is sensitive to the chlorophyll content (Immitzer et al., 2012; Schuster et al., 2012; Adelabu et al., 2013). Furthermore, using or adding different season images is also improved the tree species classification (Mickelson et al., 1998; Hill et al., 2010) since the tree has its phenological event which is related to the dynamic of chlorophyll content and leaf pigment (Horler et al., 1983). However, based on our knowledge, there are still limited studies that optimize the combination of red-edge channel and different season image for tree species classification. Some studies optimized the red-edge channel only in a single season image, particularly summer with fully developed leaves (leaf-on) since that is a decent condition for classification (Immitzer et al., 2012; Waser et al., 2014). On the other hand, Hill et al. (2010) reported that adding leaf-off and partially leaf-on/off images without red-edge channel improve the classification accuracy. Therefore, investigating the combination of red-edge bands in different season images may obtain a comprehensive description of Sentinel-2 capability to improve the accuracy of tree species classification through multitemporal approach.

1.3. Research objectives

The aim of the study is to evaluate the performance of three different machine learning algorithms, i.e., Artificial Neural Network (ANN), Random Forest (RF) and Support Vector Machine (SVM) in mapping dominant tree species using multi-season Sentinel-2 multispectral imagery. The specific objectives are as follow:

1. To compare the accuracies of the ANN, RF and SVM algorithms in mapping dominant tree species using a single month in the summer period of Sentinel-2 imagery without using red-edge bands (i.e., using B2-B4; B8; B11-B12 bands only).

2. To compare the accuracies of the ANN, RF and SVM algorithms for mapping dominant tree species using a single month in the summer period of Sentinel-2 imagery with the inputs of red-edge bands (i.e., using B2-B8; B8A; B11-B12 bands).

3. To compare the accuracies of the ANN, RF and SVM algorithms for mapping dominant tree species using all bands (i.e., using B2-B8; B8A; B11-B12 bands) of multi-season Sentinel-2 imagery (i.e., winter, spring, summer and autumn).

4. To compare the overall accuracies of the ANN, RF and SVM algorithms for mapping dominant tree species using Sentinel-2 imagery in all experiments (red-edge and multi-season).

5. To evaluate the computational efficiency (i.e., CPU time usage and parameter settings) of the three machine learning algorithms (i.e., ANN, RF and SVM) for mapping dominant tree species using Sentinel-2 imagery.

1.4. Research questions

1. Does adding red-edge bands of the Sentinel-2 imagery significantly improve the dominant tree species mapping accuracy compared to the accuracy derived from the Sentinel-2 image without red-edge bands?

2. Does the use of multi-season Sentinel-2 imagery significantly improve the dominant tree species mapping accuracy compared to the accuracy derived from single season image?

3. Are there significant differences in dominant tree species mapping accuracies between the ANN, RF and SVM algorithms using Sentinel-2 imagery?

4. Which classification algorithm (i.e., ANN, RF and SVM) is computationally more efficient in mapping dominant tree species?

1.5. Research hypotheses

Hypothesis 1
H_0: There are no significant differences in dominant tree species mapping accuracies with or without the inputs of the red-edge bands of Sentinel-2 imagery.
H_1: Adding the red-edge bands of the Sentinel-2 imagery can significantly improve the dominant tree species mapping accuracy.

Hypothesis 2
H_0: There are no significant differences in dominant tree species mapping accuracies between the use of the one single season image (summer) and the multi-season Sentinel-2 imagery.
H_1: Adding the multi-season Sentinel-2 imageries can significantly improve the dominant tree species accuracy rather than the use of any single season Sentinel-2 imagery.

Hypothesis 3
H_0: There are no significant differences in dominant tree species mapping accuracies between the ANN, RF and SVM algorithms using Sentinel-2 imagery.
H_1: The mapping accuracy derived from the SVM algorithm is significantly higher than the ANN and RF algorithms.

Hypothesis 4
H_0: There are no differences in computational efficiency for mapping dominant tree species between the ANN, RF and SVM algorithms.
H_1: The RF algorithm is computationally more efficient than the ANN and SVM algorithms for mapping dominant tree species.

2. MATERIALS AND METHODS

2.1. Study Area

The study was undertaken in the Bavarian Forest National Park (BFNP), located in south-eastern Germany (48°54' N, 13°29' E) within the two rural districts of Regen and Freyung-Grafenau bordering The Sumava National Park in the Czech Republic (Figure 2.1). The BFNP area covers 242 km² and is characterized by montane areas in altitudes between 300 to 1,450 m. The area belongs to the temperate climate zone which is influenced by Atlantic and continental characteristics. The annual precipitation is between 900 and 1800 mm, and the mean annual temperature varies from high elevations (3.5°C) to low elevations (9.0°C). Geologically, it is relatively homogenous with granitic and gneissic bedrock. The soil is acidic with sandy and loamy soils dominating (Bässler et. al., 2015).

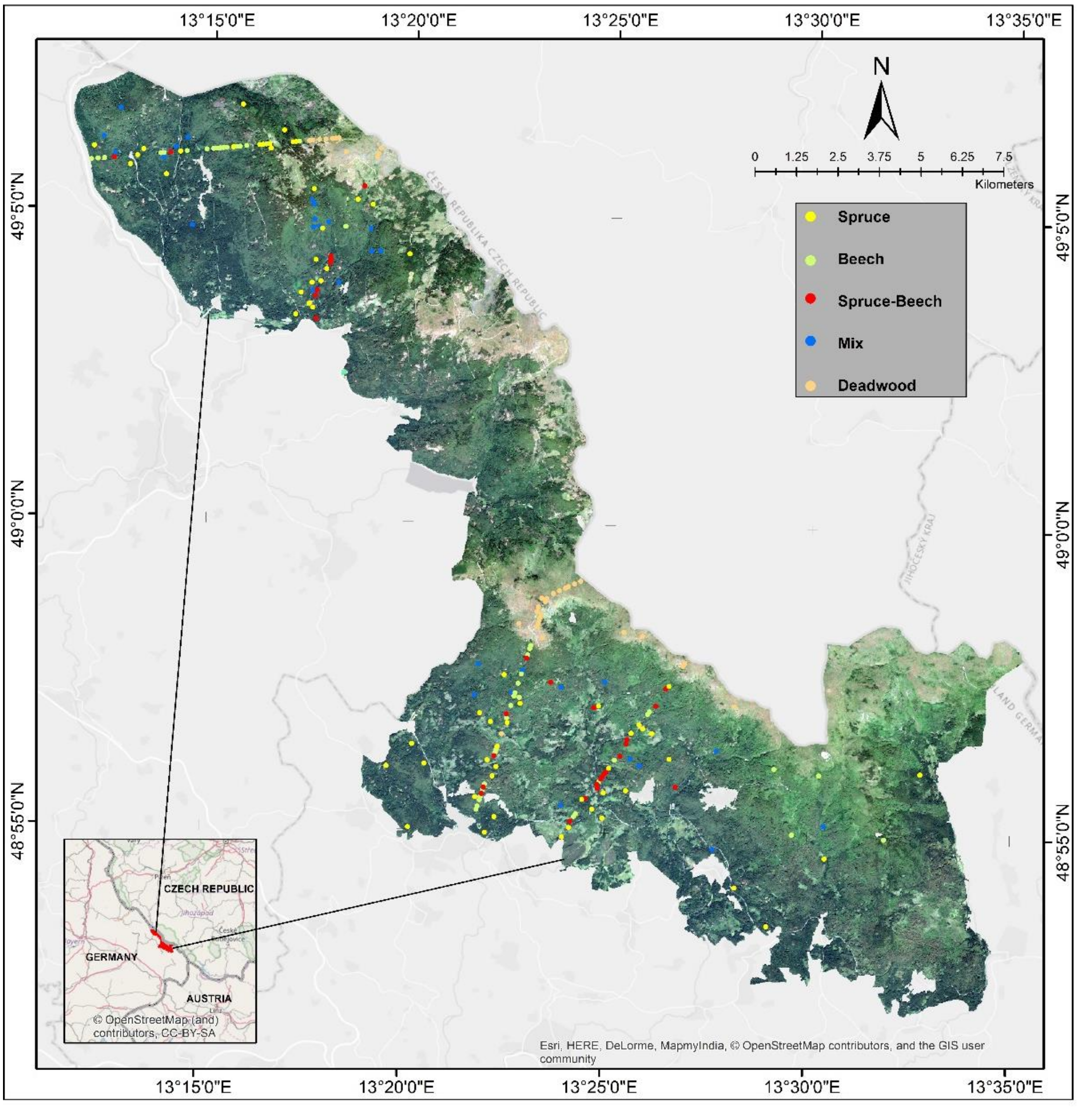


Figure 2.1. The location map of Bavarian Forest National Park with the distribution of sample plots. The presented image is Sentinel-2 with true colour composite (band 432) acquired at August 27, 2016.

The composition and the distribution of tree species in BFNP had changed along with the history of the forest management within BFNP. In the beginning of 16^{th} century, Norway spruce (*Picea abies*) was the most common tree species followed by beech *(Fagus sylvatica*) and silver fir (*Abies Alba*). Previously, the composition was dominated by beech-fir forest type. As time went by, the spruce became more dominant since the forest management was handed over to the Kingdom of Bavaria at 19^{th}-20^{th} century and during World War I and II, it formed highest proportion. In this time, the beech was relatively stable while the fir was decreased (Heurich & Englmaier, 2010). When the National Park was established in 1970, the proportion of spruce had increased to 72% while beech was still stable at 25% and fir decreased to only 3.2%. Then, 20-30 years after establishment, the composition of tree species within BFNP changed a little where the spruce decreased due to the bark beetle infestation, fir increased and beech remains stable (Heurich & Englmaier, 2010). In present days, the spruce is still dominating the species in BFNP forming around 67% of the forest, as shown in Table 2.1 which taken from Sommer et al., (2015).

Sommer et al., (2015) divided the forest types in BFNP into three characteristics with regards to the elevation as follows:

1. Region at elevation > 1,100 m (16% of the BFNP area) is dominated by Norway spruce.
2. Region at elevation between 600-1,100 m (68% of the area) consists of montane mixed forests with Norway spruce, silver fir, European beech and sycamore maple.
3. Region at elevation < 600 m, at the bottom of the valley (16% of the area), consists of Norway spruce, mountain ash and birch .

Table 2.1. Relative percentage of tree species composition within BFNP based from forest inventory in 2002/2003 (taken from Sommer et al., 2015)

Scientific Name	Common Name	Relative Amount (%)
Picea abies	Norway spruce	67
Fagus sylvatica	European beech	24.5
Sorbus aucuparia	Common rowan	3.1
Abies alba	Silver fir	2.6
Acer pseudoplantanus	Sycamore maple	1.2
Betula pendula	European white birch	0.7
Pseudotsuga menziesii	Douglas fir	0.2
Larix decidua	European larch	0.1
Pinus sylvestris	Scots pine	0.1
Alnus glutinosa	European alder	0.1
Populus tremula	European aspen	0.1
Fraxinus excelsior	European ash	0.1

2.2. Data

2.2.1. Tree species plot

This study used an existing sample of tree species plots which was designed in 2006 under the BIOKLIM project, with fixed coordinate position to allow annual biodiversity monitoring (Bässler et al., 2009). The BIOKLIM sampling design uses four transect plots and represents the different ranges of altitude environment within the BFNP. The sample consists of 293 plots, each measuring 30 m × 30 m (Figure 2.1). Twenty-three replications of measurement within the altitudinal range are needed to overcome the environmental effect. Therefore, to reach the number of replications, the distance between the plots is 100 m (Bässler et al., 2009). 13 more plots from Wang et al. (2017) were added to add more information about the species distribution. These 13 plots have same dimensions as the BIOKLIM's plots and were taken during fieldwork from mid-July to mid-August 2013.

A threshold 75% of canopy coverage was used to define a species as dominant within a plot. The 75% threshold is based on the FAO National Forest Inventory Guideline (2004). To assign plot into certain class in the classification scheme, several conditions were used to define the class boundary (Table 2.2). Refer to the BFNP forest inventory 2002/2003 (Table 2.1), there are two dominant species (i.e., spruce and beech) within BFNP. The other species were categorized as less dominant species in this study. If a plot is covered whether by spruce or beech canopy for ≥ 75%, then it will be assigned as a spruce or beech class. If the plot consists of less dominant species outside spruce and beech for ≥ 40%, then it will be assigned to a mixed class. This 40% canopy coverage is assumed to be the boundary between the vegetated area and the deadwood area (non-vegetated) since based on the BFNP land use map and BIOKLIM database, mostly the plots which are lied on the deadwood area have < 40% of canopy coverage. A class called spruce-beech was developed since several plots consist of spruce and beech with canopy cover of ≥ 75%. The deadwood class was considered to be added to classification scheme as it contains information about the evidence of bark beetle infestation (Heurich & Englmaier, 2010) and that this class can also give a picture of the distribution of BFNP area which previously was vegetated. The deadwood class is defined by canopy coverage of < 40% for all species both dominant and less dominant.

Table 2.2. The class boundary condition for defining and assigning species into a class in the classification system. All the conditions are referred to the FAO National Forest Inventory Guideline (2004) where the threshold of canopy coverage within a plot is 75%. Since the BFNP dominated by spruce and beech, other species was assigned as mixed species, besides the number of sample plots is too few. The class configuration setting also assisted by land use map of BFNP which derived from aerial photograph (AP) interpretation.

No	Class Name	Class Boundary Conditions
1	Spruce	the canopy coverage of spruce is ≥ 75% and other species are < 75%.
2	Beech	the canopy coverage of beech is ≥ 75% and other species are < 75%.
3	Spruce-Beech	the canopy coverage of spruce and beech are ≥ 75% and other species are < 75%.
4	Mix Species	the canopy coverage of: 1. less dominant species is ≥ 40%, and spruce and or beech do not exist. 2. spruce and or beech is ranging from ≥ 40% to <75%, with or without less dominant species existence. 3. all species both dominant (spruce and beech) and less dominant are ≥ 75%.
5	Deadwood	the canopy coverage of all species both dominant (spruce and beech) and less dominant are < 40%. **Note,** if the plot is not lie on the deadwood area based on AP Land-Use map, then it considered to assign as mix class.

The canopy cover percentage information is available on the BIOKLIM plots database. It also should be noted that the analyzed object within the plot is a group of stands/species instead of an individual stand. The defining and setting of the species configuration plots were assisted by a very high-resolution image (0.5 m) interpretation from an aerial photograph (AP) which was taken on August 19, 2012, and existing land cover/land use map of BFNP which was also derived from the AP for references.

To develop a classification scheme that consists of classes which spectrally have a minimum correlation, a class separability test was proposed using Transformed Divergence (TD) to find the best classification scheme. This process is achieved through trial and error process by selecting and assigning several pixels as a region of interest (ROI) within the plots. As Jensen (2005) proposed, the index of TD should has value at least 1.7 to imply that the classes have good separation/minimum spectral correlation. From the developed classification scheme, 2,572 original pixels were created for classification process. Further, the original pixels were split randomly into training set (two-thirds of original pixels) and test set (one-third of original pixels), refer to the suggestion from Ghosh et al. (2014) and Raczko & Zagajewski (2017). Thus, the training set has 1,712 pixels, meanwhile test set has 860 pixels. The sizes of training and test sets were varied for each class due to a different number of samples per class. The detail of the classification scheme and the number of training and test sets can be seen in Table 2.3.

Table 2.3. The classification scheme with number of training and test pixels for each class.

ID	Class Name	Number of Pixels	
		Training	Test
1	Spruce	529	261
2	Beech	386	193
3	Spruce-Beech	252	168
4	Mix	252	110
5	Deadwood	293	128

2.2.2. Sentinel-2 Pre-Processing

In this study, Sentinel-2 imagery were acquired at four different months. These images include October 2015, December 2015, May 2016 and August 2016, representing different seasons from autumn, winter, spring, and summer respectively. Following the study objectives, various acquisition images will address the phenology of tree species within BFNP as input features for the classifiers. Sentinel-2 has 13 different wavelength bands with various resolutions (10 m, 20 m and 60 m). This study uses only ten bands with 10 m (B2-4; B8) and 20 m (B5-7; B8A; and B11-12) resolution. The other three bands (B1, B9, B10) are not utilized for this study because they are extremely coarse (60 m in spatial resolution) and are considered as 'atmospheric bands' which are not useful for this study. All the bands were resampled to the resolution of 20 m to preserve the information in the red-edge bands (B5, B6, B7) since the red-edge wavelength is essential for capturing chlorophyll content information within vegetation. Table 2.4 shows the selected bands that will be used for this study.

All the Sentinel-2 images used in this research are at level 1-C where the images are geometrically corrected, and per-pixel radiometric measurements are provided in Top of Atmosphere (ToA) reflectance (ESA, 2015). However, not all the images are in the cloud-free condition, particularly for May 2016 (spring) image. Therefore, a cloud masking method was applied by using *"Cloud Mask Model"* which is run in ERDAS Imagine® software to mask the cloud. The BFNP land cover map was used to mask the area outside our study area from the images. Since this study is not focusing on land cover classification, other objects such as built-up area and water bodies are masked out. This activity was achieved through visual interpretation from very high-resolution aerial photograph.

To create multi-season image, all bands from four different season images were stacked into one file image. This multi-season image consists of 40 bands derived from 10 bands of each single season image. Furthermore, the images are divided into three datasets, i.e., single season (summer) without red-edge bands (6 bands); single season images (summer, autumn, winter and spring, each image has 10 bands); and Multi-season image (40 bands).

Table 2.4. The characteristics of selected bands of Sentinel-2 which is used in this study (adapted from ESA, 2015).

