Machine Learning for Monitoring Forest Restoration

Mark Swan

4th Year Project Report
Computer Science
School of Informatics
University of Edinburgh

2021

Abstract

Forest restoration is the process of planting, protecting and conserving trees. The recent investment in global forest restoration initiatives has huge potential in mitigating the effects of climate change. However, at present, restoration projects are surveyed using methods which are ineffective for monitoring projects at scale. Furthermore, previous approaches to apply machine learning for forest monitoring have focused on using low-resolution satellite imagery which are not well suited for identifying small planted trees in forest restoration sites. The primary objective of this project is therefore to investigate the possibility of monitoring reforestation projects using machine learning and high-resolution satellite imagery. To achieve this research aim, a unique dataset was manually constructed using high-resolution satellite imagery. Here, we present a random forest machine learning model which is trained on this unique dataset and is evaluated on a real-life forest restoration site in northern Fiji. The results demonstrate that the proposed model, RF-Ranger, can achieving promising 96.4% classification accuracy and 96.3% Macro-F1 score on unseen pixels in a test dataset. This study was conducted in partnership with the non-profit environmental organisation, Conservation International (CI), who manage the active reforestation site assessed in this report. Overall, the results indicate that the proposed model, RF-Ranger, can successfully identify trees in forest restoration sites and can be used to monitor these projects at scale.

Acknowledgements

First, I would like to thank my supervisor, James Garforth for supporting me in this self-proposed project. His constant encouragement, advice and unwavering guidance was instrumental in helping me navigate the challenges of undertaking this undergraduate dissertation. I am most grateful for his ability to uplift my spirits during meetings and conversations, and for his thoughtful supervision through unprecedented times.

Second, I would like to thank Dr. David Milodowski, from the School of GeoSciences, for sharing his expert knowledge in remote-sensing and forest monitoring with me. From the beginning, his investment in time was instrumental in helping me access satellite imagery, and for that I am immensely grateful. This dissertation is a testament to how much I have developed, grown and learned about environmental science under his support.

A special thank you to the wonderful team at Conservation International, including Mark Ferguson (Board Member), Sebastian Troeng (Executive Vice President) and Mariano Gonzales-Roglich (Director, Ecosystem Analysis), for agreeing to partner with me in this dissertation. To the Conservation International team in Fiji, including Isaac Rounds and Eliki Senivasa, to whom I sincerely hope this dissertation can help with your current forest monitoring work - vinaka vaka levu!

Lastly, none of this work would have been possible without the fantastic encouragement and support from my family. To my girlfriend Elaine for always keeping a smile on my face. To my parents, for their continued encouragement, generous love and valued support. And finally, to my Gran, who sadly passed away of Covid-19 during writing this dissertation and who I hope I have made proud.

Table of Contents

Chapter 1

Introduction

Planting trees is crucial to protecting our planet. Trees help mitigate the effects of climate change through carbon sequestration, the natural process whereby trees absorb carbon dioxide (CO2) from the air, capturing and storing carbon, whilst releasing clean oxygen back into the atmosphere (Lal, 2008). Forests also regulate global temperatures and provide a natural home to 80% of the world's terrestrial wildlife and biodiversity (WWF, 2021). Despite the evident value of preserving forests, the current rate of global deforestation is alarming. Every minute, 36 football fields' worth of trees are lost to deforestation (CI, 2021). Every day, 137 species of life forms are driven to extinction due to logging, cattle ranching and burning of tropical rainforests (Silber and Velton, 2020). And, since 1960, nearly half of the world's rainforests have been completely destroyed (CI, 2021). Planting more trees and monitoring forest change is therefore critically important.

Over the past decade, there has been a major global push toward expanding forested regions through restoration initiatives. Active forest restoration (reforestation) is the process of planting trees in project sites which are closely managed by a specific organisation (Morrison and Lindell, 2011; Philipson et al., 2020). Most typically these projects are managed by environmental non-profit organisations, such as Conservation International or The Nature Conservancy to name a few (CI, 2021; TNC, 2021).

Active forest restoration is highly sought after because it offers a cheap, scalable and effective method to offset carbon. Recent research by Philipson et al. (2020) found that active forest restoration projects not only reduce carbon dioxide emissions but they generate higher rates of carbon accumulation than naturally regenerating forests. As well as being an effective solution for carbon storage, active reforestation is also considerably more cost-effective and easier to implement than alternative technological carbon capture solutions (Griscom et al., 2017).

As a consequence of the evident environmental and cost-effective benefits, there has been major influx of global investment in active reforestation projects. For instance, initiatives such as The Bonn Challenge are focused on restoring 350 million hectares (around 10x the size of Germany) of forests by 2030 (IUCN, 2011). Moreover, last year, the World Economic Forum launched a global initiative to grow, restore and

conserve 1 trillion trees over the next decade (WEF, 2020). To put this into perspective, this roughly equates to planting 3000 trees, every second, for the next 10 years[1].

Monitoring the success of forest restoration projects is crucial to assessing whether interventions make a difference. Forest monitoring can also hold governments, non-profit organisations and private companies to account for carbon offsetting initiatives. For instance, under the Paris Agreement, countries are now required to frequently measure and record changes in forest cover (UNFCC, 2015; Milodowski et al., 2017). However, as forest restoration projects continue to scale it becomes increasingly more challenging to monitor sites effectively. Traditional methods of reforestation monitoring, including manual on-foot field surveys, are simply not feasible for examining large projects which are often in remote or challenging natural terrains (de Almeida et al., 2020).

Recent improvements in machine learning and remote sensing can help address the challenges of large scale forest monitoring. Over the past decade, research contributions have demonstrated the ability of supervised machine learning methods, such as random forest classifiers, to detect large changes in tree cover using low-resolution (30m per pixel) satellite imagery. For instance, the leading paper offered by Hansen et al. (2013) is currently used by policy-makers and conservation charities to monitor large tree cover changes in contiguous, closed-canopy forests (Brandt and Stolle, 2020). Whilst these approaches using low-resolution imagery are appropriate for detecting substantial deforestation in large areas such as the Amazon Rainforest, they are not well suited for identifying trees in reforestation sites because the size of the planted trees in these regions are often smaller than the size of the 30m pixel resolution.

In September 2020, coinciding with the start of this dissertation project, high-resolution (4.77m per pixel) satellite imagery was publicly released for the first time from Planet Labs (Planet, 2020). Until then, high-resolution satellite imagery was restricted for private commercial access and so it could not be used publicly for forest monitoring efforts. The release of this new dataset brought with it exciting opportunities to monitor small-scale changes in tree cover in forest restoration sites for the first time and this is the area of research focused on in this dissertation.

The primary goal of this project therefore is to investigate to what extent could machine learning methods be used to identify trees in high-resolution satellite imagery and help monitor forest restoration projects. The secondary objective of this dissertation is to evaluate the suitability of newly released high-resolution PlanetScope satellite imagery for forest mapping using machine learning.

To achieve these research aims, a partnership was established with the non-profit environmental organisation, Conservation International and a supervised machine learning model was developed to help address the difficulties of monitoring a real-life forest restoration site in northern Fiji.

[1]Calculated using: 1 trillion / seconds in a decade $\rightarrow$1e12 / 3.154e+8 $\approx$ 3170 trees per second

1.1 Contributions

This self-proposed project contributed to the first in-depth analysis of monitoring reforestation projects using machine learning and high-resolution satellite imagery. In summary, the contributions of this project were the following:

- Established a research partnership with the non-profit environmental organisation, Conservation International (CI). Arranged frequent meetings with CI team members across the world: including conversations with the CI Reforestation Research Team in Washington D.C., meetings with the on-ground forest management team in Fiji and a call with CI's Executive Vice President, Sebastian Troeng in Colombia.
- Analysed previous forest monitoring approaches that used machine learning and satellite imagery. Discussed limitations in the current research and identified areas for improvement with high resolution imagery.
- Constructed a novel dataset containing examples of "trees" and "no trees" class labels in PlanetScope satellite imagery using the Azevea GroundWork image labelling tool.
- Implemented a spatial block cross validation method to prevent data leakage in the training data. This method was coded by myself and was a major contribution to the project in helping prevent the supervised machine learning models from overfitting.
- Developed four random forest models for predicting the presence of trees in satellite imagery with varying multi-spectral optical bands as input features: (1) RF-Green, (2) RF-RGB, (3) RF-RGBN+ and (4) RF-Ranger.
- Evaluated the final model on a test set, including assessing the performance on seasonal sub-sets of the test data (Summer vs Winter). The proposed model achieved 96.3% Macro-F1 score and 96.4% classification accuracy on the combined unseen test set.
- Analysed the final model performance in identifying planted trees in a forest restoration site in northern Fiji. The results demonstrated that the proposed model, RF-Ranger, can predict the presence of planted and growing trees in forest restoration sites, and achieved improvements over the current state-of-the-art model in forest restoration monitoring.

Chapter 2

Background

First, this chapter will discuss the challenges of forest restoration monitoring with reference to a real-world project managed by CI in northern Fiji. Then, this chapter will provide a background in remote sensing and machine learning which are necessary to understanding the technical details of this project. Following this, the task of forest mapping and tree detection using machine learning and remote sensing data will be discussed with reference to related literature in this research domain. In particular, this section aims to discuss the limitations with the existing literature and use this to support the high-resolution data and per-pixel machine learning methods utilised in this project.

2.1 Restoration Monitoring

2.1.1 Industry Partnership with CI

A primary goal of this dissertation was to carry out research that could help address some of the real-world challenges faced by non-profit organisations in monitoring forests. For this reason, a partnership with the non-profit environmental organisation Conservation International (CI) was established. In particular, the objective of this project was to develop a supervised machine learning model using high-resolution PlanetScope satellite imagery to help monitor a CI forest restoration site in northern Fiji, known locally as the Nakauvadra Reforestation Project.

2.1.1.1 Nakauvadra Reforestation Site

The Nakauvadra Community Based Reforestation Project in Fiji was identified and developed by CI to increase forest cover and to expand critical habitat for native endangered species. The project is comprised of 27 reforestation sites, covering a total tree planting area of 1,135 hectares (ha) across the northern region of the Vitu Levu mainland island. The reforestation sites are located along the southern and northern slopes of the Nakauvadra forest mountain range, an area covering 11,387 hectares (ha) (Figure 2.1). This is roughly equivalent to the land area of the City of Edinburgh (see Appendix B.2).

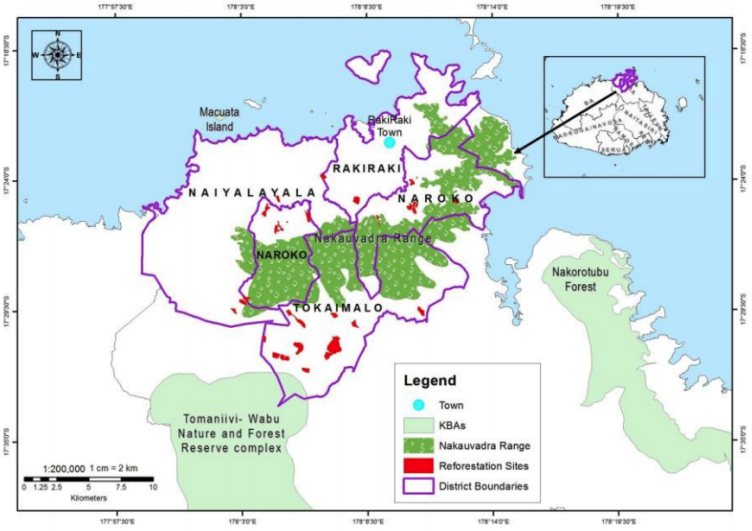


Figure 2.1: Map showing the geographical location of the Nakauvadra Range and distinct boundaries (Image sourced from private document shared by CI under an NDA agreement so original document cannot be directly cited or shared)

The Nakauvadra site was chosen for this study because it is one of the largest, oldest and most successful restoration projects managed by CI. Moreover, the site was selected because there was many technical and environmental challenges to effective forest monitoring in this region: (i) the site is currently monitored by a team of forest rangers who manually survey the reforestation projects using on-foot land surveys, (ii) it is a mountainous region, therefore it is extremely difficult to manually measure tree-regrowth using on-foot land surveys, and finally, (iii) it is a region frequently hit by natural disaster events such as cyclones, therefore frequent forest monitoring efforts are required to assess the potential damage.

For all of these reasons outlined above, the Nakauvadra reforestation site was chosen because machine learning and remote sensing are well suited technologies to help address the challenges of forest monitoring in this area.

2.2 Technical Primer

2.2.1 Remote Sensing

Remote sensing is to ability learn information about an object from a distance. This is achieved by using remote sensors, most commonly land observation satellites or aircrafts, to measure how much light is reflected from earth. More specifically, remote satellite sensors do not just detect light but they measure the reflectance of electromagnetic (EM) radiation for a particular location on earth.

As defined in Campbell and Wynne (2011), remote sensing is “the practice of deriving information about the Earth’s land and water surfaces using images acquired from an overhead perspective, using electromagnetic radiation in one or more regions of the

electromagnetic spectrum, reflected or emitted from the Earth's surface" (Campbell and Wynne, 2011, p. 6).

Interpretation of remote sensing imagery therefore relies on a comprehensive understanding of how particular objects on earth reflect EM radiation (Campbell and Wynne, 2011). For this reason, the EM spectrum and specifically its relationship with multispectral satellite imagery is described in detail in the following sub-sections.

2.2.1.1 The Electromagnetic Spectrum

The electromagnetic (EM) spectrum covers the range of wavelengths that represent radiation (Crockett, 2019). From low-frequency radio waves to high-frequency cosmic rays, radiation emitted or reflected from earth can be quantitatively measured and categorised on the EM spectrum depending on it's wavelength distance (λ). Visible light, what humans recognise as colour reflecting from objects, represents only a small subset of the EM spectrum (see Figure 2.2).

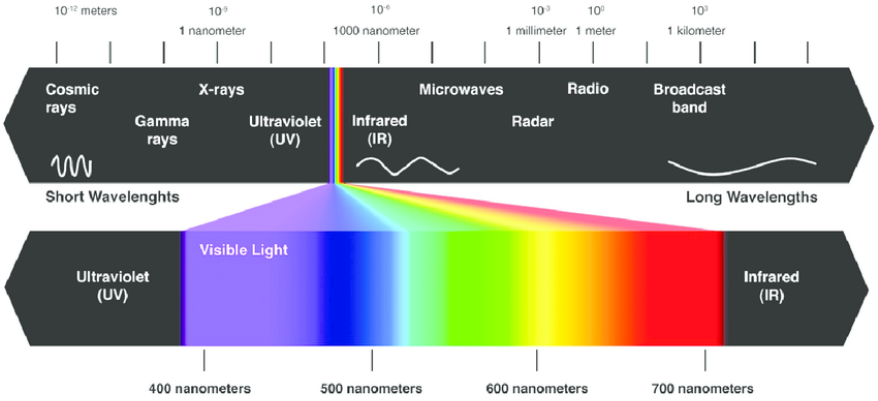


Figure 2.2: Diagram representing the spectral band of visible light in the electromagnetic spectrum. Source: (Mishra et al., 2013).

All objects on earth emit and reflect EM radiation. So much so, that we can learn a lot about an object by looking at what types of EM radiation reflect from it. For instance, we know the colour of an object by considering which visible light reflects from it. Similarly, infrared radiation reveals how much heat is emitted from the object. This is why infrared cameras are used to detect if someone has a temperature (Chan et al., 2004).

