Xu et al. *New Zealand Journal of Forestry Science* (2023) 53:12
https://doi.org/10.33494/nzjfs532023x314x
E-ISSN: 1179-5395
Published on-line: 30/10/2023

RESEARCH ARTICLE **Open Access**

New Zealand Journal of Forestry Science

Mapping minor plantation species for New Zealand's small-scale forests using Sentinel-2 satellite data

Cong Xu*, Bruce Manley and Ning Ye

New Zealand School of Forestry, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

*Corresponding author: cong.xu@canterbury.ac.nz

(Received for publication 24 May 2023; accepted in revised form 2 October 2023)

Abstract

Background: Relying solely on radiata pine (*Pinus radiata* D.Don) leaves New Zealand's plantation forest industry vulnerable to fluctuations in market demand and at risk from a potentially devastating pest or disease outbreak. Therefore, the New Zealand government and forestry industry urge to diversify the forest resource and wood supply beyond the reliance on radiata pine. Unfortunately, the lack of accurate information on minor species' area, composition, and location poses challenges to forecasting potential log supply and logistics planning.

Methods: The objective of this study is to classify minor species in New Zealand using imagery and phenological features extracted from data collected by the Copernicus Sentinel-2 satellite. The study collected reference data of minor species from large-scale forest owners and applied Random Forest classification using Sentinel-2 imagery to classify nine minor species classes in the Hawke's Bay region of New Zealand.

Results: The study achieved an overall classification accuracy of 92.2% for minor species in New Zealand's Hawke's Bay region. Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) and *Eucalyptus* species had the highest accuracies, exceeding 90% for both producer's and user's accuracies. Acacia, larch, and other mixed species had lower accuracies, likely due to their lower occurrence. The most important input variable for classification was the Digital Elevation Model, indicating the significance of elevation in differentiating plantation species. The Greenness Index (GI) and Red edge bands also proved useful in the classification. The phenological measure Mean-EVI2 was found useful in classifying deciduous species such as larch and poplar.

Conclusions: To the best of our knowledge, this study is the first to map the spatial extent and distribution of minor plantation species in New Zealand at the regional level, providing promising results for potentially expanding the study to national-level species mapping.

Keywords: Minor species; small-scale forests; Sentinel-2 satellite; random forest; species classification

Introduction

Plantation forests in New Zealand cover an estimated 1.74 million hectares (MPI 2021); nearly 90% of the forests are radiata pine (*Pinus radiata* D.Don). Relying on a single species potentially leaves the forestry industry vulnerable to fluctuations in market demand and at risk from a devastating pest or disease outbreak. There is increasing interest in diversifying forest resources in New Zealand. The Specialty Wood Products Research Partnership (SWP) was established as a partnership between government and industry aiming to develop a high-value speciality wood products industry based on species other than radiata pine, such as Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), *Eucalyptus* species and cypress species (e.g. *Cupressus macrocarpa* (Hartw.) and *Cupressus lusitanica* Mill.). In order to model the potential sustainable log supply from these minor species, it is critical to understand the area and location of existing resources.

The National Exotic Forest Description (NEFD) is compiled by the Ministry for Primary Industries (MPI) to maintain an authoritative database of New Zealand's production forests. In 2021, the NEFD recorded a total of 168,000 hectares for New Zealand's minor species forest plantations, which comprises Douglas-fir, cypresses, eucalypts and other softwoods and hardwoods (MPI 2021). However, the NEFD is a non-spatial database and lacks reliability and accuracy for describing the small-scale plantation forests (less than 1000 ha), especially the forests that are under 40 ha (Manley et al. 2017; Manley et al. 2020). These limitations particularly restrict the understanding of minor species resources and complicate modelling the log supply.

Often, plantations of minor species are small and fragmented with limited accessibility, which makes it impractical to inventory the minor species through ground measurements. Therefore, a remote sensing approach which acquires information on resources without physical contact offers an alternative approach for describing forest resources. It provides opportunities to efficiently and cost-effectively identify forest species information, determine spatial distribution, and allow for frequent updates (White et al. 2016). Forest species mapping with remote sensing has been conducted worldwide, employing a range of technologies such as low- to high-resolution multispectral satellite imagery (Fassnacht et al. 2016; Grabska et al. 2020), Light Detection and Ranging data (Shi et al. 2018), unmanned aerial vehicle (UAV) images (Schiefer et al. 2020) and fusion of different sensors (Immitzer et al. 2018; Wang et al. 2018).

Although very high resolution imagery would be ideal in classifying multiple forest species, the cost and limited coverage have hindered the potential to map small-scale plantations across a large geographic area (Williams et al. 2021). Free 10-m resolution imagery obtained from the Copernicus Sentinel-2 satellite has gained popularity in forest mapping studies worldwide due to its relatively high spatial and spectral resolution. Alonso et al. (2020) classified fragmented chestnut plantations in Northwest Spain using images obtained from the Sentinel-2 satellite and achieved 81.5% accuracy. Wan et al. (2021) fused high-resolution aerial images with Sentinel-2 images to segment forest stands and classified 11 tree species on a forest farm, with classification accuracies ranging from 90% to 91.3%.

Forest species mapping can be challenging in large areas with diverse forest compositions and diverse environmental conditions. Several studies have demonstrated successful classification of forest species over larger geographic scales by incorporating temporal variations in input image data. Punalekar et al. (2021) applied the ExtraTree classifier to Sentinel-2 images collected over 4.5 years to classify national-level larch forest plantations in Wales, with all mapping accuracies above 90% when compared against an independent reference dataset. Hamrouni et al. (2021) proposed an active learning-based approach to map national-level poplar plantations in France using the Sentinel-2 time series. Schindler et al. (2021) applied a random forest classifier to map the national extent of southern beeches using a temporal stack of Sentinel-2 imagery acquired between 2016-2019 and achieved an accuracy of 87.7%. Hościło and Lewandowska (2019) used the random forest classifier to identify eight tree species in a sizable forest in southern Poland; by merging topographic data with multi-temporal Sentinel-2 data, they improved the overall classification from 75.6% to 81.7%.

Previous studies utilising Sentinel-2 imagery for tree species classification have primarily focussed on multiple species but within a small geographical extent (Grabska et al. 2019; Immitzer et al. 2016; Karasiak et al. 2017), or on classifying a single species over a large extent (Alonso et al. 2020; Punalekar et al. 2021; Schindler et al. 2021). In addition, species classification studies worldwide have predominantly concentrated on large forests, with only a limited number of studies addressing the classification of multiple species in small-scale forests dispersed across a broad geographic scale. Therefore, this study aims to explore the feasibility of classifying multiple species for small forests dispersed over a large regional scale in New Zealand, in order to understand the spatial distribution and area of these minor tree species.

Specifically, the objectives of this study are to:

(1) Classify minor tree species in New Zealand's Hawke's Bay Region using input features extracted from Sentinel-2 satellite data; and

(2) Identify the important features in classifying minor plantation species.

Methods

Study area

The study area is in the Hawke's Bay region of New Zealand, which is located on the east coast of New Zealand's North Island (39°25′S, 176°49′E). The region covers 1.42 million hectares and consists of the Wairoa District, Hastings District, Napier City and Central Hawke's Bay District (Figure 1). Forests mainly occupy the roughest terrain on the northern and eastern side of the region (Hawke's Bay Regional Council 2022). There are around 139,000 hectares of plantation forests in Hawke's Bay, owned by companies, investors, individual landowners and a small amount by the Hawke's Bay Regional Council. Nearly 20% of the forests are less than 100 hectares in size. The NEFD reported 3,190 ha of minor species in the Hawke's Bay region, comprising 445 hectares of Douglas-fir, 961 hectares of *Eucalyptus* species, 368 hectares of cypress species, 917 hectares of other softwoods and 499 ha of other hardwoods (MPI 2021). However, the spatial distribution of these minor species and a detailed species breakdown are unknown.

Reference data collection

Ground reference data plays a crucial role in the classification process. Ground reference data with known location and species are required to perform species classification. The data should be representative and cover all relevant species and all existing age classes and growing conditions so that the classification algorithm can learn what a forest species 'looks like' spectrally and texturally from remote sensing imagery and then classify pixels with similar features accordingly. Therefore, the first step of the project was to investigate the availability of ground reference data.

