In-situ aerial mapping of New Zealand Myrtaceae affected by Myrtle Rust (*Austropuccinia psidii*) using deep learning

by

Robin Pfaff

2026

School of Science

A thesis submitted to Auckland University of Technology in partial fulfilment of the requirements for the degree of Master of Science (Research) - Geospatial Science

Supervisors:
- Dr. Rebecca Rogers
- Graham Hinchliffe

Imprint

Form of citation
APA 7th edition

Keywords
Myrtle Rust, *Austropuccinia psidii*, *Syzygium maire*, Deep Learning, LiDAR, Multispectral, High-Resolution UAV-imagery

Abstract

Invasive fungal pathogens pose a significant threat to forest ecosystems worldwide and have far-reaching consequences for tree species. The rust fungus (*Austropuccinia psidii*) causes the disease commonly known as myrtle rust and threatens susceptible Myrtaceae populations on several continents. This includes *Syzygium maire*, a rare taonga (treasure) species endemic to New Zealand that is significant in Māori culture and ecologically important, but is now threatened with extinction. Accurate spatial mapping of threatened populations is essential for targeted management and conservation efforts, but traditional ground-based survey methods are logistically challenging and time-consuming.

The practical application of unmanned aerial vehicles (UAVs) in combination with deep learning was evaluated to detect *S. maire* in dense, species-rich native forests. High-resolution RGB (1.5 cm) and multispectral (2.5 cm) imagery were captured from four urban forest reserves on New Zealand's North Island using consumer-grade imaging sensors. A fully convolutional neural network for semantic segmentation (U-Net) was trained to classify *S. maire* from background vegetation. Furthermore, dataset composition and hyperparameter configurations were systematically tested including loss functions, learning rates, and different spectral band combinations. Point cloud segmentation approaches using a UAV-mounted LiDAR system were also qualitatively evaluated to assess the potential for three-dimensional tree instance detection.

Site-specific models showed moderate to good detection performance (F1 scores: 0.46–0.81), with RGB images performing comparably or marginally better than multispectral images. Dice loss outperformed pixel-wise approaches in handling severe class imbalances, and an aggressive learning rate of 0.02 with adaptive scheduling led to significantly better performance. However, generalising models across multiple sites proved more difficult due to site-specific differences (best F1=0.51). LiDAR-based instance segmentation algorithms developed for managed forests have potential for the dense, structurally complex context of natural forests, but are insufficient without further development.

The results show that deep learning can successfully identify *S. maire* under optimal, site-specific conditions. At the same time, critical limitations for operational real-world use were identified. The limited availability of training samples, the severe class imbalance of 3.4±0.7% (mean±SE) target class and the insufficient radiometric calibration capabilities of low-cost multispectral sensors fundamentally limit the approach. These findings highlight the need for standardised frameworks governing multispectral data capture and radiometric calibration.

The findings expose significant challenges in translating remote sensing methods from simplified scenarios to operational conservation monitoring in structurally complex forests. The methodological insights regarding hyperparameter optimisation, spectral band selection, and calibration challenges can be transferred to analogous applications for species detection. As invasive pathogens increasingly threaten forest biodiversity worldwide, the development of robust detection methods for (rare) species in complex environments remains essential to support proactive conservation and biosecurity measures.

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor used artificial intelligence tools or generative artificial intelligence tools (unless it is clearly stated and referenced, along with the purpose of use), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signed:

Robin Pfaff

Date: 19/02/2026

Statement of Contribution

All data collection, processing, and analysis were completed by R.P., with assistance from G.H. and R.R. during UAV operations; initial LiDAR data processing was completed by G.H., with further processing and analysis by R.P. The thesis was written by R.P., with feedback from R.R. and G.H.

Acknowledgements

I would like to thank the following people who provided guidance and support during this thesis project. My supervisors, Dr. Rebecca Rogers and Graham Hinchliffe, for their continuous feedback, support, and encouragement throughout the project, especially for the open and uncomplicated communication. I would also like to thank Derek Craig (formerly of Kaipātiki Project) for initial guidance on site selection, and Maria Valkova (Kaipātiki Project), Kathy McCormack (Friends of Bushglen Reserve), and Robert Beresford (Bioeconomy Science Institute) for directing me to the *S. maire* during fieldwork and providing me with valuable contextual information around *S. maire*. I would also like to thank Jamie Chapman and Andrea Campos, not only for their help with UAV operations, but also for their ongoing mental support. Furthermore, I would like to thank Laurin Frey for his help with proofreading and feedback. Finally, and most of all, I would like to thank my Godmother (Therese Rudolph von Rohr) and my parents (Doris and René Pfaff) without whom I would not have been able to realise my dream of pursuing a Master's degree in beautiful Aotearoa.

List of Figures

List of Tables

Table of contents

Chapter 1 Introduction

1.1 Global Emergence and Spread of Myrtle Rust

The spread of invasive fungal pathogens is increasingly impacting forest ecosystems worldwide, with their dispersion to new areas partially facilitated by ever-increasing global trade and travel (Grgurinovic et al., 2006; Paap et al., 2022; Wingfield et al., 2017). Rust fungi belong to the diverse phylum Basidiomycota and are biotrophic plant parasites (Helfer, 2014). One member of the rust fungi, *Austropuccinia psidii*, formerly *Puccinia psidii* (Beenken, 2017), has a broad host range outside its native habitat. Originating from South and Central America, one biotype of this fungus is considered pandemic and spread to Hawaii, Australia, Southeast Asia, China, Japan, and New Zealand (Du Plessis et al., 2019; Granados et al., 2017; Stewart et al., 2018). Globally, *A. psidii* infects approximately 480 Myrtaceae species (Soewarto et al., 2019), causing the disease commonly known as myrtle rust. Due to the lack of co-evolution, Oceanic Myrtaceae did not have the opportunity to adapt to myrtle rust, making them epidemiologically naïve to the pathogen and, as a result, more susceptible (Makinson, 2018). Whilst it is thought that *A. psidii* has arrived in New Zealand in 2017 via wind, its urediniospores can also spread by humans, birds or insects (Magyar et al., 2016; Schmid et al., 2021).

1.2 Consequences of Myrtle Rust for New Zealand

Since the pathogen's arrival in 2017 in New Zealand, at least 13 out of 30 native (De Lange et al., 2024) and 17 exotic Myrtaceae species are confirmed as hosts (Toome-Heller et al., 2020). *A. psidii* infects actively growing tissue and can cause yellow pustules, necrotic lesions, shoot dieback, and defoliation (Boufleur et al., 2023). This reduces regeneration and increases mortality risk in susceptible Myrtaceae (Fensham & Radford-Smith, 2021). Given these potential consequences of *A. psidii* infection, myrtle rust poses far-reaching implications across ecological, cultural and economic systems of Aotearoa. In woody ecosystems successional plant communities and Myrtaceae-dominated old-growth forests are at most risk (McCarthy et al., 2024). Widespread loss of Myrtaceae from myrtle rust would cause substantial changes to the functional structure of New Zealand's forests, particularly if multiple Myrtaceae species become infected at the same site and co-occurring species lack the capacity to compensate for their loss (McCarthy et al., 2024)

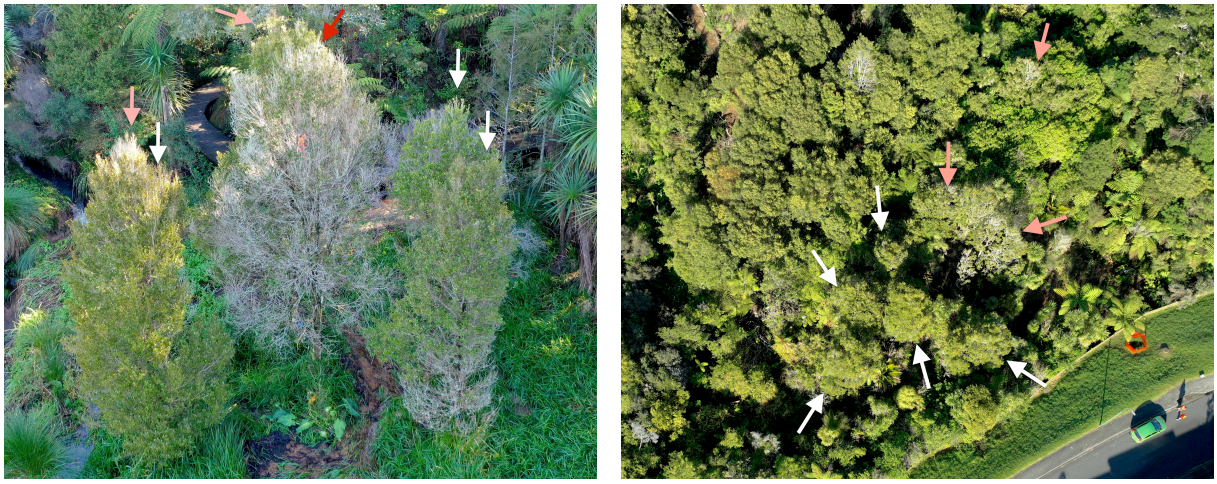
Myrtaceae species are keystone species, and birds, reptiles, bats and invertebrates feed on their nectar (Gailbraith & Large, 2017; Jo et al., 2022; Sutherland et al., 2020). Furthermore, bats are reported to hunt for invertebrates above Myrtaceae (Gailbraith & Large, 2017). However,

while birds, bats and reptiles are projected to be flexible enough to overcome decline of Myrtaceae, some invertebrates are thought to have an obligate relationship with Myrtaceae species (Gailbraith & Large, 2017). Beyond ecological concerns, many Myrtaceae are taonga (treasure) species which hold profound cultural and spiritual significance for Māori (Teulon et al., 2015). Furthermore, Myrtaceae species have significant economic value (Essien et al., 2019). Mānuka (*Leptospermum scoparium*) honey generated NZ$538 million in 2022 and is projected to rise to over NZ$1 billion by 2030 (Apiculture New Zealand, 2024). Consequently, there is considerable concern about the long-term impact of myrtle rust on the Myrtaceae species, because of their importance to Māori culture, the New Zealand economy, and its ecosystems.

1.3 *S. maire*: Ecology, Significance, and Conservation Challenges

1.3.1 Biology and Taxonomy

Syzygium is the most diverse tree genus globally, comprising over 1000 species (Low et al., 2022), and serves as an important food source for birds, mammals and insects across the tropical Indo-Pacific (Parnell et al., 2007). Within New Zealand, *Syzygium maire* is the only native species of the Syzygium genus (De Lange et al., 2024). *S. maire* is a slender tree and can grow up to 16 m tall with numerous branches and a compact crown, comprised of ~1.5 cm wide and ~5cm long leaves (Figure 1). It is characterised by specialised pneumatophores (aerial roots), enabling it to inhabit waterlogged soils. Between January and April, healthy and mature individuals produce white brush flowers and bright red, one-seeded edible berries alongside immature green fruit (De Lange et al., 2024; Mahuta et al., 2021; Taranaki Regional Council, n.d.).



(a) A stand of six *S. maire* trees with variable infection severity. **(b)** Aerial view of *S. maire* within mixed native forest canopy.

Figure 1: Examples of *S. maire* in urban parks of a separated stand (**a**) and surrounded by vegetation (**b**). Arrows indicate tree locations; colours represent dieback severity (white = limited, light red = advanced, red = severe mortality).

1.3.2 Distribution, Habitat, and Conservation Status

Species distribution models (SDMs) indicate that *S. maire*'s historical habitat spanned from the north of the South Island across most parts of the North Island (McCarthy et al., 2019). *S. maire* is most commonly found in lowland wetlands with low water tables and along stream gullies (De Lange et al., 2024; McCarthy et al., 2019). Its pneumatophores enable it to be a dominant component of swamp forests, where it often occurs alongside pukatea (*Laurelia novae-zelandiae*), a more widespread generalist species with similarly small leaves (3 × 6 cm; Mahuta et al., 2021; Manaaki Whenua - Landcare Research, 2025). Other co-occurring species include tawa (*Beilschmiedia tawa*), which also possess small leaves (1.4 × 6 cm), though narrower than those of *S. maire* (Manaaki Whenua - Landcare Research, 2026), and climbing plants such as kiekie (*Freycinetia banksii*), and supplejack (*Ripogonum scandens*; Singers & Rogers, 2014), which may establish within the canopy of *S. maire*. In addition, human-driven deforestation and land drainage have restricted *S. maire* to a fraction of its original distribution. It is estimated that only 10% of New Zealand's wetlands are still intact (Dymond et al., 2021). Consequently, *S. maire* is in decline and classified as "Threatened–Nationally Vulnerable" (De Lange et al., 2024), New Zealand's highest extinction risk category (Townsend et al., 2008).

1.3.3 Ecological and Cultural Significance

Ecologically, *S. maire* plays a critical role in forest structure and function. Its fruits provide a food source for native birds including kererū (*Hemiphaga novaeseelandiae*), tūī (*Prosthemadera*

novaeseelandiae), and korimako (*Anthornis melanura*; Van Der Walt et al., 2020), and its presence provides structural habitat and shelter for some of New Zealand's most threatened freshwater fish species (Mahuta et al., 2021). Notably, *S. maire* can achieve reproductive maturity early, with fruit production documented in individuals five to ten years old (Balkwill et al., 2025). Like other Myrtaceae species, *S. maire* holds profound cultural and spiritual significance for Māori as taonga species, valued as a source of kai (food), rongoā (medicine), dye, perfume or air-freshener, and timber weapons for self-defence (Balkwill et al., 2024; Gould et al., 2006; Mahuta et al., 2021; Taranaki Regional Council, n.d.). This cultural importance, combined with its ecological role in swamp forests, ecosystems that have suffered widespread loss in Aotearoa, underscores the conservation urgency for this species.

1.3.4 Susceptibility to Myrtle Rust and Research Gaps

A. psidii infection severity varies widely among host species (Fernandez Winzer et al., 2019). *S. maire* is one of the most susceptible Myrtaceae species to *A. psidii* infection (Beresford, 2025; Schmid et al., 2021), with mature trees capable of dying within three years of initial infection (De Lange et al., 2024). Upon *A. psidii*'s arrival in New Zealand, the biology of *S. maire* remained scarcely researched and understood, contributing to the absence of a robust conservation strategy. However, this research gap is being progressively closed, with recent publications exploring genome assembly (Balkwill et al., 2024), adaptive potential (Balkwill et al., 2025), and reproductive strategies (Bettoni et al., 2024). Ex-situ cryopreservation of *S. maire* seeds is also being investigated as a potential long-term conservation option, though practical application remains unfeasible (Nadarajan et al., 2021; Van Der Walt et al., 2020, 2022, 2023; Van Der Walt, 2021). Furthermore, *S. maire* seeds exhibit characteristics that prevent successful long-term freezing (Van Der Walt et al., 2023), limiting the viability of seed banking as a standalone conservation strategy. These constraints place further emphasis on the urgent need for in-situ conservation efforts, including improved mapping and monitoring of remaining populations.

1.4 UAV-based Deep Learning for Plant Species Mapping

Deep learning has transformed species mapping from remote sensing by enabling models to learn hierarchical spectral–spatial features directly from remotely sensed imagery and structural features from 3D point clouds, supporting end-to-end detection and classification workflows (LeCun et al., 2015; L. Zhong et al., 2024). Convolutional neural network (CNN) and more recent encoder–decoder segmentation architectures applied to high-resolution UAV and satellite imagery can capture fine-grained canopy texture and spectral signals that are often indicative

of species identity at the crown and sub-crown scales (Kattenborn et al., 2019; Lobo Torres et al., 2020). Complementary 3D approaches that operate on light detection and ranging (LiDAR)-derived point clouds (including PointNet++ and voxel-based models) extract crown geometry, vertical structure, and other structural traits to distinguish species with subtle foliar differences (Briechle et al., 2020; Qi et al., 2017). Furthermore, multimodal approaches fuse 2D with 3D data to improve discrimination in complex, multi-species canopies, though they introduce challenges around sensor co-registration and increased computational complexity (Briechle et al., 2020; L. Zhong et al., 2024).

Despite these advances, operational deployment faces constraints including limited instances for rare species, strong class imbalance, site-to-site transferability, and the need for careful validation against ground truth. Given the complexity of fusing multimodal data, this thesis focuses on single-source data, while treating the 3D workflow as an exploratory complement to the primary 2D approach.

1.5 Research Aims and Scope

The need for automated *S. maire* mapping is driven by the species' ecological importance, its vulnerability to myrtle rust, and the practical limits of field-based monitoring. This thesis tests whether current unmanned aerial vehicle (UAV)-deep learning approaches can support conservation monitoring of a rare, morphologically subtle species in dense, biodiverse native forest. Accordingly, the research focuses on four urban forest reserves in the North Island of New Zealand with known *S. maire* populations to explicitly evaluate operational capabilities and limitations when detecting non-dominant canopy species.

This thesis addresses four research questions (RQs): (1) Can current deep learning approaches successfully detect and delineate *S. maire* in complex, dense native forest despite severe class imbalance, morphological similarity to co-occurring species, and limited training data? (2) How do spectral characteristics (RGB imagery vs. multispectral (MS) imagery) and hyperparameter configurations affect detection performance? (3) Can models that are trained on multiple sites generalise across all locations and outperform models trained on a single site? (4) Can readily available 3D LiDAR-based point-cloud segmentation algorithms supplement 2D-based fully convolutional neural network (FCNN) detection?

To address these questions, two complementary methodologies were evaluated to test whether structural and spectral information could improve detection under the given constraints: (1) a 2D FCNN semantic segmentation approach using RGB imagery and MS imagery data, which

exploits spectral and spatial texture information at high resolution, and (2) a 3D point cloud-based approach using LiDAR data with tree instance segmentation (Figure 2). We hypothesise that structural information (3D crown geometry, vertical position relative to other trees) and spectral information (foliage reflectance, vegetation index (VI)) may provide complementary signals for species discrimination, particularly for small-leaved species where morphological distinctiveness is subtle. While the deep learning portion of this thesis ultimately focusses on the first method, the evaluation aimed to assess how both these methods could benefit species discrimination. Furthermore, evaluating both methods provides insights into whether complex forest environments require multi-modal data fusion or whether single modalities suffice under current sensor and algorithmic constraints.

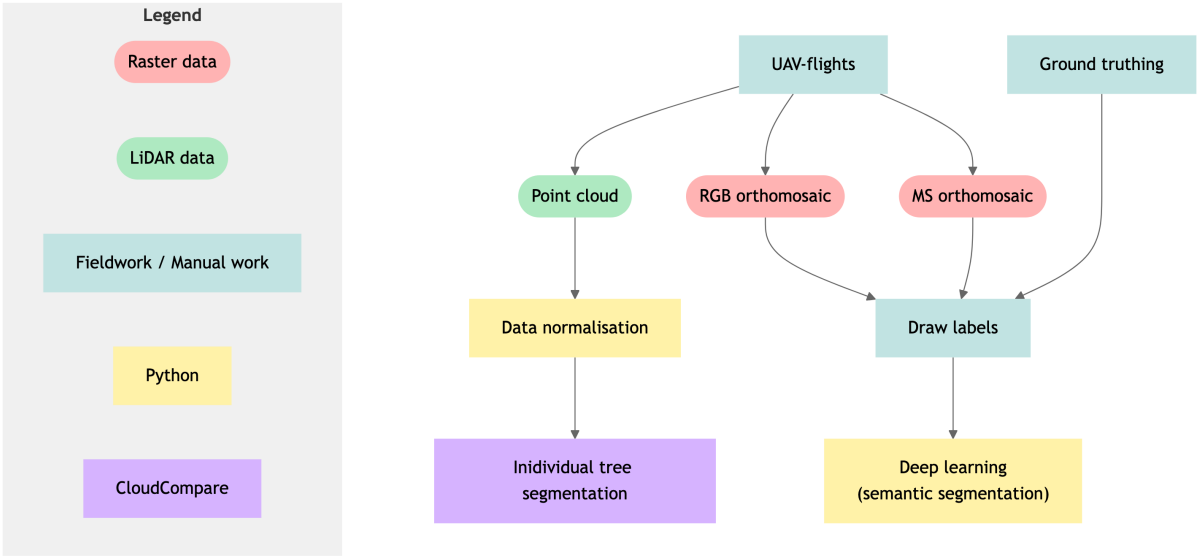


Figure 2: Overview of the UAV-based data workflow used in this thesis, showing field acquisition and the two analytical branches: 2D semantic segmentation of RGB/MS imagery and exploratory LiDAR-based individual tree segmentation.

Chapter 2 Literature Review

2.1 Plant-level Symptoms of Myrtle Rust

A. psidii spores infect the host-plant only on actively growing tissue like young shoots, developing leaves and fruits (Boufleur et al., 2023; Chock, 2020; Glen et al., 2007). After germination and a latent period of approximately two weeks, the symptoms start to show (Figure 3; Beresford et al., 2020). Initially, the growing uredinial spores form yellow pustules (Figure 3a). These growth areas turn into red blotches on leaves, killing the photosynthetic tissue and preventing the growth of new meristematic tissue, leading to the death of shoot tips (Figure 3b) and flowers. Sustained infection ultimately leads to substantial defoliation (Boufleur et al., 2023). Hence, infection hinders the host to produce new photosynthetic tissue and development of seeds, disrupting the reproductive cycle (Fensham & Radford-Smith, 2021). In severe cases, sustained tissue loss and inability to produce new seeds can render the host infertile, ultimately leading to mortality (Fensham & Radford-Smith, 2021). Even when seed production is not entirely suppressed and seedlings germinate, they are highly susceptible to infection, severely limiting regeneration (Fernandez-Winzer et al., 2020). Such reproduction failure has raised concerns that highly susceptible species may face extinction within a single generation (Carnegie et al., 2016; Fensham & Radford-Smith, 2021). As a consequence, this may have cascading effects across trophic levels, impacting birds, insects, and other organisms that depend on these species as habitat or food (Sutherland et al., 2020).

(a) Yellow pustules on leaves. Image: Copyright by Bioeconomy Science Institute. Reprinted with permission.

(b) Red blotches on leaves leading to necrotic lesions and shoot tip dieback.

Figure 3: The myrtle rust symptoms showing the advancement of *A. psidii* infection on *S. maire*. Initial growth of yellow urediniospores on leaves (**a**), and necrotic lesions and shoot tip dieback (**b**).

2.2 Management and Mitigation Strategies

Myrtle rust is considered to be established in New Zealand since 2018, and efforts are focused on long-term mitigation (Ministry for Primary Industries, 2020). Current strategies are targeted at protecting susceptible species and reducing the rate of spread (Beresford, 2025). These include removal of infected plant parts, or entire plants with minimal disturbance, and disposal in a manner that minimises the risk of further spore dispersal. Where possible, pruning is recommended during cooler seasons to avoid stimulating new growth. Additionally, clothing and equipment should be disinfected between sites. Synthetic fungicides offer an additional option for reducing infection severity and have shown to be highly effective in treating myrtle rust with low phytotoxicity (Beresford & Wright, 2022; Chng et al., 2019). However, their effectiveness depends on regular application, making them costly and impractical for large-scale use (Sutherland et al., 2020). Nevertheless, fungicide treatment may be viable in cases where only a small number of trees require protection, provided that their exact locations are known.

2.3 Current Mapping of *S. maire*

Currently, *S. maire* is predominantly mapped using community science platforms like iNaturalist. While iNaturalist is increasingly recognised as a valuable tool in biodiversity research and used to develop SDMs (Mason et al., 2025), its data is inherently biased to areas with higher human

activity and accessibility. Geurts et al. (2023) found that 94% of observations are within 1 km of roads, demonstrating the spatial bias of such opportunistic datasets. Additionally, local conservation groups, such as the Kaipātiki Project or Friends of Bushglen Reserve, maintain records of *S. maire* locations within their management areas. These known populations can provide the basis for developing and validating automated detection methods. Beyond discovering novel stands, deep learning-based detection offers the advantage of automating repeated monitoring to track tree health, growth, and infection progression over time. Another approach consists of SDMs, in which a multifactorial GIS analysis is developed to identify their most probable habitat. Such a model already exists for the greater Wellington region for *S. maire* (Herbert et al., 2025). However, such models only predict the likelihood of occurrence, but not their exact location, which is critical for current management approaches for myrtle rust.

remote sensing has become a widely used tool for tree species identification (Chang et al., 2025; L. Zhong et al., 2024) and represents a promising option to close the mapping gap for *S. maire*, which could be built on SDM outputs. However, applying these methods to *S. maire* presents inherent challenges: the species' rarity makes gathering of reference data challenging, its morphological overlap with co-occurring species, such as pukatea and tawa, complicates identification, and the potential for climbing plants to establish within its canopy further obscures individual trees for detection on aerial imagery.

2.4 Deep Learning for Species Mapping

Tree detection in aerial imagery has followed the broader computer vision trend toward deep learning solutions (LeCun et al., 2015). Even though still used (Bosch, 2020; Malek et al., 2014; Miraki et al., 2021; Safonova et al., 2021), conventional machine learning methods, such as random forests (RFs) or support vector machines (SVMs) that classify individual pixels independently, have been outperformed by deep learning methods (Liu et al., 2018), which can even outperform humans in some cases (Alzubaidi et al., 2021). Many studies focus on tree segmentation as their primary objective: detecting individual trees regardless of species (Ball et al., 2023; Gan et al., 2023; Ulku et al., 2022). Beyond this foundational step, a more advanced and challenging task involves species-level classification, where trees are identified and assigned to their respective species (Briechle et al., 2020; Kattenborn et al., 2019; Lobo Torres et al., 2020; H. Zhong et al., 2024). To achieve this species-level accuracy, methodologies broadly fall into two categories: 2D-sensed architectures, such as CNNs, and 3D-sensed architectures for LiDAR point clouds. Both are widely used; however, because 2D approaches can draw on a broader range of data sources (RGB imagery, MS imagery, hyperspectral (HS) imagery) and are

often more cost-effective to acquire, there is substantially more deep learning research in remote sensing based on 2D-sensed than on 3D-sensed data (Figure 4; L. Zhong et al., 2024).

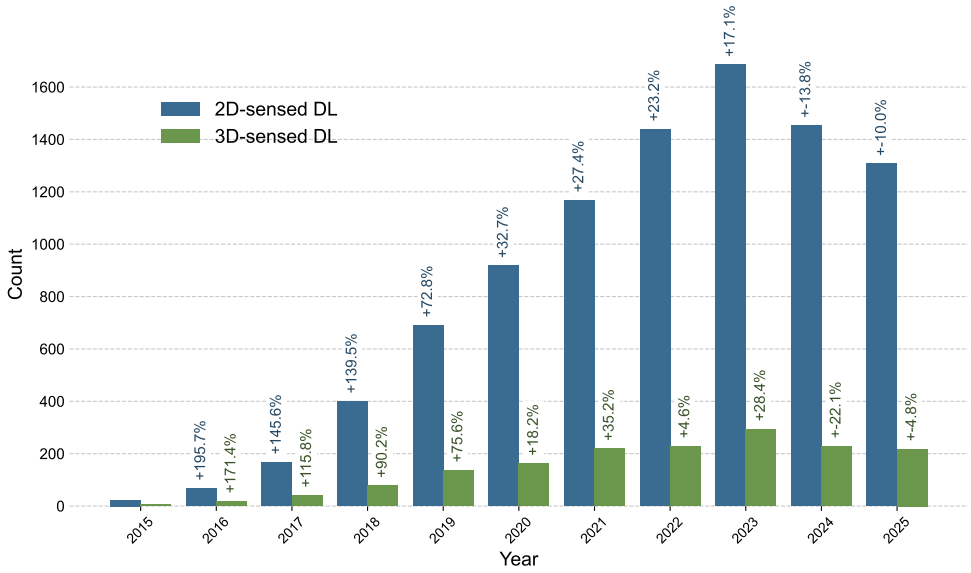


Figure 4: Total count and percentage increase between successive years of peer-reviewed scientific articles in the field of remote sensing related to deep learning from 2015 - 2025. Blue represents publications related to 2D-sensed and green to 3D-sensed data. Search terms can be found in Appendix B and were retrieved from OpenAlex (Priem et al., 2022).

