Biosecurity New Zealand
Tiakitanga Pūtaiao Aotearoa

MPI 18607 Project Report

Developing improved methods for mapping Metrosideros species in New Zealand

Biosecurity New Zealand Technical Paper No: 2019/23

Prepared for the Ministry for Primary Industries
By Dr Grant Pearse[1], Dr Julia Soewarto[1], Dr Michael Watt[1], Honey Estarija[1]

[1]Scion, New Zealand Forest Research Institute Limited

ISBN No: 978-0-9951272-7-2
ISSN No: 2624-0203

July 2019

New Zealand Government

Ministry for Primary Industries
Manatū Ahu Matua

Disclaimer

While every effort has been made to ensure the information in this publication is accurate, the Ministry for Primary Industries does not accept any responsibility or liability for error of fact, omission, interpretation or opinion that may be present, nor for the consequences of any decisions based on this information.

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries website at http://www.mpi.govt.nz/news-and-resources/publications/

Executive summary

The problem

Existing methods for tree species classification using remote sensing often require LiDAR, multispectral or hyperspectral data. These data can be prohibitively expensive and complex for large-scale species mapping exercises such as those required for biosecurity responses. Traditional methods often rely on the derivation of variables (metrics) that attempt to capture the unique spectral or textural features that make the species distinct. Development of methods using this approach can also be complex, subjective and time-consuming – delaying the usefulness of this approach for rapid biosecurity responses.

The advent of deep convolutional neural networks has revolutionised machine capabilities in object recognition and classification. These models do not rely on spectral or other traits captured through hand-picked variables. Instead, with enough data, they can learn the hard-to-quantify features and appearance that make objects visually distinctive in low-cost imagery, achieving state-of-the-art accuracies across a range of domains. Application of these methods to classify species from high-resolution aerial imagery remains largely unexplored despite many species being visually distinct and relatively easy for trained observers to identify in this type of imagery.

This project

The objectives of this study were to trial rapid, accurate tree species classification from low-cost aerial imagery using deep convolutional neural networks. These deep learning models were compared to traditional approaches relying on textural and spectral indices input to XGBoost classification algorithms.

We demonstrate a novel approach to train these two state-of-the-art classifiers to detect myrtle rust host species in the *Metrosideros* genus (*Metrosideros excelsa* in particular) across Tauranga. We combined the strong phenology of this species (extensive flowering in the summer) with inspection records collected during the myrtle rust response to develop an extensive dataset to address three key objectives:

1. To test the classifiers on aerial imagery (10 cm resolution, three-band RGB) utilising the strong phenology of the target species (distinctive and extensive flowering in summer).
2. To test the classifiers on historical imagery of the same type and over the same area but without the assistance of strong phenology (little to no flowering observed).
3. To test the classifiers on a combined dataset containing imagery from different years showing a mixture of flowering and nonflowering trees to simulate conditions that might be encountered in a real-world mapping exercise where phenology and available data sources will vary.

Key results

The classifiers trained to recognise *Metrosideros* spp. using the strong and distinctive flowering both performed well. The XGBoost model using the spectral and textural traits achieved an accuracy of 88.1% driven by the strong red flowers present on many canopies. In the absence of strong phenology, the performance fell to 80.0% and the models relied instead on the blueish hue and texture of the canopies in relation to other species. The accuracy for the combined dataset was 84.9% and leveraged a combination of features important in the previously described models.

The deep learning model produced a near-perfect classification when using the strong phenology (99.7% accuracy) even though the level of flowering varied substantially. In the absence of phenology, the deep learning model accuracy remained high at 95.3% accuracy despite significant variation in the appearance of the *Metrosideros* spp. and degraded imagery quality. The performance of the combined model was similarly high at 98.1% accuracy. This suggests the potential to generalise well to real-world, large-scale biosecurity mapping applications where phenology and available data sources will vary.

Implications and future work

Deep learning models show substantial potential for rapid, large-scale tree species classification using low-cost aerial imagery. Species phenology can be leveraged to help develop the substantial datasets required to train these models. The demonstrated approach should be expanded to test classification

of multiple additional species. As a general rule, species that can be positively identified in aerial imagery by experienced analysts may be good candidates. Alternative deep learning frameworks that have the ability to simultaneously find and classify multiple targets (canopies) within large images should be explored.

Contents

1 Project background

To better understand myrtle rust and limit its impact in New Zealand, the Ministry for Primary Industries commissioned a comprehensive research programme in 2017 with more than 20 projects valued at over $3.7 million. Projects in this programme were completed by June 2019.

The projects covered research in the following themes:

- Theme 1 - Understanding the pathogen, hosts, and environmental influence.
- Theme 2 – Building engagement and social licence: Improved understanding of public perceptions and behaviours to allow better decisions about investment, improved design of pathway control strategies and maintain social license for use of management tools.
- Theme 3 – Te Ao Māori: Greater understanding of Te Ao Māori implications of myrtle rust in order to support more effective investments, and improved use of Mātauranga, specific Māori knowledge, and kaupapa Māori approaches in management regimes.
- Theme 4 – Improving management tools and approaches: Improved diagnostic and surveillance speed, accuracy and cost-effectiveness, supporting eradication efforts and enabling scaling up of surveillance efforts for a given resource. More effective treatment toolkits to avoid emergences of MR resistance to treatments and to enable disease control over increasingly large scales that will lead to reduced or avoided impacts.
- Theme 5 - Evaluating impacts and responses: Improved understanding of environmental, economic, social and cultural, impacts to inform risk assessment and management and to communicate implications to decision/makers and stakeholders.

This report is part of the MPI commissioned research under contract MPI18607 which addressed research questions within Theme 2, 4 and 5.

Text in the report may refer to other research programmes carried out under the respective theme titles.

2 Introduction

Background

The early stages of a biosecurity response to a newly arrived plant pathogen can have a significant bearing on the final outcome and cost (Goldson et al., 2015; Kriticos, Phillips, Suckling, & others, 2005). Once an unwanted pathogen has been positively identified, mapping and identification of potential host species become essential for managing the incursion (Kalaris et al., 2014). Identification of host plants must be carried out by trained personnel and the hosts may be located across a mixture of public and private property or in hard to access areas. Hosts may also be concealed by other vegetation or cryptic in appearance. For these reasons, carrying out large-scale searches for host plants can be very costly and challenging to resource.

The level of host detection and surveillance required in the face of an incursion is usually defined by the response plan. For example, planning to eradicate the pathogen necessitates exhaustive detection of host species to monitor spread and enable destruction of infected plants and, in some cases, potential hosts showing no signs of infection to limit future spread. Attempts to manage or monitor the spread of a pathogen may require only identification of key indicator species to define the infection front and to monitor the long-term ecosystem impacts. Finally, long-term management strategies for newly established pathogens may require large-scale but inexhaustive host identification to identify resistant individuals within a population for breeding programmes or other approaches to biological control (DiTomaso et al., 2017; Mundt, 2014).

Remote sensing offers one approach to identify host species that is potentially cost-saving and efficient for large-scale applications (Asner et al., 2018; Huang & Asner, 2009). Imagery acquired from UAVs, aircraft or even space-borne optical sensors can be used to identify both potential hosts as well as the symptoms of pathogen infection on susceptible host species (Asner et al., 2018; Y. He, Chen, Potter, & Meentemeyer, 2019). However, detection and classification of species from remotely sensed data is a complex sub-discipline. Fassnacht et al. (2016) carried out a comprehensive review of the current state-of-the-art for tree species classification using remotely sensed data. The approaches reviewed encompassed a diverse range of data sources and methodologies with varying degrees of accuracy reported. however, clear themes were evident in the literature. Multispectral and hyperspectral data were highlighted as being the most useful data sources for accurate species classification with LiDAR data being highly complementary. Multi/hyperspectral data sources observe reflected light outside of the visible light spectrum with the increased chance of observing patterns of reflectance related to structural or biochemical traits that may be unique or distinctive to species or groups.