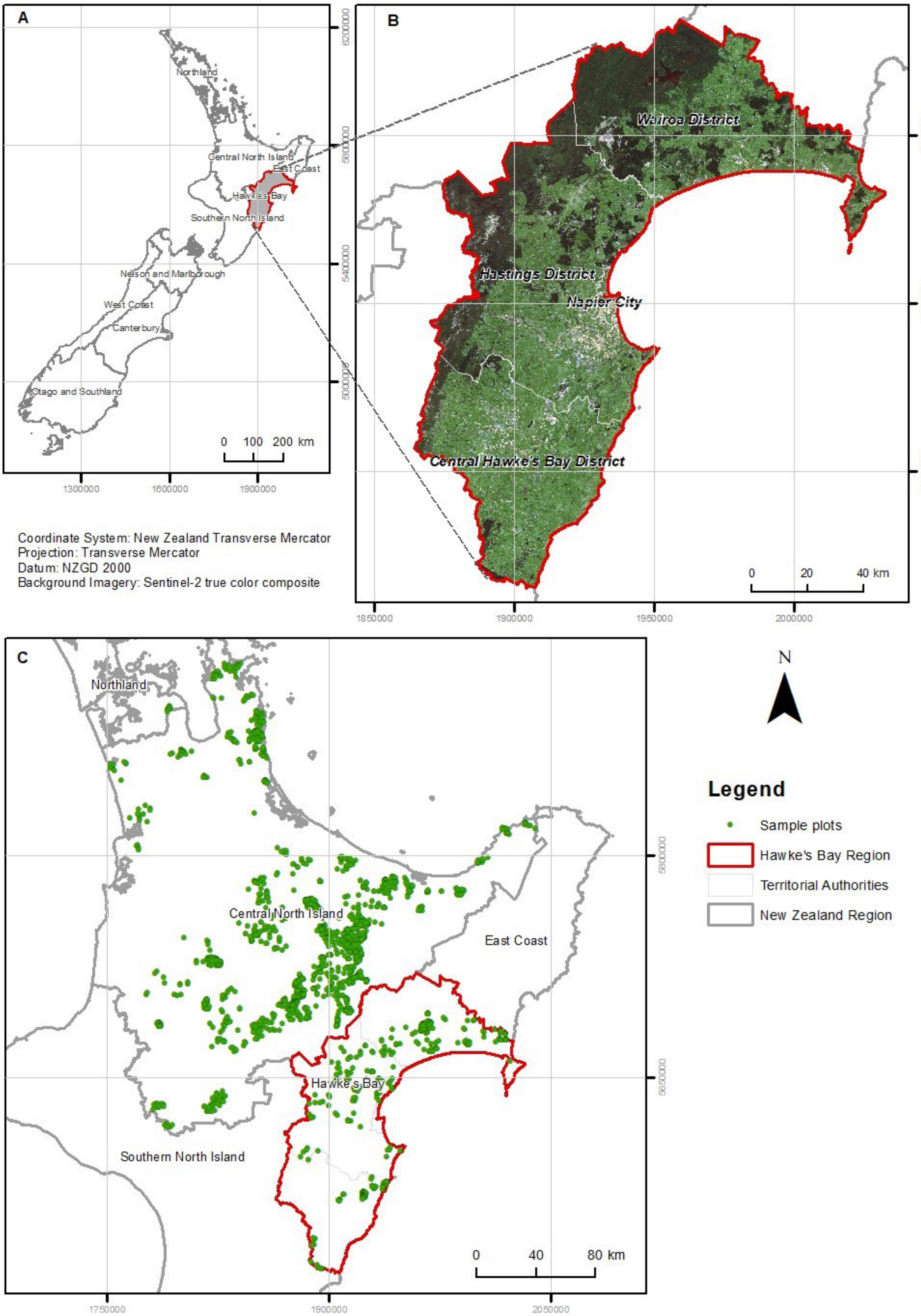

FIGURE 1: Location of the study area. Map A gives an overview of New Zealand regions. Map B shows the boundary of Hawke's Bay region and four territorial authorities in the region. Map C shows the location of the sample plots, including the reference data from the Central North Island region.

Due to lack of an official record of minor species and the impracticality of visiting each plantation in New Zealand, ground reference data were intended to be collected from owners of minor species. The spatially explicit reference data were collected from large-scale owners in geographic information system (GIS) format, whereas no data was collected from the small-scale owners due to the absence of spatial records.

An email request was sent to eleven large-corporate owners in Hawke's Bay enquiring about the spatial location of their minor species plantations in June 2021. All owners responded and provided locations of 1,130 hectares of minor species in the region, of which 53% (598 ha) were *Eucalyptus* species; Douglas-fir (*Pseudotsuga menziesii*), cypress *(Cupressus* species*)*, redwood (*Sequoia sempervirens* (D.Don) Endl.), other pine (non-*Pinus radiata*), Acacia *(Acacia* species*)*, larch (*Larix* species); while, other minor species made up the remaining 47% (Table 1). In addition, over one-third of the resource was aged five years or below. That means they are less likely to be detected from satellite imagery compared with mature forests.

We considered the reference data received from large-scale owners in Hawke's Bay, imbalanced and insufficient as we are concerned the reference data only cover small areas and may not be representative for these minor species. Therefore, we requested additional data from large-scale owners from the Central North Island (CNI) region which is located adjacently (Figure 1). The ground reference data from CNI and Hawke's Bay were combined. Within the provided GIS boundaries, 2,788 circular plots with 50 m radius were automatically and randomly generated at least 100 metres apart from each other. Occasionally, there are plots that extend beyond the stand boundary, which happens in the case of very small forest stands that are less than 50 metres in width. In such instances, these plots were adjusted to align with the forest stand boundaries to ensure that they cover only one species at a time. These plots were then used as the sample data for species classification (Map C on Figure 1). A quick visual inspection was conducted to the plots to ensure there were trees present so that the reference data was valid. The sample data was randomly split into 70% for training and 30% for validation. A summary of the ground reference data for each target classification class is described in Table 2.

Sentinel imagery

The national Sentinel-2 mosaic was processed by Manaaki Whenua - Landcare Research based on workflow developed by Shepherd et al. (2020) and distributed by the Ministry for the Environment (MfE), New Zealand. The image product is a 10 m, ten-band multispectral, cloud-minimised mosaic of multiple Sentinel-2A and -2B satellite images over New Zealand and was acquired from late 2019 to early 2022 (Table 3). The mosaic went through pan-sharpening, atmospheric and bidirectional reflectance distribution function correction, cloud clearing and a minimising process.

The national mosaic imagery was then clipped to the extent of the study area to only include Hawke's Bay and Central North Island. The latest Land Cover Database (LCDB v5.0) (Landcare Research 2021), which is a multi-temporal, thematic classification of New Zealand's land cover, was used to mask out non-plantation forest areas.

Input features

Vegetation indices (VIs), which are the spectral transformation of two or more spectral bands, are useful in detecting spectral response variations in foliage and have considerable advantages in cellular structure evaluation, stress prediction, moisture content estimate, pigment content detection, and stress estimation (Immitzer et al. 2019). In total, 33 vegetation indices, which are sensitive to vegetation properties and have been previously used in vegetation classification studies (Immitzer et al. 2019; Ye et al. 2021), were extracted from the Sentinel-2 mosaic.

Textural features are mainly related to the variability of stand density, forest type (broadleaved, coniferous), crown size, crown closure, crown form, and crown closure (Fassnacht et al. 2016). They can considerably enhance the classification accuracy when combined with spectral features (Mallinis et al. 2008). For this study, due to multiple input bands from the Sentinel-2 mosaic, a Principle Component Analysis (PCA) was

TABLE 1: Area of minor plantation species in the Hawke's Bay and Central North Island regions based on data received from large-scale forest owners.

Species Group	Scientific Name	Hawke's Bay (ha)	Central North Island (ha)
Acacia	*Acacia* species	25	65
Cypress	*Cupressus* species	74	858
Douglas-fir	*Pseudotsuga menziesii*	183	12,841
Eucalyptus	*Eucalyptus* species	598	2,303
Larch	*Larix* species	23	43
Other pine	*Pinus* species other than radiata pine	26	267
Redwood	*Sequoia sempervirens*	57	967
Other species	Other minor species	144	482
Total (ha)		**1,130**	**17,826**

TABLE 2: Description of training and validation data for each species class. Each pixel represents a 10 × 10 m grid. Radiata pine samples were manually added as placeholders in the classification. 'Other pine' are pine species other than radiata pine. Other species include other minor species that are not listed in the table.

Species Class	No. of Training pixels	No. of Validation pixels	Total No. of Truthing pixels
Acacia	1,770	758	2,528
Cypress	9,264	3,970	13,234
Douglas-fir	40,432	17,328	57,760
Eucalyptus	14,376	6,160	20,536
Larch	1,465	627	2,092
Other pine	4,813	2,062	6,875
Poplar	1,356	580	1,936
Radiata pine	27,399	11,742	39,141
Redwood	4,792	2,053	6,845
Other species	6,835	2,928	9,763
Total	**112,502**	**48,208**	**160,710**

performed. The first principle band, which holds most of the data variance, was used to extract textural features. Textural metrics including the Grey Levels Co-Occurrence Matrix (GLCM) of mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation were computed at a 3 pixel × 3 pixel window size.

Phenological features were derived from analysing the temporal variation of Enhanced Vegetation Index-2 (EVI2) using Sentinel-2 data (Level-2A product – Surface Reflectance) collected from 1 January 2019 to 31 December 2020 in Google Earth Engine (GEE). EVI2 was chosen because it is one of the most commonly used VIs for phenological studies, as reviewed by Caparros-Santiago et al. (2021). It was developed by Jiang et al. (2008) to address the saturation issue of the Normalized Difference Vegetation Index (NDVI) in areas with high biomass and to avoid using the blue band, which lacks vegetation characteristic information in the calculation (Wang et al. 2018). In addition, when comparing the differences and consistency of the remote sensing data (AVHRR, MODIS, VIIRS, SPOT-VGT, and SeaWiFS) for the characterisation of tropical forest phenology, Kim et al. (2007) discovered that EVI2 outperformed NDVI and EVI.

Harmonic (Fourier) analysis has a low sensitivity to non-systematic noise so it was used to represent the seasonal dynamic of the land surface and to extract phenological information (Derwin et al. 2020; Wu et al. 2021). By linearly mixing the sine and cosine functions, the algorithm can simulate symmetric seasonal variation (Shumway et al. 2000). The original GEE code was created by Clinton (2016) for time-series Landsat-8 data analysis, and it was modified to retrieve Sentinel-2 image collection for this study. Three seasonal metrics, amplitude (AMP), phase (PH) and mean EVI2 of the period, were extracted. Phase measures the length of the change's time window, whereas amplitude shows the size of the shift relative to a baseline.