The comparison matters because 2D methods are easier to scale; 3D methods exploit crown geometry and vertical context, but they need more preprocessing and tree isolation before inference. Neither line of work has fully resolved rare-species mapping in dense forest under severe class imbalance.

2.4.1 2D Semantic Segmentation

CNNs were originally designed for whole-image classification tasks (Krizhevsky et al., 2017) and have been applied by community science tools like iNaturalist, Flora Incognita, and Pl@ntNet to identify plant species from photographs (Dyrmann et al., 2016; Wäldchen et al., 2018). These networks revolutionised object detection through their capacity to analyse image texture (contextual information from clusters of neighbouring pixels) rather than evaluating pixels in isolation (Chollet, 2017). They autonomously learn which textural patterns (such as leaf morphology or canopy characteristics) are relevant for distinguishing vegetation communities and species (Dyrmann et al., 2016). In these architectures, sequential pooling operations aggregate feature maps to progressively coarser spatial scales, with the final layer indicating only whether distinct (species) features appear anywhere within the image, without preserving their spatial locations.

This self-learning capacity provides substantial advantages in computational efficiency and automation, eliminating manual feature engineering required for traditional machine learning (Maggiori et al., 2017). Combined with advances in high-resolution sensors, CNNs currently transform vegetation mapping capabilities.

However, remote sensing-based vegetation mapping demands more than whole-image classification can provide. For mapping applications, success depends not on detecting whether the target class exists, but on determining its precise spatial location throughout the imagery, requiring spatially-continuous, detailed classification across entire image extents (Kattenborn et al., 2019). FCNNs address this limitation by retaining and reconstructing the spatial locations of contextual features through encoder-decoder structures. These networks extract contextual features whilst preserving spatial origin, enabling fine-grained, spatial segmentation at the original imagery resolution (Maggiori et al., 2017; Ronneberger et al., 2015).

U-Net, originally developed for biomedical image segmentation (Ronneberger et al., 2015), has become a widely adopted FCNN architecture for species detection in remote sensing (Freudenberg et al., 2019; Kattenborn et al., 2019; Lake et al., 2022; Ulku et al., 2022; Wagner et al., 2019; C. Zhang et al., 2020). The U-shaped encoder-decoder architecture progressively reduces spatial resolution whilst increasing feature channels, enabling efficient pixel-wise semantic classification. Although numerous FCNN architectures exist, their performance tends to be comparable to U-Net's (Lobo Torres et al., 2020). In MS approaches, imagery is predominantly used with traditional machine learning methods such as SVMs, yet is barely employed in CNN-based studies (L. Zhong et al., 2024).

This is the main trade-off in the 2D literature: U-Net-style models restore the localisation that whole-image CNNs lose, but most published successes still rely on relatively accessible targets or less severe class imbalance than this thesis faces.

2.4.2 3D Instance Segmentation and Point Cloud Processing

For 3D LiDAR data, specialised neural network architectures have been developed to process forest point clouds. These methods leverage the vertical dimension and 3D tree crown geometry to discriminate between species, offering complementary information to spectral approaches (Henrich et al., 2024; H. Zhong et al., 2024). Two primary architectural approaches exist: voxel-based methods that convert point clouds into regular 3D grids suitable for convolutional operations (Lecigne et al., 2018; Xiang et al., 2024), and point-based architectures such as PointNet++ that process raw point coordinates directly whilst preserving geometric detail (Qi et

al., 2017). Recent studies have demonstrated successful species identification across diverse forest types using these approaches (Briechle et al., 2020).

However, point cloud data presents unique preprocessing challenges compared to imagery. Unlike raster data where spatial dimensions are standardised through tile size, point clouds contain variable numbers of points depending on forest structure and acquisition parameters. For species-level classification, individual trees must first be isolated (Briechle et al., 2020). This tree instance segmentation can be achieved through graph-based methods such as Treeiso (Xi & Hopkinson, 2022) or deep learning approaches like TreeLearn (Henrich et al., 2024). Once individual trees are extracted, point counts must be normalised to provide consistent neural network inputs, either through voxelisation into fixed grid sizes or through resampling to a standard number of points per tree. These preprocessing requirements add complexity but enable the exploitation of 3D structural information to potentially use for species identification.

That preprocessing requirement is the key limitation: LiDAR adds structural information that helps separate similar crowns, but it also makes the workflow less straightforward to operationalise for repeated monitoring.

2.5 Synthesis and Research Gap

A critical challenge for detecting rare species is severe class imbalance combined with limited training data. Most published studies report target classes comprising 30–50% of annotated pixels (Lobo Torres et al., 2020), whereas rare species in complex, biodiverse forests present significantly different characteristics. These challenges create a substantial gap between studies on readily detectable species (Pearse et al., 2021) and operational requirements for rare species in complex ecosystems. *S. maire* shares morphological similarities with co-occurring native species, possesses small leaves that are difficult to resolve at typical UAV flight altitudes, and is a rare and conservation-critical New Zealand tree species highly vulnerable to myrtle rust. Taken together, these factors make it an appropriate case study for testing whether current deep learning approaches can successfully map rare, morphologically inconspicuous species in complex forest environments.

The literature therefore suggests three unresolved issues relevant to this thesis: many studies are developed around comparatively detectable target classes and do not fully represent the imbalance and ambiguity of rare-species monitoring; 2D and 3D approaches each capture different information, but the literature does not yet clearly show which modality is more robust under dense, species-rich forest conditions; and existing studies rarely assess whether models

trained on one site can generalise across other locations with different canopy structure and background complexity. These gaps motivate the thesis objective and the methods evaluated in the next chapter.

Chapter 3 Methodology

This chapter describes the methodological approach used to address the thesis research questions, and is organised into four parts: (1) experimental design and data collection, (2) data processing and model development, (3) evaluation and analysis, and (4) exploratory LiDAR tree segmentation. The experiments were designed to investigate detection accuracy, the contribution of different sensor and band combinations, and model generalisability across sites.

3.1 Experimental Design and Data Collection

3.1.1 Research Sites

The thesis was conducted in four urban parks in Auckland and Hamilton, New Zealand (Figure 5). Site selection was guided by known occurrences of *S. maire* as reported on iNaturalist and by local conservation groups. Additional sites were scouted around Tauranga, Auckland and Hamilton but were not considered further due to limited presence (< five individuals) or complete or highly advanced mortality of *S. maire*. The final four sites were selected based on a minimum of five trees and an age range of approximately 20–60 years, ensuring cross-site comparability and supporting the cross-site generalisability analysis (RQ2).

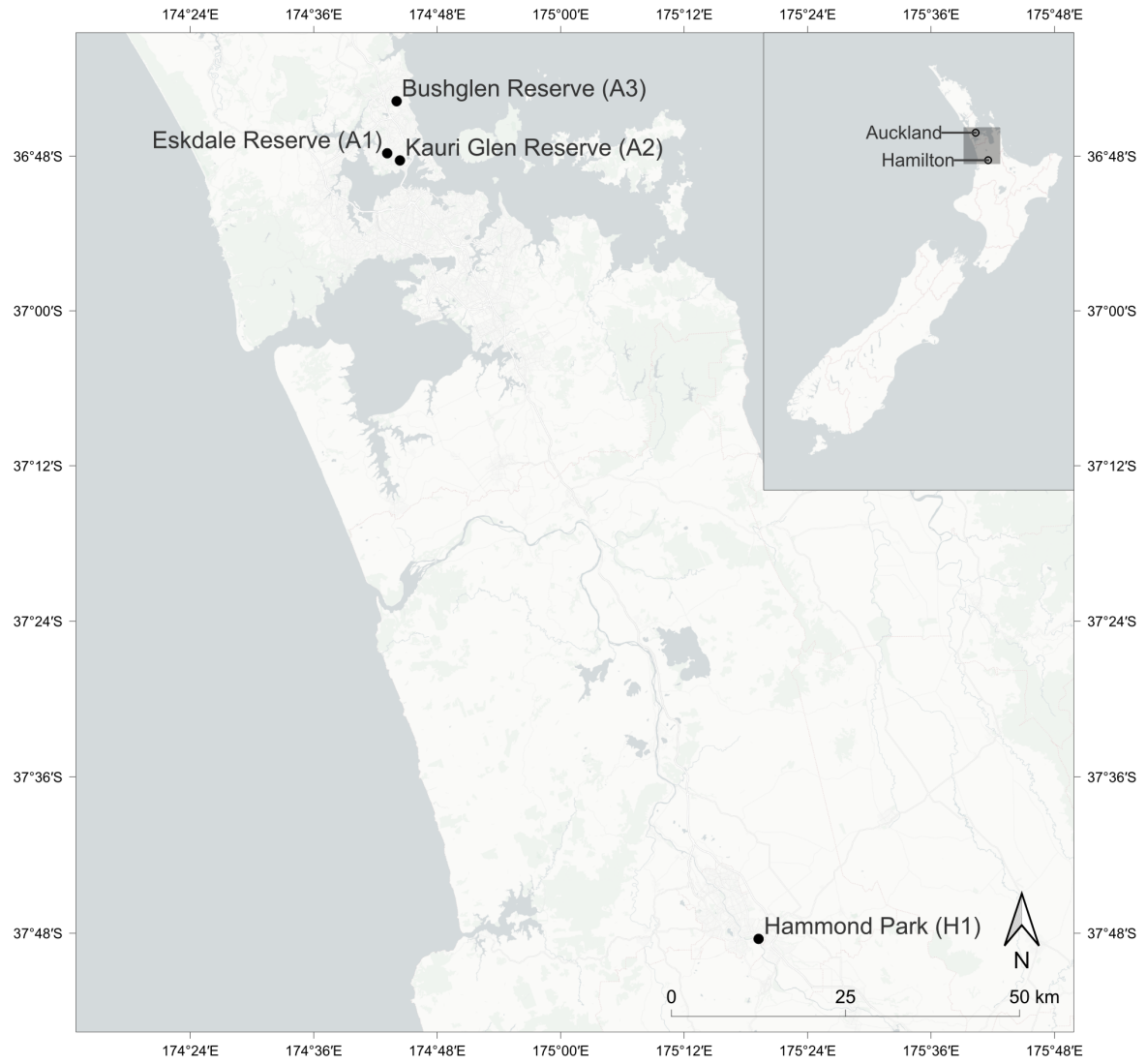
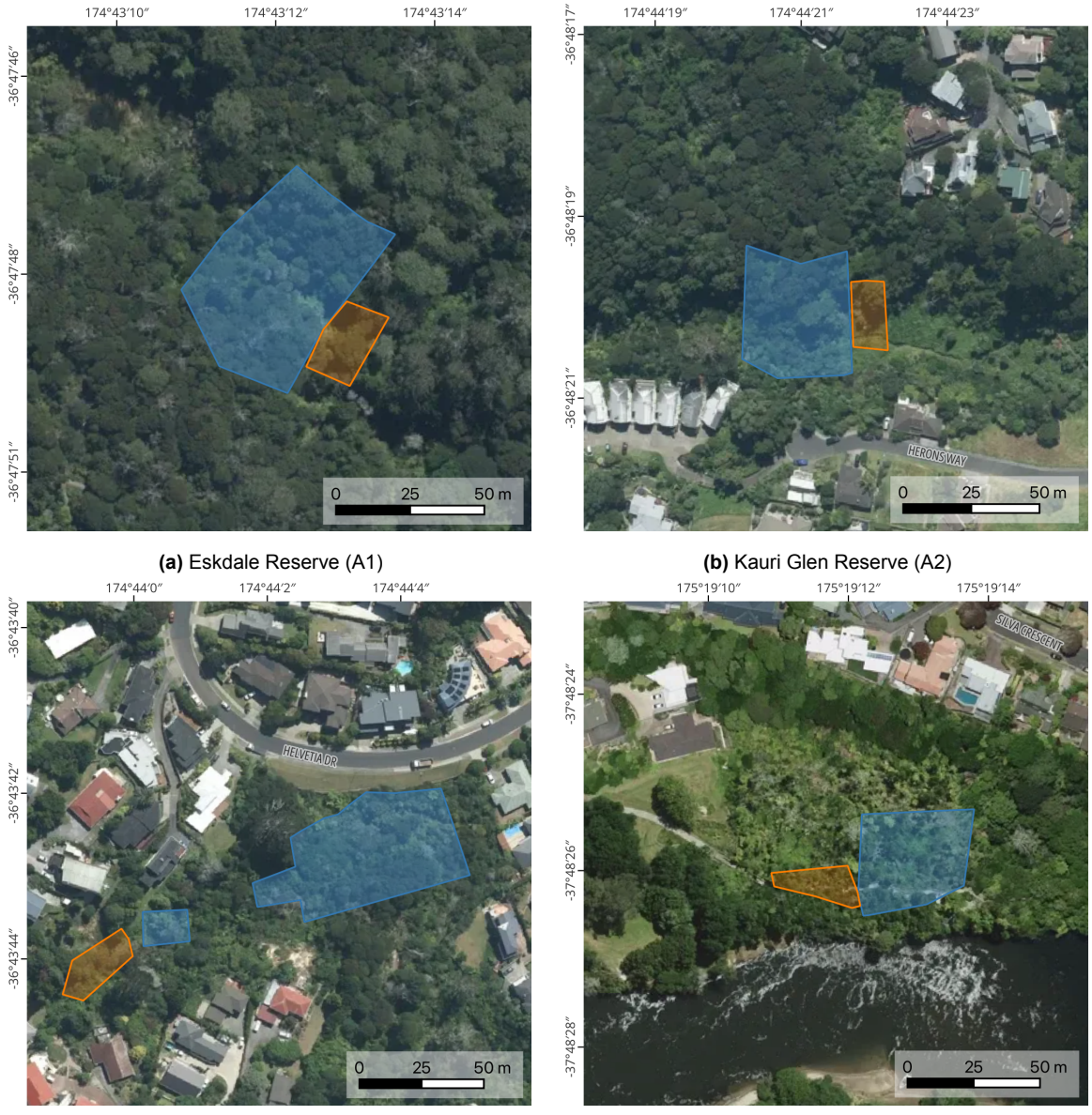


Figure 5: Overview of the four research sites, of which all are located within urban parks. Three are located in Auckland and one in Hamilton. Basemap: Copyright CARTO and OpenStreetMap contributors under ODbl by OSFM.

The regrowth forest at all sites is classified as warm temperate forest type 8 (WF8): Kahikatea, pukatea forest (Singers & Rogers, 2014). This forest type comprises a podocarp and broadleaved community dominated by kahikatea (*Dacrycarpus dacrydioides*) with scattered pukatea, kiekie and supplejack, and occasional rimu (*Dacrydium cupressinum*), tawa and *S. maire*, predominantly on organic and gley soils characterised by elevated water tables. The canopy consists of a dense, complex matrix of tree species, making visual identification of *S. maire* challenging, especially given the similar appearance of co-occurring native trees. The sites are surrounded by residential housing, ranging from 2.5–40 ha (Table 1), though *S. maire* was only present in small portions of each site (Figure 6). On three sites (A1–A3), *S. maire* is under fungicide spray management, whilst H1 is unmanaged (Table 1; R. Beresford, personal communication, October 29, 2025).

Table 1: Summary of the research sites with *S. maire* populations. LiDAR captures at Eskdale Reserve (A1) consisted of one nadir and four oblique flight legs, while those at Kauri Glen (A2) consisted of one nadir and two oblique flight legs. Number refers to the number of *S. maire* canopies visible on aerial imagery. Area is for the entire forest within that park. Train refers to the area of the training zone and te st to the area of the test zone (Figure 6).

Name	Code	Num.	Fungicide	UAV-data	Area (ha)	Train (m^2)	Test (m^2)
Eskdale Reserve	A1	15	2023	RGB/MS, LiDAR (1+4)	40	2,916	426
Kauri Glen Res.	A2	17	2023	RGB/MS, LiDAR (1+2)	21	1,667	306
Bushglen Res.	A3	23	2022	RGB/MS	2.5	3,017	405
Hammond Park	H1	15	No	RGB/MS	6	1,428	324



(a) Eskdale Reserve (A1)

(b) Kauri Glen Reserve (A2)

(c) Bushglen Reserve (A3)

(d) Hammond Park (H1)

Figure 6: Site extents showing training (blue) and test (orange) zones used for model evaluation; maps showing full park extent and surrounding area can be found in Figure S1. Site areas are listed in Table 1. Basemap: Copyright LINZ aerial imagery under CC BY 4.0.

3.1.2 RGB and MS Acquisition

Flights were carried out with a DJI Mavic 3 Multispectral (DJI, n.d.-a) UAV (Figure 7a), equipped with a RGB visible imaging sensor (20 MP, 5280 × 3956) and a MS imaging sensor (5 MP, 2592 × 1944) across four spectral bands: green (G) (560±16 nm), red (R) (650±16 nm), red edge (RE) (730±16 nm), and near-infrared (NIR) (860±26 nm). Initial RGB imagery and MS imagery captures were conducted in April 2025 but had to be repeated due to variable lighting conditions, resulting in poor data quality (Figure S2). Final RGB imagery and MS imagery were acquired on 19 September (A1–A3) and 20 September (H1) 2025.

(a) The DJI Mavic 3 Multispectral UAV with the RGB and MS sensors. Image: From wikimedia by *ZLEA* (2024). CC-BY-SA-4.0.

(b) The DJI Matrice 350 UAV-platform equipped with the Zenmuse L2 LiDAR sensor. Image: Copyright by Auckland University of Technology (AUT). Reprinted with permission.

Figure 7: The two UAVs used for RGB/MS (**a**) and LiDAR (**b**) data collection.

All flights were carried out under full sunlight with wind speeds below 6 m/s. To ensure optimal exposure, flights were conducted within two hours of solar noon (Robbins et al., 2025). Flight paths were automatically generated in DJI Pilot 2 (DJI, n.d.-b), and the RTK system was connected to the nearest GNSS base station from Toitū Te Whenua PositioNZ Real Time Service (PositioNZ-RT) to ensure high positional accuracy. Flight altitude was set to 50 m above ground level (AGL), based on a 1 m digital elevation model (DEM), corrected to the geoid (LINZ, 2012, 2025). Forward and side overlap were set to 85%, and flight speed was specified at 3 m/s. These settings were chosen and adapted based on the recommendations in Heim et al. (2019). As suggested by Olsson et al. (2021), a photo of a MicaSense RP03 Calibrated Reference Panel (CRP) was taken after takeoff and before landing with all other settings left at their defaults.

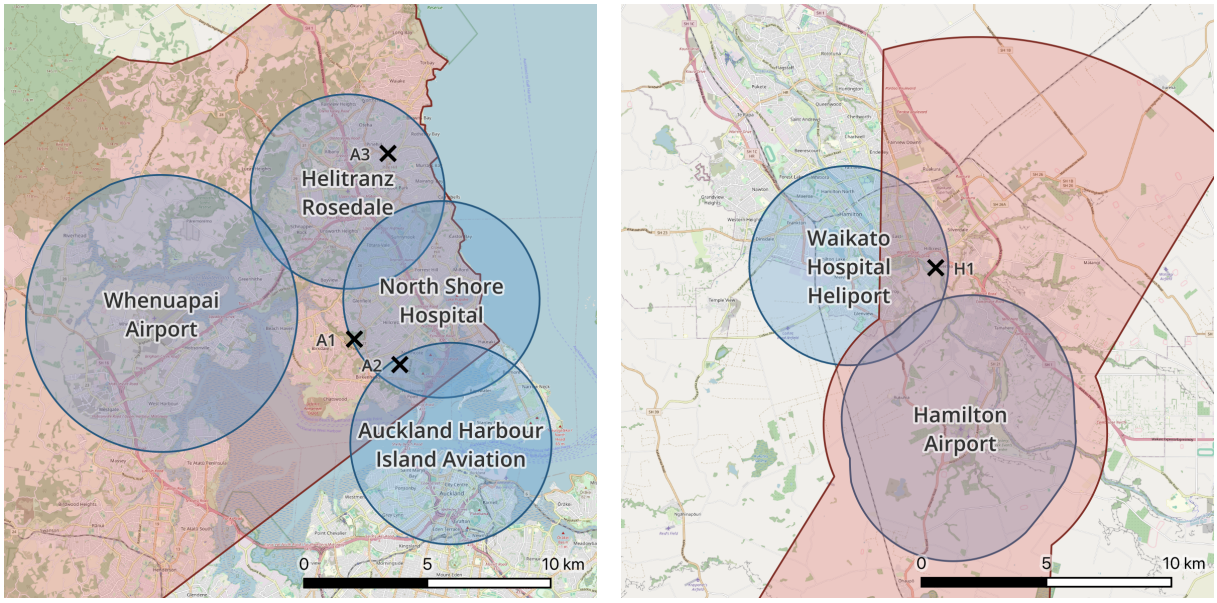
3.1.3 UAV-LiDAR Acquisition

LiDAR flights were carried out on 22 May 2025 using a DJI Matrice 350 RTK (DJI, n.d.-c) equipped with a Zenmuse L2 sensor (Figure 7b; DJI, n.d.-e). The Zenmuse L2 is a LiDAR sensor, which emits laser impulses and measures the time for the laser pulse to be reflected

by obstacles (e.g. vegetation, buildings or the ground) and return to the sensor. The time is then converted to distance in the given direction, which results in a three-dimensional point cloud. Flight altitude was set to 60 m AGL at A1 and 70 m AGL at A2. The reduced flight height at site A1 was chosen to maintain operational safety margins with a co-occurring flight at this location. Each site included one nadir flight and four (A1) or two (A2) oblique flights at a 20° angle (Table 1). This difference in oblique flights was due to operational limitations of carrying out flights in mixed urban-forest environment to avoid overflight of private properties and comply with regulations. Forward overlap was set to 80%, side overlap to 50%, and flight speed to 7 m/s. Five returns were recorded per pulse, and the scan mode was non-repetitive. The RTK setup mirrored that used for the MS/RGB imagery flights. The point density was at 8630 points/m² for site A1 and 4116 points/m² for site A2. This LiDAR acquisition was only used for the exploratory tree-segmentation assessment (RQ4). The main reason for the difference in point density is the number of oblique flights (Table 1), which results in more overlapping flight paths and thus variable point density. Furthermore, site A2 had a section with more open swampy areas, meaning less vegetation and therefore further reduced point density.

Regulatory Compliance and Safety Considerations

UAV-flights were conducted in accordance with Civil Aviation Authority (CAA) Part 101 operation rules and in compliance with relevant local regulations. Prior to each flight, operations were logged and approved via the Airshare airspace management platform. This was required, since all sites were located within the control zone (CTR) airspace from Whenuapai Airport (A1-A3) or Hamilton Airport (H1). When operations were conducted within 4km of an aerodrome, approval was gained from all relevant aerdrome operators prior to any UAV flights (Figure 8). For site A1, flight paths were planned to avoid overflight of sensitive areas, including a generous distance to a Horse paddock with animals present and the Glenfield cemetery at Eskdale Reserve (A1).



(a) Auckland Airspace.

(b) Hamilton Airspace.

Figure 8: The maps for Auckland (**a**) and Hamilton (**b**) airspaces. Blue area shows zones within 4km of an aerodrome and red are Control Zones. Black crosses indicate research sites. All airspace boundaries are approximations based on data from the New Zealand Aeronautical Information Package (AIP) issued 25 December 2025. Basemap: Copyright OpenStreetMap contributors under ODbl by OSFM.

Landownership requirements were verified for all sites. As all locations were on council-managed land (Auckland Council for A1–A3, Hamilton City Council for H1), formal consent was not required (Auckland Council, n.d.; Hamilton City Council, n.d.), except for one privately owned property boundary at Bushglen Reserve (A3), for which verbal consent was obtained. A briefed observer was present during all flight operations to maintain situational awareness and ensure safe operations in these urban environments.

Scope and justification of the LiDAR component

The LiDAR work in this thesis is presented as a targeted, exploratory component intended to assess whether 3D tree-segmentation outputs could potentially complement the raster-based deep learning workflow. Due to operational constraints (different oblique flight counts and resulting point-density variation), time limitations and initial segmentation quality, the LiDAR analysis was scoped to a pilot evaluation and qualitative assessment rather than a full quantitative integration with the U-Net experiments. Hence, the scope of the LiDAR component and its results are exploratory: useful for identifying promising directions, but not used as a primary source for the main comparative model evaluations. Recommendations for deeper, quantitative follow-up work are given in the discussion.

3.1.4 Reference Data Collection

Ground-truthing of *S. maire* on sites A2 and A3 was carried out on 10 May and 14 June 2025 for site A1. *S. maire* locations were shown by volunteers from the local conservation group. To ensure exact positioning, the Emlid Reach RS3 RTK receiver (Emlid, n.d.) was used and connected to the closest PositioNZ-RT base station. The receiver was then connected to QField (OPENGIS.ch). In the field, the receiver was positioned next to the trunk. Locations were only stored once the receiver had a confirmed accuracy of <30 cm. When travelling between sites, footwear was changed to reduce risk to spread *A. psidii* spores. Site H1 was not ground-truthed with the receiver. Since all trees, except two, were directly adjacent to the path, their rough locations were noted and later identified on the orthomosaic.

3.2 Data Processing and Model Development

All data processing and analysis was performed using a combination of Python (v3.13.2), Agisoft Metashape (Agisoft LLC, v2.2.1, 2025), DJI Terra (DJI, v4.13.3.0, n.d.) and CloudCompare (CloudCompare, v2.14.alpha, 2022). Computations were carried out on a workstation with a 13th Gen Intel® Core™ i7-13700 CPU, NVIDIA GeForce RTX 4070 Ti GPU, and 128 GB RAM running Windows 11. The following sections describe the data preprocessing, training and inference workflow in detail. The source code of this thesis is available on www.github.com/pfaffrob/MSc_Thesis_RP.

3.2.1 Orthomosaic

RGB and MS orthomosaics were generated from raw imagery in Agisoft Metashape, following workflows proposed by Geospatial Tips (2022), with some modifications for our data requirements. This process involved first calibrating reflectance and aligning of images, followed by iterative tiepoint filtering based on reconstruction uncertainty (<27), reprojection error (<0.83), and projection accuracy (>3.8), with alignment optimisation after each step. A dense point cloud was constructed and filtered by confidence (>2), which was then used to build a 2.5D surface (a point cloud which is triangulated into a mesh without overhang). The orthomosaics were exported at 2.5 cm/pixel resolution for MS imagery and 1.5 cm/pixel for RGB imagery, with blending mode disabled (Figure 9). The blending mode was disabled to ensure that no smoothing or interpolation occurred during the orthomosaic generation process, and only raw image values were assigned to each pixel.