Band	Resolution	Wavelength	Description	Information
B2	10 m	490 nm	Blue	Resampled to 20 m
B3	10 m	560 nm	Green	Resampled to 20 m
B4	10 m	665 nm	Red	Resampled to 20 m
B5	20 m	705 nm	Red-Edge 1	-
B6	20 m	740 nm	Red-Edge 2	-
B7	20 m	743 nm	Red-Edge 3	-
B8	10 m	842 nm	Near Infrared (NIR)	Resampled to 20 m
B8A	20 m	865 nm	Near Infrared 2 (NIR)	-
B11	20 m	1610 nm	Shortwave Infrared (SWIR)	-
B12	20 m	2190 nm	Shortwave Infrared (SWIR)	-

2.3. Methods

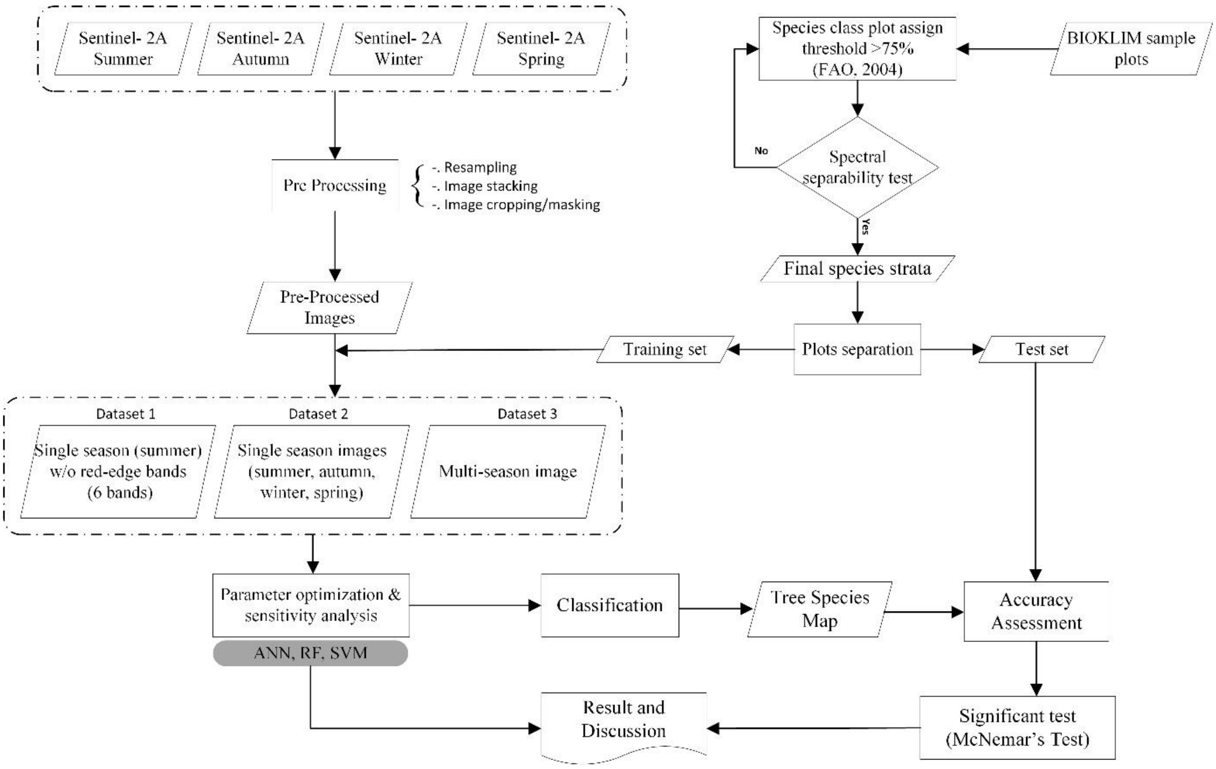


Figure 2.2 A flow chart of the methodology and classification process. Three datasets were produced from different season images of Sentinel-2. The dataset 3 (multi-season image) were used to determine the optimum value of ANN, RF and SVM parameters. After the optimum value was found, then it will be applied to datasets 1 and 2 in classification process. To assess the performance of the algorithms, the sensitivity analysis was applied to the algorithms parameters. This is also done by using dataset 3.

2.3.1. Classification procedure

Regarding the study objectives, we applied three machine learning algorithm classifiers (i.e., ANN, RF and SVM) for tree species classification and then evaluated the performance of these algorithms on different datasets (i.e., single season (summer) without red-edge bands; single season images; and multi-season image). The evaluation of the performance of the algorithms was done through sensitivity analysis of the parameters (Figure 2.2). The purpose of conducting sensitivity analysis in algorithm parameter settings, other than performance evaluation, is to find the optimum value for each parameter for each algorithm, given the input data set. The sensitivity analysis is only applied to one dataset, i.e. multi-season dataset, since accuracy of multi-season image potentially outperforms the other datasets, and further, the defined optimum value will be applied to all datasets for each algorithm.

2.3.2. Parameters optimization

Artificial Neural Network (ANN)

As mentioned before in §1.1.3, The MLP is the most common ANN architecture that used by remote sensing application (Tso & Mather, 2009; Cracknell & Reading, 2014; Kumar et al., 2015; Raczko & Zagajewski, 2017). Therefore, this study is applied ANN-MLP for classifying dominant tree species. Five parameters need to be optimized in using ANN-MLP, i.e., training threshold contribution, learning rate, momentum, the number of hidden layers and number of epochs (iterations). The training and the classification processes were applied in ENVI-IDL® software using the *Neural Net* classification tool. In this study, the number of iterations was determined by the lowest root-mean-square (RMS) error by running the model in 1000 iterations. As seen in figure 2.3, the RMS tended to converge beyond approximately 400 iterations. Thus, 400 iterations were decided to be used in this study since beyond this value there are no substantial changes in the errors, and the model was tended to overfit, and further needs more computational effort. The other parameters were also determined through experimental testings, and the sensitivity analysis is used to find the optimum value. Since the parameters in ANN are user-defined based,

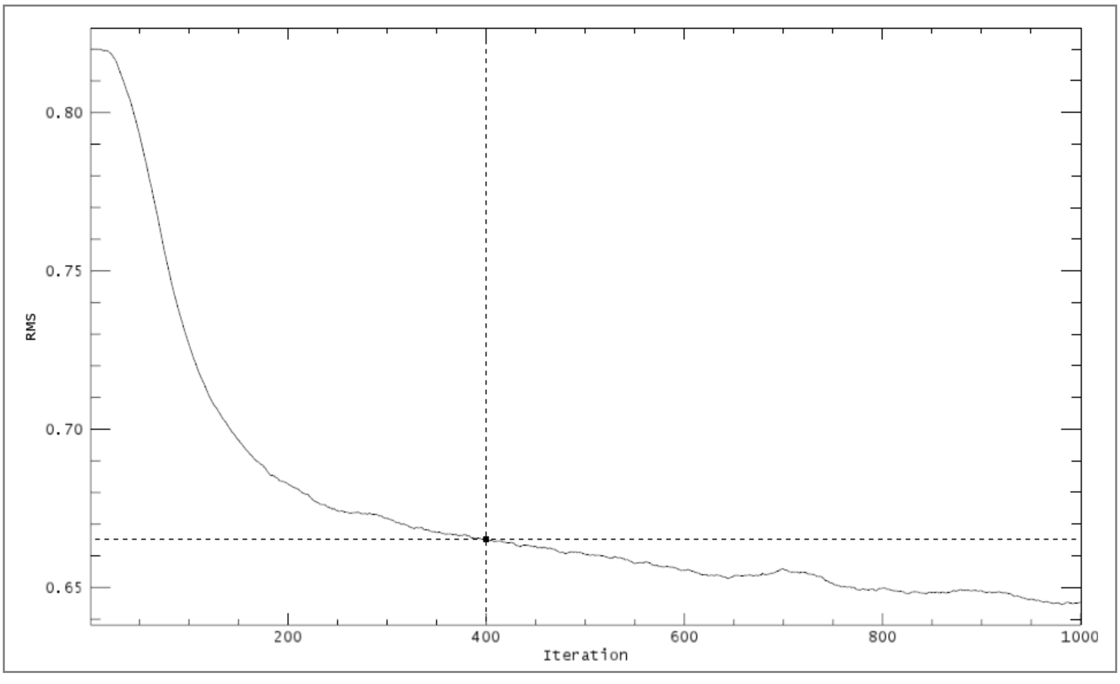


Figure 2.3. The result of optimum iteration number for ANN algorithm in this study. The target point shows the number of iteration that been chosen for this study is 400 with RMS 0.6652. The test was running in 1000 iterations.

often the optimum number was found based on the user experience and experimental testing (Skidmore et al., 1997; Kavzoglu & Mather, 2003). Thus, based on the experimental results, this study uses the value of 0.1, 0.2, 0.2, 2 for training threshold contribution, learning rate, momentum and number of hidden layers respectively (Table 2.5). All the training and the classification processes used logistic activation function.

Random Forest (RF)

The training and classification processes for RF classifier were performed in R environment using the *randomForest* package (Liaw & Wiener, 2002). Two parameters are needed to optimize; those are Ntree (number of trees) and Mtry (number of input features). These two parameters are user-defined which allow user to used arbitrary values. However, this is not an issue since RF is insensitive to overfitting and the use of Out-Of-Bag (OOB) produced unbiased estimation (Breiman, 2001; Pal, 2005; Belgiu & Drăguţ, 2016). The OOB is based on the bagging approach by drawing a subset of the training set from replacement (Breiman, 1996). Through the bagging approach, the same sample can be selected many times while the others might never chosen (the OOB). Breiman (2001) suggests using one-third of samples be used as OOB for internal cross-validation to estimate the RF performance. Based on the advantage of the OOB, this study is using an experimental activity corresponding to Rodriguez-Galiano et al., (2012b), by setting different Mtry and running the algorithm with different Ntree settings (this study used 0-1500 trees) to find the lowest OOB error. Although there is a function in the package called *tuneRF,* it was not suggested to use it since there is no research yet about the effects of selecting Mtry to optimize OOB error rates for RF using remote sensing data (Rodriguez-Galiano et al., 2012b).

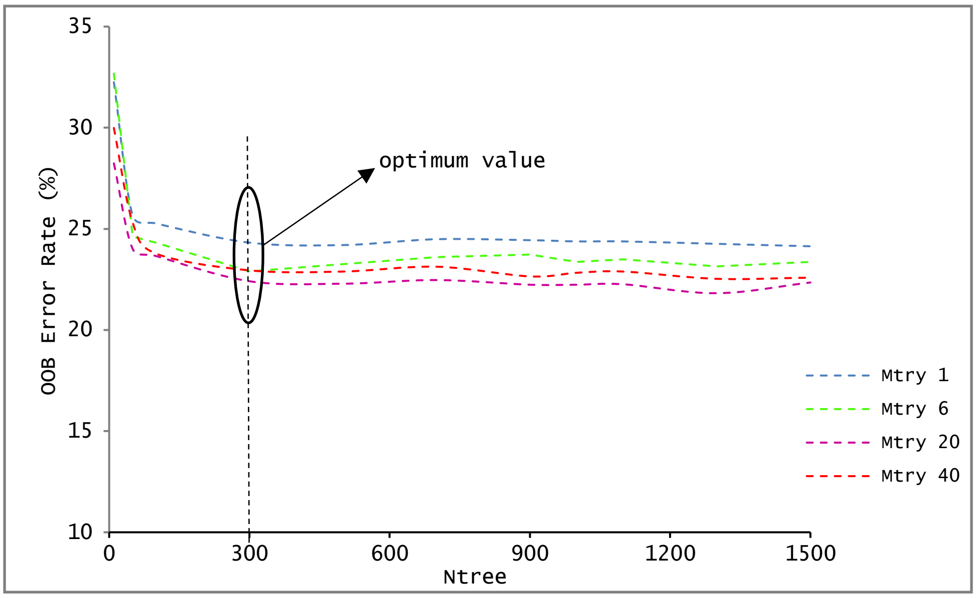


Figure 2.4. The Out-Of-Bag (OOB) error estimation of random forest algorithm. The estimation is using four different Mtry values with Ntree that is set from 0-1500 trees. From this estimation, 300 trees and Mtry 6 were found as optimum values as the lowest error reached. Beyond Ntree 300, the error tends to converge.

As shown in figure 2.4, the estimation used different values of Mtry from multi-season image which consists of 40 bands. The Mtry value was set to four different values, i.e., Mtry 1, Mtry 6, Mtry 20 and Mtry 40. The Mtry 6 is derived from the square root of the total input features (40 bands), as suggested by Gislason et al. (2006) to use the square root of total input features for defining Mtry. From the estimation, it was found that after 100-300 trees the error rates are reached the lowest rate and tended to converge in all Mtry, meaning that the increase in number of trees does not affect the model accuracy. Therefore, this study decides to use 300 trees instead of 100 trees as suggested by Breiman (2001) to run past the point where the error converges to avoid OOB bias. In respect to Mtry, it was found that Mtry 6, 20, and 40 had slightly different in error rate. Even more, in 300 trees the values of Mtry 6 and 40 are similar. Thus, this study decided to use the square root of input features (6 bands) for defining the Mtry (Table 2.5).

Support Vector Machine (SVM)
Based on the results from Huang et al.,(2002) and Kavzoglu & Colkesen (2009), this study applied non-linear SVM with the Radial Basis Function (RBF) kernel to the classification process using the *e1071* package from Meyer et.al (2015) in R environment. As the consequence of using RBF kernel, two parameters were needed to be set, i.e., cost (C) and gamma (γ). Both parameters are user-defined and the value can be arbitrarily chosen. This study uses a certain range of values for both C and γ and then applies it with a function from *e1071* package called *best.tune* with 10-fold cross validation to obtain the optimum parameters combination (Li et. al., 2016). There are 8 values of C parameter (10^{-2}, 10^{-1}, 10^{0}, 10^{1},10^{2},10^{3},10^{4} and 10^{5}) and 20 values of γ parameter which ranging from 0-1 at an interval of 0.05. Based on *best.tune* result, the optimum combination was found to be C-10^{1} and γ-0.05 (Table 2.5).

Table 2.5. Summary of parameters optimization for ANN, RF and SVM. All the values were optimized in multi-season dataset and then will be applied to other datasets.

Algorithms	Parameters	Values
ANN	Training Threshold Contribution (weight)	0.1
	Learning rate	0.2
	Momentum	0.2
	Hidden Layers	2
	Iteration	400
RF	Mtry	Square root of input features
	Ntree	300
SVM-RBF	Cost (C)	10
	Gamma (γ)	0.05

2.3.3. Accuracy assessment and significance test

The accuracy performance of each experiment and algorithm are assessed by using a calculation based on the confusion matrix. The confusion matrix is commonly used for accuracy assessment and widely used in remote sensing applications as a simple cross-tabulation which gives information of the mapped class against the ground truth for a sample data at certain locations (Congalton, 1991; Foody, 2002). Besides being the common assessment method in remote sensing, the confusion matrix also gives an adequate representation of the present error on the map since the individual accuracies in each map category were described (Congalton & Green, 2009). The calculations used for this study are overall accuracy and the Kappa statistic. The overall accuracy is derived from the sum of the number of samples that are correctly classified in both mapped classes and ground truth data divided by the total number of all samples (Congalton, 1991; Stehman & Czaplewski, 1998).

To assess the reliability of the accuracy assessment results, this study uses Kappa statistics (Cohen, 1960) to measure the map accuracy agreement. Kappa measures the differences between the actual agreement in error matrix and the chance agreement that is indicated by the row and column total (Congalton,1991). The value of Kappa can range from -1 to 1 where 0 represents the amount of agreement that is expected from random chance and 1 represents perfect agreement. Cohen (1960) noticed that Kappa value of less than 0 is also possible although unlikely in practice. In this study, the interpretation of Kappa value follows the benchmark from Landis & Koch (1977), where < 0.00 is poor agreement; 0.00-0.20 is slight; 0.21-0.40 is fair; 0.41-0.60 is moderate; 0.61-0.80 is substantial, and 0.81-1.00 is almost perfect agreement.

This study uses the same sample sites for both training and test sets in accuracy assessment to guarantee the comparability between the experiments and or classifiers. This is a common procedure that has been applied in many remote sensing studies (de Leeuw et al., 2006; Raczko & Zagajewski, 2017). The consequence of this procedure is that samples used in both classifiers are not independent. Therefore, the testing of significant difference from two experiments or classifiers can be done by McNemar's test instead of Kappa z-test (Foody, 2004; de Leeuw et al., 2006). The McNemar's test is based on chi-square statistics which computed from two error matrices (usually 2×2 contingency table). The computation calculates total cases that correctly classified by classifier 1 but wrongly classified by classifier 2 and vice versa (Manandhar et al., 2009). The McNemar's test is quite preferable since this test is non-parametric, easy to execute and simply understandable (Manandhar et al., 2009; Adam et al., 2014). It should be noted that this study is using 95% ($\alpha = 0.05$) as the level of confidence. Thus, all the tests will be considered to have a significant difference if p-value < 0.05.

3. RESULTS

3.1. The classification results

According to the research objectives, the classification results are divided into two parts. The first part describes the accuracy of the tree species classification in a single season, in this case summer season, by optimizing red-edge bands in Sentinel-2 imagery. The second part describes how classification accuracy is improved by combining different images of different season to optimize the information from tree phenology.

3.1.1. Accuracies with and without red-edge bands

Table 3.1 provides a comparison of the accuracy result from three machine learning algorithms with and without the red-edge bands. The result shows that the accuracy from SVM outperformed other algorithms in both experiments (with or without red-edge bands) and followed by ANN and RF respectively. The mapping accuracy is improved by adding red-edge bands in all classifiers by 1-2%, although the accuracy in each algorithm is considered low in both experiments (with and without red-edge bands). This low accuracy may be caused by the spruce-beech and mix class. As seen in error matrix example (Table 3.3) from one of the algorithm (SVM), both spruce-beech and mix class yielded an extremely low producer's and user's accuracy compare to the other classes. However, the difference of the accuracies (producer's, user's, overall) from both experiments are not significantly difference for all algorithms. This result is confirmed by the McNemar's test result (Table 3.2) which shows that there is no a significant difference between the experiments where p-value > 0.05 for all algorithms. It should be noted that this experiment is only applied to a single season image (i.e., summer image).

Table 3.1. Comparison of overall mapping (OA) accuracy and Kappa for tree species classification from different algorithms (i.e., SVM, RF and ANN) in image without and with red-edge bands.

