Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

Sandino Mora, Juan, Pegg, Geoff, Gonzalez, Felipe, & Smith, Grant (2018)
Aerial mapping of forests affected by pathogens using UAVs, hyperspectral sensors, and artificial intelligence.
Sensors, *18*(4), Article number: 9441-17.

This file was downloaded from: https://eprints.qut.edu.au/223634/

© Consult author(s) regarding copyright matters

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

License: Creative Commons: Attribution 2.5

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

https://doi.org/10.3390/s18040944

MDPI

Article

Aerial Mapping of Forests Affected by Pathogens using UAVs, Hyperspectral Sensors and Artificial Intelligence

Juan Sandino[1,*], Geoff Pegg[2], Felipe Gonzalez[1] and Grant Smith[3]

[1] Insitute for Future Environments; Robotics and Autonomous Systems, Queensland University of Technology (QUT), 2 George St, Brisbane City, QLD 4000, Australia; felipe.gonzalez@qut.edu.au (F.G.)
[2] Horticulture & Forestry Science, Department of Agriculture & Fisheries, 80 Ann St, Brisbane City, QLD 4000, Australia; geoff.pegg@daf.qld.gov.au (G.P.)
[3] BioProtection Technologies, The New Zealand Institute for Plant & Food Research Limited, Gerald St, Lincoln 7608, New Zealand; grant.smith@plantandfood.co.nz (G.S.)
* Correspondence: j.sandinomora@qut.edu.au; Tel.: +61-7-3138-1363

Version March 17, 2018 submitted to Sensors

Abstract: The environmental and economic impacts of exotic fungal species on natural and plantation forests has been historically catastrophic. Recorded surveillance and control actions are challenging because they are costly, time-consuming, and hazardous in remote areas. Prolonged periods of testing and observation of site-based tests have limitations to verify the rapid proliferation of exotic pathogens and deterioration rates in hosts. Recent remote sensing approaches have offered fast, broad-scale and affordable surveys, and additional indicators that can complement on-ground tests. This paper proposes a framework that consolidates site-based insights and remote sensing capabilities to detect and segment deteriorations by fungal pathogens in natural and plantation forests. This approach is illustrated with an experimentation case of myrtle rust (*Austropuccinia psidii*) on paperbark tea trees (*Melaleuca quinquenervia*) in New South Wales (NSW), Australia. The method integrates unmanned aerial vehicles (UAVs), hyperspectral image sensors, and data processing algorithms using machine learning. Imagery is acquired using a Headwall Nano-Hyperspec® camera, orthorectified in Headwall SpectralView®, and processed in Python programming language using eXtreme Gradient Boosting (XGBoost), Geospatial Data Abstraction Library (GDAL) and Scikit-learn third-party libraries. In total, 11,385 samples were extracted and labelled into five classes: two classes for deterioration status and three classes for background objects. Insights reveal individual detection rates of 95% for healthy trees, 97% for deteriorated trees and a global multiclass detection rate of 97%. The methodology is versatile to be applied to additional datasets taken with different image sensors, and the processing of large datasets with freeware tools.

Keywords: *Austropuccinia psidii*; drones; hyperspectral camera; machine learning; *Melaleuca quinquenervia*; myrtle rust; non-invasive assessment; paperbark; unmanned aerial vehicles (UAV); xgboost.

1. Introduction

Exotic pathogens have caused irreversible damage to flora and fauna within a range of ecosystems worldwide. Popular outbreaks include the enormous devastations of chestnut blight (*Endothia parasitica*) on American chestnut trees (*Castanea dentata*) in the U.S. [1–3], sudden oak death (*Phytophthora ramorum*) on oak populations (*Quercus agrifolia*) in Europe, California and Oregon [4–6], dieback (*Phytophthora cinnamomi*) on hundreds of hosts globally [7–9], and myrtle

rust (*Austropuccinia psidii*) on Myrtaceae family plants in Australia [10–13]. The effects of the latter case have raised national alerts and response programmes given the extensive host range and the ecological and economic importance of Myrtaceae plants in the Australian environment [14–17]. As a result, various surveillance and eradication programmes have been applied in an attempt to minimise the impacts invasive pathogens cause on local hosts such as dieback in the Western Australia Jarrah forests [18], sudden oak death in the tan oak forests of the U.S. [19] and rapid ohia death (*Ceratocystis fimbriata*) on ohia trees (*Metrosideros polymorpha*) in Hawaii [20].

Modern surveillance methods to map hosts vulnerable to and affected by exotic pathogens can be classified in site-based and remote sensing methods, according to Lawley *et al.* [21]. Site-based approaches are commonly small regions used to collect exhaustive compositional and structural indicators of vegetation condition with a strong focus on biophysical attributes of single vegetation communities [22,23]. These methods, nonetheless, require deep expertise and time to conduct experimentation, data collection and validation that, along with their limited area they can cover, represent a challenge while assessing effects on a broad scale [24]. Research has also suggested the design of decision frameworks to monitor and control the most threatened species [17,25,26]. Although these models can determine flora species that require immediate management control, limitations on the amount of tangible, feasible and broad quantified data of vulnerable host areas [21] have resulted in lack of support from state and federal governments [27].

The role of remote sensing methods to assess and quantify the impacts of invasive pathogens in broad scale has increased exponentially [28,29]. Standard approaches comprise the use of spectral and image sensors through satellite, manned and unmanned aircraft technology [30]. Concerning sensing technology by itself, applied methods by research communities range from the use of non-imaging spectroradiometers, fluorescence, multi-, hyperspectral and thermal cameras, and light detection and ranging (LiDAR) technology [31–34]. These equipment are usually employed for the calculation of spectral indexes [35–38] and regression models in the host range [39,40]. Nevertheless, these methods are mainly focused on the quantification, distribution, among other physical properties of flora species.

