Enhancing Survival Surveys: Multispectral vs. RGB UAV Imaging.

Summary: This report evaluates a survival survey methodology employing multispectral data acquired from Unmanned Aerial Vehicles (UAVs). The study utilised a commercial deep learning-based seedling detection service to identify seedlings, comparing this to ground truth carried out with traditional field plotting techniques. Results were compared with additional data captured using a high-resolution full colour (RGB) camera to evaluate any potential advantaged of multispectral sensors versus standard RGB sensors.

The study revealed that the multispectral camera enabled the detection of more seedlings than the RBG camera. The multispectral camera also displayed a notably heightened sensitivity to other weed species resulting in increased erroneous detections, reducing accuracy and precision. Attempts to improve the accuracy of the detections through filtering the detection confidence levels of the annotations saw significant reductions in the number of erroneous detections and increases in precision and accuracy. Nevertheless, these refinements remained inferior to results obtained through RGB-based detections.

A proof of concept was also conducted to assess seedling health using vegetation indices, such as simple ratio (SR) and normalised difference vegetation index (NDVI), with machine learning methods. Results showed that different health classes were able to be discerned with a high level of accuracy and moderate Kappa scores. The proof of concept demonstrated the potential of these methods for deriving additional insights from multispectral imagery increasing its utility to foresters.

Overall, while multispectral imagery was deemed to be less effective than high resolution RGB imagery for seedling detection, the results of the health assessment indicate that with further research, multispectral imagery holds potential to enhance survival survey methodologies for more informed and efficient forest management practices.

Keywords: UAV imagery, multispectral imagery, seedling detection, health assessment, forest management.

Authors: Blake A. Singleton, Robin J.L. Hartley and Sam J. Davidson.

1. Introduction

The successful establishment of seedlings following planting is a critical factor in the overall outcome of forestation efforts (Grossnickle, 2012). This early phase of growth involves significant expenses and often lacks immediate economic returns, making it a pivotal stage in forest development (Montagnini, et al., 1995). Accurate assessment of tree health and stocking during this period is essential for informed forest management decisions and provides nurseries with valuable insights into seedling performance (Neumann, et al., 1995).

Recent years have witnessed a growing adoption of geospatial technologies in the forestry industry, with Unmanned Aerial Vehicles (UAVs) gaining traction, especially in New Zealand's plantation forest management (De Gouw, et al., 2020). UAVs offer a cost-effective and efficient alternative to traditional ground-based surveying methods (Feduck, et al., 2018). Additionally, they provide a potentially safer and more accurate means of assessment compared to traditional approaches. UAVs also enable more comprehensive coverage of forest stands. A previous study demonstrated that high resolution full-colour (red/green/blue, RGB) imagery and deep learning algorithms could be used for detecting seedlings with a high level of accuracy (Pearse, et al., 2020).

This study aims to evaluate the accuracy of utilising multispectral UAV imagery for seedling detection. It also seeks to conduct a proof of concept looking at the additional benefits of multispectral imagery, i.e. health assessments. Recognising the lack of standard operating procedures (SOPs) for seedling health assessment and the need for more robust techniques,

this research endeavours to contribute to improved practices in the field.

This research aligns with current trends in forest management, emphasising the potential for enhanced precision and efficiency in the use of multispectral UAV imagery in modern forest management practices.

2. Objective

The primary objective of this study was to evaluate the potential advantages of using multispectral imagery in contrast to high-resolution RGB imagery obtained through UAVs for seedling detection. This entailed a comprehensive exploration of the benefits and potential limitations of multispectral imagery in the context of precise seedling detection.

The secondary objective of study was to conduct a proof of concept investigating the feasibility of employing the random forest (RF) (Breiman, 2001) machine learning (ML) technique for the multi-class classification of seedling health status.

3. Methods

3.1. Study sites

This study was conducted at two locations within the North Island of New Zealand, as depicted in Figure 1. These sites were managed by Manulife Forest Management (NZ) Ltd (MFMNZ), and their selection was deliberate to demonstrate efficacy of these technologies across a range of environmental conditions and terrain classes.