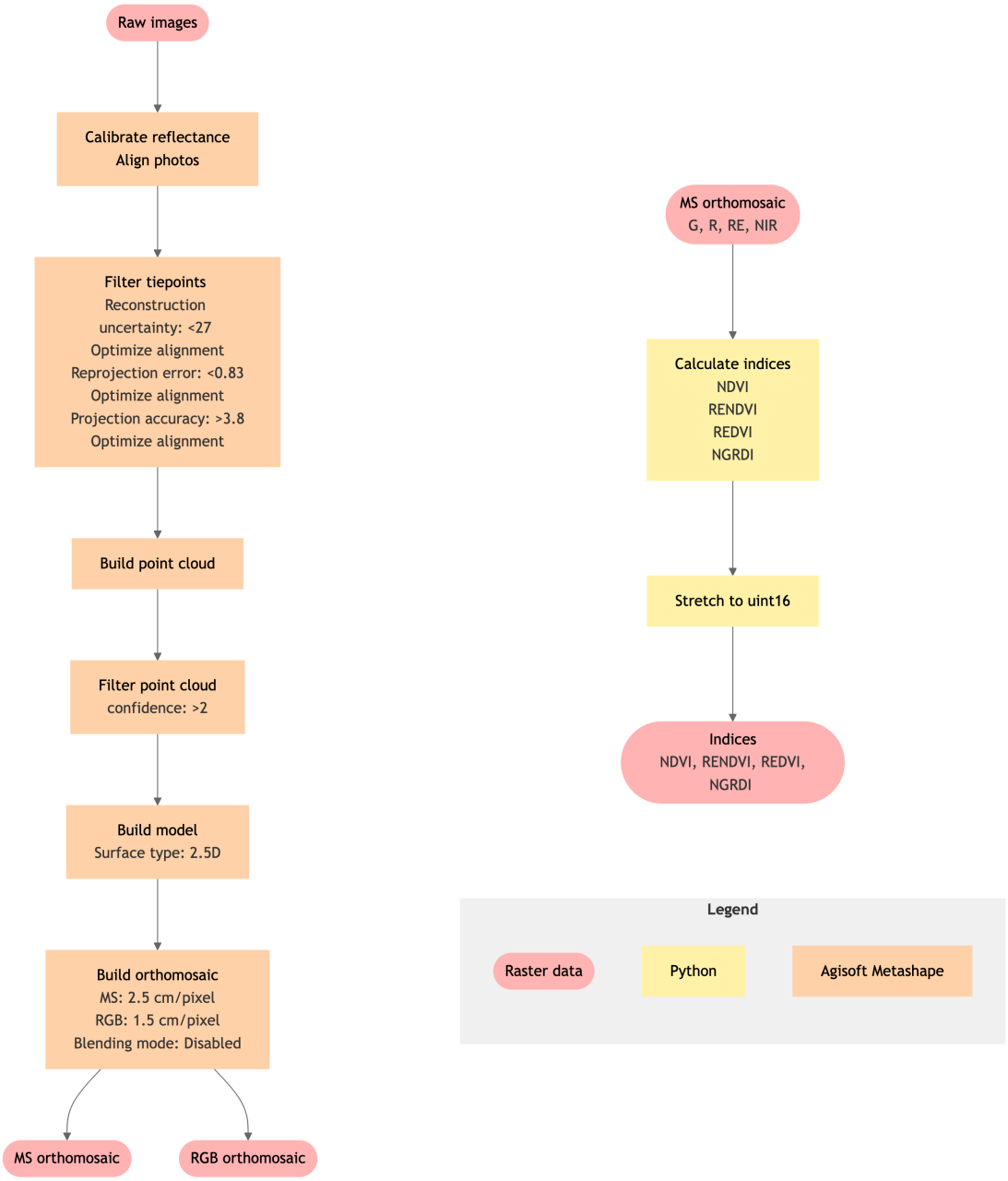


Figure 9: Workflow to generate the orthomosaics and further processing to prepare for model input.

To enhance spectral discrimination, four normalised VIs were computed from the four-channel MS orthomosaics (Table 2) and stretched from float32 to uint16 format (-1–1 to 0–65535), which is a common approach to normalise data for deep learning (Wagner et al., 2019). NDVI was calculated as a fundamental index sensitive to vegetation health and chlorophyll content. RENDVI utilises both RE and NIR channels to detect vegetation stress and subtle canopy changes (Sims & Gamon, 2002). REDVI applies a similar normalised difference approach but substitutes the RE channel for NIR, offering an alternative canopy response measurement (Briechle et al.,

2020). NGRDI, derived from G and R channels, provides vegetation signal independent of NIR. These indices were selected because they have shown potential for species identification where spectral signatures are needed due to limited morphological features (Barrero & Perdomo, 2018; Briechle et al., 2020).

Table 2: The formulas for the VIs used in this thesis.

Index Formula	Source
NDVI $= \frac{NIR - R}{NIR + R}$	Rouse Jr et al. (1974)
RENDVI $= \frac{NIR - RE}{NIR + RE}$	Gitelson & Merzlyak (1994)
REDVI $= \frac{RE - R}{RE + R}$	Briechle et al. (2020)
NGRDI $= \frac{G - R}{G + R}$	Tucker (1979)

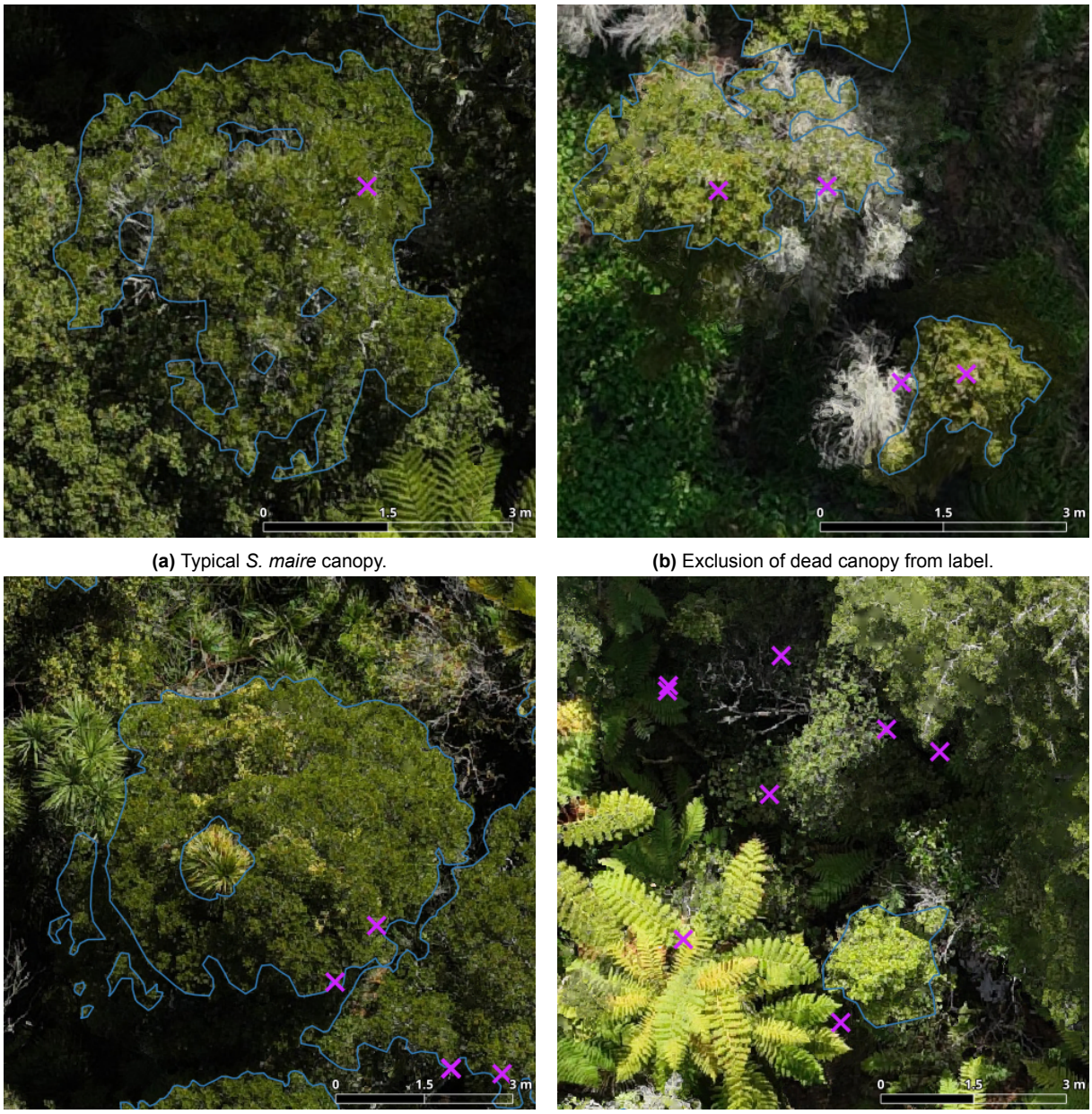
Due to saturation of the R and G bands on the CRP photo (Figure S3), MS imagery and VIs were processed in two calibration schemes: absolute (using the sun sensor and CRP) and relative (using only the sun sensor; Table 3). Absolute calibration adjusts for exposure differences between sites, whilst relative calibration was retained to evaluate whether the burned CRP pixels (Figure S3) used for radiometric calibration influenced model performance. A burned pixel means that it is assigned the maximum value (65535 in uint16 format), which means that the pixel value does not represent the true reflectance of the surface, but rather the maximum value that the sensor can record with the specified exposure settings (Figure S4).

3.2.2 U-Net

Annotation

Ground-truthed *S. maire* locations were overlaid on the orthomosaics. The annotator first examined unambiguous *S. maire* individuals to establish their characteristic visual appearance (Figure 10a). Using this visual knowledge base, the visible canopy of each *S. maire* instance was digitised, which included more complex cases such as where visible canopy was offset from the ground-truth point (e.g., due to trunk angle; Figure 10d) or growth of other vegetation within canopy (Figure 10c). Given the limited training samples, dead foliage was excluded (Figure 10b) to keep the training data as uniform as possible and reduce within-class variability. Furthermore, *S. maire* present at ground-truth points but with no visible canopy were not annotated (Figure 10d).

Where canopy crowns were segmented by overlapping vegetation or shadowed/poorly lighted, only areas clearly identifiable as *S. maire* were digitised, which could result in fragmented annotation polygons (Figure 10c). To account for processing differences and geometric offsets between sensors, RGB and MS imagery were annotated independently. Following digitisation, annotation polygons were rasterised to match orthomosaic resolution.



(a) Typical *S. maire* canopy.

(b) Exclusion of dead canopy from label.

(c) Fragmented label and growth ingrowth of other plant, presumably kiekie.

(d) Canopy offset from the ground-truthed position (bottom right) and overgrowth of other vegetation.

Figure 10: Examples of *S. maire* canopy annotation: (**a**) typical crown, (**b**) canopy offset from the ground-truth position with adjacent overgrowth, (**c**) dead canopy excluded from annotation, and (**d**) fragmented annotation where other vegetation grows within the crown. Annotation polygons are shown in blue and ground-truth points as in pink.

3.2.3 Dataset Preparation

In deep learning, the optimisation of model weights is based on the training data, while validation data is used during training to monitor model performance on independent data and prevent overfitting. Test data, in contrast, is used only for final performance evaluation, ensuring an unbiased assessment of how the model performs on unseen data (Goodfellow et al., 2016). Therefore, to evaluate model performance, training, validation, and test data must be separated carefully. This subset is also required because the U-Net model cannot process these full-resolution orthomosaics directly due to memory and computational constraints, and must therefore be further divided into smaller chunks. In geospatial applications, the partitioning is critical with tiling frameworks where adjacent image tiles overlap and share boundary pixels. If training and test data were derived from overlapping tiles within the same area, performance metrics would be artificially inflated because the model would have effectively "seen" the test data during training (Kattenborn et al., 2022).

To avoid this bias, orthomosaics were cropped into two spatially separated areas with zero overlap: a training zone and a test zone (Table 1; Figure 6). Following the tiling framework by Chen et al. (2024), train and test zones were clipped into 576x576 pixel tiles with 25% overlap. The tiles of the training zone were then further separated into a 80/20 train/validation split (Kattenborn et al., 2019). The selection was not fully random to ensure similar proportion of *S. maire* pixels in both subsets. The resulting dataset exhibited severe class imbalance, with *S. maire* representing 3.4±0.7% (mean±SE) of pixels in training tiles and 4.9 ± 0.5% in validation tiles (Table S1). Datasets with seven band combinations were created to assess their relative performance (Table 3).

Table 3: Band combinations used for training with mirroring relative and absolute combination for MS bands.

Band Type	Bands	Description
RGB	red-green-blue (RGB)	Standard RGB channels
MS_{rel}	R, G, RE, NIR	MS with relative calibration
MS_{abs}	R, G, RE, NIR	MS with absolute calibration
IND_{rel}	NDVI, NGRDI, REDVI, RENDVI	Vegetation indices
IND_{abs}	NDVI, NGRDI, REDVI, RENDVI	Vegetation indices
MS+IND_{rel}	R, G, RE, NIR, RENDVI	MS_{rel} + $RENDVI_{rel}$
MS+IND_{abs}	R, G, RE, NIR, RENDVI	MS_{abs} + $RENDVI_{abs}$

3.2.4 Data Augmentation

To improve model generalisation and mitigate the drawbacks of limited training data, the dataset was artificially expanded by augmenting the image tiles using Albumentations (Buslaev et

al., 2020). Augmentations included horizontal and vertical flips (probability 0.5), geometric transformations via ShiftScaleRotate with ±10% translation, ±15% scale, and ±45° rotation (probability 0.5), grid distortion to handle terrain and canopy variations (probability 0.2), and multiplicative noise to simulate illumination changes (multipliers 0.95–1.05, probability 0.2). This approach applied only n-channel-safe augmentations whilst creating substantial variation in training samples.

3.2.5 Model Architecture

The U-Net CNN architecture was selected for semantic segmentation based on its established efficacy in remote sensing applications for vegetation detection (Abreu-Dias et al., 2025; Freudenberg et al., 2019; Kattenborn et al., 2019; Lobo Torres et al., 2020; Shahi et al., 2023). The model follows an encoder-decoder structure with skip connections (Ronneberger et al., 2015). In the contracting path, two 3×3 convolutions followed by a rectified linear unit (ReLU) activation are applied, then a 2×2 max pooling operation with stride 2 downsamples by halving spatial resolution whilst doubling feature channels. This pattern repeats through four encoder blocks. The expanding path mirrors this: a 2×2 transposed convolution upsamples the feature map and halves the channels, and is merged with its corresponding feature map from the contracting path through skip connections that preserve spatial detail. Two 3×3 convolutions and ReLU activations follow, continuously reconstructing spatial resolution. A final 1×1 convolution maps each feature vector to output classes, producing the segmentation map. The architecture is displayed in Figure 11.

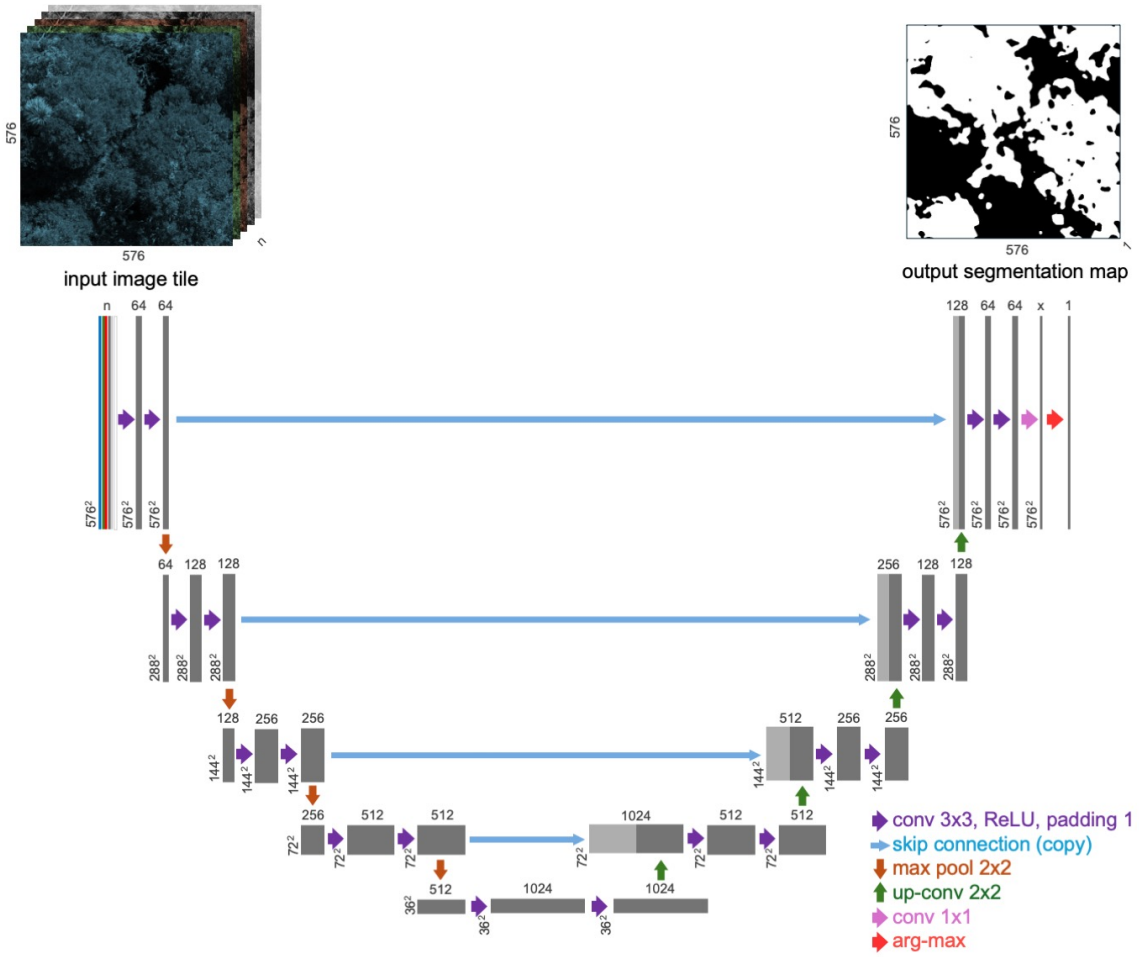


Figure 11: U-Net architecture which accepts n-channels input image of 576×576 pixels. Each grey box represents a multichannel feature map. Channel numbers are shown on top of each box, with spatial resolution indicated at the lower left. Light grey boxes display feature map copies, and arrows represent the various operations.

The implementation was adjusted by adding zero-padding to the 3×3 conv/ReLU blocks to retain input-to-output spatial resolution, eliminating the cropping required in the original design and simplifying inference post-processing. Whilst zero-padding could introduce artefacts at tile boundaries (Huang et al., 2019), this was not observed with the applied tiling approach. Therefore, other options, such as using a buffer for predictions which is clipped post-processing (Ball et al., 2023; Freudenberg et al., 2019), were not further explored. The model was implemented in PyTorch (Paszke et al., 2019) and core functionalities used from Pfaff (unpublished).

3.2.6 Loss Functions

Selection of a suitable loss function is critical for training a deep learning model (Azad et al., 2023). The loss function calculates the error between the annotation mask and the predicted output, thereby guiding the optimisation of network weights during training. Various loss functions

are used in binary semantic segmentation to address different challenges, such as the class imbalanced present in this study (target class representing only 3.4±0.7% (mean±SE) of pixels; Table S1). Such a class imbalance has been identified as a crucial factor influencing deep learning model performance (Ghosh et al., 2024). Selecting a loss function which applies class weighting is a common mitigation strategy (Bourday et al., 2024). Given the importance of this selection, three loss functions were evaluated: L_{WBCE} with class weights of 1, 10, and 50; unweighted L_{Dice} loss; and L_c, a combination of the two, which has shown good results in imbalanced binary segmentation tasks (Koishiyeva et al., 2025).

In the following formulations, $y^{(i)}$ denotes the ground-truth label for pixel i (0 = background, 1 = target class). The model's predicted probability for that pixel is represented by $\hat{y}^{(i)}$. Additionally, N indicates the total number of pixels, and β is the weight assigned to the target class.

L_{WBCE} adjusts the standard binary cross-entropy loss by applying a target class-specific weight (β), increasing the penalty for misclassifying minority class pixels (Li et al., 2022):

$$L_{WBCE} = -\sum_{i=0}^{N} \beta\, y^{(i)} \log \hat{y}^{(i)} + \left(1 - y^{(i)}\right) \log \left(1 - \hat{y}^{(i)}\right) \tag{1}$$

L_{Dice} is a region-based overlap optimisation loss function that is less sensitive to class imbalance, as it directly optimises the overlap between predicted and ground-truthed regions:

$$L_{Dice} = 1 - \frac{2\, y^{(i)} \hat{y}^{(i)}}{y^{(i)} + \hat{y}^{(i)}} \tag{2}$$

The combined L_c loss function balances pixel-wise accuracy with region overlap optimisation (Li et al., 2022):

$$L_C = L_{WBCE} + L_{Dice} \tag{3}$$

3.2.7 Training Configuration

Separate models were trained on each site independently (single-site) and on all sites combined (multi-site). The main limitation of site-specific models is that they cannot be deployed to new locations. This means that either a new model must be trained for each location (requiring time-expensive ground-truthing and re-annotation, defeating the purpose of a method aimed at identification of unknown *S. maire* locations), or a single model must be developed that

generalises across sites. Therefore, a single model was trained on combined data from all four sites to assess whether comparable or better performance could be achieved relative to site-specific models (RQ2).

For model training, stochastic gradient descent optimisation with initial learning rates (LRs) of 0.02 or 5×10^{-5} were used. A ReduceLROnPlateau scheduler monitored target class intersection over union (IoU) on the validation split, reducing the LR by a factor of 0.5 after 15 epochs without improvement (Srivastava et al., 2024). Training ran for a maximum of 300 epochs with batch size of 2 and early stopping after 20 epochs without improvement in training loss. The best-performing model (by highest F1 on the validation dataset) was used for subsequent inference.

For single-site models, all combinations of LRs (0.02, 5×10^{-5}), class weights (1, 10, 50), loss functions (L_{Dice}, L_{WBCE}, L_c), and band combinations (MS_{rel}, RGB) were evaluated individually on each reserve. For the remaining band combinations, base hyperparameters were used (weight=10, LR=0.02, loss=L_c), resulting in 120 single-site models.

For multi-site models trained on all available training data, LR was set to 0.02. Three hyperparameter combinations were tested across five band combinations (MS_{abs}, MS_{rel}, MS_{abs} + RENDVI, IND_{abs}, RGB), with unweighted L_{Dice}, L_c with weights 10 and 50, resulting in 15 multi-site models.

3.3 Evaluation and Analysis

Predictions were generated by applying identical preprocessing steps as applied to the training zone (Figure 6) and running the trained model in inference mode. Predicted tiles were stitched back into a single orthomosaic, combining overlapping regions. In the stitching process, any pixel predicted as target class was retained as target class in the final output.

Model performance was evaluated on the spatially separated test zone (Figure 6) to ensure metrics reflected unseen data. The following evaluation metrics were calculated exclusively for the target class. Precision, which measures the proportion of correctly predicted target pixels among all pixels predicted as target:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

where TP is true positives and FP is false positives. Recall measures the proportion of correctly predicted target pixels among all actual target pixels:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

where FN is false negatives. The primary aggregating metric was F1, the harmonic mean of precision and recall:

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

To assess whether additional *S. maire* canopy area improves model performance, a Pearson correlation test was performed between annotated area of the training zones and F1 scores. The hypotheses were defined as H_0: increasing available *S. maire* canopy area does not improve model performance (no linear relationship), and H_1: increasing available *S. maire* canopy area improves model performance (positive linear relationship). A simple linear regression was also fitted to quantify the relationship.

3.4 LiDAR Tree Segmentation

Two methods of LiDAR based tree segmentation were trialled to determine whether they could supplement the raster based FCNN workflow: TreeLearn, which is deep learning-based (Henrich et al., 2024), and Treeiso, which is graph-based (Xi & Hopkinson, 2022). Due to time constraints and preliminary assessment of segmentation results indicating challenges with segmentation accuracy (Figure 16), these approaches were only assessed qualitatively to determine whether they are worth exploring further in future work. This exploratory LiDAR work (RQ4) is therefore distinct from the main raster-based experiments. The following describes the processing steps undertaken, and the workflow is shown in Figure 12.

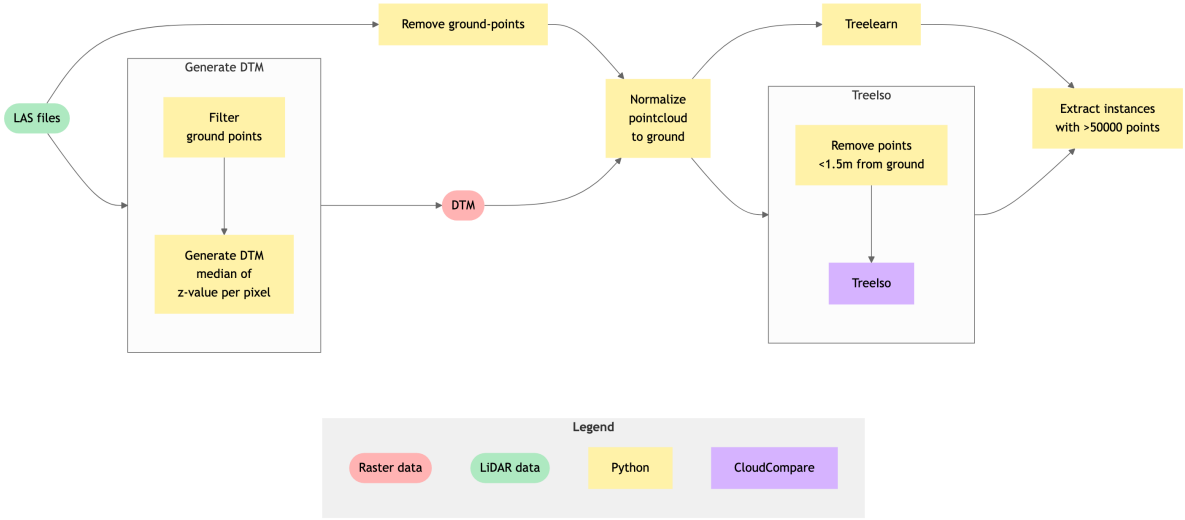


Figure 12: Workflow to test the LiDAR tree segmentation methods (TreeLearn and Treeiso). This included point cloud processing, including the generation of a DTM, and application of the two segmentation methods.

3.4.1 Point Cloud Processing

Raw LiDAR data was processed to LAS files using DJI Terra. A 10cm digital terrain model (DTM) was generated from the point cloud to enable height normalisation. Point clouds were filtered to non-ground points and then normalised to ground level by subtracting the DTM elevation from each point's z-coordinate. This ensured the height AGL of the represented vegetation structure rather than terrain elevation. For Treeiso segmentation, points lower than 1.5 m above the ground were removed to reduce noise from dense understory vegetation (Xi & Hopkinson, 2022).

3.4.2 Segmentation Methods

Both TreeLearn and Treeiso algorithms were applied to segment individual tree crowns from the normalised point clouds. TreeLearn voxelises the point cloud by dividing the 3D space into small cubes (voxels) and assigning points to these cubes, then applying a deep learning model to classify each voxel as belonging to a specific tree instance, effectively segmenting individual tree crowns based on learned spatial features and patterns in the point cloud data (Henrich et al., 2024). In contrast, Treeiso employs a graph-based approach, treating the point cloud as a network of connected points where each point is a node and connections exist between nearby points. The algorithm identifies clusters or groups of connected points associated with the same tree crown by analysing the structural relationships and connectivity patterns between points (Xi & Hopkinson, 2022). Segmentation outputs were visually assessed for accuracy by comparing segmented tree boundaries with canopy extents visible in orthomosaics (Figure 16).