Multispectral data (4-12 bands) is relatively easy to capture and has been widely used in combination with machine learning methods such as random forests and support vector machines to perform species classification (Dash, Watt, Pearse, & Dungey, 2017; Ferreira, Wagner, Aragão, Shimabukuro, & de Souza Filho, 2019). However, accurate classification is often limited to broad groups such as conifer vs. deciduous forest types (Ballanti, Blesius, Hines, & Kruse, 2016). Hyperspectral data contains many more (>12) narrow spectral bands – enhancing the ability to observe small differences that may be present between the spectra of tree species. This approach has been particularly well studied for fine-grained species classification tasks (Clark, Roberts, & Clark, 2005; M. Dalponte, H. O. Ørka, T. Gobakken, D. Gianelle, & E. Næsset, 2013). The idea of unique spectral ‘signatures’ for species has been present in the literature for several decades; however, Fassnacht et al. (2016) conclude that these signatures appear to be rare in practice and, when present, require observation of a wide portion of the spectrum using sophisticated sensors (Hesketh & Sánchez-Azofeifa, 2012). Regardless of these challenges, hyperspectral data has been successfully used to classify as many as 42 species and application of these sophisticated methods has allowed large-scale host species identification (Asner et al., 2018; Fassnacht et al., 2016).

These large-scale applications of hyperspectral-based species classification face challenges related to practicality and cost. The increased spectral resolution usually demands very careful acquisition from expensive sensors that can only be operated under a narrow set of conditions. The post-processing of these data can also be complex and requires careful correction of atmospheric impacts and noise reduction. Finally, the substantial volumes of data must often be subjected to dimensionality reduction before analysis can proceed (Ballanti et al., 2016; J. M. Bioucas-Dias et al., 2013).

Spectral-based methods reduce the problem of species identification to discriminating patterns in the reflected light. Hence these methods require high-quality data calibrated to reflectance rather than radiance otherwise the classifiers developed can translate poorly to new areas or data sources (Bannari, Morin, Bonn, & Huete, 1995). Other information content such as the structure, shape, texture and other distinctive but hard to quantify characteristics are often only partially utilised or completely neglected. Some efforts to characterise the texture or the shape of the crown or other attributes have been made but these typically rely on deriving a small number of variables (metrics) that attempt to summarise complex properties (Ballanti et al., 2016; Ferreira et al., 2019).

In contrast, the human visual system allows experienced individuals to distinguish a large number of species by visual inspection alone. Some cryptic species remain hard to tell apart visually, but trained experts (and even non-experts) can discriminate a surprising number of species (Lacerda & Nimmo, 2010). This has led to the development of systems that allow interested members of the public to upload high-resolution imagery taken from consumer cameras or smartphones to platforms where experts can review and identify species within their areas of expertise such as iNaturalist (Van Horn et al., 2018).

Recently, the advent of deep convolutional networks has transformed the capability of machines to perform fine-grained classification of images, often reaching or exceeding human-level accuracy (De Fauw et al., 2018; Krizhevsky, Sutskever, & Hinton, 2012). These advances have been extended to include fine-grained species classification from simple imagery acquired from consumer cameras or smartphones (Affouard, Goëau, Bonnet, Lombardo, & Joly, 2017; Cui, Song, Sun, Howard, & Belongie, 2018). These 'deep learning' approaches utilise artificial neural networks consisting of many layers with different functions that operate on the input imagery. The architecture of these networks, combined with efficient training algorithms, allows these networks to effectively learn the mappings between potentially very large numbers of possible input images and output classes (Mallat, 2016). In this approach, the features important for classification are not engineered or pre-selected but rather learned by the network from labelled training examples and the method relies comparatively little on pre-processing of the imagery. In practice, these approaches have demonstrated a remarkable capacity to perform fine-grained classification from images that are largely unsuitable for classical remote sensing methods. However, they can demand a substantial number of training examples to enable the network to learn effectively.

Examples of this technology for host species detection remain rare in the literature and rarer still are examples using methods for large-scale host species detection from data such as aerial imagery. This data is simple – usually consisting of only RGB colour channels - and is typically delivered as uncalibrated radiance values rather than being processed to represent the fraction of light reflected from the surface (bottom of atmosphere reflectance). The simplicity of the data means that is it practical to regularly capture large areas at high-resolution (<10 cm) for a relatively low cost. While aerial imagery cannot be used to identify small or occluded individuals, it offers advantages in terms of scale and largely bypasses issues such as access and permissions facing ground surveillance teams. Applying deep learning to develop methods for large-scale host species identification using readily available and low-cost imagery has the potential to support future biosecurity eradication and management efforts.

Research Context

The fungal plant pathogen *Austropuccinia psidii* (myrtle rust) affects a broad range of hosts in the *Myrtaceae* family, causing lesions, dieback and, in some cases, mortality (Carnegie et al., 2016; Glen, Alfenas, Zauza, Wingfield, & Mohammed, 2007). The pathogen is airborne and has spread rapidly around the globe (Carnegie & Cooper, 2011; Coutinho, Wingfield, Alfenas, & Crous, 1998; McTaggart et al., 2016; Roux, Greyling, Coutinho, Verleur, & Wingfield, 2013). New Zealand is home to at least 37 native myrtaceous species (Lange et al., 2018) and many more exotic members of the family such as *Eucalyptus* spp. have been introduced for amenity and production values. Within this family, 12 species of the *Metrosideros* genera are present in New Zealand. Of these, pōhutukawa (*Metrosideros excelsa*) is especially highly valued culturally and has been widely planted for amenity purposes. This coastal evergreen tree has a sprawling habit of up to 20 m and produces dense masses of red flowers over the Christmas period (Allan, 1961) – earning it the name 'the New Zealand Christmas tree'. Observations from pōhutukawa growing in other countries where myrtle rust is present indicate that the species is susceptible to myrtle rust (Loope, 2010; Sandhu & Park, 2013).

In May 2017, myrtle rust was detected on the New Zealand mainland for the first time (Ho et al., 2019). The disease has spread rapidly to the majority of at-risk regions and has established on numerous native and exotic host species (Beresford et al., 2018). The arrival of the disease triggered a major biosecurity response involving a broad range of public and private institutions. In the early stages of the response, an intensive effort was made to identify and inspect host species around the sites of infection. Host identification continued as the pathogen spread to new regions and has been ongoing as the management plan transitioned from eradication to long-term management. These efforts produced a substantial volume of surveillance data including GPS locations and positive identification of *Metrosideros* spp. by trained inspectors.

This report presents the results of a study carried out to test novel methods suitable for large-scale identification and mapping of key host species (*Metrosideros* spp.) of myrtle rust. To achieve this, we focused on three key objectives:

- **Objective 1:** Test two state-of-the-art classification methods (XGBoost and deep convolutional neural networks) applied to 3-band aerial imagery by leveraging the strong phenology of the target species (distinctive and extensive flowering in summer).

- **Objective 2:** Test classification of the identical trees using the same type of imagery but without the assistance of phenology.
- **Objective 3:** Test the practicality and generalisability of these techniques in real-world conditions by creating a combined dataset from Objectives 1 & 2 that contained imagery from different years and sources and a mixture of flowering and non-flowering trees.

3 Materials and methods

3.1 Ground truth data

New Zealand maintains an extensive biosecurity surveillance system including regularly monitored sites across the country and an established incursion response protocol. In the first months after the incursion of myrtle rust, the response was centred on positively identified cases at the monitored surveillance sites and locations reported through a domestic biosecurity network (e.g. plant producers). Once a site was confirmed infected, intensive ground surveillance efforts were carried out to identify and inspect all potential host species within a fixed radius from the infected site. New, confirmed infections triggered additional searches around the new site and, in some cases, at-risk areas would be intensively surveyed. As the response continued, the surveillance efforts expanded in order to monitor the spread of the disease within regions and across the country.

The surveillance response, including collation and management of the GIS data and inspection records, was coordinated by AsureQuality New Zealand on behalf of the Ministry for Primary Industries. A mobile app was used to log each inspection and to embed the GPS coordinates of the observer. Trained observers recorded the confirmed genus, location, infection status and other key information for every inspection carried out. By May 2019, surveillance inspections had been carried out on *Metrosideros* spp. across the country, with most of the effort concentrated in the higher-risk North Island regions. The database included several thousand inspections carried out in urban areas such as Tauranga. Inspection of the ground records quickly revealed that many the records for *Metrosideros* spp. were pōhutukawa. Large numbers of these trees are present within Tauranga and many were located and inspected between 2017-2019 (Figure 1).

