UNIVERSITY OF CALIFORNIA SAN DIEGO

Remote Sensing of Mangroves using Machine Learning based on Satellite and Aerial Imagery

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Electrical Engineering (Machine Learning & Data Science)

by

Stanley Dillon Hicks

Committee in charge:

Professor Curt Schurgers, Chair
Professor Ryan Kastner
Professor Karcher Morris

2023

The Thesis of Stanley Dillon Hicks is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my family and friends who have always supported me,

And to my father and sister who were able to see my journey, but not the destination

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Vegetation Index
UAV	Unmanned Aerial Vehicle
P4P	Phantom 4 Pro
GIS	Geographic Information System
GMW	Global Mangrove Watch
RTK	Real-time Kinematic positioning
GCP	Ground Control Point
ERT	Extremely Randomized Trees
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DEM	Digital Elevation Model
IoU	Intersection over Union
LR	Low Resolution
HR	High Resolution
BCS	Baja California Sur

ACKNOWLEDGEMENTS

Thank you to Dr Curt Schurgers and Dr Ryan Kastner for providing the opportunity to work on this project for so many years though the Engineers for Exploration program. I would also like to thank past and present members of the Aburto lab for providing expertise, insight, and labelers to accomplish these surveys.

ABSTRACT OF THE THESIS

Remote Sensing of Mangroves using Machine Learning based on Satellite and Aerial Imagery

by

Stanley Dillon Hicks

Master of Science in Electrical Engineering (Machine Learning & Data Science)

University of California San Diego, 2023

Professor Curt Schurgers, Chair

Mangrove forests are critical to mitigating climate change and provide many essential benefits to their ecosystems and local environments but are under threat due to deforestation. However, monitoring mangroves through remote sensing can help pinpoint and alleviate the causes of their deforestation. Machine learning can be used with remotely sensed low-resolution satellite or high-resolution aerial imagery to automatically create mangrove extent maps with higher accuracy and frequency than previously possible. This study explores and offers

recommendations for two practical scenarios. In the first practical scenario, where only low-resolution hyperspectral satellite imagery is acquired, we implemented several classical machine learning models and applied these results to data acquired in the Clarendon parish of Jamaica. We found that utilizing extensive feature engineering and hyperspectral bands can result in strong performance for mangrove extent classification, with an accuracy of 93% for our extremely randomized trees model. In the second practical scenario, we explored when there is full coverage of both low-resolution satellite and high-resolution aerial imagery over a survey area. We created a hybrid model which fuses low-resolution pixels and high-resolution imagery, achieving an accuracy of 97% when applied to a dataset based in Baja California Sur, Mexico, offering another high-performance method to automatically create mangrove extent maps if both high- and low-resolution imagery is available. Overall, the methods tested over these two scenarios provide stakeholders flexibility in data and methods used to achieve accurate, automatic mangrove extent measurement, enabling more frequent mangrove monitoring and further enabling the protection of these important ecosystems.

1. Introduction

1.1. Background

Mangroves are a type of highly important tree species that live in tropical areas across the world. They live in the intertidal zones of many coasts, being able to uniquely live in brackish water where many other categories of trees are unable to live. The unique environment that mangroves live in limits the area that they can inhabit, leaving them to be contained within a relatively small area compared to the vast forests in other parts of the world (Spalding, 2010). Despite mangrove forests' small relative size, they offer many benefits to both the world and their local areas. Mangroves are exceptional at sequestering carbon out of the atmosphere (Donato, et al., 2011), as per cubic meter, they can store approximately twice as much carbon in their branches and roots when compared to tropical forests (Tomlinson, 2016). They are also very important to the economies of the local areas they inhabit, providing protection from coastal storms and supporting food production in the form of fisheries (Alongi, 2002) (Das & Vincent, 2009). Due to these reasons, a single hectare of mangrove forest can provide over $50,000 per year in value to local economies (Mukherjee, et al., 2014).

Even though Mangrove forests provide a boon to the local economies in the areas they inhabit and are exceptional at fighting the causes and effects of climate change, they are at risk due to many factors. Such factors include deforestation from agricultural development, encroaching coastal cities (Richards & Friess, 2016), and damage from tropical storms (Thomas, et al., 2017) which are made increasingly more powerful due to climate change (Bhatia, Vecchi, Murakami, Underwood, & Kossin, 2018). Due to these factors, 20-35% of global mangrove

extent has been lost over the last 50 years (Goldberg, Lagomasino, Thomas, & Fatoyinbo, 2020), and policymakers across the world have created goals to reduce mangrove deforestation.

1.2. Mangrove Remote Sensing

1.2.1. Low-Resolution - Satellites

For policymakers to make informed decisions on how to best conserve mangrove ecosystems, mangrove extent is often used as a direct measure of mangrove health. Many scientists have turned to remote sensing methods for deriving mangrove extent, deeming in-situ surveys too labor-intensive (Pham, Yokoya, Bui, Yoshino, & Friess, 2019). Remote sensing involves “detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance” (U.S. Geological Survey , n.d.). In the case of mangrove remote sensing, light inside and outside of the visual spectrum is captured through images taken above the mangrove canopy.

One image source commonly used to remotely monitor mangroves' extent is low-resolution satellite imagery, which offers many advantages compared to traditional in-situ surveys (Purnamasayangsukasih, Norizah, Ismail, & Shamsudin, 2016). Satellite imagery is easy to acquire and has a low cost, with providers such as SkyWatch providing imagery in the ranges of $0 - $30 per kilometer (SkyWatch, n.d.). This imagery can be captured as frequently as daily for most areas of the world, allowing vast areas of mangrove ecosystems to be captured with little effort, enabling remote sensing scientists to derive extent maps with less resources. Also, modern satellite imagery sources have the capability to schedule the recording of imagery in the future or access the vast archives of already recorded imagery to see long term changes in mangrove extent (Planet, n.d.).

Satellites also have the advantage that they are often able to also capture data in a variety of spectral bands, outside of the visible panchromatic spectrum. For mangrove remote sensing in particular, light from hyperspectral bands such as Near-infrared (NIR) are reflected from mangroves differently than nonmangrove areas. These hyperspectral bands can be utilized to generate additional important features that can give clues to determining mangrove extent (Vaiphasa, Ongsomwang, Vaiphasa, & Skidmore, 2005). Such features include the Normalized Difference Vegetation Index (NDVI) (Pettorelli, et al., 2005) and Normalized Difference Water Index (NDWI). As illustrated in Figure 1, these features can help detect vegetation and water in hyperspectral imagery (Gao, 1996), further aiding in the process of classifying mangrove extent (Valderrama-Landeros, Flores-de-Santiago, Kovacs, & Flores-Verdugo, 2017).

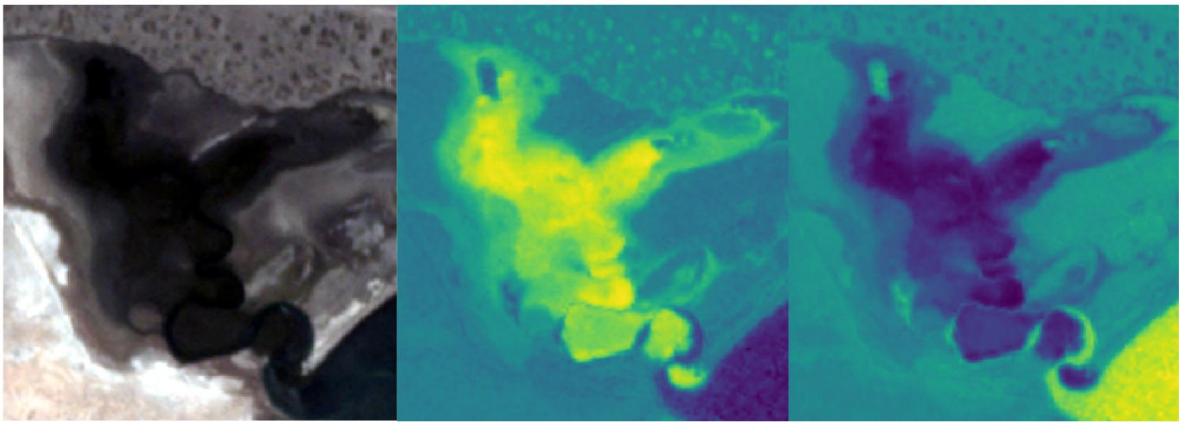

Figure 1 - *Example Planetscope satellite imagery showing original imagery and calculated NDVI and NDWI Features*

While satellites have become a primary tool in a remote sensing scientist's kit to calculate mangrove extent, they do have some disadvantages. One disadvantage when using visual imagery is that cloud cover can pose a significant challenge to satellite image acquisition as clouds can cover areas of interest. This forces remote sensing scientists to wait until there is little

or no cloud cover such that useful imagery can be captured. Another major disadvantage is their low spatial resolution, or the large geographic area represented by a single pixel in their imagery. Some commonly used satellites such as Planetscope and Sentinel-2A have a spatial resolution of ~3m and ~10m, respectively (The European Space Agency, n.d.). While one may be able to discern between vegetation and non-vegetation easily at such resolutions, it is virtually impossible to visually distinguish between mangrove and nonmangrove vegetation with visual-spectrum imagery. This is because as important details such as leaf shape and texture, which can be used to visually distinguish mangrove from nonmangrove vegetation are lost, as shown in Figure 2. To compare the resolution and spectrum count, we have provided a list of commonly used satellites below in Table 1.