The first study site was situated in Pipiwai Forest in Northland. The ~24 ha stand was planted in May 2022 and was characterised by its diverse topography ranging from steep slopes to rolling terrain, providing a valuable setting for examining the impact of varying topography on seedling detection and health assessment. The second study site, a ~24 ha stand located on Topaz Road in Kinleith Forest, Waikato, was established in September 2022 and features terrain that ranges from flat to rolling. This site was chosen as its distinct topographic characteristics complemented those found in the Pipiwai Forest site.
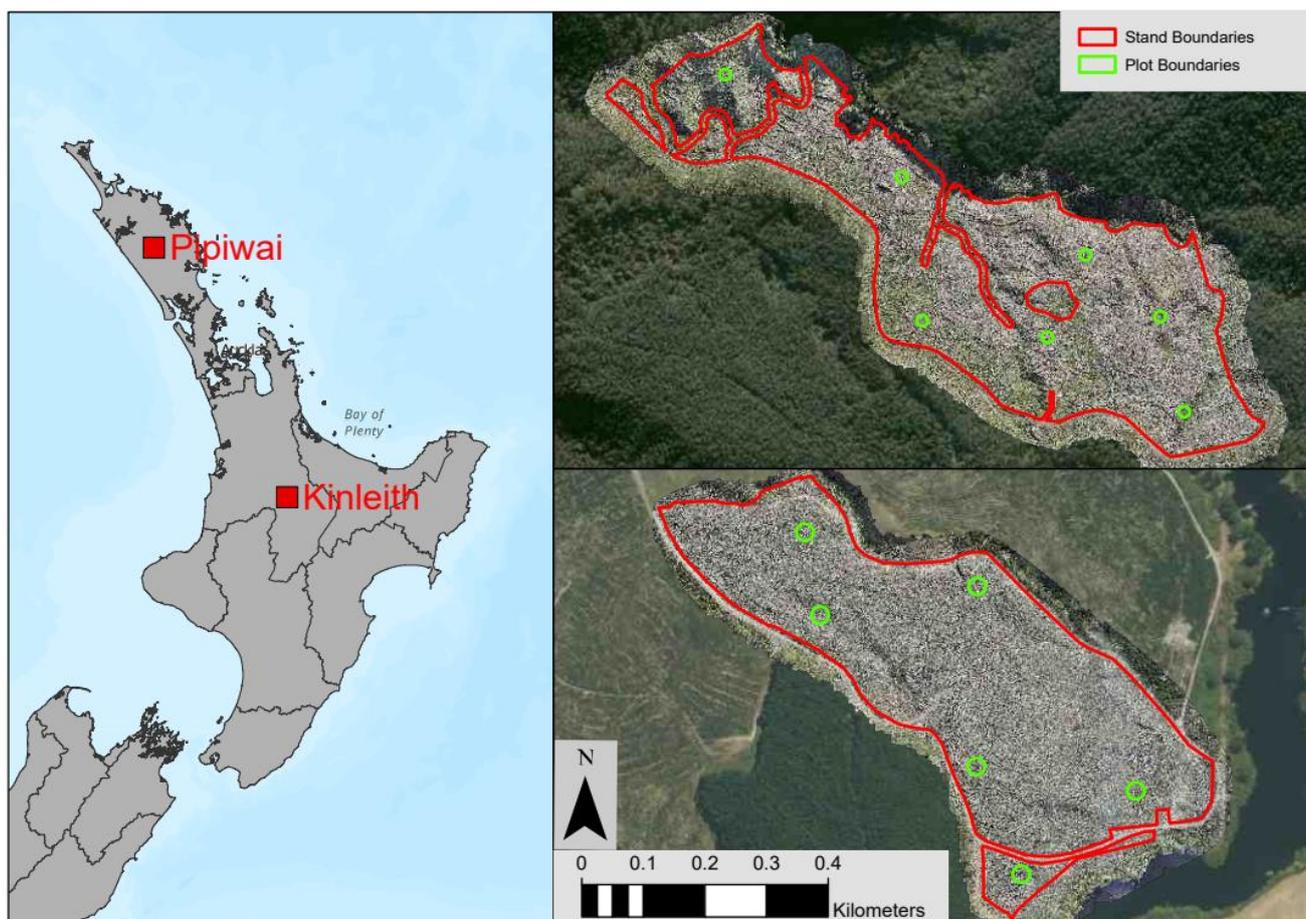
Figure 1. Map showing the high resolution UAV orthomosaics of the two study sites Pipiwai Forest (upper right) and Kinleith Forest (lower right) and their location within the North Island of NZ (left). Ground plot locations for the study are also indicated in green.

3.2. Field data capture methods

Seven 0.02 ha ground truth plots were measured at the Pipiwai site while six 0.06 ha ground truth plots were measured at the Kinleith site. Whilst the plots were erroneously measured to different sizes, this had little bearing on the results.

Each plot consisted of a survival count of the trees within the bounds of the plot. Each tree was assigned a code from the following five-point scale: 'gone', 'dead', 'dying', 'alive', or 'vigorous'. Tree assessments were taken starting at the western end of the row that was due north of the plot centre as shown in Figure 1.
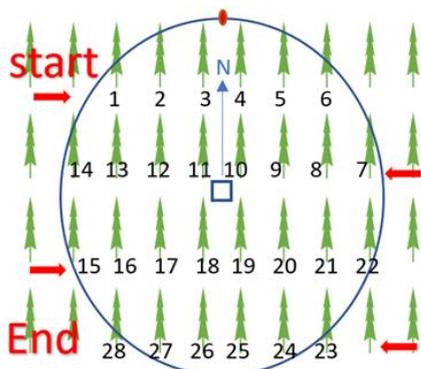
Figure 2. Plot assessment diagram for ground truth data capture, showing the start and end points for measuring plots relative to the plot centre.

When rows did not run east to west, as per Figure 2, the corrected bearing was recorded to assist the co-registration of the UAV data with ground truth data. Except for marking the plot centres with paint or high contrast targets, no paint or other markings were applied within or surrounding ground truth plots to avoid bias in subsequent analyses.

A total of 148 and 267 tree crowns (or absence thereof) were delineated within the Pipiwai and Kinleith sites, respectively. Table 1 displays a breakdown of the health status at each site.