Satellite and manned aircraft surveys have reported limitations concerning resolution, operational costs and unfavourable climate conditions (e.g. cloudiness and hazard winds) [41]. In contrast, the continuous development of unmanned aerial vehicles (UAVs) designs, navigation systems, portable image sensors and cutting-edge machine learning methods allow unobtrusive, accurate and versatile surveillance tools in precision agriculture and biosecurity [42–44]. Many studies have positioned UAVs for the collection of aerial imagery in applications such as weeds, diseases and pests mapping, and wildlife monitoring [21,45–47]. More recently, unmanned aerial systems (UASs) have been deployed in cluttered and global positioning system (GPS)-denied environments [48].

Approaches to the use of UAVs, hyperspectral imagery and artificial intelligence are gaining popularity. For example, Aasen *et al.* [49] deployed UAS to boost vegetation monitoring efforts using hyperspectral three-dimensional (3D) imagery. Nasi *et al.* [50] developed techniques to assess pest damages at canopy levels using spectral indexes and k-nearest neighbour (k-NN) classifiers, achieving global detection rates of 90%. Similar research focused on diseases monitoring, however, has been limited. Calderon *et al.* [51] for instance, evaluated the early detection and quantification of Verticillium Wilt in Olive plants using support vector machines (SVM) and linear discriminant analysis (LDA), obtaining mixed accuracy results among the evaluated classes of infection severity (59% to 75%) [52]. Albetis *et al.* [53] presented a system to discriminate asymptomatic and symptomatic red and white vineyards cultivars by *Flavescence doree*, using UAVs, multispectral imagery and up to 20 data features, collecting contrasting results between the cultivars and maximum accuracy rates of 88%. In sum, the integration of site-based and remote sensing frameworks have boosted the capabilities of these surveillance solutions by combining data from abiotic and biotic factors and spectral responses, respectively [54]. However, this synthesis is still challenging due to the high number of singularities, data correlation and validation procedures presented in each case study.

Considering the importance of site-based and remote sensing methods to obtain reliable and broader assessments of forest health, specifically, for pest and fungal assessments [55], this paper presents an integrated system that classifies and maps natural and plantation forests exposed and deteriorated by fungal pathogens using UAVs, hyperspectral sensors and artificial intelligence. The framework is exemplified by a case study of myrtle rust on paperbark tea trees (*Melaleuca quinquenervia*) in a swamp ecosystem of north-east New South Wales (NSW), Australia.

2. System Framework

A novel framework was designed for the assessment of natural and plantation forests exposed and potentially exposed to pathogens as presented in Figure 1. It comprises four sections linked each other denominated Data Acquisition, Data Preparation, Training and Prediction. The system interacts directly and indirectly with the surveyed area to acquire information, preprocess and arrange obtained data into features, fit a supervised machine learning classifier, tune the accuracy and performance indicators to process vast amounts of data, and provide prediction reports through segmented images.

Figure 1. Pipeline process for the detection and mapping of alterations in natural and plantation forests by fungal diseases.

2.1. Data Acquisition

The data acquisition process involves an indirect data collection campaign using an airborne system, and direct ground assessments of the studied pathogen through exclusion trials, which are controlled by biosecurity experts. Airborne data is compiled using a UAV, image sensors and a ground workstation in the site to acquire data above the tree canopy. Similarly, ground data is collected through field assessment insights by biosecurity experts. Field assessments bring several factors such as growth, reproduction and regeneration on coppiced trees exposed to any specific pathogen. A database is created by labelling, georeferencing and correlating relevant insights into every tested plant. Details on the studied area, flight campaign and field assessments can be found in Sections 3.1, 3.2 and 3.3.

2.1.1. UAV and Ground Station

This methodology incorporates a hexa-rotor DJI S800 EVO (DJI, Guangdong, China) UAV. The drone features high-performance brushless rotors, a total load capacity of 3.9 kg and dimensions of 118 × 100 × 50 cm. It follows an automatic mission route using the DJI Ground Station 4.0 software, controlling the route, speed, height above the ground, and overlapping values remotely. It is worth mentioning, however, that other UAVs with similar characteristics can also be used and included into the airborne data collection process of Figure 1.

2.1.2. Sensors

Spectral data is collected using a Headwall Nano-Hyperspec® hyperspectral camera (Headwall Photonics Inc., MA, USA). This visible near infra-red (VNIR) sensor provides spectral wavelength responses up to 274 bands, a wavelength range from 385 to 1000 nm, a spectral resolution of 2.2 nm, a frame rate of 300 Hz, spatial bands of 640 pixels and a total storage limit of 480 GB. The camera is mounted in a customised gimbal system that augments data quality by minimising external disturbances such as roll, pitch and yaw oscillations, as depicted in Figure 2.

Figure 2. Assembly of the gimbal system and the Headwall Nano-Hyperspec® camera into the DJI S800 EVO unmanned aerial vehicle (UAV).

2.2. Data preparation

Imagery is downloaded and fed into a range of software solutions to transform raw data into filtered and orthorectified spectral bands in reflectance. Using Headwall SpectralView® software, raw hypercubes from the surveyed area were automatically processed. The preprocessing operations included radiance, orthorectification (through ground control points) and the white reference illumination spectrum (WRIS). The WRIS spectrum comes from the extraction of the radiance signature of a Spectralon target from an acquired image in the surveyed area. Using the orthorectified imagery and the WRIS, reflectance data and scene shading are calculated through CSIRO | Data61 Scyven 1.3.0. software [56,57]. Considering the massive amount of data contained for any orthorectified hyperspectral cube (4-12 GB), each cube is cropped to process regions of interest only (0.5-1 GB) and handle with system resources efficiently, as shown in Figure 3.