An interesting question we might ask ourselves is "why are trees green?". Trees are green because they *reflect* green light and mostly absorb other types of EM radiation (Figure 2.3). Similarly, water is perceived to be blue to humans because it reflects blue light. How much light an object sends back to our eyes as humans is known as **surface reflectance**. The greater the surface reflectance value, the more intense we perceive the object showing that particular subset of EM radiation.

Over the past two decades, EM radiation has been used extensively in research to distinguish objects from space using remote sensors. For instance, recent research

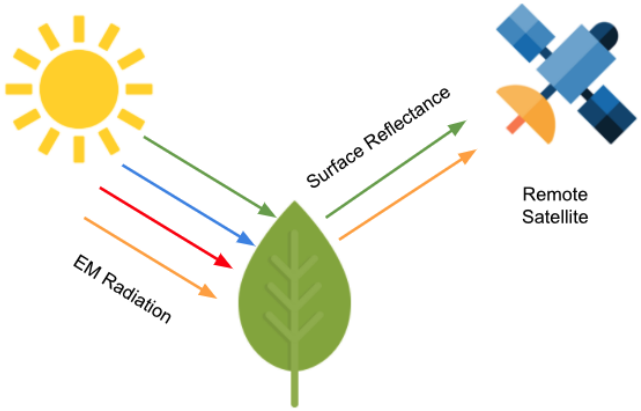


Figure 2.3: Diagram representing the surface reflectance of EM radiation in vegetation.

classified different species of vegetation using only EM radiation (Hennessy et al., 2020). The assumption goes that, the wider the spectrum of EM radiation, the easier it is to distinguish objects from space. For this reason, in this project it was optimal to use **multi-spectral** satellite imagery for analysis. This was preferable because multi-spectral imagery collects both visible {Red, Green and Blue} and non-visible {Near Infrared (NIR)} EM radiation.

2.2.1.2 Multi-spectral Satellite Imagery

Images can be thought of as a grid of pixels. In satellite imagery, each pixel measures the surface reflectance of EM radiation for a particular geolocation on earth. The **spatial resolution** of each pixel represents the physical square meter distance on earth covered by that single pixel. For instance, the spatial resolution of **low-resolution** imagery is often $30m^2$ (Landsat) or $250\text{-}1000m^2$ (MODIS). By contrast the spatial resolution of **high-resolution** covers a significantly smaller area, often $<5m^2$ (PlanetScope, WorldView-2). The **temporal resolution** refers to the period frequency between satellite land observation images, e.g., 1-day, 6-months, 1-year.

Unlike images that you take on your mobile phone, where pixels represent a composite of visible light (Red, Green and Blue), each pixel in multi-spectral satellite imagery contains numerous surface reflectance **bands** which represent both visible and non-visible EM radiation. For instance, the PlanetScope Surface Reflectance multi-spectral imagery used later in this report measures the following four spectral bands: Red, Green, Blue and Near-Infrared (NIR). Each pixel in this imagery can therefore be thought of as a (1x4) array where each element represents the surface reflectance values for each spectral band {blue, green, red, NIR} (see Figure 2.4).

Looking at these four surface reflectance values for each pixel can provide a wealth of information about what objects are contained within that area on earth. Our hypothesis is that pixels containing trees will return different surface reflectance band values than pixels containing no trees (buildings, roads, agricultural farming land and waterbodies). Our aim therefore is to train a supervised machine learning model to predict

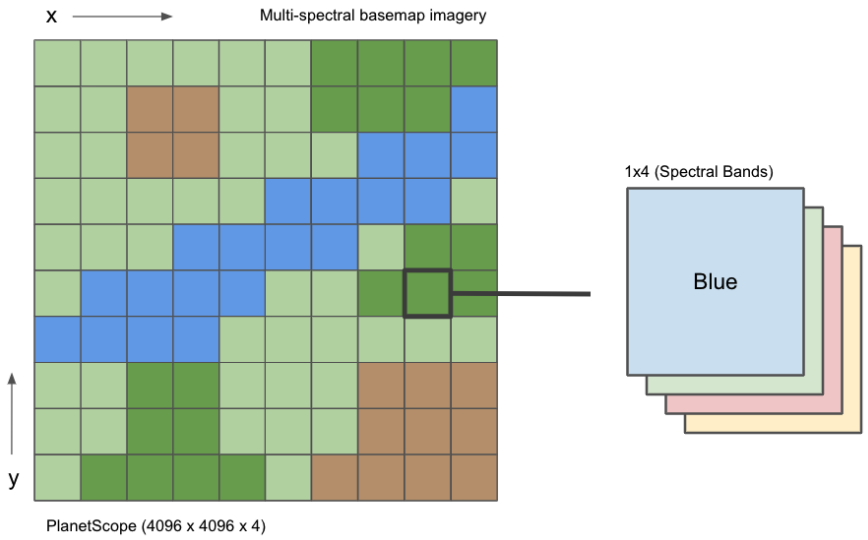


Figure 2.4: Each pixel in multi-spectral satellite imagery contains surface reflectance values for multiple spectral bands - e.g., Red, Green, Blue and NIR.

the per-pixel presence of trees within multi-spectral satellite imagery using only these spectral band values as input features.

2.2.2 Machine Learning

2.2.2.1 Supervised Learning

A machine learning algorithm is an algorithm that can learn from input data (Goodfellow et al., 2016). This learning can be achieved broadly through two methods: supervised or unsupervised learning (Caruana and Niculescu-Mizil, 2006). Supervised learning is the process whereby a machine learning algorithm learns patterns from labelled training data and uses this information to make reasoned predictions about future unseen data. For instance, the task of supervised learning is outlined by Russell and Norvig (2009) as follows.

Definition 2.2.1 (Supervised Learning)**.** Given a training set of **N** example input-output pairs $(x_1, y_1), (x_2, y_2), ...(x_N, y_N)$, where each y_j was generated by an unknown function $y = f(x)$, discover a function h that approximates the true function f. (Russell and Norvig, 2009, p. 695).

Common supervised learning algorithms include: logistic regression (Wright, 1995), random forests (Breiman, 2001), support vector machines (Drucker et al., 1997) and artificial neural networks (Yegnanarayana, 2009). For the purposes of this dissertation, our objective is to train a supervised machine learning algorithm on labelled satellite images that contain examples of “tree” and “no tree” classes. The model will then be able to predict the class of each pixel in the image by considering only the surface reflectance values of those pixels.

2.2.2.2 Classification

Classification relates to the task of using a machine learning model to predict k distinct class labels from the problem domain (Brownlee, 2020). Classification is defined in Goodfellow et al. (2016) as the following:

Definition 2.2.2 (Classification)**.** Given a vector representing input data $\mathbf{x}$, the learning algorithm produces a mapping function $f : \mathbb{R}^n \rightarrow \{1...k\}$, where $y = f(\mathbf{x})$ (Goodfellow et al., 2016, p. 98).

Here, we present a binary classification task because we have only two labels "tree" or "no tree". These predictions are made on a per-pixel basis, meaning that each pixel in the multi-spectral imagery is classified depending on the surface reflectance value of that pixel.

As discussed extensively in Section 2.3, per-pixel classification methods have proven effective for predicting the presence of trees within low-resolution and medium-resolution satellite imagery (Hansen et al., 2013; Ottosen et al., 2020; Bolyn et al., 2018). We therefore expect the success to extend to high-resolution imagery and this is one of the major research questions we want to investigate in this project.

2.2.2.3 Overfitting

A common problem in supervised machine learning is the issue of overfitting the model to the training data. This occurs when the machine learning model learns noise in the training data to the extent of negatively impacting the generalisation performance of the model on new data (Brownlee, 2019). Overfitting becomes more likely as the number of input features increases and less likely as the number of training examples increases (Russell and Norvig, 2009).

Overfitting and data leakage is especially common in machine learning models trained on ecological data with inherent temporal, spatial and hierarchical structures (Roberts et al., 2017). For satellite imagery, this means that objects in pixels typically share similar characteristics to objects in neighbouring pixels. However, often these inherent similarities are ignored when training models, resulting in overestimation of the predictive power of the models when random validation methods are adopted (Roberts et al., 2017).

Consider a hypothetical example where training and testing sets of labelled satellite imagery pixels are randomly assigned to train and evaluate a supervised machine learning model. Neighbouring pixels in satellite imagery are likely to share similar ecological dependency structures. For instance, if pixel at position (i, j) contains a tree, then pixel at $(i+1, j+1)$ is also likely to contain a tree with similar surface reflectance values. Therefore, if pixels in the test set contain neighbouring pixels in the training set, then we would expect the model to dramatically overfit. To solve this issue, research from Roberts et al. (2017) recommends using a spatial block validation method to prevent such data leakage from occuring. As discussed in detail later in this report, spatial block validation is the approach we follow in this dissertation.

2.3 Previous Work

In recent years, there have been various attempts to apply machine learning to detect the presence of trees in medium-resolution or low-resolution satellite imagery. These papers have revolutionised how we detect large-scale forest changes, such as detecting mass deforestation in the Amazon Rainforest or quantitatively measuring primary forest loss over the past decade (GFW, 2021). Having said this however, there is limited research on machine learning applied to publicly available high-resolution imagery. Also, to the best of my knowledge and research, no work has been done specifically for the use of identifying planted trees in active forest restoration sites.

In this section we discuss the previous research conducted in this space. Specifically we aim to evaluate the advantages and limitations of previous papers, using this information to motivate the methods and data utilised later in this report.

2.3.1 Why Per-Pixel Random Forest?

2.3.1.1 Per-Pixel vs Deep Neural Networks

Previous attempts to monitor forest cover with satellite imagery have focused on two main machine learning approaches: per-pixel classifiers (Hansen et al., 2013; Bolyn et al., 2018; Radoux et al., 2016; Crouzeilles et al., 2020) or deep neural network classifiers (Brandt and Stolle, 2020; Ulmas and Liiv, 2020). Here, we decided to follow a per-pixel classifier method because this was a technical requirement imposed by the forest monitoring research team at CI.

For a model to be used by CI in future projects, the research team would need to know how to re-purpose the model to new geographical locations. However, training deep learning models would require GPU hardware dependencies that are expensive and are not typically accessible to non-profit environmental organisations including CI (Zhu et al., 2017). By contrast, per-pixel approaches, such as random forest classifiers, require limited technical overheads and are used extensively across the earth science research community, including being previously used by researchers at the Conservation International Moore Centre for Science [1]. For these reasons, a per-pixel method was chosen.

There is also significant merit in having a simplified pipeline which per-pixel machine learning methods have to offer. It is true that deep neural networks have received significant attention in recent years, achieving new state-of-the-art results in complex tasks, such as image recognition (Krizhevsky et al., 2012), speech detection (Deng and Platt, 2014) and time-series forecasting (Fawaz et al., 2019). However, all of these papers involve using deep learning to discover complex patterns within high-dimensional data and large datasets (LeCun et al., 2015). This process can be highly computationally expensive and simply is not necessary for tasks with low-dimensional input data. Given the dataset used in this project contains multi-spectral band values $\{$Red, Green, Blue, NIR$\}$ and vegetation indices $\{$NDVI, NDWI, MSAVI2, BI$\}$ (a total

[1] I was informed of this during meetings with the CI Reforestation Research Team who are based in Washington D.C.

input-dimension of [1x8]), it therefore could be considered unnecessary to apply deep learning in this case. As a consequence, a per-pixel classification method was prefered.

2.3.1.2 Random Forest Per-Pixel Classification

Various per-pixel machine learning classification methods have been proposed for the task of forest monitoring using multi-spectral imagery, including using algorithms such as random forest classfiers (Immitzer et al., 2016; Bolyn et al., 2018; Hościło and Lewandowska, 2019; Crouzeilles et al., 2020) and support vector machines (SVMs) (Huang et al., 2002; Mountrakis et al., 2011). Here, we develop a per-pixel random forest classification model because they have been shown to outperform other per-pixel methods in remote sensing tasks.

Results from a meta-review of machine learning classification methods using remote sensing data, found that random forests algorithms consistently outperform other per-pixel classification methods (Lawrence and Moran, 2015). Specifically, in an evaluation across 30 multi-spectral datasets, Lawrence and Moran (2015) show that random forest machine learning classifiers obtained the highest average classification accuracy 73.19%, achieving notably better results than other per-pixel methods, such as SVMs (62.28%) and Logistic Model Trees (64.82%) (Lawrence and Moran, 2015).

One explanation for why random forest classifiers outperform other per-pixel methods is because of their ability to generalise and reduce overfitting. Random forests are an example of an ensemble machine learning algorithm, where multiple base models are combined to produce a single optimal prediction. For instance, they work by combining several decision tree classifiers, thereby creating a "forest", whose results are then aggregated to predict a single decision in a process known as ensemble learning (Ho, 1995) (see Section 4.1.2). Recent research from Maxwell et al. (2018) found strong evidence that ensemble machine learning methods are more effective than models that use a single classifier in remote sensing classification tasks.

For all the reasons outlined above, we decided to focus solely on implementing per-pixel random forest models for the task of identifying trees in satellite imagery. In this project, we therefore attempt to "predict natural forests with random forests".

2.3.2 Existing Solutions

Recent contributions have demonstrated the ability of detecting the presence of large areas of trees using random forest models and coarse resolution satellite imagery. These attempts have been used largely for mapping tree cover in contiguous, closed-canopy forests. In the following subsection we outline some of the major research in this space.

2.3.2.1 Global Forest Watch

The leading paper in per-pixel **global** forest monitoring is offered by Hansen et al. (2013), who developed a machine learning random forest method for mapping global

tree cover using low-resolution (30m per pixel) Landsat satellite imagery. The paper received significant worldwide attention and it is currently the model used to inform policy makers, conservation organisations and private companies of forest cover changes through an online platform called Global Forest Watch (GFW) (Global Forest Watch, 2002). Although the global per-pixel method proposed by Hansen et al. (2013) achieves promising results in detecting changes in large closed-canopy forests, their data is not suitable for monitoring changes in tree growth in small-scale active forest restoration projects. This is because the crown width of planted trees are typically smaller than the coarse 30m pixel resolution of Landsat imagery, meaning small trees are not detected by the classification model (Brandt and Stolle, 2020). Furthermore, recent research has found numerous issues with Hansen et al. (2013), such as examples of where the model underestimates tree cover in settings outside closed-canopy forests and is not accurate in regions with heterogeneous biodiversity such as urban or semi-urban environments (Brandt and Stolle, 2020; Ottosen et al., 2020; Milodowski et al., 2017).

2.3.2.2 Regional Models

To address these challenges, there have been various follow-on efforts to detect changes in tree cover at **regional** scales with both low-resolution and medium-resolution imagery. These contributions have continued to use random forests machine learning classifiers (Immitzer et al., 2016; Bolyn et al., 2018; Hościło and Lewandowska, 2019; Crouzeilles et al., 2020). For instance, Crouzeilles et al. (2020) mapped natural forest regeneration in the Brazilian Atlantic Forest region with low-resolution (30 meter) data and a random forest classifier, achieving 76.9% accuracy in detecting the presence or absence of natural tree re-growth. Comparably, Immitzer et al. (2016) developed a random forest model with 65% accuracy in classifying different tree species in Germany with medium-resolution (10 meter) Sentinel-2 satellite imagery. Lastly, Bolyn et al. (2018) achieved over 90% accuracy with a binary classification per-pixel random forest model, distinguishing between forests and non-forests in Belgium with medium-resolution (10 meter) Sentinel-2 imagery. However, whilst the results outlined above provide some preliminary evidence that regional-based, per-pixel random forest models achieve promising results in detecting trees in satellite imagery, they still all suffer from the same problem as Hansen et al. (2013) in using low-resolution and medium-resolution imagery. For these reasons, these regional models still do not address the difficulty of monitoring small planted trees within active forest restoration sites.