In addition, a Digital Elevation Model (DEM) was retrieved from Land Information New Zealand (LINZ) (LINZ 2020) and was re-sampled to 10 m to be consistent with the rest of the input features. In total, 55 features were extracted using the remote-sensing software ENVI version 5.6 (ENVI 2021); Table 4. More details of the vegetation indices can be found in Appendix 1.

TABLE 3: Bands specification of Sentinel-2 mosaic.

Band	Band Name	Short Name	Wavelength (nm)
2	Blue	B	490
3	Green	G	560
4	Red	R	665
5	Red Edge 1	RE705	705
6	Red Edge 2	RE740	740
7	Red Edge 3	RE783	783
8	Near Infrared wide	NIR842	842
8A	Near Infrared narrow	NIR865	865
11	Short Wave Infrared 1	SWIR1610	1610
12	Short Wave Infrared 2	SWIR2190	2190

TABLE 4: List of 55 input features used for species classification.

Abbreviation	Name	Abbreviation	Name
Spectral bands		***Vegetation Indices***	
Blue	Blue band	LAI	Leaf Area Index
Green	Green band	MCARI_I	Modified Chlorophyll Absorption Ratio Index – Improved
Red	Red band	MNLI	Modified Non-Linear Index
RE705	Red Edge 705 nm	MNDWI	Modified Normalised Difference Water Index
RE740	Red Edge 740 nm	MSR	Modified Simple Ratio
RE783	Red Edge 783 nm	MSAVI2	Modified Soil Adjusted Vegetation Index 2
NIR842	Near Infrared 842 nm	MTVI_I	Modified Triangular Vegetation Index – Improved
NIR865	Near Infrared 865 nm	NDVI	Normalised Difference Vegetation Index
SWIR1610	Short-wave infrared 1610 nm	OSAVI	Optimized Soil Adjusted Vegetation Index
SWIR2190	Short-wave infrared 2190 nm	RENDVI	Red Edge Normalised Difference Vegetation Index
Textural		REPI	Red Edge Position Index
GLCM_Mean	Local mean of Gray-Level Co-Occurrence Matrix (GLCM)	RGRI	Red Green Ratio Index
GLCM_Variance	Local variance of GLCM	RDVI	Renormalised Difference Vegetation Index
GLCM_Homogeneity	GLCM Homogeneity	SAVI	Soil Adjusted Vegetation Index
GLCM_Contrast	GLCM Contrast	NIR_R	Simple Ratio NIR/red
GLCM_Dissimilarity	GLCM Dissimilarity	B_RE705	Simple Ratio blue/RE705
GLCM_Entropy	GLCM Entropy	B_RE740	Simple Ratio blue/RE740
GLCM_2ndMoment	GLCM 2nd Moment	B_RE783	Simple Ratio blue/RE783
GLCM_Correlation	GLCM Correlation	NIR_B	Simple Ratio NIR/blue
Phenology		NIR_G	Simple Ratio NIR/green
Mean EVI2	The average Enhanced Vegetation Index 2 (EVI2)	NIR_RE705	Simple Ratio NIR/RE705
EVI2 phase	The phase of EVI2	NIR_RE740	Simple Ratio NIR/RE740
EVI2 amplitude	The amplitude of EVI2	NIR_RE783	Simple Ratio NIR/RE783
Topography		TCARI	Transformed Chlorophyll Absorption Reflectance Index
DEM	Resampled 10 m Digital Elevation Model	TVI	Triangular Vegetation Index
Vegetation Indices		VARI	Visible Atmospherically Resistant Index
EVI2	Enhanced Vegetation Index 2	WDRVI	Wide Dynamic Range Vegetation Index
GEMI	Global Environmental Monitoring Index		
GARI	Green Atmospherically Resistant Index		
GCI	Green Chlorophyll Index		
GI	Greenness Index		
GNDVI	Green Normalised Difference Vegetation Index		

Species classification

Random forest is a machine learning algorithm applied widely in image classification because of its high prediction accuracy and the ability to handle high-dimensional data. The classifier is an ensemble of independent individual decision trees, each individual decision tree in the classifier casts a vote for the class that should be applied to the given sample, and the class that receives the most votes wins (Breiman 2001). The algorithm does not require distributional assumption and is less sensitive to the number of input variables and overfitting (Fassnacht et al. 2016). Pelletier et al. (2016)

compared classification algorithms and concluded random forest is most robust in mapping land cover over large areas by producing the highest classification accuracy with the shortest training time, as well as being less affected by parametrisation and number of training samples. Therefore, pixel-based classification with the random forest classifier was applied using the "randomForest" package (Liaw & Wiener 2002) in statistical package R (R Core Team 2013).

Due to high dimensional input features and target species classes, a feature selection process using the "VSURF" package (Genuer et al. 2015) was applied to eliminate redundant variables and reduce computation time for classification. Based on findings from Speiser et al. (2019), Variable Selection Using Random Forests (VSURF) outperformed other feature selection methods for random forest classification. After the classification, a majority filter (with 3 x 3 neighbours) was applied to the classification image to minimise the occurrence of small isolated pixels.

Accuracy assessment

The accuracy of classification was assessed using the confusion matrix (Congalton 2001), which compares the classified and reference species classes based on the validation dataset. Measures such as the overall accuracy, producer's accuracy (PA) and user's accuracy (UA) were calculated for individual classes. The overall accuracy indicates the proportion of pixels that were correctly classified out of all the reference pixels. The PA, which is related to omission error, reflects the probability of a species class being correctly classified. The UA relates to the commission error, which represents the probability that a pixel classified into a given species actually represents that species on the ground.

Application of classification

The random forest classification was applied to the forested areas in Hawke's Bay region. The spatial extent of the small-scale minor forests, which was manually delineated based on LINZ 30 cm aerial photos by the School of Forestry, University of Canterbury, New Zealand was used to clip the classification result so that the area summary of classified species only applies to minor species. Furthermore, the area summary was conducted for small-scale and large-scale plantations separately, and the area of each species class was summarised and compared with the NEFD record of minor species.

Results and Discussion

Spectral signature of tree species

Different land covers absorb, emit and reflect different wavelengths of the electromagnetic spectrum. A predictive model known as "spectral signatures" was created using multivariate statistical algorithms using reference data and multi-spectral satellite data for the same sites in order to categorise the satellite image into different types of land cover (Laborte et al. 2010).

Prior to species classification, the spectral signature of each tree species indicating how species' reflectance differs between the wavelength bands, was examined to understand the potential separability of different species (Figure 2). The spectral signature suggests that generally, all tree species reflect similarly within

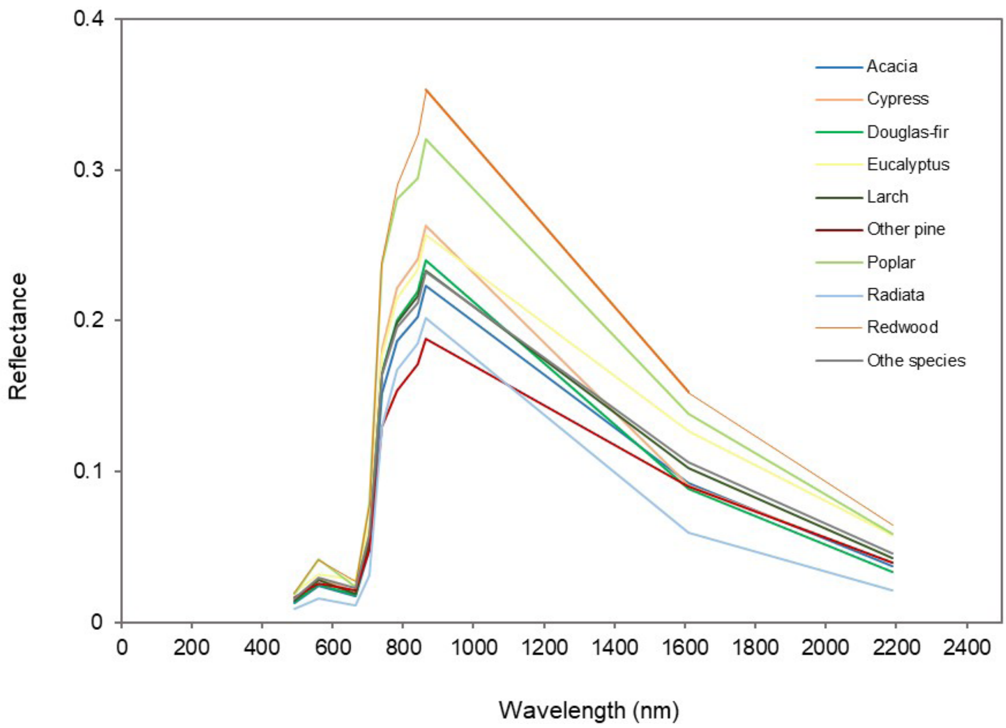

FIGURE 2: Spectral signature of different species.

the visible wavelength (400-700 nm) but illustrate the higher separation between the reflectance in the red edge and NIR spectrum (700-1300 nm). The SWIR spectrum (1300 -2500 nm) also indicated some level of separation of reflectance. The spectral signature of all species showed a preliminary possibility of separating tree species at the RE, NIR and SWIR spectrum. The reflectance characteristics of individual leaf components play the main role in how radiation interacts with vegetation. Chlorophyll, carotenoids, and anthocyanins, which are pigments found in leaves, absorb incident light to produce the majority of the visible spectrum's signal. Water absorption is the main factor in the NIR spectrum. Water has a major role in determining the reflectance in the SWIR region, although nitrogen and different types of carbon also contribute significantly to the reflectance (Asner 1998).