Figure 3. Creation of regions of interest in hyperspectral cubes.

An additional image that contains the localisation of tested trees in the exclusion trial is created. From a recovered red, green, blue colour model (RGB) image of the cropped hypercube, each tree is graphically labelled using their tracked GPS coordinates in Argis ArcMap 10.4. To handle all the data processing from the proposed system framework, Algorithm 1 was developed. These tasks were conducted with Python 2.7.14 programming language and several third-party libraries for data manipulation and machine learning, including Geospatial Data Abstraction Library (GDAL) 2.2.2 [58], eXtreme Gradient Boosting (XGBoost) 0.6 [59], Scikit-learn 0.19.1 [60], OpenCV 3.3.0 [61] and Matplotlib 2.1.0 [62].

Algorithm 1 Detection and mapping of vegetation alterations using spectral imagery and sets of features.

Required: orthorectified layers (bands) in reflectance I. Labelled regions from field assessments L.
Data Preparation
Load I data.
$S \leftarrow$ Spectral indexes array from I.
$X \leftarrow$ Features array $[I, S]$.
Training
$Y \leftarrow$ Labels array from dataset L.
$D \leftarrow$ filtered dataset of features X with corresponding labelled pixels from Y.
Split D into training data D_T and testing data D_E.
Fit an XGBoost classifier C using D_T.
$R \leftarrow$ List of unique relevance values of processed features X from C.
for all values **in** R **do**
    $D_{TF} \leftarrow$ Filtered underscored features from D_T.
    Fit C using D_{TF}.
    Append accuracy values from C into T.
end for
Fit C using the best features threshold from T.
Validate C with k-fold cross validation from D_T. ▷ number of folds = 10
Prediction
$P \leftarrow$ Predicted values for each sample in X.
Convert P array into a 2D orthorectified image.
$O \leftarrow$ Displayed/overlayed image.
return O

Reflectance cube bands are loaded into the program to calculate suitable spectral indexes and improve the detection rates as mentioned in Step 2. For this approach, the most traditional indexes are calculated such as the normalised difference vegetation index (NDVI) [63], the green normalised difference vegetation index (GNDVI) [64], the soil-adjusted vegetation index (SAVI) [65] and the second modified adjusted vegetation index (MSAVI2) [66]. Additionally, two-dimensional (2D) smoothing kernels are applied into such indexes, following Equation (1).

$$K = \frac{1}{w^2} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{1}$$

where K is the kernel of the filter and w the size of the window. Here, up to three kernels for $w = 3$, $w = 7$ and $w = 15$ are calculated per vegetation index. All the hypercube bands in reflectance, as well as calculated vegetation spectral indexes, are denominated data features. In Step 3, an array of features is generated from all the retrieved bands I and the calculated indexes S.

2.3. Training and Prediction

The labelled regions from the ground-based assessments are exported from ArcMap and loaded into an array Y. In Step 5, an array D is created by filtering the features from X with their corresponding labels in Y only. Filtered data is separated into a training (80%) and testing (20%) data array. In step 7, data is processed into a supervised XGBoost classifier. This model is utilised considering the moderate amount of labelled data (insufficient to run a deep learning model), the amount of information to be processed for a single hypercube and the nature of data from the exclusion trial (ground-based test). This classifier is currently a cutting-edge decision tree and gradient boosting model optimised for large tree structures, excellent performance and fast execution speeds, outperforming detection rates of standard non-linear models such as random forests, k-NN, LDA and SVM [59]. Moreover, input data features do not require any scaling (normalisation) in comparison with other models. Once the model is fitted, Step 8 retrieves the relevance of each processed feature (reflectance bands, spectral indexes and transformed images) from the array X in a list R. The list is sorted to discard irrelevant features, increment the detection rates of the algorithm, avoid over-fitting, decrease the complexity of the model and reduce computer processing loads.

Algorithm 1 executes a loop to evaluate the optimal number of ranked features that can offer the best balance between accuracy and data processing load. At each instance of the loop, specific features from the training set D_T is filtered based on a threshold of relevance scores (Step 10). Later, the XGBoost model is fit using the filtered training set (Step 11) to records all the accuracy values per combination. In Step 14, the best combination of accuracy and number of features is retrieved to re-fit the classifier. Finally, the fitted model is validated using k-fold cross validation (Step 15).

In the prediction stage, unlabelled pixels are processed in the optimised classifier and their values displayed into the same 2D spatial image from the orthorectified hyperspectral cube. Ultimately, classified pixels are depicted using distinguishable colours and exported in tagged image file (TIF) format, a compatible file with georeferencing-based software.

3. Experimentation Setup

3.1. Site

As displayed in Figure 4a, the experimentation occurred in a paperbark tea tree forest located near 71 Boggy Creek Rd, Bungawalbin, NSW 2469, Australia (29°04′42.9″ S 153°15′10.0″ E). Data acquisition was conducted on the 25th of August, 2016 at 11:40 a.m. Weather Observations for that day stated conditions of a partially cloudy day, with a mean temperature of 18.8 °C, a relative humidity of

46%, west-north-west winds of 20 km/h, a pressure of 1009.5 hPa, and no precipitation [67]. As seen in Figure 4b, the site includes selected paperbark trees which are monitored to assess the effects of myrtle rust on them.

(a) **(b)**

Figure 4. Site of the case study. (**a**) Location and covered area of the farm in red and experiment trees in blue. (**b**) Overview of paperbark tea trees under examination.

3.2. Flight campaign

The acquired data from flight campaign incorporated a single mission route. The UAV was operated with a constant flight height of 20 m above the ground, an overlap of 80%, side lap of 50%, mean velocity of 4.52 km/h and a total distance of 1.43 km. The acquired hyperspectral dataset had a vertical and horizontal ground sample distances (GSD) of 4.7 cm/pixel.