Table 1 - *Table containing commonly used satellite names, resolution, number of bands, and launch year (first year where data is available)*

Name	***Resolution***	***# of Bands***	***Launch Year***
Planetscope Dove	3m	4	2016
Planetscope Superdove	3m	8	2021
Sentinel 2A	10m	13	2015
WorldView 2	1.84m	8	2009
Pleiades	0.5m	5	2011

Although there are some higher resolution satellites available, such as the Pleiades constellation at a resolution of 0.5m (Airbus Intelligence, n.d.), high-resolution satellites are much more costly than their lower resolution counterparts, relegating some scientists to lower resolution image sources such as Planetscope and Sentinel 2A. To solve this issue of low spatial

resolution, we have investigated utilizing extensive feature engineering of our input satellite imagery to create accurate mangrove extent maps, which will be discussed in Chapter 4.

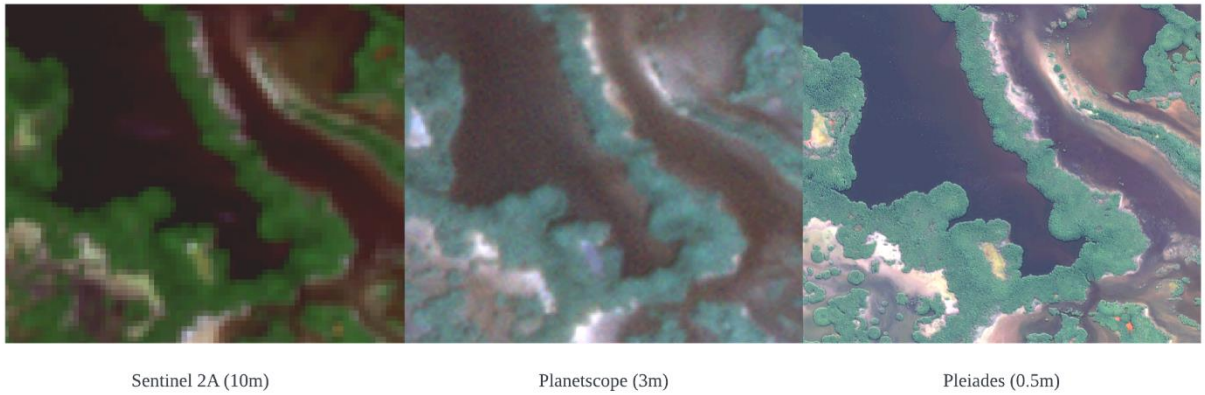

Figure 2- *Example imagery of commonly used satellites*

1.2.2. High-Resolution - Drones

As a popular alternative to low- to moderately low-resolution imagery sources such as satellites, high-resolution aerial imagery can be acquired. Aerial imagery can be most commonly acquired through remotely controlled unmanned air vehicles (UAVs), often referred to as drones. (Tang & Shao, 2015) (Dash, Watt, Pearse, Heaphy, & Dungey, 2017). Drones can fly over an area of interest, capturing many images which are then stitched together into a single high-resolution image using orthorectification. This high-resolution imagery proves to be drone's main advantage when compared to satellites. While satellite imagery spatial resolution is often measured in the magnitude of meters, high-resolution imagery is measured in centimeters. In fact, despite their low cost, standard consumer drones such as the DJI Phantom 4 Pro (P4P) can record at a spatial resolution of 3cm at an altitude 120m as highlighted in Figure 3, which compares satellite and drone imagery of the same area. High-resolution imagery can be used to great effect when measuring mangrove extent, as leaf shape and texture can be utilized when

creating mangrove extent maps. With lower resolution, visual spectrum imagery, it can be difficult or impossible to visually distinguish mangrove from nonmangrove using imagery in the visual spectrum, as illustrated in Figure 3. Lastly, high-resolution sources can give remote sensing a lot of flexibility in surveys, as with drones, scientists can customize what drones to use along with their payload, allowing survey teams to cater their equipment to differing needs in a survey. With this, UAVs can collect images in more targeted way, as without impediments to flying, UAVs can fly and record imagery at any time, where tasked satellites must wait until they fly over their survey area to record imagery.

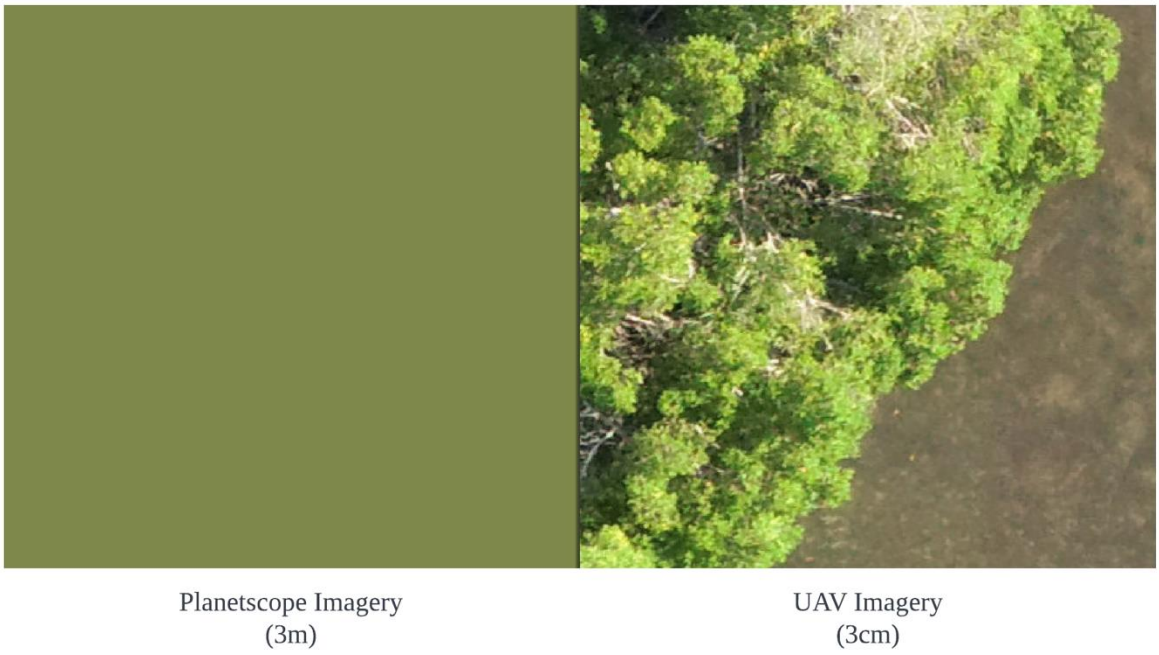

Figure 3 - *Example Planetscope and UAV imagery highlighting differences in resolution*

On the other hand, while drones have advantages that make them a popular alternative to satellite imagery, high-resolution image sources, most notably drones, have some downsides that prevent them from replacing satellites altogether. For one, the cost of drones themselves and their maintenance can be quite high, with costs increasing to the tens of thousands in USD for specialized hyperspectral cameras, such as the MicaSense RedEdge, if imagery outside of

panchromatic bands is required. While hyperspectral bands are desirable to have when measuring mangrove extent, they may be unobtainable to many teams with limited resources. In addition, since drones need operators, the cost of transportation and compensation for the survey team can add additional costs.

