

Deep Learning with Northern Australian Savanna Tree Species: A Novel Dataset

Andrew J. Jansen ^{1,*} , Jaylen D. Nicholson ¹ , Andrew Esparon ¹, Timothy Whiteside ¹ , Michael Welch ¹, Matthew Tunstill ¹, Harinandanan Paramjyothi ¹, Varma Gadhiraaju ², Steve van Bodegraven ² and Renee E. Bartolo ¹ 

¹ Department of Climate Change, Energy, Environment and Water, Environmental Research Institute of the Supervising Scientist, Darwin, NT 0820, Australia

² Microsoft, Sydney, NSW 2000, Australia

* Correspondence: andrew.jansen@environment.gov.au

Abstract: The classification of savanna woodland tree species from high-resolution Remotely Piloted Aircraft Systems (RPAS) imagery is a complex and challenging task. Difficulties for both traditional remote sensing algorithms and human observers arise due to low interspecies variability (species difficult to discriminate because they are morphologically similar) and high intraspecies variability (individuals of the same species varying to the extent that they can be misclassified), and the loss of some taxonomic features commonly used for identification when observing trees from above. Deep neural networks are increasingly being used to overcome challenges in image recognition tasks. However, supervised deep learning algorithms require high-quality annotated and labelled training data that must be verified by subject matter experts. While training datasets for trees have been generated and made publicly available, they are mostly acquired in the Northern Hemisphere and lack species-level information. We present a training dataset of tropical Northern Australia savanna woodland tree species that was generated using RPAS and on-ground surveys to confirm species labels. RPAS-derived imagery was annotated, resulting in 2547 polygons representing 36 tree species. A baseline dataset was produced consisting of: (i) seven orthomosaics that were used for in-field labelling; (ii) a tiled dataset at 1024 × 1024 pixel size in Common Objects in Context (COCO) format that can be used for deep learning model training; (iii) and the annotations.

Dataset: <https://doi.org/10.5281/zenodo.7094916>

Dataset License: Creative Commons Attribution 4.0 International (CC BY 4.0)

Keywords: RPAS; drones; tropical savanna; savanna woodlands; Kakadu National Park; artificial intelligence



Citation: Jansen, A.J.; Nicholson, J.D.; Esparon, A.; Whiteside, T.; Welch, M.; Tunstill, M.; Paramjyothi, H.; Gadhiraaju, V.; van Bodegraven, S.; Bartolo, R.E. Deep Learning with Northern Australian Savanna Tree Species: A Novel Dataset. *Data* **2023**, *8*, 44. <https://doi.org/10.3390/data8020044>

Academic Editors: Emma Tonkin and Kristina Yordanova

Received: 5 October 2022

Revised: 25 January 2023

Accepted: 16 February 2023

Published: 20 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Summary

The savanna woodlands of Northern Australia are highly complex systems. They are characterised by a discontinuous overstory tree composition and a continuous grassy understory [1]. They have high levels of intraspecies variability as well as low levels of interspecies variability in some cases, and are morphologically heterogenous in branching structure, crown shape and height [2]. There has been increasing effort to quantify woody cover within savanna ecosystems [3–5]. Using aerial imagery collected from satellites, light aircraft and Remotely Pilot Aircraft Systems (RPAS) or drones, woody cover measurements can be performed at the site or landscape scale. Measurements of individual trees in savanna ecosystems at the landscape scale are pertinent for quantifying reference ecosystems to assess ecosystem restoration efforts at the same scale [6] and developing standards for mine site ecosystem restoration in those environments [7]. Whilst considerable progress

has been made in determining the woody cover from aerial imagery [7], there has been limited success in assigning species-level classification to trees within savanna ecosystems. Success has mostly been achieved using multi- and hyperspectral sensors deployed often in combination with LiDAR data [8–10] which adds layers of data-processing complexity when developing automated classification systems. Due to the significant inter- and intraspecific morphological variation in some tree species when observed from above, traditional remote sensing algorithms such as support vector machine and regression trees are suboptimal when attempting to generalize classification for multiple species at the landscape scale.

Deep learning algorithms such as Convolution Neural Networks (CNNs) can detect patterns in nonlinear data and, increasingly, are being applied in the field of computer vision [11]. Supervised deep learning requires the labour-intensive generation of annotated and labelled training datasets by domain-experts that are used to train a CNN. Whilst many large training datasets are made publicly available such as ImageNet [12] which started with 3.2 million images and Common Objects in Context (COCO) with 328,000 images [13], these datasets are often developed and optimized for detecting a wide range of general objects, not specifically tree species. There are fewer publicly available training datasets curated for trees. One includes 2848 hand-annotated trees using bounding boxes supplemented with 434,551 computer-generated annotations that correspond to coregistered RGB, LiDAR and hyperspectral aerial imagery [14]. Another training dataset comprises ~30,000 labelled trees consisting of mixed aerial and ground-level photos [15]. However, these datasets were curated mostly in Northern Hemisphere forests and scenarios where tree species are distinct and easily distinguishable, compared to Northern Australia savanna woodlands with morphologically similar tree species.

The visual classification of savanna tree species from RPAS-derived imagery is a difficult task without prior knowledge or experience in matching aerial imagery with the botanical features typically used for species identification from the ground.

We aimed to generate an annotated and labelled training dataset for common savanna tree species in Kakadu National Park, Northern Territory, Australia. The classification of trees in RPAS imagery needed to be confirmed by expert botanists to ensure individual tree crowns were correctly delineated and species were labelled for every tree. This dataset can be used to train CNNs for instance segmentation, object detection and image classification. The dataset forms a baseline that can more broadly be used for: an inventory of savanna tree species at the landscape scale; assessing ecosystem restoration efforts; savanna tree ecology research; and computer vision. As the majority of individual trees remain fixed in their position through time (an exception is change through disturbance), the tree species annotations can be georeferenced to aerial imagery across different time periods (years) and seasons, and applied to higher-resolution imagery when it becomes available without having to repeat intensive ground-truth field surveys. The savanna tree AI dataset is made publicly available; orthomosaics can be accessed at www.geonadir.com (accessed on 22 June 2021) and a tiled dataset in COCO format can be accessed at <https://doi.org/10.5281/zenodo.7094916>. We provide methods to perform model training with the dataset and present preliminary results.

2. Data Description

This dataset consists of seven orthorectified mosaics (Red, Green, and Blue) and seven shape files which include hand-annotated polygons which delineate the tree canopies in the imagery. All imagery was collected from seven 1 ha sites in Kakadu National Park, Northern Territory, Australia (Figure 1).

The variation amongst sites in savanna woodland tree community composition is described in Table 1. Overall, *Eucalyptus tetradonta*, *Eucalyptus miniata* and *Acacia mimula* were the most dominant species present amongst the seven sites surveyed and recorded the greatest number of polygons compared to other tree species.