3.3. Field Assessments

In order to evaluate the effects and flora response of myrtle rust on paperbark trees, a biological study in an exclusion trial on the mentioned site was conducted. Several on-ground assessments were conducted in individual trees within a replicated block design using four treatments: trees treated with fungicides (F), insecticides (I), fungicides and insecticides (F + I) and trees without any treatment action. Figure 5 illustrates the treatment methods the studied trees received.

Figure 5. Aerial view of individual trees through on-ground assessments.

From the indicators generated, insect and disease assessments were extracted to label every tree. Overall, the assessment report showed that only the trees that received insecticide and fungicide treatments remained healthy under direct exposure to the rust. In contrast, trees treated with insecticide were affected by rust and those treated with fungicide were affected by insects. Thus, trees treated with fungicides and insecticides were consequently labelled as healthy and the others as affected in the database.

3.4. Preprocessing

As mentioned in Section 2.2, the entire surveyed area included an exclusion trial. As a result, an orthorectified hyperspectral cube in reflectance with spatial dimensions of 2652 $\times$ 1882 pixels sized 5.4 GB was generated. This area was cropped from the original hypercube, reducing computational costs and discarding irrelevant data. Eventually, a 1.7 GB cube of 1200 $\times$ 1400 pixels as depicted in Figure 6 was extracted.

Figure 6. Red, green, blue (RGB) colour model representation of the orthorectified cube with its extracted region of interest.

3.5. Training and Prediction

The XGBoost classifier contains several hyper-parameters to be set. Following a grid search technique, in which the classifier performance and detection rates were tracked with a set of possible hyper-parameters, the optimal values found for this case study were:

$$\text{estimators} = 100, \ \text{learning rate} = 0.1, \ \text{maximum depth} = 3$$

where "estimators" is the number of trees, "learning rate" is the step size of each boosting step and "maximum depth" is the maximum depth per tree that defines the complexity of the model.

4. Results and Discussion

To visualise the benefits of inserting an optimisation scheme in Step 8 of Algorithm 1, detection rates were tracked by training and running the classifier multiple times with only a set of filtered features per instance. The features were ranked with their relevance by the XGBoost classifier and sorted consequently, as illustrated in Figure 7.

Figure 7. Performance of the classifier using different filtered features. Optimal number of features: 24.

The classifier can achieve high accuracy rates exceeding 97% of global accuracy when it processes data using from 10 to 40 features only, with an optimal number of features of 24. On the other hand, the classifier merely improves their registers when the number of processed features is more substantial. With this capability, the proposed approach can process fewer data and reduce the number of calculations to achieve high detection values. Additionally, this boosts the capability of the algorithm of processing large datasets in less time, an ideal scenario for mapping vast rural areas. The most relevant features of this study case are depicted in Table 1 and Figure 8.

Table 1. Ranking of the most 30 relevant features.

#	Feature	Score	#	Feature	Score	#	Feature	Score
1	NDVI_Mean15	0.0933	11	NDVI_Mean3	0.0247	21	975.3710	0.0119
2	Shading_Mean15	0.0780	12	759.9730	0.0212	22	671.1490	0.0109
3	GNDVI_Mean7	0.0563	13	999_Mean3	0.0212	23	893.2090	0.0109
4	NDVI_Mean7	0.0558	14	999_Mean7	0.0202	24	990.9150	0.0099
5	999_Mean15	0.0504	15	997.5770	0.0188	25	877.6650	0.0094
6	444.6470	0.0494	16	764.4140	0.0148	26	966.4890	0.0094
7	Specularity_Mean15	0.0380	17	444_Mean15	0.0143	27	766.6350	0.0084
8	999.7980	0.0341	18	462.4120	0.0133	28	853.2380	0.0079
9	GNDVI_Mean15	0.0286	19	NDVI	0.0133	29	935.4000	0.0079
10	444_Mean7	0.0267	20	Shading_Mean7	0.0133	30	GNDVI_Mean3	0.0079

Figure 8. False colour representation of the first six features by relevance.

It is shown how the first four features for this classification task come from specific vegetation indexes and processed images by 2D kernels, specifically, NDVI, shading and GNDVI features (Figure 8a,8b,8c,8d). Although their illustrations show insights of distinguishable intensities between healthy and affected tree regions from Figure 5, these set of features are insufficient to segment areas of other objects. Thus, specific reflectance wavelengths bands such as 999 and 444 nm (Figure 8e,8f) are also determinant. Also, features processed with 2D kernels obtained better relevance scores than their unprocessed counterparts. That difference was even greater for processed features using big window kernels considering that high amounts of noise, common in raw hyperspectral imagery, altered the performance of the approach. Nonetheless, these rankings do not suggest that these features can be used as global indicators to detect and map similar case studies (myrtle rust); the feature ranking table showed here is relevant to the fitted XGBoost model only and results may differ if the same features are processed through other machine learning techniques. It is recommended, therefore, to perform individual analyses for every case study.

A total of 11,385 pixels contained in 23 features filtered by their relevance were read again in Step 14 of Algorithm 1. Data was divided into a training array D_E with 9108 pixels and a testing array D_T with 2277 pixels. The generated confusion matrix of the classifier and its performance report is shown in Table 2 and Table 3.

Table 2. Confusion Matrix of the eXtreme Gradient Boosting classifier.

	Predicted	Healthy	Affected	Background	Soil	Stems
Labelled	Healthy	1049	15	0	0	0
	Affected	45	531	0	0	0
	Background	0	0	158	0	0
	Soil	0	0	0	321	0
	Stems	0	0	0	1	157

Table 3. Classification report of the confusion matrix of Table 2.