Table 1. Delineated seedlings by health status and site

Site	Gone	Dying	Dead	Alive	Vigorous	Total
Pipiwai	26	24	2	90	6	148
Kinleith	3	1	3	5	255	267

Due to the high level of class imbalance at the Kinleith site, only the Pipiwai dataset was used in the machine learning classification of health status was unable to be applied to the Kinleith dataset.

3.3. Remote sensing capture methods

Multispectral datasets were collected using a MicaSense RedEdge MX dual-camera system (Micasense, Inc., Seattle, WA, USA) mounted on a DJI M300 RTK UAV (DJI Ltd., Shenzhen, China). The system was programmed to capture data at an altitude of 25 m above ground level (AGL), creating imagery with a Ground Sample Distance (GSD) of 1.5 cm. This altitude was chosen as it provided the smallest GSD that was practical to obtain with this sensor. Data was captured at Kinleith on 2nd June 2023, and Pipiwai on 26th April 2023.

Additionally, "wall-to-wall" high resolution RGB imagery datasets, capturing the entirety of each stand, were collected using a DJI P1 45 MP camera (DJI Ltd., Shenzhen, China) mounted on a DJI Matrice 300 pro. The system was configured to capture data at an altitude of 85 m AGL, capturing data with a GSD of 0.07 cm. Data was captured at Kinleith on 15th August 2023, and Pipiwai on 25th May 2023.

Flight routes were planned and controlled using the UGCS (SPH Engineering, Riga, Latvia) software to include a buffer zone around the ground truth plots. Flights were planned with a forward and side overlap between images of 80%. The key flight specifications and sensor systems used at each site are displayed in Table 2.

Table 2. UAV Flight specifications for each site.

Site	Craft	sensor	Altitude (AGL; m)	GSD (cm)	Flight speed (m/s)	Overlap % (fw:side)	Planned flight time (hh:mm)
Pipiwai	DJI Matrice 210	MicaSense RedEdge MX Dual	25	1.5	3	80:80	00:09
	DJI Matrice 300	DJI P1	85	0.7	5	80:80	01:15
Kinleith	DJI Matrice 300	MicaSense RedEdge MX Dual	25	1.5	3	80:80	00:09
	DJI Matrice 300	DJI P1	85	0.7	5	80:80	01:15

The MicaSense system comprises ten spectral bands, enabling the discrimination of different surface characteristics and the calculation of various vegetation indices. The DJI P1 sensor captured only 3 bands in the visible light range (red, green and blue - RGB). The spectral resolution and associated wavelengths of each of the sensors is shown in Table 3.

Table 3. Spectral resolution specifications for the cameras utilised in this study.

Sensor	Band Name	Central Wavelength (nm)	Band Width (nm)
	Coastal Blue	444	28
	Blue	475	32
	Green	531	14
MicaSense	Green	560	27
RedEdge	Red	650	16
MX Dual	Red	668	14
	Red Edge	705	10
	Red Edge	717	12
	Red Edge	740	18
	NIR	842	57
	Red	660	NA
DJI P1	Blue	550	NA
	Green	470	NA

Prior to each flight a calibration panel was imaged with the MicaSense sensor to ensure correct radiometric measurements were used in post processing. This was achieved by holding the UAV above the calibration panel and manually triggering a photo for each array.

3.4. Data processing

Individual images were radiometrically calibrated and combined to generate orthomosaics using Pix4Dmapper software using the default processing settings. No ground control points (GCPs) were used for the high-resolution data. These were not required as the UAV system comprised a real time kinematic (RTK) global navigational satellite system (GNSS), enabling higher positional accuracy. To keep the assessment of the data capture as realistic as possible, no GCPs were used in the multispectral data capture either.

To ensure accurate alignment and integration of these orthomosaics, we applied a 'spline' transformation. This transformation was executed using the 'spline' transformation tool within the ArcGIS Pro v3.2.0 toolset (ESRI Inc., Redlands, CA).

3.5. Data analysis: Individual tree detection and delineation

Both multispectral and High-Resolution RGB UAV imagery raster layers were imported into a geographic information system (GIS). The centre points of the plots were annotated in the GIS identified by the markers placed in the field. Plot boundaries were determined by creating a buffer around the plot centres with the appropriate radius. Individual tree crowns were manually delineated within the GIS. This process integrated information derived from multispectral and high-resolution RGB UAV imagery, with data from ground truth plots. This combined approach was necessary, as a single spatial layer often did not provide sufficient information for the definitive delineation of individual trees. The delineation focused specifically on pixels containing tree crowns. This ensured pixels belonged to tree crowns and not background objects and textures as illustrated in Figure 3.
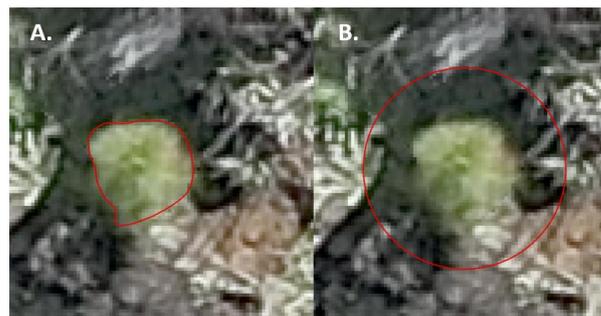
Figure 3. Example of good tree crown delineation, omitting background pixels (a), and poor tree crown delineation generated by adding a buffer to the tree detection location, including background pixels (b).