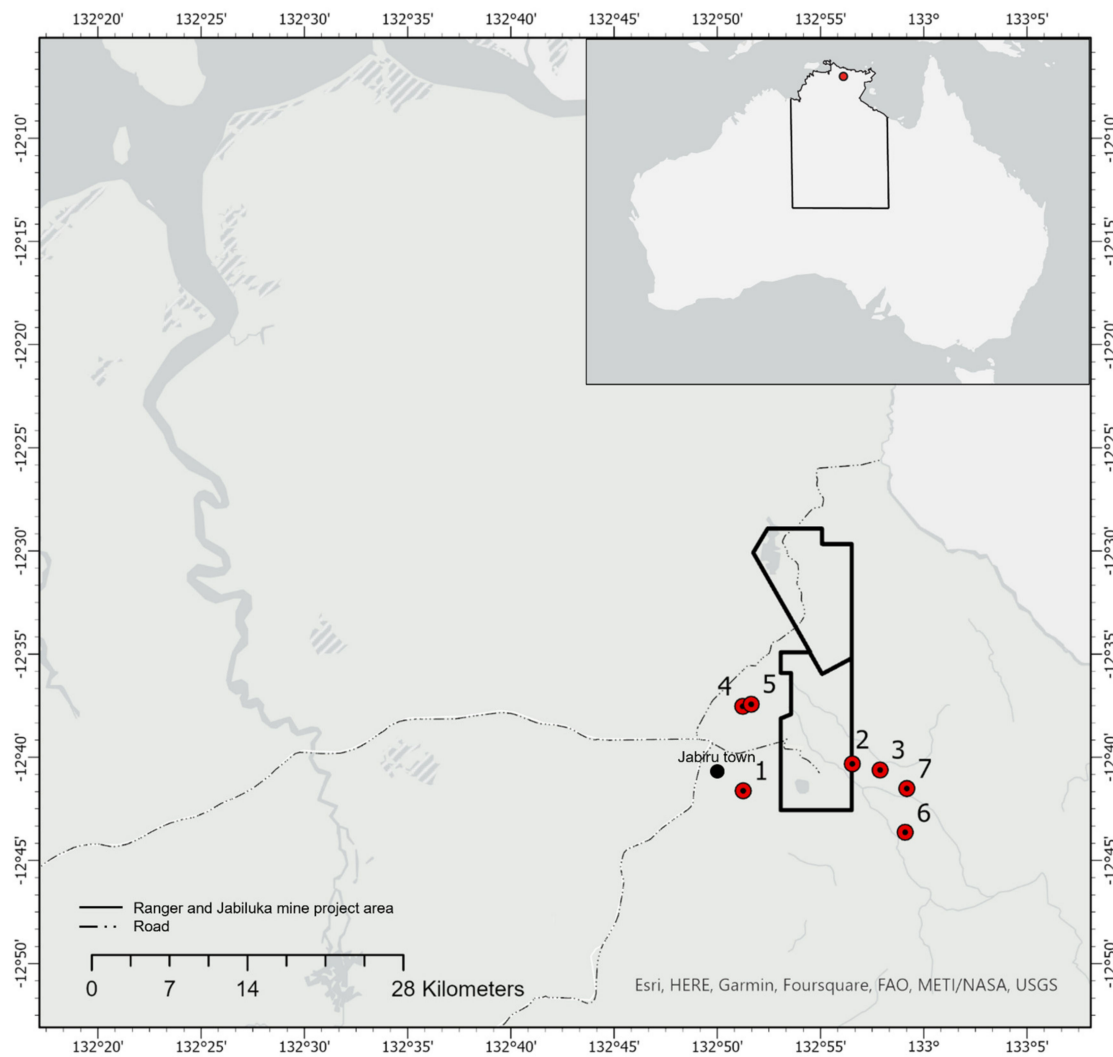


Figure 1. Location of sites for imagery collected in Kakadu National Park, Northern Territory, Australia.

Table 1. Description of plots surveyed, number of polygons, overall stem densities (trees) and top 3 dominant tree species.

Plot	Number of Polygons [Classes]	Top Three Dominant Tree Species Per Plot
Site 1	219 [13]	<i>Eucalyptus tetradonta</i> , <i>Xanthostemon paradoxus</i> , and <i>Corymbia porrecta</i>
Site 2	322 [17]	<i>Acacia mimula</i> , <i>Eucalyptus tetradonta</i> , and <i>Persoonia falcata</i>
Site 3	169 [15]	<i>Eucalyptus tetradonta</i> , <i>Eucalyptus miniata</i> , and <i>Acacia mimula</i>
Site 4	402 [18]	<i>Eucalyptus tetradonta</i> , <i>Eucalyptus miniata</i> , and <i>Erythrophleum chlorostachys</i>
Site 5	538 [20]	<i>Eucalyptus tetradonta</i> , <i>Acacia lamprocarpa</i> , and <i>Acacia mimula</i>
Site 6	365 [7]	<i>Acacia mimula</i> , <i>Eucalyptus miniata</i> , and <i>Eucalyptus tetradonta</i>
Site 7	532 [17]	<i>Eucalyptus tetradonta</i> , <i>Erythrophleum chlorostachys</i> , and <i>Corymbia porrecta</i>
Total	2547 [36]	<i>Eucalyptus tetradonta</i> , <i>Acacia mimula</i> , and <i>Eucalyptus miniata</i>

The shape file includes an attribute containing the species-level identification of trees for each polygon that was verified through on-ground vegetation surveys (Table 2). Ground-truthed species identifications were further validated by a botanist on a desktop workstation after on-ground surveys were conducted. On-ground vegetation surveys were conducted in April and June 2021 where 2547 polygons for 36 tree species were collected.

Table 2. Species distribution for all plots combined with species code used in shape file and Common Objects in Context (COCO) format. Total number of annotations (polygons) for each class is included.