Class	Precision (%)	Recall (%)	F-score (%)	Support
Healthy	95.89	98.59	97.24	1064
Affected	97.25	92.19	94.72	576
Background	100.00	100.00	100.00	158
Soil	99.69	100.00	99.68	321
Stems	100.00	99.37	99.68	158
Mean	97.32	97.32	97.35	$\sum = 2277$

In sum, most of the classes were predicted favourably. The majority of misclassifications between the "Healthy" and "Affected" classes are possibly caused by human error while labelling the regions manually in the raw imagery. Considering a weighed importance of precision and recall of 1:1, the F-support scores highlight a detection rate of 97.24% for healthy trees, 94.25% for affected trees, and an overall detection rate of 97.35%.

The performance of Algorithm 1 was tested in a computer with the following characteristics: Processor Intel® Core™ i7-4770, 256 GB SSD, 16 GB RAM, Windows 7 64bit and AMD Radeon™ HD 8490. It contains a report of the elapsed seconds for the application to accomplish the primary data processing, training and prediction tasks, as illustrated in Table 4.

Table 4. Performance in seconds of the main tasks from Algorithm 1.

Sub-section	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5	Mean	Std. dev.
Data preparation							
Loading Hypercube	11.927	10.944	11.954	11.766	11.521	11.622	0.417
Calculating indexes	46.864	51.860	51.901	52.622	47.322	50.114	2.779
Training							
Preprocessing	0.152	0.141	0.149	0.148	0.140	0.146	0.005
Fitting XGBoost	8.948	8.654	8.758	8.679	8.692	**8.746**	0.119
Features Filtering	53.236	55.433	60.364	57.253	53.446	55.946	2.962
Re-Fitting XGBoost	0.964	1.023	1.010	0.998	0.965	**0.992**	0.026
Prediction							
Predicting results	29.738	40.749	42.131	34.473	66.477	42.714	14.188
Display	0.776	0.705	1.043	0.917	0.612	0.811	0.171
Total	152.607	169.508	177.309	166.857	189.175	171.091	13.489

Taking into account the dimensions of the processed hypercube (1400 $\times$ 1200) and the initial number of bands (274), it is observed how a great demand of resources were invested to open the file itself and calculate spectral indexes, accumulating 61.7 seconds, in average. Similarly, the features filtering process in the training section also demanded considerable time, exceeding 50 seconds. On the other hand, the elapsed time executing the remaining tasks of the training phase are remarkably small. Specifically, the report highlights the benefits of filtering irrelevant features by comparing the duration of fitting the classifier for the first time with the duration of re-fitting it again with less yet relevant data from 8.74 to 0.99 seconds. Overall, the application spent 2 minutes and 51 seconds to evaluate and map an area of 338 m^2 approximately.

The GSD value of 4.7 cm/pixel from the acquired hyperspectral imagery represented a minor challenge to label individual trees, but is still problematic when highlighting specific stems or leaflets is needed. Higher resolution can assist in higher classification rates. As an illustration, the final segmented image of the optimised classifier is shown in Figure 9, where Figure 9a shows the digital labelling of every class region and Figure 9b depicts the generated segmentation output by Algorithm 1. A hypercube covering the entire area flown was also processed using the trained model with results shown in Figure 10.

(a) **(b)**

Figure 9. Segmentation results of the proposed approach. (**a**) Recovered hyperspectral image in red, green, blue (RGB) colour model.(**b**) Segmentation results.

Figure 10. Layer of mapping results of the study area in Google Earth.

Results show a segmentation output using XGBoost as the supervised machine learning classifier that works well for this task. This classifier as well as Algorithm 1 are not only important for their capabilities to offer a pixel-wise classification task, but they also allow a rapid convergence, do not involve many complex mathematical calculations and filter irrelevant data, compared to other methods. Nevertheless, it is suggested to revise their prediction performance with new data. Like any model based on decision trees, over-fitting may occur, and misleading results might be generated. In those situations, labelling misclassified data, aggregating them into the features database and rerunning the algorithm is suggested.

The availability to process and classify data with small GSD values demonstrates the potential of UASs for remote sensing equipment compared with satellite and manned aircraft for forest health assessments on forest and tree plantations, and traditional estimation methods, such as statistical regression models in regions of study. In comparison with similar approaches of non-invasive assessment techniques using UAVs and spectral sensors, this framework does not provide general spectral indexes that can be applied with different classifiers and similar evaluations. In contrast, this presented method boosts the efficiency of the classifier by receiving feedback from the accuracy scores of every feature and transforming the input data in consequence. The more explicit the data for the classifier is, the better the classification rates are. Furthermore, it is also demonstrated that a classifier which processes and combines data from multiple spectral indexes provides better performance than analysing individual components from different source sensors.

5. Conclusions

This paper describes a pipeline methodology for effective detection and mapping of indicators of poor health in forest and plantation trees integrating UAS technology and artificial intelligence approaches. The techniques were illustrated with an accurate classification and segmentation task of paperbark tea trees deteriorated by myrtle rust from an exclusion trial in NSW, Australia. Here, the system achieved detection rates of 97.24% for healthy trees and 94.72% for affected trees. The algorithm obtained a multiclass detection rate of 97.35%. Data labelling is a task that demands many resources from both site-based and remote sensing methods, affecting smoothly, the accuracy and reliability of the classifier results due to human errors.

The approach can be used to train various datasets from different sensors to improve detection rates that single solutions offer as well as the capability of processing large datasets using free-ware software. The case study demonstrates an effective approach that allows rapid and accurate indicators, and alterations of exposed areas at early stages. However, understanding disease epidemiology and interactions between the pathogen and host is still required to enable the effective use of these technologies.