Chapter 3

Data

A vast quantity of labelled satellite data was required to train a supervised machine learning model to predict the presence of trees in satellite imagery. Unfortunately, to the best of my knowledge and research, no labelled satellite imagery existed for the Nakauvadra forest restoration area studied in this project. As a result, a unique dataset had to be manually constructed.

To create this dataset, we first downloaded **unlabelled** high-resolution PlanetScope satellite imagery. Then we **manually labelled** examples of "tree" and "no tree" classes within these satellite images. The full dataset curation process was implemented, developed and validated solely by myself. Full details of collection, labelling and dataset analysis are discussed extensively throughout this chapter.

3.1 PlanetScope Satellite Imagery

Until only very recently, high-resolution satellite imagery has been restricted for private commercial use only. This has significantly limited the research on tree classification tasks using high-resolution satellite imagery and it has prevented non-profit organisations from using it to help with forest monitoring efforts.

In September 2020, coinciding with the start of this dissertation, Norway's International Climate and Forests Initiative (NICFI) partnered with Planet Labs, a commercial satellite imagery company, to provide free universal access to high-resolution satellite imagery for the purposes of forest monitoring efforts in global tropical regions (Planet, 2020). This dataset, referred to as PlanetScope, provides access to basemaps representing 4.77m per pixel resolution imagery. The release included an archive of bi-annual (June and December) satellite imagery basemaps dating back to 2015. Given the significant value of this newly released dataset and the potential to help with identifying small trees in forest restoration sites, PlanetScope satellite imagery was chosen for this project.

3.2 Collection

Two PlanetScope Surface Reflectance satellite images were downloaded using the Basemap Viewer online tool provided by Planet (see Appendix B.1). Surface Reflactance Basemaps are a special type of high-resolution satellite imagery available through Planet. Basemaps collate multiple satellite images within a given time-period (1-month) and generate optimised, cloud-free and color-balanced satellite images which are better suited for multi-spectral analysis. These basemaps are unlabelled but each pixel contains surface reflectance values {Red, Green, Blue, NIR}.

Planet downloads satellite basemaps as a GeoTiff (.tiff) file, which is a special type of format specifically used for geo-referenced multi-spectral imagery. As shown in Figure 3.1, a single satellite basemap represents a (4096 x 4096 x 4) multi-dimensional array of pixels, where each dimension represents the four spectral band values {Red, Green, Blue, NIR} that compose each pixel. Another way of representing the data is to think of satellite imagery as a table, where each row represents a pixel and the columns are spectral band values (Figure 3.1). Ultimately, this tabular format is what we use to train the random forest machine learning model.

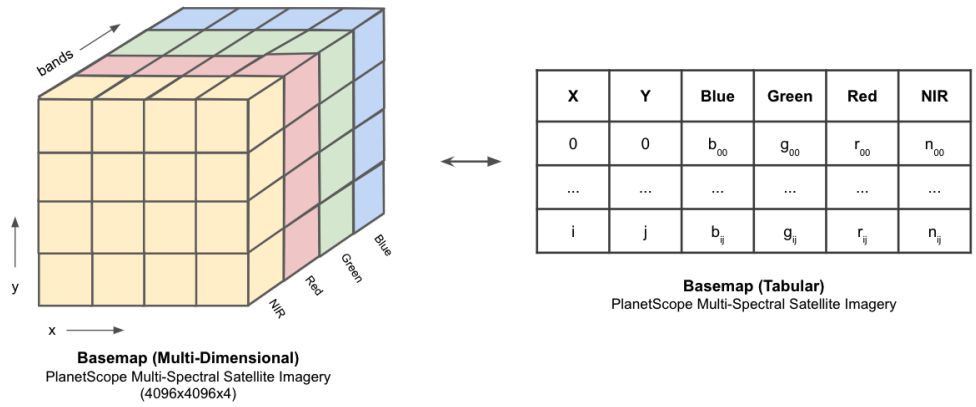
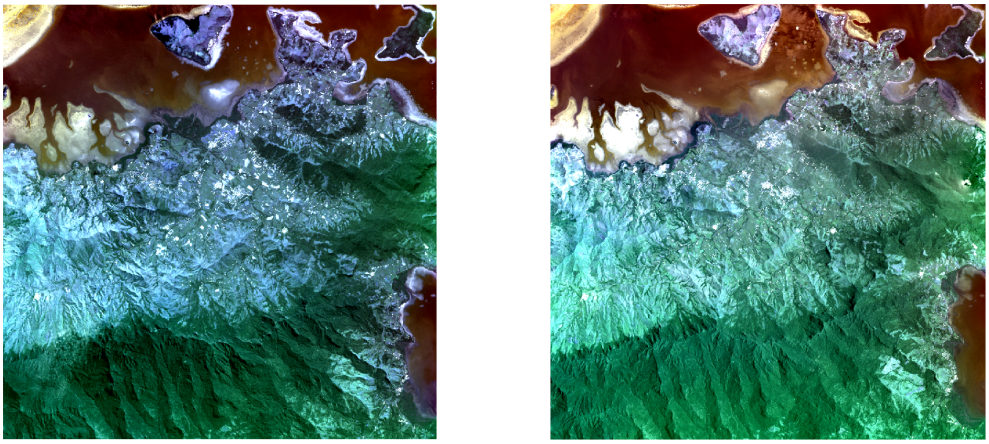


Figure 3.1: PlanetScope Multi-Spectral Satellite Imagery can be represented as a (4096x4096x4) multi-dimensional array (left) or as 2-D table where each column represents the per-pixel spectral band values (right).

The center of every pixel is assigned a unique geospatial coordinate - longitude (x) and latitude (y) - referencing a specific location on earth. Each basemap covers a total land area of approximately 20,480m^2 across northern Fiji. To account for seasonal variations in surface reflectance values between wet and dry seasons, two satellite basemaps from distinct time periods were downloaded - June 2019 and December 2019 (Figure 3.2).

At this stage the downloaded basemaps were unlabelled. Further work was therefore required to label training examples of “tree” and “no tree” classes within these images.



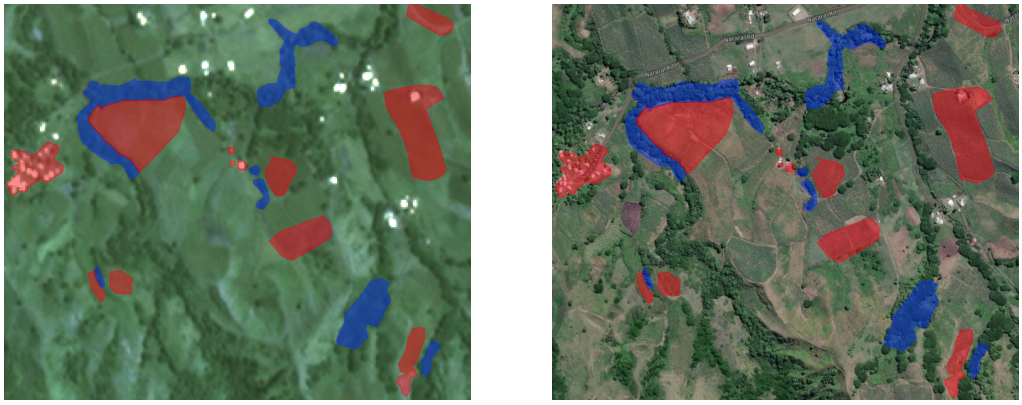
(a) June 2019 (b) December 2019

Figure 3.2: PlanetScope Surface Reflectance basemaps for Nakauvadra Forest Range in June and December 2019.

3.3 Labelling

All polygons were labelled and validated manually using GroundWork, an open-source labelling tool for creating custom datasets from satellite imagery (Azavea, 2020).

GroundWork was chosen because of the following reasons: (i) it was specifically tailored for labelling satellite imagery and was compatible with GeoTiff input file formats, (ii) the free-tier GroundWork package provided up to 10GB of free cloud storage which was suitable for the project, and finally, (iii) GroundWork provided unique free-access to commercial super-high-resolution (0.3m per pixel) optical imagery. It is important to note that this commercial imagery was not downloaded but was instead used to visually validate the polygon labels on the existing PlanetScope high-resolution (4.77m per pixel) imagery (see Figure 3.3).



(a) PlanetScope Resolution (4.77m per pixel) (b) GroundWork Resolution (0.3m per pixel)

Figure 3.3: Example polygons of classes “tree” (blue) and “no tree” (red), manually identified through visual interpretation and drawn using the GroundWork labelling tool. Best practices recommend labelling small polygons that are spatially distributed as opposed to labelling the full satellite image.

Data labelling best practices were followed to reduce the likelihood of bias, inaccuracies and human-error when labelling satellite imagery using visual interpretation. First, a similar class balance was maintained between tree and no tree polygon labels. This class balance was preserved between training, validation and test datasets (Table 3.1). This sampling method was chosen because it has been shown address the problem of superficial model performance in binary classification tasks (He and Garcia, 2009). Second, classes were labelled using satellite imagery from two time-periods (June 2019, December 2019), thereby accounting for seasonal changes in vegetation cover between wet and dry seasons in Fiji (Xie et al., 2008). Our hypothesis was that satellite imagery surface reflectance values will be highly variable to seasonal conditions, therefore polygons from satellite imagery in different seasons were collected for training, validation and testing. All polygons were sampled semi-randomly, using a randomised grid selection method, thereby reducing spatial-decadency bias in the training data (Roberts et al., 2017). In addition, attempts were made to directly label class examples affected by shade cover and surface reflectance distortion which are common issues with satellite imagery in mountainous regions with steep gradients (Thompson et al., 2019). Finally, attempts were made to label small polygons representing the broad distribution of "tree" and "no tree" class examples within the Nakauvadra forest range. For instance, the "no tree" class includes a variety of examples: buildings, roads, rivers, grasslands, agricultural farming, croplands and waterbodies (oceans). Similarly, the "tree" class contains examples of both natural forests and planted trees within active forest restoration sites.

Using the procedure above, a total of 890 polygons (429 polygons for June 2019 and 461 polygons for December 2019) were manually labelled. These geo-referenced polygons were then downloaded as a GeoJSON file format, an open standard format designed for representing geographical features. The GeoJSON file contains each polygon as a dictionary of longitude and latitude pairs with a corresponding class label value "tree" and "no tree". However, the GeoJSON does not contain the multi-spectral values corresponding to each longitude and latitude value at this stage. To complete the data curation process, the two datasets had to be merged (1) long-lat GeoTIFF basemap with multi-spectral values and (2) long-lat GeoJSON with corresponding class labels.

3.4 Preprocessing

3.4.1 Raster Cropping

To achieve the final tabular dataset, multi-spectral values for each labelled polygon were collected using a procedure called **raster cropping**. Raster cropping is the process of returning a subset of pixels (in the unlabelled PlanetScope basemap) that are covered by the labelled GeoJSON polygons (see Figure 3.4). This raster cropping procedure was implemented by myself and a Python program was written using a number of geo-spatial data science packages.

First, the Python package Rasterio was used to store each unlabelled satellite imagery basemap as a pixelated "raster" image. This converted each basemap into a multi-dimensional array or tabluar format of multi-spectral values (see Figure 3.1). Sec-

ond, the Python package GeoPandas was used to store the labelled GeoJSON file as a dataframe structure. Then, by iterating through each polygon in the GeoPandas dataframe, the original basemap was cropped and the resulting tabular output was appended to a finalised dataset - train, validation or test.

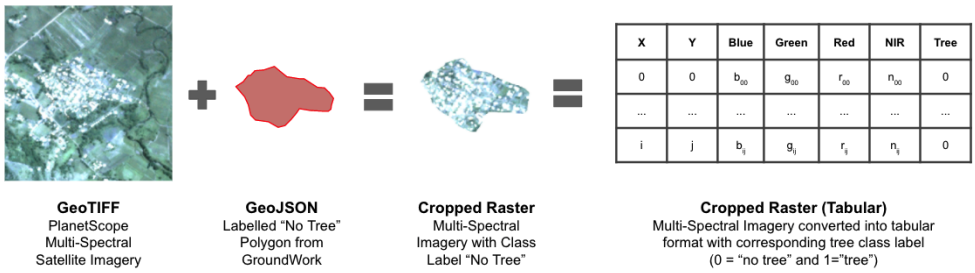


Figure 3.4: Diagram illustrating the raster cropping procedure.

3.4.2 Spatial-Blocking

As discussed briefly in Chapter 2, overfitting as a result of data leakage is common in ecological datasets with inherent temporal, spatial and hierarchical structures (Roberts et al., 2017). For instance, in the dataset allocation described above, it would be erroneous to randomly assign labelled pixels to train, validation and test data. This is because random data allocation ignores the spatial correlation between neighbouring pixels. A random split would lead to the test and validation sets containing neighbouring pixels in the training set. In this case, the model would have already "seen" similar pixels during training and so we would expect the model to overfit. To address this major issue, as recommended in Roberts et al. (2017) a spatial-blocking method was utilised to allocate train, validation and test datasets. This was achieved by assigning polygons, not pixels, to each dataset. Furthermore, during the data labelling process it was ensured that no polygons overlapped. As a result, the spatial-blocking method developed in this report helped effectively prevent neighbouring pixels from "leaking" between train, validation and test.

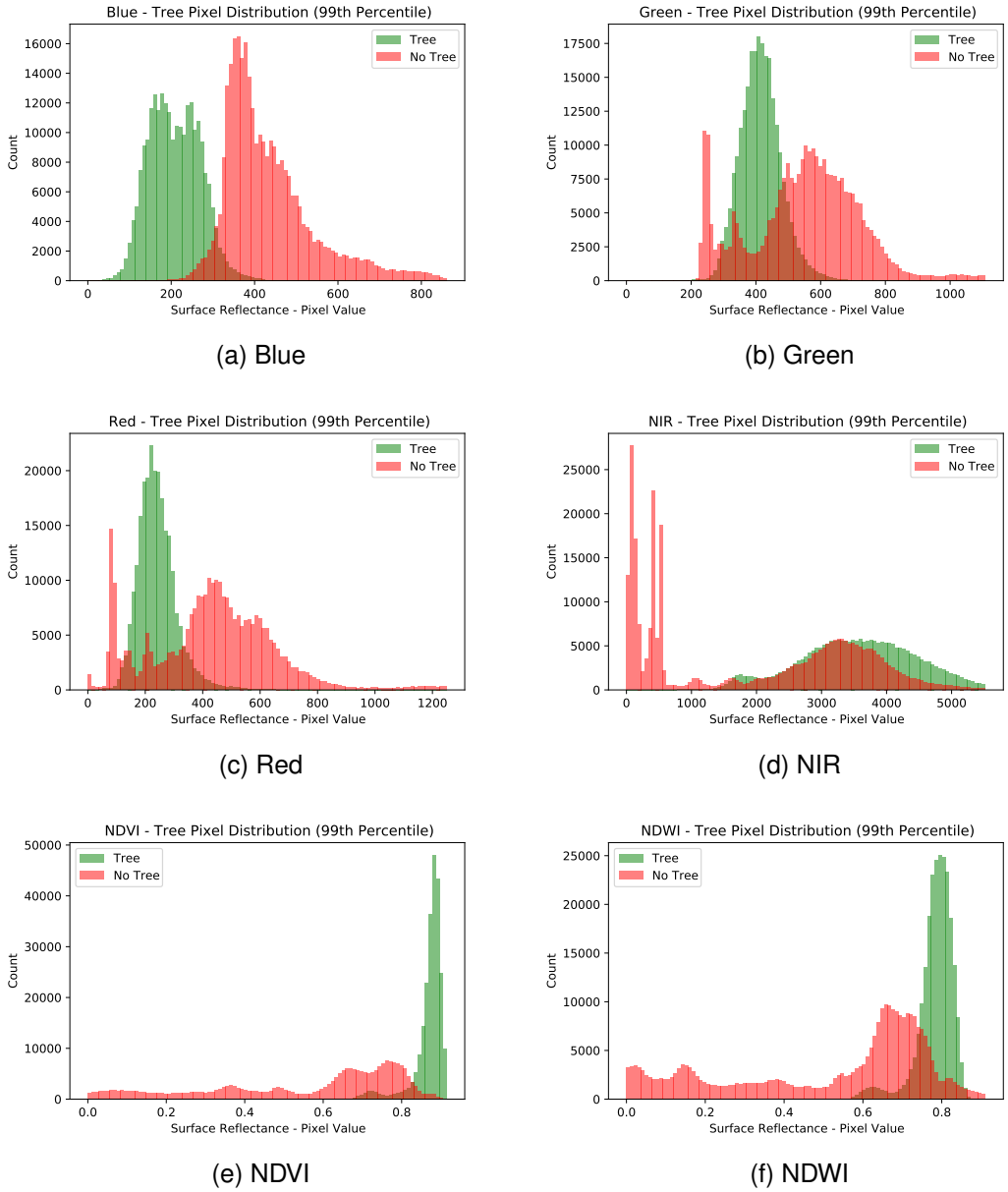
Using the raster cropping procedure described above, a total of 890 labelled polygons generated a finalised tabular dataset representing over 500,000 labelled pixels. As shown in Table 3.1, there are ∼380,000 labelled pixels in the train set, ∼81,000 labelled pixels in validation and ∼78,000 in test. In simple terms, rows in these tabular datasets represent pixels and the columns represent satellite imagery bands {Red, Green, Blue, NIR}.