Species	Code (in Shape File)	Number of Polygons
<i>Acacia dimidiata</i>	ACDIM	6
<i>Acacia lamprocarpa</i>	ACLAM	2
<i>Acacia mimula</i>	ACMIM	308
<i>Acacia oncinocarpa</i>	ACONC	8
<i>Brachychiton megaphyllus</i>	BRMEG	0
<i>Buchanania obovata</i>	BUOBO	27
<i>Calytrix exstipulata</i>	CAEXS	37
<i>Cochlospermum fraseri</i>	COFRA	34
<i>Coelospermum reticulatum</i>	CORET	3
<i>Corymbia bleeseri</i>	COBLE	2
<i>Corymbia ferruginea</i>	COFER	1
<i>Corymbia polycarpa</i>	COPOC	1
<i>Corymbia polysciada</i>	COPOS	2
<i>Corymbia porrecta</i>	COPOR	176
<i>Denhamia obscura</i>	DEOBS	3
<i>Erythrophleum chlorostachys</i>	ERCHL	261
<i>Eucalyptus miniata</i>	EUMIN	598
<i>Eucalyptus tetradonta</i>	EUTET	597
<i>Ficus aculeata</i>	FIACU	2
<i>Gardenia megasperma</i>	GAMEG	10
<i>Grevillea pteridifolia</i>	GRPTE	2
<i>Livistona humilis</i>	LIHUM	54
<i>Owenia vernicosa</i>	OWVER	7
<i>Pandanus spiralis</i>	PASPI	62
<i>Persoonia falcata</i>	PEFAL	17
<i>Petalostigma pubescens</i>	PEPUB	4
<i>Petalostigma quadriloculare</i>	PEQUA	2
<i>Planchonella arnhemica</i>	PLARN	1
<i>Planchonia careya</i>	PLCAR	14
<i>Premna acuminata</i>	PRACU	1
<i>Stenocarpus acacioides</i>	STACA	2
<i>Syzygium eucalyptoides</i> subsp. <i>bleeseri</i>	SYEUB	1
<i>Syzygium eucalyptoides</i> subsp. <i>eucalyptoides</i>	SYEUE	1
<i>Terminalia ferdinandiana</i>	TEFER	101
<i>Terminalia grandiflora</i>	TEGRA	6
<i>Xanthostemon paradoxus</i>	XAPAR	194

3. Methods

We outline the methods used to collect, preprocess, label and use the dataset to train deep neural networks for the instance segmentation, object detection and image classification of savanna woodland tree species. The architecture used for this project is outlined in Figure 2.

3.1. Plot Setup

Before aerial imagery was collected, tape was placed to mark the perimeter of each survey area, 100 m × 100 m in size. Pink, red and yellow marker cones were then positioned within the survey area in a grid-like manner, spaced ~20 m apart. The placement of tape and cones was to provide identifiable features in the mosaic that could be used by field teams to correctly identify the tree being labelled (Figure 3).

3.2. Image Capture

The aerial imagery was collected using a DJI Matrice M200 RPAS mounted with a DJI Zenmuse X5S 20.8 MP CMOS (4/3 inch) sensor (DJI Sky City, No.55 Xianyuan Road, Nanshan District, Shenzhen, China) and DJI MFT 15 mm f1.7 lens. The RPAS was operated

at approximately 80 m above ground level, resulting in a ground sample distance of almost 2 cm (Table 3). All imagery was collected between 11:00 and 15:30 to reduce the effects of shadows on the imagery. For each flight, there was an image overlap of 85% and 75%, forward and side, respectively. All imagery, except for Site 1, was acquired in mid- to late June 2021. Site 1 data were collected in April 2021.

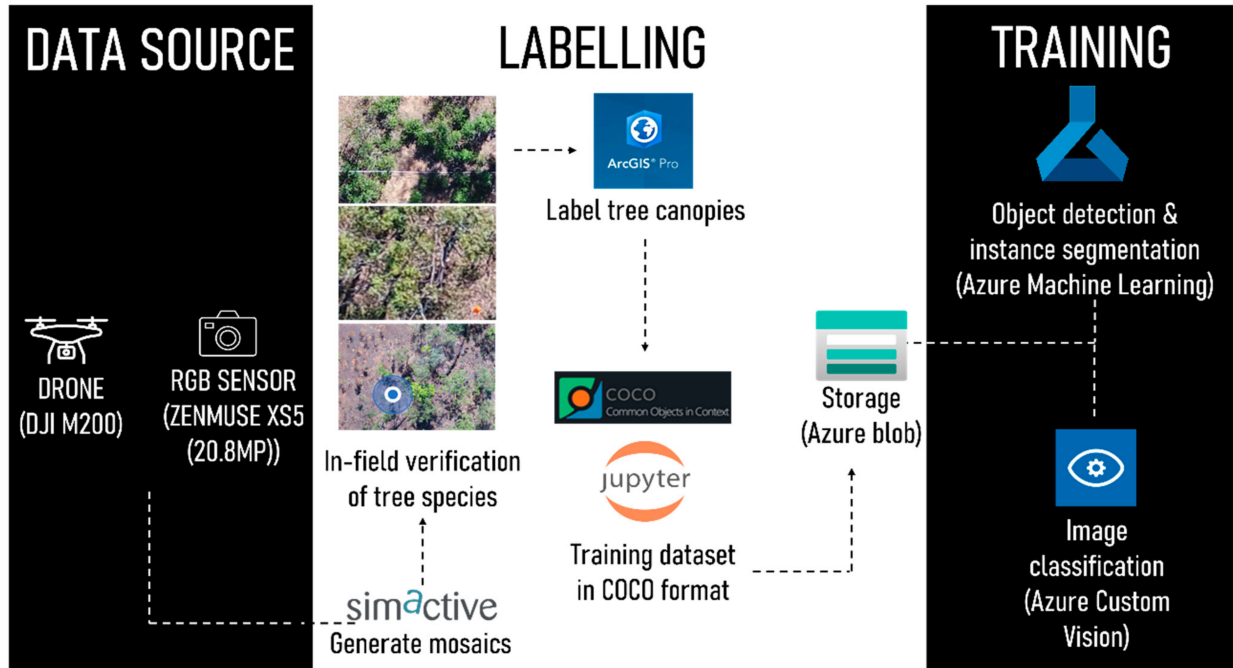


Figure 2. Data-processing pipeline for generating tree species dataset and training deep learning models.

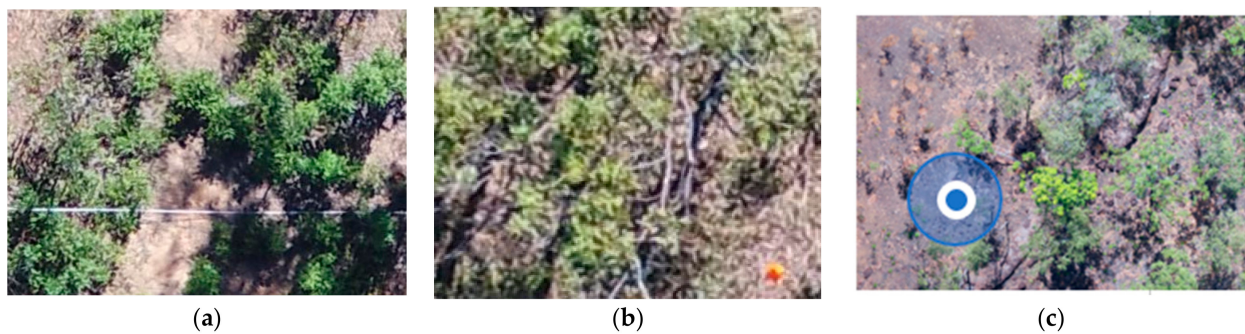


Figure 3. Examples of field referencing methods for labelling tree species: (a) tape line; (b) cones; (c) and real-time GPS.