Classification accuracy

The species classification with all 55 input features achieved an overall accuracy of 92.2% and kappa coefficient of 89.0%. Twelve input features were selected from VSURF and the classification with 12 selected variables produced almost identical overall classification accuracy (92.3%) and kappa coefficient (89.1%) as using all input variables. The differences in the user's and producer's accuracies for individual species were also minimal (Table 5). However, classification using all 55 metrics was more time-consuming. Therefore, the classification algorithm with 12 selected variables was chosen to be applied to the whole study area due to similar accuracy and reduced computation time.

Douglas-fir and *Eucalyptus* were the two most accurately classified species (Table 5), with PA over 90%. These two classes also contained more reference data than other species classes. On the other hand, acacia was the least accurately classified class, with PA of 59.4% and UA of 88.8%. Other species with potentially more than one species were also less accurately classified (PA of 75.0% and UA of 89.1%). All species classes achieved high user's accuracies (over 85%).

There is no standard definition for the minimal classification accuracy for classifying tree species because it always relies on the study's location and goals. However, it is generally agreed, following the suggestion of Thomlinson et al. (1999), that the minimum accuracy requirements were 85% overall accuracy with no class-specific accuracy below 70% (Fassnacht et al. 2016) . The overall accuracy 92.3% well exceeded the minimum overall accuracy, and all individual species classes were above the minimum accuracy for individual class except for acacia and other exotic species. The overall classification accuracy for classifying multiple tree species was comparable with other studies using Sentinel-2 imagery, e.g. Bolyn et al. (2018) classified 10 tree species with an overall accuracy of 88.9%, Persson et al. (2018) produced an overall accuracy of 88.2% for classifying five tree species in a Swedish forest, Grabska et al. (2019) achieved up to 92.38% overall accuracy for classifying nine tree species.

Overall, this study successfully classified forest species in highly fragmented forests over a large geographic area, with a wide range of input variables including spectral bands, vegetation indices, phenological features and DEM, and produced satisfactory classification accuracies. However, it is challenging to achieve high accuracies for certain tree species (e.g. acacia and other species). Similarly, Immitzer et al. (2016) also observed lower classification accuracies for those tree species which are either uncommon in the study area or within mixed stands.

Input features

After running VSURF variable selection process, twelve out of 55 input variables were selected for species classification, each contributing differently to the species classification (Figure 3). According to the importance score of all variables, DEM was the most useful variable for classifying all minor species, suggesting that elevation plays an important role in differentiating plantation species in the study area. DEM was also found as the most important contributor to land cover and forest species classification in other studies (Ye et al. 2021; Zhang & Yang 2020). Elevation influences the distribution of minor plantation species for various reasons. Firstly, it affects climate conditions, with varying elevations corresponding to distinct temperature, solar radiation, and precipitation levels to which minor species have

TABLE 5: The producer's accuracy (PA) and user's accuracy (UA) for each species. All features mean species classification uses all 55 input features, while selected features mean classification uses VSURF (12) selected features. Detailed Confusion Matrices can be found in Appendices 2 and 3.

Feature	Acacia	Cypress	Douglas-fir	Eucalyptus	Larch	Other pine	Poplar	Redwood	Other species
All features									
PA	0.590	0.899	0.980	0.956	0.762	0.872	0.872	0.851	0.754
UA	0.892	0.940	0.924	0.918	0.931	0.936	0.939	0.910	0.894
Selected features									
PA	0.594	0.910	0.979	0.954	0.766	0.878	0.886	0.853	0.750
UA	0.888	0.936	0.921	0.906	0.909	0.941	0.935	0.905	0.891

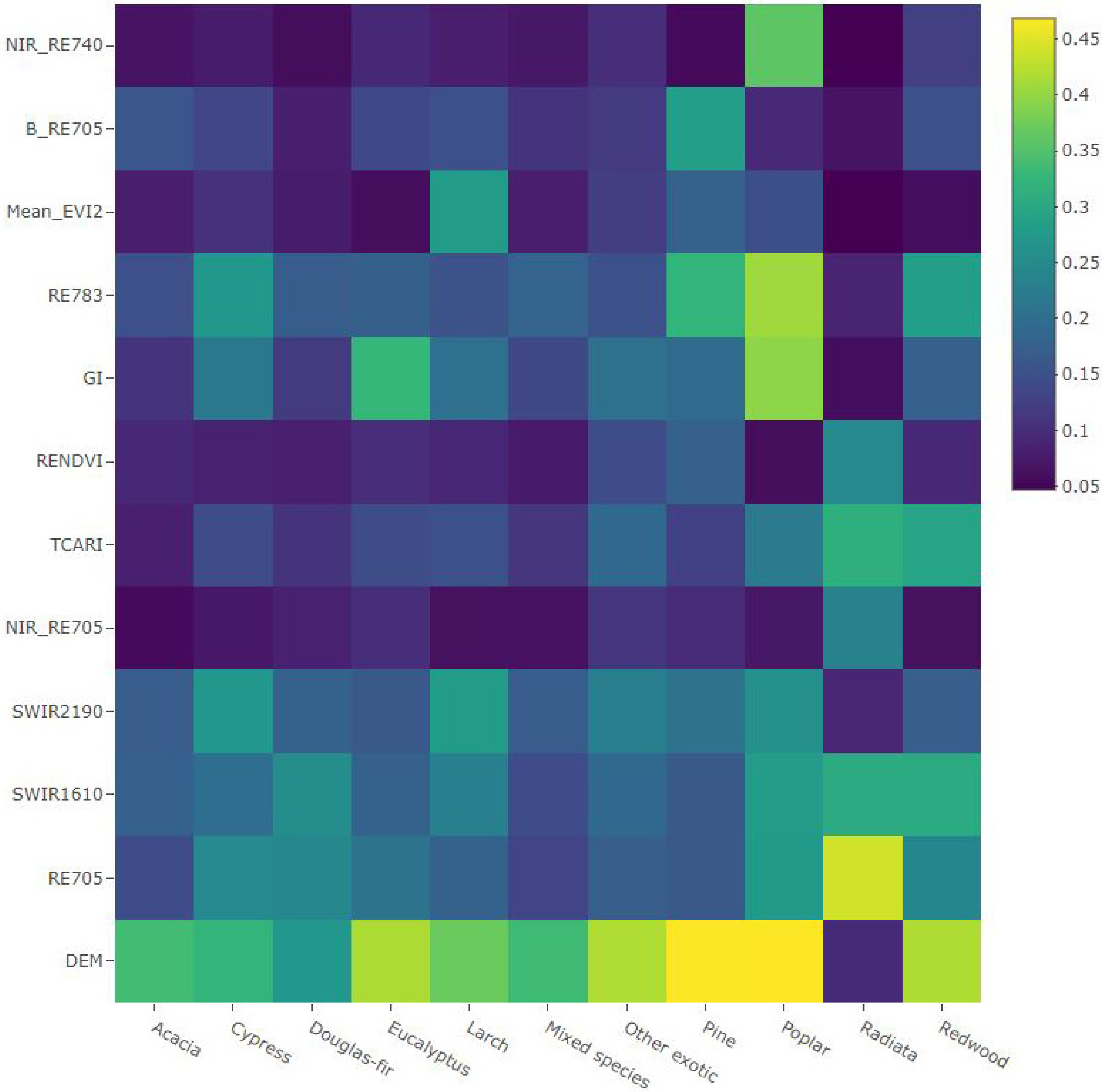

FIGURE 3: The importance score of the selected variables for each species class.

adapted (Körner 2007). Additionally, topographic features associated with elevation can impact how sunlight is distributed across slopes and aspects, further shaping trees' survival and productivity (Stage & Salas 2007).

Four original spectral bands, two RE and two SWIR bands were also identified as useful variables, which shows consistency of findings from the spectral signature of species (Figure 3). RE and SWIR bands were also identified as high-value bands for forest species mapping (Immitzer et al. 2016) and land cover classification (Schuster et al. 2012) in earlier studies. In tree canopies, the amount of radiation reflected at various wavelengths varies according to the chemical makeup of the tissue, which includes water, light-harvesting pigments, and structural carbohydrates (Asner 1998). The sharply sloping portion of the vegetation reflectance curve between 690 nm and 740 nm, which is brought on by the change from chlorophyll absorption to near-infrared leaf scattering, is known as the red edge. The amount of chlorophyll in the vegetation can be estimated by using near-infrared measurements, which have a far deeper penetration depth through the canopy than red bands (Sims & Gamon 2002). Visible and RE bands are dominated by absorption from foliar pigments, making them effective for classifying vegetation (Hennessy et al. 2020). In this study, RE bands were found particularly helpful in discriminating poplar, pine, redwood and cypress species based on the importance scores (Figure 3). SWIR bands, which are less affected by the atmosphere, and are capable of detecting water contents in soil and forest canopy, were found useful in discriminating forest types, especially evergreen coniferous forests (Murakami 2006). SWIR bands were found particularly useful in classifying redwood, radiata pine, cypress and larch in this study.