Chapter 4 Results

This section presents the outcomes of applying deep learning methods to detect *S. maire* using UAV-based imagery and point cloud data. The analysis explored two key approaches: semantic segmentation of RGB and MS imagery, and instance segmentation of LiDAR point clouds (Figure 2). The primary question was whether automated methods can reliably identify individual *S. maire* canopy in dense native forests. This is a task with direct conservation implications, as accurate mapping of rare and threatened species such as *S. maire* is essential for targeted management efforts against myrtle rust.

Detection of *S. maire* was moderately successful using semantic segmentation. Under certain conditions, models demonstrated promising detection (F1 scores of 0.65–0.81 on favourable sites), though performance declined significantly with smaller training datasets (best F1=0.56) or when training across multiple sites simultaneously (best F1=0.51). While LiDAR-based instance segmentation methods showed promise, both methods produced results that were not sufficient for inclusion within the workflows of this study. Below, we present detailed results from the semantic segmentation analysis, including investigation of hyperparameters, followed by a qualitative assessment of why the LiDAR approaches require further development.

4.1 U-Net Semantic Segmentation

4.1.1 Single-site

When trained on data from a single site, U-Net was able to detect *S. maire* canopies with moderate to good performance. However, performance varied considerably between sites, with F1 scores ranging from 0.46 to 0.81. Amount of available training data significantly improved performance (Pearson p=0.017 and R^2=0.996; Figure 13). A variety of model configurations and different band combinations were tested as input to identify the best performing configuration. The following sections present this optimisation work, followed by a section on the best models and how they performed on unseen data.

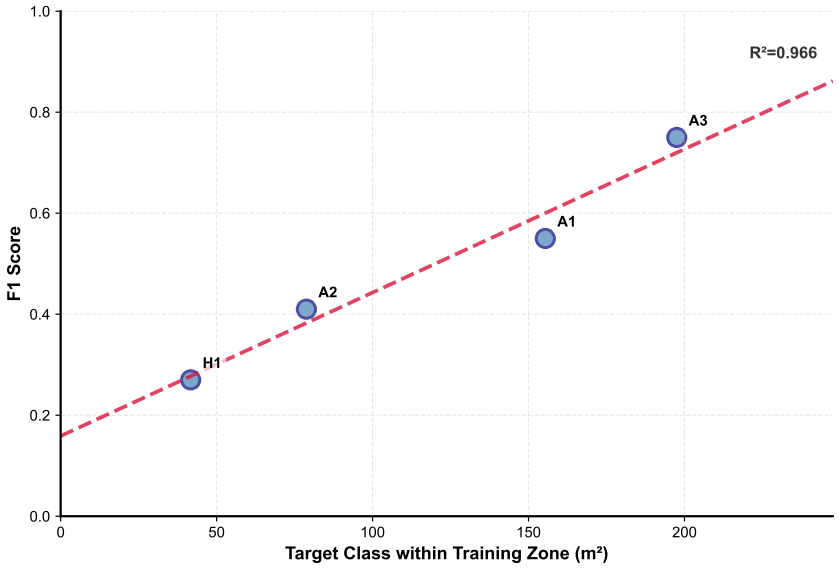


Figure 13: Linear regression of annotated *S. maire* area (m^2) within the training zone vs. F1 score trained on RGB imagery (Table 5) for each site. Each point represents a site, and the regression line is shown in red with R^2=0.966.

Hyperparameter Configuration

To identify which configuration had the greatest impact on detection accuracy, three key hyperparameters were systematically tested: the loss function (which guides how the model learns), the LR (which controls the magnitude of model weight adjustments during training), and class weights (which penalise misclassification of the underrepresented *S. maire* class more heavily). This systematic comparison of hyperparameters (Table 4) revealed differences in model performance.

L_{Dice} achieved the highest mean F1 scores (0.56±0.14 for RGB imagery, 0.54±0.25 for MS_{rel}), outperforming L_c (0.50±0.20 / 0.45±0.26) and L_{WBCE} (0.48±0.27 / 0.46±0.17). LR had the most substantial impact on performance. Models trained with an initial LR=0.02 outperformed those with LR=5×10^{-5}. When using L_{Dice} F1 scores dropped from 0.56±0.14 to 0.13±0.09 (RGB) and from 0.54±0.25 to 0.13±0.11 (MS_{rel}). The lower LR produced random predictions with extremely high recall (up to 0.98) but negligible precision (0.07–0.14). This means that virtually any pixel was classified as *S. maire* (Figure S5). Class weighting improved L_c performance: F1 gained with each incremental step adjusting the weights from 1 to 10 to 50 in both RGB and MS_{rel} (Table 4b). However, the difference from 10 to 50 is marginal when visually inspecting results (Figure S6).

Table 4: Comparison of hyperparameters on *S. maire* detection performance (mean±std), tested on each site separately. Tested hyperparameters included loss function (**a**), class weight (**b**), and LR (**c**).

(a) Results from tested loss functions (L_{WBCE}, L_{Dice} (no weight) and L_c) with weight=10 and LR=0.02.

Bands	Loss	F1	IoU	Precision	Recall
MS_{rel}	L_{WBCE}	0.46±0.17	0.31±0.14	0.58±0.22	0.56±0.36
MS_{rel}	L_{Dice}	0.54±0.25	0.40±0.23	0.65±0.18	0.59±0.35
MS_{rel}	L_c	0.45±0.26	0.31±0.20	0.36±0.14	0.69±0.42
RGB	L_{WBCE}	0.48±0.27	0.34±0.23	0.55±0.18	0.55±0.34
RGB	L_{Dice}	0.56±0.14	0.40±0.14	0.47±0.13	0.76±0.19
RGB	L_c	0.50±0.20	0.35±0.19	0.46±0.21	0.56±0.21

(b) Results from assigning weights 1, 10 and 50 for *S. maire* class using L_c.

Bands	Weight	F1	IoU	Precision	Recall
MS_{rel}	1	0.42±0.34	0.31±0.29	0.55±0.38	0.40±0.39
MS_{rel}	10	0.45±0.26	0.31±0.20	0.36±0.14	0.69±0.42
MS_{rel}	50	0.50±0.17	0.34±0.14	0.42±0.07	0.71±0.36
RGB	1	0.47±0.17	0.32±0.17	0.59±0.23	0.49±0.28
RGB	10	0.50±0.20	0.35±0.19	0.46±0.21	0.56±0.21
RGB	50	0.52±0.14	0.36±0.12	0.49±0.15	0.63±0.24

(c) Results from initial LRs (0.02 and $5x10^{-5}$) in combination with a LR-scheduler with L_c using weight=10.

Bands	Loss	LR	F1	IoU	Precision	Recall
MS_{rel}	L_c	$5x10^{-5}$	0.14±0.11	0.08±0.07	0.08±0.07	0.98±0.01
MS_{rel}	L_c	0.02	0.45±0.26	0.31±0.20	0.36±0.14	0.69±0.42
MS_{rel}	L_{Dice}	$5x10^{-5}$	0.13±0.11	0.07±0.06	0.07±0.07	0.95±0.03
MS_{rel}	L_{Dice}	0.02	0.54±0.25	0.40±0.23	0.65±0.18	0.59±0.35
RGB	L_c	$5x10^{-5}$	0.19±0.10	0.11±0.06	0.14±0.08	0.44±0.30
RGB	L_c	0.02	0.50±0.20	0.35±0.19	0.46±0.21	0.56±0.21
RGB	L_{Dice}	$5x10^{-5}$	0.13±0.09	0.07±0.05	0.10±0.08	0.23±0.13
RGB	L_{Dice}	0.02	0.56±0.14	0.40±0.14	0.47±0.13	0.76±0.19

Band Combinations

Using base hyperparameters (LR=0.02, weight=10, L_c), RGB imagery achieved the highest mean F1 score (0.50) across reserves, though, only marginally better than MS+IND_{rel} (mean F1=0.48). However, this was not consistent across all sites. At sites A1 and H1, MS band combinations outperformed RGB imagery: MS+IND_{rel} achieved a notably high F1 of 0.64 on site A1. Performance varied substantially between the other sites: A3 showed the strongest and most consistent results across all band types (F1=0.60–0.75), whilst site A2 consistently showed poor performance (F1=0.08–0.41). Adding RENDVI to MS bands did not consistently improve performance, except for site A1 where it contributed to the best result of that site (Table 5).

Table 5: F1 scores for tested band combinations (RGB, MS$_{rel}$, IND$_{rel}$, MS+IND$_{rel}$) across reserves with LR=0.02, weight=10, and L_c.

Reserve	RGB	MS$_{rel}$	IND$_{rel}$	MS+IND$_{rel}$
A1	0.55	0.47	0.38	0.64
A2	0.41	0.08	0.18	0.20
A3	0.75	0.66	0.64	0.60
H1	0.27	0.59	0.42	0.46
Mean	**0.50**	**0.45**	**0.40**	**0.48**

The best-performing model, trained on site A3 with MS$_{rel}$ bands and L_{Dice} (weight=1, LR=0.02), achieved an F1 of 0.81 and IoU of 0.68 (Table 6; Table S2, S4, S5, S6). Training metric progression is shown in Figure S7. Site A3 consistently produced the strongest results, with six of the top ten models from this reserve. Performance was more variable on other sites: A1 reached a maximum F1 of 0.65, H1 achieved 0.59, and A2 showed the lowest maximum F1 of 0.46.

Table 6: Top 10 individual single-site models ranked by F1 score. The top 10 models for each site individually can be found in Table S2.

Reserve	Bands	Loss	W	LR	F1	IoU	Prec	Rec
A3	MS$_{rel}$	L_{Dice}	1	0.02	0.81	0.68	0.74	0.88
A3	MS$_{rel}$	L_c	1	0.02	0.80	0.67	0.72	0.89
A3	RGB	L_{WBCE}	10	0.02	0.77	0.62	0.68	0.87
A3	RGB	L_c	10	0.02	0.75	0.60	0.74	0.76
A3	RGB	L_{Dice}	1	0.02	0.74	0.59	0.64	0.87
A3	RGB	L_c	1	0.02	0.72	0.57	0.62	0.86
A3	MS$_{rel}$	L_c	10	0.02	0.66	0.49	0.51	0.91
A1	MS$_{rel}$	L_c	50	0.02	0.65	0.48	0.52	0.87
A3	RGB	L_c	50	0.02	0.65	0.48	0.52	0.86
A1	MS+IND$_{rel}$	L_c	10	0.02	0.64	0.47	0.51	0.87

Visual inspection of the predictions reveals clear differences in model behaviour (Figure 14). Site A3 predictions show high correspondence with ground truth, especially for L_{Dice}, which yielded cleaner edges, whereas L_c models produced frayed edges and additional noise. Site A2 predictions were consistently smaller than ground-truthed area. Site H1 showed the poorest performance, with consistent misclassified patches and additional noise. L_c with higher weights introduced more noise in both RGB and MS imagery predictions compared to L_{Dice}.

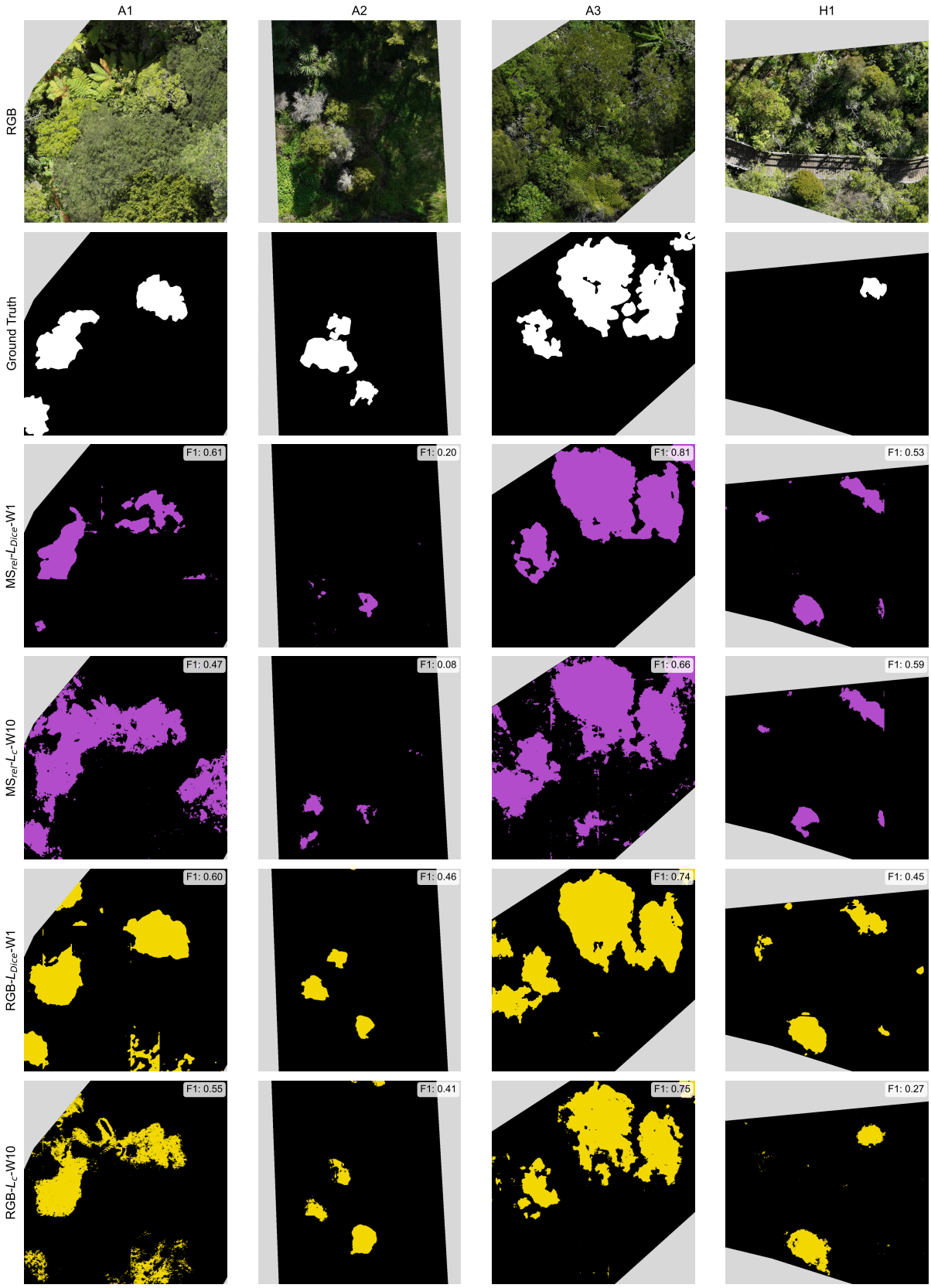


Figure 14: Predictions from the best-performing single-site models for the test zone for each reserve for a 15x15 m extent. Pink overlays indicate predictions on MS imagery, yellow overlays on RGB imagery. Model configurations are listed on the left; all were trained with a LR of 0.02. For example RGB-L_c-W10 reads as: “model trained on RGB bands using L_c with a weight of 10”.

4.1.2 Multi-site

Band and Loss Functions

RGB bands outperformed MS bands in the multi-site setting, with the three best models all using RGB imagery (Table 7). This trend confirms the single-site results, however, the performance gap between RGB and MS imagery is much more pronounced with F1=0.51 (RGB) and F1=0.35 (MS).

Table 7: The performance metrics from the 10 best performing model configurations trained on the multi-site dataset, sorted by F1 score. A table with all models is available in Table S7.

Bands	Loss	Weight	F1	IoU	Prec	Rec
RGB	L_c	50	0.51	0.34	0.39	0.72
RGB	L_{Dice}	1	0.50	0.33	0.45	0.56
RGB	L_c	10	0.45	0.29	0.31	0.81
MS_{rel}	L_c	50	0.35	0.22	0.23	0.75
MS_{rel}	L_c	10	0.30	0.18	0.27	0.34
IND_{abs}	L_c	10	0.22	0.12	0.23	0.20
IND_{abs}	L_c	50	0.19	0.10	0.12	0.52
MS_{rel}	L_{Dice}	1	0.16	0.09	0.44	0.10
MS_{abs}	L_c	50	0.16	0.09	0.17	0.15
MS+IND_{abs}	L_{Dice}	1	0.16	0.09	0.11	0.31

Absolute-calibrated band combinations (MS_{abs}, IND_{abs}, MS+IND_{abs}) showed poor generalisation in multi-site training, with F1 scores ≤0.22 and subsequently also poor visual results (Figure 15). L_{Dice} on MS imagery showed low noise but exhibited substantial under-segmentation and misclassification, where both sites A1 (F1=0.00) and A3 (F1=0.07) were subject to these issues. This was also visible in training metrics progression, where IoU plateaued after ~30 epochs (Figure S8).

Best-performing Multi-site Models

Despite systematic optimisation, multi-site models achieved substantially lower performance than single-site models. The best multi-site configuration (RGB imagery, L_c, weight=50) achieved an F1 of 0.51 and IoU of 0.34, representing a 37% reduction in F1 compared to the best single-site model. Only the RGB imagery model with L_c achieved performance comparable to some single-site models (F1=0.54). Across all configurations, multi-site predictions were notably noisier than their single-site counterparts (Figure 15), indicating that the model struggled to learn consistent *S. maire* characteristics across different forest structures and lighting conditions. Notably, visually one of the best results was on a site (A2) that showed poor results in the single-site setting, identifying *S. maire* accurately (RGB, L_c, weight=50; F1=0.54; Figure 15).

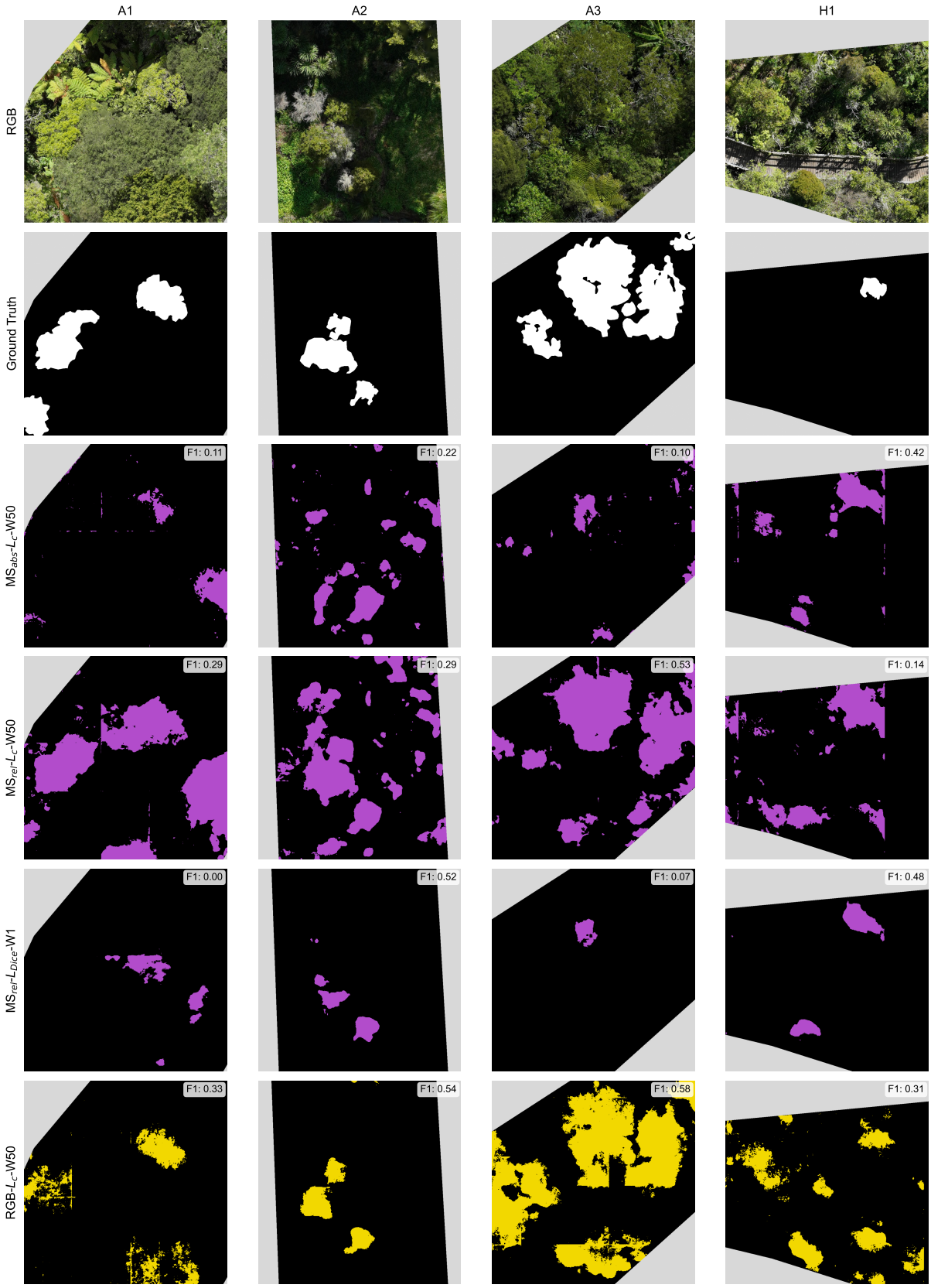
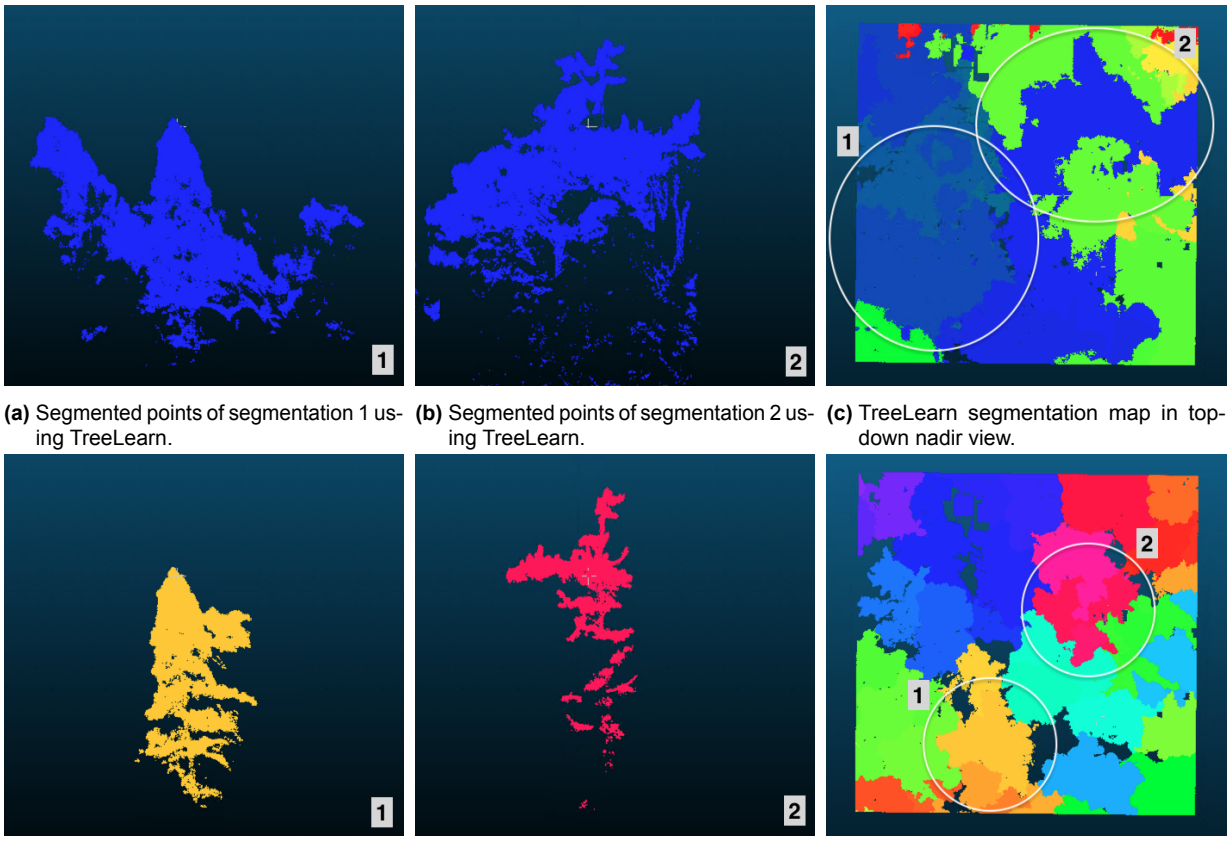


Figure 15: Predictions from the best-performing multi-site models for the test zone for each reserve for a 15x15 m extent. Pink overlays indicate predictions on MS imagery, yellow overlays on RGB imagery. Model configurations are listed on the left; all models were trained with a LR of 0.02. For example RGB-L_c-W50 reads as: "model trained on RGB bands using L_c with a weight of 50".

4.2 LiDAR Tree Segmentation

Two instance segmentation algorithms (TreeLearn and Treeiso) were tested to isolate individual tree point clouds from the forest point cloud. However, qualitative assessment of the two trialled segmentation methods revealed significant challenges with both approaches.

TreeLearn did not successfully segment individual trees and produced messy, near random looking segmentations (Figure 16 a-c). Treeiso performed better, producing recognisable tree segments, but segmentations frequently included portions of adjacent tree crowns, limiting their quality for species-specific classification (Figure 16 d-f). Given these limitations and the more promising results from MS imagery analysis, further development of LiDAR-based segmentation was not pursued within the timeframe of this thesis.



(a) Segmented points of segmentation 1 using TreeLearn.

(b) Segmented points of segmentation 2 using TreeLearn.

(c) TreeLearn segmentation map in top-down nadir view.

(d) Segmented points of segmentation 1 using Treeiso.

(e) Segmented points of segmentation 2 using Treeiso.

(f) Treeiso segmentation map in top-down nadir view.

Figure 16: Comparison between extracted segmentation groups using TreeLearn (**a**–**c**) and Treeiso (**d**–**f**).

Chapter 5 Discussion

A. psidii (which causes myrtle rust) has spread globally over the past two decades with dramatic consequences for Myrtaceae, in severe cases even leading to local extinctions (Fensham et al., 2020; Fensham & Radford-Smith, 2021). Since the pathogen's arrival in 2017 in Aotearoa, concerns have emerged for native Myrtaceae (Jo et al., 2022; Sutherland et al., 2020; Toome-Heller et al., 2020). Accurate mapping, especially of highly susceptible species like *S. maire*, is critical for conservation efforts. However, the rarity of *S. maire* can be one of the main limiting factors in research topics evolving around this species (McCarthy et al., 2024). Advances in UAV technology and deep learning architectures have enabled forest inventory monitoring globally (Abreu-Dias et al., 2025; Osco et al., 2021; Wołk & Tatara, 2024). This thesis applied these technologies in a case study to map *S. maire*, a rare native tree threatened by myrtle rust, in dense and complex native forests based on high-resolution UAV imagery and LiDAR point clouds.