For each multispectral dataset two false colour composite datasets were analysed by Indufor Asia Pacific Ltd, Auckland, who provide a commercial service for seedling detection utilising deep learning techniques to analyse aerial imagery. Additionally, high resolution RGB imagery was also analysed to provide a benchmark for detection rates. The multispectral composite images consisted of the near infrared (NIR), blue and green bands (MS1), and the NIR, green and costal blue bands (MS2; Table 2). The high-resolution composite image (HR) consisted of bands: red, green and blue.

All three detection layers were imported into a GIS and aligned with ground truth data. In cases where the detection layers provided multiple detection boxes for the same tree, a unique tree ID was assigned to avoid over-classification. This process involved matching each delineated seedling crown from the reference imagery to its corresponding polygon within the detection layers. The assignment of the unique tree ID to individual polygons within the different detection layers established a direct linkage to the ground truth data.

By linking each delineated seedling to its respective polygon in the detection layers a number of classification metrics were produced. Cross validation identified true positives, false positives and false negatives, which were then used to calculate classification metrics; accuracy, precision and recall (see section 3.8). Additionally, detection rates relative to ground truth data were obtained through recording

the number of detected boxes present in a given plot boundary.

3.6. Spectral Vegetation Indices

The delineated tree crown polygons were used to derive the mean, median, minimum (min), maximum (max), minority, count, unique, majority, standard deviation (std) and sum values for the pixels found within each tree crown per band or index raster. The 'gone' class was not included in the health analysis, as this class represented a missing tree. Additionally, classes of 'dead' and 'vigorous' were excluded from subsequent analysis due to limited observations resulting in model imbalance. 140 statistical variables, derived from various bands and indices, were categorised into seven groups to prevent model over-fitting. These groups and their contents are outlined in Table 4.

Several standard vegetation and soil indices were calculated. All indices were calculated utilising the red, green, blue, red edge and NIR bands (Table 3). The standard *formulae* used for these calculations are shown in Table 5.

Table 4. Variable groups and their associated measures, forming input for the machine ML models.

Variable Group	Measures Included
Mean	Mean values of all 14 bands and indices
Median	Median values of all 14 bands and indices
Max	Maximum values of all 14 bands and indices
NDRE	Statistical measures derived from the NDRE index (mean, median, min, max, minority, majority, std, and sum)
NDVI	Statistical measures derived from the NDVI index (mean, median, min, max, minority, majority, std, and sum)
Red	Statistical measures derived from the Red band (mean, median, min, max, minority, majority, std, and sum)
SR	Statistical measures derived from the SR index (mean, median, min, max, minority, majority, std, and sum)

Table 5. Vegetation Indices used for health assessment.

Full name	Formula	Reference
Normalised difference vegetation index	$NDVI = \dfrac{NIR - Red}{NIR + Red}$	(Tucker, 1979)
Normalised difference red edge	$NDRE = \dfrac{NIR - Red\ Edge}{NIR + Red\ Edge}$	(Gitelson, et al., 1996)
Enhanced vegetation index	$EVI = 2.5 \times \dfrac{NIR - Red}{NIR + 6 \times Red - 7.5 \times Blue + 1}$	(Huete, et al., 2002)
Green normalised difference vegetation index	$GNDVI = \dfrac{NIR - Green}{NIR + Green}$	(Pettorelli, 2013)
Soil adjusted vegetation index	$SAVI = (1 + L) \times \dfrac{NIR - Red}{NIR + Red}$ $L = 0.5$	(Huete, 1988)
Simple ratio	$SR = \dfrac{NIR}{Red}$	(Chen, 1996)
Red-green ratio index	$RGRI = \dfrac{Red}{Green}$	(Peñuelas, et al., 1995)
Modified normalised difference vegetation index	$MNDVI = \dfrac{2 \cdot NIR + 1 - \sqrt{(2 \cdot NIR + 1)^2 - 8 \cdot (NIR - Red)}}{2}$	(Qi, et al., 1994)

3.7. Machine learning for health assessment

For the health analysis, only the data from the Pipiwai site was included in the health assessment due to a large class imbalance in the Kinleith data (Table 1). The dataset of 140 variables from the 114 observations from Pipiwai were stratified by class prevalence into training and testing sets at a ratio of 90:10. A random forest classifier (Liaw, et al., 2002) was trained with a grid search to find optimal values for model hyperparameters 'number of trees' and 'mtry', or the number of features to consider at each split when growing the trees. The best model was evaluated on the test set to compute the performance metrics per class and overall. This whole process was repeated for 10 unique iterations (folds) to reduce the variance in the test set predictions to account for the small dataset. Results were average across the 10 iterations and tabulated.