Vegetation Indices (VIs) combine the surface reflectance at two or more wavelengths to emphasise a specific characteristic of vegetation, such as photosynthetic activity and canopy structure. They enhance the sensitivity of spectral properties of vegetation while reducing spectral disturbance (Glenn et al. 2008). VIs describe the biochemical and physiological properties of vegetation that could contribute to the vegetation classification. Six out of the twelve variables were vegetation indices (GI, TCARI, B_RE705, RENDVI, NIR_RE740 and NIR_RE705). GI, which is the ratio of the green and red band, is a chlorophyll index which was found empirically related to leaf chlorophyll content (Smith et al. 1995). It is believed that measurement of chlorophyll concentration may reveal information about the physiological state of the plant, as well as nitrogen levels and hence photosynthesis (le Maire et al. 2004). In this study, GI appeared to be useful in differentiating forest species, particularly for poplar, eucalyptus and cypress. Similarly, another chlorophyll index TCARI, which describes the chlorophyll's relative abundance, was also identified as an important feature, especially for classifying redwood and radiata pine. It tends to be sensitive to detecting the reflectance of the underlying soil, especially in vegetation with a low Leaf Area Index (LAI) (Haboudane et al. 2004). RENDVI is an adjustment to the standard NDVI by employing bands along the red edge to replace the primary absorption and reflectance peaks. It takes advantage of how sensitive the vegetation's red edge is to even little variations in canopy leaf content, gap fraction, and senescence (Gitelson & Merzlyak 1994). RENDVI was found useful in discriminating radiata pine and other pine species in this study.

Three simple ratios (B_RE705, NIR_RE740 and NIR_RE705) that took the advantages of the blue, RE and NIR bands, were found more useful in classifying poplar, radiata and other pine species. B_RE705 ratio was found useful in landcover classification, particularly for low vegetations (Radoux et al. 2016). Datt (1999) found that the NIR_RE705 ratio correlated well with the chlorophyll content (up to Pearson-r = 0.83) of *Eucalyptus* species and can be used as an indicator of growth for woody vegetation. NIR_740 which was derived based on Datt (1999) was also identified as an useful input feature.

One phenology feature (Mean EVI2), which is the mean value of EVI2 over a two-year period, was also selected. Phenology is another useful property for identifying different tree species. Phenology includes obvious temporal processes such as the change of leaf colour in deciduous forests in autumn due to leaf senescence (primarily related to the faster decomposition of chlorophyll pigments), the bright green hues of young leaves and needles in the spring, as well as flowering events (Chuine & Beaubien 2001). The mean EVI2, which allows observation of the temporal change in the EVI2 over two growing seasons of plantation species, is potentially useful in discriminating evergreen and deciduous species, such as larch (Figure 3). According to Jiang et al. (2008), EVI2 is a useful VI for tracking vegetation development as it can remain sensitive to variations in thick vegetation despite not being as sensitive to background soil reflectance as NDVI (Rocha & Shaver 2009).

Application of results

The classification using a random forest classifier with 12 selected features was applied to the whole Hawke's Bay region within the extent of non-plantation forest areas and pre-defined forest boundaries. This provides a spatial representation of all minor species in the Hawke's Bay region.

In Hawke's Bay, a total of 2,151 ha of minor species were classified in the small-scale forests. The classification suggests that the most common minor species for small-scale owners are *Eucalyptus* species, with 671 ha mapped accounting for 31% of all small-scale minor species (Figure 4), followed by cypress and poplar (18% and 17% respectively). Acacia, other pine and larch are the least planted minor species (less than 20 ha) in Hawke's Bay region. The other species account for 17% of all small-scale plantations (376 ha), but the actual species distribution is unknown due to limited reference information. When summarised together with the data provided by the large-scale owners, the total distribution of all minor species in Hawke's Bay can be obtained (Figure 4).

When comparing the area with the official NEFD record, the total area of minor species is 3281 ha, which is 91 ha (3%) more than the total NEFD-reported area (Table 6). We were only able to compare three species classes as NEFD did not summarise more detailed species levels. At the species level, apart from Douglas-fir, both cypress and eucalyptus were estimated to have more area than the NEFD area. It is worth noting that NEFD lacks spatial representation of plantation forests and the area summary for Hawke's Bay region may not be accurate (Manley et al. 2020).

Limitation and opportunities

This study applied a random forest classifier to automatically classify minor species and achieved promising classification accuracy for most species. Due to limited ground reference data in the study area of Hawke's Bay, data from a neighbouring region, Central North Island, were also utilised to augment the reference dataset. The classification results revealed that a larger reference dataset corresponded to higher accuracy in species classification. Consequently, the high classification accuracies observed in Hawke's Bay may be largely attributed to the abundance of reference data from Central North Island. In addition, it is important to acknowledge a limitation with the assumption that the accuracy and representativeness of the GIS data provided by the large-scale owners apply to the small-scale plantations.

This study classified nine minor species classes, but in reality, understanding the distribution within each species class is also critical. For example, there are over 700 *Eucalyptus* species, and many have been introduced to New Zealand. Each species may have different

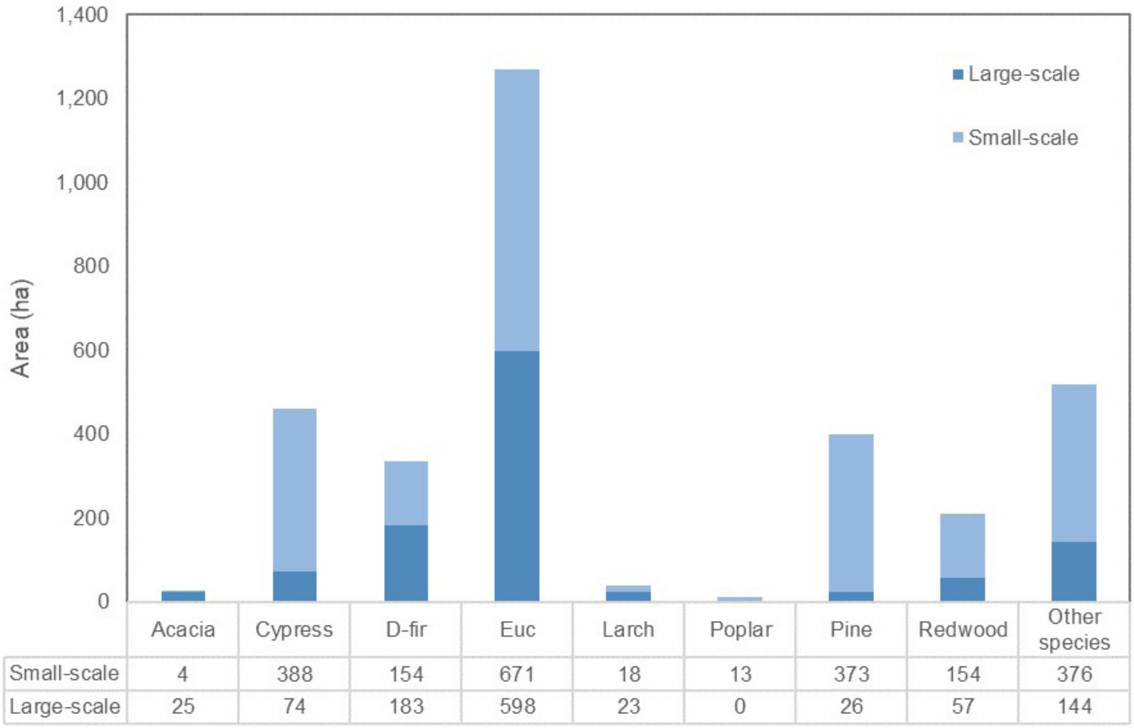

FIGURE 4: Area summary of large-scale and small-scale minor species plantation. Large-scale species was provided by owners while small-scale species was estimated from classification.

growth and product characteristics; hence identifying individual species becomes critical. In order to classify individual species, sub-metre resolution imagery (e.g. UAV multispectral imaging) and more detailed reference data will be required. This may require field verification of species by either visiting the forests with a GPS or doing drone surveys of representative minor species plantations.

In this study, classifying species within the same genus is not possible given the limited training data and 10 m resolution imagery. However, with more reference data and higher resolution imagery, it might be possible to classify individual species within the same genus (e.g. *Eucalyptus* species), given that each species 'looks' differently from higher resolution imagery. Being able to differentiate eucalypts at a species level would improve the usefulness of undertaking a national inventory.

TABLE 6: The total area of minor species in Hawke's Bay compared with NEFD 2019 area (MPI 2020). The areas estimated in this study include both large-scale and small-scale. Other species are aggregated due to different species class definition in NEFD.

Species	**Estimated in this study (ha)**	**NEFD Area (ha)**
Douglas-fir	337	445
Cypress	462	368
Eucalypt	1,269	961
Other	1,213	1,416
Total	**3,281**	**3,190**

Conclusions

This study provides proof of the concept of using remote sensing to classify minor species of the small-scale plantation at a regional level and achieved high classification accuracies for most species. The two minor species that were most correctly identified were Douglas-fir and *Eucalyptus* species, with over 90% of both producer's and user's accuracies. It was found that the classification accuracy of using random forest classifier highly depends on the availability of reference data. In total, 2151 ha of small-scale minor species were classified for Hawke's Bay, and a majority of them are eucalyptus, cypress and poplar. The Digital Elevation Model was the most significant input variable chosen for the classification, indicating that elevation is a key factor in separating plantation species. Greenness Index (GI) and red edge bands are also shown to be helpful in the classification. Mean-EVI2, a phenological feature, was only discovered to be helpful in classifying deciduous species like larch and poplar.