Also, while the acquisition of satellite imagery is only limited by cloud cover and atmospheric conditions, the acquisition of drone imagery can be impacted in more severe ways. For the best image quality, drone must record images under consistent lighting with limited reflections. In the worst case, drones must be grounded in case of strong winds or rain to prevent damage to their sensitive electronics, extending the time and cost it takes to record imagery over a survey area. Lastly, some survey areas may be inaccessible to drones due to a lack of proper takeoff and landing locations, making image acquisition by drones unfeasible. For these reasons, the acquisition of drone imagery may be impossible in such circumstances.

Because of these inherent strengths and weaknesses between satellites and drones, there is no best method for measuring mangrove extent. Instead, high-resolution image surveys can serve to ‘fill the gap’ between in-situ surveys and satellite image surveys, allowing remote sensing scientists to get more finer-grained details of mangroves and their extent. Because of these reasons, remote sensing scientists have a lot of flexibility when using remote sensing imagery to determine mangrove extent. In this work, we will consider two modalities of remote sensing data: solely low-resolution imagery with satellites, or a combination of low- and high-resolution data from satellites and drones and consider automated methods to measure mangrove extent.

1.2.3. Machine Learning

Once imagery of a mangrove survey area has been recorded, the question can then be posed: how to measure its extent? The common answer to this question is to manually annotate this imagery, using Geographic Information System (GIS) Software, such as QGIS, to painstakingly draw georeferenced polygons over every pixel in our imagery. This effort needed to label is increased when annotating drone imagery, as each individual polygon drawn can be minuscule in comparison to the total area of a survey site. Due to this effort, generating mangrove extent maps manually can take 1000s of person hours, which may be impractical in the long term when continuous monitoring is needed.

To limit the amount of laborious manual labeling, machine learning can be utilized to derive mangrove extent when there are labels are only created for a portion of a survey area, as machine learning is widely used in remote sensing for similar tasks (Lary, Alavi, Gandomi, & Walker, 2016). Using a relatively small amount of annotated mangrove and nonmangrove pixels, a machine learning model can be trained on pairs of reference mangrove extent labels and imagery to automatically classify new mangrove imagery. Such a machine learning model learns from these pairs of data what is and is not mangrove, without explicit programming. This automatic classification allows mangrove extent maps and statistics to be generated quickly and more often, giving policymakers frequent information to make more informed decisions on mangrove conservation.

Machine learning has been used with great success in previous studies to derive mangrove extent. Most notably, with Global Mangrove Watch (GMW), Bunting et. al have utilized a machine learning model to complete a previously unthinkable task, deriving an estimated mangrove extent map for the entire world using low-resolution hyperspectral satellite

imagery (Bunting, et al., 2018). While these global maps provide the area of labels needed for worldwide decision making, their low-resolution and accuracy of classifications in certain areas (Hsu, et al., 2020) and once per year label updates can leave local policymakers with inaccurate, outdated mangrove extent maps at a local scale. If machine learning methods that leverage more frequent and higher resolution drone and satellite imagery are used instead, fine-tuned mangrove conservation policies based off more recent information can be created, allowing for more frequent long-term monitoring.

Therefore, in this study, we will investigate and implement a selection of methods used to classify mangrove extent that can be used to provide higher resolution and more accurate mangrove extent maps on a local scale. We will test these methods with datasets based on two different and real survey scenarios, where one dataset only has complete low-resolution satellite imagery, while the other dataset contains complete drone imagery and satellite imagery. We aim to answer the following two questions –

1. **For the two cases of data acquired during a local survey - full high-resolution and low-resolution imagery coverage, or just low-resolution imagery coverage, what machine learning architectures should one use to generate accurate classifications for each scenario?**
2. **For each scenario, how do these recommended machine learning architectures perform, and how can each model be used to classify mangrove extent for entire future survey areas?**

To answer these questions, we will examine the two scenarios separately, discussing their background and implementations for two different machine learning methods that we will use for each scenario: classical methods and hybrid models. Next, we will test both models and their

respective baselines for both scenarios separately and discuss the results of each scenario. Using these results, we will then finally give recommendations for how to use these methods in future mangrove extent surveys. Our contribution is thus to develop practical approaches to study mangrove extent, based on machine learning models trained on to imagery obtained from satellite and drones.

2. Background: Machine Learning

In this section, we discuss the background and motivation behind the two categories of machine learning models used to automate the estimation of mangrove extent: classical machine learning and deep learning.

2.1. Machine Learning: Background

Many of the models used in remote sensing to accomplish classification and regression tasks can be categorized in two classes, classical and deep learning. Classical methods were created before the advent of deep learning models, but this does not mean that classical models have poor performance. These classical models have roots as far back as the 19th century, with linear regression serving as the foundational method for statistics and machine learning. Since the 19th century, classical methods have greatly increased in complexity and performance, allowing them to have strong performance and see usage on complex datasets and applications. Classical methods have been used widely in remote sensing, being applied in problems such as cloud cover classification (Bai, Li, Sun, Chen, & Li, 2016), predicting urban ozone concentrations in urban areas (Yi & Prybutok, 1996), and glacier detection (Brenning, 2009). Most importantly, classical methods have been utilized widely in mangrove extent classification of satellite imagery (Liu, Li, Shi, & Wang, 2008) (Valderrama-Landeros, Flores-de-Santiago,

Kovacs, & Flores-Verdugo, 2017) (Wang, et al., 2018), with the notable example being GMW (Bunting, et al., 2018).

2.1.1. Decision Trees

A commonly used classical method in remote sensing that we can use to classify mangrove satellite imagery is a decision tree. Decision trees form predictions in the form of a binary tree, where one could traverse this tree using features of a sample, with each node guiding the traversal through decision rules. For example, each node can be simplified as an if statement, where one could traverse this tree using feature values to create an output prediction at its leaves. Using Figure 4 as an illustration to motivate how a decision tree may function when operating on satellite imagery, we can traverse this example tree using a set of Boolean statements. For example, to arrive on the prediction "Mangrove", our input sample must meet meets the following decision rules in order, `Green Value == High`, `NDVI Value == High`, `NDWI Value == Low`. The feature chosen at each split is the feature that produces the best separation between samples of the dataset in each of the two nodes. From this, a decision tree can be trained through a recursive method called recursive partitioning, where splitting ends when all samples at a leaf consist of the same class in the case of classification, or when splitting no longer improves performance of the model in the case of regression.

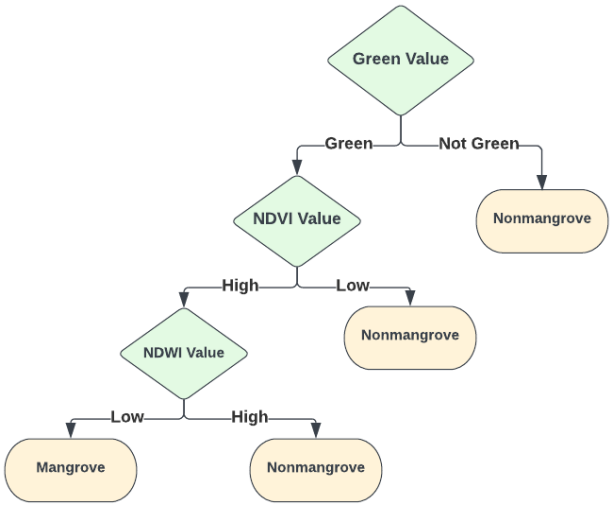

Figure 4 - *Example decision tree to classify a satellite imagery pixel*

Because of their structure, decision trees have several advantages. First, they are invariant to scaling and transformation of input feature values, meaning that data needs little preprocessing or rescaling of input features to achieve good predictions. In addition, they are robust to unneeded features, meaning that explicit feature selection is not needed to remove features that have little or negative effect on model performance. Due to these advantages, decision trees, like many other classical methods, have been employed widely in remote sensing with direct success even in mangrove classification, as for example, Liu, Li, Shi, and Wang utilized Landsat satellite data to track forest changes (Liu, Li, Shi, & Wang, 2008).