Dataset	Tree (X=1)	Tree (%)	No Tree (X=0)	No Tree (%)	Total
Train	157,249	41%	227,586	59%	384,835
Validation	35,371	43%	46,407	57%	81,778
Test	31,530	40%	46,596	60%	78,126

Table 3.1: Train, validation and test dataset sizes, including tree class (%).

3.5 Analysis

3.5.1 Tree Pixel Distribution



(a) Blue (b) Green

(c) Red (d) NIR

(e) NDVI (f) NDWI

Figure 3.5: Multi-Spectral Pixel Distribution for “Tree” vs “No Tree” classes.

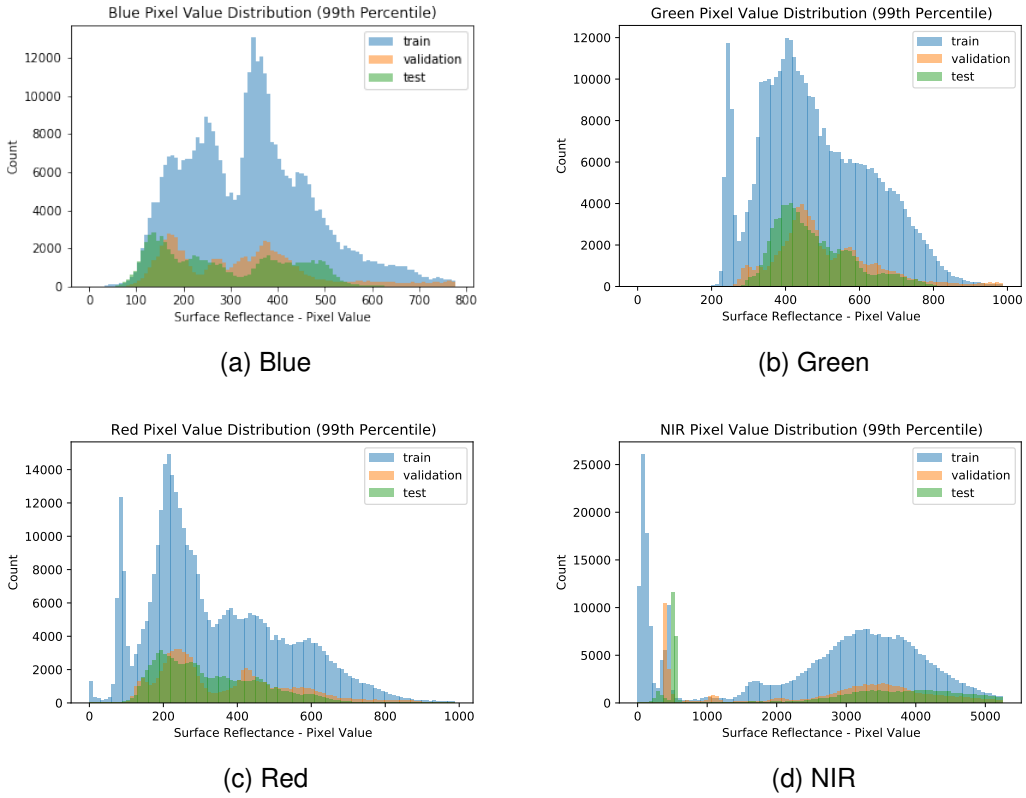
Figure 3.5 directly compares the respective distributions between “tree” or “no tree” class labels for each pixel in the final labelled dataset. Moreover, the distributions of composite pixel vegetation indices NDVI and NDWI (see Section 4.3.3) are also included in the analysis as they showed promising linear separation between class labels. For each pixel band, the distribution of surface reflectance values are plotted up to the 99th percentile to remove outliers in the pixel data.

From the results, we can see that multi-spectral pixel values can be used to linearly

separate the two classes in the data. Moreover, we can observe that certain pixels are better discriminating between "tree" and "no tree" classes. For example, the blue pixel (a) shows high surface reflectance values for the "no tree" class . This could be explained because non-tree objects, such as rivers, oceans and buildings often reflect a higher intensity of blue light. Similarly, high NDVI values > 0.8 in the dataset shows a strong likelihood of the class being "tree", whilst low NDVI values indicate the presence of "no tree". This finding confirms the literature that suggests high NDVI values (approximately 0.6 to 0.9) correspond to "dense vegetation, such as that found in temperate and tropical forests" (Pettorelli, 2013).

3.5.2 Dataset Pixel Distribution

Maintaining a similar pixel distribution between train, validation and test datasets is critically important because differences can dramatically impact model evaluation. Figure 3.6 compares the distribution of surface reflectance values between these datasets for each pixel. As before, the values are plotted up to the 99th percentile to remove outliers. The results show that a similar distribution of surface reflectance is maintained between datasets. One important observation is the training data contains a high count of low surface reflectance for Green, Red and NIR pixels (spikes on left of each chart). This means a supervised model will learn features that are not present in validation or test data. However, this is not problematic because from Figure 3.5 we know these pixels are of the "no tree" class, so it would not negatively impact classification.



(a) Blue (b) Green

(c) Red (d) NIR

Figure 3.6: Multi-Spectral Pixel Distribution for Train, Validation and Test Sets

Chapter 4

Methodology

In this section, we present the methodology used for training a supervised machine learning model to detect the presence of trees in PlanetScope satellite imagery. First, a brief introduction into decision tree and random forest algorithms is explored. Second, the methods followed in this project are described. In particular, we focus on the problem setup and present the specific random forest models used in experiments. Finally, a brief description is provided of the evaluation metrics used to quantitatively assess the performance between models.

4.1 Models

4.1.1 Decision Trees

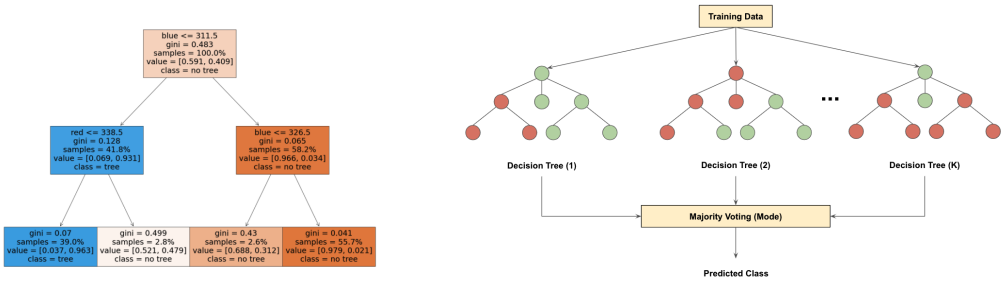
A **decision tree** is a simple machine learning algorithm used for classification and regression tasks (Quinlan, 1986). Decision trees are highly sought after because, unlike deep neural networks, they are simple to train and are easy to interpret.

Decision trees work by splitting features in training data into decision nodes. These nodes represent decision branches leading to a terminal leaf node that represents the predicted class. Decision trees are built by recursively ranking decision nodes by their level of predictive power. The most important decision nodes are priorities towards the root. In this way, as we descend down a decision tree we become more certain of the predicted class of the input data. Decision trees do have limitations. If the decision tree is very large and has many decision nodes, then the model will begin to overfit. To prevent overfitting, decision trees can be pruned therefore reducing the size of the tree.

To classify an input $\mathbf{x}$ using a decision tree, we start at the root and take the branch appropriate to the condition at each node. This cascading binary decision process continues until a terminal leaf is encountered and the final predicted class is found (Quinlan, 1986). Figure 4.1a illustrates a simple Decision Tree with the "tree_depth" parameter set to 3 to simplify visualisation. The model was trained on the data created in this project (see Appendix C.2). To classify an input pixel x_i the root node first checks whether the blue surface reflectance value is $\leq$331.5, if True then proceed to the left node and if False then proceed to the right.

4.1.2 Random Forests

The **Random Forest** algorithm is a machine learning model inspired by the theory, "wisdom of the crowd" (Géron, 2019). As presented in Géron (2019), suppose you ask a thousand people a complex question, then aggregate their answers. In many cases, the answer generated by "the crowd" is better than if you were to ask a single "expert". Also, asking multiple people at random may address the independent bias of asking a single person. Analogous to this theory, it is sensible to hypothesise that aggregating multiple prediction algorithms would return a better result than using a single predictor. This aggregation of prediction algorithms is known as **ensemble learning** and the random forest algorithm is simply an ensemble of decision trees.



(a) Decision tree model

(b) Ensemble learning of Random Forest

Figure 4.1: Example decision tree model trained on labelled training data (left) and random forest model using ensemble learning (right).

Random forest classification algorithms work by aggregating the results of multiple decision tree classifiers. They fit these on distinct sub-sets of the training dataset and aggregate predictions from each model to return a single output. In this way, random forest models use ensemble learning to make a aggregate prediction of all predictors Géron (2019). In classification tasks, typically random forest algorithms select the most frequently predicted class of k decision trees (Figure 4.1b). Random forests are presented in their original paper Breiman (2001) as the following:

Definition 4.1.1. A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(\mathbf{x}, \theta_k), k = 1, ...\}$ where the $\{\theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x. (Breiman, 2001)

This aggregation process has been shown to reduce both bias and variance, therefore Random Forest algorithms consistently perform better than a single Decision Tree model in classification tasks (Géron, 2019).

4.2 Problem Setup

In order to develop a rigorous method that utilises high-resolution PlanetScope imagery to train a random forest classification model, we divided the methodology into three major components: (i) dataset preparation (discussed in Data, chapter 3), (2) initial experiments on RF-models with varying input features (covered in Experiments, chapter

5), and finally, (3) quantitative and qualitative evaluation on the final optimised model named RF-Ranger (discussed in Evaluation, chapter 6). A flow chart summarising the method followed in this project is shown below (Figure 4.2)

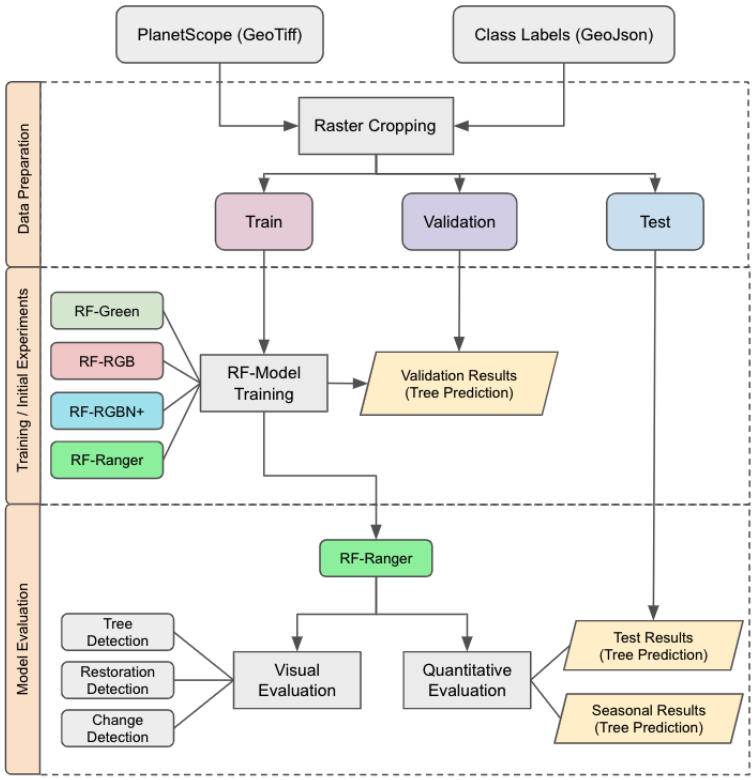


Figure 4.2: Illustration of the data preparation, model training, experiments and evaluation methodology.

4.3 RF-Models

In machine learning, there is an underlying assumption that increasing the number of input features will improve model performance because the model "learns" more information. To test this hypothesis, multiple random forest (RF) models were developed to detect the presence of trees with varying satellite imagery bands as input features. It was expected that more input features would improve classification performance.

4.3.1 RF-Green

First, a baseline random forest model named **RF-Green** was developed. Intuitively, trees reflect green light, therefore it felt important to test whether a very simple baseline model could achieve promising results by only considering the green surface reflectance value. A baseline model was trained using *only* the green pixel value. All other spectral values were dropped from the data to train and evaluate the RF-Green model. The final data sizes for the RF-Green model were (384,835 x 1) for training and (81,778 x 1) for validation.

4.3.2 RF-RGB

The second model **RF-RGB** was trained using only bands of visible light {Red, Green and Blue}. From initial visual inspection of the satellite imagery, it felt reasonable that a human observer could detect the presence of trees in each pixel. Consequently, it was important to test whether a random forest model, constrained to only the visible light spectrum, could detect the per-pixel presence of trees by considering the "colour" of the pixel. To achieve this, the RF-RGB model was trained using the Red, Green and Blue bands. The non-visible band NIR was dropped from the data, resulting in a final data size of (384,835 x 3) for training and (81,778 x 3) for validation.

4.3.3 RF-RGBN+

A third model **RF-RGBN+** was trained using all available spectral bands in the data {Red, Green, Blue and NIR}. In addition, band-derived vegetation indices {NDVI, NDWI, MSAVI2 and BI} were also calculated and included as input features to train the model. Vegetation indices are mathematical operations carried out spectral bands to detect particular types of vegetation in pixels. For RF-RGBN+ we include the following vegetation indices which were found to improve tree detection performance in Brandt and Stolle (2020). Our hypothesis is that including more spectral bands and vegetation indices will make it easier to distinguish between pixels that contain "tree" and "no tree" classes. The final data sizes for the RF-RGBN+ model were (384,835 x 8) for training and (81,778 x 8) for validation.

(1) Normalised Difference Vegetation Index (NDVI) is a reflectance measure used to detect the presence of healthy vegetation (Tucker, 1979).

$$\text{NDVI} = \frac{NIR - RED}{NIR + RED} \tag{4.1}$$

(2) Normalised Difference Water Index (NDWI) is a reflectance measure used to detect the presence of liquid water - oceans, rivers, lakes, etc. (Gao, 1996).