Red-Edge Bands Experiments	SVM		RF		ANN	
	OA	Kappa	OA	Kappa	OA	Kappa
single season (summer) without red-edge bands	61.2%	0.50	58.4%	0.48	59.0%	0.46
single season (summer) with red-edge bands	62.2%	0.51	60.4%	0.49	60.8%	0.48

Table 3.2. The McNemar's test results for tree species classification accuracy from different algorithms (i.e., SVM, RF and ANN) in red-edge experiment. f_{ij} means the number of cases that were correctly classified in classifier *i* but wrongly classified in classifier *j* (i,j = 1,2) and the confidence level is 95% (α = 0.05). The result however, did not show any significant difference since p-value > 0.05 in all algorithms.

Algorithms	Red-Edge vs. without Red-edge			
	f12	f21	Chi-Sqr	p-value
SVM	228	192	2.92	0.09
RF	339	304	1.80	0.18
ANN	109	85	2.73	0.10

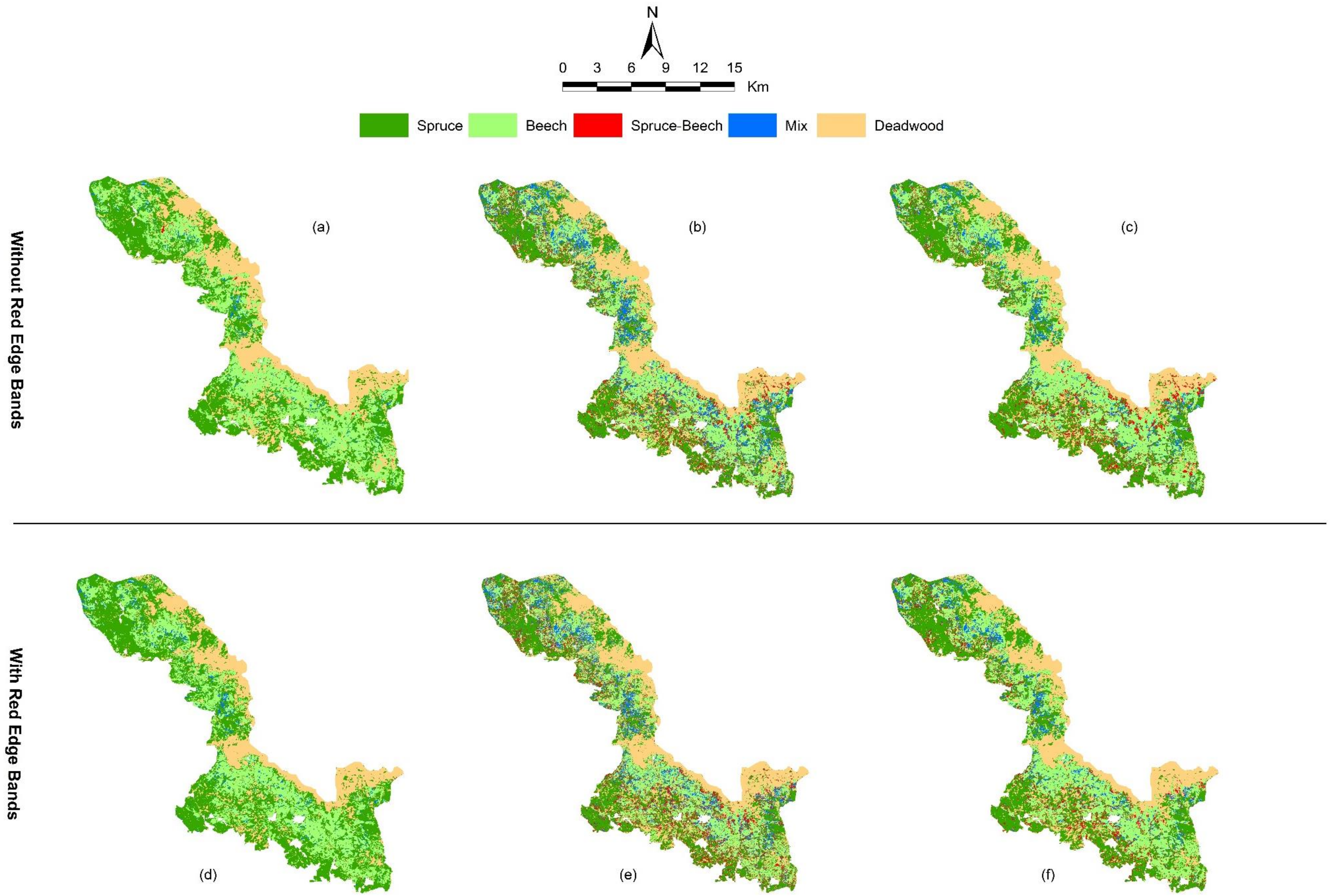


Figure 3.1. Maps of red-edge bands experiment result. Maps a, b and c above show classification results of ANN, RF and SVM respectively when red-edge bands were excluded. Maps d, e and f below show classification results of ANN, RF and SVM respectively when red-edge bands were included.

It can be seen in Figure 3.1, that maps created by each algorithm have a similar pattern. The class mix and spruce-beech are either relatively missing, or the patch size is relatively too small to appear in ANN classification; this occurred in both the experiments, with or without red-edge bands. In contrast, RF and SVM succeeded in bringing out the class mix and spruce-beech with more patches and larger size. This was the case with both experiments, with or without red-edge bands.

Table 3.3. Error matrices example of one of the algorithm (SVM) in red-edge experiment. Table (a) represents classification without red-edge bands and (b) represents classification with red-edge bands. OA: Overall Accuracy; PA: Producer's Accuracy; UA: User's Accuracy.

(a)

	Spruce	Beech	Spruce-Beech	Mix	Deadwood	**Total**
Spruce	183	0	50	24	2	259
Beech	13	126	37	43	6	225
Spruce-Beech	7	12	20	8	3	50
Mix	3	19	31	17	1	71
Deadwood	14	1	7	0	97	119
Total	220	158	145	92	109	724
PA	83.2%	79.7%	13.8%	18.5%	89.0%	
UA	70.7%	56.0%	40.0%	23.9%	81.5%	
					OA	**61.2%**

(b)

	Spruce	Beech	Spruce-Beech	Mix	Deadwood	**Total**
Spruce	182	0	50	18	1	251
Beech	10	125	40	43	5	223
Spruce-Beech	12	14	22	11	1	60
Mix	3	23	25	20	1	72
Deadwood	10	0	7	0	101	118
Total	217	162	144	92	109	724
PA:	83.9%	77.2%	15.3%	21.7%	92.7%	
UA:	72.5%	56.1%	36.7%	27.8%	85.6%	
					OA:	**62.2%**

3.1.2. Accuracies of single season vs. multi-season

It can be seen in Table 3.4, the accuracy of the SVM algorithm is higher than other classifiers in all experiment (single season and multi-season), while ANN outperformed the RF except for the autumn and multi-season cases. Thus, only the SVM consistently outperformed the other algorithms. Compared to the single season imagery, the multi-season imagery has an improved map accuracy of about 6-8% and has the highest accuracy in all three algorithms. In single season images experiment, the result shows that summer image has higher accuracy than other seasons.

Table 3.4. Comparison of overall mapping (OA) accuracy and Kappa coefficient for tree species classification from different classifiers (i.e., SVM, RF, and ANN) in four different seasons and combination of multiple seasons.

Season Experiments	SVM		RF		ANN	
	OA	Kappa	OA	Kappa	OA	Kappa
Summer	62.2%	0.51	60.4%	0.49	60.8%	0.48
Autumn	60.1%	0.45	56.6%	0.45	53.6%	0.41
Winter	57.7%	0.45	55.5%	0.43	57.0%	0.43
Spring	59.8%	0.46	56.1%	0.43	57.6%	0.45
Multi-season	63.0%	0.52	62.3%	0.51	61.1%	0.48

As can be seen in Figure 3.2 and Appendix 1A and 1B, the maps produced by each season for each classifier have similar distribution patterns for spruce and beech class, since these classes are the dominant species. However, the distribution patterns of mixed and spruce-beech class vary across seasons, as the patches are too small to visualize. For example, in Figure 3.2 the maps produced from SVM classification show that the mix and spruce-beech class can be easily identified in the autumn and winter images. Meanwhile, in ANN classification, summer image shows that spruce-beech class tends to be missing in the map (Appendix 1A). In opposite with the single season image, the multi-season image was succeeded to brought out all classes as well as increase the map accuracy regardless the algorithms.

Using McNemar's test, the significant difference between seasons were tested for each classifier. In general, it appears that single season and multi-season classification accuracies are comparable as show by the non-significant results in most tests. The experiments in ANN classifier, both in Table 3.5a and b, shows that only in 'multi-season vs. autumn' and 'summer vs. autumn' was there a significant difference with p-value < 0.05. As well as ANN, SVM classifier shows that only three seasons exhibit a significant difference in classification accuracy (i.e., 'multi-season vs. winter'; 'multi-season vs. spring' and 'winter vs. spring'). Among the ANN, RF and SVM, RF algorithms succeeded in providing significant results more often than ANN and SVM. It can be seen in Table 3.5a and b, that from 10 experiments, RF provided five significant results. In total, there were 30 experiments from multi-season and single season experiments, but in only one third or 10 experiments could be confirmed that using different seasons or combining all seasons would improve the accuracy significantly. Therefore, in context of classifying the dominant tree species in BFNP, the use of different seasons or multi-season image did not yield a significant improvement in accuracy regardless of the classifier algorithms used.

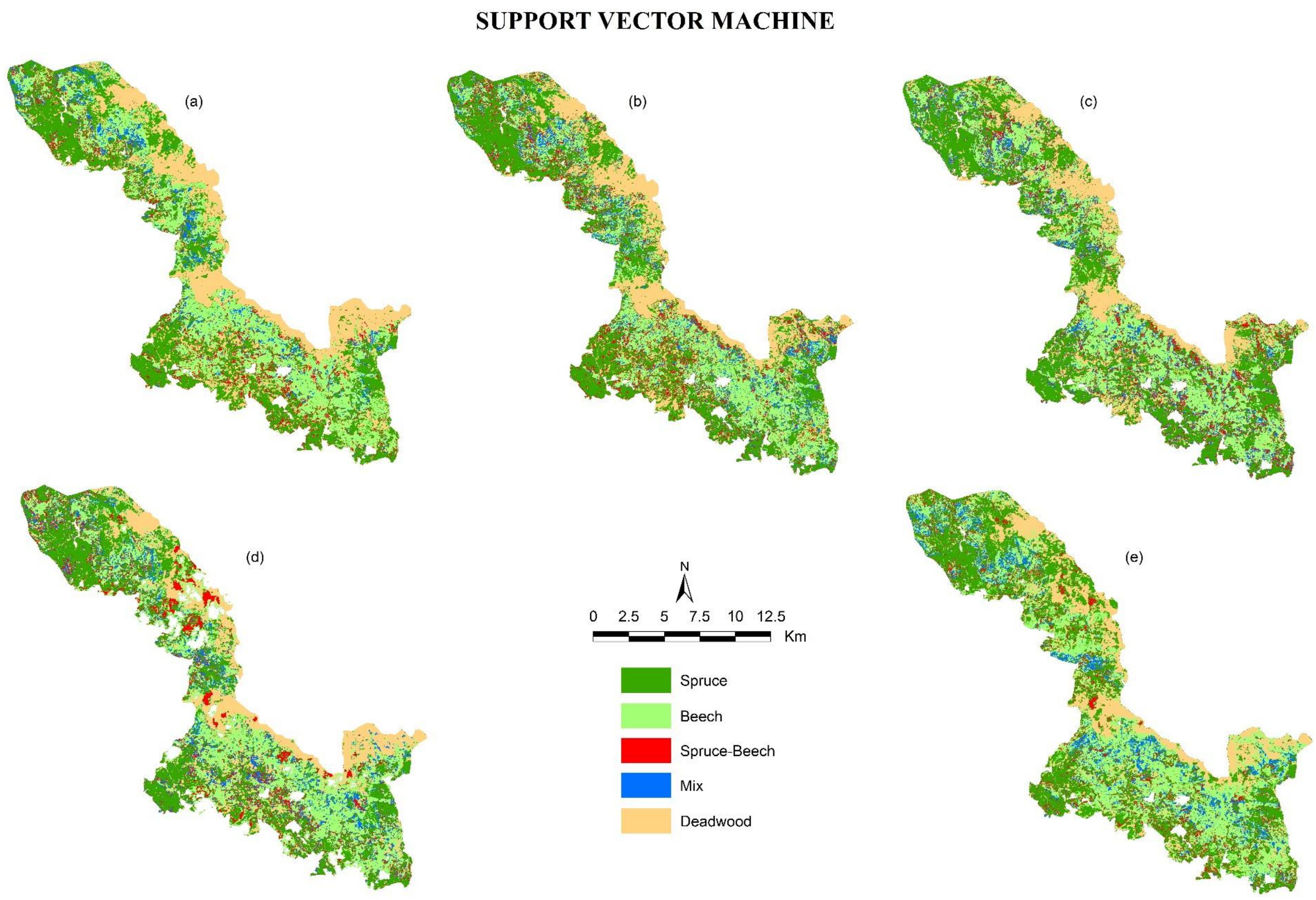


Figure 3.2. Example results from season experiment of one of the classifiers, i.e., Support Vector Machine (SVM). The maps a, b and c above represent summer, autumn, and winter images respectively. Maps d and e below represent spring and multi-season images respectively.

Table 3.5. The McNemar's test results of tree species classification from different season experiments in different algorithms (i.e.,SVM, RF and ANN). Table a and b represent McNemar's test results in multi-season and single season experiments respectively. f_{ij} means the number of cases that were correctly classified in classifier *i* but wrongly classified in classifier *j* (i,j = 1,2). The highlighted values are the experiments with a significant difference where p-value < 0.05.

(a)

Pairwise Seasonal Experiments	SVM				RF				ANN			
(Multi vs single season	$f12$	$f21$	Chi-Sqr	p-value	$f12$	$f21$	Chi-Sqr	p-value	$f12$	$f21$	Chi-Sqr	p-value
Multi-season v Summer	256	295	2.62	0.11	295	336	2.54	0.11	182	199	0.10	0.76
Multi-season v Autumn	341	376	1.61	0.20	336	254	11.12	0.00	274	226	4.42	0.04
Multi-season v Winter	327	276	4.15	0.04	377	271	17.01	0.00	237	204	2.32	0.13
Multi-season v Spring	271	366	13.87	0.00	337	207	30.59	0.00	228	244	0.48	0.49

(b)

Pairwise Seasonal Experiments	SVM				RF				ANN			
	$f12$	$f21$	Chi-Sqr	p-value	$f12$	$f21$	Chi-Sqr	p-value	$f12$	$f21$	Chi-Sqr	p-value
Summer v Autumn	287	253	2.02	0.16	359	380	0.54	0.46	256	159	22.21	0.00
Summer v Winter	253	289	2.26	0.13	372	415	2.24	0.13	188	204	0.57	0.45
Summer v Spring	243	259	0.45	0.50	383	403	0.46	0.50	167	153	0.53	0.47
Autumn v Winter	313	311	0.00	0.97	299	355	4.63	0.03	294	252	3.08	0.08
Autumn v Spring	334	294	2.42	0.12	300	319	0.52	0.47	218	254	2.60	0.11
Winter v Spring	269	331	6.20	0.01	286	375	11.72	0.00	206	178	1.90	0.17

3.1.3. Comparison of overall accuracies among ANN, RF and SVM algorithms

Overall, SVM yielded the highest accuracy, followed by RF and ANN respectively in both experiments (red-edge and season/multi-season), although the value of these three algorithms is slightly different. Two experiments (i.e., single season (summer) image and multi-season image) were used to test the significant difference of map accuracy between ANN, RF and SVM using the McNemar's test. The two experiments were chosen as they produced the highest accuracy among the other experiments and were considered to be able to represent the whole classification process between ANN, RF and SVM algorithms (Table 3.6a and 3.6b). Apparently, from two experiments, only SVM has a significantly higher mapping accuracy compared to ANN with a p-value < 0.05.

Table 3.6. The McNemar's test results of tree species classification accuracy from pairwise algorithms. Table (a) presents results from pairwise algorithms in single season (summer) image and table (b) presents results from multi-season image. f_{ij} means the number of cases that were correctly classified in classifier *i* but wrongly classified in classifier *j* (i,j = 1,2). The highlighted values are the experiments with a significant difference where p-value < 0.05.

(a)

Pairwise Algorithms	Single Season (Summer)			
	*f*12	*f*21	Chi-Sqr	p-value
SVM v RF	272	319	3.58	0.06
SVM v NN	181	142	4.47	0.03
RF v NN	247	262	0.39	0.53

(b)

Pairwise Algorithms	Multi-season			
	*f*12	*f*21	Chi-Sqr	p-value
SVM v RF	190	211	1.00	0.32
SVM v NN	338	283	4.70	0.03
RF v NN	299	169	35.56	0.00

3.2. Parameter optimization

The other aim of this research is to evaluate the performance of ANN, RF and SVM algorithms in mapping the tree species. In general, the first objective is to find which algorithm obtained the highest accuracy for tree species mapping, with the results described in section 3.1. Section 3.2 describes the evaluation of the ANN, RF and SVM parameter settings including the quality of sample data to gain an understanding of the classification process, and understand why such algorithm obtains certain accuracy. All the evaluations were applied for multi-season imagery, and the sensitivity analysis of algorithms parameter settings is applied for training set and test set to obtain comprehensive information on classifier algorithm's behaviour.

3.2.1. Class separability

Table 3.7. provides the results of class separability by using Transformed Divergence (TD) test. The test showed relatively poor separation between the classes except when the tree species classes were paired with deadwood class. The lowest separation was showed by the pair of mix and spruce-beech class. The results indicate that in general the spectral between the species is possibly correlated in the Sentinel-2 image. The visualization of the class separability also provided in 2-D scatterplot in Figure 3.3 and Appendix 2 which shows the spectral reflectance of each class between band 4 (NIR) vs. band 5 (red-edge 1). Clearly, the scatterplot shows that there are no clear boundaries between the tree species clusters. The classes, except

deadwood class, tend to be clustered and therefore the expectation of the classification accuracy should not be high.

Table 3.7. The result of tree species class separability test using Transformed Divergence (TD).