2.1.2. Bias, Variance, and Ensembling

Although decision trees can perform well on certain datasets, they do have some downsides that can make them not ideal in certain cases. For one, if nodes of decision trees are allowed to split too deep, leaves of the decision tree can represent very specific cases of input features. This results in strong performance when predicting on samples from the training dataset but could result in poor performance when predictions are generated from outside the training

dataset. This relation of high training performance, but low generalization can be modeled mathematically, where the prediction of y can be modeled in the form of a function such that

$$\varepsilon = y - \hat{f}(x).$$

where x are the input samples, $f(x)$ gives a value, or label for x, $\hat{f}(x)$is the predictive model for f, and ε is error in the prediction. We then decompose the squared expected error, such that

$$E\left[\left(y - \hat{f}(x)\right)^2\right] = (\,\mathrm{E}\left[\hat{f}(x)\right] - f(x)\,)^2 + \mathrm{E}[(\hat{f}(x) - \,\mathrm{E}\left[\hat{f}(x)\right]\,)^2] + \,\sigma^2$$

We can define $\mathrm{E}\left[\hat{f}(x)\right] - f(x)$ as bias - how close a model can predict the ground truth, $\mathrm{E}\left[\left(\hat{f}(x) - \,\mathrm{E}\left[\hat{f}(x)\right]\,\right)^2\right]$ as variance - how close the predictions over all of x is to each other, and σ^2 as some irreducible error. This phenomenon is called the bias-variance tradeoff (Kohavi & Wolpert, 1996) (Probst, Wright, & Boulesteix, 2019), where the variance in a model can be reduced by increasing its bias, and vice-versa, establishing some lower limit in error that models can achieve. Since decision trees can overtrain on certain samples in the dataset, decision trees have high bias, which greatly limits their model performance.

Therefore, using principles from the bias-variance tradeoff, the random forest model was introduced (Breiman, 2001) to improve on the weaknesses of decision trees. Random forests are a type of ensemble model, where the predictions of weak models are aggregated to create strong model. In other words, random forest combines many decision trees and aggregates their results to offer stronger performance. Each tree in this random forest is unique since the features chosen during a split are not selected from all features available, but from a random selection of features in the dataset, with replacement (Probst, Wright, & Boulesteix, 2019). This randomness is intensified with bootstrap aggregating (bagging), where each tree is constructed from a random selection of samples in the dataset, with replacement. Due to this randomness, each decision tree

has a high variance in prediction, and therefore low bias, helping with model generalization. In addition, since the output of each tree is averaged out, the aggregate output can lower bias as well since overtrained trees can be 'voted out'. For these reasons, random forests remain a popular choice for many remote sensing tasks (Belgiu & Drăguţ, 2016).

To improve on some of the weaknesses of Random Forest, we also implement the Extremely Randomized Tree (ERT) algorithm, the main model of GMW (Bunting, et al., 2018). ERT improves on Random Forest by decreasing model variance, but increasing bias, by not using bagging and using the entire dataset when constructing each tree. In addition, splits are decided from a random selection of features rather than a random selection. Because splits are randomized, ERT is trained much faster than random forest, allowing for faster prototyping of model parameters, while still having similar performance to random forest.

2.1.3. Implementing Classical Models

When implementing classical models to remote sensing applications, we can get some clues as to how they are so well suited to the task of satellite pixel classification. One clue is how both classical methods and satellite imagery are structured. Classical models typically input each sample i in the form of a vector, x^i, where each element represents the value of some feature in that sample. The classical model can be defined as a function, $\hat{f}(x)$, to predict some real value or vector $\hat{y_i}$. The classical model can then be represented in the form of, where f and $\hat{f}$ are an $\mathbb{R}^n \rightarrow \mathbb{R}^m$ mapping such that

$$\hat{y_i} = \hat{f}(x^i), \qquad \hat{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

This structure can be intuitively used with low-resolution satellite imagery, where $\mathbb{R}^n$ vectors representing single satellite pixels and additionally engineered features can be directly input to the classical model, with outputs of size $\mathbb{R}^m$ as predictions in the form of

regressions and general classifications or $\mathbb{R}$ for binary classifications. With this, a $\mathbb{R}$ output value representing the probability of the pixel being in the mangrove class can be used. We can use these single pixels classical models since little is gained from spatial information in satellite imagery, as spatial features visible in drone imagery, such as leaf shape and texture, are lost with the relatively low-resolution of satellite imagery. Thus, each pixel in an image can be classified individually, not using other surrounding pixels to guide our classification. While the lack of spatial features can result in poor performance with standard panchromatic imagery, hyperspectral bands and engineered features can enable strong performance without the clues present in higher resolution imagery. **Thus, we can test and use classical models for datasets that only have full low-resolution hyperspectral satellite imagery if high-resolution drone imagery is not available.** On the other hand, other methods such as deep learning that can take advantage of these input features when full high-resolution imagery coverage is present can be used.

2.2. Deep Learning and Convolutional Neural Networks

2.2.1. Artificial Neural Networks

Although the prevalence of deep learning is a relatively new phenomenon, the underlying technology of deep learning has existed since the turn of the century in the form of artificial neural networks (ANNs), a classical method (Abiodun, et al., 2018). ANNs, also known as multilayer perceptrons (MLPs), work through the connection of neurons which act as nonlinear functions. Inputs to a neuron are multiplied by trained weights which model the relationship between input features and neurons, and then the sum of these weighted inputs are fed to a nonlinear activation function. These neurons are often organized in fully connected dense layers,

where the inputs to a single neuron are the outputs of all neurons in the previous layer. Thus, from an input layer that has no weights or activation function, information is fed forward through the network and is distributed between neurons through the weighted connections and nonlinear activation functions. Training is done in the opposite direction through backpropagation, where the gradient is taken over each layer starting from the output of the ANN, working backwards to the input. Once an ANN is trained in this way, the nonlinearity of these trained layers allows ANNs to approximate essentially any nonlinear function, extending to any application where there is an $\mathbb{R}^n \rightarrow \mathbb{R}^m$ nonlinear mapping between input and output vectors.

ANNs can be useful remote sensing in the form of satellite pixel classification (Atkinson & Tatnall, 1997), where we can approximate the class of a pixel through a nonlinear function of an input satellite pixel and a number or vector representing classifications (Gardner & Dorling, 1998). However, because ANNs are inherently $\mathbb{R}^n \rightarrow \mathbb{R}^m$ nonlinear mappings, they can struggle when classifying whole square images such as high-resolution drone imagery, instead of single pixels. This is because when transforming a 3 dimensional square $\mathbb{R}^{i \text{ x } j \text{ x } k}$ image, where i, j, k represents the width, height, and number of bands respectively, to be compatible with the flat inputs format of an ANN, the image must be flattened as well, and all implicit information in the image, such as texture and relational position of pixels, is lost in the $\mathbb{R}^{ixjxk} \rightarrow \mathbb{R}^{i*j*k}$ transformation.

2.2.2. Convolutional Neural Networks

To effectively classify whole images in the form of a $\mathbb{R}^{ixjxk}$ matrix, Convolutional Neural Networks (CNNs) are very powerful when applied to images, as they improve on many of the weaknesses of ANNs. CNNs use spatially invariant transformations in their convolutional layers to extract implicit image features, thereby forming a $\mathbb{R}^{ixjxk} \rightarrow \mathbb{R}^m$ nonlinear mapping of input

images to output features. AlexNet was one of the first major implementations of the CNN (Krizhevsky, Sutskever, & Hinton, 2012), raising the bar for benchmarks of the ImageNet image classification dataset (Deng, et al., 2009). AlexNet improved on many existing image classification methods through creating an effective CNN architecture utilizing convolutional and pooling layers. Convolutional layers use a convolution operation, which shifts a weighted kernel over the pixels of a square image. Each convolution operation takes the inner product between pixels overlapping the kernel and the weights of the kernel, which can then be outputted to a lower dimensional feature set. The goal of this convolution is to extract implicit image features from bands of an image, such as edges and textures. The weights of these kernels, like neurons within ANNs, are trained through backpropagation, enabling these kernels to extract features present in the dataset they are being trained on. An illustration of convolution kernels within AlexNet are shown in Figure 5 below, highlighting how this feature extraction is accomplished - features present in an input image such as shape and texture can be extracted when a kernel is convoluted over the image.