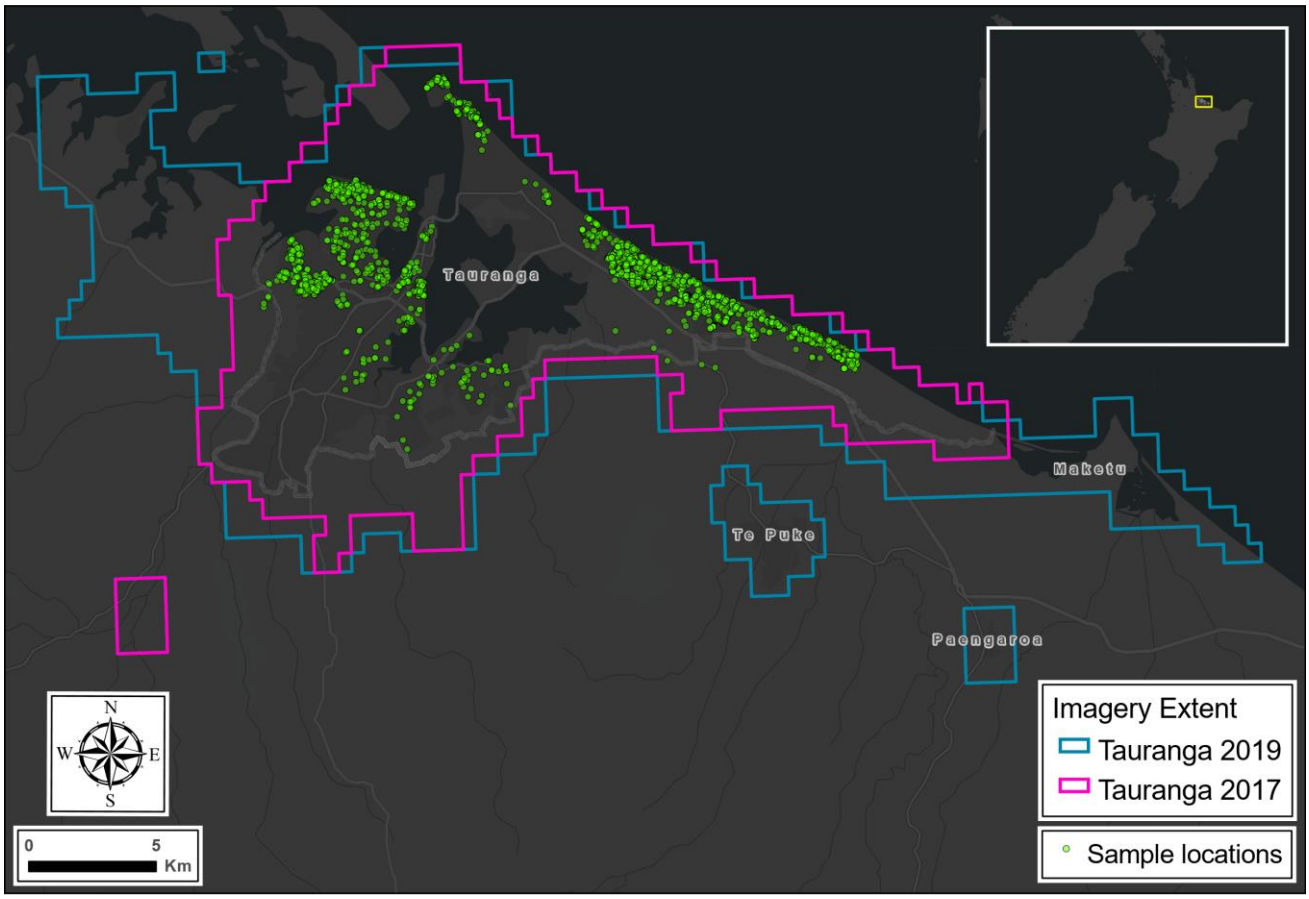

Figure 1. The extent of aerial imagery datasets and location of sample trees around Tauranga in the Bay of Plenty, New Zealand.

The ground surveillance records from Tauranga were used to seed the creation of a substantial dataset capable of addressing the study objectives. First, we focused on developing a dataset to test classification leveraging the phenology of the species. The extensive, distinctive red flowers that are easily identifiable from above made pōhutukawa good candidates for aerial mapping during the summer. For much of the year, either the buds, flowers, or seed capsules are present and visible. However, the multi-leader crown shape is also

somewhat distinctive as is the texture and blueish hue of the large waxy and elliptical leaves that are present all year round (Figure 2).

Figure 2: Images of pōhutukawa trees illustrating the distinctive features such as multi-stem form, leaf colour and texture and extensive buds and flowers usually present in summer. The bottom row shows aerial views from very high-resolution UAV imagery. Image credits: Julia Soewarto and Scion Research.

Aerial imagery captured over Tauranga during the 2018-2019 Christmas period (Table 1) was overlaid with the ground surveillance locations from the same region. Inspection locations were collected using consumer grade GPS units and could only be considered approximate. For each inspection record, a trained analyst examined the GPS point and imagery and identified the corresponding tree. If the tree was positively identified and showed at least some evidence of flowering, then the imagery was annotated by delineating a tight bounding box around the canopy extent. The distinctive features of the canopy and strong flowering observed in the imagery greatly assisted identification and annotation; however, inspection records were only at the genus level and other species with similar phenological traits such as rātā (*Metrosideros robusta*) may also be

found within this region. In addition, some cultivated 'yellow' pōhutukawa (*Metrosideros excelsa* 'Aurea') appeared to be present within the dataset but these were removed due to the small number of samples available. For fine-grained classification, these similar species would have posed a challenge; however, for the purposes of this study, the similar species noted above are also host species for myrtle rust – making misclassification less of an issue. Nonetheless, we attempted to assess the purity of the training dataset by inspecting a large subset of trees using publicly available, street-level imagery followed by physical on-site inspections and other referencing other records (e.g. iNaturalist) for a smaller subset of trees. The results showed that the vast majority of the study trees identified by combining the aerial imagery and surveillance records were unambiguously pōhutukawa. This process was laborious but ensured the purity of the dataset for model development and confirmed that the phenology and other visual traits of the species could reliably be used by a human analyst to identify new examples from aerial imagery. After this stage, we considered that the assembled training dataset consisted of only pōhutukawa and any misclassifications would have been small in number. Nonetheless, because every individual could not be verified we refer to these as *Metrosideros* spp. throughout the report.

Table 1: Imagery datasets.

Imagery Dataset	Phenology	Resolution, Colour channels
Tauranga – summer 2018-2019	Wide-spread flowering	10 cm/pixel, 3-band RGB
Tauranga – March 2017	Few flowering	10 cm/pixel, 3-band RGB

Development of the classifiers also required negative examples. The candidate negative examples were any tree other than *Metrosideros* spp. We once again leveraged the ground inspection efforts to develop this dataset. The intensity of the initial surveillance efforts meant that within inspected areas such as streets or parks the locations for nearly every *Metrosideros* spp. were recorded. We used these areas to select negative examples and cross-referenced a substantial portion of the dataset against other imagery and some field inspection. This approach reduced the chances of accidentally including *Metrosideros* spp. or biasing the training set by excluding visually similar non-*Metrosideros* species due to uncertainty. In addition, this provided a realistic set of non-target tree canopies which the classifier might encounter in the areas surveyed for the biosecurity response. Tight bounding boxes around the canopies were defined against the aerial imagery and annotation proceeded until the dataset was balanced. Examples of typical and atypical pōhutukawa and other tree species canopies as seen in the imagery are shown in Figure 3.

3.2 Imagery datasets

The aerial imagery datasets described in Table 1 were tiled together into separate orthomosaics. Aside from some minor issues, the orthomosaics were of high quality as they were captured, and quality checked for operational use by local authorities. The imagery from 2017 showed slightly lower levels of detail (likely because of poorer image matching) resulting in the trees appearing blurrier with less visible detail (Figure 3). The spatial coordinates of the bounding boxes were used to extract sub-images from the larger orthomosaics and each image 'chip' was labelled with the dataset year and class (*Metrosideros* spp. or other spp.). All images were visually checked and only those images where the same trees were present in both 2017 and 2019 were included in the dataset. Very small trees (canopy radius < ~ 1.5m) had to be excluded as these canopies were represented by too few pixels in the 10 cm resolution aerial images. The final datasets included 2278 images of tree canopies evenly split between *Metrosideros* spp. and other spp. with images available for both 2017 and 2019.