This is one of the very few studies that classified multiple forest species in highly fragmented forests over a large geographic extent. The spatial distribution of minor plantation species in New Zealand was, to the best of our knowledge, mapped for the first time at the regional level, and the results are encouraging enough to continue with species mapping at the national level. To further enhance the accuracy of minor species mapping, acquiring additional reliable reference data and employing higher resolution imagery could potentially improve the identification of more detailed species.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CX: Conceptualisation, reference data collection, remote sensing data processing and analysis, investigation, writing - original draft. BM: Conceptualisation, reference data collection, project management, funding acquisition, writing- review & editing. NY: Phenology data processing and analysis, writing - review & editing.

Acknowledgements

This project was supported by the Specialty Wood Products Partnership (SWP) New Zealand. We are grateful for forest managers who provided species truthing data for this project. Thanks to the New Zealand School of Forestry of University of Canterbury students and Bridget De Vries for manually mapping the small-scale plantation boundaries in the region. We also thank François Bissey, from Digital Services of University of Canterbury, for providing the Research Compute Cluster service to support the image classification in this study, and for the anonymous referees and journal editors involved in reviewing this paper.

Additional Files

Appendix 1: Vegetation indices extracted from Sentinel-2 imagery.

Appendix 2: Confusion matrix of classification with all input features. It was produced based on 30% validation dataset. PA stands for producer's accuracy and UA stands for user's accuracy. Overall accuracy is 0.922 and kappa coefficient is 0.890.

Appendix 3: Confusion matrix of classification with 12 selected variables. It was produced based on 30% validation dataset. PA stands for producer's accuracy and UA stands for user's accuracy. Overall accuracy is 0.923 and kappa coefficient is 0.891.

References

Alonso, L., Picos, J., Bastos, G., & Armesto, J. (2020). Detection of very small tree plantations and tree-level characterization using open-access remote-sensing databases. *Remote Sensing, 12*(14): 2276. https://doi.org/10.3390/rs12142276

Asner, G.P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment, 64*(3), 234-253. https://doi.org/10.1016/S0034-4257(98)00014-5

Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., & Bonnet, S. (2018). Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. *Biotechnology, Agronomy and Society and Environment, 22*(3), 172-187. https://doi.org/10.25518/1780-4507.16524

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Caparros-Santiago, J.A., Rodriguez-Galiano, V., & Dash, J. (2021). Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing, 171*, 330-347. https://doi.org/10.1016/j.isprsjprs.2020.11.019

Chuine, I., & Beaubien, E.G. (2001). Phenology is a major determinant of tree species range. *Ecology Letters, 4*(5), 500-510. https://doi.org/10.1046/j.1461-0248.2001.00261.x

Clinton, N. (2016). *Time Series Analysis in Earth Engine*. Retrieved from https://docs.google.com/presentation/d/1J1rUtf-bkfaJwYJY-tU17kzKI4U8FnF7Q2_VWqWdaak/edit#slide=id.g4954714e1_18

Congalton, R.G. (2001). Accuracy assessment and validation of remotely sensed and other spatial information. *International Journal of Wildland Fire, 10*(4), 321-328. https://doi.org/10.1071/WF01031

Datt, B. (1999). Visible/near infrared reflectance and chlorophyll content in *Eucalyptus* leaves. *International Journal of Remote Sensing, 20*(14), 2741-2759. https://doi.org/10.1080/014311699211778

Derwin, J.M., Thomas, V.A., Wynne, R.H., Coulston, J.W., Liknes, G.C., Bender, S., Blinn, C.E., Brooks, E.B., Ruefenacht, B., Benton, R., Finco, M.V., & Megown, K. (2020). Estimating tree canopy cover using harmonic regression coefficients derived from multitemporal Landsat data. *International Journal of Applied Earth Observation and Geoinformation, 86:* 101985. https://doi.org/10.1016/j.jag.2019.101985

ENVI. (2021). *Exelis Visual Information Solutions.* Boulder, CO, USA.

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

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2015). VSURF: An R package for variable selection using Random Forests. *The R Journal 7/2*.

Gitelson, A., & Merzlyak, M.N. (1994). Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology, 143*(3), 286-292. https://doi.org/10.1016/S0176-1617(11)81633-0

Glenn, E.P., Huete, A.R., Nagler, P.L., & Nelson, S.G. (2008). Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors, 8*(4), 2136-2160. https://doi.org/10.3390/s8042136

Grabska, E., Frantz, D., & Ostapowicz, K. (2020). Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. *Remote Sensing of Environment, 251: 112103*. https://doi.org/10.1016/j.rse.2020.112103

Grabska, E., Hostert, P., Pflugmacher, D., & Ostapowicz, K. (2019). Forest stand species mapping using the Sentinel-2 time series. *Remote Sensing, 11*(10): 1197. https://doi.org/10.3390/rs11101197

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., & Strachan, I.B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment, 90*(3), 337-352. https://doi.org/10.1016/j.rse.2003.12.013

Hamrouni, Y., Paillassa, E., Chéret, V., Monteil, C., & Sheeren, D. (2021). From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing, 171*, 76-100. https://doi.org/10.1016/j.isprsjprs.2020.10.018

Hawke's Bay Regional Council (2022). *About our region. Hawke's Bay stats and facts on our climate, people, environment and economy.* Napier, New Zealand: Hawke's Bay Regional Council. Retrieved from https://www.hbrc.govt.nz/hawkes-bay/about-our-region/

Hennessy, A., Clarke, K., & Lewis, M. (2020). Hyperspectral classification of plants: A review of waveband selection generalisability. *Remote Sensing, 12*(1): 113. https://doi.org/10.3390/rs12010113

Hościło, A., & Lewandowska, A. (2019). Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data. *Remote Sensing, 11*(8): 929. https://doi.org/10.3390/rs11080929

Immitzer, M., Böck, S., Einzmann, K., Vuolo, F., Pinnel, N., Wallner, A., & Atzberger, C. (2018). Fractional cover mapping of spruce and pine at 1ha resolution combining very high and medium spatial resolution satellite imagery. *Remote Sensing of Environment, 204*, 690-703. https://doi.org/10.1016/j.rse.2017.09.031

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., & Atzberger, C. (2019). Optimal input features for tree species classification in Central Europe based on multi-temporal Sentinel-2 data. *Remote Sensing, 11*(22): 2599. https://doi.org/10.3390/rs11222599

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sensing, 8*(3): 166. https://doi.org/10.3390/rs8030166

Jiang, Z., Huete, A.R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment, 112*(10), 3833-3845. https://doi.org/10.1016/j.rse.2008.06.006

Karasiak, N., Sheeren, D., Fauvel, M., Willm, J., Dejoux, J., & Monteil, C. (2017). Mapping tree species of forests in southwest France using Sentinel-2 image time series. Paper presented at the *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, 27-29 June 2017. https://doi.org/10.1109/Multi-Temp.2017.8035215

Kim, Y., Huete, A.R., Jiang, Z., & Miura, T. (2007). Multisensor reflectance and vegetation index comparisons of Amazon tropical forest phenology with hyperspectral Hyperion data. Paper presented at *Remote Sensing and Modeling of Ecosystems for Sustainability IV*. 28–29 August 2007, San Diego, California, USA https://doi.org/10.1117/12.734974

Körner, C. (2007). The use of 'altitude' in ecological research. *Trends in Ecology & Evolution, 22*(11), 569-574. https://doi.org/10.1016/j.tree.2007.09.006

Laborte, A.G., Maunahan, A.A., & Hijmans, R.J. (2010). Spectral signature generalization and expansion can improve the accuracy of satellite image classification. *Plos One, 5*(5): e10516. https://doi.org/10.1371/journal.pone.0010516

Land Information New Zealand (LINZ). (2020). *NZ 8m Digital Elevation Model (2012)*. Retrieved from https://data.linz.govt.nz/layer/51768-nz-8m-digital-elevation-model-2012/

Landcare Research. (2021). *LCDB v5.0—Land Cover Database version 5.0*. Retrieved from https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/

le Maire, G., François, C., & Dufrêne, E. (2004). Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment, 89*(1), 1-28. https://doi.org/10.1016/j.rse.2003.09.004

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18-22.

Mallinis, G., Koutsias, N., Tsakiri-Strati, M., & Karteris, M. (2008). Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. *ISPRS Journal of Photogrammetry and Remote Sensing, 63*(2), 237-250. https://doi.org/10.1016/j.isprsjprs.2007.08.007

Manley, B., Morgenroth, J., Visser, R., Final year BForSc students of 2015, & Final year BForSc students of 2016. (2017). Quantifying the small-scale owners' estate in Canterbury, Otago and Southland. *New Zealand Journal of Forestry, 62*(3), 24-32.

Manley, B., Morgenroth, J., & Xu, C. (2020). Quantifying the area of the small-scale owners' forest estate in the East Coast, Hawke's Bay and Southern North Island. *New Zealand Journal of Forestry, 64*(4), 18-26.