Future research should discuss the potential of monitoring the evolution of affected species through time, the prediction of expansion directions and rates of the disease, and how data will contribute on improving control actions to deter their presence in unaffected areas. Technologically, future works should analyse and compare the efficacy of unsupervised algorithms to label vegetation items accurately, integrate the best approaches in the proposed pipeline and evaluate regression models that predict data based on other biophysical information offered by site-based methods.

Acknowledgments: This work was funded by the Plant Biosecurity Cooperative Research Centre (PBCRC) 2135 project. The authors would like to acknowledge Jonathan Kok for his contributions in co-planning the experimentation phase. We also gratefully acknowledge the support of the Queensland University of Technology (QUT) Research Engineering Facility (REF) Operations team (Dirk Lessner, Dean Gilligan, Gavin Broadbent and Dmitry Bratanov), who operated the DJI S800 EVO UAV and image sensors, and performed ground referencing. We thank Gavin Broadbent for the design, manufacturing and tuning of a customised 2-axis gimbal for the spectral cameras. We acknowledge the High-Performance Computing and Research Support Group at QUT, for the computational resources and services used in this work.

Author Contributions: Felipe Gonzalez and Geoff Pegg contributed to experimentation and data collection planning. Felipe Gonzalez supervised the airborne surveys, quality of acquired data and logistics. Juan Sandino designed the proposed pipeline and conducted the data processing phase. Felipe Gonzalez, Geoff Pegg and Grant Smith provided definitions, assistance and essential advice. Juan Sandino analysed the generated outputs, and validated and optimised the algorithm. All the authors contributed significantly to the composition and revision of the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

2D	Two-dimensional
3D	Three-dimensional
F	Fungicides
F + I	Fungicides and Insecticides
GDAL	Geospatial data abstraction library
GPS	Global positioning system
GNDVI	Green normalised difference vegetation index
GSD	Ground sampling distance
I	Insecticides
k-NN	k-nearest neighbours
LDA	Linear discriminant analysis
LiDAR	Light detection and ranging
MDPI	Multidisciplinary digital publishing institute
MSAVI2	Second modified soil-adjusted vegetation index
NDVI	Normalised difference vegetation index
NSW	New South Wales
RGB	Red, green, blue colour model
SAVI	Soil-adjusted Vegetation Index
SVM	Support Vector Machines
TIF	Tagged Image File
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
VNIR	Visible Near Infra-red
WRIS	White Reference Illumination Spectrum
XGBoost	eXtreme Gradient Boosting

References

1. Smock, L.A.; MacGregor, C.M. Impact of the American Chestnut Blight on Aquatic Shredding Macroinvertebrates. *Journal of the North American Benthological Society* **1988**, *7*, 212–221.
2. Anagnostakis, S.L. Chestnut Blight: The Classical Problem of an Introduced Pathogen. *Mycologia* **1987**, *79*, 23.
3. Burke, K.L. The effects of logging and disease on American chestnut. *Forest Ecology and Management* **2011**, *261*, 1027–1033.
4. Rizzo, D.M.; Garbelotto, M.; Hansen, E.M. Phytophthora ramorum: Integrative Research and Management of an Emerging Pathogen in California and Oregon Forests. *Annual Review of Phytopathology* **2005**, *43*, 309–335.
5. Frankel, S.J. Sudden oak death and Phytophthora ramorum in the USA: A management challenge. *Australasian Plant Pathology* **2008**, *37*, 19–25.
6. Grünwald, N.J.; Garbelotto, M.; Goss, E.M.; Heungens, K.; Prospero, S. Emergence of the sudden oak death pathogen Phytophthora ramorum. *Trends in Microbiology* **2012**, *20*, 131–138.
7. Hardham, A.R. Phytophthora cinnamomi. *Molecular Plant Pathology* **2005**, *6*, 589–604.
8. Shearer, B.L.; Crane, C.E.; Barrett, S.; Cochrane, A. Phytophthora cinnamomi invasion, a major threatening process to conservation of flora diversity in the South-west Botanical Province of Western Australia, 2007.
9. Burgess, T.I.; Scott, J.K.; Mcdougall, K.L.; Stukely, M.J.; Crane, C.; Dunstan, W.A.; Brigg, F.; Andjic, V.; White, D.; Rudman, T.; Arentz, F.; Ota, N.; Hardy, G.E.J. Current and projected global distribution of Phytophthora cinnamomi, one of the world's worst plant pathogens. *Global Change Biology* **2017**, *23*, 1661–1674.
10. Pegg, G.S.; Giblin, F.R.; McTaggart, A.R.; Guymer, G.P.; Taylor, H.; Ireland, K.B.; Shivas, R.G.; Perry, S. Puccinia psidii in Queensland, Australia: Disease symptoms, distribution and impact. *Plant Pathology* **2014**, *63*, 1005–1021.