Table 3. Drone image capture details.

Site (Mosaic)	Date and Time	Flight Height (m)	No. Images	GSD (cm)
Site 1	12 April 2021 15:10	82	233	1.8
Site 2	21 June 2021 11:30	87	260	1.9
Site 3	22 June 2021 11:20	82	203	1.8
Site 4	14 June 2021 14:20	87	112	1.9
Site 5	15 June 2021 12:30	90	342	2.0
Site 6	21 June 2021 11:35	82	270	1.8
Site 7	17 June 2021 12:30	78	198	1.7

The imagery was processed using the standard photogrammetric workflow within the Correlator 3D software application (SimActive, Montreal, QC, Canada) and geometrically corrected orthomosaics were generated.

3.3. Field-Based Image Labelling

The mosaics were visualized in ArcGIS Pro v2.5 on laptops to enable labelling in the field. Within the ArcGIS Pro tooling, a feature class was created in the project's file geodatabase. With the feature class selected, the tool to create polygons was selected within the table. Using the aerial imagery, reference markers and real-time GPS integrated with ArcGIS, field teams navigated to the trees in the field to perform species-level identification and draw a polygon around the canopy of the tree (Figure 4). Species labels were entered into a class column in the file geodatabase. To assist with navigation in the field, the GPS function was activated in ArcGIS for real-time geographical positioning. Due to the complex morphology of savanna tree branching structures, care was taken to ensure the correct boundary of the polygon matched the 'real' extent in the imagery. For multistem trees, polygons were created around the individual canopies when distinguishable in the imagery. Otherwise, one polygon was created for the cluster and it was noted that it was multistemmed. For instances where canopies of different tree species were overlapping, the field team attempted to keep polygons from overlapping and tried to distinguish the canopies as best as practicable.

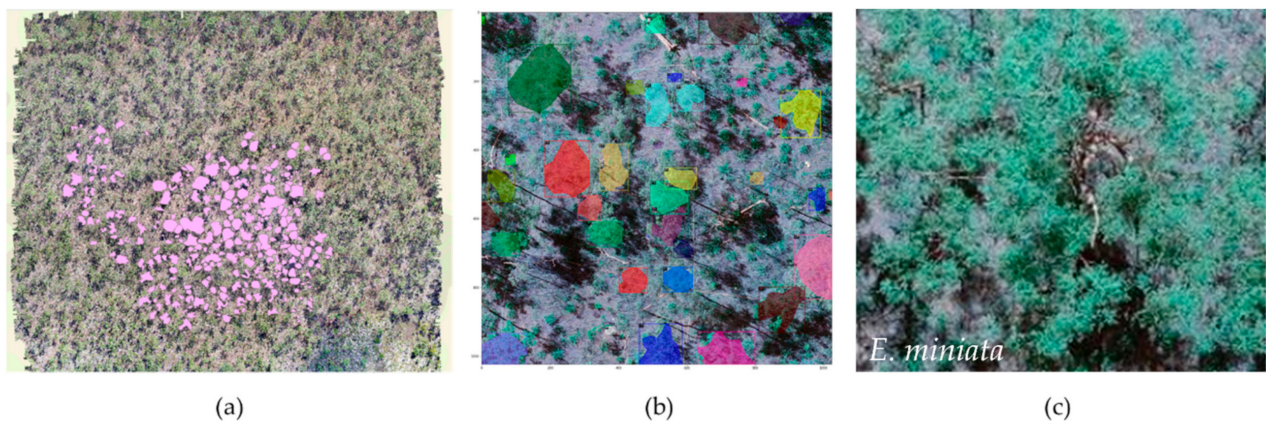


Figure 4. Example of tree annotations and labels: (a) at the scale of a plot; (b) species-level polygons; and (c) *Eucalyptus miniata* at the dimensions of the bounding box annotation.

3.4. Preparing Plot Data for Deep Learning

Deep neural networks are commonly trained on smaller image sizes such as 256×256 or 512×512 pixels to maintain computational efficiency and as a standard scale for objects of interest in an image to be learnt by a network [16,17]. As such, the mosaics which ranged in size from $16,840 \times 15,112$ to $30,170 \times 28,020$ pixels needed to be tiled for model training.

Mosaic images were tiled at 1024×1024 pixel size with a 512-pixel step size (overlap) to attempt to include the larger-sized tree canopies in at least one of the tiles. The tiling process resulted in some images without polygons bound around every tree in the image due to sparsity in the annotations generated on the edge of the survey (Figure 4a) and small shrubs (<1.5 m) not targeted during in-field labelling (Figure 5). Species annotations in the shape file that corresponded to each tile were converted to Microsoft Common Objects in Context (MS COCO) format and combined into one .json file. For object detection and instance segmentation model training the data from plots 1, 2, 3, 5, 6 and 7 were aggregated into one training dataset and plot 4 data were used as a validation dataset.

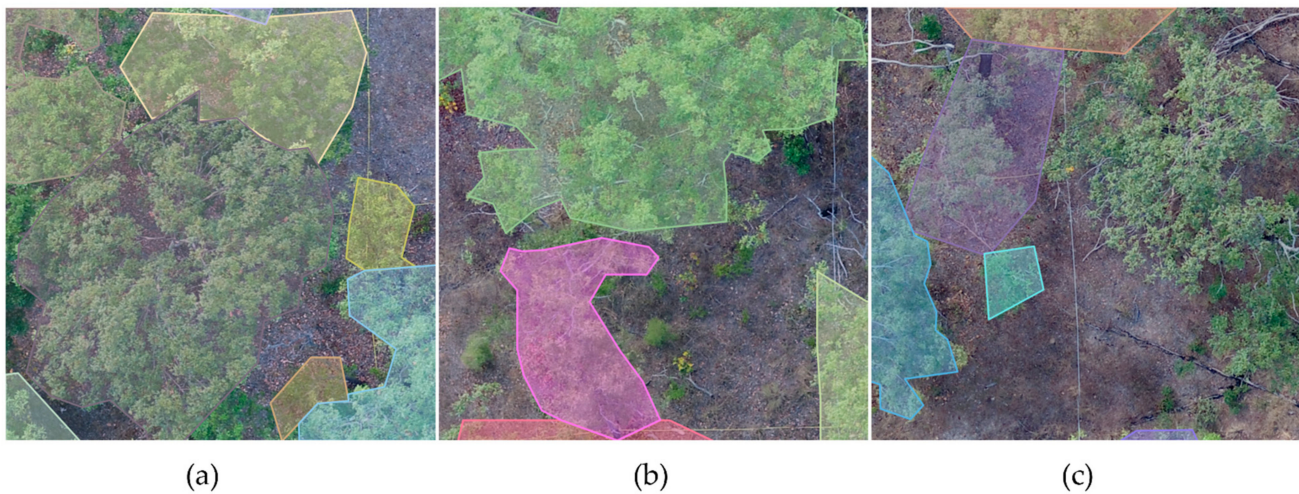


Figure 5. Examples of tiled (1024×1024) training data with: (a) all trees bound by polygons; (b) missing some shrubs; and (c) on the edge of the survey area including canopies that were not annotated.