Detection of *S. maire* was attempted using a 2D deep learning approach with U-Net semantic segmentation. Model performance varied significantly from site to site (F1 ranging from 0.46–0.81), and results did not transfer well to multi-site detection (best F1=0.51). Regarding spectral and hyperparameter configurations, RGB imagery at higher spatial resolution (1.5 cm) produced marginally better results (F1=0.50) than lower resolution MS imagery (2.5 cm; mean F1=0.45–0.48). Hyperparameters configuration had substantial impact on model performance: L_{Dice} and a LR of 0.02 in combination with adaptive scheduling generally led to best results. For the point cloud approach, the TreeLearn algorithm was unsuccessful, whilst the Treeiso segmentation method produced recognisable tree segments and showed promise for improving the deep learning-based *S. maire* detection workflow (Figure 16). These findings indicate that UAV-based deep learning methods can detect *S. maire* at individual sites under optimal conditions, but additional research, particularly in acquiring training data and addressing radiometric calibration issues, is necessary before such an approach can be reliably applied for conservation purposes.

Supporting this conclusion, analysis of the relationship between training data area and detection performance (Figure 13) demonstrates a strong positive trend: a Pearson correlation test showed a significant linear relationship between available *S. maire* annotation area and F1 scores ($p = 0.017$, $n = 4$), with linear regression explaining 96.6% of the variance in performance across sites ($R^2 = 0.966$). While the sample size of $n = 4$ was too small to meet conventional statistical robustness thresholds, the consistency of this trend across sites is strong evidence

that acquisition of additional training data would substantially improve detection capabilities. Future research expanding to additional *S. maire* populations should therefore directly address this identified bottleneck, enabling the development of more generalisable detection models suitable for operational conservation applications.

In the following sections, data availability and collection, and processing constraints are discussed in detail, followed by the methodological approach and hyperparameter configurations, spectral considerations, and challenges for model generalisation across sites. Comparison of performance to other studies embeds the results within existing literature, whilst broader implications for species monitoring and conservation are explored. Limitations and recommendations for future research are discussed throughout these sections.

5.1 Data Collection and Processing

5.1.1 Temporal Considerations

Initial data collection was planned for April, but due to weather constraints and temporal limitations of this thesis, the MS imagery was collected in September. April was originally chosen because the maximum symptom occurrence for MR is between March and April (Beresford et al., 2019), but had to be redone because of uncontinuous lighting conditions leading to artefacts in the generated orthomosaic (Figure S2). While the scope of this thesis did not investigate disease detection directly, it is likely that this would be more successful if data were collected between March and April as intended.

Outside of disease detection, the peak growing season for *S. maire* is between December and February. This is when flowering rates are highest. This means data collection at this time could enhance detection performance through more distinctive spectral signatures of new growth tissue or phenological traits. However, since flowering is limited (e.g. tree is not fully covered in flowers) and berries are small, leveraging these phenological traits would be challenging at practical flight heights. Furthermore, Robbins et al. (2025) observed that data collection at higher latitudes requires timing closer to summer solstice for optimal results due to lower sun angles and reduced vegetation activity. Future investigations into species detection should therefore prioritise data acquisition during peak growing season (December–February) to maximise spectral differentiation between species.

5.1.2 Sensor and Processing Constraints

The DJI M3M is a relatively low-cost MS UAV platform that is promoted as a practical option for aerial surveying, crop growth monitoring and natural resource surveys (DJI, n.d.-d). The sensor

collects RGB imagery plus four MS bands (G, R, RE, and NIR), bands that have proven useful for vegetation mapping and species detection in deep learning studies using other platforms and sensors (Barreto et al., 2023; Ulku et al., 2022; T. Zhang et al., 2021). However, especially in myrtle rust mapping, the blue band and indices derived from it, have been identified as an important factor to identify the disease (Heim et al., 2019). Nonetheless, the low cost and ease of use renders it a viable option for the detection of rare and threatened species, particularly when budgets are as limited as it is the case for conservation efforts of species like *S. maire*, that are often managed by budget-constrained community restoration groups.

Accurate radiometric calibration of MS imagery is essential for cross-site comparability (Olsson et al., 2021; Swaminathan et al., 2024). This thesis faced several limitations regarding these calibrations. Firstly, Agisoft Metashape does not officially support the DJI M3M for reflectance calibration, although it supports the identical DJI Phantom 4 MS UAV calibration workflow (Agisoft LLC, 2025). Furthermore, CRP saturation in R and G bands (Figure S3) compromised absolute calibration and thus likely contributing to poor multi-site generalisation.

Best practices to mitigate calibration issues include thermal stabilisation of the sensor prior to data collection, CRP capture at mission altitude, larger panels for adequate pixel coverage, and usage of multiple CRPs with different reflectances below 20% to prevent saturation (Olsson et al., 2021). The M3M used in this thesis permitted only automatic exposure, resulting in variable exposure settings for individual images (Figure S4), i.e., preventing the use of consistent gain and shutter speed recommended for reproducible results (Swaminathan et al., 2024). The M3M's absence from deep learning species detection literature (Abreu-Dias et al., 2025) likely reflects these calibration challenges in non-research grade UAVs.

For multi-site generalisation, the model trained on relatively calibrated MS images was presumably limited by not correcting for different exposure factors, whereas absolute calibration failed due to CRP saturation from overexposed images (Figure S3). Therefore, future research should carefully select MS imaging sensors and prioritise those with official software support. Calibration panels should be selected to avoid saturation issues, and the best practices for radiometric calibration mentioned above should be followed where possible. Additionally, normalisation techniques such as histogram stretching (0.01–99.9%) applied to RGB imagery by Schiefer et al. (2020) may improve cross-site generalisation.

Finally, artefacts in both the processed RGB and MS orthomosaics were observed, presumably caused by wind-induced canopy movement between the capture of subsequent images (Figure S9). These artefacts were not present in the raw images, indicating they were introduced

during the orthomosaic generation process. Such artefacts distort canopy features and can therefore negatively impact model performance due to misrepresentation. Future research should therefore consider adopting a methodology in which training is based on raw images rather than orthomosaics, to avoid such artefacts and thus improve model performance. This approach has been applied previously for species detection in environmental monitoring (James & Bradshaw, 2020). However, the annotation workflow would be considerably more time-consuming than working with orthomosaics. Additionally, such an approach would require more complex post-processing to georeference detection results correctly. A notable advantage lies in the expanded training dataset, where multiple images of the same tree from different viewing angles could be taken. Therefore, as long as provided orthomosaic software cannot overcome these artefacts in high-resolution imagery, training on raw images may offer better detection performance at the cost of complicated results processing.

5.1.3 Reference Data Considerations

Site reconnaissance relied on iNaturalist records and local expert knowledge to identify candidate *S. maire* populations. Local experts were consulted and a selection of the most promising sites between Auckland, Hamilton and Tauranga was checked, with only four sites deemed suitable (Section 3.1.1). Therefore, acquiring a larger dataset was not possible, which is a challenge that may also persist in a large-scale study, due to the rareness of the species. Thus, the most prominent issue for a future study lies in identifying and accessing more sites with larger *S. maire* populations. However, finding substantially larger populations may prove challenging given the rarity and scattered distribution of *S. maire*, and if population numbers on other sites prove equally limited, the logistical burden of conducting UAV surveys across numerous sites may become impractical to manage alongside other conservation priorities.

Whilst iNaturalist observations were valuable for identifying populations, their spatial precision was unsuitable for direct use for annotating, and the records were incomplete. Collecting reliable reference data required comprehensive field verification across sites. Ground-truthing then provided GNSS positions of individual trees. Although manual annotation is usually a bottleneck, it was manageable given the limited dataset size. However, if a scaled approach were attempted, this might become a challenge. Annotations were generated through manual digitisation of *S. maire* canopy from orthomosaics guided by these field-collected positions. Spatial offsets between RGB and MS orthomosaics required separate annotation of each orthomosaic. This introduced additional complexity and prevented the creation of a combined dataset. The accuracy

and confidence in these annotations would be improved with the inclusion of additional people to validate the reference data (Kattenborn et al., 2019).

5.1.4 Site Limitations

Time and resource constraints meant this thesis was restricted to four research sites. This limited development of models generalising across the diverse conditions where *S. maire* occurs. Site-specific characteristics substantially affected performance. For example: sites A1 and A3 were expected to yield similar results given proximity and comparable dataset size, yet A3 outperformed A1. This suggests that site-specific variations like vegetation composition or foliage health may have impacted results. It is assumed that lower performance at sites A2 and H1 can be traced to extremely limited training data (6–9 images compared to 21–26 at better-performing sites), or potentially a false annotation at site H1, which was consistently predicted as *S. maire*. Even with multi-site training, generalisation proved problematic. Therefore, a more extensive study should also consider including sites containing *S. maire* populations within natural replanted sites, different urban parks, and old-growth swamp sites containing remnant trees to account for site-specific differences.

5.2 U-Net Implementation

Given the varying performance achieved with the standard U-Net architecture, it is worth considering whether more advanced architectures could have led to better results. While further developed U-Net architectures are available, including variants with attention mechanisms, irregular segmentation capabilities, and enhanced feature extraction (Chen et al., 2024; T. Zhang et al., 2021; T. Zhang et al., 2022), the performance gains are marginal compared to the standard U-Net model tested here. Furthermore, this adds considerable complexity and computational processing requirements. For example, Zhang et al. (2021) developed an irregular segmentation U-Net (Ir-UNet) for wheat yellow rust detection that incorporated additional convolutional layers and attention modules, achieving modest improvements over standard U-Net. Considering the limited training data in the present study, the standard architecture was deemed most appropriate for this exploratory application, prioritising model simplicity and training efficiency over marginal performance gains.

5.3 Hyperparameter Optimisation

5.3.1 Loss Function

Three loss functions were evaluated: L_{WBCE}, L_{Dice}, and their combination: L_c. L_{WBCE} computes loss by summing pixel-wise classification errors while applying different weights to account for class imbalance. L_{Dice} directly optimises IoU by measuring overlap between predicted and ground truth regions as a whole. Results partially confirm findings by Li et al. (2022), where L_{Dice} outperformed L_{WBCE}. However, contrasting their findings, L_{Dice} outperformed L_c, suggesting that spectral information, which per-pixel loss functions like L_{WBCE} leverage, is not significant for *S. maire* discrimination. This is underlined by L_c introducing more misclassifications on MS imagery than RGB imagery (Figure 14). In addition, implementation of a weight map-based loss function assigning reduced penalties to annotation pixels near annotation edges could reduce the implications of frayed boundaries of vegetation (James & Bradshaw, 2020). Loss function comparisons are scarce in remote sensing deep learning studies, highlighting the need for further research evaluating optimal loss functions for semantic segmentation in this field.

5.3.2 Learning Rate Configuration

LR exerts substantial influence on model performance (Wu et al., 2019). Two configurations were evaluated: conservative (5×10^{-5}) and aggressive (0.02) with ReduceLROnPlateau scheduling. Other remote sensing U-Net studies typically employ a LR of 5×10^{-4} (Kattenborn et al., 2019; Lake et al., 2022; Lobo Torres et al., 2020; Schiefer et al., 2020; Ulku et al., 2022). Only Freudenberg et al. (2019) mentioned a comparable approach with LR=3×10^{-5} reduced to 5×10^{-6} during training. The results demonstrate that an aggressive LR with adaptive reduction mechanisms substantially outperform lower initial LRs. Testing multiple static LR values was waived due to the already large number of test configurations. Future deep learning-based remote sensing studies should consider such adaptive LR scheduling to improve training efficiency and model performance.

5.4 Spectral Band Analysis

Even though RGB imagery proved marginally better in a single-site setting, the superior performance of RGB over MS imagery is most likely due to calibration issues of MS imagery (Section 5.1.2). This suggests that spatial texture and structural features captured at higher resolution (1.5 cm for RGB imagery vs. 2.5 cm for MS imagery) and additional spectral resolution contribute similarly to *S. maire* canopy discrimination. This contrasts with previous research

showing improved species identification using MS data (Safonova et al., 2021), but aligns with Kattenborn et al. (2019) which demonstrated that increased spectral resolution at this spatial scale is not essential for species identification. Overall, the results suggest that optimal band selection is site-dependent, and that no single imagery type consistently outperforms all others across sites. Thus, it appears to be a near-even tradeoff between the additional spectral information provided by MS bands and the higher spatial resolution in RGB imagery.

Incorporation of RENDVI with MS bands did not consistently improve performance, suggesting that derived indices may not provide discriminative information beyond raw spectral bands. This thesis employed RENDVI because the sensor lacked blue band availability; earlier studies using blue bands have shown improved performance in vegetation monitoring (Heim et al., 2019). In a dataset where MS imagery was combined with its derived VIs, performance marginally outperformed RGB imagery (Ulku et al., 2022), a finding that could be confirmed by the results of this study.

Absolute-calibrated MS bands (MS_{abs}) performed poorly in multi-site models, indicating that CRP limitations, specifically saturation in R and G bands, compromised cross-site model performance. Relative calibration, using only the sun sensor, achieved comparable or superior single-site performance and substantially better multi-site generalisation, suggesting the sun sensor alone provided more consistent reflectance estimation than the combined sun sensor and (saturated) CRP calibration approach. Additionally, at this high resolution, it has been found that flying under sunny conditions actually reduces model performance (Barreto et al., 2023). Although single-site results indicate that properly calibrated MS imagery can perform well, the overall outcomes suggest it is unlikely to outperform high-resolution RGB imagery for *S. maire* detection. This has practical implications: resources could be concentrated on acquiring a high-quality RGB camera rather than a higher-cost MS setup, reducing data volume, processing time, and the complexity of workflows for community groups. Nonetheless, MS data may still provide value for direct disease detection, which was not explored in this thesis.

5.5 Model Performance

Model performance exhibited substantial variability across sites and hyperparameter configurations. Single-site models achieved F1 scores ranging from 0.81 at the best site to <0.5 at other locations, whilst multi-site training proved less effective (best F1=0.51), leading to numerous misclassifications (Figure 15). Despite these limitations, the results demonstrate successful application of the methodology under optimal, site-specific conditions.

5.5.1 Site-Specific Performance Variation

The performance of models varied significantly across sites. The vegetation composition at lower-performing sites (A2, H1) was comparable to better-performing sites (A1, A3). However, site A3 benefited from both greater *S. maire* foliage density and larger training datasets that have been more representative (Figure 13). These dataset size differences had substantial impact on performance. However, even the best-performing sites' training datasets are minimal compared to other studies employing hundreds of annotated features (Ball et al., 2023) or thousands of training images (Pearse et al., 2021), making the strong results for these sites promising for scaling to larger datasets.

Despite larger combined training datasets, multi-site models achieved substantially lower performance (best F1=0.51 for RGB imagery and F1=0.35 for MS imagery). This further indicates the presence of site-specific characteristics and the inherent difficulty in developing universally applicable detection methods for complex forest structures, where differences in phenology and vegetation composition occur.

5.5.2 Data Constraints and Forest Structure

Class imbalance represents a major challenge in remote sensing applications due to skewed distribution of ground features. Whilst research lacks consensus on optimal approaches, the strategies employed here (assigning higher weights to minority classes and implementing data augmentation) represent current best practice (Sharma & Gosain, 2025). The class imbalance (3.4±0.7% (mean±SE) target class pixel) was substantially more severe than in comparable studies. For example, Lobo Torres et al. (2020) reported 44% target class representation, making the present task considerably more challenging.

The complex canopy structure posed additional challenges. *S. maire* can occupy understory positions beneath taller canopy species (Figure 10d), resulting in occlusion that constrained top-down detection. Consequently, not all specimens were visible on orthomosaics, further limiting sample size. High species diversity, morphological similarity among co-occurring trees, and the relatively small leaf size of *S. maire* (~5 cm) complicate visual discrimination. Achieving higher resolution to fully leverage canopy morphology would require impractically low flight heights, with increased computational and data storage requirements. Nonetheless, Barreto et al. (2023) used ultra-high-resolution UAV-imagery (0.25–0.41cm) to detect disease, showing that such an approach could be successful.

The combination of limited training data, forest complexity, morphological similarity, small leaf dimensions, and partial understory positioning proved too substantial for developing a robust, generalisable *S. maire* mapping methodology. Comparative studies benefited from substantially larger datasets: Schiefer et al. (2020) utilised 51 ha with significantly more samples. Future studies on *S. maire* could therefore benefit from larger datasets, more sites and even the inclusion of forest-wide inventory mapping to help the models delineate *S. maire* from other species.

5.5.3 Comparative Analysis of Findings

Understanding performance metric selection for imbalanced datasets is crucial and requires careful consideration. The F1 score was selected to ensure comparability with existing literature, where it remains the predominant evaluation standard (Ball et al., 2023; Li et al., 2022; Schiefer et al., 2020; L. Zhong et al., 2024). However, F1 can distort model performance evaluation with increasing class imbalance (Holzmann & Klar, 2024). Additionally, F1 may not fully capture performance when predictions are spatially correct, but boundary delineation is imprecise. See RGB prediction on site A2 in Figure 14 for an example of this issue, which applies to any semantic segmentation task on vegetation given the frayed boundaries of canopies.

This thesis represents the first deep learning application for *S. maire* detection. Whilst other New Zealand Myrtaceae (*Metrosideros excelsa*) have been successfully classified using deep learning, direct comparison is limited. Pearse et al. (2021) achieved 0.947 accuracy for phenologically distinct flowering trees and 0.927 during non-flowering periods. They employed substantially larger training datasets with 4600 manually extracted images despite using lower-resolution RGB imagery at 10 cm. Furthermore, they used an object detection method rather than semantic segmentation. Given the limited number of trees available in this thesis and their morphological variability, the semantic segmentation approach adopted here was designed to emphasise detailed growth patterns and morphology rather than tree crown boundary shape.

Studies also applying a FCNN approach on more abundant species achieve stronger results (James & Bradshaw, 2020; Kattenborn et al., 2019; Lobo Torres et al., 2020; Shahi et al., 2023). Alternatively to the F1 score, accuracy can also be used, which is defined as:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{7}$$

When identifying the shrub species *Ulex europaeus*, Kattenborn et al. (2019) achieved an accuracy of 0.84 on a substantially larger dataset (4000 image tiles) with a more balanced class distribution. However, comparison between F1 score and accuracy is challenging as in

imbalanced datasets accuracy is more distorted than F1; in this thesis, with 95% background and 5% target class, predicting all pixels as background would result in an accuracy of 0.95. James & Bradshaw (2020) achieved F1=0.83 in binary semantic segmentation on shrubs from the *Hakea* genus. However, their training data consisted of raw UAV imagery without fixed ground resolution, avoiding orthomosaic merging artefacts that degrade tree canopy quality. Such artefacts were also an issue in this thesis (Figure S9). Lobo Torres et al. (2020) achieved F1=0.96 for single tree species (*Dipteryx alata*) detection in urban contexts on 3 cm resolution RGB imagery.

All comparable studies target species that are easier to distinguish from surrounding vegetation. The present results are therefore promising given both the limited training data for *S. maire* and the complexity of the vegetation structure. Future research should explore complementary methodologies including object detection approaches, ultra-high resolution UAV imagery in the mm-scale focusing on leaf morphology, or even lesion-level disease detection if abaxial leaf surfaces become visible (Barreto et al., 2023).

5.6 LiDAR Tree Segmentation

Two tree instance segmentation algorithms were evaluated for extracting individual tree point clouds: TreeLearn (Henrich et al., 2024) and Treeiso (Xi & Hopkinson, 2022). Both algorithms were developed primarily for managed or plantation forests, predominantly targeting coniferous species, rather than the present complex forest type. Furthermore, both models were trained on terrestrial laser scanning (TLS) data rather than UAV acquisitions (Henrich et al., 2024; Xi & Hopkinson, 2022). TLS provides substantially more sub-canopy information than nadir platforms. Although an advanced UAV-based approach exists (Xiang et al., 2024), it was similarly developed for managed coniferous forests.

Both TreeLearn and Treeiso would benefit from training or fine-tuning on representative native forest data (Henrich et al., 2024; Xi & Hopkinson, 2022). Treeiso is explicitly not designed as a complete solution; the segmentation results would require iterative processing of problematic areas or manual correction. Given resource constraints and time limitations, the decision was made to discontinue LiDAR-based species identification in favour of 2D-sensed data approaches. Furthermore, full-waveform LiDAR systems (Reitberger et al., 2009) combined with MS imagery have achieved strong classification results in other studies (Briechle et al., 2020); however, the Zenmuse L2 sensor used here captures only intensity values without waveform data.

The main limitation appears to be algorithmic: existing methods struggle with dense canopy layering and understory complexity in this forest type. Ground-based LiDAR could provide higher density, but would face similar structural challenges without algorithmic improvements. Future research should initially explore further training of these existing algorithms. Especially Treeiso seems promising, considering the complexity of the forest type. Furthermore, multi-modal approaches combining spectral and structural data should be explored. Fusing MS bands onto LiDAR point clouds would create a unified dataset containing both spectral and geometric information, which has shown to be viable (Briechle et al., 2020).

5.7 Broader Implications and Direct Disease Detection

The ultimate objective of *S. maire* mapping is to support myrtle rust management. In an optimal case, the methodology would even allow assessment of individual tree health, by either directly detecting myrtle rust or assessing foliage density to draw conclusions about advancement of infection. However, direct detection of disease symptoms presents additional challenges beyond species identification. *A. psidii* pustules develop on leaf undersides, rendering them largely invisible to nadir UAV imagery. Hence, LiDAR point densities in outer canopy regions may be a good indicator of disease advancement based on foliage dieback. However, considering the cost and complexity of LiDAR and MS data acquisition and processing, RGB imagery-based approaches remain the most practical and desirable for operational applications.

The annual symptom maximum occurs during March–April (Beresford et al., 2019); however, data collection in this thesis occurred in September, a period of minimal symptom occurrence. Furthermore, the 2024/25 summer exhibited reduced symptom levels due to unfavourable weather conditions for new infections (R. Beresford, personal communication, October 29, 2025), limiting potential disease signals in the acquired data. Even under conditions of active disease expression, the small leaf dimensions and most symptom occurrence on leaf undersides, make disease detection extremely challenging.

Nonetheless, myrtle rust identification has been achieved in a plantation setting with an accuracy of over 95% (Heim et al., 2019); however, this is a much simpler task than field-based detection as the species is already known. This thesis has shown that the complexity of *S. maire*'s natural environment and the rarity of the species made the application of a deep learning approach inherently challenging. Hence, future research should first focus on improving the identification of *S. maire* only, before additionally trying to detect the disease. To achieve this, larger and more diverse training samples are recommended. For example, the inclusion of natural old-growth swamp, different urban parks, replanted sites and remnant trees could lead

to a more representative and extended dataset. Additionally, UAV flights should be carried out in March–April if disease detection is the priority.

5.8 Scalability and Conservation Applications

The methodologies from this thesis have broader applicability beyond *S. maire*, with potential to support conservation of other threatened Myrtaceae affected by myrtle rust. New Zealand hosts 13 native Myrtaceae species, several of which also face conservation challenges. For example, Pōhutukawa, whilst more abundant (Pearse et al., 2021), exhibits high myrtle rust susceptibility (De Lange et al., 2024).

The transferability of this methodology to other Myrtaceae depends on species-specific factors including canopy position, morphological distinctiveness, and phenological traits. Canopy-dominant species like pōhutukawa and rātā would be substantially more detectable than a partial understory species due to reduced occlusion, and their distinctive morphology and phenology offer additional detection opportunities (Pearse et al., 2021). However, transfer learning from the constrained *S. maire* models is not recommended; instead, species-specific models should be trained from scratch with substantially larger training datasets. This approach allows optimisation of data collection protocols (flight timing, sensor selection, radiometric calibration) and avoids transferring the fundamental limitations of dataset size, class imbalance, and site-specificity inherent to the current thesis.

UAV-deep learning approaches integrated with existing biosecurity surveillance frameworks could enable landscape-scale monitoring programmes. New Zealand's myrtle rust surveillance is currently mostly reactive, relying on opportunistic reporting and targeted ground surveys at known infection sites. A proactive surveillance system using either manned aircraft that can hold large sensors or UAV-based monitoring (which is more cost-efficient) of susceptible populations could enable earlier detection of infection, supporting timely fungicide application before severe damage occurs. Such systems would be particularly valuable for isolated or remote populations where ground-based monitoring is logistically challenging.

Species distribution modelling integrated with UAV reconnaissance offers a hierarchical detection framework: coarse-scale SDMs identify high-probability occurrence areas and UAV surveys map individuals within those areas. This thesis provides a methodological foundation and proof-of-concept for developing more robust UAV-deep learning approaches, which would be particularly valuable for expanding surveillance capacity beyond known populations. For example, recent SDM development for *S. maire* in the Wellington region (Herbert et al., 2025)

identified potential habitat that could be systematically surveyed using UAV approaches to locate previously unknown populations, expanding the known populations and enabling proactive myrtle rust management.

The methodology's applicability extends internationally to other regions affected by *A. psidii*. Australia faces comparable myrtle rust impacts on native Myrtaceae, with local extinctions already documented (Fensham & Radford-Smith, 2021). The technical approaches developed here provides a methodological template adaptable to similar conservation challenges globally. However, each regional context requires separate consideration of species traits, abundance and forest structure to determine optimal detection strategies.

5.9 Future Work

Several lines of work follow directly from the limitations identified above and from the LiDAR experiments. First, future research should initially focus on improving the identification of *S. maire* before attempting direct disease detection. This requires larger and more diverse training samples drawn from additional sites and seasons (for example inclusion of natural old-growth swamp, different urban parks, replanted sites and remnant trees) to improve representativeness. In cases where disease detection is the priority, targeted acquisitions during the March–April symptom peak are recommended.

Second, LiDAR-based methods require further algorithmic development and domain-adaptive training: both TreeLearn and Treeiso would benefit from fine-tuning on representative native-forest UAV/TLS datasets. Exploring higher-density or full-waveform sensors could also improve structural signal for both instance segmentation and health assessment.

Third, multi-modal fusion warrants focused development: combining spectral information from MS orthomosaics with structural data from LiDAR point clouds (for example by projecting MS bands onto segmented tree point clouds) is a promising avenue shown to be viable in prior work (Briechle et al., 2020).

Finally, operational deployment should prioritise integration with SDM outputs to guide targeted UAV surveys and standardise radiometric calibration workflows across sites to reduce multi-site generalisation issues.

Chapter 6 Conclusions

In this study, we demonstrated that deep learning can successfully identify *S. maire* under optimal conditions using UAV-derived imagery and deep learning-based semantic segmentation. However, performance was highly variable across sites, and multi-site generalisation proved substantially more challenging than site-specific models. This highlights the difficulty of developing universally applicable detection methods for complex forest structures and rare species with a lack of distinct morphological features. Whilst deep learning is often promoted as a universal solution for ecological detection tasks, its effectiveness fundamentally depends on the availability and representativeness of training data. This is a constraint especially severe for rare, cryptic, or canopy-hidden species. Therefore, multi-modal data-fusion approaches should be further explored.

This work revealed several methodological insights for *S. maire* detection and semantic segmentation in general. L_{Dice} outperformed pixel-wise L_{WBCE} for severe class imbalance. RGB imagery performed comparably to or better than MS imagery, suggesting that spatial resolution and morphological features contribute equally or more to detection performance than additional spectral bands. An aggressive LR with adaptive scheduling substantially improved results. Calibration issues of MS imagery likely contributed to poor multi-site generalisation, indicating that future research requires improved radiometric calibration protocols and best practice recommendations. LiDAR-based approaches showed promise for improving deep learning detection workflows and should be further investigated, particularly for detecting myrtle rust infection and assessing plant health.

The methodological insights from this study, the critical importance of radiometric calibration, hyperparameter sensitivity, and limitations of multi-site generalisation, are transferable to analogous species and disease detection applications. This work reveals a substantial disparity between controlled detection tasks commonly reported in literature and operational monitoring requirements in complex, species-rich forest systems. As invasive pathogens increasingly affect forest ecosystems globally, research must address species identification challenges in complex environments to support biosecurity response and conservation efforts.