11. Carnegie, A.J.; Kathuria, A.; Pegg, G.S.; Entwistle, P.; Nagel, M.; Giblin, F.R. Impact of the invasive rust Puccinia psidii (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. *Biological Invasions* **2016**, *18*, 127–144.
12. Howard, C.; Findlay, V.; Grant, C. Australia's transition to management of myrtle rust. *Journal of Forest Science* **2016**, *61*, 138–139.
13. Fernandez Winzer, L.; Carnegie, A.J.; Pegg, G.S.; Leishman, M.R. Impacts of the invasive fungus Austropuccinia psidii (myrtle rust) on three Australian Myrtaceae species of coastal swamp woodland. *Austral Ecology* **2017**.
14. Dayton, L.; Higgins, E. Myrtle rust 'biggest threat to ecosystem', 2011.
15. Carnegie, A.J.; Cooper, K. Emergency response to the incursion of an exotic myrtaceous rust in Australia. *Australasian Plant Pathology* **2011**, *40*, 346–359.
16. Carnegie, A.J. First Report of Puccinia psidii (Myrtle Rust) in Eucalyptus Plantations in Australia. *Plant Disease* **2015**, *99*, 161–161.
17. Pegg, G.; Taylor, T.; Entwistle, P.; Guymer, G.; Giblin, F.; Carnegie, A. Impact of Austropuccinia psidii (myrtle rust) on Myrtaceae-rich wet sclerophyll forests in south east Queensland. *PloS one* **2017**, *12*, e0188058.
18. Government of Western Australia. Phytophthora dieback - Parks and Wildlife Service. http://www.webcitation.org/6xLA86qjW. (Archived on 19-02-2018).
19. U.S. Forest Service. Sudden Oak Death (SOD) | Partnerships | PSW Research Station | Forest Service. http://www.webcitation.org/6xLDwPURd. (Archived on 19-02-2018).
20. State of Hawaii. Department of Agriculture | How to report suspected Ohia Wilt/Rapid Ohia Death. http://www.webcitation.org/6xLCVG70h. (Archived on 19-02-2018).
21. Lawley, V.; Lewis, M.; Clarke, K.; Ostendorf, B. Site-based and remote sensing methods for monitoring indicators of vegetation condition: An Australian review. *Ecological Indicators* **2016**, *60*, 1273 – 1283.
22. Oliver, I.; Smith, P.L.; Lunt, I.; Parkes, D. Pre-1750 vegetation, naturalness and vegetation condition: What are the implications for biodiversity conservation? *Ecological Management & Restoration* **2002**, *3*, 176–178.
23. Lawley, V.; Parrott, L.; Lewis, M.; Sinclair, R.; Ostendorf, B. Self-organization and complex dynamics of regenerating vegetation in an arid ecosystem: 82 years of recovery after grazing. *Journal of Arid Environments* **2013**, *88*, 156 – 164.
24. Ostendorf, B. Overview: Spatial information and indicators for sustainable management of natural resources. *Ecological Indicators* **2011**, *11*, 97 – 102. Spatial information and indicators for sustainable management of natural resources.
25. Roux, J.; Germishuizen, I.; Nadel, R.; Lee, D.J.; Wingfield, M.J.; Pegg, G.S. Risk assessment for Puccinia psidii becoming established in South Africa. *Plant Pathology* **2015**, *64*, 1326–1335.
26. Berthon, K.; Esperon-Rodriguez, M.; Beaumont, L.; Carnegie, A.; Leishman, M. Assessment and prioritisation of plant species at risk from myrtle rust (Austropuccinia psidii) under current and future climates in Australia. *Biological Conservation* **2018**, *218*, 154 – 162.
27. Carnegie, A.J.; Kathuria, A.; Pegg, G.S.; Entwistle, P.; Nagel, M.; Giblin, F.R. Impact of the invasive rust Puccinia psidii (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. *Biological Invasions* **2016**, *18*, 127–144.
28. Lausch, A.; Erasmi, S.; King, D.J.; Magdon, P.; Heurich, M. Understanding Forest Health with Remote Sensing -Part I –A Review of Spectral Traits, Processes and Remote-Sensing Characteristics. *Remote Sensing* **2016**, *8*.
29. Tuominen, J.; Lipping, T.; Kuosmanen, V.; Haapanen, R. Remote Sensing of Forest Health. In *Geoscience and Remote Sensing*; Ho, P.G.P., Ed.; InTech: Rijeka, 2009; chapter 02.
30. Lausch, A.; Erasmi, S.; King, D.J.; Magdon, P.; Heurich, M. Understanding Forest Health with Remote Sensing-Part II–A Review of Approaches and Data Models. *Remote Sensing* **2017**, *9*.
31. Cui, D.; Zhang, Q.; Li, M.; Zhao, Y.; Hartman, G.L. Detection of soybean rust using a multispectral image sensor. *Sensing and Instrumentation for Food Quality and Safety* **2009**, *3*, 49–56.
32. Candiago, S.; Remondino, F.; De Giglio, M.; Dubbini, M.; Gattelli, M. Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. *Remote Sensing* **2015**, *7*, 4026–4047.