3.8. Statistical Methods

Simply taking the sum of detected seedlings and dividing it by the sum of the seedlings present in the ground truth plots is not an accurate way of assessing detection accuracy. This is because these amounts could be equal if some other objects were erroneously detected as seedlings (false positives), and actual seedlings were not detected (false negatives). For this reason, accuracy (I), precision (II), and recall (III) are commonly used to for detection accuracy. Accuracy quantifies the percentage of the total seedling detections that were correct. Precision is the proportion of the seedlings that were detected that were actually seedlings. Recall is the proportion of the total number of seedlings in the ground truth data set that were correctly identified by seedling detections. These metrics were calculated as follows:

$$\text{I.} \quad Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{II.} \quad Precision = \frac{TP}{TP + FP}$$

$$\text{III.} \quad Recall = \frac{TP}{TP + FN}$$

where:

TP (True positives): Are the cases where the detection model correctly identified healthy trees (a positive case)

FP (False positives): Are cases where the model incorrectly identified a tree (a positive case) when there was no tree present. This might represent situations where the model wrongly detected a tree that does not exist.

FN (False Negatives): Are cases where the model failed to identify a tree (a positive case) when a tree actually existed in the ground truth data. I.e., the model missed detecting a tree.

TN (True Negatives; not applicable): Are the cases where the model correctly identified the absence of healthy trees. I.e., the model correctly identified that there was no tree (a negative case) when indeed there was no tree. Therefore, these were not included in the analysis.

In evaluating the health assessment of trees, the additional statistical measures of Cohen's Kappa (Kappa), was calculated through standard confusion matrix calculations. Kappa is a statistical measure that assesses the agreement between the model's classifications and the ground truth data while taking into account the agreement that might occur by chance. The Kappa statistic was calculated as follows:

$$Kappa\ (Cohen's\ Kappa) = \frac{P_o - P_e}{1 - P_e}$$

Where:

P_o represents the observed agreement between the model and the ground truth data and P_e represents the expected agreement by chance.

4. Results

4.1. Detection Performance

When comparing the detection accuracy between the three data types (MS1, MS2 and RGB), the overall results for the RGB were the strongest at both sites (Table 6). There was no significant difference between the two multispectral datasets at Pipiwai, however, MS1 performed better than MS2 at Kinleith. MS2 also produced a significantly larger number of FPs (772) than MS1 (595) at Kinleith. Both MS dataset produced significantly more FPs at both sites than the RGB data (Table 6).

Table 6. number of TP, FP and FN observed across all plots at each site, for the three data types, along with levels of accuracy, precision and recall, coloured in 0.2 point increments.

Site	Data	TP	FP	FN	Accuracy	Precision	Recall
Kinleith	MS1	231	595	39	0.27	0.28	0.86
	MS2	243	772	29	0.23	0.24	0.89
	RGB	215	28	51	0.73	0.88	0.81
Pipiwai	MS1	102	489	48	0.16	0.17	0.68
	MS2	105	482	46	0.17	0.18	0.70
	RGB	57	19	90	0.34	0.75	0.39

The results were notably site specific for all data streams. The best results, from the RGB data were notably stronger at Kinleith (0.73, 0.88 and 0.81 for accuracy, precision and recall, respectively), than at Pipiwai (0.34, 0.75 and 0.39, for accuracy, precision and recall, respectively).

When a deep learning algorithm detects an object, it assigns a value to each detection conferring the level of confidence the model had in making that prediction. These detection confidence levels were utilised to attempt to increase the accuracy of the results.

There was a significant increase in the precision and accuracy of both multispectral models at Kinleith when

Forest Growers Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: forestgrowersresearch@fgr.nz Web: www.fgr.nz