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \tag{4.2}$$

(3) Modified Soil-Adjusted Vegetation Index (MSAVI2) addresses the limitations of NDVI when applied to areas with a high degree of exposed soil surface (TLT, 2012).

$$MSAVI2 = \frac{2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - RED)}}{2} \tag{4.3}$$

(4) Bare Soil Index (BI) detects soil variations and is useful for distinguishing between trees and other forms of soil-based vegetation (e.g., grasslands, agriculture) (Brandt and Stolle, 2020).

$$BI = \frac{(BLUE + RED) - GREEN}{BLUE + RED + GREEN} \tag{4.4}$$

4.4 Evaluation Metrics

In this report, each random forest model is evaluated based on the following evaluation metrics: Accuracy and Macro-F1 score. For binary classification tasks, evaluation metrics are composed of the following four possible outcomes: (i) **True Positive (TP)** - model correctly predicts the "tree" class, (ii) **False Positive (FP)** - model incorrectly predicts the "tree" class, (iii) **True Negative (TN)**- model correctly predicts the "no tree" class and (iv) **False Negative (FN)** - model incorrectly predicts the "no tree" class.

4.4.1 Accuracy

Accuracy measures the fraction of predictions the model gets correct over all possible outcomes.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.5}$$

In binary classification tasks, however, if there is a class imbalance in the train, validation and test data, then accuracy can be a misleading evaluation metric (He and Garcia, 2009). For instance, consider a trivial dataset with 90% tree and 10% non-tree labels. A model that *always* predicted the presence of tree for every pixel would achieve 90% accuracy. As shown in Table 3.1, we have a minor class imbalance in the novel data - 41% tree and 59% no-tree. This class-imbalance is not dramatic and is maintained across train, validation and test datasets, so it should therefore not overestimate the model performance. To be cautious, in addition to accuracy, another evaluation metric that accounts for class imbalances known as Macro-F1 score is also used in this study.

4.4.2 Macro-F1

Macro-F1 measures the average F1-Score between all class labels. The F1-Score is the harmonic mean of precision and recall, thereby accounting for any class imbalance in the data distribution. Macro-F1 is useful in multi-class tasks, as it can help determine how well the model predicts the more rare classes. The Macro-F1 is computed as the following, where i is the class index and N represents the number of classes:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}; \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.6}$$

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.7}$$

$$\text{Macro-F1} = \frac{1}{N} \sum_{i=0}^{N} \text{F1-Score}_i \tag{4.8}$$

Chapter 5

Experiments and Results

This chapter provides a detailed examination of the random forest model experiments and results conducted in this project. First, this chapter investigates three random forest models, testing the respective model performance on different spectral-band input features. Following this, hyperparameter tuning is then carried out on the best overall model from these tests. Finally, the best performing model is evaluated on an unseen test set.

The python machine learning library Scikit-learn (Pedregosa et al., 2011) was used to implement all machine learning models and to run experiments.

5.1 Feature Comparison

Initial experiments were conducted to evaluate the performance of random forest models in detecting the presence of trees in pixels with varying levels of satellite imagery bands as input features. Our hypothesis is that models with more input features will have better tree classification performance. To conduct these experiments the RF-Green, RF-RGB and RF-RGBN+ models were developed. Each model was trained on their specific input features from the training set. The respective performance of each model was then evaluated on 81,778 held-out pixels in the validation set.

Random forest hyperparameters were kept constant across all models to enable fair comparison. Each RF-model contained 32 decision tree estimators. The "bootstrap" parameter was set set to True to ensure training data points were re-sampled in each decision tree *with replacement*. The "max_features" parameter (as default) was set to the square-root of the total number of features, thereby limiting the depth and improving the efficiency of each model. For initial experiments, 32 estimators were used in each model because it was fast to train and evaluate models. All model performance results on the validation set are shown in Table 5.1.

Model		Validation Set	
		Accuracy (%)	Macro-F1 (%)
Experiments	RF-Green (n=32)	82.7	82.6
	RF-RGB (n=32)	96.4	96.3
	RF-RBN+ (n=32)	**97.7**	**97.7**

Table 5.1: Initial experiments results. Validation set performance comparison for the three random forest models. Best results are in **bold**.

5.1.1 Results

The results indicate that increasing the number of input features improved model performance. In particular, the RF-RGBN+ model with access to both visible and non-visible bands as input features was consistently the best performing model on the validation data. This model achieved 97.7% on both accuracy and Macro-F1 evaluation metrics, resulting in a +15.1% Macro-F1 improvement over RF-Green and +1.4% improvement over RF-RGB.

5.1.2 Discussion

These findings demonstrate that the non-visible spectral band (NIR) and vegetation indices (NDVI, NWDI, MSAVI2 and BI) could be crucial to predicting the existence of trees in pixels in satellite imagery. At this stage of understanding, we speculate that vegetation indices have more predictive power than visible spectral bands such as Red, Green and Blue light. To test this hypothesis, it would be beneficial to carry out permutation importance to compare the relative contribution of each feature to the final RF-RGBN+ model (discussed in next section).

The results indicate that the RF-Green model, which only considers one single spectral value (e.g. how green is a pixel?) appears to perform worse than models trained on multiple spectral values. This finding was in line with expectations given that various non-tree classes, such as grasslands, agricultural farming and shrubs, can also reflect a high intensity of green light.

5.2 Model Optimisation

In this next set of experiments, we aim to fine-tune the RF-RGBN+ model to boost its performance on the validation set. This is achieved by conducting two experiments. First, we carry out permutation importance to learn what spectral bands have the greatest predictive power in the RF-RGBN+ model. Using this information, our aim is to remove redundant features, thereby further optimising the RF-RGBN+ model. Second, hyperparameter tuning is carried out to find the optimal number of estimators within the RF-RGBN+ model.

After conducting these two experiments, a final optimised random forest model is proposed, named **RF-Ranger** (which is named after the profession "forest ranger", a person who protects and monitors reforestation projects).

5.2.1 Feature Selection

Permutation importance is a process that is used to identify which input features in a random forest model have the most predictive power. As shown in the initial experiments, machine learning models trained on a greater range of input features can often lead to better classification performance. However, not all features used in training may be relevant. Therefore, permutation importance can be used to reduce the feature space and further optimise the model.

As presented in Parr et al. (2018), permutation importance identifies the most important features in a trained model by randomly shuffling the input features and calculating the effect on model accuracy. In this way, permutation importance can be used to investigate whether all input features in a random forest model contribute to the classification accuracy or whether some should be removed. Permutation importance was carried out by adapting code from ELI5 Permutation Importance Documentation (ELI5, n.d.).

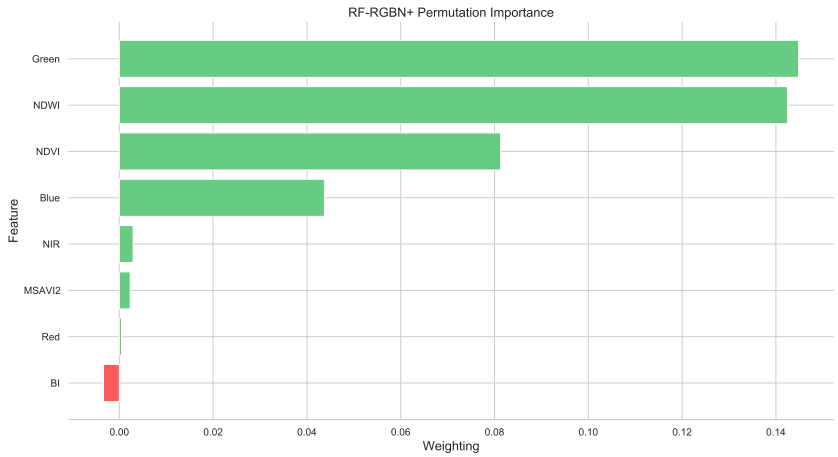


Figure 5.1: Permutation Importance for the RF-RGBN+ model. Not all vegetation indices contribute to model performance, Bare Soil Index (BI) negatively impacts model.

Figure 5.1 shows permutation importance applied to the RF-RGBN+ model. The results indicate that Green, Blue, NDWI and NDVI input features contribute the most to the RF-RGBN+ model. Meanwhile, the Bare Soil (BI) vegetation index actually negatively impacts performance and the Red band has only minimal contribution. One rational explanation is that NDWI and Blue input features are most commonly associated with "no tree" classes such as rivers and oceans. Similarly, the Green and NDVI input features are most useful for detecting the presence of healthy vegetation, such as natural forests and trees within active forest restoration sites. Together, these four input features carry the vast majority of the predictive power of the classification model.

Following recommendations outlined in Parr et al. (2018), input features that return negative permutation importance can be removed from the model without hindering performance. Therefore, for the final optimised RF-Ranger model, the BI input feature is dropped.

Another interesting commentary is to mention the relationship between permutation importance and the tree pixel distributions shown previously in Section 3.5.1. The tree pixel histograms generated for the Green, Blue, NDWI and NDVI show the greatest amount of separation between "tree" and "no tree" classes. This shows that the current RF-RGBN+ model is successfully identifying the linear separability in the pixel distributions between these two classes in these input features. Comparing these two charts is an effective measure to validate that the proposed random forest model is working constructively and identifying the underlying patterns in the training data.

5.2.2 Hyperparameter Tuning

As discussed in Section 4.1.2, a random forest model is fundamentally an ensemble of k decision trees. Intuitively, as the number of trees increases, the training time slows and the model performance improves. However, as $k \to \infty$ model performance follows the law of diminishing returns: classification accuracy marginally improves but at the expense of computational cost of training more decision trees. As a consequence, one important hyperparameter in building a random forest model is to determine the optimal number of k estimators. To optimise the RF-Ranger model further, hyperparameter tuning was carried out to investigate the optimal number of estimators in the RF-RGBN+ model. Figure 5.2 plots the Macro-F1 score against different numbers of estimators $k = [1, 2, 4, 8, 16, 32, 64, 128, 256]$. The results show the effect of diminishing returns and indicate the optimal number of estimators is 128.

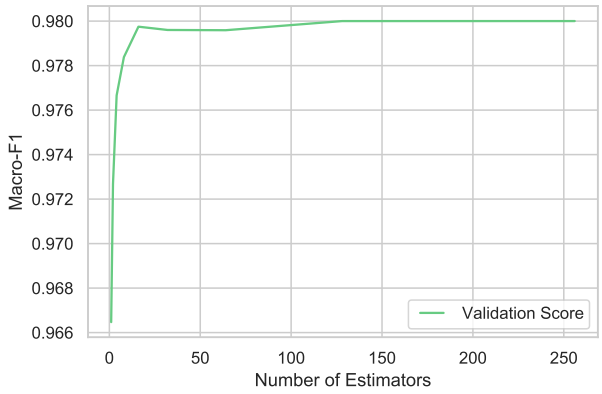


Figure 5.2: Plotting Macro-F1 score to number of decision tree classifiers used in RF-RGBN+ model. Macro-F1 has elbow at 16 estimators and increases whilst plateauing at 128 estimators.

5.2.3 Results

The final optimised RF-Ranger model is evaluated on 81,778 pixels from the validation and 78,126 pixels from the test datasets respectively. The model performance is compared against RF-Green, RF-RGB and RF-RGBN+ to illustrate the marginal performance improvements after fine-tuning.

Model	**Validation Set**		**Test Set**	
	Accuracy (%)	Macro-F1 (%)	Accuracy (%)	Macro-F1 (%)
RF-Green (n=32)	83.0	82.5	75.4	74.4
RF-Visible (n=32)	96.4	96.3	93.1	92.9
RF-RBN+ (n=32)	97.7	97.7	96.3	96.2
RF-Ranger (n=128)	**98.0**	**97.9**	**96.4**	**96.3**

Table 5.2: Results on the validation and test set for the random forest models. Best results are in **bold** and n = num_estimators.

As shown in Table 5.2, the optimised RF-Ranger model achieves the best results on the unseen test set, recording a Macro-F1 score of 96.3% and an accuracy of 96.4%. These results show only marginal improvements over the RF-RGN+ model, +0.1% gains on both accuracy and Macro-F1. However, the final model achieves significant improvements over RF-Green (+21.9% Macro-F1) and RF-RGBN+ (+3.4% Macro-F1). The RF-Ranger Macro-F1 performance on the test set is quite similar validation, therefore suggesting the final model can successfully generalise to new unseen pixels within the domain of northern Fiji.

5.2.4 Discussion

One initial concern about the findings of Table 5.2, is that the three best performing models all achieve high +90% accuracy and Macro-F1 scores. However, this should not be a cause for concern because high accuracy is expected for simple classification tasks with only two classes. If we were to expand this problem into detecting more classes, such as predicting the species of tree (“reforested” vs “natural forest”) and type of non-tree (“ocean”, “building”, “river”, “road”, etc) then we would expect the accuracy and Macro-F1 scores to be lower.

In the context of the related literature, the results observed above are closely aligned to the accuracy scores of other papers focused on binary tree classification tasks. For instance, Bolyn et al. (2018) achieved 93.3% accuracy using a random forest classifier to classify “tree” and “no tree” classes using Sentinel-2 satellite imagery for a forested region in Belgium. However, the model accuracy dropped to 88.9% when predicting the species of tree from 11 forest classes (Bolyn et al., 2018). Whilst this is evidently in a different geographical location, it does confirm that high +90% accuracies can be achieved using a random forest model for tree detection.

Chapter 6

Evaluation

This chapter is devoted to evaluating the suitability of the RF-Ranger model in helping effectively monitor reforestation projects. Two distinct methods were chosen for evaluation: (i) quantitative evaluation and (ii) visual evaluation. The first section is dedicated to measuring differences in model performance between seasons. The second section focuses on visually evaluating the model on a number of tasks relating to monitoring forest restoration sites. These task include (i) detecting the presence of "trees" and "no tree" classes in satellite imagery, (ii) identifying planted trees in reforestation sites, and finally, (iii) detecting changes in forest growth between two distinct time periods.

For the purposes of evaluation, we test the model on a real-world active forest restoration site, managed by CI, in the Nakauvadra Reforestation Project in northern Fiji. We also compare the performance of the proposed RF-Ranger model against the current state-of-the-art model by Hansen et al. (2013) which was accessed using the forest explorer tool on the Global Forest Watch online platform (GFW, 2021).

6.1 Quantitative Evaluation

6.1.1 Seasonal Performance

To test the seasonal performance of the RF-Ranger model, subsets of the test data representing 41,077 pixels from June 2019 and 37,049 pixels from December 2019 were created. The RF-Ranger model then predicted the presence of "tree" and "no tree" classes in these unseen seasonal datasets and the results are shown in Table 6.1.

The results verify that the RF-Ranger model is largely season invariant, thereby confirming that the model can achieve promising results all year round. This is an important finding because it would allow the CI forest restoration monitoring team to identify tree re-growth more frequently. For instance, this could enable the projects to be surveyed between months, not years.

Table 6.1 splits the performance into individual F1-Scores for "tree" and "no tree" classes respectively. By comparing the results between seasons, we can see that the

Model	Jun 2019 (F1-Score)		Dec 2019 (F1-Score)	
	No Tree	Tree	No Tree	Tree
RF-Ranger (n=128)	**0.96**	0.93	0.94	**0.98**
Macro-F1	0.95		0.96	

Table 6.1: Performance evaluation of the RF-Ranger model on seasonal subsets of the test data. Best results are included in **bold**.

model achieves better performance in identifying the "no tree" class in June and the "tree" class in December. One reasonable explanation for this is that June represents the end of wet season in Fiji, meaning grasslands and agricultural lands may been "greener", thus making it harder to distinguish between trees and non-tree green vegetation. Conversely, December typically describes dry summer season in Fiji, therefore detecting the presence of trees would be expected to be easier in this setting.