Generalisability is the goal of any species classification model. This is especially true for biosecurity applications where accurate detection is required for effective incursion management. To validate the models, the datasets were randomly split into training (75%) and validation (25%) sets. Model training was carried out using only the images in the training sets and the validation sets were completely isolated from model development, only being used at test time to evaluate final model performance. This approach was extended by forming a combined dataset that included imagery from both the 2019 (with phenology) and 2017 (without phenology) datasets (Table 2). The combined dataset can be thought of as an attempt to test the technique in real-world conditions. Expression of phenological traits such as flowering is inherently variable and irregular in nature. Pōhutukawa can switch from flowering to vegetative growth only, making flowering alone an unreliable characteristic. In addition, large-scale mapping for biosecurity responses or other purposes may require the use of aerial imagery datasets from different sources or seasons.

Figure 3: Examples of canopy images used to train classification models. *Metrosideros* spp. canopies from the 2019 imagery are shown in panel a). The same canopies from the 2017 imagery are shown in panel b). Examples of non-*Metrosideros* spp. including some harder examples are shown in panel c) while the same canopies from the 2017 imagery are shown alongside in panel d).

Table 2: Summary of datasets used to train and validate classification models.

Dataset	Description	Number of Trees (*Metrosideros* spp./ Other spp.)	Training / Validation
Tauranga 2019	Test classification using phenology	2278 (1139/1139)	1709/569 (75/25%)
Tauranga 2017	Test classification without phenology	2278 (1139/1139)	1709/569 (75/25%)
Tauranga 2017 & 2019 combined	Test combined classification with and without phenology	4556 (2278/2278)	3417/1139 (75/25%)

3.3 Deep learning models

Various candidate deep learning architectures were evaluated for the classification tasks. A 50-layer residual neural network was identified as offering the best compromise between training time and accuracy. Residual networks are a biologically inspired enhancement to deep neural network architectures that allow a much greater number of network layers to be utilised, ultimately producing substantial gains in accuracy on many image classification tasks (K. He, Zhang, Ren, & Sun, 2015, 2016; Veit, Wilber, & Belongie, 2016).

Learning rate is a crucial hyperparameter in deep learning model development that can dramatically impact model performance. We used an iterative approach to select a learning rate that balanced convergence and stability for each model developed against the number of epochs required to descend to the optimum network state. Initially, only the final, connected layers of the models were trained on the tree canopy imagery with a relatively high learning rate ($3x10^{-3}$ – $1x10^{-4}$) for a smaller number of epochs (iterations) (10 – 15). Thereafter, we fine-tuned deeper layers of the network using a reduced learning rate ($1x10^{-5}$-$1x10^{-6}$) applied for a greater number of epochs (15 – 30). At this point, indicators of performance for the models showed no further decrease in model error using the internal data. All deep learning models were implementation using the PyTorch 1.0.0 deep learning library (Paszke et al., 2017) and all training was carried out on an Nvidia Tesla K100 GPU with 12GB of onboard memory.

3.4 XGBoost models

For the deep learning models, the canopy images required no pre-processing or feature generation and could be used directly in the model development pipeline. The architecture of the model is left to determine important features and characteristics within the imagery. However, for XGBoost or other more traditional approaches, the imagery must be used to generate variables (metrics) that capture features or characteristics that can be used to discrimination different species (Ballanti et al., 2016; Fassnacht et al., 2016). These variables often take the form of vegetation indices computed from specific wavelengths; however, very few indices can be computed from RGB imagery and fewer still can be reliably computed without converting the imagery from radiance (raw brightness values) to reflectance values representing the fraction of incoming light reflected. Reflectance conversion requires calibration and atmospheric data that are usually not available during aerial surveys that can span months. We carefully chose a range of variables targeting the distinctive properties of the *Metrosideros* spp. canopies. These included spectral metrics aimed at capturing the blueish hue of the leathery, elliptical leaves and the strong and distinctive sprays of flowers and buds present in the canopies during summer. The canopies also exhibit distinctive textural properties arising from the multi-stem structure and leaf and bud arrangements independent of the presence or absence of flowers (Figures 2 & 3). Texture analysis using grey-level co-occurrence matrices (GLCMs) (Haralick & Shanmugam, 1973) was carried out to try and capture these characteristics within predictive variables. Computation and extraction of the texture images were carried out separately using the 'glcm' package (Zvoleff, 2019) in the R statistical software environment (R Core Team, 2019). Selection of the GLCM metric classes and parameters was carried out using the guidelines and methods proposed by Hall-Beyer (Hall-Beyer, 2017). The raw digital numbers (pixel values) within each defined bounding box were used to generate patch-level mean values for the predictive metrics detailed in Table 3.

The values for every labelled tree canopy were then inputted into the XGBoost algorithm and the algorithm was trained for a maximum of 400 iterations (rounds), with early stopping parameters set to prevent over-fitting. Subsampling of variables and observations for individual tree learners was also implemented alongside fine-tuning of the *gamma* hyperparameter to further guard against over-fitting.

Table 3: Vegetation Indices and metrics computed from 3-band (RGB) aerial imagery of tree canopies for use in the XGBoost classification model. All metrics used the raw image digital numbers (DNs) (0-255) from the input pixels.

Variable Name	Description	Definition	Source
Mean_red	Mean of red channel DNs	$\frac{\sum Red}{Num\ Red}$	NA
Mean_green	Mean of green channel DNs	$\frac{\sum Green}{Num\ Green}$	NA
Mean_blue	Mean of blue channel DNs	$\frac{\sum Blue}{Num\ Blue}$	NA
Sd_red	Standard deviation of red channel DNs	$\sqrt{\frac{\sum(Red - Mean\ Red)^2}{Num\ Red - 1}}$	NA
Sd_green	Standard deviation of green channel DNs	$\sqrt{\frac{\sum(Green - Mean\ Green)^2}{Num\ Green - 1}}$	NA
Sd_blue	Standard deviation of blue channel DNs	$\sqrt{\frac{\sum(Blue - Mean\ Blue)^2}{Num\ Blue - 1}}$	NA
Rg_ratio	Red Green Ratio Index	$\frac{Red}{Green}$	Gamon & Surfus (1999)
Normdiff_rg	Normalised difference red/green ratio	$\frac{Red - Green}{Red + Green}$	NA
Scaled_red	Scaled red ratio	$\frac{Red}{(Red + Green + Blue)}$	NA
Scaled_green	Scaled green ratio	$\frac{Green}{(Red + Green + Blue)}$	NA
Scaled_blue	Scaled blue ratio	$\frac{Blue}{(Red + Green + Blue)}$	NA
Sd_gi	Std. deviation of the scaled green index	$\sqrt{\frac{\sum(SGreen - Mean\ SGreen)^2}{Num\ SGreen - 1}}$	NA
glcm_correlation	Textural metric computed on RGB channels	Grey-level co-occurrence correlation	Haralick & Shanmugam (1973)
glcm_homogeneity	Textural metric computed on RGB channels	Grey-level co-occurrence homogeneity	Haralick & Shanmugam (1973)
glcm_mean	Textural metric computed on RGB channels	Grey-level co-occurrence mean	Haralick & Shanmugam (1973)
glcm_entropy	Textural metric computed on RGB channels	Grey-level co-occurrence entropy	Haralick & Shanmugam (1973)

3.5 Performance metrics

The predictions made by the two modelling approaches on the withheld portion of the 3 different datasets were used to compute the number of true positives (correct classifications) and false positives (incorrect classifications) for the two classes (*Metrosideros* spp. and other spp.). These values were then used to compute the classification performance metrics shown in Table 4.

Table 4: Performance metrics used to assess classification models. TP=true positive, FP = false positive, TN = true negative, FN = false negative.

Metric	Description	Definition
Accuracy	A measure of how often the classifier's predictions were correct.	$\frac{TP + TN}{TP + FP + TN + FN}$
Error	A measure of how often the classifier's predictions were wrong.	$1 - Accuracy$
Cohen's kappa	A measure of a classifier's prediction accuracy that accounts for chance agreement.	$\frac{observedAgreement - chanceAgreemeent}{1 - changeAgreement}$
Precision (Positive predictive value)	A measure of the proportion of positive predictions that were correct.	$\frac{TP}{TP + FP}$
Sensitivity (Recall)	The proportion of actual positives (*Metrosideros*) that were correctly identified by the classifier.	$\frac{TP}{TP + FN}$
Specificity	The proportion of actual negatives (not-*Metrosideros*) that were correctly identified by the classifier.	$\frac{TN}{TN + FP}$

4 Results and discussion

The results from the XGBoost and deep learning models applied to the withheld portions of the datasets used to test classification with phenology (2019 imagery), without phenology (2017 imagery) and classification of the combined datasets are shown in Table 5.