References

Abreu-Dias, R., Santos-Gago, J. M., Martín-Rodríguez, F., & Álvarez-Sabucedo, L. M. (2025). Advances in the Automated Identification of Individual Tree Species: A Systematic Review of Drone- and AI-Based Methods in Forest Environments. *Technologies*, *13*(5), 187. https://doi.org/10.3390/technologies13050187

Agisoft LLC. (2025). *Agisoft Metashape User Manual: Professional Edition, Version 2.2* [Manual]. https://www.agisoft.com/pdf/metashape-pro_2_2_en.pdf

Agisoft LLC, v2.2.1. (2025). *Agisoft Metashape Professional (Version 2.2.1)* [Computer Software]. https://www.agisoft.com

Agisoft Metashape - Complete Tutorial (Cloud, Mesh, DSM, DTM, Classify, Orthoimage - No GCPs). (2022). https://www.youtube.com/watch?v=je79gV8HsZI

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Apiculture New Zealand. (2024). *New Zealand Honey Strategy 2024-2030*. https://apinz.org.nz/wp-content/uploads/2024/02/NZ-Honey-Strategy-2024-2030-FINAL-version-for-web-20-Feb.pdf

Auckland Council. (n.d.). *Where you can fly a drone or an unmanned aerial vehicle (UAV)*. Retrieved April 5, 2025, from https://www.aucklandcouncil.govt.nz/en/parks-recreation/get-outdoors/drones-unmanned-aerial-vehicles/where-you-can-fly-your-drone-uav.html

Austropuccinia psidii (G. Winter) Beenken. (2017). https://biotanz.landcareresearch.co.nz/scientific-names/66c070c6-6ea1-4a04-923a-9d34faf95653

Azad, R., Heidary, M., Yilmaz, K., Hüttemann, M., Karimijafarbigloo, S., Wu, Y., Schmeink, A., & Merhof, D. (2023). *Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook* (arXiv:2312.05391). arXiv. https://doi.org/10.48550/arXiv.2312.05391

Balkwill, C. G., Deslippe, J. R., Horton, P., David, C., Wu, C., Koot, E., Ritchie, P., Blissett, W., & Chagné, D. (2024). De-novo assembly of a reference genome for a critically threatened Aotearoa New Zealand tree species, Syzygium maire (Myrtaceae). *Tree Genetics & Genomes*, *20*(5), 24. https://doi.org/10.1007/s11295-024-01659-5

Balkwill, C. G., Koot, E., Ritchie, P., Chagné, D., & Deslippe, J. R. (2025). Adaptive Potential of *Syzygium Maire* , a Critically Threatened Habitat Specialist Tree Species in Aotearoa New Zealand. *Evolutionary Applications*, *18*(10), e70161. https://doi.org/10.1111/eva.70161

Ball, J. G. C., Hickman, S. H. M., Jackson, T. D., Koay, X. J., Hirst, J., Jay, W., Archer, M., Aubry-Kientz, M., Vincent, G., & Coomes, D. A. (2023). Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN. *Remote Sensing in Ecology and Conservation*, *9*(5), 641–655. https://doi.org/10.1002/rse2.332

Barrero, O., & Perdomo, S. A. (2018). RGB and multispectral UAV image fusion for Gramineae weed detection in rice fields. *Precision Agriculture*, *19*(5), 809–822. https://doi.org/10.1007/s11119-017-9558-x

Barreto, A., Reifenrath, L., Vogg, R., Sinz, F., & Mahlein, A.-K. (2023). Data Augmentation for Mask-Based Leaf Segmentation of UAV-Images as a Basis to Extract Leaf-Based Phenotyping Parameters. *KI - Künstliche Intelligenz*, *37*(2-4), 143–156. https://doi.org/10.1007/s13218-023-00815-8

Beenken, L. (2017). Austropuccinia: A new genus name for the myrtle rust *Puccinia Psidii* placed within the redefined family Sphaerophragmiaceae (Pucciniales). *Phytotaxa*, *297*(1), 53. https://doi.org/10.11646/phytotaxa.297.1.5

Beresford, R. (2025). *Guide to myrtle rust removal and treatment* (27020). Plant & Food Research.

Beresford, R. M., Shuey, L. S., & Pegg, G. S. (2020). Symptom development and latent period of *Austropuccinia Psidii* (myrtle rust) in relation to host species, temperature, and ontogenic resistance. *Plant Pathology*, *69*(3), 484–494. https://doi.org/10.1111/ppa.13145

Beresford, R., Grant, S., Ganley, B., & Campbell, R. (2019). Impacts of myrtle rust in New Zealand since its arrival in 2017. *New Zealand Garden Journal*, *22*(2), 10. https://myrtlerust.org.nz/assets/news/NZ-Garden-Journal-Dec-2019-p5-10.pdf

Beresford, R., & Wright, P. (2022). *Risk-based fungicide management for myrtle rust in nurseries* (22715; p. 37). Plant & Food Research. https://www.mpi.govt.nz/dmsdocument/54247-Risk-based-fungicide-management-for-myrtle-rust-in-nurseries

Bettoni, J. C., Van Der Walt, K., Souza, J. A., McLachlan, A., & Nadarajan, J. (2024). Sexual and asexual propagation of *Syzygium Maire*, a critically endangered Myrtaceae species of New Zealand. *New Zealand Journal of Botany*, *62*(1), 35–52. https://doi.org/10.1080/002885X.2022.2158110

Bosch, M. (2020). DetecTree: Tree detection from aerial imagery in Python. *Journal of Open Source Software*, *5*(50), 2172. https://doi.org/10.21105/joss.02172

Boufleur, T. R., Morales, J. V. P., Martins, T. V., Gonçalves, M. P., Júnior, N. S. M., & Amorim, L. (2023). A Diagnostic Guide for Myrtle Rust. *Plant Health Progress*, *24*(2), 242–251. https://doi.org/10.1094/PHP-06-22-0057-DG

Bourday, R., Aattouchi, I., & Ait Kerroum, M. (2024). A Comparative Study of Deep Learning Loss Functions: A Polyp Segmentation Case Study. In F. P. García Márquez, A. Jamil, I. S. Ramirez, S. Eken, & A. A. Hameed (Eds.), *Computing, Internet of Things and Data Analytics* (Vol. 1145, pp. 68–78). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53717-2_7

Briechle, S., Krzystek, P., & Vosselman, G. (2020). Classification of Tree Species and Standing Dead Trees by Fusing UAV-Based LiDAR Data and Multispectral Imagery in the 3D Deep Neural Network PointNet++. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *V-2-2020*, 203–210. https://doi.org/10.5194/isprs-annals-V-2-2020-203-2020

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and Flexible Image Augmentations. *Information*, *11*(2), 125. https://doi.org/10.3390/info11020125

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

Chang, B., Li, F., Hu, Y., Yin, H., Feng, Z., & Zhao, L. (2025). Application of UAV remote sensing for vegetation identification: A review and meta-analysis. *Frontiers in Plant Science*, *16*, 1452053. https://doi.org/10.3389/fpls.2025.1452053

Chen, H., He, Y., Zhang, L., Yang, W., Liu, Y., Gao, B., Zhang, Q., & Lu, J. (2024). A Multi-Input Channel U-Net Landslide Detection Method Fusing SAR Multisource Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 1215–1232. https://doi.org/10.1109/JSTARS.2023.3339294

Chng, S., Soewarto, J., Adusei-Fosu, K., Rolando, C. A., Ganley, R., Padamsee, M., Waipara, N., Grant, A., Wegner, S., & Gee, M. (2019). *Potential disease control tools most likely to be effective against Austropuccinia psidii*. Ministry for Primary Industries.

Chock, M. K. (2020). The global threat of Myrtle rust (*Austropuccinia Psidii*): Future prospects for control and breeding resistance in susceptible hosts. *Crop Protection*, *136*, 105176. https://doi.org/10.1016/j.cropro.2020.105176

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

CloudCompare, v2.14.alpha. (2022). *CloudCompare (Version 2.14.alpha)* [Computer Software]. https://www.cloudcompare.org/

De Lange, P. J., Gosden, J., Shannel P. Courtney, Fergus, A. J., Barkla, J., Beadel, S. M., Champion, P. D., Hindmarsh-Walls, R., Makan, T., & Michel, P. (2024). *Conservation status*

of vascular plants in Aotearoa New Zealand, 2023. https://doi.org/10.13140/RG.2.2.11940.49288

DJI. (n.d.-a). *DJI Mavic 3M* [Equipment]. DJI. https://ag.dji.com/mavic-3-m

DJI. (n.d.-b). *DJI Pilot - Download Center - DJI New Zealand* [Software]. https://www.dji.com/nz/downloads/djiapp/photo

DJI. (n.d.-c). *Matrice 350 RTK* [Equipment]. https://enterprise.dji.com/matrice-350-rtk

DJI. (n.d.-d). *Mavic 3 Multispectral Edition*. Retrieved February 8, 2026, from https://ag.dji.com/photo

DJI. (n.d.-e). *Zenmuse L2* [Equipment]. https://enterprise.dji.com/zenmuse-l2

DJI, v4.13.3.0. (n.d.). *DJI Terra (v4.13.3.0)* [Software]. https://enterprise.dji.com/dji-terra

Du Plessis, E., Granados, G. M., Barnes, I., Ho, W. H., Alexander, B. J. R., Roux, J., & McTaggart, A. R. (2019). The pandemic strain of *Austropuccinia Psidii* causes myrtle rust in New Zealand and Singapore. *Australasian Plant Pathology*, *48*(3), 253–256. https://doi.org/10.1007/s13313-019-0624-x

Dymond, J., Sabetizade, M., Newsome, P., Harmsworth, G., & Ausseil, A.-G. (2021). Revised extent of wetlands in New Zealand. *New Zealand Journal of Ecology*. https://doi.org/10.20417/nzjecol.45.32

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering*, *151*, 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024

Emlid. (n.d.). *Reach RS3* [Equipment]. https://au.store.emlid.com/au/product/reach-rs3/

Essien, S. O., Baroutian, S., Dell, K., & Young, B. (2019). Value-added potential of New Zealand mānuka and kānuka products: A review. *Industrial Crops and Products*, *130*, 198–207. https://doi.org/10.1016/j.indcrop.2018.12.083

Fensham, R. J., Carnegie, A. J., Laffineur, B., Makinson, R. O., Pegg, G. S., & Wills, J. (2020). Imminent Extinction of Australian Myrtaceae by Fungal Disease. *Trends in Ecology & Evolution*, *35*(7), 554–557. https://doi.org/10.1016/j.tree.2020.03.012

Fensham, R. J., & Radford-Smith, J. (2021). Unprecedented extinction of tree species by fungal disease. *Biological Conservation*, *261*, 109276. https://doi.org/10.1016/j.biocon.2021.109276

Fernandez Winzer, L., Berthon, K. A., Carnegie, A. J., Pegg, G. S., & Leishman, M. R. (2019). Austropuccinia psidii on the move: Survey based insights to its geographical distribution, host species, impacts and management in Australia. *Biological Invasions*, *21*(4), 1215–1225. https://doi.org/10.1007/s10530-018-1891-0

Fernandez-Winzer, L., Berthon, K. A., Entwistle, P., Manea, A., Winzer, N., Pegg, G. S., Carnegie, A. J., & Leishman, M. R. (2020). Direct and indirect community effects of the invasive plant pathogen *Austropuccinia Psidii* (myrtle rust) in eastern Australian rainforests. *Biological Invasions*, *22*(7), 2357–2369. https://doi.org/10.1007/s10530-020-02260-2

Freudenberg, M., Nölke, N., Agostini, A., Urban, K., Wörgötter, F., & Kleinn, C. (2019). Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net. *Remote Sensing*, *11*(3), 312. https://doi.org/10.3390/rs11030312

Gailbraith, M., & Large, M. (2017). Implications for selected indigenous fauna of Tiritiri Matangi of the establishment of *Austropuccinia Psidii* (G. Winter) Beenken (myrtle rust) in northern New Zealand. *Perspectives in Biosecurity*, *2*.

Gan, Y., Wang, Q., & Iio, A. (2023). Tree Crown Detection and Delineation in a Temperate Deciduous Forest from UAV RGB Imagery Using Deep Learning Approaches: Effects of Spatial Resolution and Species Characteristics. *Remote Sensing*, *15*(3), 778. https://doi.org/10.3390/rs15030778

Geurts, E. M., Reynolds, J. D., & Starzomski, B. M. (2023). Turning observations into biodiversity data: Broadscale spatial biases in community science. *Ecosphere*, *14*(6), e4582. https://doi.org/10.1002/ecs2.4582

Ghosh, K., Bellinger, C., Corizzo, R., Branco, P., Krawczyk, B., & Japkowicz, N. (2024). The class imbalance problem in deep learning. *Machine Learning*, *113*(7), 4845–4901. https://doi.org/10.1007/s10994-022-06268-8

Gitelson, A., & Merzlyak, M. N. (1994). Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. And Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology*, *143*(3), 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0

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

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Gould, K. S., Thodey, K., Philpott, M., & Ferguson, L. R. (2006). Antioxidant activities of extracts from traditional Maori food plants. *New Zealand Journal of Botany*, *44*(1), 1–4. https://doi.org/10.1080/0028825X.2006.9513001

Granados, G. M., McTaggart, A. R., Barnes, I., Rodas, C. A., Roux, J., & Wingfield, M. J. (2017). The pandemic biotype of *Austropuccinia Psidii* discovered in South America. *Australasian Plant Pathology*, *46*(3), 267–275. https://doi.org/10.1007/s13313-017-0488-x

Grgurinovic, C. A., Walsh, D., & Macbeth, F. (2006). Eucalyptus rust caused by *Puccinia Psidii* and the threat it poses to Australia. *EPPO Bulletin*, *36*(3), 486–489. https://doi.org/10.1111/j.1365-2338.2006.01048.x

Hamilton City Council. (n.d.). *Flying a drone*. Retrieved July 12, 2025, from https://hamilton.govt.nz/community-spaces/flying-a-drone

Heim, R. H. J., Wright, I. J., Scarth, P., Carnegie, A. J., Taylor, D., & Oldeland, J. (2019). Multispectral, Aerial Disease Detection for Myrtle Rust (*Austropuccinia Psidii*) on a Lemon Myrtle Plantation. *Drones*, *3*(1), 25. https://doi.org/10.3390/drones3010025

Helfer, S. (2014). Rust fungi and global change. *New Phytologist*, *201*(3), 770–780. https://doi.org/10.1111/nph.12570

Henrich, J., Van Delden, J., Seidel, D., Kneib, T., & Ecker, A. S. (2024). TreeLearn: A deep learning method for segmenting individual trees from ground-based LiDAR forest point clouds. *Ecological Informatics*, *84*, 102888. https://doi.org/10.1016/j.ecoinf.2024.102888

Herbert, S. M., Tomscha, S. A., Lai, H. R., Benavidez, R., Balkwill, C. G., Ruston, P. R., Jackson, B., & Deslippe, J. R. (2025). Identifying potentially suitable and accessible refugia to mitigate impacts of an emerging disease on a rare tree. *Conservation Biology*, e70088. https://doi.org/10.1111/cobi.70088

Holzmann, H., & Klar, B. (2024). *Robust performance metrics for imbalanced classification problems* (arXiv:2404.07661). arXiv. https://doi.org/10.48550/arXiv.2404.07661

Huang, B., Reichman, D., Collins, L. M., Bradbury, K., & Malof, J. M. (2019). *Tiling and stitching segmentation output for remote sensing: Basic challenges and recommendations*. https://doi.org/10.48550/arXiv.1805.12219

James, K., & Bradshaw, K. (2020). Detecting plant species in the field with deep learning and drone technology. *Methods in Ecology and Evolution*, *11*(11), 1509–1519. https://doi.org/10.1111/2041-210X.13473

Jo, I., Bellingham, P. J., McCarthy, J. K., Easdale, T. A., Padamsee, M., Wiser, S. K., & Richardson, S. J. (2022). Ecological importance of the Myrtaceae in New Zealand's natural forests. *Journal of Vegetation Science*, *33*(1), e13106. https://doi.org/10.1111/jvs.13106

Kattenborn, T., Eichel, J., & Fassnacht, F. E. (2019). Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific Reports*, *9*(1), 17656. https://doi.org/10.1038/s41598-019-53797-9

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M. D., & Dormann, C. F. (2022). Spatially autocorrelated training and validation samples inflate performance assessment

of convolutional neural networks. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, *5*, 100018. https://doi.org/10.1016/j.ophoto.2022.100018

Koishiyeva, D., Kang, J. W., Iliev, T., Bissembayev, A., & Mukasheva, A. (2025). Analysis of Loss Functions for Colorectal Polyp Segmentation Under Class Imbalance. *EEPES 2025*, 17. https://doi.org/10.3390/engproc2025104017

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Lake, T. A., Briscoe Runquist, R. D., & Moeller, D. A. (2022). Deep learning detects invasive plant species across complex landscapes using Worldview-2 and Planetscope satellite imagery. *Remote Sensing in Ecology and Conservation*, *8*(6), 875–889. https://doi.org/10.1002/rse2.288

Lecigne, B., Delagrange, S., & Messier, C. (2018). Exploring trees in three dimensions: VoxR, a novel voxel-based R package dedicated to analysing the complex arrangement of tree crowns. *Annals of Botany*, *121*(4), 589–601. https://doi.org/10.1093/aob/mcx095

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, X., He, M., Li, H., & Shen, H. (2022). A Combined Loss-Based Multiscale Fully Convolutional Network for High-Resolution Remote Sensing Image Change Detection. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5. https://doi.org/10.1109/LGRS.2021.3098774

LINZ. (2012). *NZ Geoid 2009* (Dataset 6639c917-555f-23b0-1419-9cee41907cd3). https://data.linz.govt.nz/layer/51051-nz-geoid-2009/

LINZ. (2025). *New Zealand LiDAR 1m DEM* (Dataset 5d3a81ea-1e69-8430-c654-7993af1c11d9). https://data.linz.govt.nz/layer/121859-new-zealand-lidar-1m-dem/

Liu, T., Abd-Elrahman, A., Morton, J., & Wilhelm, V. L. (2018). Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience & Remote Sensing*, *55*(2), 243–264. https://doi.org/10.1080/15481603.2018.1426091

Lobo Torres, D., Queiroz Feitosa, R., Nigri Happ, P., Elena Cué La Rosa, L., Marcato Junior, J., Martins, J., Olã Bressan, P., Gonçalves, W. N., & Liesenberg, V. (2020). Applying Fully Convolutional Architectures for Semantic Segmentation of a Single Tree Species in Urban Environment on High Resolution UAV Optical Imagery. *Sensors*, *20*(2), 563. https://doi.org/10.3390/s20020563

Low, Y. W., Rajaraman, S., Tomlin, C. M., Ahmad, J. A., Ardi, W. H., Armstrong, K., Athen, P., Berhaman, A., Bone, R. E., Cheek, M., Cho, N. R. W., Choo, L. M., Cowie, I. D., Crayn, D., Fleck, S. J., Ford, A. J., Forster, P. I., Girmansyah, D., Goyder, D. J., … Albert, V. A. (2022). Genomic insights into rapid speciation within the world's largest tree genus Syzygium. *Nature Communications*, *13*(1), 5031. https://doi.org/10.1038/s41467-022-32637-x

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). High-Resolution Aerial Image Labeling With Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(12), 7092–7103. https://doi.org/10.1109/TGRS.2017.2740362

Magyar, D., Vass, M., & Li, D.-W. (2016). Dispersal Strategies of Microfungi. In D.-W. Li (Ed.), *Biology of Microfungi* (pp. 315–371). Springer International Publishing. https://doi.org/10.1007/978-3-319-29137-6_14

Mahuta, R., van Schravendijk-Goodman, C., & van der Weerden, A. (2021). Kei konei tonu te maire tawake – keeping the kete full for the ultimate swamp tree. In *Te reo o te rep kei konei tonu au – The voice of the wetland I am still here* (pp. 89–100). Manaaki Whenua – Landcare Research.

Makinson, R. O. (2018). *Myrtle Rust reviewed: The impacts of the invasive pathogen* Austropuccinia Psidii *on the Australian environment.* Plant Biosecurity Cooperative Research Centre, Canberra. https://www.apbsf.org.au/wp-content/uploads/2018/11/Myrtle-Rust-reviewed-June-22-2018-web.pdf

Malek, S., Bazi, Y., Alajlan, N., AlHichri, H., & Melgani, F. (2014). Efficient Framework for Palm Tree Detection in UAV Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(12), 4692–4703. https://doi.org/10.1109/JSTARS.2014.2331425

Manaaki Whenua - Landcare Research. (2026). *New Zealand Plant Names Database - Beilschmiedia tawa (A.Cunn.) Benth. & Hook.f. Ex Kirk*. https://biotanz.landcareresearch.co.nz/scientific-names/E8CD2202-3B4D-4B8D-A881-2AE10996183A

Mason, B. M., Mesaglio, T., Barratt Heitmann, J., Chandler, M., Chowdhury, S., Gorta, S. B. Z., Grattarola, F., Groom, Q., Hitchcock, C., Hoskins, L., Lowe, S. K., Marquis, M., Pernat, N., Shirey, V., Baasanmunkh, S., & Callaghan, C. T. (2025). iNaturalist accelerates biodiversity research. *BioScience*, biaf104. https://doi.org/10.1093/biosci/biaf104

McCarthy, J. K., Richardson, S. J., Cooper, J. A., Bellingham, P. J., & Wiser, S. K. (2019). *Species distribution models of the native New Zealand Myrtaceae* (Report MPI 18607). Ministry for Primary Industries.

McCarthy, J. K., Richardson, S. J., Jo, I., Wiser, S. K., Easdale, T. A., Shepherd, J. D., & Bellingham, P. J. (2024). A Functional Assessment of Community Vulnerability to the Loss of Myrtaceae From Myrtle Rust. *Diversity and Distributions*, *30*(11), e13928. https://doi.org/10.1111/ddi.13928

Ministry for Primary Industries. (2020). New approach to manage myrtle rust | NZ Government. In *New approach to manage myrtle rust*. https://www.mpi.govt.nz/news/media-releases/new-approach-to-manage-myrtle-rust

Miraki, M., Sohrabi, H., Fatehi, P., & Kneubuehler, M. (2021). Individual tree crown delineation from high-resolution UAV images in broadleaf forest. *Ecological Informatics*, *61*, 101207. https://doi.org/10.1016/j.ecoinf.2020.101207

Nadarajan, J., Van Der Walt, K., Lehnebach, C. A., Saeiahagh, H., & Pathirana, R. (2021). Integrated *Ex Situ* conservation strategies for endangered New Zealand Myrtaceae species. *New Zealand Journal of Botany*, *59*(1), 72–89. https://doi.org/10.1080/0028825X.2020.1754245

Olsson, P.-O., Vivekar, A., Adler, K., Garcia Millan, V. E., Koc, A., Alamrani, M., & Eklundh, L. (2021). Radiometric Correction of Multispectral UAS Images: Evaluating the Accuracy of the Parrot Sequoia Camera and Sunshine Sensor. *Remote Sensing*, *13*(4), 577. https://doi.org/10.3390/rs13040577

OPENGIS.ch. *QField for QGIS* [Mobile App]. https://qfield.org/

Osco, L. P., Marcato Junior, J., Marques Ramos, A. P., De Castro Jorge, L. A., Fatholahi, S. N., De Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., & Li, J. (2021). A review on deep learning in UAV remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102456. https://doi.org/10.1016/j.jag.2021.102456

Paap, T., Wingfield, M. J., Burgess, T. I., Wilson, J. R. U., Richardson, D. M., & Santini, A. (2022). Invasion Frameworks: A Forest Pathogen Perspective. *Current Forestry Reports*, *8*(1), 74–89. https://doi.org/10.1007/s40725-021-00157-4

Parnell, J. A. N., Craven, L. A., & Biffin, E. (2007). Matters of Scale: Dealing with One of the Largest Genera of Angiosperms. In *Reconstructing the tree of life: Taxonomy and systematics of species rich taxa* (Vol. 72). CRC Press LLC.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library* (arXiv:1912.01703). arXiv. https://doi.org/10.48550/arXiv.1912.01703

Pearse, G. D., Watt, M. S., Soewarto, J., & Tan, A. Y. S. (2021). Deep Learning and Phenology Enhance Large-Scale Tree Species Classification in Aerial Imagery during a Biosecurity Response. *Remote Sensing*, *13*(9), 1789. https://doi.org/10.3390/rs13091789

Pfaff, R. (unpublished). *Privet detection in Auckland Parks using deep learning*. Auckland University of Technology.

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. arXiv. https://doi.org/10.48550/ARXIV.2205.01833

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). *PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space* (arXiv:1706.02413). arXiv. https://doi.org/10.48550/arXiv.1706.02413

Reitberger, J., Schnörr, Cl., Krzystek, P., & Stilla, U. (2009). 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(6), 561–574. https://doi.org/10.1016/j.isprsjprs.2009.04.002

Robbins, A., Carlos Montes-Herrera, J., Lucieer, A., & Sparrow, B. (2025). *NatureScan: RGB and Multispectral Drone Data Collection for Ecological Monitoring*. protocols.io. https://doi.org/10.17504/protocols.io.3byl463q2go5/v1

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Rouse Jr, W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Monitoring Vegetation Systems in the Great Plains with ERTS*, *SP-351(I)*, 309–317. https://ntrs.nasa.gov/api/citations/19740022614

Safonova, A., Hamad, Y., Dmitriev, E., Georgiev, G., Trenkin, V., Georgieva, M., Dimitrov, S., & Iliev, M. (2021). Individual Tree Crown Delineation for the Species Classification and Assessment of Vital Status of Forest Stands from UAV Images. *Drones*, *5*(3), 77. https://doi.org/10.3390/drones5030077

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

Schmid, L. M. H., Large, M. F., Galbraith, M., & de Lange, P. J. (2021). Short Communication: Observation of western honeybee (Apis mellifera) foraging urediniospores from myrtle-rust-infected maire tawake (Syzygium maire), Ōwairaka/Mt Albert, Tāmaki Makaurau/Auckland, New Zealand. *Perspectives in Biosecurity*, *6*, 1–7.

Shahi, T. B., Dahal, S., Sitaula, C., Neupane, A., & Guo, W. (2023). Deep Learning-Based Weed Detection Using UAV Images: A Comparative Study. *Drones*, *7*(10), 624. https://doi.org/10.3390/drones7100624

Sharma, S., & Gosain, A. (2025). Addressing class imbalance in remote sensing using deep learning approaches: A systematic literature review. *Evolutionary Intelligence*, *18*(1), 23. https://doi.org/10.1007/s12065-024-01012-8

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, *81*(2-3), 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X

Singers, N. J. D., & Rogers, G. M. (2014). A classification of New Zealand's terrestrial ecosystems. *Science for Conservation 325*, 87.