3.5. Deep Learning Model Training and Preliminary Results

Deep CNNs were trained using Azure Machine Learning services AutoML [18] and Custom Vision platforms [19]. AutoML is an end-to-end machine learning tool that enables automated CNN model training for computer vision tasks. It supports state-of-the-art CNN models for computer vision and automated hyperparameter tuning to sweep across all the possible parameters and optimize performance. AutoML was used to train instance segmentation and object detection models.

Azure's Custom Vision platform was used to train an image classification model. Custom Vision is an AI service and end-to-end platform for labelling and training deep learning models. The platform optimizes model training and hyperparameter tuning based on the data. An image classification project with multiclass (single tag per image) tagging was created for model training.

Three evaluation metrics were reported to assess the performance of models trained: (1) precision: the probability that a detected class was the correct class; (2) recall: the probability that out of all the classes that should be predicted, the model made the correct prediction; and (3) average precision: the area under the curve of precision and recall. The mean average precision is reported as the mean of the average precision calculated for all classes in the model.

3.5.1. Instance Segmentation

An instance segmentation model was trained using 2885 images, each with 1024×1024 pixel size, on one class (tree). The model architecture used was Mask R-CNN with a ResNet50-FPN backbone [20]. A learning rate of 0.005 and training batch size of 2 were used in conjunction with the early stopping feature which ceases training when the model is no longer learning from the dataset. The model was validated against a validation dataset of 449 images consisting of unseen images from the training data. Measures of precision and recall by the model at each training epoch step are presented in Figure 6. One epoch represents a forward and backward pass of the entire dataset through the neural network. The best combination of precision, recall and mean average precision were 25.5%, 61.4% and 34.5% measured at epoch 4 and the early stopping feature was initialized at epoch 10. As model training progressed, precision increased, while recall decreased.

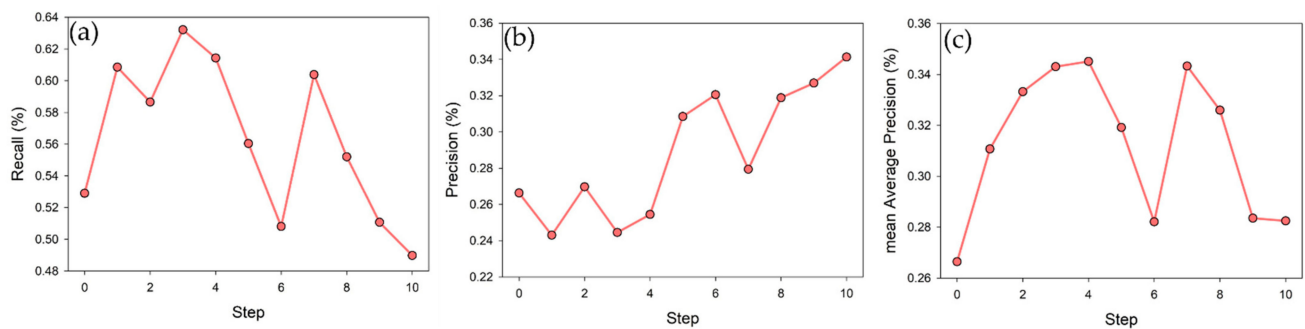


Figure 6. Calculations of: (a) recall; (b) precision; and (c) mean average precision at each epoch step during instance segmentation model training on one class (tree).

When visualizing model predictions on the test imagery, patterns were evident in the predictive performance. When the trees were sparse, smaller trees were often undetected (Figure 7a). When the tree canopies were dense, and with minimal overlap, the model was able to detect and mask most tree crowns (Figure 7b), and when there was a high degree of canopy overlap, the model attempted to combine multiple tree canopies into one (Figure 7c). Most large tree canopies were detected in the test imagery when visualizing predictions, contrasting with the low metrics of precision generated during the model training. This may be due to predictions made on the unannotated trees present in the validation dataset, resulting from the tiling process (see Section 3.4), which were recorded as False Positive (FP) predictions and lowered the calculations of precision. The calculations of recall (61.4%) were double that of precision, indicating more than half the trees in the validation dataset were detected by the model.

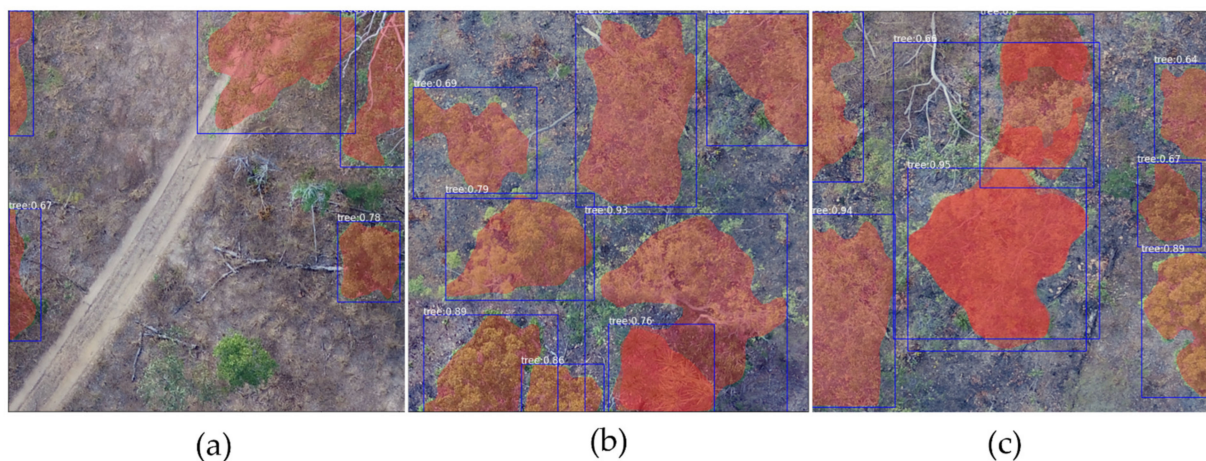


Figure 7. Instance segmentation model predictions on 1024×1024 test images demonstrating examples of: (a) sparse tree coverage; (b) dense but isolated tree canopies; and (c) connected or overlapping tree canopies.