Table 5: Classification results obtained by applying the trained models to the withheld portion of each dataset.

	Classification with phenology	Classification without using phenology	Classification of combined datasets with and without phenology
XGBoost			
Accuracy	88.1%	80.0%	84.9%
Error	11.9%	20.0%	15.1%
kappa	0.762	0.600	0.698
Precision (PPV)	0.944	0.818	0.862
Sensitivity	0.814	0.779	0.832
Specificity	0.950	0.822	0.866
Deep Learning			
Accuracy	99.7%	95.3%	98.1%
Error	0.3%	4.7%	1.9%
kappa	0.993	0.905	0.961
Precision	0.997	0.939	0.974
Sensitivity	0.997	0.969	0.988
Specificity	0.996	0.936	0.974

The XGBoost classifiers showed high levels of accuracy for all three datasets. The strong phenological traits of the *Metrosideros* species (nearly all pōhutukawa in this study) in summer greatly improved the classification accuracy, with the model improving to 88.1% compared to the dataset without phenology (80.0%). Examination of the variable importance scores extracted from the models showed that the *rg_ratio* metric capturing the ratio of red to green pixels had the highest importance in the model utilising phenology (Figure 4). This metric, and the scaled green and blue metrics, likely responded to the extensive showers of red flowers present on many of the trees at this time. Examining the misclassified canopies from this model showed that many of the misclassified *Metrosideros* spp. canopies showed limited flowering or were small with more background visible (Figure 5, a) while the canopies of other spp. incorrectly classified as *Metrosideros* spp. were often reddish or dark in colour (Figure 5, b). This suggests that variables capturing the strong flowering patterns drove the high performance of this model but provided poor ability to separate other species with reddish or darker canopies.

In the absence of the most distinctive phenological traits, the XGBoost model performance decreased and the model relied heavily on the scaled red values (*scaled_red*) and textural metrics computed from the blue channel (Figure 4). In the 2017 data, there were several hundred non-*Metrosideros* tree canopies with reddish hues while many of the remaining canopies were somewhat greener than the blueish-green appearance of the *Metrosideros* species. The combination of these metrics appeared to allow the model to separate out the *Metrosideros* spp. with a relatively high level of accuracy. The misclassified canopies varied considerably, making it harder to understand the causes of these failures. Several misclassified *Metrosideros* spp. appeared to show buds or seed capsules, altering their appearance from the other examples (Figure 5, c). Examples of other species misclassified as *Metrosideros* spp. included a range of different canopy types, but some examples appeared to have darker green canopies, similar to many of the *Metrosideros* spp. canopies (Figure 5, d).

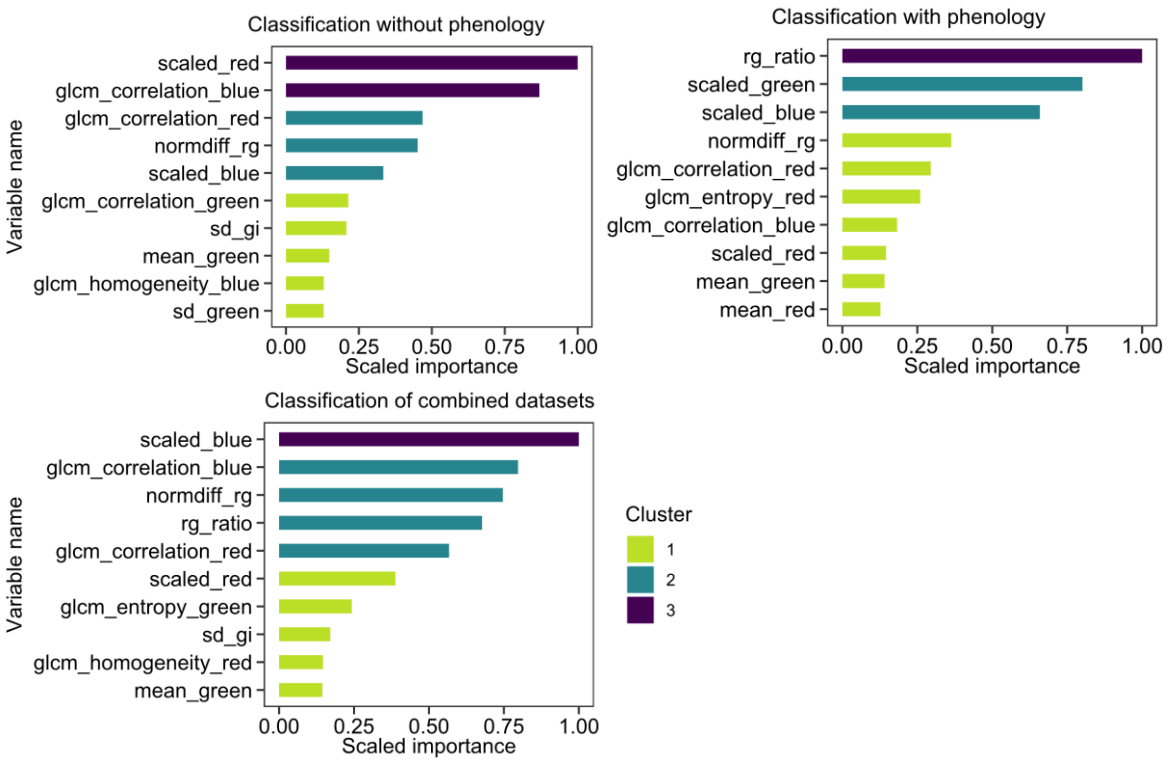

Figure 4: Plots of scaled variable importance metrics from the XGBoost classification models. Higher scores indicated features important to the accuracy of the respective models.

For the combined datasets that included both flowering and non-flowering *Metrosideros* spp, the model accuracy increased modestly to 84.9%. This suggests that the model gained little additional advantage from the doubling of the available training data as the performance was roughly the average of the two separate models. The combined model had a greater number of variables with high importance and the variables selected were largely a combination of the variables highlighted in the separate models (Figure 4). Examining the misclassified canopies from the combined XGBoost model, it was clear that 2017 imagery without phenology caused the greatest number of misclassifications. Over two-thirds of the misclassified canopies were from 2017 and some of the false negatives were due to shadows but others were reasonably easy to identify visually (Figure 5,e). Some of the false positives were distinctive from *Metrosideros* spp. but shared a reddish hue (Figure 5, f).

Figure 5: a) Examples of *Metrosideros* spp. canopies incorrectly classified as other species (false negatives) by the XGBoost classifier using the 2019 imagery (with phenology). b) Examples of other spp. canopies incorrectly classified as *Metrosideros* spp. (false positives) by the XGBoost classifier using the 2019 imagery. c) False negatives from the 2017 imagery (without phenology). d) False positives from the 2017 imagery. False positives e) and false negatives f) from the combined XGBoost classifier. The bottom row shows the false negative g) and false positive h) from the 2019 deep learning classifier alongside the false negatives i) and false positives j) from the 2017 imagery.

The deep learning models performed substantially better than the XGBoost models on all three datasets. The classifier developed using the dataset with phenology made only 2 mistakes out of 569 test canopies (Table 5) achieving an accuracy of 99.7%. The 'loss' values associated with each canopy image allow them to be ranked according to how difficult the model considered classification with a high loss associated with misclassified or difficult (lower probability score) canopies. The single false negative was of a *Metrosideros* spp. canopy with few flowers (Figure 5 g) while the single false-positive image was of a unique canopy not present in the training dataset (Figure 5, h). The feature discovery process for deep learning is intrinsic to the model but the classification performance

indicated that the model was highly effective at discriminating flowering trees from other species with reddish canopies or other flowering species present in the data. Without using the strong flowering, the classifier's performance dropped marginally to 95.3% (Table 5). The false positives often displayed similar colour and canopy shape to the *Metrosideros* spp. canopies. The classifier also appeared to struggle more with the canopies affected by the lower quality of the imagery – small, blurry canopies without the characteristic appearance visible in other images frequently appeared in the misclassified images (Figure 5, i). The model trained on the combined dataset (with and without phenology) showed high accuracy (98.1%). All of the misclassified canopies were from the 2017 dataset and once again these images showed blurry and indistinct features relative to other correctly classified examples.