MPI. (2020). *2020 National Exotic Forest Description*. Wellington, New Zealand: Ministry for Primary Industries (MPI). Retrieved from https://www.canopy.govt.nz/assets/content-blocks/downloads/43540-NEFD-2020-12-18-14-10.pdf

MPI. (2021). *2021 National Exotic Forest Description.* Wellington, New Zealand: Ministry for Primary Industries (MPI). Retrieved from https://www.canopy.govt.nz/assets/content-blocks/downloads/2021-NEFD-Report.pdf

Murakami, T. (2006). How is short-wave Infrared (SWIR) useful to discrimination and classification of forest types in warm temperate region? *Journal of Forest Planning, 12*(2), 81-85. https://doi.org/10.20659/jfp.12.2_81

Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment, 187*, 156-168. https://doi.org/10.1016/j.rse.2016.10.010

Persson, M., Lindberg, E., & Reese, H. (2018). Tree species classification with multi-temporal Sentinel-2 data. *Remote Sensing, 10*(11): 1794. https://doi.org/10.3390/rs10111794

Punalekar, S.M., Planque, C., Lucas, R.M., Evans, D., Correia, V., Owers, C.J., Poslajko, P., Bunting, P., & Chognard, S. (2021). National scale mapping of larch plantations for Wales using the Sentinel-2 data archive. *Forest Ecology and Management, 501:* 119679. https://doi.org/10.1016/j.foreco.2021.119679

R Core Team. (2013). *R: A language and environment for statistical computing*.

Radoux, J., Chomé, G., Jacques, D.C., Waldner, F., Bellemans, N., Matton, N., Lamarche, C., D'Andrimont, R., & Defourny, P. (2016). Sentinel-2's potential for sub-pixel landscape feature detection. *Remote Sensing, 8*(6): 488. https://doi.org/10.3390/rs8060488

Rocha, A.V., & Shaver, G.R. (2009). Advantages of a two band EVI calculated from solar and photosynthetically active radiation fluxes. *Agricultural and Forest Meteorology, 149*(9), 1560-1563. https://doi.org/10.1016/j.agrformet.2009.03.016

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., & Schmidtlein, S. (2020). Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 170*, 205-215. https://doi.org/10.1016/j.isprsjprs.2020.10.015

Schindler, J., Dymond, J.R., Wiser, S.K., & Shepherd, J.D. (2021). Method for national mapping spatial extent of southern beech forest using temporal spectral signatures. *International Journal of Applied Earth Observation and Geoinformation, 102:* 102408. https://doi.org/10.1016/j.jag.2021.102408

Schuster, C., Förster, M., & Kleinschmit, B. (2012). Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing, 33*(17), 5583-5599. https://doi.org/10.1080/01431161.2012.666812

Shepherd, J.D., Schindler, J., & Dymond, J.R. (2020). Automated mosaicking of Sentinel-2 satellite imagery. *Remote Sensing, 12*(22): 3680. https://doi.org/10.3390/rs12223680

Shi, Y., Wang, T., Skidmore, A.K., & Heurich, M. (2018). Important LiDAR metrics for discriminating forest tree species in Central Europe. *ISPRS Journal of Photogrammetry and Remote Sensing, 137*, 163-174. https://doi.org/10.1016/j.isprsjprs.2018.02.002

Shumway, R.H., Stoffer, D.S., & Stoffer, D.S. (2000). *Time Series Analysis and its Applications* (Vol. 3): Springer. https://doi.org/10.1007/978-1-4757-3261-0

Sims, D.A., & Gamon, J.A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment, 81*(2), 337-354. https://doi.org/10.1016/S0034-4257(02)00010-X

Smith, R.C.G., Adams, J., Stephens, D.J., & Hick, P.T. (1995). Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. *Australian Journal of Agricultural Research, 46*(1), 113-125. https://doi.org/10.1071/AR9950113

Speiser, J.L., Miller, M.E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications, 134*, 93-101. https://doi.org/10.1016/j.eswa.2019.05.028

Stage, A.R., & Salas, C. (2007). Interactions of elevation, aspect, and slope in models of forest species composition and productivity. *Forest Science, 53*(4), 486-492.

Thomlinson, J.R., Bolstad, P.V., & Cohen, W.B. (1999). Coordinating methodologies for scaling landcover classifications from site-specific to global: Steps toward validating global map products. *Remote Sensing of Environment, 70*(1), 16-28. https://doi.org/10.1016/S0034-4257(99)00055-3

Wan, H., Tang, Y., Jing, L., Li, H., Qiu, F., & Wu, W. (2021). Tree species classification of forest stands using multisource remote sensing data. *Remote Sensing, 13*(1), 1-24. https://doi.org/10.3390/rs13010144

Wang, B., Jia, K., Liang, S., Xie, X., Wei, X., Zhao, X., Yao, Y., & Zhang, X. (2018). Assessment of Sentinel-2 MSI spectral band reflectances for estimating fractional vegetation cover. *Remote Sensing, 10*(12): 1927. https://doi.org/10.3390/rs10121927

Wang, D., Wan, B., Qiu, P., Su, Y., Guo, Q., Wang, R., Sun, F., & Wu, X. (2018). Evaluating the performance of Sentinel-2, Landsat 8 and Pléiades-1 in mapping mangrove extent and species. *Remote Sensing, 10*(9): 1468. https://doi.org/10.3390/rs10091468

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., & Tompalski, P. (2016). Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing, 42*(5), 619-641. https://doi.org/10.1080/07038992.2016.1207484

Williams, P.T., Wynne, R.H., Thomas, V.A., & DeFries, R. (2021). Mapping smallholder forest plantations in Andhra Pradesh, India using multitemporal harmonized landsat sentinel-2 S10 data. *Land Degradation and Development, 32*(15), 4212-4226. https://doi.org/10.1002/ldr.4027

Wu, N., Shi, R., Zhuo, W., Zhang, C., Zhou, B., Xia, Z., Tao, Z., Gao, W., & Tian, B. (2021). A Classification of Tidal Flat Wetland Vegetation Combining Phenological Features with Google Earth Engine. *Remote Sensing, 13*(3): 443. https://doi.org/10.3390/rs13030443

Ye, N., Morgenroth, J., Xu, C., & Chen, N. (2021). Indigenous forest classification in New Zealand – A comparison of classifiers and sensors. *International Journal of Applied Earth Observation and Geoinformation, 102:* 102395. https://doi.org/10.1016/j.jag.2021.102395

Zhang, F., & Yang, X. (2020). Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment, 251:* 112105. https://doi.org/10.1016/j.rse.2020.112105

New Zealand Journal of Forestry Science (2023) 53:12 https://doi.org/10.33494/nzjfs532023x314x

New Zealand Journal of Forestry Science

Supplemental Data

Mapping minor plantation species for New Zealand's small-scale forests using Sentinel-2 satellite data

Cong Xu*, Bruce Manley and Ning Ye

Appendix 1: Vegetation indices extracted from Sentinel-2 imagery

Abbreviation	Name	Formula	Reference
EVI2	Enhanced Vegetation Index 2	$2.5 * \frac{NIR - Red}{NIR + 2.4 \times Red + 1}$	Jiang et al. (2008)
GEMI	Global Environmental Monitoring Index	$\eta - 0.25\eta^2 - \frac{Red - 0.125}{1 - Red}$ $\eta = \frac{2(NIR1^2 - Red^2) + 1.5NIR1 + 0.5Red}{NIR1 + Red + 0.5}$	Pinty and Verstraete (1992)
GARI	Green Atmospherically Resistant Index	$\frac{NIR - (Green - 1.7(Blue - Red)}{NIR + (Green - 1.7(Blue - Red)}$	Gitelson et al. (1996)
GCI	Green Chlorophyll Index	$\left(\frac{NIR}{Green}\right) - 1$	Gitelson et al. (2003)
GI	Greenness Index	$\frac{Green}{Red}$	Le Maire et al. (2004)
GNDVI	Green Normalised Difference Vegetation Index	$\frac{NIR - Green}{NIR + Green}$	Gitelson and Merzlyak (1998)
LAI	Leaf Area Index	$3.618 * EVI - 0.118$	Boegh et al. (2002)

Abbreviation	Name	Formula	Reference
MCARI_I	Modified Chlorophyll Absorption Ratio Index – Improved	$\frac{1.5[2.5(RE783 - Red) - 1.3(RE783 - Green)]}{\sqrt{(2 * RE783 + 1)^2 - (6 * RE783 - 5 * \sqrt{\mathrm{Red}})} - 0.5}$	Haboudane et al. (2004)
MNLI	Modified Non-Linear Index	$\frac{(NIR^2 - Red) * (1 + L)}{NIR^2 + Red + L}$	Yang et al. (2008)
MNDWI	Modified Normalised Difference Water Index	$\frac{Green - RE865}{Green + RE865}$	Gitelson et al. (1996)
MSR	Modified Simple Ratio	$\frac{\left(\frac{NIR}{Red}\right) - 1}{(\sqrt{\frac{NIR}{Red}}) + 1}$	Chen (1996)
MSAVI2	Modified Soil Adjusted Vegetation Index 2	$\frac{2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8(\mathrm{NIR} - \mathrm{red})}}{2}$	Qi et al. (1994)
MTVI_I	Modified Triangular Vegetation Index – Improved	$\frac{1.5[1.2(RE783 - Green) - 2.5 * (Red - Green)]}{\sqrt{(2 * RE783 + 1)^2 - (6 * RE783 - 5 * \sqrt{\mathrm{Red}})} - 0.5}$	Haboudane et al. (2004)
NDVI	Normalised Difference Vegetation Index	$\frac{NIR - Red}{NIR + Red}$	Tucker (1979)
OSAVI	Optimized Soil Adjusted Vegetation Index	$\frac{(NIR - Red)}{(NIR + Red + 0.16)}$	Rondeaux et al. (1996)
RENDVI	Red-Edge Normalised Difference Vegetation Index	$\frac{RE740 - RE705}{RE740 + RE705}$	Gitelson and Merzlyak (1994)
REPI	Red Edge Position Index	Maximum derivation between RE740 and RE705	Curran et al. (1990)
RGRI	Red Green Ratio Index	$\frac{\sum_{i=600}^{699} Ri}{\sum_{j=500}^{599} Rj}$	Gamon and Surfus (1999)
RDVI	Renormalized Difference Vegetation Index	$\frac{(NIR - Red)}{\sqrt{(NIR + Red)}}$	Roujean and Breon (1995)
SAVI	Soil Adjusted Vegetation Index	$\frac{NIR - Red}{NIR + Red + 0.5} * 1.5$	Huete (1988)
NIR_R	Simple Ratio NIR/red	$\frac{NIR}{Red}$	Birth and McVey (1968)