detections were filtered from the datasets based on confidence level (Figure 4). The best results were observed with the MS1 data set, where levels of precision and accuracy for MS1 increased from 0.27 to 0.41 and from 0.28 to 0.44, respectively at Kinleith when filtering for only detections with >99% confidence (Figure 4). Improvements to accuracy and precision for multispectral data were not as significant at Pipiwai (Figure 4). Notably, accuracy and recall dropped for the RGB data at each site with increased filtering of detections based on confidence.
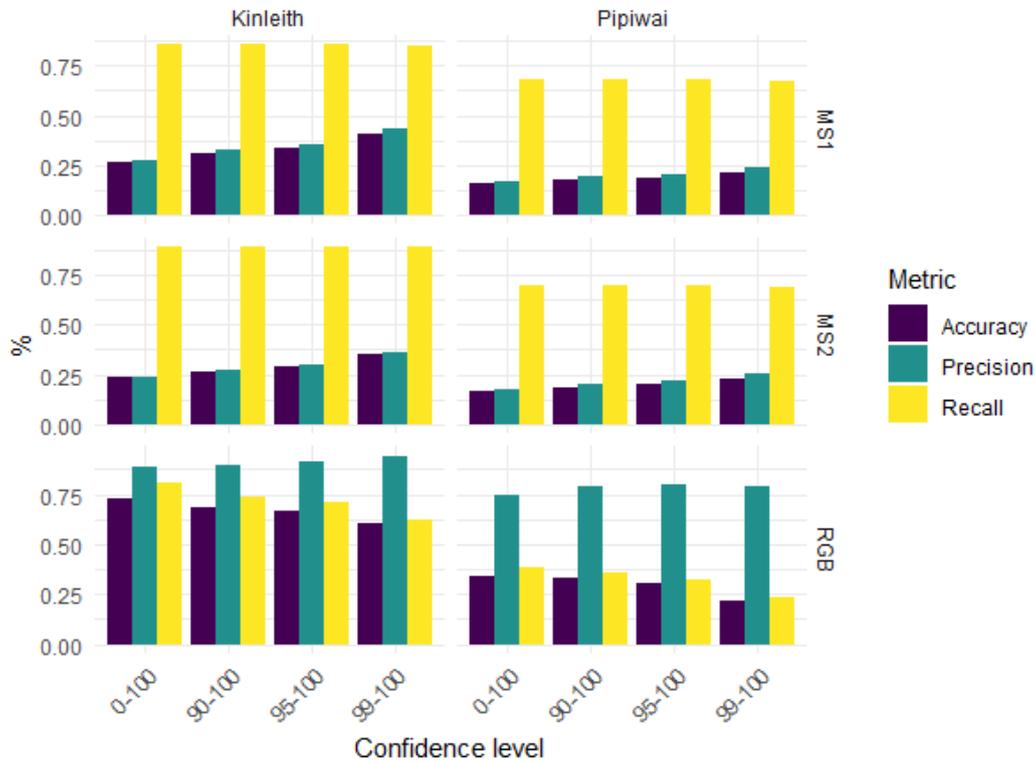
Figure 4. Impact of refining the number of detections on accuracy and precision. Facets are split into three rows for data types, and two columns for site. Bars are coloured according to the metric they represent.
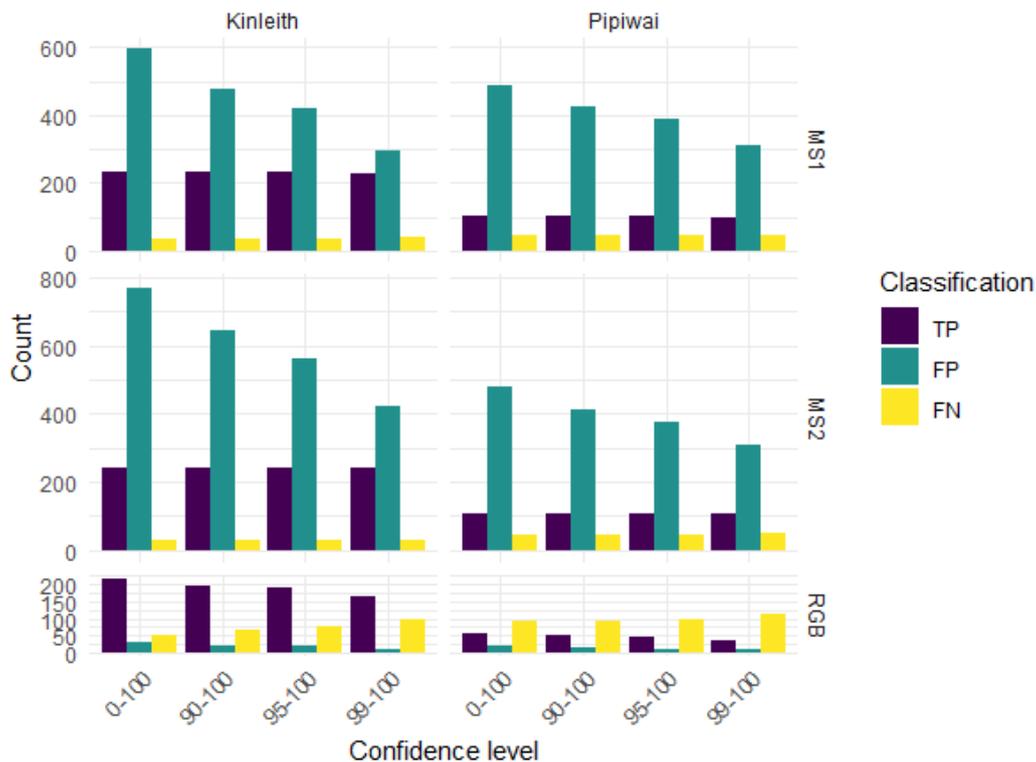
Figure 5. Impact of refining the number of detections on the numbers of TP, FP, and FN. Facets are split into three rows for data types, and two columns for site. Bars are coloured according to the metric they represent.

Forest Growers Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: forestgrowersresearch@fgr.nz Web: www.fgr.nz

The number of FP detections for both multispectral data sets was significantly reduced at each site through increased filtering of detection confidence levels (Figure 5). The largest impact was demonstrated by filtering the MS2 detections at Kinleith to include only detections with >99% confidence (FP count reduced by 465). It is notable, however, that there was a reduction in the number of TP observed in the RGB dataset for each site with an increase in confidence filtering (Figure 5).

4.2. Health assessment

The results of the proof of concept with machine learning indicate that it is possible to assess the health status of seedlings utilising multispectral imagery with a high level of accuracy, precision and recall (Table 8): The top model produced by the RF classifier included only SR variables (accuracy = 0.86, Kappa = 0.61),

and the best model including all indices used the Max statistical values (accuracy = 0.85, Kappa = 0.54; Table 7). A model fit using all 140 variables gave similar results (accuracy = 0.86, Kappa = 0.56; Table 7).

Table 7. Average accuracy and Kappa values of the top three RF models for predicting health, ranked by Kappa score.

Model composition	Average Accuracy	Average Kappa
SR Index (all statistics mean, max, majority, count etc)	0.86	0.61
All 140 variables (features)	0.86	0.56
Max (statistic)	0.85	0.54

Table 8. Precision and recall values for the top three models for predicting alive and dying seedling health classes, ranked by precision scores for 'alive' class.