	Spruce	Beech	Spruce-Beech	Mix	Deadwood
Spruce	-	1.4372	0.8203	1.0487	1.9988
Beech	-	-	0.6747	0.8410	1.9996
Spruce-Beech	-	-	-	0.3225	1.9999
Mix	-	-	-	-	2.0000
Deadwood	-	-	-	-	-

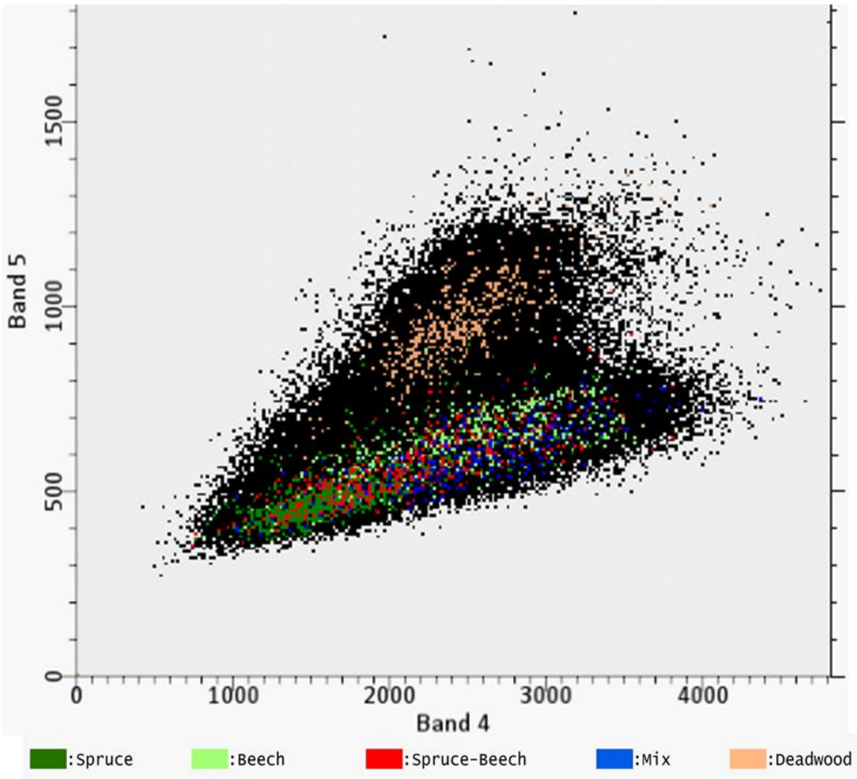


Figure 3.3. The 2-D scatter plot of Sentinel-2 band 4 vs. band 5. Band 4 represented NIR spectrum and band 5 represented red-edge 1 spectrum.

3.2.2. Sensitivity analysis to the size of training samples and class definition threshold

It can be seen in Figure 3.4 that all algorithms show sensitivity to the sample size by the increases of sample size from 170 to 340 training pixels. ANN was the most sensitive with an increase of accuracy by 0.09%, while SVM and RF increased by 0.04% and 0.05% respectively (Table 3.8). The interesting finding from this analysis is after the training sample increase above 340 pixels, the pattern of each algorithm is different from one another. ANN shows insensitivity to the increase in sample sizes after 340 pixels and tends to decrease slightly by 0.01%, but then it turns sensitive when the number of training pixels' crosses 1,360 by showing increases of 0.04%. It found that the optimum sample size for training the ANN algorithm is 340-680 pixels or 20-40% of total training pixels since the overall accuracy outperformed the overall accuracies of SVM and RF. RF accuracy shows insensitivity by slightly increase in accuracy beyond 1,020-1,360 pixels or around 60-80% of total training samples. Therefore, the optimal number of training pixels appears to be around 1,020 pixels to train the RF. In contrast with ANN and RF, SVM is more sensitive indicated by the increase of overall accuracy as the training sample size increases. The other interesting result is that SVM outperformed both RF and ANN only when all the training samples were used. In general, this study shows that beyond 1,020 training pixels might be the optimum training size for use in all algorithms, since the algorithms steadily obtain the higher overall accuracy.

Table 3.8. The overall accuracy (OA) and Kappa of trees species classification from SVM, RF and ANN with different subsets of training sample size.

Sample Size (Pixel)	SVM		RF		ANN	
	OA	Kappa	OA	Kappa	OA	Kappa
170	0.51	0.38	0.53	0.41	0.49	0.35
340	0.55	0.42	0.58	0.46	0.58	0.45
680	0.56	0.43	0.57	0.44	0.58	0.45
1020	0.59	0.47	0.61	0.50	0.57	0.46
1360	0.58	0.46	0.61	0.50	0.57	0.44
1700	0.63	0.52	0.62	0.51	0.61	0.48

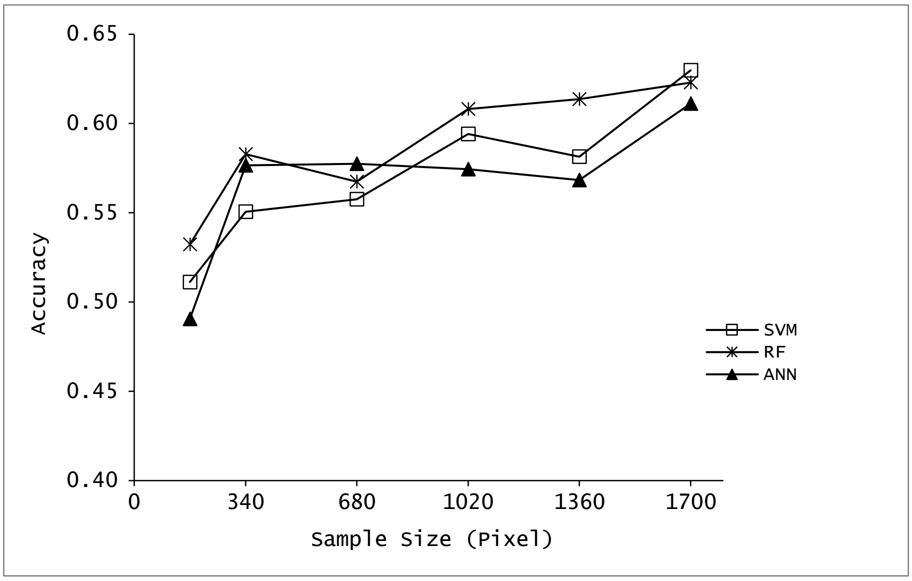


Figure 3.4. Sensitivity analysis of SVM, RF and ANN overall accuracies to the different training sample sizes. The SVM and RF tend to be sensitive to the alteration of sample size, meanwhile ANN shows an insensitive event.

As this study used a threshold of 75% of canopy coverage based on the National Forest Inventory Guideline from FAO (the detail is provided in § 2.2.1) to assign species class, it might be interesting to observe if different threshold value will affect the accuracy of tree species classification. Figure 3.5 shows the results of Kappa by changing class thresholds in ANN, RF and SVM. Based on the graph, SVM outperformed ANN and RF in all different threshold. In general, all algorithms demonstrate the same pattern with regards to the changing threshold. The Kappa decreases when the threshold is changed from 60% to 70%, and then rises gradually until the class threshold is 90%. In general, this event was expected since the threshold is increased, the variance between the classes will decrease. Thus, the consequences of the increase of class threshold besides the higher accuracy, are that the chances of less dominant classes (i.e., mix class) to appear is getting higher since the dominant class such as spruce and beech were merged into mix class, as the canopy coverage was not meet the higher threshold criteria.

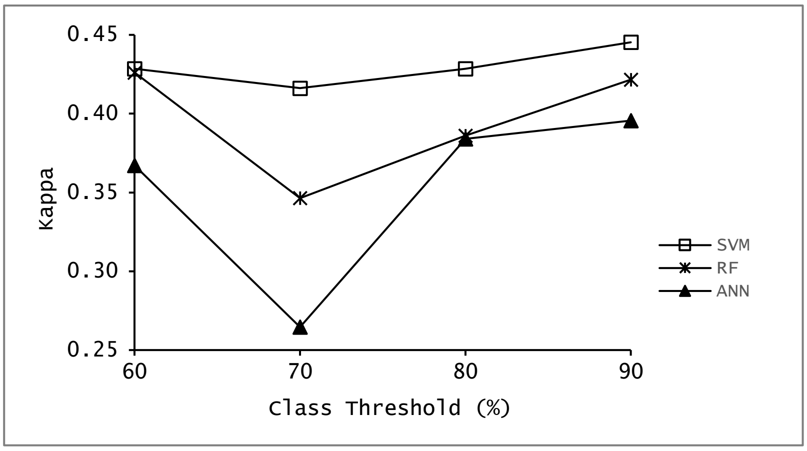


Figure 3.5. Sensitivity analysis of SVM, RF and ANN Kappa analysis to the different class thresholds. In general, all algorithms show similar pattern to the alteration of class thresholds.

3.2.3. Sensitivity analysis to the parameters of ANN, RF and SVM

Besides investigating the performance of SVM, RF and ANN algorithms in tree species classification, the sensitivity analysis to the algorithms parameter settings was also carried out to find the optimum value for the algorithm parameter settings. The other reason why parameters sensitivity needs an investigation, is to observe where and when the classifier is overfitted/over-trained. For further, a prevention can be taken to avoid this situation. Follow sections describe the obtained results from sensitivity analysis for each parameter per classifier.

Sensitivity to ANN parameters

In this study, the parameters of ANN that need to be tuned are epoch (iterations), training thresholds contribution (TTC), the number of hidden layers, learning rate and momentum, where the sensitivity analysis to these parameters is applied to multi-season imagery. The sensitivity of Kappa to epoch (iterations) is shown in Figure 3.6. As can be seen that the training Kappa improves while the test Kappa gradually decreases as the number of iterations increases. The training set becomes relatively diverged beyond approximately 400 iterations where the change in the kappa values is small, thus it can be inferred that above 400 iterations the model becomes over-trained. The over-training was also clarified by test set with the declination after 400 iterations.

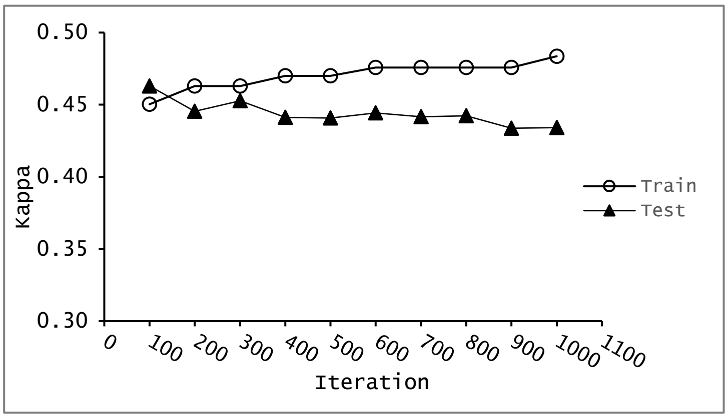


Figure 3.6. The accuracy of training and test set against of epoch (iterations).

The result of sensitivity analysis to the number of hidden layers is shown in Figure 3.7. As can be seen the accuracy of the training set increases gradually as the number of hidden layers increase reaches 2, after which, it begins to decline. In contrast, the accuracy of the test set is declines as the number of hidden layer increase to 1, and remains steady from that point on. Based on the training set sensitivity analysis, the optimum number of hidden layers is 2 since that is when the highest accuracy is obtained.

The sensitivity results of Training Threshold Contribution (TTC) can be seen in Figure 3.8. From the Figure, the graph describes that the increase in TTC does not significantly affect training set as is shown by a relatively plain curve with regards to change in TTC. The test set gave a different result to the training set. The test set accuracy result shows decline as the TTC increases and drops when TTC goes above 0.5. By this result, the optimum range value of TTC is identified to be 0.1-0.2 where relatively higher accuracies can be obtained.

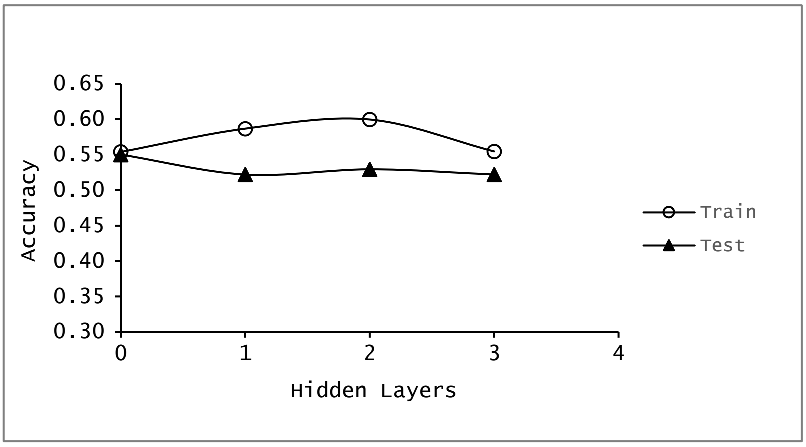


Figure 3.7. The accuracy of training and test set against the number of hidden layers.

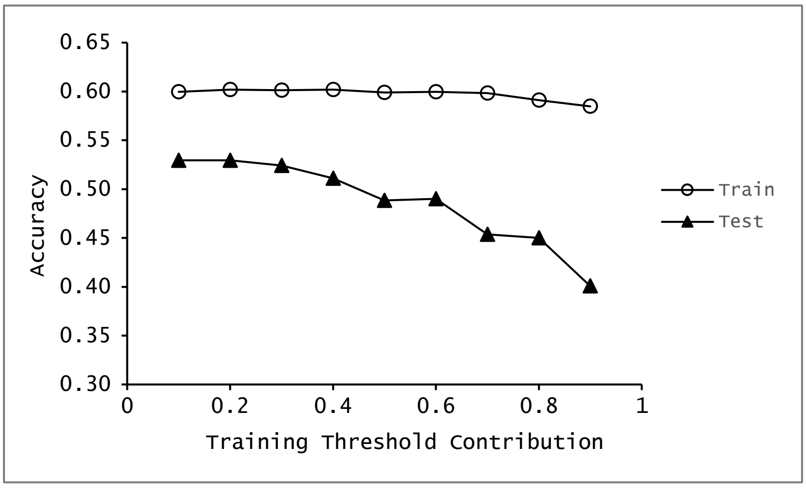
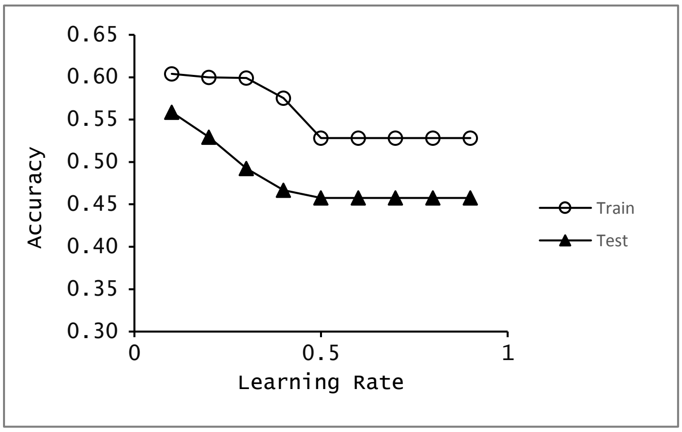


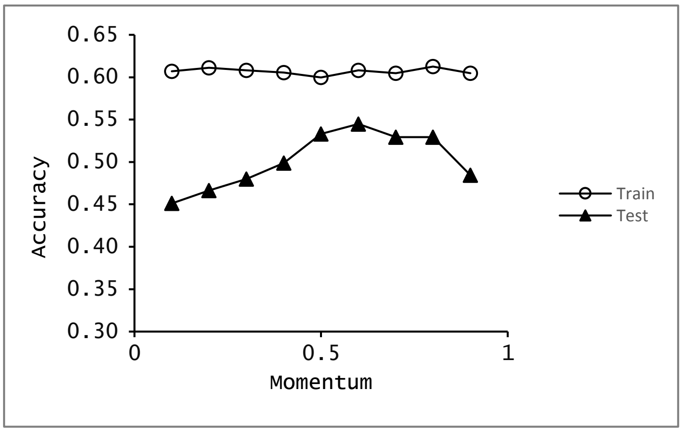
Figure 3.8. Sensitivity of training and test set to training threshold contribution parameter.

Figure 3.9a and b provide results from learning rate parameter and momentum parameter respectively. These sensitivity analyses were run by keeping one of each parameter constant at a certain value. In this case, when the sensitivity test of learning rate was run, the momentum was held constant at 0.2, and for the momentum sensitivity analysis, the learning rate is also constant at 0.2. The pattern of the training set and test set accuracies appear to be similar while testing sensitivity to learning rate parameter (Figure 3.9a). Both decrease as the learning rate increases and both reach asymptotic state after the learning rate goes above 0.5. From this result, the optimum learning rate value is 0.1-0.2 with the use of the constant value of momentum at 0.2. The lower accuracies that were obtained after 0.5 might be caused by the large steps taken by the system as the learning rate was set too high.

The curves of the training set and test set while testing sensitivity to momentum parameter show a contrasting pattern (Figure 3.9b). In training set, the accuracy changes only slightly as the momentum increase up to 1; in fact, it appears that the accuracy is almost constant. However, in the test set, the accuracy gradually increases as momentum moves up to 0.6, after which it begins to decline momentum reaches 1. In general, the momentum did not improve the accuracy for the training set. However, based on this sensitivity result showed by Figure 3.9b, the optimum value for momentum is identified to be 0.2 with the use of the constant value of learning rate at 0.2.



(a)



(b)

Figure 3.9. The sensitivity analysis of learning rate parameter (a) and momentum (b). The learning rate sensitivity is run in constant value of momentum at 0.2 and the momentum sensitivity is run in constant value of learning rate at 0.2.