6.1.2 Limitations

The primary limitation of quantitative evaluation with satellite imagery is the lack of data - both in terms of quality and quantity.

The ideal quantitative evaluation method would be to cross-validate the coordinates of the random forest model predictions with real "on the ground" locations of trees on earth. Unfortunately however, this geo-referenced data was not available in the undertaking of this project because CI do not collect coordinates for each individual tree planted in the Nakauvandra Reforestation Project.

The second best method for quantitative evaluation is to label satellite imagery and then measure the performance of the model on these labelled datasets. However, as discussed in Chapter 3, labelling training examples in satellite imagery can be time consuming. Also, even although best practices can be followed to reduce bias when labelling, there still may be minor human inaccuracies in the labelling process.

For all of these reasons outlined above, we believe that quantitative evaluation can (and should) be supported by further visual validation. This qualitative evaluation method can be used to manually validate the random forest model predictions overlaid onto satellite imagery, checking whether the model appropriately identifies trees from space. This is the approach we follow in this report.

6.2 Visual Evaluation

In addition to quantitative analysis, visual evaluation was used extensively to manually validate the model performance in detecting the presence of trees in PlanetScope satellite imagery.

To visually evaluate the RF-Ranger model, we downloaded recent PlanetScope basemaps (December 2019) and generated tree cover maps for an area of interest within the satel-

lite imagery. Typically these areas were small, representing 1km x 1km subsections of the PlanetScope imagery over northern Fiji.

As shown in Figure 6.1, tree cover maps were generated by the RF-Ranger model by predicting the class for each pixel in a subset of PlanetScope data. Pixels that intersect with a "tree" class are labelled 1 and pixels that intersect with "no tree" classes are labelled 0. These per-pixel binary labels can the be used to generate a simple two-colour mask which can be overlaid onto the original basemap satellite imagery to visually validate the model performance.

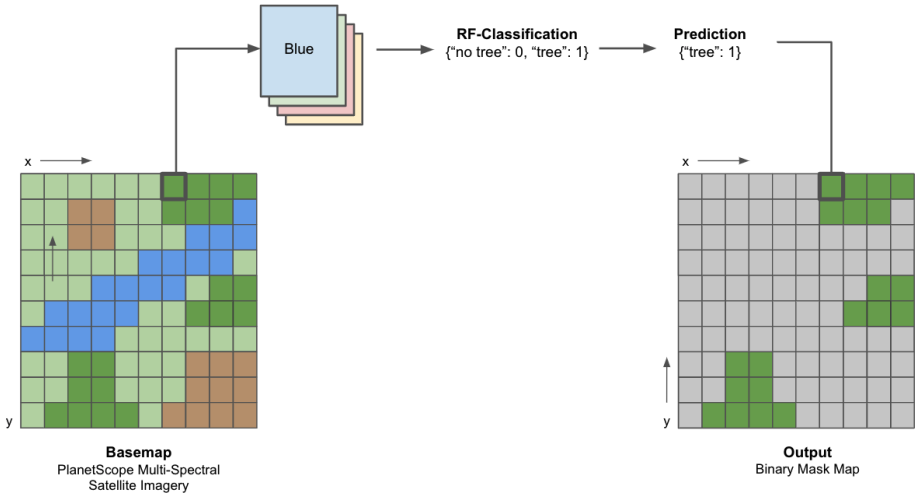


Figure 6.1: Diagram of Per-Pixel Classification Method and Binary Mask Generation.

Using this method outlined above, binary masks were generated to investigate various research areas of interest. In the following subsections we set out to answer the following questions.

1. Can the proposed RF-Ranger model correctly identify trees in PlanetScope satellite imagery?
2. To what extent can the RF-Ranger model identify planted or growing trees in active forest restoration sites?
3. Can the RF-Ranger model detect changes in tree cover between two distinct time periods?
4. To what extent is the RF-Ranger model better than existing models for Tasks (1-3) outlined above?

In the following subsections, the predicted binary masks are overlaid onto very high-resolution imagery (0.3m per pixel) using the latest satellite imagery filters on Google Earth Pro. However, it is important to note that it was not possible to download the per-pixel spectral band values from Google Earth. For this reason, Google Earth imagery was not used to train any of the random forest models and was not used for quantitative analysis. Instead, it was only used to facilitate better visual comparisons between different models.

6.2.1 Tree Prediction

Figure 6.2 shows a binary forest cover map generated by the RF-Ranger model for a 1km x 1km sub-region within the Nakauvadra mountain range. This figure presents a scene of sparsely distributed tree cover. The image on the **top left** of Figure 6.2 shows the input PlanetScope imagery fed into the RF-Ranger model. The image on the **top right**, shows the tree cover predictions generated by the RF-Ranger model overlaid onto the same PlanetScope imagery. It is important to note, the image on the **bottom right** shows the same tree cover predictions (at PlanetScope spatial resolution) but overlaid onto Google Earth Pro imagery to enable better visual validation.

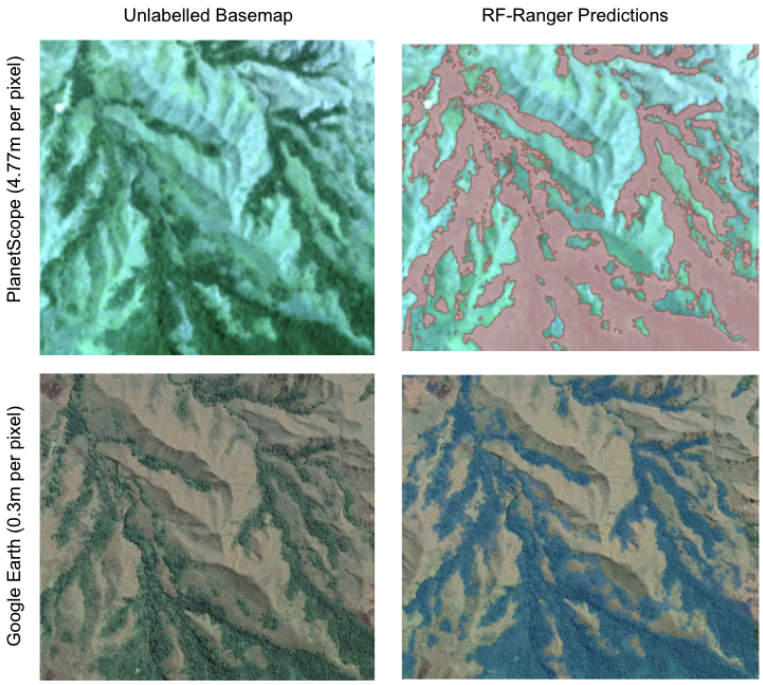


Figure 6.2: Tree predictions of the proposed RF-Ranger model overlaid onto PlanetScope (pink) and Google Earth Pro imagery (blue). Both the pink and blue binary masks represent the same RF-Ranger predictions (at 4.77m per pixel resolution) just overlaid onto satellite images with different spatial resolutions.

The tree cover maps demonstrate that the RF-Ranger model can successfully locate trees in PlanetScope satellite imagery. The proposed random forest model can detect large natural forests and is even capable of identifying sparsely distributed patches of trees (Figure 6.2) (see also Appendix A for more examples).

When visually compared with the tree predictions from Hansen et al. (2013), the RF-Ranger model achieves significantly better performance in detecting trees in this scene with sparsely distributed tree cover. As illustrated in Figure 6.3, the proposed model predictions using high-resolution PlanetScope satellite imagery provide much clearer mapping of tree cover compared to the Hansen et al. (2013) model. In particular, the results show that Hansen et al. (2013) predictions routinely underestimate tree cover. These results thereby confirm the research hypothesis that random forest mod-

els trained on high-resolution PlanetScope satellite imagery can improve tree detection. Moreover, the results confirm the findings detailed in the relevant research that mention the current limitations of Hansen et al. (2013) with respect to detecting tree cover in heterogeneous landscapes.

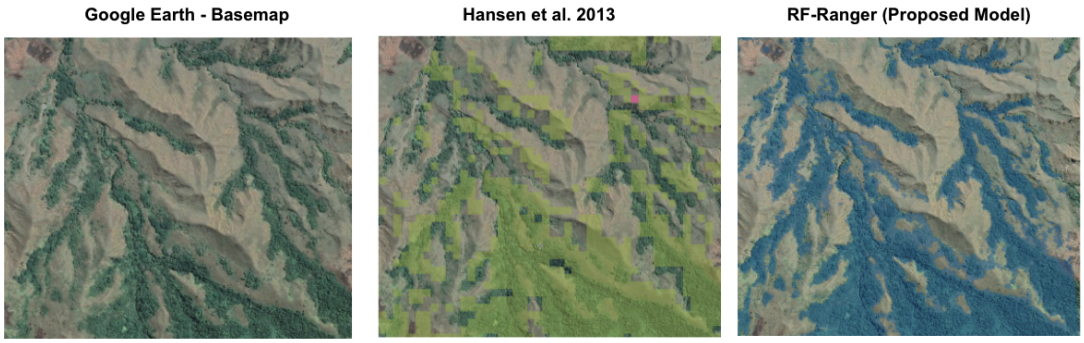


Figure 6.3: Tree predictions of the RF-Ranger model (blue pixels). For comparison, tree predictions generated by Hansen et al. (2013) are added for the same region (light green pixels). All binary maps are overlaid onto very high resolution Google Earth Pro imagery to enable better visual interpretation.

6.2.2 Reforestation Monitoring

The primary research focus of this dissertation was to investigate to what extent could a machine learning algorithm detect the presence of planted trees in active forest restoration sites. In this subsection, we explore this question by generating a binary tree cover map for a planting region in the Nakauvadra Reforestation Project.

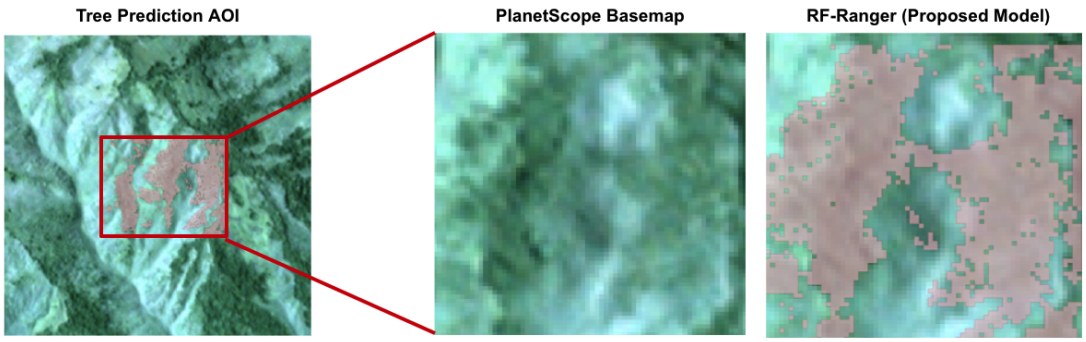


Figure 6.4: Tree predictions for a single **active reforestation site** in northern Fiji.

To run this experiment, on-ground tree planting data from the CI Forest Monitoring Team in Fiji was used. This data included 27 GPS locations, representing the center point of 27 active forest reforestation planting sites across the Nakauvadra Reforestation Project. These GPS locations do not show the exact location of each planted tree but instead represented the centre point of each reforestation site. This meant the data

could not be used for quantitative analysis because there was no way to align the co-ordinates of each planted tree with the pixels in the PlanetScope imagery. Having said this however, the GPS locations were used to generate tree prediction binary masks for the area 0.5km x 0.5km surrounding the center point of the planting site GPS location. Then validation was carried out using visual interpretation. For the purpose of evaluation, one planting site out of 27 was focused on. This planting site was chosen because it represents one of the largest planting regions within the Nakauvadra Reforestation project. All predictions used PlanetScope satellite imagery from December 2019.

Figure 6.4 was included in this report to show what the RF-Ranger model "sees" during training and when making predictions. However, as can be seen above, it is difficult for humans to visually validate whether planted trees exist within these images. Therefore, to facilitate easier visual interpretation of results, the exact same RF-Ranger prediction for the same reforestation area is overlaid onto Google Earth Pro imagery in Figure 6.5 below.

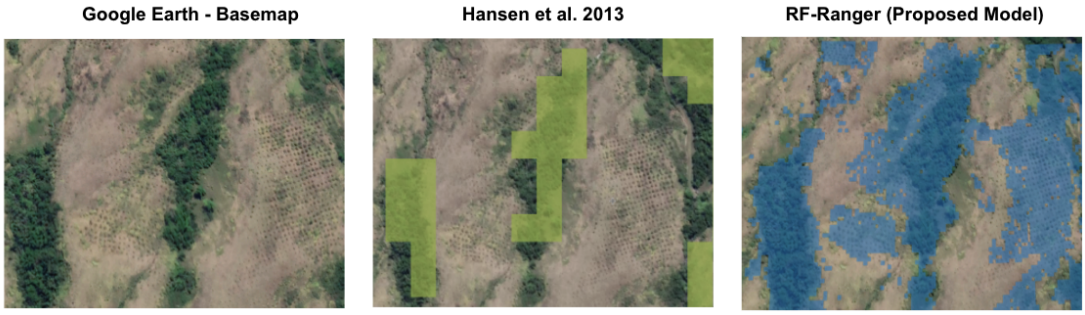


Figure 6.5: Tree prediction of the proposed model RF-Ranger (blue pixels) for a single **active forest restoration site** in northern Fiji. Predictions of Hansen et al. (2013) are also included (light green pixels) for comparison. All binary maps are overlaid onto very high resolution Google Earth Pro imagery to enable better visual interpretation.

The results indicate the proposed RF-Ranger model can successfully identify planted trees in the chosen reforestation site. When compared to Hansen et al. (2013), the proposed model is significantly better at identifying planted trees. Specifically, the Hansen et al. (2013) model underestimates tree cover in the region, failing to detect planted trees and only identifying the main forest blocks in the image. The success of the RF-Ranger model is likely largely attributed to the fact that it has access to higher resolution satellite imagery. Nonetheless, the results demonstrate the potential for newly available PlanetScope satellite imagery to help identify planted trees in forest restoration sites using machine learning methods. Until now, this area of research has simply not been explored.

One limitation of the RF-Ranger model is it struggles to identify very small trees to the south edge of the planting region. Most likely there is a minimum limit on tree width to be able to effect the spectral composition of the pixel. It would be interesting to investigate what this limit is, however, this research is beyond the scope of this dissertation.

6.2.3 Change Detection

A potentially highly useful application of the per-pixel tree prediction model RF-Ranger is the ability to quantitatively analyse changes in tree cover between two time periods. This tree cover gain and loss information is incredibly valuable to non-profit environmental organisations for measuring the quality of restoration efforts and to assess whether interventions are making a difference.

Figure 6.6 below shows the RF-Ranger tree predictions for the chosen forest restoration site between December 2016 and December 2019. Predictions were made using PlanetScope imagery for these two dates as input data. It is important to note, Google Earth imagery was not possible to access for 2016 or 2019 and so the binary tree cover predictions are overlaid onto the same imagery date (September 2017) to improve visual validation.