Model composition	Class	Precision	Recall
SR Index (all statistics mean, max, majority, count etc):	Alive	0.94	0.89
	Dying	0.71	0.77
All 140 variables (features)	Alive	0.91	0.92
	Dying	0.74	0.63
Max (statistic)	Alive	0.91	0.91
	Dying	0.77	0.62

When looking at individual classes, the models were able to predict 'alive' seedlings with a high level of precision (0.91-0.94) and recall (0.89-0.92; Table 8). The difference between the SR model and the model using all variables was not significant, with a gain in precision and a loss in recall for 'alive' seedlings of 0.03 and 0.03 respectively, and a loss in precision and a gain in recall for 'dying' seedlings of 0.03 and 0.14 (Table 8). All models demonstrated higher confidence in predicting the 'alive' seedling class.

5. Discussion

5.1. Detection Performance

The analysis of detection performance revealed that the multispectral datasets allowed the detection of more TPs than the high resolution RGB dataset (Table 6). However, the multispectral data delivered significantly poorer performance overall. This appears to be caused by the inflated number of FPs for the multispectral data (Table 6). Utilising different combinations of near infrared wavelength bands had a minimal impact on the resulting detections, with slightly improved results being noted when using the NIR and green bands with blue (MS1) as opposed to coastal blue (MS2). It was notable that the MS2 dataset performed worse than the MS1 data at Kinleith, where it resulted in a higher number of false positives. This perhaps infers that the coastal blue

band is less sensitive to the difference between the spectral signature of radiata pine seedlings and the signatures of other vegetation. Future research should assess whether using alternative NIR or red edge wavelengths from this camera would improve the detections.

Detections were found to be highly influenced by site, with a significant decrease in precision, accuracy and recall observed for all datasets at each site. As observed previously, there were significantly increased rates of FPs at Kinleith, however, this site was characterised as having lower overall weed cover. Deep learning models are noted for being highly site specific and it is probable that the model had not been exposed to sites of this nature before.

The high recall, and lower accuracy and precision indicate that the high levels of FPs were having the greatest impact on the data. To address this, the confidence level for each detection was filtered. This had a positive impact on the multispectral data, reducing the number of FPs without altering the number of TPs and FNs. This had the opposite effect on the RGB data, where two- to three-fold reductions in FPs at Pipiwai and Kinleith respectively, coincided with reductions of ~25% and ~50% in TPs at Kinleith and Pipiwai respectively.

FPs, or erroneous detections, can be improved through an alternate process of "hard negative mining", in which the model is retrained with examples of FP detections. This trains the model to recognise the difference between TP and FPs. Alternately, to increase the number of TPs, additional training data can be added for similar sites and data types to improve overall accuracy. As the detections in this project were made using a commercial service, these improvements could not be made and would need to be discussed with the service provider in an operational scenario.

The results imply that, although multispectral data is capable of detecting more seedlings than RGB data, the number of erroneous detections that it makes highly hinder the utility of this technology. Further research is required to assess methods of refining this type of data to derive more accurate results. Multispectral cameras are also more sensitive to lighting conditions than RGB cameras, requiring spectral calibration and more careful planning of weather conditions. The processing and analysis of multispectral data can also be more complex due to the additional wavelengths and spatial layers that they produce. To truly harness the benefits of these cameras, deriving health status alongside stocking will be essential.

5.2. Health assessment

Within this project, an attempt was made to classify health utilising machine learning. The random forest classifier (RF) provided results that were encouraging, demonstrating that seedling health status could be classified with a strong level of accuracy (0.85-0.86) and moderate level of agreement between the model's predictions and the ground truth health assessments (Kappa = 0.54-0.61).

Although two of the four health classes had to be discounted from the analysis owing to class imbalance, individual health classes were able to be discerned by the model with high levels of accuracy, precision and recall (Table 8). The three top models all performed best at predicting the 'alive' health class, with high levels of precision and recall (0.91-0.94 and 0.89-0.92 respectively; Table 8). This is likely due to the class imbalance, in which there were more than threefold the number of seedlings in the alive class than the dying class (Table 1). Future research should ensure that there is a more even distribution of classes within the sample or collect a large enough dataset to ensure there are enough samples per class.

It was notable that the simple ratio (SR) index performed better, compared to models composed of multiple bands and indices. This result indicates that a relatively affordable multispectral camera solution, such as the DJI Mavic 3 Multispectral UAV (DJI, Shenzhen, China) that has a less complex sensor with fewer spectral bands, could be utilised to similar effect. It also underscores the value of employing multispectral sensors capable of capturing the near infrared (NIR) wavelength.

The machine learning process, however, is relatively complex and requires a sufficiently large and balanced dataset to achieve good results. Due to the limited size and imbalanced distribution of health classes within the dataset, along with varying levels of experience in the field crews and time constraints, little more than a proof of concept could be achieved within this study. The results were, however, strong enough to warrant further investigation, and so future research should be focused on exploring the use of machine learning techniques for health assessments in survival surveys using multispectral imagery. The methodology assessed is included in an appendix so that future studies can see what has been trialled to date.