3.5.2. Object Detection

An object detection model was trained using 2885 images, each with 1024×1024 pixel size, on one class (tree). The model architecture used was Faster R-CNN with a ResNet50-FPN backbone [20]. A learning rate of 0.005 and training batch size of 2 were used with the early stopping feature which ceases training when the model is no longer learning from the dataset. The model was validated against a validation dataset of 449 images consisting of unseen images from the training data.

Measures of precision and recall by the model at each training epoch step are presented in Figure 8. The best combination of precision, recall and mean average precision were

24.1%, 65.1% and 37.0% measured at epoch 5 and the early stopping feature was initialized at epoch 11. As the model training progressed, precision increased, while recall decreased.

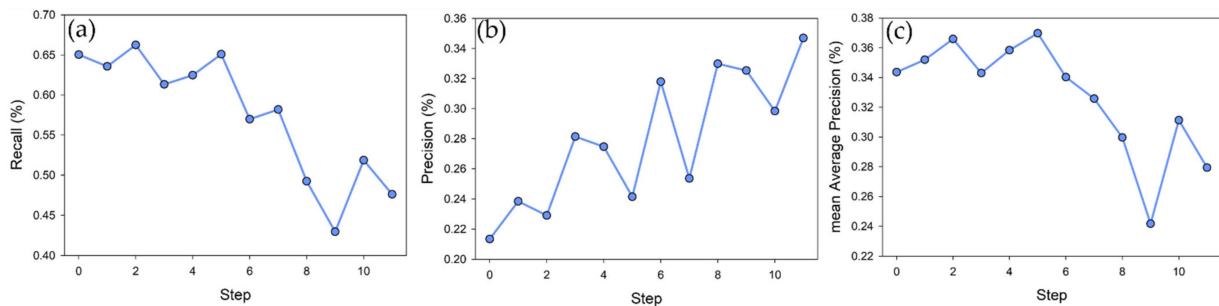


Figure 8. Calculations of: (a) recall; (b) precision; and (c) mean average precision at each epoch step during object detection model training on one class (tree).

The object detection model performed similarly to instance segmentation (see Section 3.5.1) with slightly higher recall. The visualisation of model predictions also revealed similar detection patterns with the occasional missed tree canopy when tree crowns were closer or overlapping (e.g., the tree canopy in the top of Figure 9c).

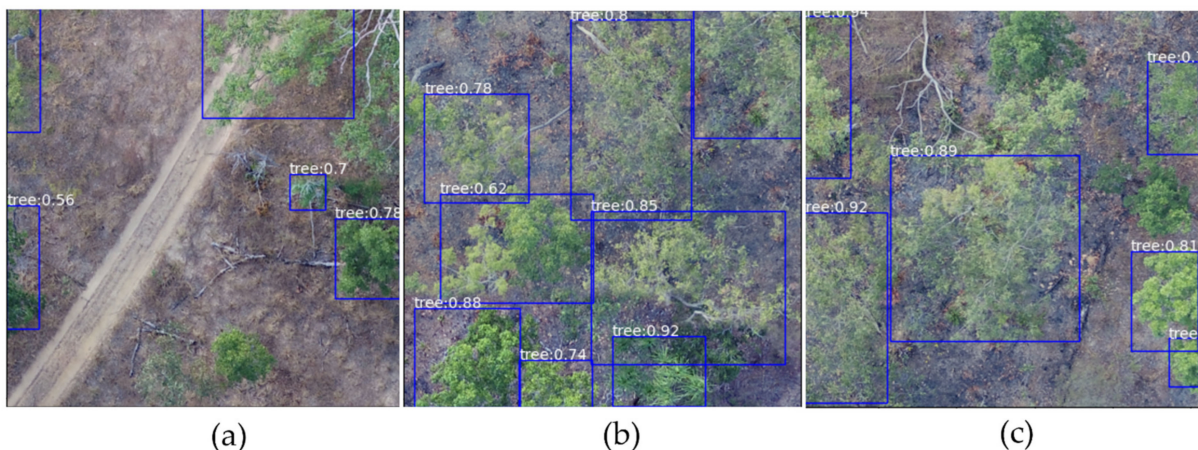


Figure 9. Object detection model predictions on 1024×1024 test images demonstrating examples of: (a) sparse tree coverage; (b) dense but isolated tree canopies; and (c) connected or overlapping tree canopies.

3.5.3. Image Classification (Multiclass)

An image classification model was trained using 2280 images with 17 classes (tree species) that were selected with sufficient training data (>5 images per species) to train CNN models in the Custom Vision platform. Training data were generated by clipping each tree at the dimension of the bounding box for each species annotation (Figure 4c). The model selection, validation dataset and hyperparameters were chosen by the platform and training was undertaken for 2 h of computing time. The mean average precision, precision and recall were 76.7%, 75.4% and 67.8%, respectively.

Measures of precision and recall were calculated at 1% confidence intervals between 0 and 100% (Figure 10). The area under the curve (AUC) or trade-off between precision and recall demonstrates a generalized model was trained with ~50% recall when constraining predictions to a high (99%) confidence threshold.

Metrics for each of the tree species are presented in Table 4. The probability threshold for accepting predictions with the test dataset was 50%, which is supported by the precision–recall curve (Figure 10) as a suitable threshold for high levels of recall or classification of tree species.

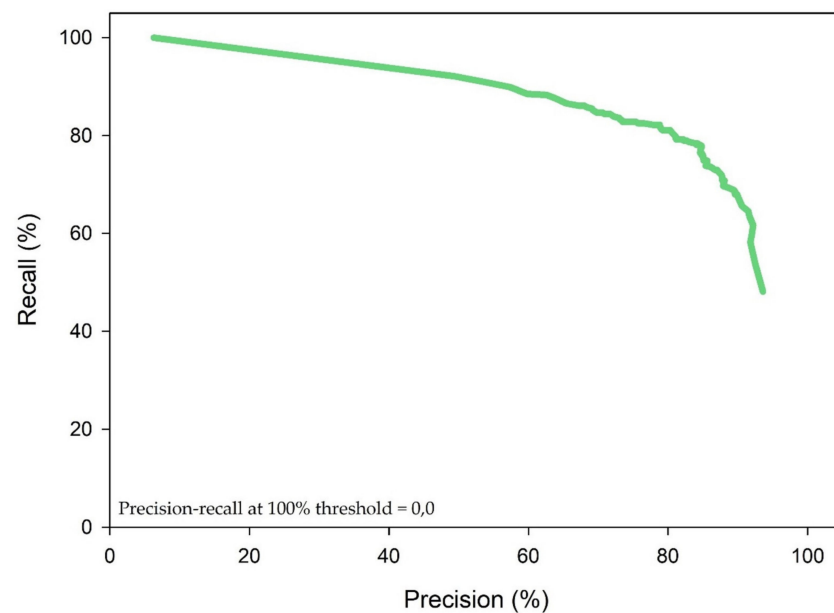


Figure 10. Calculations of precision and recall at confidence thresholds between 0 and 100% at 1% intervals.