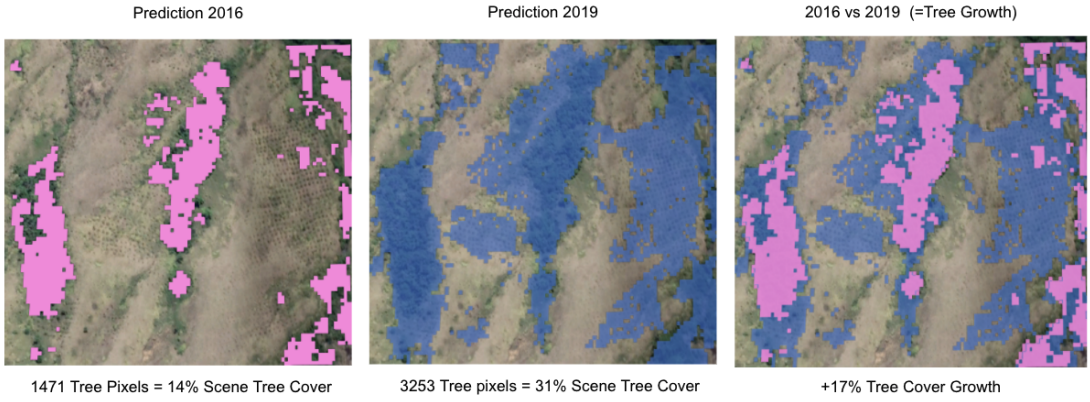


Figure 6.6: Detecting tree cover gain in the active reforestation site between December 2016 and December 2019. All predictions are from the RF-Ranger model onto an image of 10,350 existing pixels. Taking this into account, tree prediction changes represent a +17% gain in per-pixel tree cover over a 3 year period.

The results indicate that on a pixel-level (tree predictions are per-pixel), there is a +17% increase in tree cover in the reforestation planting region between these two time periods. It was previously not possible to quantitatively analyse this kind of planted tree growth information at this small scale using satellite imagery. Thus, the results show promising signs that random forest models such as RF-Ranger could be used to effectively monitor tree regrowth in active forest restoration sites.

Beyond assessing the quality of forest restoration, these change detection maps can potentially help non-profit organisations decide where the protect and plant trees. Intuitively, these organisations want to plant trees in areas which are suitable for tree growth. A significant amount of time, effort and resources go into deciding where to plant trees and protect regions experiencing forest regrowth. The findings above show that RF-Ranger could potentially assist with this decision process by mapping where tree growth has *already* occurred with a large region. Overall, the visual evaluation results indicate that the RF-Ranger model has potentially multiple useful applications to help with active forest restoration site monitoring.

Chapter 7

Discussion

Predicting the presence of trees in active forest restoration sites using satellite imagery is a challenging task. Specifically it is difficult because planted trees have small crown widths which are often smaller than the pixel resolution of the satellite imagery. The primary research objective of this project was to investigate to what extent could machine learning models, trained on recently available open-source PlanetScope satellite imagery, be used to identify trees in active reforestation projects. To achieve this research aim, a random forest model was proposed which showed promising results during both quantiative and visual evaluation experiments. In this section we discuss broader conclusions from the study, explore potential implications for supporting forest restoration efforts and provide suggestions for future areas of work.

7.1 Comparison to State-Of-The-Art

Overall, the experiment results demonstrate that the RF-Ranger model can achieve promising 96.4% classification accuracy and 96.3% Macro-F1 score on unseen pixels in a test set. Furthermore, additional visual evaluation illustrated that the proposed model outperforms the current state-of-the-art model developed by Hansen et al. (2013) in predicting the existence of planted trees in reforestation projects. As discussed in Chapter 6, the difference between these two models can be largely attributed to the fact the RF-Ranger model has access to higher resolution satellite imagery. For instance, the model is trained on 4.77m per pixel PlanetScope resolution imagery, whilst the existing model by Hansen et al. (2013) uses 30m Landsat imagery.

Even so, this research has contributed to the first evaluation of PlanetScope satellite imagery for the task of identifying planted trees in forest restoration sites. The results demonstrate the feasibility of predicting the presence of planted and growing trees: a task that was previously not achievable with low-resolution satellite imagery and yet could have major impact in addressing some of the real-world challenges of monitoring forest restoration projects using remote sensing data.

7.2 Implications for Active Reforestation Monitoring

As a result of the research conducted in this project, a forest monitoring system was successfully developed to predict the presence of trees in the Nakauvadra Reforestation Project in northern Fiji. In their current state, the materials developed in this project including the unique labelled dataset, the random forest methods and the visual evaluation experiments will help the CI forest monitoring team to quantitatively and qualitatively assess the performance of forest intervention efforts for this region. After sharing the initial results from this dissertation with Isaac Rounds, the Nakauvadra restoration site manager, I was told that "this sure will be a great tool for restoration work in Fiji"[1]. As a result, his comments confirm that the materials generated in this report will hopefully have real-world impact.

In addition, this project presents an outline for how CI (and other non-profit organisations) could develop similar systems for new geographical regions. Our goal here is that the CI forest monitoring team will follow a similar machine learning development pipeline to the template outlined in this project and this will in turn help researchers at CI to monitor future forest restoration projects in new areas.

7.3 Future Work

7.3.1 Dataset Quality

The quality of a supervised machine learning model is highly dependent on the quality of the data samples used for training. In this project, training samples were manually labelled using visual interpretation of "tree" and "no tree" classes in PlanetScope satellite imagery. A potential significant improvement to the data used in this project, would be to have access to geo-referenced labelled examples of real-world trees from "on-the-ground" inventory surveys in forest restoration sites. To train a model on this data, the geo-referenced pixels in the PlanetScope imagery could be successfully aligned with these real-world trees, therefore providing a more robust dataset for training and evaluation. To generate this dataset, it would admittedly require significant human involvement and the costs involved would likely limit the number of geo-referenced trees that could be measured. Nevertheless, this data *in combination* with labelled satellite imagery could be beneficial for improving the quality of the models.

7.3.2 Multi-Class Classification

In this report, we present a simplified binary classification task. Trees grow and evolve through time, therefore one limitation of the approach covered in this project is that it does not address the challenges of distinguishing between trees at different stages of growth. One possible development could be to extend predictions into a hierarchical classification task: first, predict the presence of "tree" vs "no tree", then distinguish between multiple "tree" classes such as "planted trees" vs "pre-existing trees". This

[1] Quote directly taken from a private email exchange with Isaac Rounds, Carbon-Offset Manager at CI in Fiji after sharing initial results of tree predictions generated by the random forest model.

would be a useful distinction for helping forest monitoring efforts. Intuitively, planted trees should be harder to identify than more established forests, so we would expect classification error rates to be higher for these classes. Currently, the RF-Ranger model is not suitable to address these challenges and it remains unclear whether the high classification accuracies for the "tree" class are primarily enhanced by the number of labels containing dense natural forests in the evaluation data. Further work in this area could be interesting and would help assess a more true representation of classification performance for identifying "planted" trees.

7.3.3 Global Model

Here we developed a random forest model for a specific region in northern Fiji. Having shown promising classification results in a small region, the next logical step would be to extend the model to new geographical regions. PlanetScope satellite imagery is freely available for 96 countries across the tropics. Future research into developing a globally consistent random forest model, trained on multiple regions of PlanetScope satellite imagery would be useful to a wider community of conservation charities, governments and private organisations. The data collection, model training and evaluation pipeline outlined in this project provides a useful template for how a global method might be developed. If achieved, a global model trained on PlanetScope satellite imagery would be well placed to displace Hansen et al. (2013) as the leading global forest mapping resource. As shown in this report, RF-Ranger demonstrates improvements over Hansen et al. (2013) and therefore we would expect this classification accuracy improvements to extend to different geographical regions and potentially at a global scale.

Chapter 8

Conclusion

The primary goal of this project was to investigate to what extent could machine learning methods be used to identify trees in satellite imagery and help monitor forest restoration projects. The secondary objective of this dissertation was to evaluate the suitability of high-resolution PlanetScope satellite imagery for forest mapping using machine learning.

To achieve these research aims, a partnership was established with the non-profit environmental organisation, CI and a final random forest machine learning model was presented that could help address the challenges of monitoring a real-life active forest restoration site in northern Fiji.

A unique dataset was created by manually labelling examples of "tree" and "no tree" classes in newly released, high-resolution PlanetScope satellite imagery. This dataset was then used to train and evaluate a number of random forest machine learning algorithms for the task of identifying the per-pixel presence of trees in satellite imagery. Initial experiments were carried out to test the performance of random forest models with varying input features and it was found that both visible and non-visible bands of multi-spectral satellite imagery enhanced classification performance.

The final random forest model, RF-Ranger, was found to have successful performance in identifying trees in PlanetScope imagery. Specifically, the model achieved 96.4% classification accuracy on more than 75,000 unseen pixels in a labelled test set. Visual evaluation was also conducted which demonstrated the feasibility of the RF-Ranger model in assisting with forest monitoring tasks, including (i) identifying trees in PlanetScope satellite imagery, (ii) predicting the presence of planted trees in forest restoration sites, and finally, (iii) detecting changes in tree cover between two distinct time periods.

Overall, this research has contributed to the first evaluation of PlanetScope satellite imagery for the task of identifying trees in forest restoration sites using machine learning. Not only has this research demonstrated the feasibility of using random forest models for helping forest restoration projects, but this project also helped a non-profit environmental organisation tackle a real-world problem.

Bibliography

Azavea (2020), 'Finally, image segmentation and classification labeling for geospatial imagery'.
URL: *https://www.azavea.com/announcements/announcing-groundwork/*

Bolyn, C., Adrien, M., Gaucher, P., Lejeune, P. and Bonnet, S. (2018), 'Forest mapping and species composition using supervised per pixel classification of sentinel-2 imagery', *Biotechnology, Agronomy and Society and Environment* **22**, 172–187.

Brandt, J. and Stolle, F. (2020), 'A global method to identify trees outside of closed-canopy forests with medium-resolution satellite imagery', *International Journal of Remote Sensing* **42**(5), 1713–1737.
URL: *http://dx.doi.org/10.1080/01431161.2020.1841324*

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

Brownlee, J. (2019), 'Overfitting and underfitting with machine learning algorithms'.
URL: *https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/*

Brownlee, J. (2020), '4 types of classification tasks in machine learning'.
URL: *https://machinelearningmastery.com/types-of-classification-in-machine-learning/*

Campbell, J. B. and Wynne, R. H. (2011), *Introduction to remote sensing*, Guilford Press.

Caruana, R. and Niculescu-Mizil, A. (2006), An empirical comparison of supervised learning algorithms, *in* 'Proceedings of the 23rd international conference on Machine learning', pp. 161–168.

Chan, L.-S., Cheung, G. T., Lauder, I. J. and Kumana, C. R. (2004), 'Screening for fever by remote-sensing infrared thermographic camera', *Journal of travel medicine* **11**(5), 273–279.

CI (2021), 'Deforestation: 11 facts you need to know'.
URL: *https://www.conservation.org/stories/11-deforestation-facts-you-need-to-know*

Crockett, C. (2019).
URL: *https://earthsky.org/space/what-is-the-electromagnetic-spectrum*

Crouzeilles, R., Beyer, H., Monteiro, L., Feltran-Barbieri, R., Moreira Pessôa, A. C., Barros, F., Lindenmayer, D., Lino, E., Grelle, C., Chazdon, R., Matsumoto, M., Rosa, M., Latawiec, A. and Strassburg, B. (2020), 'Achieving cost-effective landscape-scale forest restoration through targeted natural regeneration', *Conservation Letters* **13**.

de Almeida, D. R., Stark, S. C., Valbuena, R., Broadbent, E. N., Silva, T. S., de Resende, A. F., Ferreira, M. P., Cardil, A., Silva, C. A., Amazonas, N. et al. (2020), 'A new era in forest restoration monitoring', *Restoration Ecology* **28**(1), 8–11.

Deng, L. and Platt, J. C. (2014), Ensemble deep learning for speech recognition, *in* 'Fifteenth annual conference of the international speech communication association'.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., Vapnik, V. et al. (1997), 'Support vector regression machines', *Advances in neural information processing systems* **9**, 155–161.

ELI5 (n.d.), 'eli5-org/eli5'.
URL: *https://github.com/eli5-org/eli5*

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L. and Muller, P.-A. (2019), 'Deep learning for time series classification: a review', *Data Mining and Knowledge Discovery* **33**(4), 917–963.

Gao, B.-C. (1996), 'Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space', *Remote sensing of environment* **58**(3), 257–266.

Géron, A. (2019), *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*, O'Reilly Media.

GFW (2021), 'What happened to global forests in 2020?: Global forest watch blog'.
URL: *https://blog.globalforestwatch.org/data-and-research/global-tree-cover-loss-data-2020/*

Global Forest Watch, G. (2002), 'Global forest watch', *World Resources Institute, Washington, DC Available from http://www. globalforestwatch. org (accessed March 2002)* .

Goodfellow, I., Bengio, Y. and Courville, A. (2016), *Deep Learning*, MIT Press. http://www.deeplearningbook.org.

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., Herrero, M., Kiesecker, J., Landis, E., Laestadius, L., Leavitt, S. M., Minnemeyer, S., Polasky, S., Potapov, P., Putz, F. E., Sanderman, J., Silvius, M., Wollenberg, E. and Fargione, J. (2017), 'Natural climate solutions', *Proceedings of the National Academy of Sciences* **114**(44), 11645–11650.
URL: *https://www.pnas.org/content/114/44/11645*

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A.,

Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O. and Townshend, J. R. G. (2013), 'High-resolution global maps of 21st-century forest cover change', *Science* **342**(6160), 850–853.
URL: *https://science.sciencemag.org/content/342/6160/850*

He, H. and Garcia, E. A. (2009), 'Learning from imbalanced data', *IEEE Transactions on Knowledge and Data Engineering* **21**(9), 1263–1284.

Hennessy, A., Clarke, K. and Lewis, M. (2020), 'Hyperspectral classification of plants: a review of waveband selection generalisability', *Remote Sensing* **12**(1), 113.

Ho, T. K. (1995), Random decision forests, *in* 'Proceedings of 3rd international conference on document analysis and recognition', Vol. 1, IEEE, pp. 278–282.

Hościło, A. and Lewandowska, A. (2019), 'Mapping forest type and tree species on a regional scale using multi-temporal sentinel-2 data', *Remote Sensing* **11**(8).
URL: *https://www.mdpi.com/2072-4292/11/8/929*

Huang, C., Davis, L. and Townshend, J. (2002), 'An assessment of support vector machines for land cover classification', *International Journal of remote sensing* **23**(4), 725–749.

Immitzer, M., Vuolo, F. and Atzberger, C. (2016), 'First experience with sentinel-2 data for crop and tree species classifications in central europe', *Remote Sensing* **8**(3).
URL: *https://www.mdpi.com/2072-4292/8/3/166*

IUCN (2011), 'The bonn challenge'.
URL: *https://www.iucn.org/theme/forests/our-work/forest-landscape-restoration/bonn-challenge*

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012), Imagenet classification with deep convolutional neural networks, *in* F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds, 'Advances in Neural Information Processing Systems', Vol. 25, Curran Associates, Inc.
URL: *https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf*

Lal, R. (2008), 'Carbon sequestration', *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1492), 815–830.

Lawrence, R. L. and Moran, C. J. (2015), 'The americaview classification methods accuracy comparison project: A rigorous approach for model selection', *Remote Sensing of Environment* **170**, 115–120.
URL: *https://www.sciencedirect.com/science/article/pii/S0034425715301310*

LeCun, Y., Bengio, Y. and Hinton, G. (2015), 'Deep learning', *nature* **521**(7553), 436–444.