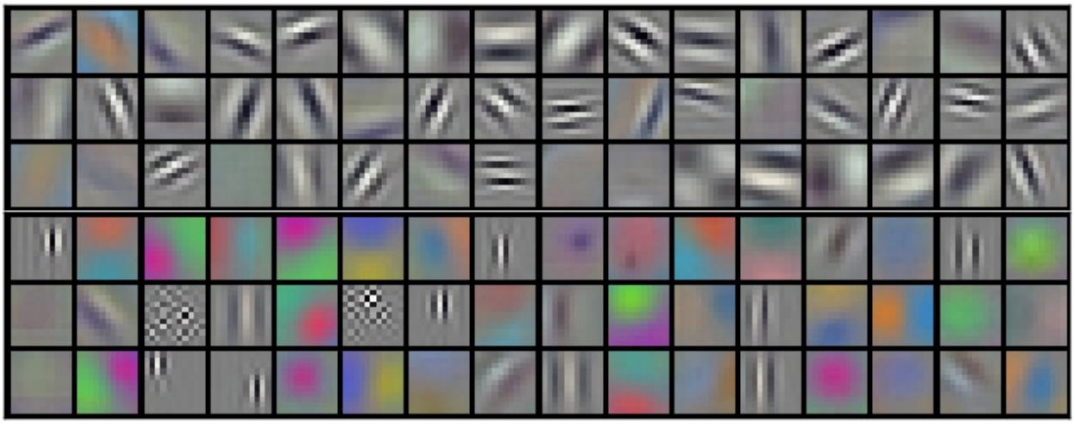

Figure 5 - *Image of kernel activations from the first convolutional layer from (Krizhevsky, Sutskever, & Hinton, 2012*

Pooling layers can then lower the dimensionality of the output of these convolutional layers by applying a function over each kernel. For example, with max pooling, the outputs of

the lower dimension feature set are the maximum value over each shifted kernel. Thus, pooling layers allow for the model to extract dominant features and remove noise in neuron activations, all while decreasing the dimensionality of input features.

This strategy of convolutional layers followed by pooling layers establishes the basis for CNN feature extraction, where an input image is fed through sequential convolutional and pooling layers, with the output representing implicit features of the input image. This feature extraction creates a desired $\mathbb{R}^{ixjxk} \rightarrow \mathbb{R}^{m}$ nonlinear mapping, such that spatial information between pixels is preserved statistically in the output $\mathbb{R}^{m}$ feature set. This implicit $\mathbb{R}^{m}$ feature set can then be inputted to a final densely connected layer, which then uses these features to classify the image.

The architecture of AlexNet, shown below in Figure 6, consists of this same strategy, layering convolutional and max pooling layers for each band to accomplish high accuracy on the ImageNet Dataset. Many of these convolutional operations are easily parallelizable, allowing for rapid prototyping of neural network architecture. AlexNet was one of the first CNN architectures to take advantage of GPU parallelization, with a final architecture of 5 convolutional layers and 3 max pooling layers, followed by an ANN in the form of 3 densely connected layers.

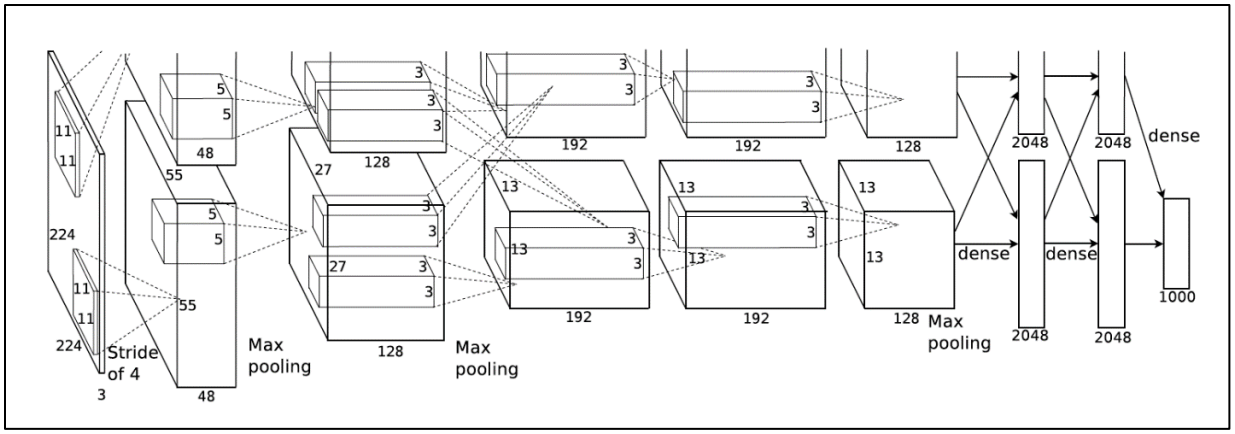

Figure 6 - *Architecture of AlexNet CNN from Krizhevsky, Sutskever, & Hinton, 2012*

2.2.3. EfficientNet

Since the advent of AlexNet, CNNs have increased in performance, and consequently in number of layers and complexity. One would think that deeper CNNs would lead to better performing classification performance, as a more complex neural network could theoretically model more complex nonlinear systems. However, larger, deeper models in fact led to worse performance due overfitting on training datasets. Since deeper networks can model more complex nonlinear systems, such networks can better fit to noise or outliers reducing their ability to generalize. Using the learnings from the bias-variance tradeoff, CNN architectures can be designed to balance bias and variance by changing the depth and width of their architecture. Newer models such as ResNet implement these improvements (He, Zhang, Ren, & Sun, 2015), reducing both bias and variance simultaneously with the use of residual blocks and trading width for depth, allowing for even more complex and deeper architectures.

However, as models broke through previous depth limitations and their performance reached closer to an irreducible error, computational needs for both training and predictions increased due to an increase of trainable parameters. Because of these computational needs, both training and prediction using these larger models can be unrealistic for many uses because of high costs. EfficientNet aims to improve on the computational needs of these large models by offering varying architectures for different computational limitations, while still offering state-of-the-art classification performance (Tan & Le, 2019). EfficientNet accomplishes this by scaling its depth, width, and resolution uniformly by using a compound coefficient related to computational usage, with this scaling illustrated in Figure 7. From this, an increase of depth by α^N, width by β^N, and input image size by γ^N, where α, β, and γ are some constant coefficients found by a grid search, can increase the number of network parameters by 2^N. Through this,

increasing model size can be done optimally and automatically in a way that limits the number of trainable parameters, also limiting computational cost.

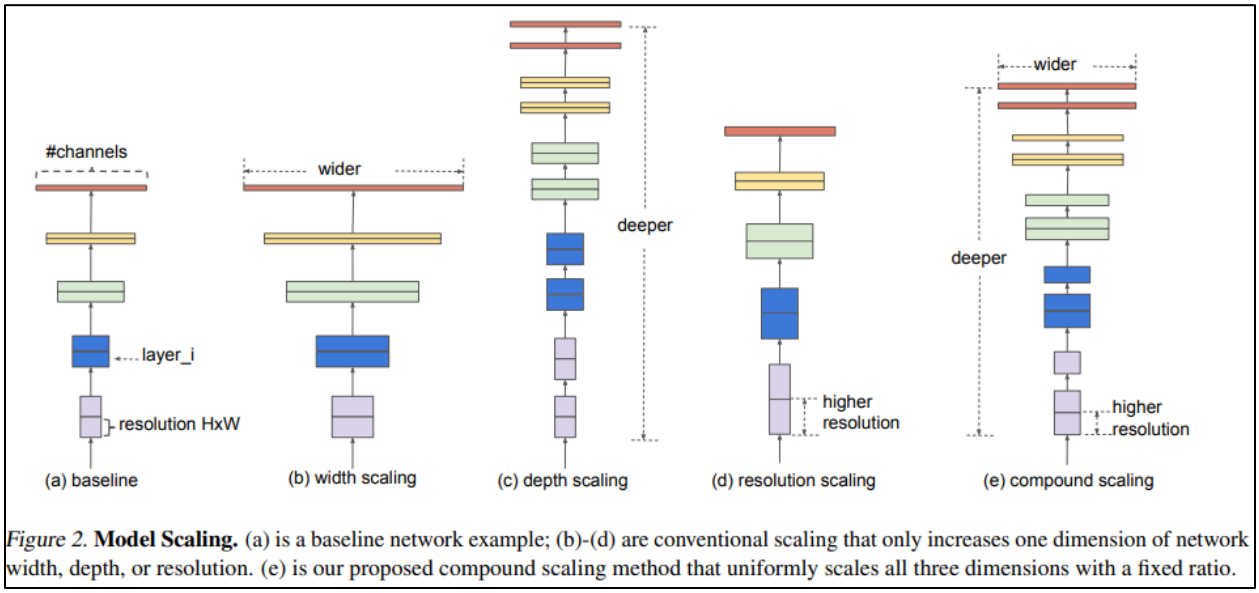

Figure 2. **Model Scaling.** (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) is our proposed compound scaling method that uniformly scales all three dimensions with a fixed ratio.