Abbreviation	Name	Formula	Reference
B_RE705	Simple Ratio blue/RE705	$\frac{Blue}{RE705}$	Le Maire et al. (2004)
B_RE740	Simple Ratio blue/RE740	$\frac{Blue}{RE740}$	Lichtenthaler et al. (1996)
B_RE783	Simple Ratio blue/RE783	$\frac{Blue}{RE783}$	Radoux et al. (2016)
NIR_B	Simple Ratio NIR/blue	$\frac{NIR}{Blue}$	Blackburn (1998)
NIR_G	Simple Ratio NIR/green	$\frac{NIR}{Green}$	Le Maire et al. (2004)
NIR_RE705	Simple Ratio NIR/RE705	$\frac{NIR}{RE705}$	Datt (1999)
NIR_RE740	Simple Ratio NIR/RE740	$\frac{NIR}{RE740}$	Radoux et al. (2016)
NIR_RE783	Simple Ratio NIR/RE783	$\frac{NIR}{RE783}$	Radoux et al. (2016)
TCARI	Transformed Chlorophyll Absorption Reflectance Index	$3[(RE705-Red)-0.2(RE705-Green)\left(\frac{RE705}{Red}\right)]$	Haboudane et al. (2004)
TVI	Triangular Vegetation Index	$\frac{120(RE740-Green)-200(Red-Green)}{2}$	(Broge & Leblanc 2001)
VARI	Visible Atmospherically Resistant Index	$\frac{Green-Red}{Green+Red-Blue}$	(Gitelson et al. 2002)
WDRVI	Wide Dynamic Range Vegetation Index	$\frac{(0.2*NIR-Red)}{0.2*NIR+Red)}$	(Gitelson 2004)

Appendix 2: Confusion matrix of classification with all input features. It was produced based on 30% validation dataset. PA stands for producer's accuracy and UA stands for user's accuracy. Overall accuracy is 0.922 and kappa coefficient is 0.890.

	Reference									
Prediction	**Acacia**	**Cypress**	**Douglas-fir**	**Eucalyptus**	**Larch**	**Other pine**	**Poplar**	**Redwood**	**Other species**	**Total**	**UA**
Acacia	438	18	11	7	1	1	0	1	14	491	0.892
Cypress	32	3517	32	14	34	17	1	51	45	3743	0.940
Douglas-fir	155	215	16892	142	69	160	29	161	455	18278	0.924
Eucalyptus	95	56	110	5762	16	63	18	54	104	6278	0.918
Larch	0	14	8	2	474	2	2	1	6	509	0.931
Other pine	8	14	22	32	2	1786	1	11	33	1909	0.936
Poplar	0	1	1	4	4	0	506	2	21	539	0.939
Redwood	4	37	57	37	4	1	7	1719	22	1888	0.910
Other species	10	39	101	29	18	19	16	21	2144	2397	0.894
Total	742	3911	17234	6029	622	2049	580	2021	2844	36032	
PA	0.590	0.899	0.980	0.956	0.762	0.872	0.872	0.851	0.754		**0.922 (0.890)**

Appendix 3: Confusion matrix of classification with 12 selected variables. It was produced based on 30% validation dataset. PA stands for producer's accuracy and UA stands for user's accuracy. Overall accuracy is 0.923 and kappa coefficient is 0.891.

	Reference										
Prediction	**Acacia**	**Cypress**	**Douglas-fir**	**Eucalyptus**	**Larch**	**Other pine**	**Poplar**	**Redwood**	**Other species**	**Total**	**UA**
Acacia	443	16	14	13	2	5	0	0	5	499	0.888
Cypress	22	3559	39	29	36	11	0	38	47	3803	0.936
Douglas-fir	176	165	16842	140	54	147	26	160	464	18277	0.921
Eucalyptus	81	71	113	5754	25	63	15	64	110	6354	0.906
Larch	0	17	10	0	478	7	2	3	8	526	0.909
Other pine	10	7	19	31	3	1798	0	11	31	1911	0.941
Poplar	0	1	1	2	2	0	514	3	27	550	0.935
Redwood	2	42	59	37	3	2	9	1718	20	1898	0.905
Other species	12	35	101	27	21	14	14	17	2136	2397	0.891
Total	746	3913	17198	6033	624	2047	580	2014	2848	36003	
PA	0.594	0.910	0.979	0.954	0.766	0.878	0.886	0.853	0.750		**0.923 (0.891)**

References

Birth, G.S., & McVey, G.R. (1968). Measuring the color of growing turf with a reflectance spectrophotometer 1. *Agronomy Journal, 60*(6), 640-643. https://doi.org/10.2134/agronj1968.00021962006000060016x

Blackburn, G.A. (1998). Quantifying chlorophylls and caroteniods at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sensing of Environment, 66*(3), 273-285. https://doi.org/10.1016/S0034-4257(98)00059-5

Boegh, E., Soegaard, H., Broge, N., Hasager, C.B., Jensen, N.O., Schelde, K., & Thomsen, A. (2002). Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sensing of Environment, 81*(2), 179-193. https://doi.org/10.1016/S0034-4257(01)00342-X

Broge, N.H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment, 76*(2), 156-172. https://doi.org/10.1016/S0034-4257(00)00197-8

Chen, J.M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing, 22*(3), 229-242. https://doi.org/10.1080/07038992.1996.10855178

Curran, P.J., Dungan, J.L., & Gholz, H.L. (1990). Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiology, 7*(1-2-3-4), 33-48. https://doi.org/10.1093/treephys/7.1-2-3-4.33

New Zealand Journal of Forestry Science (2023) 53:12 https://doi.org/10.33494/nzjfs532023x314x

Datt, B. (1999). A new reflectance index for remote sensing of chlorophyll content in higher plants: Tests using eucalyptus leaves. *Journal of Plant Physiology, 154*(1), 30-36. https://doi.org/10.1016/S0176-1617(99)80314-9

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

Gitelson, A., & Merzlyak, M.N. (1994). Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology, 143*(3), 286-292. https://doi.org/10.1016/S0176-1617(11)81633-0

Gitelson, A.A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology, 161*(2), 165-173. https://doi.org/10.1078/0176-1617-01176

Gitelson, A.A., Gritz, Y., & Merzlyak, M.N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology, 160*(3), 271-282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A.A., Kaufman, Y.J., & Merzlyak, M.N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment, 58*(3), 289-298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gitelson, A.A., & Merzlyak, M.N. (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research, 22*(5), 689-692. https://doi.org/10.1016/S0273-1177(97)01133-2

Gitelson, A.A., Stark, R., Grits, U., Rundquist, D., Kaufman, Y., & Derry, D. (2002). Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction. *International Journal of Remote Sensing, 23*(13), 2537-2562. https://doi.org/10.1080/01431160110107806

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., & Strachan, I.B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green lai of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment, 90*(3), 337-352. https://doi.org/10.1016/j.rse.2003.12.013

Huete, A.R. (1988). A soil-adjusted vegetation index (savi). *Remote Sensing of Environment, 25*(3), 295-309. https://doi.org/10.1016/0034-4257(88)90106-X

Jiang, Z., Huete, A.R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment, 112*(10), 3833-3845. https://doi.org/10.1016/j.rse.2008.06.006

Le Maire, G., Francois, C., & Dufrene, E. (2004). Towards universal broad leaf chlorophyll indices using prospect simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment, 89*(1), 1-28. https://doi.org/10.1016/j.rse.2003.09.004

Lichtenthaler, H.K., Lang, M., Sowinska, M., Heisel, F., & Miehé, J.A. (1996). Detection of vegetation stress via a new high resolution fluorescence imaging system. *Journal of Plant Physiology, 148*(5), 599-612. https://doi.org/10.1016/S0176-1617(96)80081-2

Pinty, B., & Verstraete, M. (1992). Gemi: A non-linear index to monitor global vegetation from satellites. *Vegetation, 101*(1), 15-20. https://doi.org/10.1007/BF00031911

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment, 48*(2), 119-126. https://doi.org/10.1016/0034-4257(94)90134-1

Radoux, J., Chomé, G., Jacques, D.C., Waldner, F., Bellemans, N., Matton, N., Lamarche, C., D'Andrimont, R., & Defourny, P. (2016). Sentinel-2's potential for sub-pixel landscape feature detection. *Remote Sensing, 8*(6), 488. https://doi.org/10.3390/rs8060488

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment, 55*(2), 95-107. https://doi.org/10.1016/0034-4257(95)00186-7

Roujean, J.-L., & Breon, F.-M. (1995). Estimating par absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment, 51*(3), 375-384. https://doi.org/10.1016/0034-4257(94)00114-3

Tucker, C. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*(2) 127-150. https://doi.org/10.1016/0034-4257(79)90013-0

Yang, Z., Willis, P., & Mueller, R. (2008). *Impact of band-ratio enhanced awifs image to crop classification accuracy.* Paper presented at the Pecora 17 Remote Sensing Symposium, Denver, CO, USA.