Soewarto, J., F., G., & J., C. A. (2019). Austropuccinia Psidii *(myrtle rust) global host list. Version 2. Australian Network for Plant Conservation, Canberra, ACT.* https://www.anpc.asn.au/myrtle-rust/

Srivastava, N., Rai, A., Prasad Kushwaha, S. K., & Jain, K. (2024). Advancing Multi-Class Semantic Segmentation of High-Resolution Satellite Imagery through Enhanced ASPP and Attention Mechanisms. *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 8423–8427. https://doi.org/10.1109/IGARSS53475.2024.10640382

Stewart, J. E., Ross-Davis, A. L., Graca, R. N., Alfenas, A. C., Peever, T. L., Hanna, J. W., Uchida, J. Y., Hauff, R. D., Kadooka, C. Y., Kim, M.-S., Cannon, P. G., Namba, S., Simeto, S., Pérez, C. A., Rayamajhi, M. B., Lodge, D. J., Arguedas, M., Medel-Ortiz, R., López-Ramirez, M. A., … Klopfenstein, N. B. (2018). Genetic diversity of the myrtle rust pathogen (*Austropuccinia*

Psidii) in the Americas and Hawaii: Global implications for invasive threat assessments. *Forest Pathology*, *48*(1), e12378. https://doi.org/10.1111/efp.12378

Sutherland, R., Soewarto, J., Beresford, R., & Ganley, B. (2020). Monitoring Austropuccinia psidii (myrtle rust) on New Zealand Myrtaceae in native forest. *New Zealand Journal of Ecology*, *44*(2). https://doi.org/10.20417/nzjecol.44.23

Swaminathan, V., Thomasson, J. A., Hardin, R. G., Rajan, N., & Raman, R. (2024). Selection of appropriate multispectral camera exposure settings and radiometric calibration methods for applications in phenotyping and precision agriculture. *The Plant Phenome Journal*, *7*(1), e70000. https://doi.org/10.1002/ppj2.70000

Syzygium maire (A.Cunn.) Sykes & Garn.-Jones. (1979). https://floraseries.landcareresearch.co.nz/taxa/73c686a7-ceae-4edd-98e8-e8af13d77fac

Taranaki Regional Council. (n.d.). *Biodiversity Information Sheet - Swamp Maire, Waiwaka*. https://www.trc.govt.nz/assets/Documents/Guidelines/Biodiversity-infosheets/BioInfo-28-SwampMaireWaiwaka.pdf

Teulon, D. A. J., Alipia, T. T., Ropata, H. T., Green, J. M., Rollinson, S. L. H. V.-., Cromey, M. G., Arthur, K., MacDiarmid, R. M., Waipara, N. W., & Marsh, A. T. (2015). The threat of myrtle rust to Māori taonga plant species in New Zealand. *New Zealand Plant Protection*, *68*, 66–75. https://doi.org/10.30843/nzpp.2015.68.5869

Toome-Heller, M., Ho, W. W. H., Ganley, R. J., Elliott, C. E. A., Quinn, B., Pearson, H. G., & Alexander, B. J. R. (2020). Chasing myrtle rust in New Zealand: Host range and distribution over the first year after invasion. *Australasian Plant Pathology*, *49*(3), 221–230. https://doi.org/10.1007/s13313-020-00694-9

Townsend, A. J., Lange, P. J. de, Duffy, C. A. J., Miskelly, C. M., Molloy, J., & Norton, D. A. (2008). *New Zealand Threat Classification System manual*. Science & Technical Publishing, Dept. of Conservation. https://www.doc.govt.nz/documents/science-and-technical/sap244.pdf

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

Ulku, I., Akagunduz, E., & Ghamisi, P. (2022). Deep Semantic Segmentation of Trees Using Multispectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 7589–7604. https://doi.org/10.1109/JSTARS.2022.3203145

Van Der Walt, K. (2021). Investigating cryopreservation options for 'Syzygium maire', a threatened endemic New Zealand tree. *Australasian Plant Conservation: Journal of the Australian Network for Plant Conservation*, *29*(3), 22–25. https://doi.org/10.5962/p.373864

Van Der Walt, K., Burritt, D. J., & Nadarajan, J. (2022). Impacts of Rapid Desiccation on Oxidative Status, Ultrastructure and Physiological Functions of Syzygium maire (Myrtaceae) Zygotic Embryos in Preparation for Cryopreservation. *Plants*, *11*(8), 1056. https://doi.org/10.3390/plants11081056

Van Der Walt, K., Kemp, P., Sofkova-Bobcheva, S., Burritt, D. J., & Nadarajan, J. (2020). Seed development, germination, and storage behaviour of *Syzygium Maire* (Myrtaceae) *,* a threatened endemic New Zealand tree. *New Zealand Journal of Botany*, *59*(2), 198–216. https://doi.org/10.1080/0028825X.2020.1794911

Van Der Walt, K., Nadarajan, J., Mathew, L., Bettoni, J. C., & Souza, J. A. (2023). Advances in cryopreservation of Syzygium maire (swamp maire, maire tawake) zygotic embryos, a critically endangered tree species endemic to New Zealand. *Frontiers in Conservation Science*, *4*, 1269881. https://doi.org/10.3389/fcosc.2023.1269881

Wagner, F. H., Sanchez, A., Tarabalka, Y., Lotte, R. G., Ferreira, M. P., Aidar, M. P. M., Gloor, E., Phillips, O. L., & Aragão, L. E. O. C. (2019). Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, *5*(4), 360–375. https://doi.org/10.1002/rse2.111

Wäldchen, J., Rzanny, M., Seeland, M., & Mäder, P. (2018). Automated plant species identification—Trends and future directions. *PLOS Computational Biology*, *14*(4), e1005993. https://doi.org/10.1371/journal.pcbi.1005993

Wingfield, M. J., Slippers, B., Wingfield, B. D., & Barnes, I. (2017). The unified framework for biological invasions: A forest fungal pathogen perspective. *Biological Invasions*, *19*(11), 3201–3214. https://doi.org/10.1007/s10530-017-1450-0

Wołk, K., & Tatara, M. S. (2024). A Review of Semantic Segmentation and Instance Segmentation Techniques in Forestry Using LiDAR and Imagery Data. *Electronics*, *13*(20), 4139. https://doi.org/10.3390/electronics13204139

Wu, Y., Liu, L., Bae, J., Chow, K.-H., Iyengar, A., Pu, C., Wei, W., Yu, L., & Zhang, Q. (2019). *Demystifying Learning Rate Policies for High Accuracy Training of Deep Neural Networks* (arXiv:1908.06477). arXiv. https://doi.org/10.48550/arXiv.1908.06477

Xi, Z., & Hopkinson, C. (2022). 3D Graph-Based Individual-Tree Isolation (Treeiso) from Terrestrial Laser Scanning Point Clouds. *Remote Sensing*, *14*(23), 6116. https://doi.org/10.3390/rs14236116

Xiang, B., Wielgosz, M., Kontogianni, T., Peters, T., Puliti, S., Astrup, R., & Schindler, K. (2024). Automated forest inventory: Analysis of high-density airborne LiDAR point clouds with 3D deep learning. *Remote Sensing of Environment*, *305*, 114078. https://doi.org/10.1016/j.rse.2024.114078

Zhang, C., Atkinson, P. M., George, C., Wen, Z., Diazgranados, M., & Gerard, F. (2020). Identifying and mapping individual plants in a highly diverse high-elevation ecosystem using UAV imagery and deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, 280–291. https://doi.org/10.1016/j.isprsjprs.2020.09.025

Zhang, T., Xu, Z., Su, J., Yang, Z., Liu, C., Chen, W.-H., & Li, J. (2021). Ir-UNet: Irregular Segmentation U-Shape Network for Wheat Yellow Rust Detection by UAV Multispectral Imagery. *Remote Sensing*, *13*(19), 3892. https://doi.org/10.3390/rs13193892

Zhang, T., Yang, Z., Xu, Z., & Li, J. (2022). Wheat Yellow Rust Severity Detection by Efficient DF-UNet and UAV Multispectral Imagery. *IEEE Sensors Journal*, *22*(9), 9057–9068. https://doi.org/10.1109/JSEN.2022.3156097

Zhong, H., Zhang, Z., Liu, H., Wu, J., & Lin, W. (2024). Individual Tree Species Identification for Complex Coniferous and Broad-Leaved Mixed Forests Based on Deep Learning Combined with UAV LiDAR Data and RGB Images. *Forests*, *15*(2), 293. https://doi.org/10.3390/f15020293

Zhong, L., Dai, Z., Fang, P., Cao, Y., & Wang, L. (2024). A Review: Tree Species Classification Based on Remote Sensing Data and Classic Deep Learning-Based Methods. *Forests*, *15*(5), 852. https://doi.org/10.3390/f15050852

ZLEA. (2024). *English: DJI Mavic 3 Enterprise FA373WYPN9 at 2024 Sun 'n Fun, Lakeland Linder International Airport.* [Photograph]. https://commons.wikimedia.org/wiki/File:DJI_Mavic_3M_Multispectral_FA373WYPN9_%284-9-2024%29_%28cropped%29.jpg

Appendix A: List of Acronyms

AGL above ground level

CAA Civil Aviation Authority

CNN convolutional neural network

CRP Calibrated Reference Panel

DEM digital elevation model

DSM digital surface model

DTM digital terrain model

F1 The harmonic mean of precision and recall, used as a performance metric for classification models.

FCNN fully convolutional neural network

G green

GIS geographic information system

GNSS Global Navigation Satellite System

GPS global positioning system

HS hyperspectral

IoU intersection over union

LAS An open, binary file format for storing 3D point cloud data from LiDAR surveys.

LINZ Toitū Te Whenua Land Information New Zealand

LR learning rate

L_{Dice} dice loss

L_{WBCE} weighted binary cross-entropy loss

L_c combined weighted binary cross-entropy + dice loss

LiDAR light detection and ranging

MS multispectral

NDVI normalised difference vegetation index

NGRDI normalised green-red difference index

NIR near-infrared

P+F Plant & Food Research

PositioNZ-RT Toitū Te Whenua PositioNZ Real Time Service

R red

RE red edge

REDVI red edge difference vegetation index

RENDVI red edge normalised difference vegetation index

RF random forest

RGB red-green-blue

ROI region of interest

RTK Real-Time Kinematic

ReLU rectified linear unit

SDM species distribution model

SGD stochastic gradient descent

SVM support vector machine

TLS terrestrial laser scanning

UAV unmanned aerial vehicle

VI vegetation index

WF8 warm temperate forest type 8

Appendix B: Supplemental Materials

Search Terms

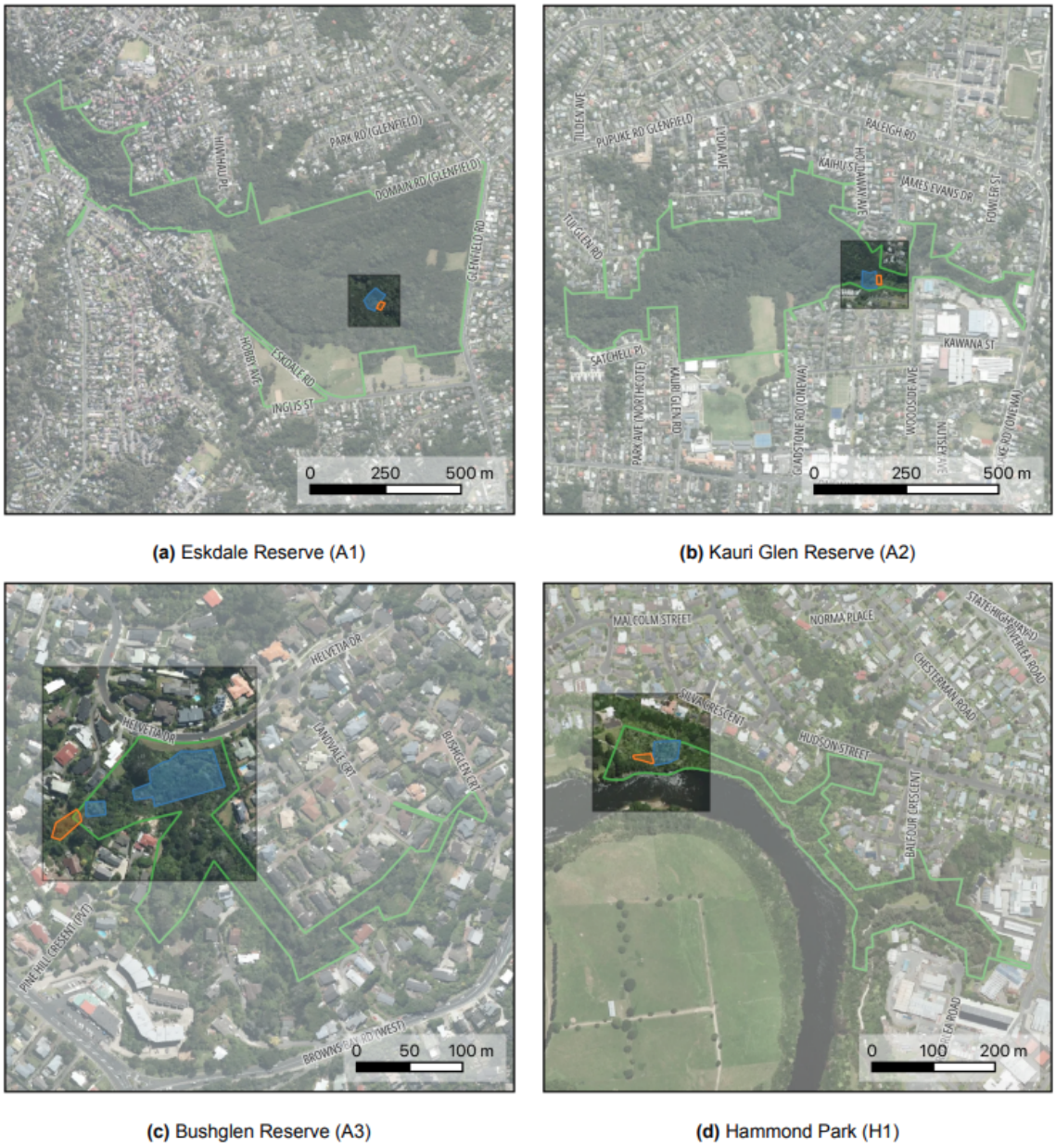
Grid DL

(“deep learning” OR “deep neural network” OR “CNN” OR “convolutional neural network”) AND (“remote sensing” OR “GIS” OR “geospatial” OR “earth observation” OR “EO” OR “mapping” OR “environmental monitoring”) AND (“satellite” OR “aerial” OR “UAV” OR “UAS” OR “drone” OR “multispectral” OR “MS” OR “hyperspectral” OR “HS” OR “RGB”)

Point-DL

(“deep learning” OR “deep neural network” OR “neural network” OR “geometric deep learning” OR “3D deep learning” OR “point-based network” OR “PointNet” OR “PointNet++” OR “PointCNN” OR “DGCNN” OR “Graph Neural Network” OR “PointTransformer”) AND (“remote sensing” OR “GIS” OR “geospatial” OR “earth observation” OR “EO” OR “mapping” OR “environmental monitoring”) AND (“point cloud” OR “pointcloud” OR “LiDAR” OR “lidar” OR “light detection and ranging” OR “3D data” OR “3D point cloud” OR “laser scanning” OR “ALS” OR “TLS”)

Site Extent Inset Maps



(a) Eskdale Reserve (A1)

(b) Kauri Glen Reserve (A2)

(c) Bushglen Reserve (A3)

(d) Hammond Park (H1)

Figure S1: Inset maps showing the location of the maps of Figure 6. The map shows park extent (green), training (blue) and test (orange) zones used for model evaluation; site areas are listed in Table 1. Basemap: LINZ aerial imagery under CC BY 4.0. Park extents: Auckland Council (A1–A3) and Waikato OneView (H1) under CC BY 4.0.

Multispectral Reflectance Calibration

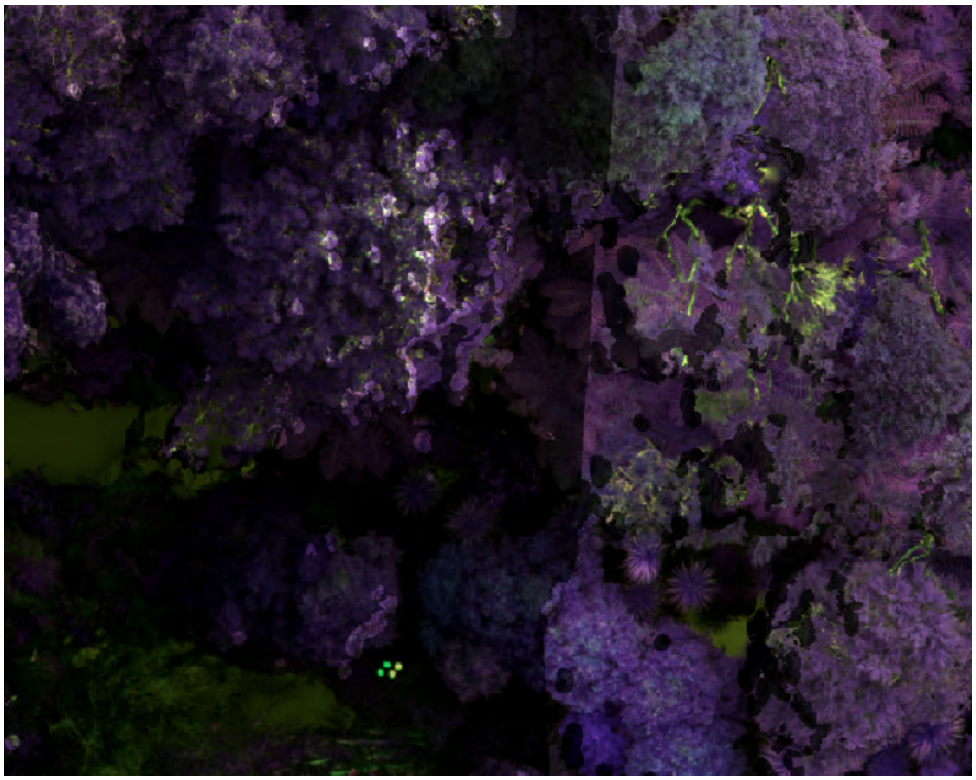
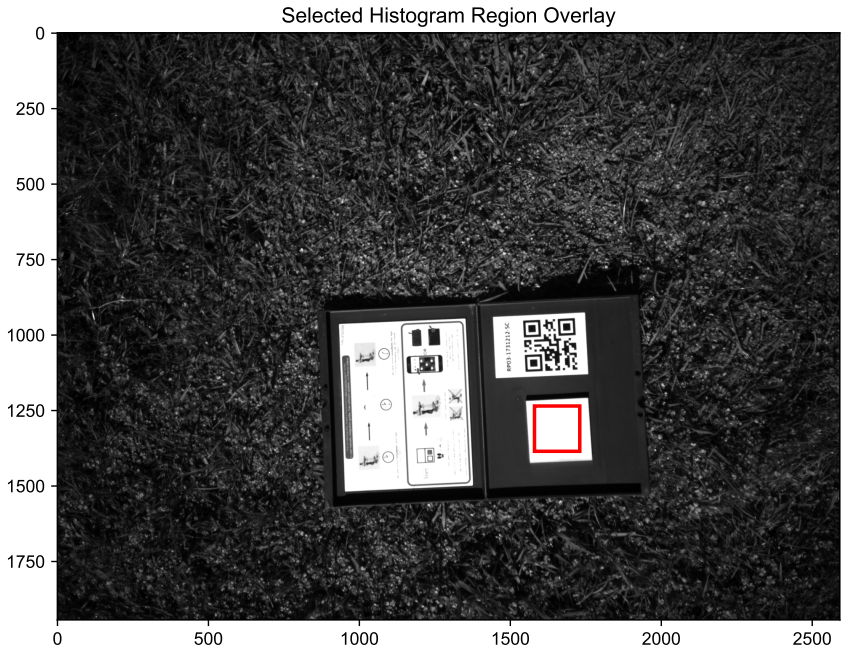
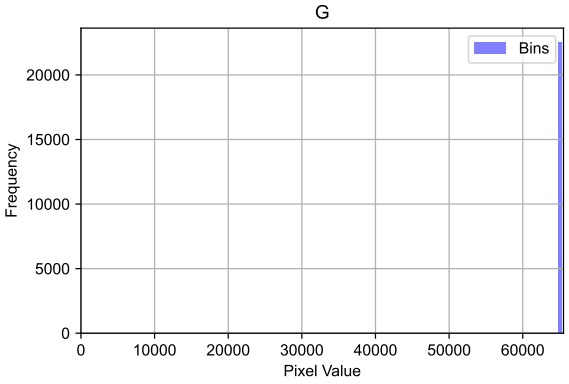


Figure S2: Example of different reflectance on calibrated MS imagery from preliminary flights in April, most likely caused by changing cloud coverage, leading to the decision to carry out flights again with more consistent exposure to ensure high-quality data.

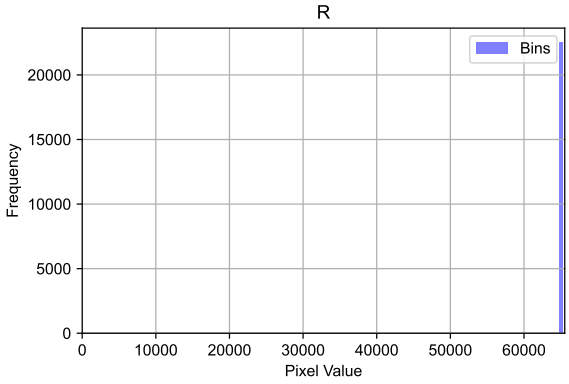
Calibration Panel Histogram Analysis



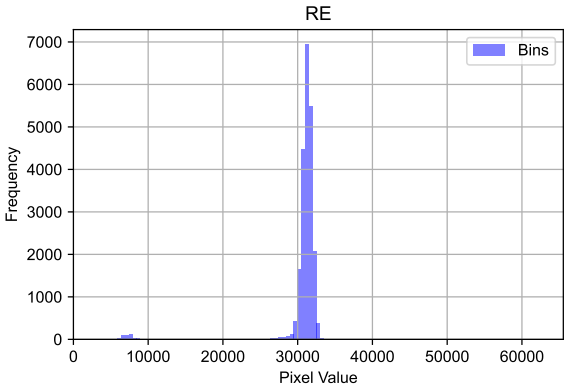
(a) Histogram area



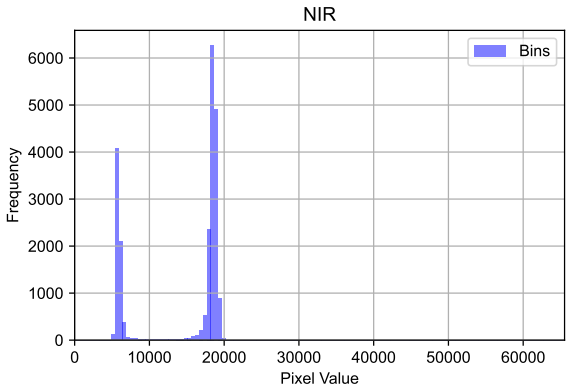
(b) G band, which is fully saturated.



(c) R band, which is fully saturated.



(d) RE band, which can be used for calibration.



(e) NIR band, which can be used for calibration.

Figure S3: Histogram of the CRP within the calibrated area (histogram is calculated only for the area in the red box). Bands G and R (**b**, **c**) are fully saturated and therefore cannot lead to proper calibration.

Image Metadata and Exposure Settings

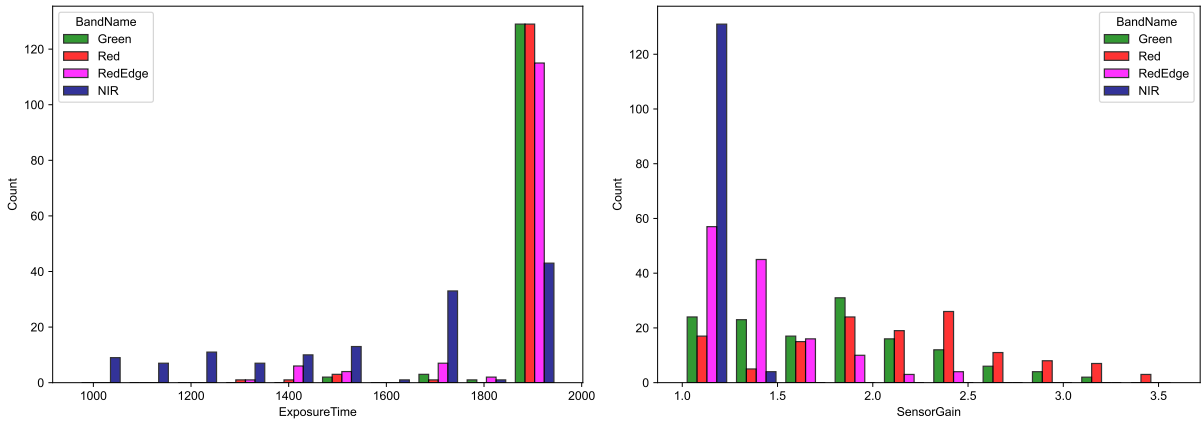


Figure S4: Metadata for site A2 showing exposure time and sensor gain, showing that the auto-settings of the M3M are inconsistent during a capture.

Model Performance and Hyperparameter Exploration

Learning Rate Example: 5×10^{-5}

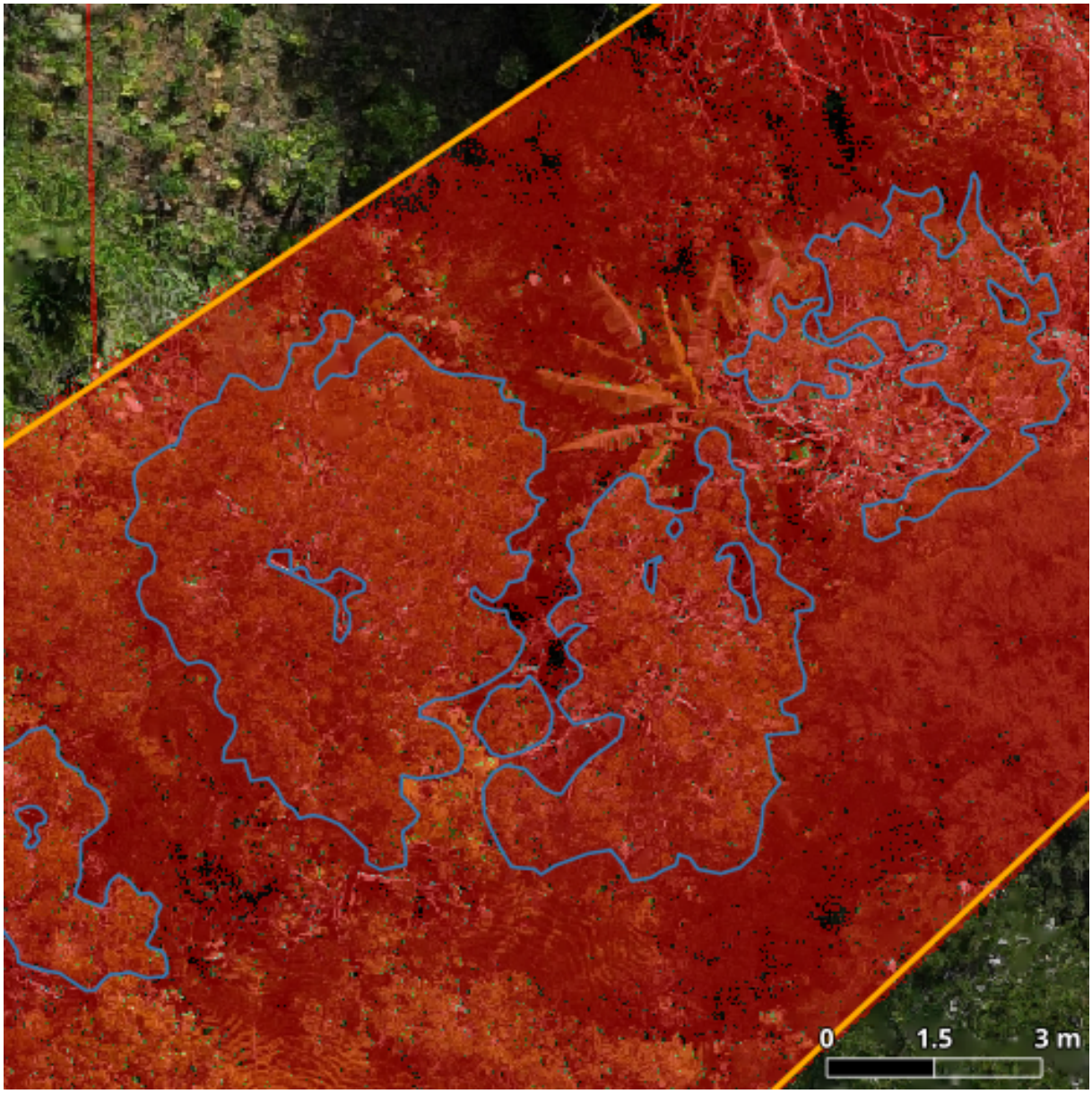


Figure S5: Example of random predictions from a model trained with LR $5x10^{-5}$. Orange shows the extent of the test area, blue are the *S. maire* annotations and red are the model predictions.