In addition, to understand the pattern of learning rate and momentum relationship, Figure 3.10 provides a graph which shows the accuracies produced by different combinations of learning rate and momentum along the increases in the iteration. Specifically, there are 5 combinations of learning rate-momentum, i.e., 0.1-0.2; 0.1-0.6; 0.2-0.2; 0.2-0.6; and 0.5-0.9. In general, the combination of small learning rate and momentum values produced the highest map accuracy. The opposite result is shown by the largest value of learning rate-momentum combination (0.5-0.9), where the map accuracy is low, and the graph is shown an unstable pattern (high oscillation). The accuracies produced by 0.5-0.9 combination show a fluctuation with an extreme difference, particularly at iterations of 200 to 300 and 400 to 500, where the accuracy rose and declined by approximately 0.3 or 30%. This condition corresponds to the result from Figure 3.9a where a high value of learning rate cause instability and oscillations of the system/model and produced low accuracy result. Moreover, this combination also uses a high value of momentum (0.9) which causing an increase of oscillations effect.

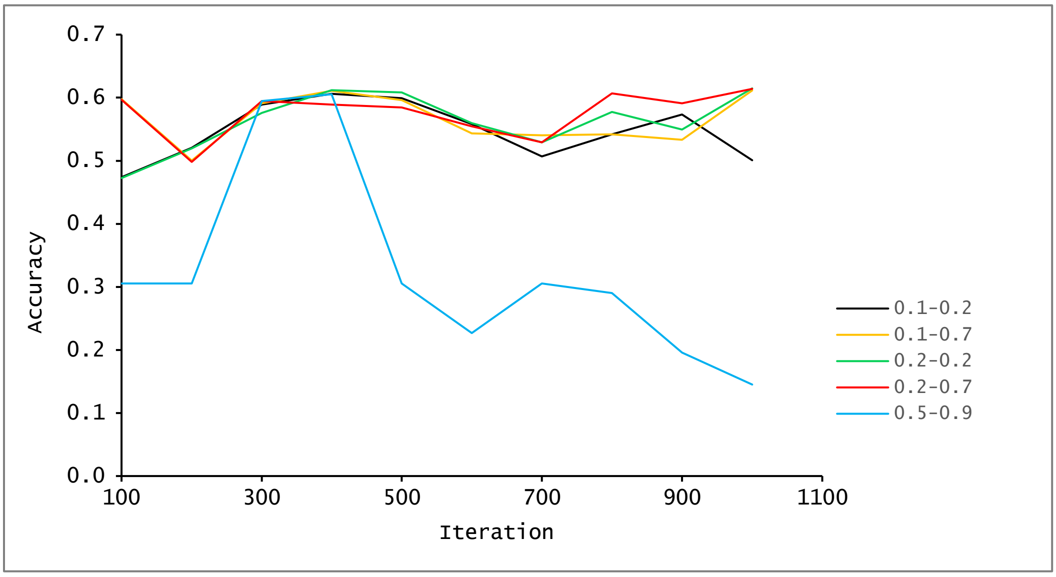


Figure 3.10. The accuracies resulted from different pairwise combinations of learning rate (first value) and momentum (second value). The combinations were running in 1000 iterations.

Sensitivity to Random Forest parameters

In random forest classification, the sensitivity analysis was applied into two parameters i.e. number of features (Mtry) and number of trees (Ntree). Figures 3.11 and 3.12 show the sensitivity of the map accuracy in response to changes in Mtry and Ntree. All tests were conducted on multi-season image which consists of 40 bands as input features in RF classification.

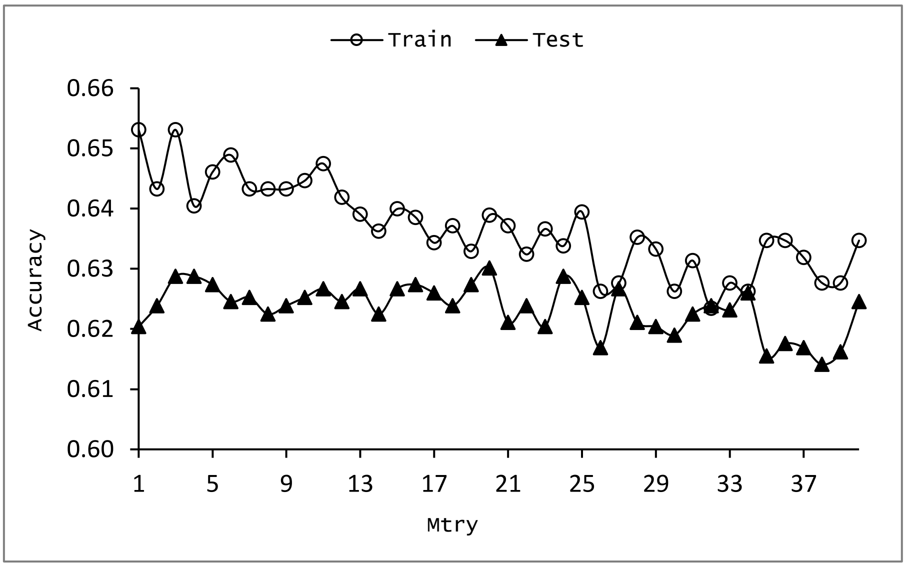


Figure 3.11. The accuracy of training and test set against the number of features (Mtry). Note that there are 40 bands of multi-season image that used as input features.

As can be seen in Figure 3.11, that in general, both training and test sets fluctuate with the change in feature numbers although the accuracies differ slightly. However, overall accuracies show a decline in both training and test sets. This might be caused by the features in RF model are more correlated and became redundant as the Mtry value is getting large. Therefore, the optimum value of Mtry can be considered to be 6 or approximately square root from total input features. Based on the Figure 3.12, the overall accuracy seems to be less sensitive to Ntree than to Mtry parameter shown by the relatively smooth curve as Ntree changes from 100 to 5000, both for trained and test sets. Moreover, the differences between the accuracies are also small which may confirm that the number of trees is not influencing the accuracy substantially. Thus, Ntree can be as large as possible, but then it will affect the computer's memory usage and the computational time. From these results, the optimum number of trees to run the classification is 300 considering the obtained accuracy in sensitivity analysis and that the OOB error rate (Figure 2.3) stabilizes after Ntree value of 300. Beyond this number, in general the accuracy has not affected at all.

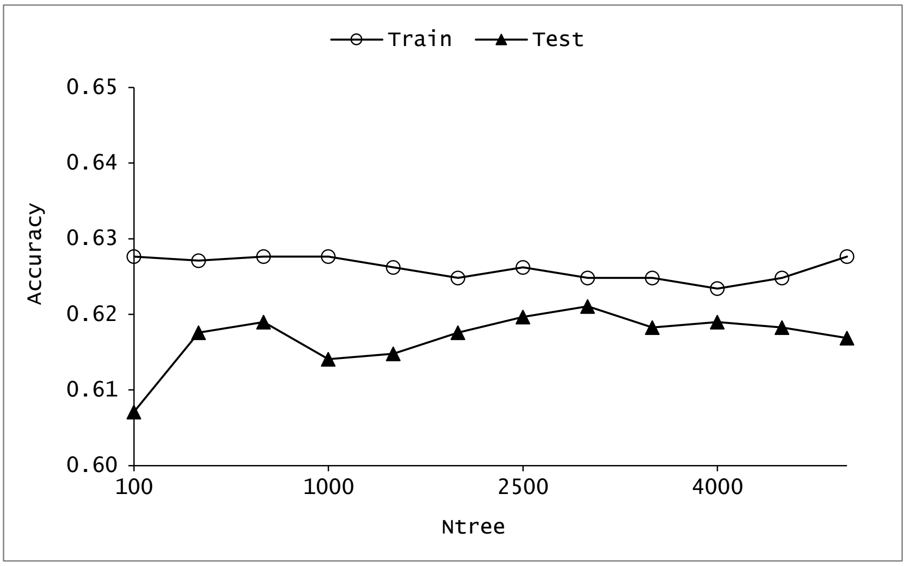


Figure 3.12. The accuracy of training and test set against the number of trees (Ntree).

Sensitivity to Support Vector Machine parameters.

As explained in methods section this research used the SVM algorithm with radial basis function (RBF) kernel. Thus, two parameters will be observed here are cost (C) and gamma (γ). Figures 3.13 and 3.14 show the impact of increase in C and γ values to the obtained overall accuracy respectively. Specifically, the sensitivity test of parameter C uses 8 different values of C, i.e., 10^{-2}, 10^{-1}, 10^{0}, 10^{1},10^{2},10^{3},10^{4} and 10^{5}. Based on this range, Figure 3.13 shows that both the training set and test set have a similar pattern where the

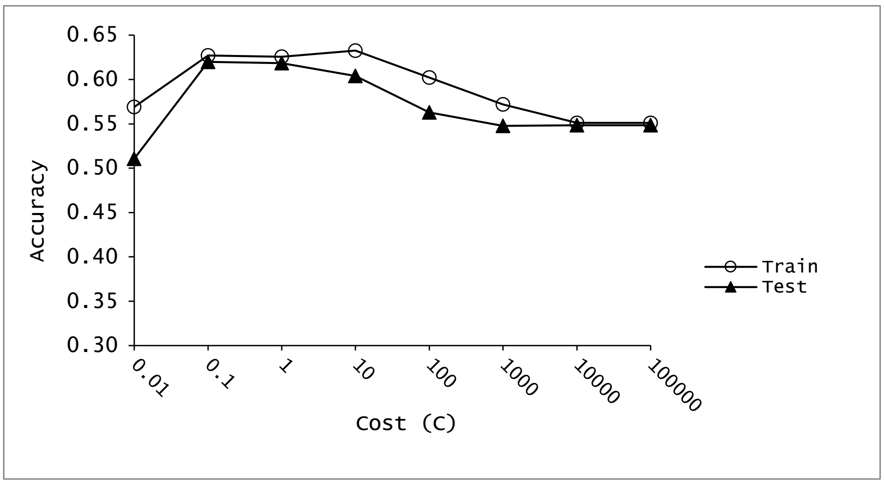


Figure 3.13. The accuracy of training and test set against the cost (C) parameter.

accuracies increase as C increases from 0.01 to 0.1 and then gradually decreased after C increases beyond 10. The model also becomes insensitive to the C value when C reaches 10.000 and beyond. From this result, the application of larger C values -in this case beyond 10- tends to cause an over-fitted model and reduce the accuracy. Therefore, the optimum value of C should be 10 that can be applied to the other experiments since this value leads to the highest accuracy.

For sensitivity analysis to the gamma (γ) parameter, the values between 0-1 were used at an interval of 0.05. In general, for both training set and test set, the overall accuracy tends to decline gradually as the γ value increases (Figure 3.14). This result indicates that larger values of γ lead to an over-fitted model which reduce the accuracy of the model. Based on this sensitivity test, a value of 0.05 is found to be the optimum γ value according to the highest obtained accuracy (0.63). The value based on this sensitivity analysis also confirms to the result of SVM tuning by 10-fold cross validation where 0.05 was also found as the best γ value for this research (Table 2.5).

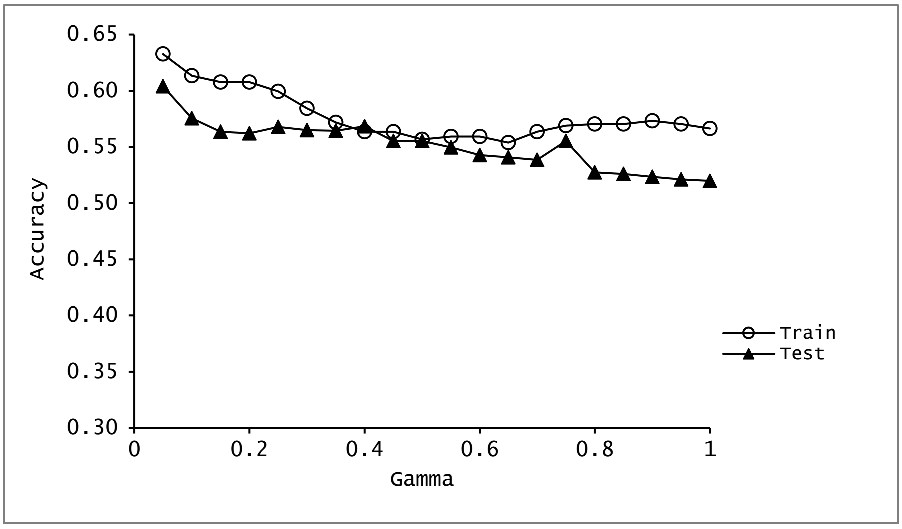


Figure 3.14. The accuracy of training and test set against the gamma (γ) parameter.

4. DISCUSSION

4.1. Uncertainty from tree species sample configuration

In the history of the Bavarian Forest National Park (BFNP), the composition and configuration of tree species have changed over the centuries (Heurich & Englmaier, 2010). During the last two decades, two main species are dominating and spreading in all elevations within BFNP with a relative total of 91.5% (i.e., Norway spruce (67%) and European beech (24.5%)) (Heurich & Englmaier, 2010; Bässler et al., 2015; Sommer et al., 2015). Therefore, since the domination of spruce and beech overspread the existence of other minor/less dominant species, assigning a plot to a certain species becomes a challenge.

Following FAO National Forest Inventory Guideline (2004), this research has used 75% of tree canopy coverage for an individual species as a threshold to assign a species class to that dominate species. This threshold is based on empirical ground observations and measurements. Since remote sensing imagery captures the object reflectance from above, the most representative reflectance from a tree is the uppermost part of the canopy. However, this uppermost canopy cover may not represent actual species configuration within the plot, since there is a possibility that the uppermost canopy consists of certain species which has coverage less than 75%. This uppermost canopy could cover the other species below (understorey) which might be more dominant (>75%). The lower canopy species may provide different spectral information than the uppermost canopy, and the information could be the actual representation of species spectral reflectance (Abdollahnejad et al., 2017).

In fact, this study finds that 75% threshold of canopy coverage did not give a satisfactory result for class spectral separation, particularly for the mix and spruce-beech class. Clearly, this research is focused only on two types of vegetation structure in BFNP (i.e., coniferous and broadleaf forests) and neglects other broad cover classes such as shrub and grass. Further, these two types of forest were broken down into specific classes which were considered as the dominant species, i.e., spruce (coniferous type) and beech (broadleaf type). The other three classes are mix (an aggregate of less dominant species and dominant species < 75%), spruce-beech (if the plot consists of spruce and beech with threshold ≥ 75%) and deadwood (non-vegetated area). The Transformed Divergence (TD) index shows that most of the pairwise classes have low separability except where the classes were paired with the deadwood class. The high separation of deadwood class is easily predictable since the deadwood is a non-vegetated area or has less vegetation which may produce high separability when compared to the other vegetated class.

In contrast with deadwood class, the separability test result is low when the dominant classes (spruce and beech) were paired together, particularly with mix or spruce-beech class. This is clearly depicted by feature space in figure 3.3 and appendix 2, where mix and spruce-beech class create unclear cluster boundary to the dominant species classes, or the other words the clusters tend to be mixed. This issue may occur since the definition of mix class consists of less dominant species which also have the same structure with the spruce such as fir or the beech such as maple and birch. Therefore, the spectra were not capable of differentiating the dominant classes (spruce and beech) and the mixed classes and further led to reducing the accuracy of the classification (Sesnie et al., 2010). The low separability of classes may also reveal that the spectral information of the species is overlapping even though they may not overlap on the ground situation. This issue will affect the classification process which is run by the classifier's algorithm, where it can be considered that the algorithm has failed to distinguish the species spectral information (Skidmore et al., 1988). However, this study confirms that there is no clear boundary between the spectral clusters representing the different classes, and that classes may not be completely distinguished by the classifier's algorithm.

In § 3.2.2, figure 3.5 implied that increasing the class threshold led to the decreases of spruce and beech class number since they are merged into mix class, but then the Kappa increased consequently. This evidence proves that merging the classes that have poor results in spectral signature separation, improve the classification accuracy (Manandhar et al., 2009; Laborte et al., 2010). However, the consequence is that less number of species is identified since the classes might have been missed. Determining different class threshold in the simulation may cause another problem since the training and test set are selected randomly into the size of two-thirds for the training set and one-third for the test set. This issue is creating a probability that some samples are repeatedly selected or not selected at all for training or test set. This issue may explain why the classification Kappa suddenly decreased when the threshold increased from 60%-70% in each algorithm (ANN, RF and SVM).

4.2. Classification improvements

4.2.1. Red-edge band experiment

This study finds that adding the red-edge bands improves the tree species classification accuracy in ANN, RF and SVM algorithms. This result confirmed other studies which also added red-edge bands, particularly if the classification scheme consists of vegetation existence (Schuster et al., 2012; Adam et al., 2014; Massetti et al., 2016; Laurin et al., 2016). The improvement from the red-edge bands of Sentinel-2 imagery may be caused by the sensitivity of red-edge bands to the chlorophyll concentration in the vegetation, particularly in the leaf parts (Horler et al., 1983; Mutanga & Skidmore, 2007). Further, chlorophyll concentration is also related to the leaf pigment which controls the optical properties of leaves at canopy level which vary from one species to another (Horler et al., 1983; Carter & Knapp, 2001; Sims & Gamon, 2002). Thus, the combination of optimizing the sensitiveness of red-edge spectrum and different behaviour of leaves optical properties makes tree species distinguishable in remote sensing imagery which consequently improves the classification accuracy. Interestingly, by adding the red-edge bands, the accuracy improves not more than 3% based on the results from Schuster et al. (2012) and Massetti et al. (2016), as well as this study. Furthermore, this study shows that there is no significant difference by adding the red-edge bands in any of the algorithms. It might be that the red-edge spectrum did not outperform the other spectrum particularly in red and NIR region where the absorption and the reflection of the spectral in vegetation objects occurred. Thus, it can be inferred that red-edge may have a role if it is incorporated with other spectral bands (Adelabu et al., 2013). Another option which is not tested in this study is incorporated red-edge based vegetation indices, where several studies have proven that adding red-edge based vegetation indices may also improve the accuracy of the classification (Schuster et al., 2012; Li et al., 2016; Laurin et al., 2016).