5 Discussion

This study demonstrated that deep learning algorithms could detect *Metrosideros* spp. (nearly all pōhutukawa) in the study area with a very high level of accuracy using only 3-band (RGB) aerial imagery – with or without the use of phenology to enhance detection. Existing approaches to tree species classification from remotely sensed data rely extensively on the use of calibrated multi or hyperspectral data that can be expensive and complex to capture over larger areas (Ballanti et al., 2016; Fassnacht et al., 2016; J. M. Bioucas-Dias et al., 2013). In contrast, simple 3-band aerial imagery is routinely captured over large areas, with many areas in New Zealand currently covered by this class of data (LINZ, 2019). Our results suggest that these data may be suitable inputs to a large-scale mapping campaign of this important genus which is threatened by myrtle rust.

The phenology of tree species has been widely used to enhance remote sensing classification in previous research (Dymond, Mladenoff, & Radeloff, 2002; Wolter, Mladenoff, Host, & Crow, 1995). However, it is still rare in the literature to attempt classification using only 3-band imagery – with or without phenology (Fassnacht et al., 2016). This is primarily because this type of imagery lacks the spectral bandwidth required by traditional methods to discriminate species and the few indices that can be derived are not widely generalisable as the imagery represents sensor radiance (recorded as digital numbers) rather than being standardised to the fraction of incident light reflected from the canopy surface (surface reflectance). To overcome this limitation, we derived features such as textural metrics and simple band ratios aimed at capturing the bright-red, extensive flowering of these species and the characteristic blueish hue and textural properties of the canopies. This appeared successful when examining the results from the XGBoost classifier as the model leveraging phenology performed well on the validation set. These features were less useful in the absence of extensive flowering and the XGBoost model did not perform as well. The performance was still high relative to other examples in the literature. For example, Pham et al. (2016) et al. achieved 68.3% classification accuracy of pōhutukawa using multispectral satellite data from the Coromandel region and similar machine learning methods. The addition of LiDAR-derived features improved this result to 81.7% but pōhutukawa were noted to be more difficult to detect than several other species targeted in that study. It is likely that having multispectral and LiDAR data would have further improved the XGBoost results in our study, but this would come with higher costs for data acquisition, storage and processing.

Using the distinctive phenology of the target *Metrosideros* spp. enhanced the classification accuracy for both the XGBoost and deep learning models. However, relying on phenology leaves a narrow window for data collection and the possibility of missing trees that flower earlier, later or not at all. The approach demonstrated in this study used the strong and distinctive flowering characteristics of the species alongside ground truth data to identify sample trees in multitemporal imagery. In this way, the model was able to be trained to recognise *Metrosideros* spp. from a much broader range of phenological traits and characteristics present at different times of the year. The performance of the combined model suggested that the deep learning approach could be used to accurately classify *Metrosideros* spp. under a broad range of conditions.

The deep learning approach differed in fundamental ways to traditional remote sensing methods. By showing these models a substantial number of *Metrosideros* and other spp. the models effectively learn a function mapping the input RGB imagery to a binary outcome. While the models will certainly utilise the colour properties of the images, as demonstrated by the near-perfect classification of flowering canopies, the learning approach is also capable of capturing harder-to-quantify features. For example, the characteristic appearance of the multi-stem canopy, distinctive canopy texture and extensive budding are relatively easy for knowledgeable analysts to identify and the deep learning models can effectively learn that these or similar attributes are important and distinctive without human

input. This makes the models harder to interpret but extremely powerful for complex classification tasks.

In this study, we focused on a binary classification problem (*Metrosideros* spp. vs. all other species); however, deep learning frameworks have been widely and successfully developed to handle classification of thousands of different categories (Cui et al., 2018). It is likely that with higher-resolution imagery capturing finer detail and additional training data the demonstrated approach could be expanded to a broader range of species. For example, the omitted white pōhutukawa or other species of interest could be included with the addition of sufficient training data. The performance of this approach on other species is likely to be determined by two main factors. Firstly, how visually distinctive the species in question are and at what resolution these features can be resolved will be important. This is especially important for fine-grained classification. For example, separating rata from pōhutukawa may require imagery of a resolution that is impractical to achieve from aerial platforms. Secondly, the availability of sufficient training data to develop the models is also a crucial factor. While the number of samples (1139) in this study was very large in comparison to many other tree classification studies (Fassnacht et al., 2016; Pham et al., 2016), this number is quite low by the standards of deep learning model development. A corollary of this is that additional samples from a broader range of conditions (e.g. from natural reserves or other regions) would enhance the chances that the model will generalise well.

It is important to acknowledge the limitations of the methodology used in this study. One key methodological limitation was the need to manually delineate individual tree canopies before training and inference could be carried out. This requirement is present in many traditional remote sensing approaches to species classification. A common workflow is to use a combination of LiDAR-derived elevation or intensity values and multispectral data (especially the NIR band) to identify threshold values or object-based rulesets to separate tree canopies from other features in the imagery (M. Dalponte et al., 2013; MacFaden, O'Neil-Dunne, Royar, Lu, & Rundle, 2012). Figure 6 shows an example of a similar approach applied to a small area containing several trees from this study. While effective, this method introduces the need for costly LiDAR data and substantial input from analysts to perform the canopy extraction. More complex deep learning frameworks may offer an alternative option to overcome this requirement. Object detection and localisation architectures such as RCNN (Girshick, Donahue, Darrell, & Malik, 2014) and SSD (Liu et al., 2016) allow networks to be trained to simultaneously identify the location of features of interest such as all tree canopies within the image and then perform classification to predict the canopy species. Some early research has demonstrated success in segmenting tree canopies in urban, semi-urban and rural environments from satellite imagery (Iglovikov, Mushinskiy, & Osin, 2017). Segmentation of individual canopies in complex forests is likely to be a more challenging task and may force a trade-off between the objective e.g. minimising false negatives for a biosecurity response but accept lower segmentation accuracy. Nonetheless, deep learning-based segmentation is now the benchmark method in domains where the task is similarly challenging, such as medical image segmentation (Litjens et al., 2017; Ronneberger, Fischer, & Brox, 2015).

The high classification accuracies observed in this study must be interpreted with some caveats. Although model development followed best-practice by withholding a relatively large validation dataset (25% of available data), the models were exposed to the unique characteristics and properties of both aerial imagery datasets from which the validation set images were drawn. Deep learning approaches differ from traditional remote sensing in that they do not expect or require calibrated or corrected imagery and many successful applications use imagery from a wide variety of sources such as UAVs or mobile phones. Internally, the models transformed and scaled the input imagery to minimise differences in appearance, but this does not guarantee transferability to new, unseen aerial datasets. The level of flowering seen in the 2019 dataset varied widely and many trees showed only limited flowering. However, the imagery was also sharp and many of the other characteristic features of *Metrosideros* spp. were easily visible in the imagery (e.g. buds, canopy form and hue – Figures 2 & 3). This provided additional features for the deep learning models and likely contributed to the high accuracy above and beyond the flowering. The 2017 imagery had the same nominal resolution (10 cm) but had markedly lower quality and detail (Figures 2 & 3). The *Metrosideros* spp. all exhibited a blueish hue in this imagery and some of the textural attributes were still discernible – both of which are likely to have contributed to the performance of the deep learning model. For predictions to work in new areas, the features learned from these datasets would need to be discernible in the new imagery. A brief test conducted by reducing the resolution of some of the imagery (bilinear resampling) showed that the accuracy of the combined classifier declined rapidly as the distinctive features were lost, with simulated 15 cm imagery showing only a 70% accuracy rate.