Training Dashboards

Predictions with Different Model Weights

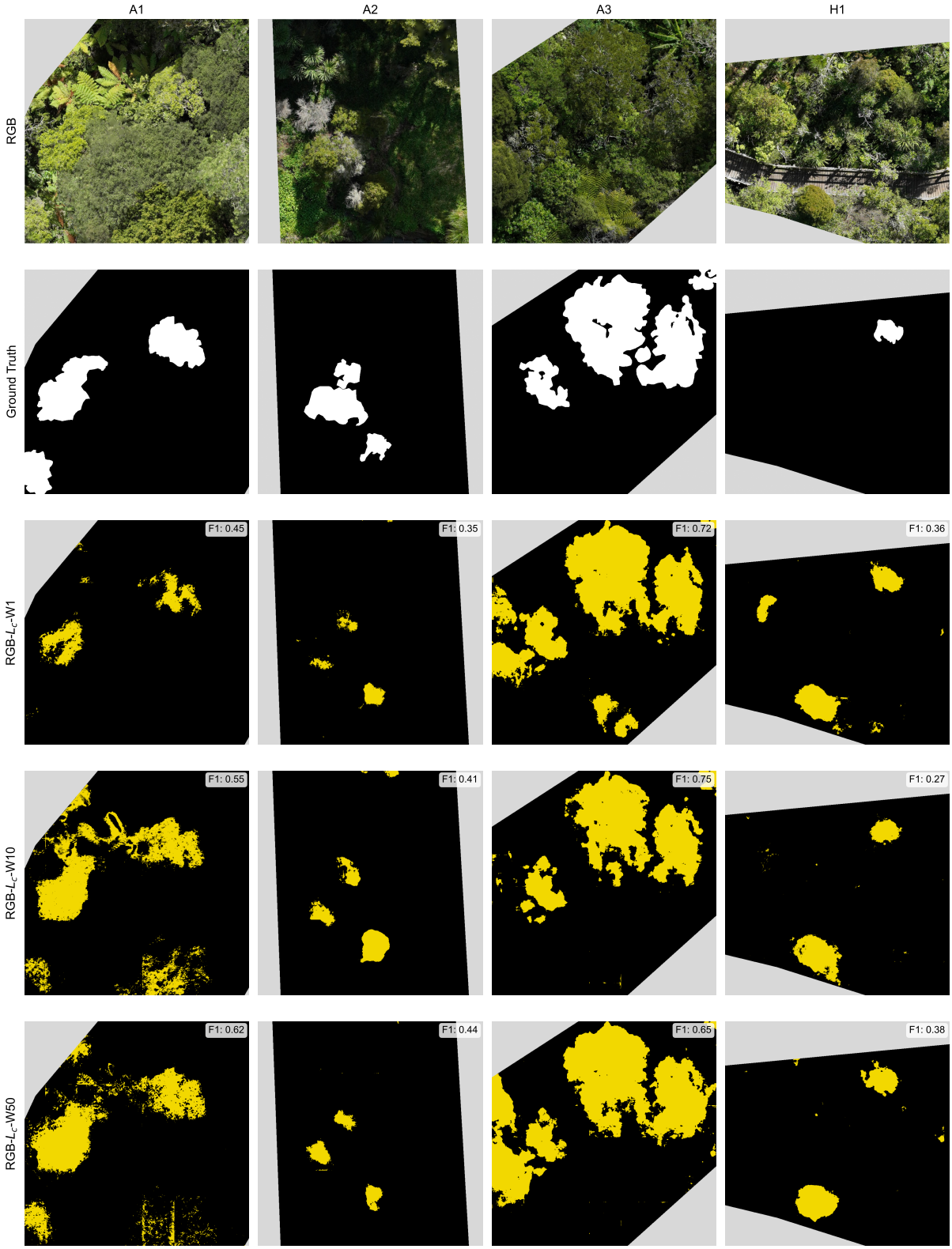


Figure S6: Predictions from the single-site models for the test zone for each reserve for a 15x15 m extent. Model configurations are listed on the left; all were trained with a LR of 0.02 and weights ranging from 1, 10 to 50. For example RGB-L_c-W50 reads as: "model trained on RGB bands using L_c with a weight of 50".

Train metrics single-site

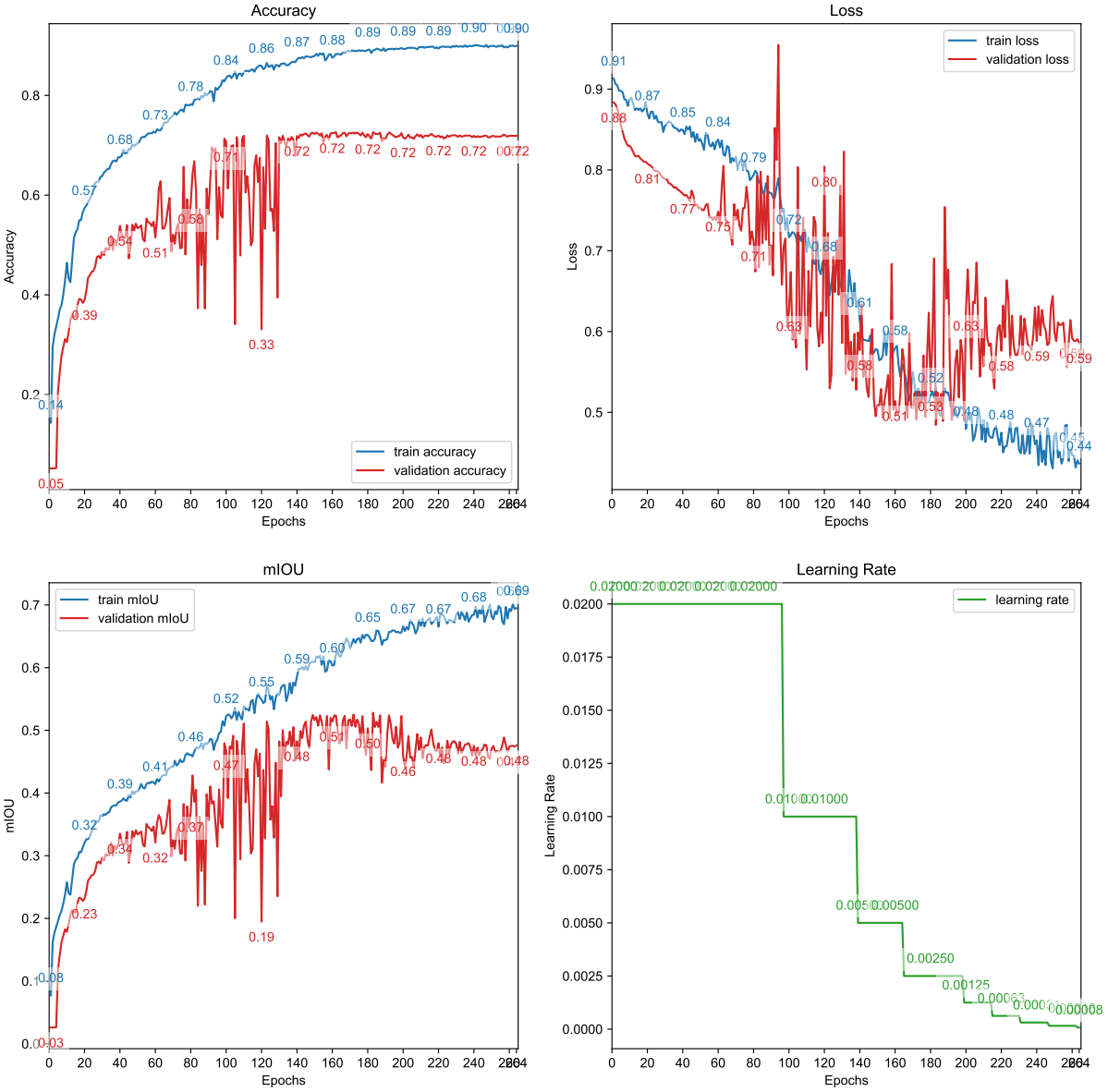


Figure S7: Train metrics from the best performing single-site model for site A2 trained on relatively calibrated MS data using L_{Dice}.

Train metrics multi-site

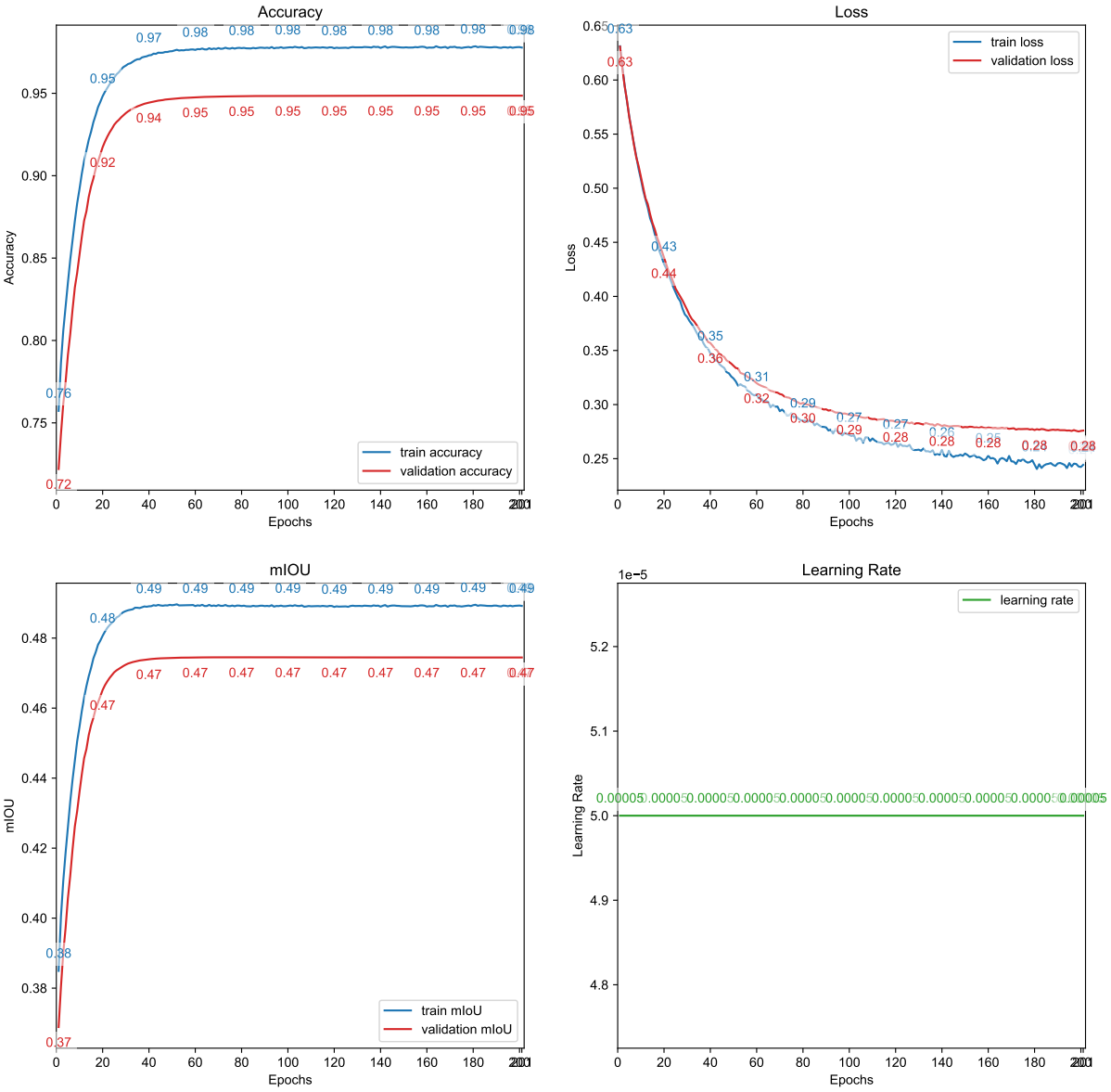
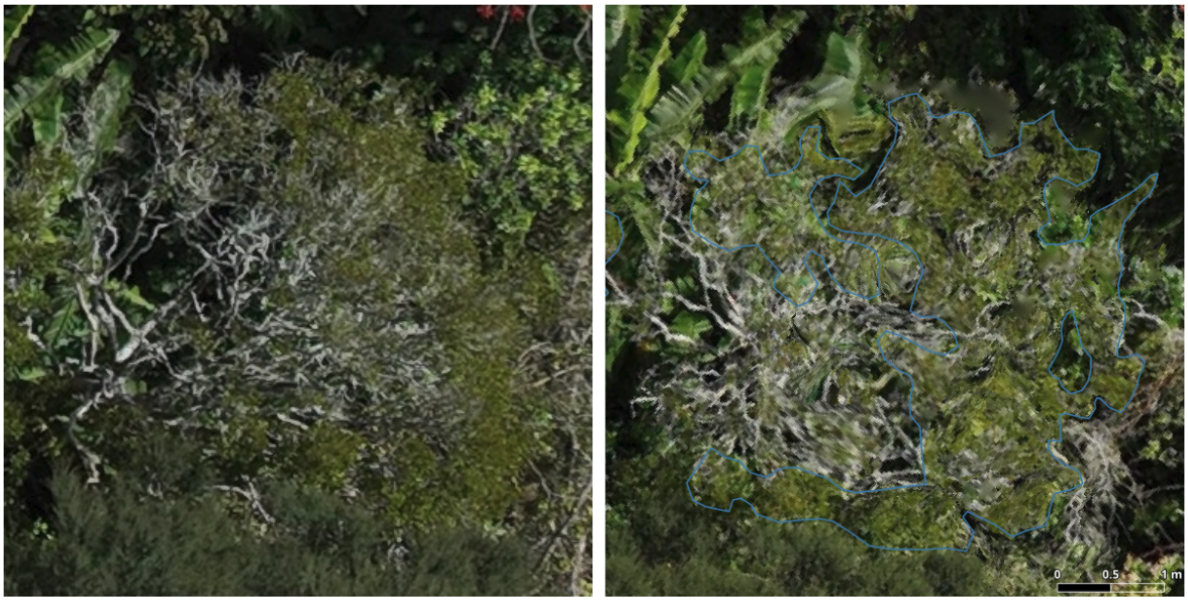


Figure S8: Train metrics from multi-site model trained on all sites with no training effect. Model was trained with a LR of $5x10^{-5}$.

Processing Artefacts



(a) Original unprocessed RGB image from the same extent.

(b) Processed RGB orthomosaic.

Figure S9: Example of artefacts on processed RGB orthomosaic (**b**), and the same extent extracted from the raw image (before processing) without artefacts (**a**).

Training Dataset Characteristics

Table S1: Dataset summary with class distribution of training (80%) and validation split (20%) showing tile counts and *S. maire* pixel percentages for each reserve.

Reserve	Train Images	Train Maire (%)	Val Images	Val Maire (%)
A1	26	3.34	8	3.94
A2	9	2.37	3	3.96
A3	21	5.64	6	6.21
H1	6	2.19	3	5.60
Combined	65	4.58	17	2.35

Best Performing Models per Site

Eskdale

Table S2: Overall best performing models for site A1.

Reserve	Bands	Loss	W	LR	F1	IoU	Prec	Rec
A1	MS_{rel}	L_c	50	0.02	0.65	0.48	0.52	0.87
A1	MS+IND_{rel}	L_c	10	0.02	0.64	0.47	0.51	0.87
A1	RGB	L_c	50	0.02	0.62	0.45	0.51	0.80
A1	MS_{rel}	L_{Dice}	1	0.02	0.61	0.44	0.70	0.55
A1	RGB	L_{Dice}	1	0.02	0.60	0.43	0.46	0.87
A1	RGB	L_{WBCE}	10	0.02	0.60	0.43	0.52	0.72
A1	MS_{rel}	L_{WBCE}	50	0.02	0.59	0.42	0.50	0.72
A1	MS_{rel}	L_c	1	0.02	0.57	0.40	0.62	0.52
A1	RGB	L_{WBCE}	50	0.02	0.56	0.38	0.41	0.86
A1	RGB	L_c	10	0.02	0.55	0.38	0.45	0.72

Kauri Glen

Table S3: Overall best performing models for site A2.

Reserve	Bands	Loss	W	LR	F1	IoU	Prec	Rec
A2	RGB	L_{Dice}	1	0.02	0.46	0.30	0.45	0.48
A2	RGB	L_c	50	0.02	0.44	0.28	0.65	0.33
A2	RGB	L_c	10	0.02	0.41	0.26	0.42	0.40
A2	RGB	L_c	1	0.02	0.35	0.21	0.67	0.24
A2	MS_{rel}	L_c	1	0.02	0.30	0.18	0.87	0.18
A2	RGB	L_c	50	$5x10^{-5}$	0.28	0.16	0.20	0.48
A2	MS_{rel}	L_{WBCE}	50	0.02	0.26	0.15	0.45	0.19
A2	RGB	L_{WBCE}	50	0.02	0.25	0.14	0.23	0.27
A2	MS_{rel}	L_c	50	0.02	0.25	0.14	0.39	0.18
A2	MS_{rel}	L_{WBCE}	10	0.02	0.23	0.13	0.85	0.13

Bushglen

Table S4: Overall best performing models for site A3.

Reserve	Bands	Loss	W	LR	F1	IoU	Prec	Rec
A3	MS_{rel}	L_{Dice}	1	0.02	0.81	0.68	0.74	0.88
A3	MS_{rel}	L_c	1	0.02	0.80	0.67	0.72	0.89
A3	RGB	L_{WBCE}	10	0.02	0.77	0.62	0.68	0.87
A3	RGB	L_c	10	0.02	0.75	0.60	0.74	0.76
A3	RGB	L_{Dice}	1	0.02	0.74	0.59	0.64	0.87
A3	RGB	L_c	1	0.02	0.72	0.57	0.62	0.86
A3	MS_{rel}	L_c	10	0.02	0.66	0.49	0.51	0.91
A3	RGB	L_c	50	0.02	0.65	0.48	0.52	0.86
A3	MS_{rel}	L_{WBCE}	10	0.02	0.64	0.47	0.55	0.77
A3	IND_{rel}	L_c	10	0.02	0.64	0.47	0.49	0.92

Hammond Park

Table S5: Overall best performing models for site H1.

Reserve	Bands	Loss	W	LR	F1	IoU	Prec	Rec
H1	MS_{rel}	L_c	10	0.02	0.59	0.42	0.44	0.89
H1	MS_{rel}	L_c	50	0.02	0.56	0.38	0.42	0.82
H1	MS_{rel}	L_{Dice}	1	0.02	0.53	0.36	0.39	0.83
H1	MS_{rel}	L_{WBCE}	10	0.02	0.48	0.32	0.32	0.94
H1	$MS+IND_{rel}$	L_c	10	0.02	0.46	0.30	0.38	0.59
H1	RGB	L_{Dice}	1	0.02	0.45	0.29	0.32	0.79
H1	IND_{rel}	L_c	10	0.02	0.42	0.27	0.28	0.80
H1	RGB	L_{WBCE}	10	0.02	0.40	0.25	0.31	0.54
H1	MS_{rel}	L_{WBCE}	50	0.02	0.39	0.24	0.25	0.94
H1	RGB	L_c	50	0.02	0.38	0.24	0.29	0.55

Best Multi-Site Models

Table S6: Overall best performing parameters for multi-site.

Bands	Loss	W	LR	F1	IoU
RGB	L_{Dice}	1	0.02	0.56±0.14	0.40±0.14
MS_{rel}	L_{Dice}	1	0.02	0.54±0.25	0.40±0.23
RGB	L_c	50	0.02	0.52±0.14	0.36±0.12
MS_{rel}	L_c	50	0.02	0.50±0.17	0.34±0.14
RGB	L_c	10	0.02	0.50±0.20	0.35±0.19

Multi-Site Models all Configurations

Table S7: The performance metrics from all model configurations trained on the multi-site dataset, sorted by F1 score.

Bands	Loss	Weight	F1	IoU	Prec	Rec
RGB	L_c	50	0.51	0.34	0.39	0.72
RGB	L_{Dice}	1	0.50	0.33	0.45	0.56
RGB	L_c	10	0.45	0.29	0.31	0.81
MS$_{rel}$	L_c	50	0.35	0.22	0.23	0.75
MS$_{rel}$	L_c	10	0.30	0.18	0.27	0.34
IND$_{abs}$	L_c	10	0.22	0.12	0.23	0.20
IND$_{abs}$	L_c	50	0.19	0.10	0.12	0.52
MS$_{rel}$	L_{Dice}	1	0.16	0.09	0.44	0.10
MS$_{abs}$	L_c	50	0.16	0.09	0.17	0.15
MS+IND$_{abs}$	L_{Dice}	1	0.16	0.09	0.11	0.31
MS+IND$_{abs}$	L_c	50	0.15	0.08	0.15	0.16
IND$_{abs}$	L_{Dice}	1	0.15	0.08	0.09	0.47
MS+IND$_{abs}$	L_c	10	0.14	0.08	0.25	0.10
MS$_{abs}$	L_c	10	0.14	0.08	0.20	0.11
MS$_{abs}$	L_{Dice}	1	0.09	0.05	0.19	0.06

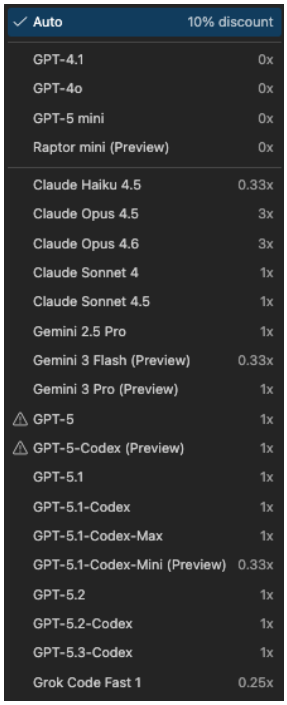
Appendix C: Reference Materials

Use of generative Artificial Intelligence

Throughout this Thesis, various AI tools were used to support literature research, data analysis and the writing process. Any output generated by generative AI tools was critically evaluated and edited by myself to ensure accuracy, coherence, and alignment with the scientific content of the thesis. It was treated as a highly efficient assistant for small well-defined tasks, whose outputs were considered potentially wrong until verified and adjusted by myself. The following sections provide a detailed account of the specific AI tools used and how they have been applied.

GitHub Copilot

GitHub Copilot was used both in the data analysis and writing process. It was set to auto mode, so the underlying model was automatically selected to best suit the given prompt. Available models include:



In the data analysis process, it was used to generate code snippets for data processing, visualisation, and analysis, and to navigate through my code space to identify relevant code sections within the convoluted U-Net module. Any provided code was thoroughly tested, adjusted where needed, and data visualised to ensure it worked as intended and produced accurate outputs.

In the writing process, personally created bullet points were discussed to point out potentially missing arguments or adjust placement under subsections. Furthermore, it was used to check

grammar and spelling, suggest synonyms, and, in some cases, improve sentence structure and phrasing.

Finally, given the nature of this thesis which was written and formatted using the Markdown language Quarto, Github Copilot was also used to help with formatting and to generate code snippets for specific formatting needs (e.g. tables, figures, or specific layout requirements).

Scite

Scite was connected to my Zotero library and used as a tool to point me to original sources within my own library. No text generated by Scite was used.

DeepL

Sentences with arguments from studies were translated to German with DeepL and paraphrased by myself in a mix of English and German words. These hybrid sentences were then translated back to English with DeepL and further adjusted if needed. This mainly included exchanging words with synonyms.

Microsoft Copilot and ChatGPT

ChatGPT and Copilot were used to generate search terms to scan scholarly databases for literature research.

LanguageTool paraphrasing tool

The rephrasing tool was scarcely used on single sentences to improve wording and sentence structure.

Reprint Permissions

Advertising & Illustrative Photographers Association

AIPA

Standard Photographic
Licence & Order Confirmation

Licence granted to: (Agency/Client) Robin Pfaff kqn7759@autuni.ac.nz Student of Robert Beresford	Photographer: Bioeconomy Science Institute Licence Number: BSI-500 Date of Licence: 09 February 2026
Advertiser: (if any)	Return Date: NA
Product: (if any)	☐ Exclusive Licence ☑ Non-Exclusive Licence

DESCRIPTION OF JOB/PHOTOGRAPH(S) COVERED BY THIS LICENCE

Myrtle Rust- Native Swamp Maire Tree _PFR2541.jpg
Myrtle Rust- Native Swamp Maire Tree _PFR2575-2.jpg

Please Note: This is not a commission. Please read clause 1 of the Standard Photographic Terms and Conditions of Engagement [v2006] available from www.aipa.org.nz

LICENSED PURPOSE

Non-Advertising	Advertising	
☐ Editorial (newspaper, magazine, periodical)	☐ Magazine/Newspaper/Periodical	☐ Billboard
☐ PR (Public Relations)	☐ Poster	☐ Catalogue
☐ Press (news)	☐ Brochure	☐ Point of Sale
☐ Book (single publication)	☐ Packaging	☐ TV/Cinema Commercial
☐ Any Non-Advertising Use	☐ Promotion	☐ Any Advertising Use
☑ Other Use (specify) For use in Robins thesis	☐ Internet (specify URL)	
☐ Internet (specify URL)	☐ Other Advertising (specify)	

LICENSED TERRITORY

☐ New Zealand	☐ Australia	☐ Australasia	☑ Worldwide	☐ Other (specify)

LICENSED PERIOD

☐ 1 year	☐ 2 years	☐ 5 years	☑ In Perpetuity	☐ Other (specify)

OTHER LICENCE DETAILS

☐ Special conditions of licence: (if any)

☐ No right to alter image (see clause 1.1.3)	☐ The Photographer waives attribution rights (see clause 4); or
☐ Photographer to act as agent (refer to clause 13.3)	☑ Attribution required as follows: Bioeconomy Science Institute
☐ Weather permitting booking (refer to clause 17.1)	

IMPORTANT NOTE

The Photographer has accepted this order on Standard Photographic Terms and Conditions of Engagement as previously supplied to you. Please read and check this order confirmation carefully as the work will proceed according to these instructions. Please request our Standard Photographic Terms and Conditions of Engagement [v2006] if you require a further copy.

Reprint permission for Figure 3a.