Table 4. Species-wise average precision, precision and recall results from image classification model training in Custom Vision.

Species	No. of Images Used for Training	Average Precision (%)	Precision (%)	Recall (%)
<i>Acacia mimula</i>	304	58.6	60.6	65.6
<i>Acacia oncinocarpa</i>	7	0.4	0	0
<i>Buchanania obovata</i>	23	10.9	33.3	20
<i>Calytrix exstipulata</i>	13	100	75	100
<i>Cochlospermum fraseri</i>	9	83.3	100	50
<i>Corymbia porrecta</i>	170	36.4	42.1	23.5
<i>Erythrophleum chlorostachys</i>	246	83.9	80	73.5
<i>Eucalyptus miniata</i>	517	85	82.7	77.9
<i>Eucalyptus tetradonta</i>	558	83.4	76.2	68.8
<i>Gardenia megasperma</i>	10	2.8	0	0
<i>Livistona humilis</i>	53	91.1	90.9	100
<i>Pandanus spiralis</i>	58	100	100	100
<i>Persoonia falcata</i>	17	38	25	25
<i>Planchonia careya</i>	14	0.0	0.0	2.5
<i>Terminalia ferdinandiana</i>	95	92.2	94.1	84.2
<i>Terminalia grandiflora</i>	5	50	0	0
<i>Xanthostemon paradoxus</i>	181	83.8	78.8	72.2

The highest average precision of 100% was observed in *P. spiralis* and *C. exstipulata* with 58 and 13 images, respectively, and the lowest average precision of 0% was observed in *P. careya* with 14 images. Species that recorded an average precision equal to or less than 50% also had a low number of images used for training (<23), except for *C. porrecta* with 170 images and 36.4% average precision. Tree species with an average precision greater than 50% generally had higher numbers of images for training (>50), except for *C. exstipulata* and *C. fraseri* with 13 and 9 images, respectively. The exceptions noted above may be explained by either their comparative morphological uniqueness or similarity in the training data. For example, *C. exstipulata* was the only species with bright-purple flowers in all training data images (Figure 11). While the number of training images was low (13), the low levels of intraspecies variation and high degree of interspecies variation for *C. exstipulata* within the training data resulted in higher evaluation metrics, meaning

it was easily learnt and distinguished by the model. Similarly, training data for *C. fraseri* represented small trees with bright-green leaves that were (Figure 11) not represented in other tree species and resulted in higher evaluation metrics.

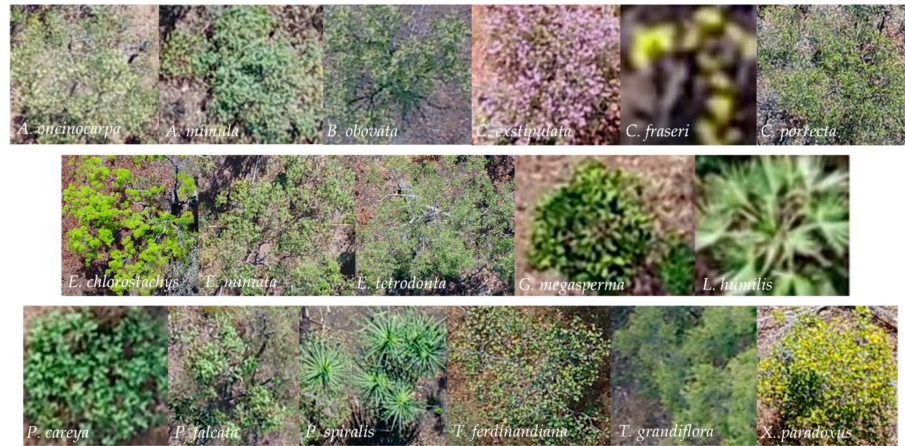


Figure 11. Examples for training data for species included in the image classification model.

C. porrecta demonstrated an opposite effect to *C. extipulata* and *C. fraseri* where there were a high number of training images (170). High degrees of intraspecies variation—due to small and large tree canopies in the training data—combined with greater interspecies similarity with taxa such as *E. tetradonta* (Figure 11) resulted in poorer model detectability. This is evident in the confusion matrix of model predictions in Figure 12 where 8 out of the 32 test images for *C. porrecta* were predicted correctly, and 13 and 7 predictions were attributed to the similar-looking species *E. tetradonta* and *E. miniata*, respectively. Visually unique tree species such as *L. humilis* and *P. spiralis* (Figure 11) had fewer misidentifications for other species (Figure 8) compared to visually similar species such as *E. tetradonta*, *A. mimula*, *C. porrecta* and *E. tetradonta*, which had higher numbers of relative misidentifications (Figure 12).

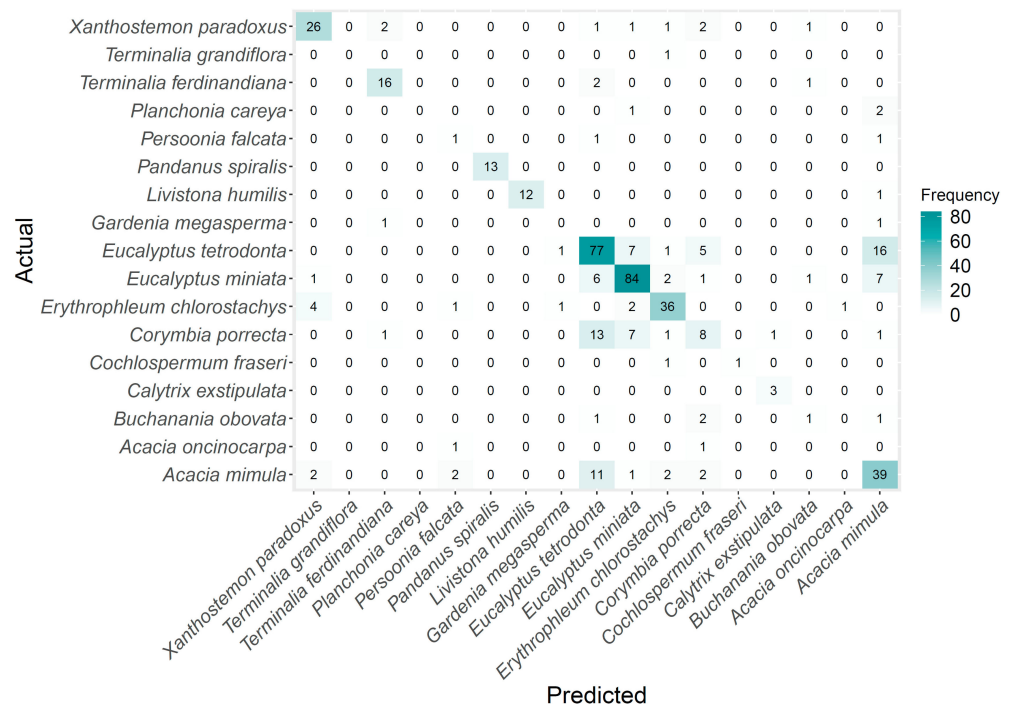


Figure 12. Confusion matrix of actual and predicted tree species from the image classification model.