Figure 7 - *Figure showing model scaling used in the EfficientNet CNN – from Tan & Le, 2019*

This automatic scaling resulted in state-of-the-art classification performance over the ImageNet dataset, while still having only a fraction of trainable parameters when compared to similar performing networks.

3. Machine Learning: Application

In this section, we will discuss design choices when applying machine learning methods mentioned in Chapter 2 to mangrove image classification. These learnings come from utilizing low-resolution or high-resolution imagery, or a combination of the two.

3.1. Low-Resolution Imagery

With strong-performing algorithms such as AlexNet and EfficientNet, CNNs remain one of, if not the primary method of image classification. CNNs have established themselves in many

applications, most notably remote sensing (Jian, Yunquan, & Yue, 2021). Like how ImageNet helped move forward the field of image classification, datasets such as the UC Merced Land Use Dataset helped further the application of CNNs in remote sensing, consisting of tiled aerial images to serve as a baseline for remote sensing image classification (Yang & Newsam, 2010). Many remote sensing CNN models follow a similar workflow, tiling an input satellite or drone image into tiles to be classified by a CNN due to computational limitations preventing large input sizes of the scale of remote sensing imagery. CNNs have been applied directly to mangrove classification in this way, with Wan et al. utilizing a tile-based CNN to do mangrove extent and species classification of WorldView 2 satellite imagery (Wan, Zhang, Lin, & Lin, 2019).

However, there are some weaknesses when specifically using satellite imagery for CNN-based classification. Although WorldView 2 comparatively has a high spatial resolution of 0.46m and 1.84m for panchromatic and hyperspectral imagery, respectively, there are too little spatial image features to make CNNs worthwhile. This is evident in Wan et al.'s visualizations of kernels within their model's convolutional layers, as shown in Figure 8. Although Wan et al.'s model has high performance, many of the kernels either consist solely of singular colors or noise, indicating that the main features extracted are color or statistical features, not spatial information such as shape and texture. From this, we can deduce that higher resolution imagery such as drone imagery may be needed to take full advantage of CNNs, but performant models can still be created with a lack of spatial features for lower resolution imagery. Thus, for when only low-resolution imagery such as satellite imagery is present, classical models can be used due to their simplicity and low computational requirements when compared to deep learning models.

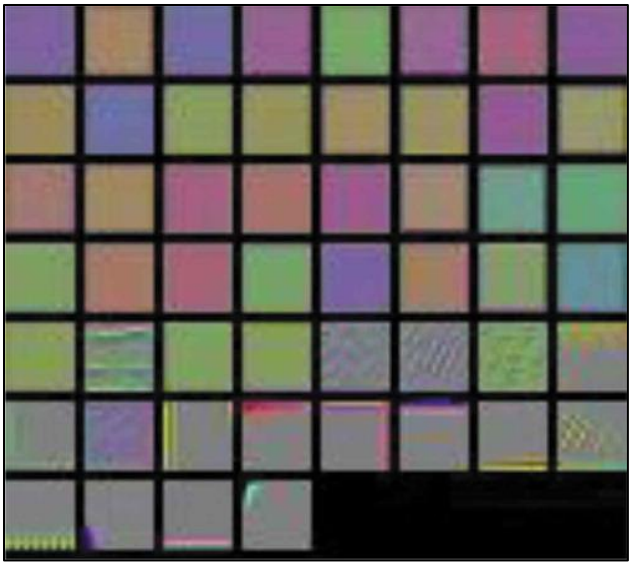

Figure 8 - *Image of kernel activations from the first convolutional layer from Wan, Zhang, Lin, & Lin, 2019*

3.2. High-Resolution Imagery

Since high-resolution imagery such as drone imagery does contain the image features needed to fully utilize the power of CNNs, we will test and utilize CNN models for datasets that have full high-resolution imagery coverage and apply other models for datasets that have lower resolution imagery. Thus, to motivate the usage of CNNs with high-resolution imagery, we initially implemented an EfficientNet-b0 CNN as a baseline method to classify mangrove extent, but several weaknesses were revealed. The first weakness is sensitivity to image brightness and colors as illustrated in Figure 9. Although textures can play a large role in the features, CNNs can still overtrain on colors within input imagery, as is evident with Wan et al.'s method. This can especially be apparent in high-resolution imagery, as although features such as texture and leaf shape can extracted, since the textures of mangrove and nonmangrove vegetation can be similar, image color still plays a large role in the decision of classifications. We attempted to fix these color and brightness issues through augmentation of our dataset to modify brightness of

input tiles randomly, but this augmentation had an adverse effect and lowered performance overall.

While manual labelers can easily account for image brightness, as they can compare brightness across an entire image when deciding between mangrove and nonmangrove areas of an image. Since CNNs only receive a single image tile, they are not afforded this luxury and can misclassify areas of mangrove canopy under cloud cover. On the other hand, if satellite imagery using no cloud coverage is used, the color of mangrove canopy can be consistent across an entire image dataset.

Figure 9 - *Maps highlighting performance of CNNs with poor lighting; Pane 1 - Input mangrove drone image, Pane 2 - EfficientNet CNN Classifications*

Also, since CNNs are sensitive to texture, overtraining on different textures can lead to incorrect classifications. This can be experienced in certain areas where imagery recorded over water visually resembled mangroves, causing incorrect classifications of water areas as mangrove. With manual annotation, human labelers can assess the context of an image and are able to discern between areas that are water and mangrove canopy, but CNNs are not afforded this luxury. At the same time, if low-resolution satellite imagery was used, hyperspectral bands

and features such as NDVI and NDWI can help filter out many of these nonvegetation areas. Lastly, since many of these image textures are not resolved in satellite imagery, models utilizing satellite imagery are typically invariant to many of these texture peculiarities. Thus, if such low-resolution satellite imagery was included in a CNN model, we could make models more invariant to high-resolution image peculiarities.

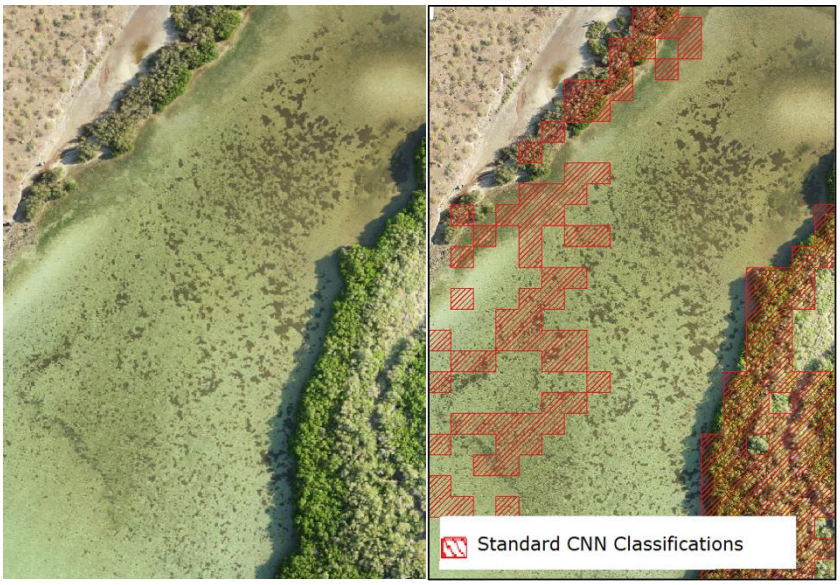

Figure 10 - *Maps highlighting CNN texture overtraining; Pane 1: Input mangrove drone image, Pane 2: EfficientNet CNN Classifications*

3.3. High- and Low-Resolution Imagery

Classical models such as Extreme Randomized Trees and deep learning-based CNN models such as EfficientNet are well suited for predicting on low-resolution and high-resolution imagery, respectively. However, using a combination of low- and high-resolution data could prove to be an effective way to make mangrove-extent classifications more generalizable. One could combine multiple sources of low-resolution data, such as multiple satellite image sources, but this would not solve the weaknesses of models trained on such data being unable to use spatial data such as texture. One could also use multiple sources of high-resolution data, such as

hyperspectral aerial imagery and RGB drone imagery, but such a survey could be unfeasible due to the resources required to acquire multiple high-resolution image sources.