Data from such health assessments could provide valuable insights for forest management practices. The effectiveness of classification methods in determining seedling health status could significantly influence management decisions. Future research should focus on determining optimal indicators for seedling health assessment, including top performing wavelengths or indices for assessing health.

5.1. Implications and Future Directions

Moving forward, several avenues for further research and improvements on the assessed methodology could be explored to enhance the accuracy and reliability of seedling health assessment methodologies.

Increasing dataset size: The drawbacks of the assessed dataset could be improved upon through the expansion of the sample size. A larger dataset could provide a more representative sample across health classes and mitigate the impact of smaller sample sizes on model performance. A larger dataset would increase the effectiveness of the classification model, potentially enhancing accuracy and precision. Furthermore, the introduction of more data can allow for the exploration of additional variables without the risk of overfitting, which may further enhance the discrimination power of the models.

Different modelling techniques: A number of machine learning techniques exist, including RF, support vector machines (Vapnik, 1995), gradient boosting (Friedman, 2001) and partial least squares regression analysis (Wold, 1982; Wold, 1987). Future study could explore different models to see whether any of these would perform better at assessing seedling health status.

Validation and generalisation: It is crucial to validate the performance of the models in different geographical regions and under varying environmental conditions. Generalising the findings to broader site conditions will ensure that the developed methodologies are robust and applicable to a wide range of forest management scenarios. Collaborations

with foresters and forestry companies could aid in the validation process and help develop robust methodologies for a range of conditions.

Exploration of applications: This initial proof of concept on small seedlings highlights the potential of multispectral imagery to various other forest management scenarios, such as disease classification (Camarretta, et al., 2024).

6. Conclusion

The analysis of detection performance revealed that multispectral data enabled the detection of significantly more seedlings than high resolution RGB data. This, however, coincided with a markedly higher number of erroneous detections, making overall performance weaker than RGB data. Attempts to refine the detections using the detection confidence level notably reduced the number of erroneous detections and improved the detection results. These improvements, however, were not big enough to match the performance of the RGB data.

Proof of concept work using machine learning methodologies demonstrated promise for utilising multispectral imagery for health assessments. With further research, multispectral sensors could prove to be useful tools for conducting stocking and health assessments during survival surveys, having valuable implications for automation of forest management practices.

The capacity to capture a wider range of spectral information, including NIR, through advanced sensors emphasises the advantages of employing such technology. With the evolving landscape of data collection and the promise of larger datasets and greater resolution, higher levels of accuracy and reliability could be possible. Multispectral UAV imagery has potential to be a valuable tool for forest managers, offering more informed and efficient decision-making in the early stages of plantation forestry establishment.

7. Acknowledgements

8. References

Breiman, L. (2001). Random forests. *Machine learning, 45*, 5-32.

Camarretta, N., Pearse, G. D., Steer, B. S., McLay, E., Fraser, S., & Watt, M. S. (2024). Automatic Detection of Phytophthora pluvialis Outbreaks in Radiata Pine Plantations Using Multi-Scene, Multi-Temporal Satellite Imagery. *Remote Sensing, 16*(2), 338.

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.

De Gouw, S., Morgenroth, J., & Xu, C. (2020). An updated survey on the use of geospatial technologies in New Zealand's plantation forestry sector. *New Zealand Journal of Forestry Science, 50*.

Feduck, C., McDermid, G. J., & Castilla, G. (2018). Detection of coniferous seedlings in UAV imagery. *Forests, 9*(7), 432.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Gitelson, A. A., & Merzlyak, M. N. (1996). Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. *Journal of plant physiology, 148*(3-4), 494-500.

Grossnickle, S. C. (2012). Why seedlings survive: influence of plant attributes. *New Forests, 43*(5), 711-738.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment, 83*(1-2), 195-213.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote sensing of environment, 25*(3), 295-309.

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

Montagnini, F., González, E., Porras, C., & Rheingans, R. (1995). Mixed and pure forest plantations in the humid neotropics: a comparison of early growth, pest damage and establishment costs. *The Commonwealth Forestry Review*, 306-314.

Neumann, R. W., & Landis, T. D. (1995). Benefits and techniques for evaluating outplanting success. *National proceedings: forest and conservation nursery associations—1995. Coordinated by Landis, TD, and Cregg, B. USDA Forest Service General Technical Report PNW-GTR-365*, 36-43.

Pearse, G. D., Tan, A. Y., Watt, M. S., Franz, M. O., & Dash, J. P. (2020). Detecting and mapping tree seedlings in UAV imagery using convolutional neural networks and field-verified data. *ISPRS Journal of Photogrammetry and Remote Sensing, 168*, 156-169.

Peñuelas, J., Filella, I., & Gamon, J. A. (1995). Assessment of photosynthetic radiation-use efficiency with spectral reflectance. *New Phytologist, 131*(3), 291-296.

Pettorelli, N. (2013). *The normalized difference vegetation index*. Oxford University Press, USA.

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.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment, 8*(2), 127-150.

Vapnik, V. (1995). The nature of statistical learning theory Springer. *New York, 10*, 978-971.

Wold, H. (1982). Soft modelling: the basic design and some extensions. *Systems under indirect observation, Part II*, 36-37.

Wold, S. (1987). *PLS modeling with latent variables in two or more dimensions*. Verlag nicht ermittelbar.