4. Conclusions

We generated an annotated and labelled training dataset for common savanna tree species in Kakadu National Park, Northern Territory, Australia using aerial imagery collected from an off-the-shelf RPAS. The dataset was created by expert botanists and remote sensing experts and can be used to train convolution neural networks for instance segmentation, object detection, and image classification. The dataset is a valuable resource for an inventory of savanna tree species at the landscape scale using RPAS-derived very-high-resolution imagery. The use of deep learning algorithms, specifically CNNs, shows promising results in classifying savanna tree species, even in the presence of significant inter- and intraspecific morphological variation. The visualisation of object detection and instance segmentation model predictions indicates the delineation of tree crowns performed best when trees were isolated and further training data were required to detect both small and large tree canopies in the same image and where significant amounts of overlap in the tree canopy occurred. Image classification models confidently distinguished 9 of 17 species included in this study at an 80% average precision or higher. Future work will focus on using higher-resolution sensors to provide increased detail in addressing some of the limitations identified in the dataset. This study highlights the potential for using CNNs to classify savanna tree species and the importance of creating annotated and labelled training datasets in the development of automated classification systems.

5. User Notes

To quickly get started with this dataset, go to [ajansenn/SavannaTreeAI \(github.com\)](https://github.com/ajansenn/SavannaTreeAI) to access the resources for model training using the Azure Machine Learning service with cloud computing. Scripts to preprocess the dataset are also available.

Author Contributions: Conceptualization, A.J.J. and R.E.B.; methodology, A.J.J., R.E.B., J.D.N., H.P., M.T., S.v.B. and V.G.; software, A.J.J. and R.E.B.; validation, J.D.N., R.E.B., H.P., M.W. and A.E.; formal analysis, A.J.J.; investigation, R.E.B.; data curation, A.E., A.J.J. and J.D.N.; writing—original draft preparation, A.J.J.; writing—review and editing, T.W., R.E.B. and J.D.N.; project administration, A.J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in GeoNadir, searchable under the names Savanna Plot (1), (2)–(7). The training dataset in COCO format is openly available in Zenodo at <https://doi.org/10.5281/zenodo.7094916>.

Acknowledgments: We pay our respects to all Traditional Owners of Kakadu National Park and the Darwin region where we conduct research and monitoring, and acknowledge Elders past, present and emerging.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hutley, L.B.; Setterfield, S.A. Savanna. In *Encyclopedia of Ecology*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 623–633.
2. Naidoo, L.; Cho, M.A.; Mathieu, R.; Asner, G. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *ISPRS J. Photogramm. Remote Sens.* **2012**, *69*, 167–179. [[CrossRef](#)]
3. Anchang, J.Y.; Prihodko, L.; Ji, W.; Kumar, S.S.; Ross, C.W.; Yu, Q.; Lind, B.; Sarr, M.A.; Diouf, A.A.; Hanan, N.P. Toward operational mapping of woody canopy cover in tropical savannas using Google Earth Engine. *Front. Environ. Sci.* **2020**, *8*, 4. [[CrossRef](#)]
4. Humphrey, G.; Eastment, C.; Gillson, L.; Timm Hoffman, M. Woody cover change in relation to fire history and land-use in the savanna-woodlands of north-east Namibia (1996–2019). *Afr. J. Range Forage Sci.* **2022**, *39*, 96–106. [[CrossRef](#)]

5. Kolarik, N.E.; Gaughan, A.E.; Stevens, F.R.; Pricope, N.G.; Woodward, K.; Cassidy, L.; Salerno, J.; Hartter, J. A multi-plot assessment of vegetation structure using a micro-unmanned aerial system (UAS) in a semi-arid savanna environment. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 84–96. [[CrossRef](#)]
6. Hernandez-Santin, L.; Rudge, M.L.; Bartolo, R.E.; Whiteside, T.G.; Erskine, P.D. Reference site selection protocols for mine site ecosystem restoration. *Restor. Ecol.* **2021**, *29*, e13278. [[CrossRef](#)]
7. Loewensteiner, D.A.; Bartolo, R.E.; Whiteside, T.G.; Esparon, A.J.; Humphrey, C.L. Measuring savanna woody cover at scale to inform ecosystem restoration. *Ecosphere* **2021**, *12*, e03437. [[CrossRef](#)]
8. Baldeck, C.A.; Asner, G.P.; Martin, R.E.; Anderson, C.B.; Knapp, D.E.; Kellner, J.R.; Wright, S.J. Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy. *PLoS ONE* **2015**, *10*, e0118403. [[CrossRef](#)]
9. Colgan, M.S.; Baldeck, C.A.; Féret, J.-B.; Asner, G.P. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. *Remote Sens.* **2012**, *4*, 3462–3480. [[CrossRef](#)]
10. Sothe, C.; Dalponte, M.; Almeida, C.M.d.; Schimalski, M.B.; Lima, C.L.; Liesenberg, V.; Miyoshi, G.T.; Tommaselli, A.M.G. Tree species classification in a highly diverse subtropical forest integrating UAV-based photogrammetric point cloud and hyperspectral data. *Remote Sens.* **2019**, *11*, 1338. [[CrossRef](#)]
11. Almeida, J.S. Predictive non-linear modeling of complex data by artificial neural networks. *Curr. Opin. Biotechnol.* **2002**, *13*, 72–76. [[CrossRef](#)] [[PubMed](#)]
12. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
13. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
14. Weinstein, B.G.; Marconi, S.; Bohlman, S.; Zare, A.; White, E. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sens.* **2019**, *11*, 1309. [[CrossRef](#)]
15. Wegner, J.D.; Branson, S.; Hall, D.; Schindler, K.; Perona, P. Cataloging public objects using aerial and street-level images-urban trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 30 June 2016; pp. 6014–6023.
16. Rukundo, O. Effects of Image Size on Deep Learning. *arXiv* **2021**, arXiv:2101.11508.
17. Sabottke, C.F.; Spieler, B.M. The effect of image resolution on deep learning in radiography. *Radiol. Artif. Intell.* **2020**, *2*, e190015. [[CrossRef](#)] [[PubMed](#)]
18. Azure Machine Learning. Available online: <https://azure.microsoft.com/en-gb/products/machine-learning/#product-overview> (accessed on 1 November 2021).
19. Custom Vision. Available online: <https://azure.microsoft.com/en-us/products/cognitive-services/custom-vision-service/> (accessed on 1 December 2019).
20. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.