On the other hand, in the more common case where both low-resolution satellite and high-resolution drone imagery are available, the creation of a model that takes advantage of both types of data has the potential to increase performance of mangrove extent classification even further. **Therefore, to take advantage of both drone and satellite data, we will create a model, which we call the 'hybrid model' that combines a CNN deep learning network with a classical ANN network, allowing high-resolution and low-resolution imagery to be used concurrently to the greatest effect**. Such a hybrid network can be fed drone and satellite imagery corresponding to a single area, effectively fusing features both data types implicitly with a neural network to achieve high accuracy when classifying that area (Hicks, Kastner, Schurgers, Hsu, & Aburto).

3.4. Hybrid Model Architecture

To establish the architecture of our Hybrid model, we can first look at how low- and high-resolution data can be ingested into the hybrid model. With low-resolution imagery such as satellite, qualitatively, little is gained by utilizing spatial information of satellite imagery, so we can simply use a single satellite pixel as an input. Since all features are self-contained within each satellite pixel after feature generation, we can utilize a perceptron in the form of a single fully connected dense layer within hybrid model. Using this dense layer not only allows the model ingest satellite features, but also to find implicit relations between input features.

As an input to provide high-resolution spatial information for the hybrid model, we used the drone imagery corresponding to the same geographical area as the satellite pixel. To ingest this drone imagery, we can use a feature extractor, which is a trained CNN with the final fully

connected classification layers removed. We can use such a feature extractor in our application to extract spatial information important to mangrove image classification such as texture. We utilized an EfficientNet-b0 feature extractor, which was selected for its combination of fast training and prediction times, while still having good classification accuracy compared to other Efficient-net variants. The features outputted by this feature extractor can then be combined with satellite features output from the perceptron to establish the basis for our hybrid model.

Next, to combine both the drone and satellite image features, the outputs of the EfficientNet-b0 feature extractor and the fully connected input dense layer can then be concatenated and fed into another dense layer. This combines features extracted from the single drone tile and the features from the satellite pixel, effectively fusing the data from low-resolution and high-resolution sources. The output of this fully connected layer is then fed into two neurons with a sigmoid activation, representing the probability of mangrove or nonmangrove (Hicks, Kastner, Schurgers, Hsu, & Aburto). From these probabilities, we can then create labels for mangrove extent at the resolution of the input satellite image features. This architecture is illustrated in Figure 11.

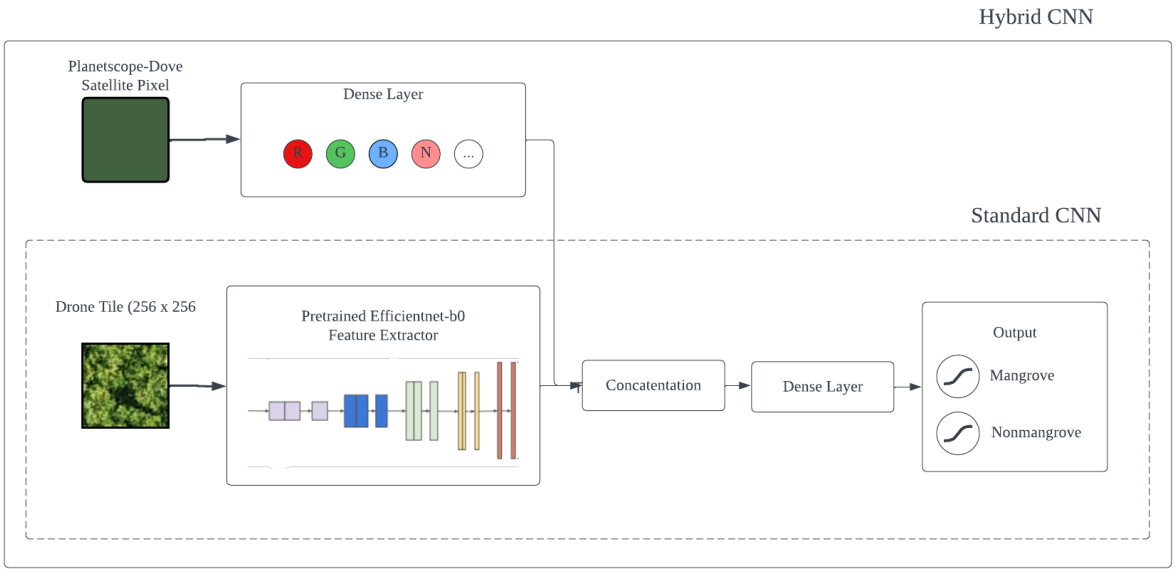

Figure 11 - *Diagram of the Hybrid Model, highlighting the input and concatenation of dense layer and separate feature extractor for mangrove classification*

4. Expeditions

With the motivation behind our machine learning development established, we can now look at the data used to train our mangrove extent classifiers was acquired. In this section, we will be describing the two expeditions in which data was acquired to train our mangrove extent classification models. These two expeditions provide the motivation of our work, highlighting real world challenges when acquiring mangrove remote sensing data. These challenges further motivate the development of our machine learning classification models due to practical differences between surveys impacted by real world conditions.

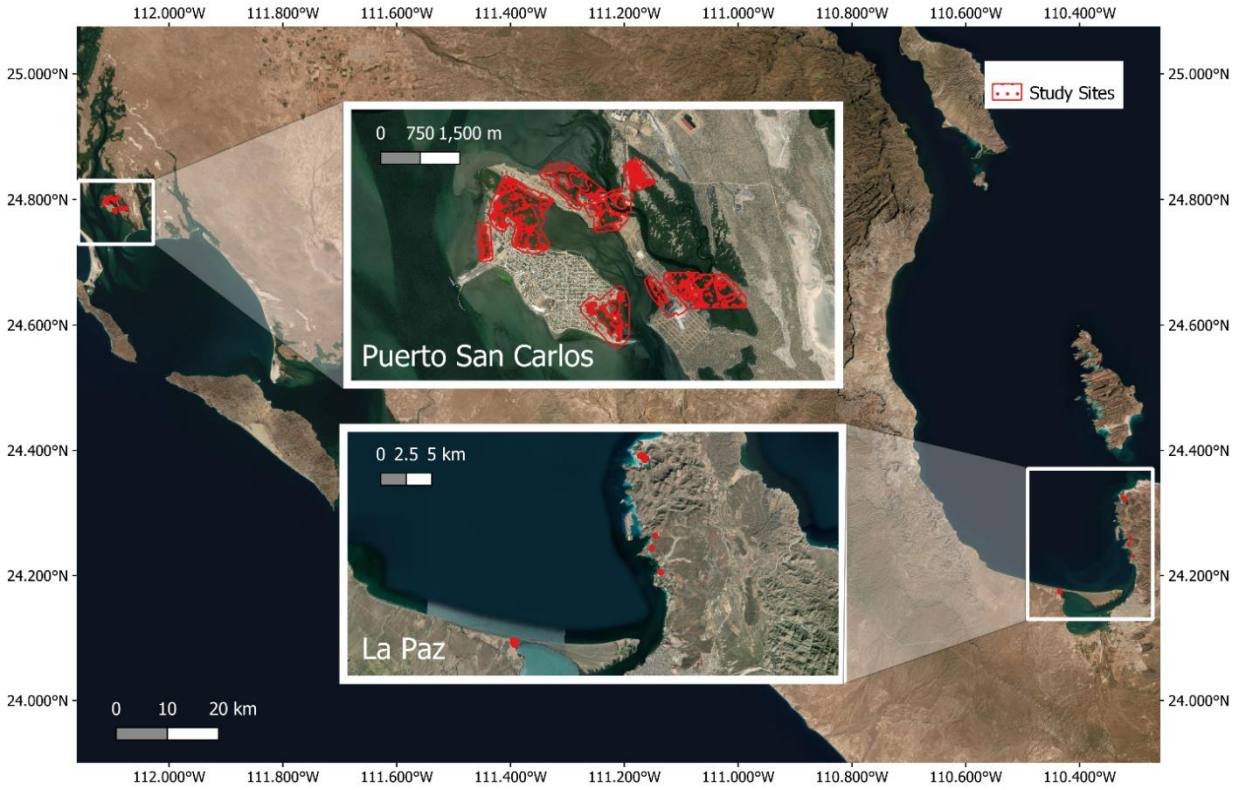

Figure 12 - *Map highlighting the survey area of the Mexico Expedition*

4.1. Mexico Expedition

Our first dataset was acquired in Baja California Sur (BCS), Mexico around the La Paz and Puerto San Carlos regions during three separate surveys between May 2018 and July 2019, illustrated in Figure 12. We will refer to all surveys during this time as the "Mexico expedition". This survey was completed in collaboration with the Gulf of California Marine Program to conserve mangroves in BCS under threat due to coastal development and deforestation. The climates of La Paz and Puerto San Carlos are arid with an average temperature of 24.5 – 29.7 °C and 11cm of rain during summer months. Because of BCS's dry, arid climate, mangrove ecosystems are fragmented along the coasts with desert and beaches separating these mangrove

ecosystems. The survey areas were separated into separate survey sites with most survey sites representing a single group of mangroves less than 1 km^2 in size.