4.2.2. Single season and multi-season experiments

This study finds that combining several images from the different seasons (i.e., summer, autumn, winter and spring) improves the map accuracy in each classifier, as also reported by Hill et al. (2010) and Li et al. (2015). The improvement of the accuracy may be caused by the different phenological stages of each species captured by the images which lead to the higher separation of the spectral signature (Hill et al., 2010; Lisein et al., 2015; Li et al., 2015). Although the accuracy is improved, the interesting thing is that the accuracies between the single season and multi-season images are relatively comparable among the algorithms (ANN, RF and SVM). The McNemar's test shows that in general there is no a significant difference between the experiments of multi-season vs. single season. In single season experiments, the McNemar's test also shows that among the different seasons, the accuracies were not significantly different in general. There are only four experiments out of eighteen that had a significant difference. These findings can be explained by the characteristics of tree species distribution within BFNP. As reported by Bässler et al. (2015), BFNP was dominated by almost 70% of evergreen coniferous type, mainly spruce which has no leaf-off state regardless the season. Therefore, from one season image to another, the variability of species composition information that can be captured by the image are quite low. This condition occurred since the spruce appearance dominates the images and almost covers the other classes appearance, and therefore provide a relatively similar condition.

Since in summer the leaf structure was fully developed for all species, the summer image provides sufficient information of the tree canopy for the algorithms to classify all species classes both coniferous and deciduous types (Immitzer et al.,2012). Therefore, the summer image had the highest accuracy compared to the other season. In contrast to the summer image, leaf condition varies for the deciduous tree species in autumn, winter and spring where the leaf pigment changes due to leaf senescence and leaf-off state in winter. This condition somehow cannot be completely captured by the images since the dominant evergreen spruce covered the other species (mostly deciduous type), and thus the information from the less dominant species is lacked. These results, are also corresponded to the study by Voss & Sugumaran (2008) and Li et al. (2015) where summer image obtains higher accuracy than another season. Meanwhile, the suggestions from Fassnacht et al. (2016), Hill et al. (2010) and Mickelson et al. (1998) to use the beginning of spring or late autumn images cannot be confirmed by the results of this study, since the characteristics of the study area are also different.

4.2.3. Performance comparison of ANN, RF and SVM

The impact from parameter settings in each algorithm and the quality of training sample yielded a range of accuracy from 54%-63% depending on the algorithms in each experiment. The accuracy result did not meet the criteria of 85% accuracy, as was suggested by Anderson et al. (1976) for operational classification map. Therefore, the Kappa was observed to see how reliable the accuracy result is. The Kappa itself yielded values ranging from 0.41-0.52 which according to benchmark from Landis & Koch (1977), have a moderate agreement. The other studies from Wang et al. (2009) and Fleiss (2013) stated that values ranging 0.41-0.75 have a good strength agreement. Thus, the accuracy yielded by each algorithm can be considered as a reliable result. The relatively low map accuracy confirmed the explanation of the poor quality of the training samples since they do not meet satisfactory criteria for class separation. Moreover, it should be noted that this study distinguishes tree species which can be considered as the derivation of one single land cover class (i.e., vegetation/forest). The samples for mix and spruce-beech class do not seem sufficient for training the algorithms, although there is a possibility of mislabelling the training samples as well.

With respect to the map accuracy, this study found that SVM outperformed RF and ANN for all experiments and ANN outperformed the RF except for autumn and multi-season cases. Although these findings correspond to the studies from Huang et al. (2002), Dalponte et al. (2012), Adelabu et al. (2013), and Shang & Chisholm (2013), the accuracy of the algorithms are relatively comparable in each experiment. Moreover, the McNemar's test shows that in general, the differences among the algorithms statistically were not significant. Table 3.5 shows that in general, RF was not significantly different from both SVM and ANN, but SVM had a significant difference to ANN. These results, however, cannot be generalized to be stated that SVM outperforms ANN and RF (Khatami et al., 2016). Since the other studies found that several times either RF and or ANN has outperformed SVM (Adam et al., 2014; Raczko & Zagajewski, 2017), while some studies found that these classifiers are relatively comparable (Attarchi & Gloaguen, 2014; Omer et al., 2015). Thus, it can be considered that these machine learning algorithms, particularly SVM, RF and ANN, are stochastic-based methods. This means that SVM, RF and ANN can produce a different accuracy result and draw a different conclusion depending on several components such as data quality, training purposes, heuristic approaches in the parameter settings and even the characteristics of study site (Nieddu & Patrizi, 2000; Li et al., 2013; Li et al., 2016).

According to the sensitivity to the training sample size, this study found that SVM and RF are relatively more sensitive to training sample size than ANN. This result slightly contrasts with other studies. For example, Li et al. (2016) stated RF is insensitive to the training size which was not confirmed by this study. This study also found that training size for ANN does not have to be large. As shown in figure 3.4, ANN became insensitive beyond 40% (340 pixels) of training size which also corresponds to the statement by Kavzoglu & Mather (2003), where larger training size is only needed by ANN if the data is too noisy. However, in general, the classification accuracy of ANN tends to increase with increasing training size as also reported by Staufer & Fischer (1997). The training size sensitivity of SVM corresponds to the result from Foody et al. (2016), where SVM is sensitive to the training size particularly if class mislabelling has occurred. It can be explained that there are some possibilities why such results were obtained in this study. Firstly, there is a possibility that the quality of the training data was relatively poor regarding the spectral separability. Secondly, the training data for less dominant species classes is insufficient, although this might need a profound investigation. Last, there are some mistakes occurred in the setting of the algorithms parameter and training data mislabelled which may affect the classification process (Foody et al., 2016).

4.3. The impact of algorithms parameter optimization

4.3.1. Artificial Neural Network (ANN)

Compared to RF and SVM, the parameter settings for ANN are more complex, wherein this study five parameters needed to be set. Most of the results from the ANN parameter sensitivity tests show that the pattern of the value of each parameter has a similar tendency as seen in previous studies by Kavzoglu & Mather (2003) and Skidmore et al. (1997). Although, in some parameters, the sensitivity tests show that there is a possibility of mistakes in the parameter settings which produced the over-fitted model. For example, according to Kavzoglu & Mather (2003), the higher momentum should increase the possibility of the model to run in the wrong direction over the error surface leading to lower classification accuracy. However, this study finds that the accuracy in the training set was not affected by an increase in momentum value. On the other hand, the test set was affected by the increases of momentum shown by the relative fluctuation of the curve (figure 3.9b). As the ANN parameter settings is user-defined based, often the optimum number was not found and lead to producing low map accuracy since the parameter optimization was failed. Therefore, in many case, user experience is a key role in determining the optimum value for ANN parameters (Skidmore et al., 1997; Kavzoglu & Mather, 2003; Pal & Mather, 2005; Mas & Flores, 2008).

It was proposed by Mas & Flores (2008) that using few training samples in ANN may disable neural network from deriving classes, while large training samples may cause overfitting and longer time for the system to learn. According to the figure 3.4, ANN in this study only needed 340 training pixels and tended to converge beyond 340. Instead of using 340 training pixels, this study used all training pixels (1,712 pixels) which may have caused the overfitted model. Moreover, as explained in § 3.2.1 and 4.1, generally the quality of the training data is poor with regards to class separability and lack of training samples for less dominant classes (mix and spruce-beech class). These issues can also explain why the ANN classifier is not obtained the optimum accuracy. As suggested by Skidmore et al. (1997), to obtain credible results, the neural networks require good training data which could not be provided in this study.

4.3.2. Random Forest (RF)

Compared to ANN and SVM, this study reveals that RF classifier has the easiest parameters to control. In many studies as well as this study, two parameters are needed to be set, i.e., Mtry (number of features) and Ntree (number of trees). Although both Mtry and Ntree also can be considered as user-defined parameters, finding the optimum value for those parameters is relatively quantifiable and can be done through a certain test (Belgiu & Drăguţ, 2016; Rodriguez-Galiano et al., 2012b). Regarding the Mtry, this study confirms Gislason et al. (2006) suggestion that using a square root of the total number of features as the Mtry value produced higher map accuracy than using all features. In this case, the use of 6 bands (square root of 40 bands) as an Mtry value in multi-season image, obtained the highest map accuracy. Moreover, the sensitivity analysis shows that beyond the Mtry-6 the accuracy tended to decrease. This event corresponds to the observation by Rodriguez-Galiano et al. (2012a) that reducing the value of Mtry will increase the accuracy since the correlation between the trees is reduced. However, Guan et al., (2013) reported that eliminating Mtry can be counterproductive if its excessively reduced leading to the lower accuracy. Therefore, finding the optimum Mtry can be the challenging part of the RF parameters settings, though it can be assisted by the measurement of Variable Importance (VI) by calculating the Mean Decrease in Accuracy (MDA) (Belgiu & Drăguţ, 2016).

This study used 300 trees as a value for Ntree which is different from several studies. In many studies, Ntree was set to 500 trees which is also the default value in the *randomForest* package in R environment. In general, Ntree values ranging from 100 to 5000 trees (Belgiu & Drăguţ, 2016). There is no exact guideline to determine how many trees that are needed to run the RF classification since RF algorithm is insensitive to overfitting, thus the number of trees can be grown as large as possible (Pal, 2005; Guan et al., 2013). However, this will affect the use of computer's memory and computational time (Belgiu & Drăguţ, 2016). To find the optimum value of Ntree, Rodriguez-Galiano et al. (2012b) proposed the Out-Of-Bag (OOB) error estimation. Through this test, it was found that after 300 trees the error is relatively low and converged beyond this value. It was proved in this study that RF classification is insensitive to the overfitting and thus Ntree can be as large as possible, and as long we optimize the Mtry through VI measurement, the accuracy would not decrease.

4.3.3. Support Vector Machine (SVM)

Similar to ANN, the disadvantage of SVM is that the selection of value in parameters setting depends on the user expertise and experience, considered as a subjective selection (Pal & Mather, 2005). This study used SVM with RBF kernel which is commonly used in remote sensing applications and has several times yielded higher accuracy than the other kernels (Huang et al., 2002; Kavzoglu & Colkesen, 2009; Omer et al., 2015). SVM-RBF, as well as RF, has only two parameters, i.e., cost (C) and Gamma (γ). However, since there are no clear guidelines to determine the range value of both C and γ, this study tried to use the range that seemed to be reasonable based on the previous studies, particularly studies from Huang et al. (2002), Samadzadegan et al. (2005) and Qian et al. (2015). Based on the sensitivity results, the accuracies from both C and γ tended to decrease as their values increased. This finding confirmed the studies from Foody & Mathur, (2006) and Qian et al. (2015), that the use of large value in both C and γ would overfit the training data and yield poorly generalization shown by low accuracy. Although, there is a possibility that if the value of the parameters was set too low, the classifier could not capture the complexity of the data shape. Thus, setting the C and γ parameters must be taken carefully. In case we are using the larger γ, then we should decrease the value of C and vice-versa, even though the subjectivity from the user still plays the key role (Foody & Mathur, 2006). Therefore, to reduce the subjectivity, this study used grid search method with 10-fold cross validation to determine the value of C and γ (Foody & Mathur, 2006; Kavzoglu & Colkesen, 2009). The grid search method yielded value of 10 for C and 0.05 for γ, and those values are similar to the finding of this study from sensitivity analysis (figure 3.13 and 3.14).

4.3.4. Which algorithm is the most efficient?

Regarding the operations of ANN, RF and SVM, this study reveals that in general, RF is the most efficient and the easiest algorithm to use since the number of parameters that need to be set is the least as compared to ANN for example. Although the number of RF parameters is same as SVM, it is much easier to find and set the optimum value, and it is less time consuming through Variable Importance (VI) analysis and Out-Of-Bag (OOB) error estimation. The SVM parameters need to be set through trial and error process which may be more time consuming, and the range of parameter values to choose is more arbitrary than RF. According to the speed of training process, RF is the fastest followed by SVM and ANN respectively. The classification and training process was performed on a computer with specifications of Intel® Core™ i5-4200U (2.30 GHz), 4 Gb RAM running Windows ©7 64 bits. The averages time that the classifiers needed to finish the training and predicting processes are 6.51; 6.70; and 8.87 minutes for RF, SVM and ANN respectively. However, this time benchmarking cannot be generalized as an indicator of algorithms speed since the results might vary as different systems also have many factors that influence the training process (Pirotti et al., 2016).

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

This study has demonstrated the performance of three machine learning algorithms, i.e., Artificial Neural Network (ANN), Random Forest (RF) and Support Vector Machine (SVM) for mapping dominant tree species from Sentinel-2 imagery. The general conclusion that can be drawn from the entire process of this study is that the performance of ANN, RF and SVM in mapping dominant tree species is comparable. This shown by the McNemar's test that the map accuracy yielded by the ANN, RF and SVM statistically does not have a significant difference. To compare the performance of each algorithm and to achieve the research objectives, an investigation was carried out in six aspects: 1) compared the overall accuracy of all algorithms; 2) compared the overall accuracy from single season (summer) image with and without red-edge band; 3) compared the overall accuracy from single season image and multi-season image; 4) the capability of Sentine-2 imagery in mapping tree species was examined; 5) the efficiency of each algorithm were evaluated; 6) the quality of the training data set/tree plot samples configuration was evaluated. The specific conclusions of this study can be summarized as follow:

- In overall, SVM outperformed RF and ANN in the accuracy of classification. However, this study finds that the difference between these algorithms is not significantly different in general. Therefore, it cannot be determined which one of these algorithms is more suitable for tree species mapping. Moreover, since these algorithms follow a stochastic-based method, the selection between ANN, RF and SVM may need to consider several aspects such as input data quality, training purposes, parameter settings and characteristics of study site.

- Adding the red-edge bands improved the classification accuracy in each algorithm. The sensitivity of red-edge region to the chlorophyll concentration is useful for discriminating tree species since the chlorophyll concentration controls the optical properties of leaves at canopy level. However, the improvement to the map accuracy is not significant for all algorithms. The accuracy only improves by 1-2% in each classifier, and the McNemar's test shows that there is no a significant difference in the accuracy of each algorithm. Apparently, the red-edge did not outperform the red and NIR spectral bands where the absorption and reflection of spectral in vegetation objects occurred. Therefore, the red-edge region can be useful if it is incorporated with other spectral bands.

- The accuracy obtained from multi-season imagery are higher (by 1-8%) than single season imagery in each algorithm. It can be concluded that this study succeeded to demonstrate that combining red-edge bands from different seasons improved the accuracy. This study states that the phenological event of the tree which captured by red-edge band provides valuable information which can enhance the tree species classification. However, in general, the McNemar's test shows that there is no a significant difference between the use of single image and multi-season image since only one-third of the experiments were found to be statistically significant. The condition of the BFNP which is mainly dominated by evergreen spruce might have caused this result since the images from season to season relatively capture the same conditions. Therefore, the selection of the season imagery for addressing tree phenology in mapping tree species should consider the characteristics of the forest structure.

- The Kappa yielded by the accuracy assessment shows that Sentinel-2 has the greatest potential for tree species mapping. Although the accuracies from ANN, RF and SVM algorithms do not meet the criteria for the operational map, the Kappa showed that the classification results have a moderate agreement (0.41-0.52) and they can be considered as reliable classification result.

- The efficiency evaluation of all algorithms from two aspects (i.e., parameter settings and computational time), found that RF outperformed SVM and ANN. RF has only two parameters that need to be set. Even the RF parameters are user-defined (similar to ANN and SVM), the selection of the value for the parameters is easier to be quantified and measured through Variable Importance (VI) and Out-Of-Bag (OOB) error estimation. The computational time for the RF classification process is also lesser than SVM and ANN by 12 seconds and 140 seconds respectively. Thus, RF is the most efficient compared to the other algorithms. Although the RF's accuracy was lower than SVM and sometimes ANN, the differences were not significant. Similarly, the McNemar's test shows that statistically, the difference among the algorithms accuracy was also not significant. Further, RF is less sensitive to the overfitting than SVM and ANN.

- The criteria of 75% threshold for assigning species within a plot did not meet a satisfactory result in class separability especially in mix class and spruce-beech class and further lead to the low accuracy map. Since remote sensing imagery captured information from the uppermost canopy, this criterion creates a gap between the species composition information where the observation from the ground/below canopy has a different perspective than from above canopy. The uppermost canopy may not represent the actual composition of certain species since there is a possibility that the species below is more dominant, but they are covered by the uppermost canopy. The domination of spruce and beech within BFNP may also cause the low map accuracy since the number of less dominant species is too few then it is difficult to find them and to assign into certain classes, and therefore the number of samples for less dominant species may not sufficient for classification process.

5.2. Recommendations for future studies

As a relatively new satellite remote sensing imagery, not many studies have used Sentinel-2 for identifying the tree species. Therefore, this study was brought to demonstrate the capability of Sentinel-2 imagery for mapping the tree species by using machine learning algorithms. In general, the accuracy results can be considered low to moderate although the Kappa shows that the accuracy results are reliable. Similar results were also obtained by a study from Immitzer et al. (2016) where the study was conducted in the similar forest type and relatively the similar region characteristic. Based on these findings, the basic issues such as spectral resolution, spatial resolution and ancillary data are still relevant to be considered before deciding to use Sentinel-2 for tree species classification study (Fassnacht et al., 2016; Immitzer et al., 2016). Regarding spectral resolution, apparently Sentinel-2 still lacks of spectral bands or has less narrow wavelength even though it can be said the spectral resolution is higher than other multispectral images such as Landsat for example. The uses of narrower wavelength bands such as hyperspectral sensor were suggested, and in several prior studies, this sensor has proven to obtain higher map accuracy in tree species mapping. However, the challenging issue will be on the operational of the hyperspectral sensor, since it is still limited to the experimental circumstances and the operation may costly and not practical in terms of handling hundreds of bands (Immitzer et al., 2012).

Ancillary data can also be useful to improve the classification accuracy (Schmidt et al., 2004). There are many ancillary data which can be added to tree species classification processes such as topography, DEM, and other sensors such as LiDAR and radar (Fassnacht et al., 2016). This study also tried to incorporate the ancillary data by combining different season images. The accuracy of classification is improved but still considered low. It seems that the classification should add more different season images with denser acquisition time differences to capture more information of tree phenology (Hill et al., 2010; Sheeren et al., 2016). It is now becoming possible with the launch of Sentinel-2B recently, with higher temporal resolution of five days' acquisition when it is combined with Sentinel-2A acquisition. The other suggestion is to combine it with another sensor such as LiDAR data, which many studies have done and produced satisfactory results (Voss & Sugumaran, 2008; Dalponte et al., 2012; Alonzo et al., 2014; Ghosh et al., 2014). By using LiDAR data, the information of the tree species can be enriched by capturing both the height and the structure of trees such as canopy, branches and trunk. Further, a comprehensive information can be obtained for each species both from the spectral signature and the tree structure. Using vegetation indices such as NDVI also proved to improve the tree species classification accuracy (Schuster et al., 2012; Sheeren et al., 2016). Moreover, Sentinel-2 added three red-edge bands which may potentially be incorporated in vegetation indices besides traditional red and NIR bands, which were not used in this study since the purpose was solely optimize the Sentinel-2 spectral bands.