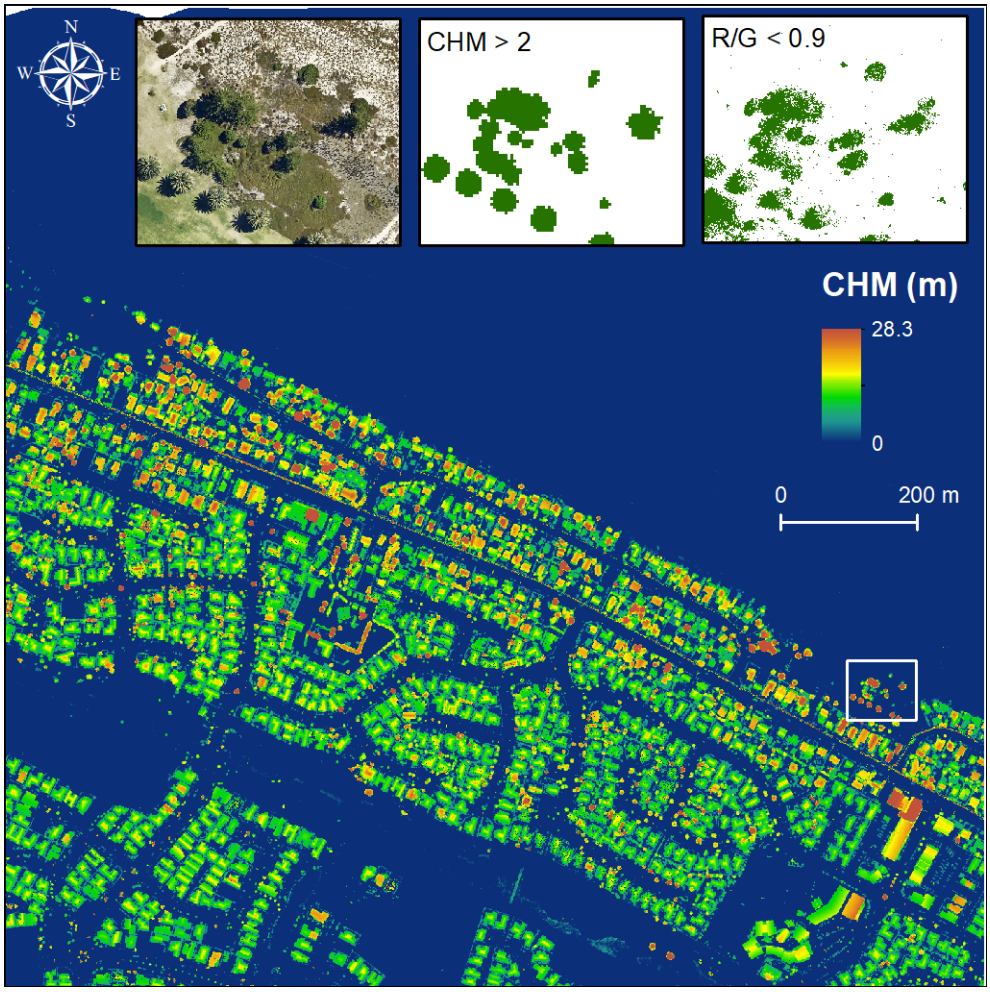

Figure 6: Example of a canopy height model (CHM) derived from aerial LiDAR in Tauranga to segregate individual tree canopies using thresholds applied to the canopy height model compared to a threshold of the red-green ratio (R/G) derived from aerial imagery.

The resolution of the imagery also placed a limit on the size of the trees that could be classified. Many canopies fell between 30-60 pixels in size. This made many of the characteristic traits difficult for a human observer to discern and the models would also have had limited information to learn on. While this was less of a problem when phenology could be utilised to enhance detection, the misclassification seemed to be biased towards smaller canopies. It is very likely that higher-resolution imagery would have improved the classification accuracy still further and may enhance the transferability of the models. This behaviour is expected as deep learning models generally perform well in situations where a trained observer could also be expected to perform well. Outside of this domain, for example, where only moderate to low resolution imagery is available, traditional multispectral or hyperspectral methods are more appropriate as they attempt to recover and utilise the spectral attributes of the canopy that can persist at coarser resolutions.

6 Recommendations and conclusions

In this study, we combined distinctive phenological traits and biosecurity surveillance records to develop a high-quality dataset to train and test novel algorithms to detect *Metrosideros* spp. from simple 3-band (RGB) aerial imagery. Both modelling approaches performed well when the dataset included distinctive phenological traits (extensive, bright red flowers). However, the deep learning algorithm was able to achieve very high accuracies even in the absence of some key traits such as the distinctive flowers. The results of this study suggest that deep learning-based approaches could be used to rapidly and accurately map certain species over large areas using only 3-band (RGB) aerial imagery. Candidate species include those where classification is achievable by an experienced analyst using the same input data. The deep learning approach did appear sensitive to image

resolution and quality and higher resolution imagery would likely expand the range of species suitable for classification using this method.

Future work should focus on enabling large-scale (regional or even national) mapping of *Metrosideros* spp. to assist stakeholders to identify and map these important species. This would enable monitoring and management strategies to understand the long-term impacts of myrtle rust as well as enabling other activities such as identifying resistant populations for breeding programmes. To achieve this, the following research questions should be addressed:

- The potential for training and applying object localisation and detection models should be explored in a range of settings (e.g. urban and dense forest) to enable canopies to be identified, segmented and classified using only RGB imagery.
- A follow-on study should look to define the limits of generalisability by testing classification using a wider variety of data sources and by expanding the training data to include the large number of *Metrosideros* spp. inspected in other areas of New Zealand.
- The demonstrated approach should be expanded to test the feasibility of performing multi-species classification possibly using higher-resolution imagery sources (<10 cm).

7 Acknowledgements

This work has been undertaken under contract MPI 18607 led by Scion, Plant and Food Research and Manaaki Whenua Landcare Research and the following collaborators: AsureQuality, Biosecurity Research Ltd., Learning for Sustainability and Te Tira Whakamātaki. The study received additional support from the Strategic Science Investment Funding provided by the Ministry of Business, Innovation and Employment to Scion Research.

We gratefully acknowledge the provision of data and research input from Deirdre Nagle and Quenten Higgan of AsureQuality Ltd. The 2017 data were sourced from Land Information New Zealand and licensed for reuse under the Creative Commons license (CC BY 4.0). Tauranga District Council and the Bay of Plenty Regional council funded and provided early access to the 2019 dataset (BOPLAS2019) also licensed for re-use under the Creative Commons license (CC BY 4.0).

8 References

Affouard, A., Goëau, H., Bonnet, P., Lombardo, J.-C., & Joly, A. (2017). Pl@ ntnet app in the era of deep learning. In *ICLR 2017-Workshop Track-5th International Conference on Learning Representations* (pp. 1–6).

Allan, H. H. (1961). Flora of New Zealand. Volume I. *Flora of New Zealand. Volume I.*, (8 x 5 1/2).

Asner, P. G., Martin, E. R., Keith, M. L., Heller, P. W., Hughes, A. M., Vaughn, R. N., … Balzotti, C. (2018). A Spectral Mapping Signature for the Rapid Ohia Death (ROD) Pathogen in Hawaiian Forests. *Remote Sensing*, *10*(3). https://doi.org/10.3390/rs10030404

Ballanti, L., Blesius, L., Hines, E., & Kruse, B. (2016). Tree Species Classification Using Hyperspectral Imagery: A Comparison of Two Classifiers. *Remote Sensing*, *8*(6). https://doi.org/10.3390/rs8060445

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, *13*(1–2), 95–120. https://doi.org/10.1080/02757259509532298

Beresford, R. M., Turner, R., Tait, A., Paul, V., Macara, G., Yu, Z. D., … Martin, R. (2018). Predicting the climatic risk of myrtle rust during its first year in New Zealand. *New Zealand Plant Protection*, *71*, 332–347. https://doi.org/10.30843/nzpp.2018.71.176

Carnegie, A. J., & Cooper, K. (2011). Emergency response to the incursion of an exotic myrtaceous rust in Australia. *Australasian Plant Pathology*, *40*(4), 346. https://doi.org/10.1007/s13313-011-0066-6

Carnegie, A. J., Kathuria, A., Pegg, G. S., Entwistle, P., Nagel, M., & Giblin, F. R. (2016). Impact of the invasive rust Puccinia psidii (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. *Biological Invasions*, *18*(1), 127–144. https://doi.org/10.1007/s10530-015-0996-y

Clark, M. L., Roberts, D. A., & Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, *96*(3), 375–398. https://doi.org/10.1016/j.rse.2005.03.009

Coutinho, T. A., Wingfield, M. J., Alfenas, A. C., & Crous, P. W. (1998). Eucalyptus Rust: A Disease with the Potential for Serious International Implications. *Plant Disease*, *82*(7), 819–825. https://doi.org/10.1094/PDIS.1998.82.7.819

Cui, Y., Song, Y., Sun, C., Howard, A., & Belongie, S. (2018). Large scale fine-grained categorization and domain-specific transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4109–4118).