33. Lowe, A.; Harrison, N.; French, A.P. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* **2017**, *13*, 80.
34. Khanal, S.; Fulton, J.; Shearer, S. An overview of current and potential applications of thermal remote sensing in precision agriculture. *Computers and Electronics in Agriculture* **2017**, *139*, 22 – 32.
35. Devadas, R.; Lamb, D.W.; Simpfendorfer, S.; Backhouse, D. Evaluating ten spectral vegetation indices for identifying rust infection in individual wheat leaves. *Precision Agriculture* **2009**, *10*, 459–470.
36. Ashourloo, D.; Mobasheri, M.; Huete, A. Developing Two Spectral Disease Indices for Detection of Wheat Leaf Rust (Pucciniatriticina). *Remote Sensing* **2014**, *6*, 4723–4740.
37. Wang, H.; Qin, F.; Liu, Q.; Ruan, L.; Wang, R.; Ma, Z.; Li, X.; Cheng, P.; Wang, H. Identification and Disease Index Inversion of Wheat Stripe Rust and Wheat Leaf Rust Based on Hyperspectral Data at Canopy Level. *Journal of Spectroscopy* **2015**, *2015*, 1–10.
38. Heim, R.H.J.; Wright, I.J.; Chang, H.C.; Carnegie, A.J.; Pegg, G.S.; Lancaster, E.K.; Falster, D.S.; Oldeland, J. Detecting myrtle rust (Austropuccinia psidii) on lemon myrtle trees using spectral signatures and machine learning. *Plant Pathology* **2018**, pp. n/a–n/a.
39. Booth, T.H.; Jovanovic, T. Assessing vulnerable areas for Puccinia psidii (eucalyptus rust) in Australia. *Australasian Plant Pathology* **2012**, *41*, 425–429.
40. Elith, J.; Simpson, J.; Hirsch, M.; Burgman, M.A. Taxonomic uncertainty and decision making for biosecurity: spatial models for myrtle/guava rust. *Australasian Plant Pathology* **2013**, *42*, 43–51.
41. Salami, E.; Barrado, C.; Pastor, E. UAV Flight Experiments Applied to the Remote Sensing of Vegetated Areas. *Remote Sensing* **2014**, *6*, 11051–11081.
42. Glassock, R.; Hung, J.Y.; Gonzalez, L.F.; Walker, R.A. Design, modelling and measurement of a hybrid powerplant for unmanned aerial systems. *Australian Journal of Mechanical Engineering* **2008**, *6*, 69–78.
43. Gonzalez, L.; Whitney, E.; Srinivas, K.; Periaux, J. Multidisciplinary aircraft design and optimisation using a robust evolutionary technique with variable fidelity models. 2004, Vol. 6, pp. 3610–3624.
44. Ken, W.; H., H.C. Remote sensing of the environment with small unmanned aircraft systems (UASs), part 1: a review of progress and challenges. *Journal of Unmanned Vehicle Systems* **2014**, *02*, 69–85.
45. Gonzalez, L.; Montes, G.; Puig, E.; Johnson, S.; Mengersen, K.; Gaston, K. Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation. *Sensors* **2016**, *16*, 97.
46. Sandino, J.; Wooler, A.; Gonzalez, F. Towards the Automatic Detection of Pre-Existing Termite Mounds through UAS and Hyperspectral Imagery. *Sensors* **2017**, *17*, 2196.
47. Vanegas, F.; Bratanov, D.; Powell, K.; Weiss, J.; Gonzalez, F. A Novel Methodology for Improving Plant Pest Surveillance in Vineyards and Crops Using UAV-Based Hyperspectral and Spatial Data. *Sensors* **2018**, *18*.
48. Vanegas, F.; Gonzalez, F. Enabling UAV Navigation with Sensor and Environmental Uncertainty in Cluttered and GPS-Denied Environments. *Sensors* **2016**, *16*.
49. Aasen, H.; Burkart, A.; Bolten, A.; Bareth, G. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *108*, 245 – 259.
50. Nasi, R.; Honkavaara, E.; Lyytikainen-Saarenmaa, P.; Blomqvist, M.; Litkey, P.; Hakala, T.; Viljanen, N.; Kantola, T.; Tanhuanpaa, T.; Holopainen, M. Using UAV-Based Photogrammetry and Hyperspectral Imaging for Mapping Bark Beetle Damage at Tree-Level. *Remote Sensing* **2015**, *7*, 15467–15493.
51. Calderon, R.; Navas-Cortes, J.; Lucena, C.; Zarco-Tejada, P. High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sensing of Environment* **2013**, *139*, 231 – 245.
52. Calderon, R.; Navas-Cortes, J.A.; Zarco-Tejada, P.J. Early Detection and Quantification of Verticillium Wilt in Olive Using Hyperspectral and Thermal Imagery over Large Areas. *Remote Sensing* **2015**, *7*, 5584–5610.
53. Albetis, J.; Duthoit, S.; Guttler, F.; Jacquin, A.; Goulard, M.; PoilvÃ©, H.; FÃ©ret, J.B.; Dedieu, G. Detection of Flavescence dorÃ©e Grapevine Disease Using Unmanned Aerial Vehicle (UAV) Multispectral Imagery. *Remote Sensing* **2017**, *9*.
54. Pause, M.; Schweitzer, C.; Rosenthal, M.; Keuck, V.; Bumberger, J.; Dietrich, P.; Heurich, M.; Jung, A.; Lausch, A. In Situ/Remote Sensing Integration to Assess Forest Health – A Review. *Remote Sensing* **2016**, *8*.
55. Stone, C.; Mohammed, C. Application of Remote Sensing Technologies for Assessing Planted Forests Damaged by Insect Pests and Fungal Pathogens: a Review. *Current Forestry Reports* **2017**, *3*, 75–92.

56. Habili, N.; Oorloff, J. Scyllarus: From Research to Commercial Software. Proceedings of the ASWEC 2015 24th Australasian Software Engineering Conference; ACM Press: New York, New York, USA, 2015; Vol. II, pp. 119–122.
57. Gu, L.; Robles-Kelly, A.A.; Zhou, J. Efficient Estimation of Reflectance Parameters From Imaging Spectroscopy. *IEEE Transactions on Image Processing* **2013**, *22*, 3648–3663.
58. GDAL Development Team. *GDAL - Geospatial Data Abstraction Library, Version 2.1.0*. Open Source Geospatial Foundation, 2017.
59. Chen, T.; Guestrin, C. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16; ACM Press: New York, New York, USA, 2016; KDD '16, pp. 785–794.
60. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
61. Bradski, G. The OpenCV library. *Dr. Dobb's Journal of Software Tools* **2000**.
62. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9*, 90–95.
63. Rouse, Jr., J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with Erts. *NASA Special Publication* **1974**, *351*, 309.
64. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* **1996**, *58*, 289 – 298.
65. Huete, A. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **1988**, *25*, 295 – 309.
66. Laosuwan, T.; Uttaruk, P. Estimating Tree Biomass via Remote Sensing, MSAVI 2, and Fractional Cover Model. *IETE Technical Review* **2014**, *31*, 362–368.
67. Australian Government. Evans Head, NSW - August 2016 - Daily Weather Observations, 2016.

© 2018 by the authors. Submitted to *Sensors* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).