Wulder et al. (2009) stated that spatial resolution is arguably essential and gave the greatest impact the tree species classification. There are three types of Sentinel-2 spatial resolution (i.e., 10 m, 20 m, 60 m) depend on the bands. This study uses 10 m and 20 m, and further, the 10 m bands were resampled to 20 m to preserve the spectral information of red-edge bands. Although it is considerably higher than other multispectral data such Landsat (30 m), the spatial resolution of Sentinel-2 appears to be insufficient to classify certain classes, particularly the mix class which consist of less dominant species. The 20 m resolution still has the issue of spectral mixture where spectral of the less dominant species was covered by the dominant species spectral (Immitzer et al., 2016). Taking into consideration of the Sentinel-2 resolution, Wulder et al. (2009) proposed that this type of resolution is suitable for stand level characteristics recognition such as deciduous vs. coniferous forest type instead of species. The other way to improve the accuracy is by improving the classifier algorithms and methods. Since there are limitations of Sentinel-2 imagery resolutions both spectral and spatial, it might be unrealistic to assign pixel into single class membership through hard classification which is applied among the algorithms (ANN, RF and SVM) in this study. Thus, for future research, another approach could be taken that may tackle this issue by using fuzzy/soft-classification approach, where each pixel is associated with a class membership in various degrees (Rocchini et al., 2013). Also, it can be considered to use pan-sharpening methods by incorporating higher resolution images to downscale the spatial resolution (Wang et al., 2016).

Based on the evaluation in this study, we may not state whether Sentinel-2 is suitable or not for classifying tree species since it depends on several factors. However, the accuracies of the classification from ANN, RF and SVM algorithms show that Sentinel-2 has potential to identify the tree species if the setting of the training sample configuration and classifiers parameters are done through proper methods and techniques (Immitzer et al., 2016). Therefore, with the obtained accuracy and Kappa, we may state that the results are quite optimistic which may be improved in future to a more satisfactory output by incorporating ancillary data and other approaches.

LIST OF REFERENCES

Abdollahnedjad, A., Panagiotidis, D., Joybari, S. S., & Surový, P. (2017). Prediction of dominant forest tree species using QuickBird and environmental data. *Forests*, *8*(2), 42–60.

Adam, E., Mutanga, O., Odindi, J., & Abdel-Rahman, E. M. (2014). Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: Evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, *35*(10), 3440–3458.

Adelabu, S., Mutanga, O., Adam, E., & Cho, M. A. (2013). Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. *Journal of Applied Remote Sensing, 7*(1),1-13.

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sensing of Environment, 148*, 70–83.

Anderson, B. J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data. In USGS Professional Paper No. 964. Washington DC.

Ardö, J., Pilesjö, P., & Skidmore, A. K. (1997). Neural networks, multitemporal Landsat Thematic Mapper data and topographic data to classify forest damages in the Czech Republic. *Canadian Journal of Remote Sensing, 23*(3), 217–229.

Attarchi, S., & Gloaguen, R. (2014). Classifying complex mountainous forests with L-Band SAR and Landsat data integration: A comparison among different machine learning methods in the Hyrcanian forest. *Remote Sensing, 6*(5), 3624–3647.

Ballanti, L., Blesius, L., Hines, E., & Kruse, B. (2016). Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sensing, 8*(6), 445.

Barrett, F., McRoberts, R. E., Tomppo, E., Cienciala, E., & Waser, L. T. (2016). A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. *Remote Sensing of Environment, 174*, 279-289.

Bässler, C., Förster, B., Moning, C., & Müller, J. (2009). The BIOKLIM project: Biodiversity research between climate change and wilding in a temperate montane forest - the conceptual framework. *Waldokologie Online, 7*, 21–34.

Bässler, C., Seifert, L., & Müller, J. (2015). The BIOKLIM project in the National Park Bavarian Forest: Lessons from a biodiversity survey. *Silva Gabreta, 21*(1), 81–93.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24–31.

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman & Hall.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*, 123–140.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Burges, C. J. C. (1998). A Tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*(2), 121–167.

Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany, 88*(4), 677–684.

Chaudhary, A., Kolhe, S., & Kamal, R. (2013). Machine learning classification techniques: A comparative study. *International Journal of Advance Computer Theory Engineering, 2*(4), 2319–2526.

Cohen, J. (1960). A coefficient of agreement for nominal scale. *Educational and Psychological Measurement, 20*(1), 37–46.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*(1), 35–46.

Congalton, R. G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data: principles and practices 2nd edition*. Boca Raton: CRC Press.

Cracknell, M. J., & Reading, A. M. (2014). Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers and Geosciences, 63*, 22–33.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2012). Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of Environment, 123*, 258–270.

de Leeuw, J., Jia, H., Yang, L., Liu, X., Schmidt, K., & Skidmore, A. K. (2006). Comparing accuracy assessments to infer superiority of image classification methods. *International Journal of Remote Sensing, 27*(1), 223–232.

European Space Agency (ESA). (2015). *Sentinel-2 user handbook*. Retrieved from https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook

FAO. (2004). *National forest inventory field manual template*. Rome: Food and Agriculture Organization of the United Nations.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Modzelewska, A., Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment, 186*, 64-87.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*(1), 185–201.

Foody, G. M. (2004). Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing, 70*(5), 627–633.

Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 42*(6), 1335–1343.

Foody, G. M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment, 103*(2), 179-189.

Foody, G., Pal, M., Rocchini, D., Garzon-Lopez, C., & Bastin, L. (2016). The sensitivity of mapping methods to reference data quality: Training supervised image classifications with imperfect reference data. *ISPRS International Journal of Geo-Information, 5*(11), 199-218.

Franklin, S. E. (2001). *Remote sensing for sustainable forest management*. New York: CRC Press LLC.

Ghosh, A., Fassnacht, F. E., Joshi, P. K., & Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation, 26*(1), 49–63.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters, 27*(4), 294-300.

Guan, H., Li, J., Chapman, M., Deng, F., Ji, Z., & Yang, X. (2013). Integration of orthoimagery and LiDar data for object-based urban thematic mapping using random forests. *International Journal of Remote Sensing, 34*(14), 5166–5186.

Guyot, V., Castagneyro, B., Vialatte, A., Deconchat, M., Selvi, F., Bussotti, F., & Jactel, H. (2015). Tree diversity limits the impact of an invasive forest pest. *PLoS ONE, 10*(9), e0136469.

Heurich, M., & Englmaier, K. H. (2010). The development of tree species composition in the Rachel – Lusen region of the Bavarian Forest National Park. *Silva Gabreta, 16*(3), 165–186.

Hill, R. A., Wilson, A. K., George, M., & Hinsley, S. A. (2010). Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Applied Vegetation Science, 13*(1), 86–99.

Hoffer, R. M. (1978). Biological and Physical Considerations in Applying Computer Aided Analysis Techniques to Remote Sensor Data. In *Remote Sensing: The Quantitative Approach*. P. H. Swain and S. M. Davis (Eds) (pp. 227-289). USA: McGraw-Hill.

Horler, D. N. H., Dockray, M., Barber, J., & Barringer, A. R. (1983). Red edge measurements for remotely sensing plant chlorophyll content. *Advances in Space Research, 3*(2), 273–277.

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing, 23*(4), 725–749.

Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. *Remote Sensing, 4*(9), 2661–2693.

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing, 8*(3), 166-192.

Innes, J. L., & Koch, B. (1998). Forest biodiversity and its assessment by remote sensing. *Global Ecology & Biogeography Letters, 7*(6), 397–419.

Jensen, J. R. (2005). *Introductory digital image processing: A remote sensing perspective 3rd edition*. New Jersey: Pearson Prentice Hall.

Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation, 11*(5), 352–359.

Kavzoglu, T., & Mather, P. M. (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing, 24*(23), 4907–4938.

Khatami, R., Mountrakis, G., & Stehman, S. V. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment, 177*, 89–100.

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., Siemens, G. (2016). *Towards automated content analysis of discussion transcripts: A cognitive presence case*. Retrieved from: https://www.slideshare.net/vitomirkovanovic/towards-automated-classification-of-discussion-transcripts-a-cognitive-presence-case

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica, 31*, 249–268.

Kumar, P., Gupta, D. K., Mishra, V. N., & Prasad, R. (2015). Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data. *International Journal of Remote Sensing, 36*(6), 1604–1617.

Laborte, A. G., Maunahan, A. A., & Hijmans, R. J. (2010). Spectral signature generalization and expansion can improve the accuracy of satellite image classification. *PLoS ONE, 5*(5), 1–9.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Laurin, G. V., Puletti, N., Hawthorne, W., Liesenberg, V., Corona, P., Papale, D., Chen, Q., Valentini, R. (2016). Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. *Remote Sensing of Environment, 176*, 163–176.

Li, D., Ke, Y., Gong, H., & Li, X. (2015). Object-based urban tree species classification using bi-temporal WorldView-2 and WorldView-3 images. *Remote Sensing, 7*(12), 16917–16937.

Li, M., Im, J., & Beier, C. (2013). Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest. *GIScience & Remote Sensing, 50*(4), 361–384.

Li, X., Chen, W., Cheng, X., & Wang, L. (2016). A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. *Remote Sensing, 8*(6), 514-540.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(December), 18–22.

Lisein, J., Michez, A., Claessens, H., & Lejeune, P. (2015). Discrimination of deciduous tree species from time series of unmanned aerial system imagery. *PLoS ONE, 10*(11), 1–20.

Lohbeck, M., Bongers, F., Martinez-Ramos, M., & Poorter, L. (2016). The importance of biodiversity and dominance for multiple ecosystem functions in a human-modified tropical landscape. *Ecology, 97*(10), 2772–2779.

Manandhar, R., Odeh, I. O. A, & Ancev, T. (2009). Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement. *Remote Sensing, 1*(3), 330–344.

Mandanici, E., & Bitelli, G. (2016). Preliminary comparison of Sentinel-2 and Landsat 8 imagery for a combined use. *Remote Sensing, 8*(12), 1014.

Marsland, S. (2015). *Machine learning: An algorithmic perspective 2nd edition*. Boca Raton: CRC Press.

Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data sensed data. *International Journal of Remote Sensing, 29*(3), 617–663.

Massetti, A., Sequeira, M. M., Pupo, A., Figueiredo, A., Guiomar, N., & Gil, A. (2016). Assessing the effectiveness of RapidEye multispectral imagery for vegetation mapping in Madeira Island (Portugal). *European Journal of Remote Sensing, 49*, 643–672.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien, R package version 1.6-7, http:// CRAN.R-project.org/package=e1071

Mickelson, J. G., Civco, D. L., & Silander, J. A. (1998). Delineating forest canopy species in the Northeastern United States using multi-temporal TM imagery. *Photogrammetric Engineering & Remote Sensing, 64*(9), 891–904.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(3), 247–259.

Mutanga, O., & Skidmore, A. K. (2007). Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing, 62*(1), 34–42.

Nieddu, L., & Patrizi, G. (2000). Formal methods in pattern recognition: A review. *European Journal of Operational Research, 120*(3), 459–495.

Omer, G., Mutanga, O., Abdel-Rahman, E. M., & Adam, E. (2015). Performance of support vector machines and artifical neural network for mapping endangered tree species using WorldView-2 data in Dukuduku Forest, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(10), 4825–4840.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*(1), 217–222.

Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing, 26*(5), 1007–1011.

Périé, C., & de Blois, S. (2016). Dominant forest tree species are potentially vulnerable to climate change over large portions of their range even at high latitudes. *PeerJ, 4*, e2218.

Pirotti, F., Sunar, F., & Piragnolo, M. (2016). Benchmark of machine learning methods for classification of a Sentinel-2 Image. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XLI*, pp. 335–340.

Qian, Y., Zhou, W., Yan, J., Li, W., & Han, L. (2015). Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing, 7*(1), 153–168.

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing, 50*(1), 144–154.

Rocchini, D., Foody, G. M., Nagendra, H., Ricotta, C., Anand, M., He, K. S., Amici, V., Kleinschmit, B., Förster, M., Schmidtlein, S., Feilhauer, H., Ghisla, A., Metz, M., Neteler, M. (2013). Uncertainty in ecosystem mapping by remote sensing. *Computers and Geosciences, 50*, 128–135.

Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012a). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment, 121*, 93–107.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012b). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*(1), 93–104.

Samadzadegan, F., Hasani, H., & Partovi, T. (2005). Sensitivity analysis of support Vector Machine in Classification of Hyperspectral Imagery. *ISPRS Proceedings XXXVIII, 2005*, pp.187–192.

Schmidt, K. S., Skidmore, A. K., Kloosterman, E. H., van Oosten, H., Kumar, L., & Janssen, J. A. M. (2004). Mapping coastal vegetation using an expert system and hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing, 70*(6), 703–715.

Schulze, E. D., Aas, G., Grimm, G. W., Gossner, M. M., Walentowski, H., Ammer, C., Kühn, I., Bouriaud, O., von Gadow, K. (2016). A review on plant diversity and forest management of European beech forests. *European Journal of Forest Research, 135*(1), 51–67.

Schuster, C., Förster, M., & Kleinschmit, B. (2012). Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing, 33*(17), 5583–5599.

Sesnie, S. E., Finegan, B., Gessler, P. E., Thessler, S., Bendana, Z. R., & Smith, A. M. S. (2010). The multispectral separability of Costa Rican rainforest types with support vector machines and random forest decision trees. *International Journal of Remote Sensing, 31*(11), 2885–2909.

Shang, X., & Chisholm, L. A. (2013). Classification of Australian native forest species using Hyperspectral remote sensing and machine-learning classification algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*(6), 2481–2489.

Sheeren, D., Fauvel, M., Josipović, V., Lopes, M., Planque, C., Willm, J., & Dejoux, J. F. (2016). Tree species classification in temperate forests using Formosat-2 satellite image time series. *Remote Sensing, 8*(9), 734-762.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment, 81*(2–3), 337–354.

Skidmore, A. K., Forbes, G. W., & Carpenter, D. J. (1988). Non-parametric test of overlap in multispectral classification. *International Journal of Remote Sensing, 9*(4), 777–785.

Skidmore, A. K., Turner, B. J., Brinkhof, W., & Knowles, E. (1997). Performance of a neural network: mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering & Remote Sensing, 63*(5), 501–514.

Sommer, C., Holzwarth, S., Heiden, U., & Heurich, M. (2015). Feature-based tree species classification using hyperspectral and LiDAR data. 9th EARSeL Imaging Spectroscopy Workshop: *EARSeL eProceedings, 14* (2), 49–70.

Staufer, P., & Fischer, M. M. (1997). Spectral pattern recognition by a two-layer perceptron: effects of training set size. In *Neurocomputation in remote sensing data analysis*. I. Kanellopoulos, G. G. Wilkinson, F. Roli, & J. Austin (Eds) (pp. 105–116). London: Springer.

Stehman, S. V, & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: An application of satellite imagery. *Remote Sensing of Environment*, *64*(January), 331–344

Swain, P. H., & Hauska, H. (1977). The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics, 15*(3), 142–147.

Thampi, B., Lukashin, C., Wong, T. (2013). *Novel application of random forest method in CERES scene type classification*. Retrieved from: https://ceres.larc.nasa.gov/documents/STM/2013-10/27_Bijoy_Random_Forest.pdf

Tso, B., & Mather, P. M. (2009). *Classification methods for remotely sensed data: 2nd edition*. Boca Raton: CRC Press.

Voss, M., & Sugumaran, R. (2008). Seasonal effect on tree species classification in an urban environment using hyperspectral data, LiDAR, and an object-oriented approach. *Sensors, 8*(5), 3020–3036.

Wang, T. J., Skidmore, A. K., & Toxopeus, A. G. (2009). Improved understorey bamboo cover mapping using a novel hybrid neural network and expert system. *International Journal of Remote Sensing, 30*(4), 965–981.

Wang, Q., Shi, W., Li, Z., & Atkinson, P. M. (2016). Fusion of Sentinel-2 images. *Remote Sensing of Environment, 187*, 241–252.