Maxwell, A. E., Warner, T. A. and Fang, F. (2018), 'Implementation of machine-learning classification in remote sensing: an applied review', *International Journal*

of Remote Sensing **39**(9), 2784–2817.
URL: *https://doi.org/10.1080/01431161.2018.1433343*

Milodowski, D. T., Mitchard, E. T. A. and Williams, M. (2017), 'Forest loss maps from regional satellite monitoring systematically underestimate deforestation in two rapidly changing parts of the amazon', *Environmental Research Letters* **12**(9), 094003.
URL: *https://doi.org/10.1088/1748-9326/aa7e1e*

Mishra, A., SINGH, R., Tandon, R., KESARWANI, S. and TIWARI, G. (2013), 'Pigment and protein analysis of certain representatives of cyanobacteria (page 11-13)', **11**, 11–13.

Morrison, E. B. and Lindell, C. A. (2011), 'Active or passive forest restoration? assessing restoration alternatives with avian foraging behavior', *Restoration Ecology* **19**(201), 170–177.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1526-100X.2010.00725.x*

Mountrakis, G., Im, J. and Ogole, C. (2011), 'Support vector machines in remote sensing: A review', *ISPRS Journal of Photogrammetry and Remote Sensing* **66**(3), 247–259.
URL: *https://www.sciencedirect.com/science/article/pii/S0924271610001140*

Ottosen, T.-B., Petch, G., Hanson, M. and Skjøth, C. A. (2020), 'Tree cover mapping based on sentinel-2 images demonstrate high thematic accuracy in europe', *International Journal of Applied Earth Observation and Geoinformation* **84**, 101947.
URL: *https://www.sciencedirect.com/science/article/pii/S0303243419306087*

Parr, T., Turgutlu, K., Csiszar, C. and Howard, J. (2018), 'Beware default random forest importances'.
URL: *https://explained.ai/rf-importance/4*

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* **12**, 2825–2830.

Pettorelli, N. (2013), *The Normalized Difference Vegetation Index*, OUP Oxford.
URL: *https://books.google.co.uk/books?id=FXywAAAAQBAJ*

Philipson, C. D., Cutler, M. E. J., Brodrick, P. G., Asner, G. P., Boyd, D. S., Moura Costa, P., Fiddes, J., Foody, G. M., van der Heijden, G. M. F., Ledo, A., Lincoln, P. R., Margrove, J. A., Martin, R. E., Milne, S., Pinard, M. A., Reynolds, G., Snoep, M., Tangki, H., Sau Wai, Y., Wheeler, C. E. and Burslem, D. F. R. P. (2020), 'Active restoration accelerates the carbon recovery of human-modified tropical forests', *Science* **369**(6505), 838–841.
URL: *https://science.sciencemag.org/content/369/6505/838*

Planet (2020), 'Nicfi program'.
URL: *https://www.planet.com/nicfi/*

Quinlan, J. R. (1986), 'Induction of decision trees', *Machine learning* **1**(1), 81–106.

Radoux, J., Chomé, G., Jacques, D., Waldner, F., Bellemans, N., Matton, N., Lamarche, C., d'Andrimont, R. and Defourny, P. (2016), 'Sentinel-2's potential for sub-pixel landscape feature detection', *Remote Sensing* **8**, 488.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F. and Dormann, C. F. (2017), 'Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure', *Ecography* **40**(8), 913–929.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.02881*

Russell, S. and Norvig, P. (2009), *Artificial Intelligence: A Modern Approach*, 3rd edn, Prentice Hall Press, USA.

Silber, S. and Velton, W. (2020), 'Fact sheet - rainforest animals'.
URL: $https://www.ran.org/fact_sheet_rainforest_animals/ : : text = A : Anaverageof 137, intheworld'stropicalrainforests.$

Thompson, D. R., Guanter, L., Berk, A., Gao, B.-C., Richter, R., Schläpfer, D. and Thome, K. J. (2019), 'Retrieval of atmospheric parameters and surface reflectance from visible and shortwave infrared imaging spectroscopy data', *Surveys in Geophysics* **40**(3), 333–360.

TLT (2012).
URL: $https://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods : modified_soil - adjusted_vegetation_index$

TNC (2021), 'Conserving the lands and waters on which all life depends'.
URL: *https://www.nature.org/en-us/*

Tucker, C. J. (1979), 'Red and photographic infrared linear combinations for monitoring vegetation', *Remote Sensing of Environment* **8**(2), 127–150.
URL: *https://www.sciencedirect.com/science/article/pii/0034425779900130*

Ulmas, P. and Liiv, I. (2020), 'Segmentation of satellite imagery using u-net models for land cover classification'.

UNFCC (2015), 'Adoption of the paris agreement'.
URL: *http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf*

WEF (2020), 'One trillion trees - uniting the world to save forests and climate'.
URL: *https://www.weforum.org/agenda/2020/01/one-trillion-trees-world-economic-forum-launches-plan-to-help-nature-and-the-climate/*

Wright, R. E. (1995), 'Logistic regression.'.

WWF (2021), 'Forest habitat'.
URL: *https://www.worldwildlife.org/habitats/forest-habitat*

Xie, Y., Sha, Z. and Yu, M. (2008), 'Remote sensing imagery in vegetation mapping: a review', *Journal of plant ecology* **1**(1), 9–23.

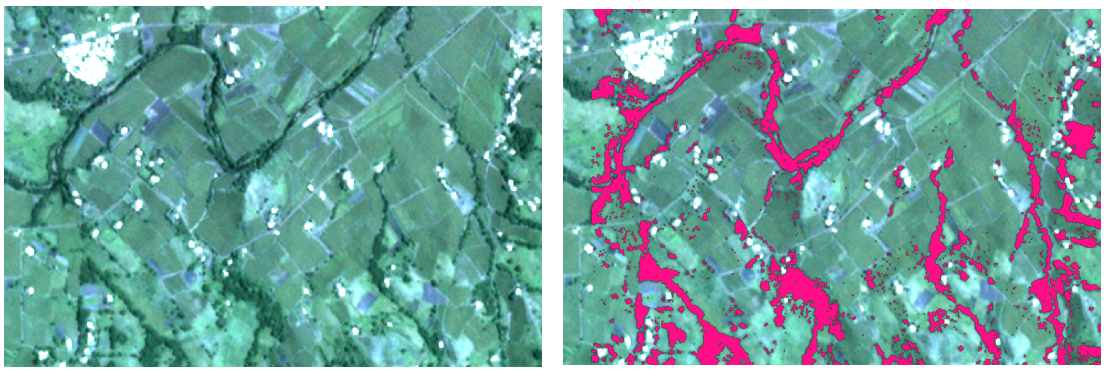
Yegnanarayana, B. (2009), *Artificial neural networks*, PHI Learning Pvt. Ltd.

Zhu, X. X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F. and Fraundorfer, F. (2017), 'Deep learning in remote sensing: A comprehensive review and list of resources', *IEEE Geoscience and Remote Sensing Magazine* **5**(4), 8–36.

Appendix A

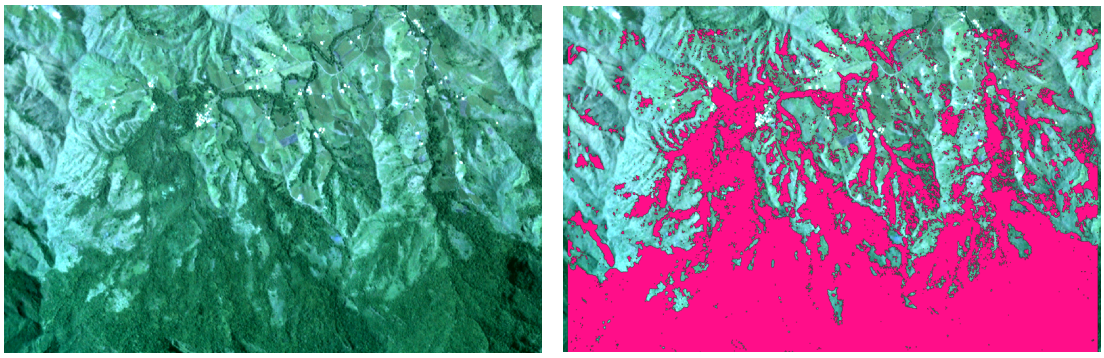
Additional Tree Predictions

Please see below for additional examples of binary mask tree predictions generated by the proposed RF-Ranger model using PlantetScope satellite imagery. Each right side image shows the tree prediction map (pink) for the area shown on the left side image.



(a) PlanetScope Basemap (No Labels) (b) Tree Predictions (Tree=Pink)

Figure A.1: Predictions show trees in urban and agricultural areas.



(a) PlanetScope Basemap (No Labels) (b) Tree Predictions (Tree=Pink)

Figure A.2: Predictions show trees at edge of natural forest region.

Appendix B

Data Preparation

B.1 Planet Basemap Viewer

All PlanetScope basemaps were collected using the Planet Basemap Viewer tool. The online tool allows users to download high-resolution basemap imagery through the NICFI program. To select a basemap, first a user has to select an area of interest, then a GeoTiff file representing that satellite imagery will be downloaded. The Planet Basemap Viewer can be accessed at - https://www.planet.com/basemaps/

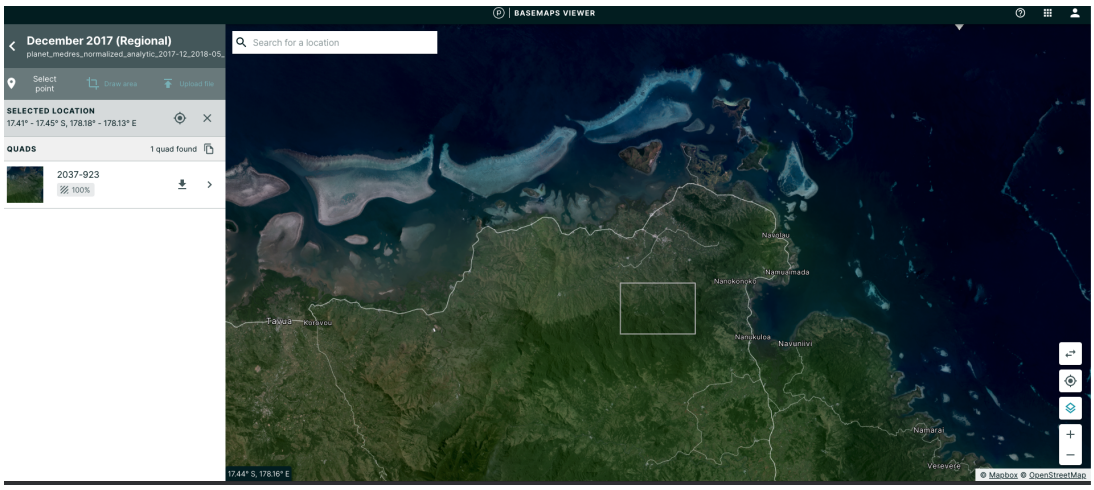


Figure B.1: Example view of downloading basemap satellite imagery.

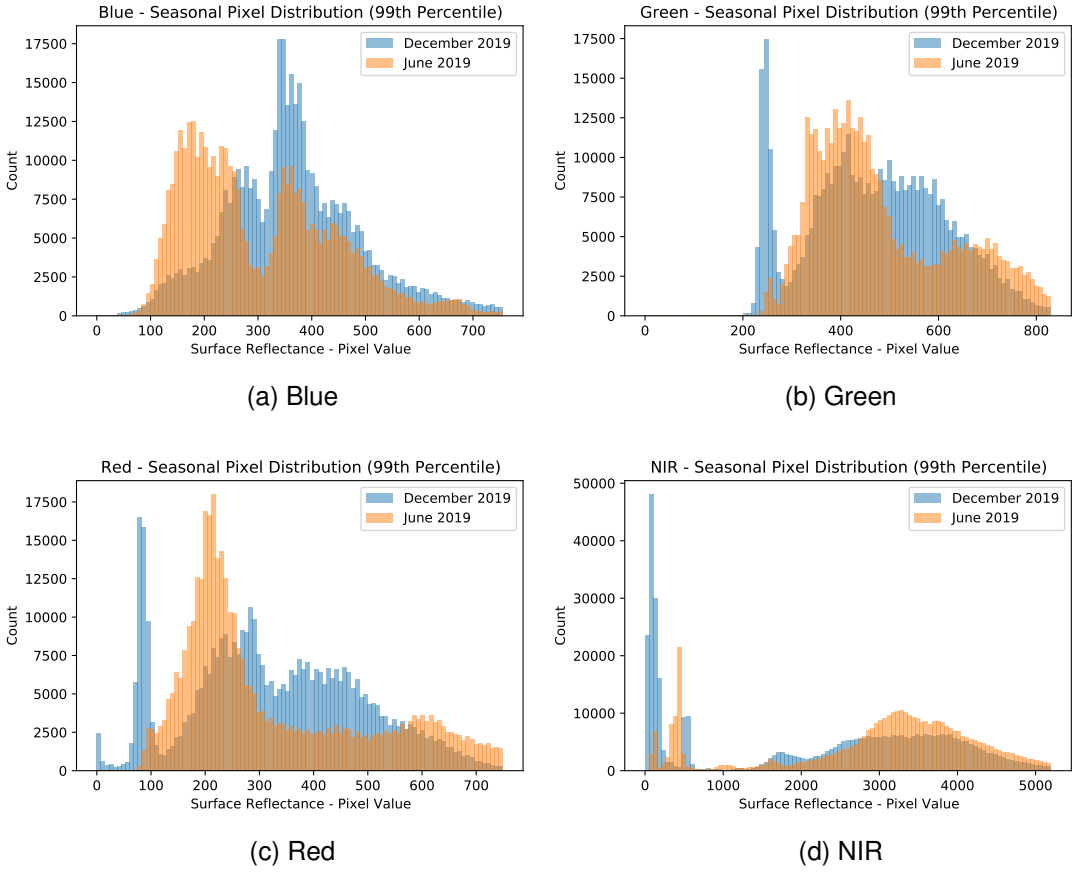
B.2 Nakauvadra vs Edinburgh

It is often difficult to visualise land area. Therefore in Figure B.2, we compare the 11,387 hectares of land area of the Nakauvadra Reforestation Project with the land area of the City of Edinburgh. The polygon and satellite imagery was produced using the measuring tool from Google Earth Pro - https://earth.google.com/web/.

Figure B.2: Example of Nakauvadra forest mountain range area (11,387 hectares) overlaid onto the City of Edinburgh, Scotland for scale.

B.3 Seasonal Pixel Distribution

Figure B.3, shows the pixel distributions of the seasonal datasets conducted for evaluation in this study. The results indicate that a similar pixel distribution is largely maintained across June 2019 and December 2019 labels.



(a) Blue (b) Green

(c) Red (d) NIR

Figure B.3: Multi-Spectral Pixel Distribution for Seasonal Datasets

Appendix C

Code Examples

C.1 Code Submission

All code for this project is shared in a ZIP file during submission. Please read the "ReadMe-Dissertation.pdf" file in the root directory for an explanation of all the Jupyter Notebooks and data files used in this report. See also all relevant code examples in the sub-directory "/dissertation_clean".

C.2 Decision Tree Visualisation

```
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
from matplotlib import pyplot as plt

import pandas as pd

# Get training data
X_train = pd.read_csv('data/train.csv', index_col=0).iloc[:, :4]
y_train = pd.read_csv('data/train.csv', index_col=0).iloc[:, -1]

# Create a single decision tree classifier
# max_depth = 2 for clearer visualisation
dt = DecisionTreeClassifier(random_state=1234, max_depth=2)

# Fit the model to the training data
model = dt.fit(X_train, y_train)

# Print the decision tree features
fig = plt.figure(figsize=(25,20))
_ = tree.plot_tree(dt,
                   feature_names=['blue', 'green', 'red', 'NIR'],
                   class_names=['no tree', 'tree'],
                   filled=True,
                   label='all',
                   proportion=True)
```