Dash, J. P., Watt, M. ., Pearse, G. D., & Dungey, H. S. (2017). *UAV Based Monitoring of Physiological Stress in Trees is Affected by Image Resolution and Choice of Spectral Index*.

De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., … Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, *24*(9), 1342–1350. https://doi.org/10.1038/s41591-018-0107-6

DiTomaso, J. M., Van Steenwyk, R. A., Nowierski, R. M., Vollmer, J. L., Lane, E., Chilton, E., … Dionigi, C. P. (2017). Enhancing the effectiveness of biological control programs of invasive species through a more comprehensive pest management approach. *Pest Management Science*, *73*(1), 9–13. https://doi.org/10.1002/ps.4347

Dymond, C. C., Mladenoff, D. J., & Radeloff, V. C. (2002). Phenological differences in Tasseled Cap indices improve deciduous forest classification. *Remote Sensing of Environment*, *80*(3), 460–472. https://doi.org/10.1016/S0034-4257(01)00324-8

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

Ferreira, M. P., Wagner, F. H., Aragão, L. E. O. C., Shimabukuro, Y. E., & de Souza Filho, C. R. (2019). Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *149*, 119–131. https://doi.org/10.1016/j.isprsjprs.2019.01.019

Gamon, J. A., & Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytologist*, *143*(1), 105–117. https://doi.org/10.1046/j.1469-8137.1999.00424.x

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587).

Glen, M., Alfenas, A. C., Zauza, E. a. V., Wingfield, M. J., & Mohammed, C. (2007). Puccinia psidii: a threat to the Australian environment and economy —a review. *Australasian Plant Pathology*, *36*(1), 1–16. https://doi.org/10.1071/AP06088

Goldson, S., Bourdôt, G., Brockerhoff, E., Byrom, A., Clout, M., McGlone, M., … Templeton, M. (2015). New Zealand pest management: current and future challenges. *Journal of the Royal Society of New Zealand*, *45*(1), 31–58. https://doi.org/10.1080/03036758.2014.1000343

Hall-Beyer, M. (2017). Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *International Journal of Remote Sensing*, *38*(5), 1312–1338. https://doi.org/10.1080/01431161.2016.1278314

Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*(6), 610–621.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *ArXiv:1512.03385 [Cs]*. Retrieved from http://arxiv.org/abs/1512.03385

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity Mappings in Deep Residual Networks. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 630–645). Springer International Publishing.

He, Y., Chen, G., Potter, C., & Meentemeyer, R. K. (2019). Integrating multi-sensor remote sensing and species distribution modeling to map the spread of emerging forest disease and tree mortality. *Remote Sensing of Environment*, *231*, 111238. https://doi.org/10.1016/j.rse.2019.111238

Hesketh, M., & Sánchez-Azofeifa, G. A. (2012). The effect of seasonal spectral variation on species classification in the Panamanian tropical forest. *Remote Sensing of Environment*, *118*, 73–82. https://doi.org/10.1016/j.rse.2011.11.005

Ho, W. H., Baskarathevan, J., Griffin, R. L., Quinn, B. D., Alexander, B. J. R., Havell, D., … Pathan, A. K. (2019). First Report of Myrtle Rust Caused by Austropuccinia psidii on Metrosideros kermadecensis on Raoul Island and on M. excelsa in Kerikeri, New Zealand. *Plant Disease*, PDIS-12-18-2243-PDN. https://doi.org/10.1094/PDIS-12-18-2243-PDN

Huang, C., & Asner, P. G. (2009). Applications of Remote Sensing to Alien Invasive Plant Studies. *Sensors*, *9*(6). https://doi.org/10.3390/s90604869

Iglovikov, V., Mushinskiy, S., & Osin, V. (2017). Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *ArXiv Preprint ArXiv:1706.06169*.

J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, & J. Chanussot. (2013). Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geoscience and Remote Sensing Magazine*, *1*(2), 6–36. https://doi.org/10.1109/MGRS.2013.2244672

Kalaris, T., Fieselmann, D., Magarey, R., Colunga-Garcia, M., Roda, A., Hardie, D., … Whittle, P. (2014). The role of surveillance methods and technologies in plant biosecurity. In *The handbook of plant biosecurity* (pp. 309–337). Springer.

Kriticos, D., Phillips, C., Suckling, D., & others. (2005). Improving border biosecurity: potential economic benefits to New Zealand. *New Zealand Plant Protection*, *58*, 1.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lacerda, A. E. B. de, & Nimmo, E. R. (2010). Can we really manage tropical forests without knowing the species within? Getting back to the basics of forest management through taxonomy. *Forest Ecology and Management*, *259*(5), 995–1002. https://doi.org/10.1016/j.foreco.2009.12.005

Lange, P. J., Rolfe, J., Barkla, J., P. Courtney, S., Champion, P., Perrie, L., … Ladley, K. (2018). *Conservation status of New Zealand indigenous vascular plants, 2017*. Department of Conservation.

LINZ. (2019). LINZ Data Service - NZ Aerial Imagery. Retrieved 13 June 2019, from https://data.linz.govt.nz/set/4702-nz-aerial-imagery/

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21–37). Springer.

Loope, L. (2010). *A summary of information on the rust Puccinia psidii Winter (guava rust) with emphasis on means to prevent introduction of additional strains to Hawaii*. US Geological Survey.

M. Dalponte, H. O. Ørka, T. Gobakken, D. Gianelle, & E. Næsset. (2013). Tree Species Classification in Boreal Forests With Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(5), 2632–2645. https://doi.org/10.1109/TGRS.2012.2216272

MacFaden, S. W., O'Neil-Dunne, J. P. M., Royar, A. R., Lu, J. W. T., & Rundle, A. G. (2012). High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *Journal of Applied Remote Sensing*, *6*(1), 063567. https://doi.org/10.1117/1.JRS.6.063567

Mallat, S. (2016). Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150203.

McTaggart, A. R., Roux, J., Granados, G. M., Gafur, A., Tarrigan, M., Santhakumar, P., & Wingfield, M. J. (2016). Rust (Puccinia psidii) recorded in Indonesia poses a threat to forests and forestry in South-East Asia. *Australasian Plant Pathology*, *45*(1), 83–89. https://doi.org/10.1007/s13313-015-0386-z

Mundt, C. C. (2014). Durable resistance: A key to sustainable management of pathogens and pests. *Infection, Genetics and Evolution*, *27*, 446–455. https://doi.org/10.1016/j.meegid.2014.01.011

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., … Lerer, A. (2017). Automatic differentiation in PyTorch. In *NIPS-W*.

Pham, L. T. H., Brabyn, L., & Ashraf, S. (2016). Combining QuickBird, LiDAR, and GIS topography indices to identify a single native tree species in a complex landscape using an object-based classification approach. *International Journal of Applied Earth Observation and Geoinformation*, *50*, 187–197. https://doi.org/10.1016/j.jag.2016.03.015

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv:1505.04597 [Cs]*. Retrieved from http://arxiv.org/abs/1505.04597

Roux, J., Greyling, I., Coutinho, T. A., Verleur, M., & Wingfield, M. J. (2013). The Myrtle rust pathogen, puccinia psidii, discovered in Africa. *IMA Fungus*, *4*(1), 155–159. https://doi.org/10.5598/imafungus.2013.04.01.14

Sandhu, K. S., & Park, R. F. (2013). Genetic basis of pathogenicity in Uredo rangelii.

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., … Belongie, S. (2018). The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8769–8778).

Veit, A., Wilber, M. J., & Belongie, S. (2016). Residual networks behave like ensembles of relatively shallow networks. In *Advances in neural information processing systems* (pp. 550–558).

Wolter, P. T., Mladenoff, D. J., Host, G. E., & Crow, T. R. (1995). Improved forest classification in the Northern Lake States using multi-temporal Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, *61*(9), 1129–1144.

Zvoleff, A. (2019). *glcm: Calculate Textures from Grey-Level Co-Occurrence Matrices (GLCMs)*. Retrieved from https://CRAN.R-project.org/package=glcm