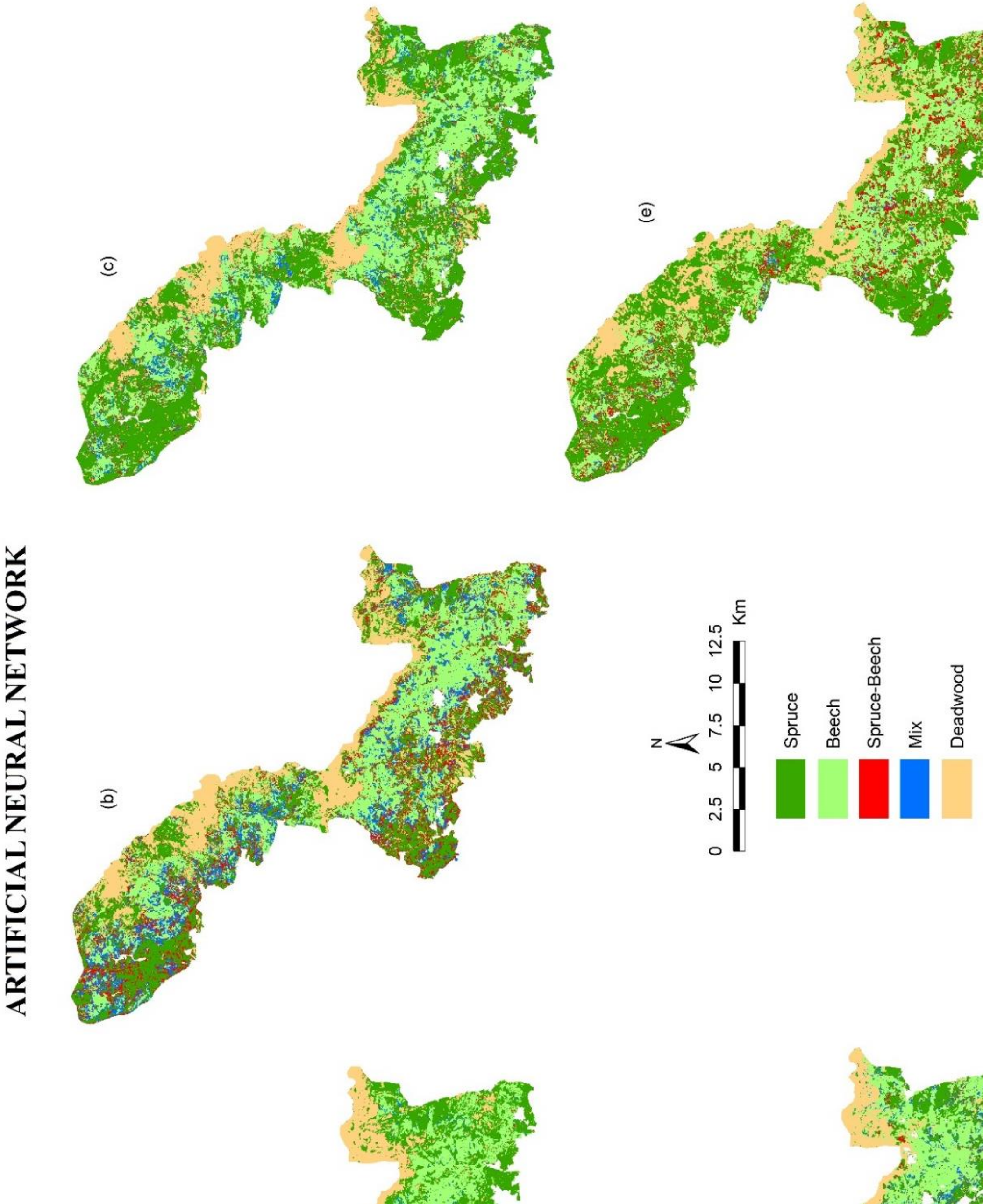
Wang, Z., Skidmore, A. K., Wang, T., Darvishzadeh, R., Heiden, U., Heurich, M., Latifi, H., Hearne, J. (2017). Canopy foliar nitrogen retrieved from airborne hyperspectral imagery by correcting for canopy structure effects. *International Journal of Applied Earth Observation and Geoinformation, 54*, 84–94.

Waser, L. T., Küchler, M., Jütte, K., & Stampfer, T. (2014). Evaluating the potential of Worldview-2 data to classify tree species and different levels of ash mortality. *Remote Sensing, 6*(5), 4515–4545.

Wulder, M. A., White, J. C., Coops, N. C., Ortlepp, S. (2009). Remote sensing for studies of vegetation condition: Theory and application. In *The SAGE handbook of remote sensing*. T.A. Warner, M. D. Nellis and G. M. Foody (Eds) (pp. 357-367). London: SAGE Publications Ltd,

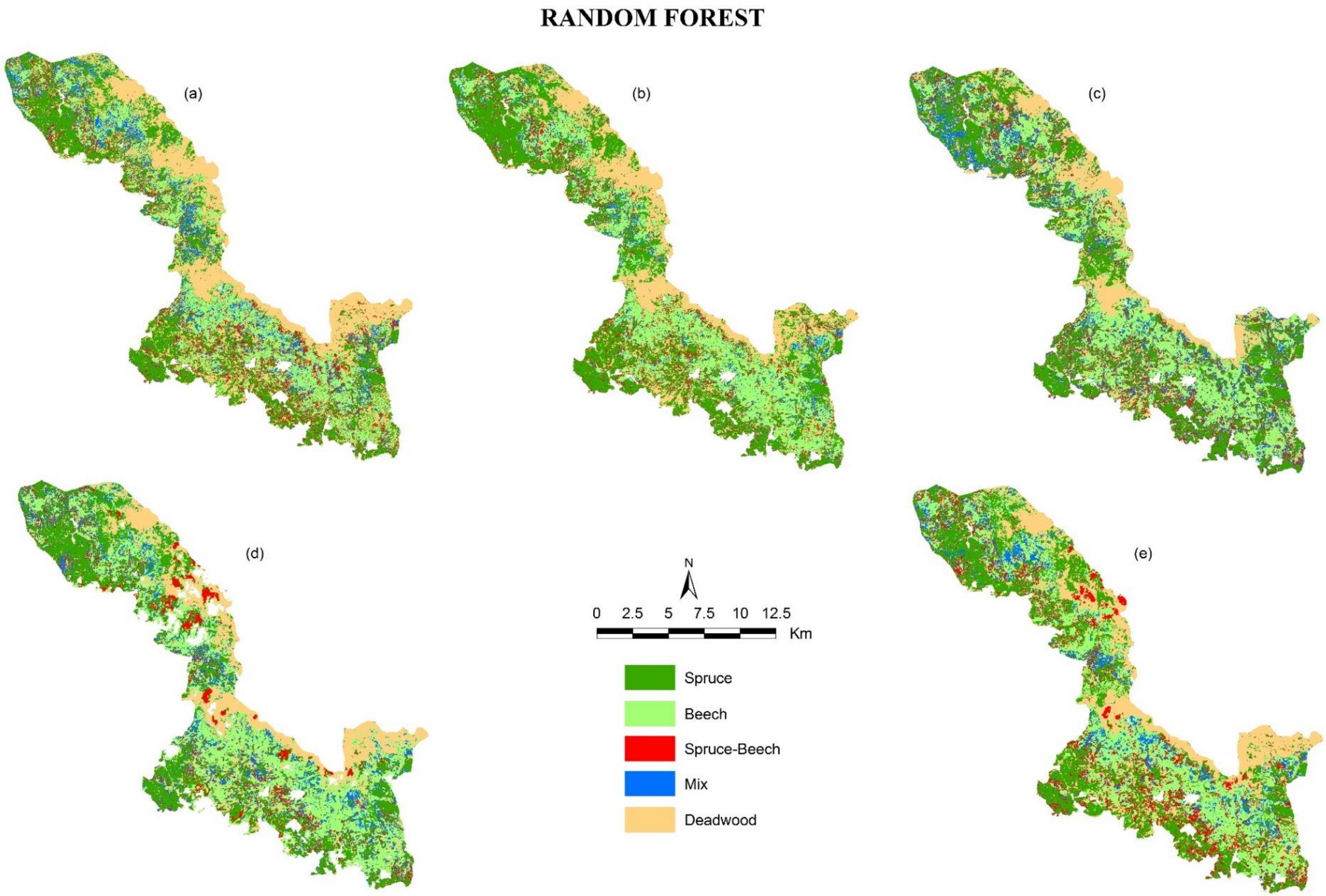
APPENDICES

Appendix 1-A



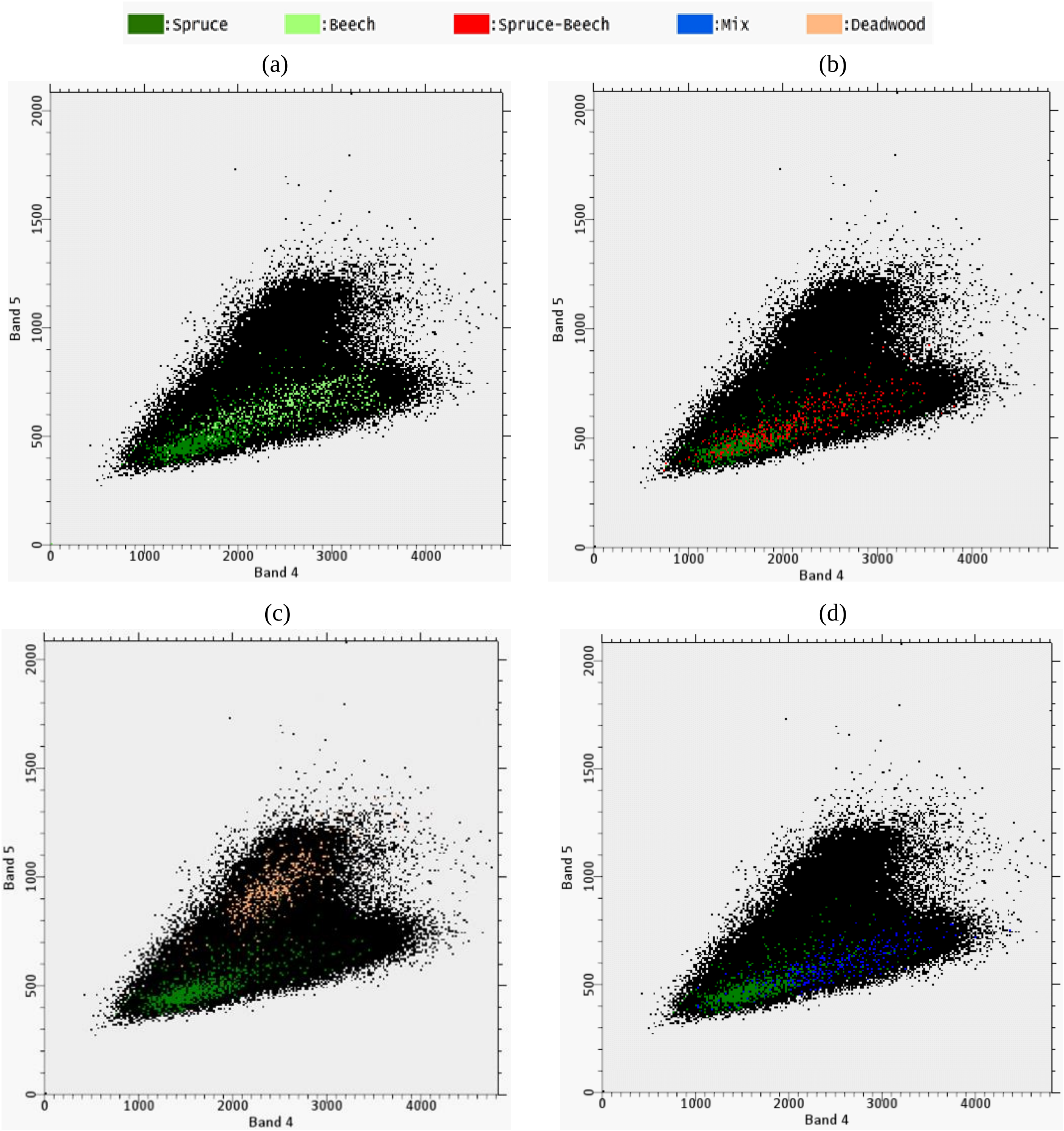
App. 1A.. Results maps from season experiments of Artificial Neural Network (ANN) classification. The maps a, b and c above represented summer, autumn and winter respectively, while d and e below represented spring and multi-season respectively

Appendix 1-B

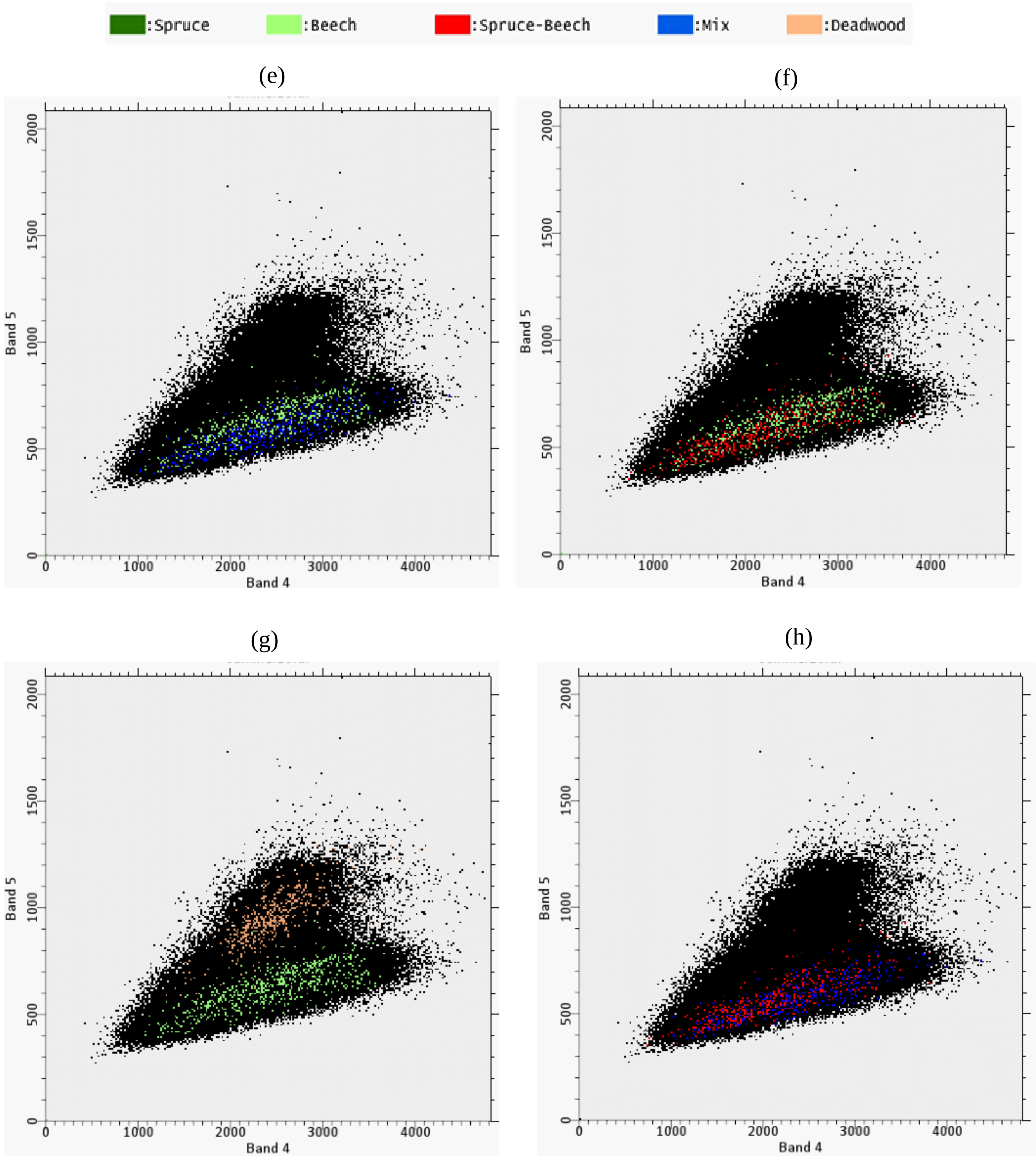


App. 1B. Results maps from season experiments of Random Forest (RF) classification. The maps a, b and c above represented summer, autumn and winter respectively, while d and e below represented spring and multi-season respectively

Appendix 2

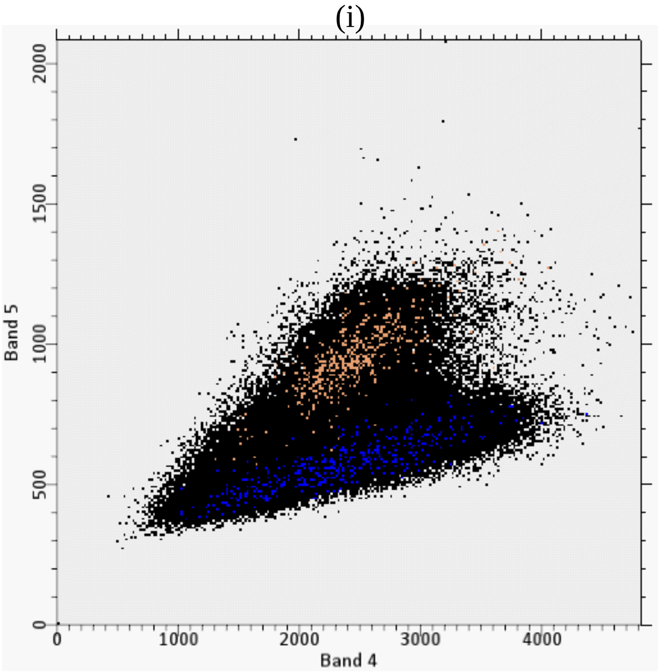


App.2. The 2-D scatter plot of Sentinel-2 band 4 vs. band 5 with different pairwise of class. (a) spruce and beech; (b) spruce and spruce-beech; (c) spruce and deadwood; (d) spruce and mix.

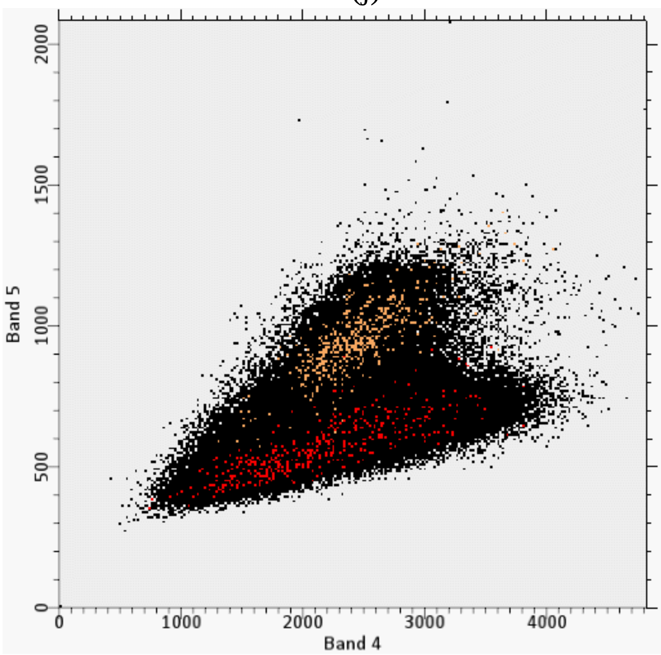


App. 2. The 2-D scatter plot of Sentinel-2 band 4 vs. band 5 with different pairwise of class. (e) beech and mix; (f) beech and spruce-beech; (g) beech and deadwood; (d) spruce-beech and mix.

(i)



(j)



App. 2. The 2-D scatter plot of Sentinel-2 band 4 vs. band 5 with different pairwise of class. (i) deadwood and mix; (j) deadwood and spruce-beech.