Drone images were captured using a Phantom 4 Pro, photographed in Figure 13, using its onboard camera, which has a 1-inch sensor capturing visual spectrum imagery at a resolution of 20 megapixels. Imagery was recorded with the assistance of the DJI Ground Station Pro software, at an altitude of 120m with an overlap of 85% in a lawnmower pattern. In addition, images of color calibration cards were taken at the beginning of every flight for white balance correction. Detailed information on this surveying procedure is outlined by Hsu et. Al. (Hsu, Lo, Dorian, & Guerrero Martinez, 2019). Because of good weather conditions and easy access to mangrove ecosystems, full coverage of 3cm spatial resolution drone imagery was able to be achieved over the survey area. This high-resolution imagery can then be used to train our CNN-based models.

Lastly, we were also able to acquire archive satellite data in the form of Planetscope Dove RGB/NIR imagery and 8 band Sentinel 2A imagery over the entire survey area. However, since this survey was completed before the launch of Planetscope Superdove, we were unable to utilize the bands present in Superdove imagery in our Mexico dataset. Despite the nonideal satellite imagery, we can still use the low-resolution hyperspectral and high-resolution visual-spectrum data acquired during this survey to train our hybrid models.

Figure 13 - *Photograph of Phantom 4 Pro taken during a survey expedition in La Paz, 2018*

4.2. Jamaica Expedition

Our second dataset was acquired during a single survey in the Clarendon parish of Jamaica during March 2022. This survey was completed in collaboration with the University of the West Indies and Coastal Dynamics Limited to conserve mangrove forests under threat due to damage from tropical storms and deforestation. The Clarendon region has a tropical and humid climate with an average temperature of 28.4 degrees °C during the month of March. Mangrove areas are larger in comparison to areas surveyed during the Mexico expedition, as mangrove ecosystems in Clarendon cover much of the coast in contiguous strips, with the study area illustrated in Figure 14. Mangrove areas were separated into individual survey sites to coordinate drone flight areas, but many of these sites proved difficult to survey due to their size, with some sites only accessible via boat.

Figure 14 - *Map highlighting the survey area of the Jamaica Expedition*

Difficulties in data acquisition during the Jamaica survey created a stark juxtaposition to the conditions experienced during the Mexico expedition. Drone imagery was captured using the same methodology as the Mexico expedition at 120m but using 80% overlap instead of 85% within Ground Station Pro to achieve a faster survey speed. However, because of intermittent weather that occurred during the survey, drone image quality was qualitatively lower when compared to drone imagery recorded during the Mexico expedition, as illustrated in Figure 15. High winds were common, which often caused drones to fly off course, resulting in poor orthorectification in final drone imagery. In addition, because of cloud cover often quickly changing midflight, some output images have poor white balance correction due to differences in lighting conditions from when color calibration images were recorded.

Figure 15 - *Example selections of orthomosaics from the Jamaica Expedition. Pane 1: Sunspot reflections circled in red, Pane 2: orthorectification errors due to poor weather conditions, Pane 3: white balance inconsistencies due to intermittent weather conditions*

To assist in the georeferencing and correction of images, ground control points (GCPs) were captured using an RTK GPS, with their locations shown in Figure 16. However, despite their usefulness in standard surveys, GCPs were limited in quantity due to the large size of the survey area and did little to combat orthorectification errors in imagery, with images highlighting these non-ideal drone imagery results shown in Figure 15.

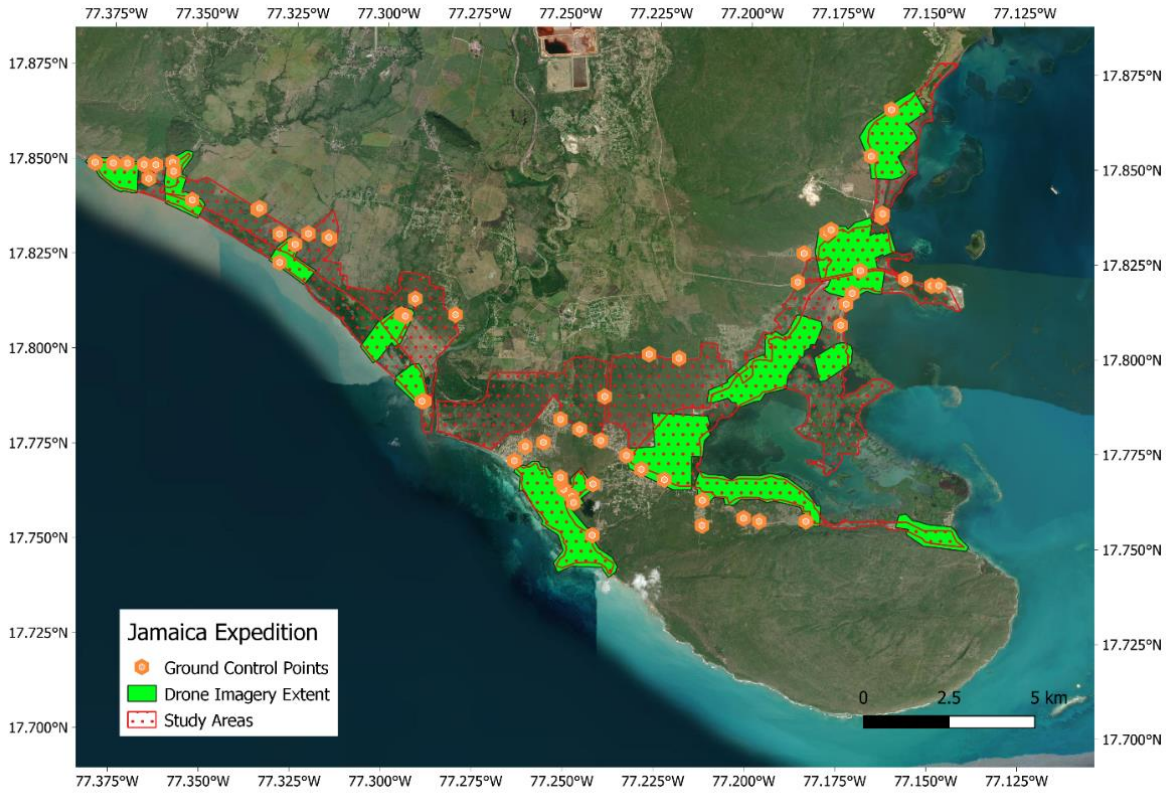

Figure 16 - *Map drone imagery coverage and GCP locations captured during the Jamaica Expedition*

Also, because of consistently high wind speeds at or near recommended limits set by DJI, all three P4Ps used during this survey achieved structural damage to their airframes, as shown in Figure 17. Repairs to one P4P was completed, but the lack of available airframes limited the survey's team ability to capture imagery of the entire survey area. Consequently, only 45% of the total survey area has coverage of drone imagery, with an illustration total drone imagery coverage shown in Figure 16.

Figure 17 - *Damage to two DJI Phantom 4 Pros during the Jamaica Survey due to high winds*

Fortunately, the remaining 55% of the survey area was available as satellite imagery. Specifically, Planetscope Superdove data was available, so we acquired 8 band Superdove archive imagery at a spatial resolution of 3m use for the entire survey site. Sentinel 2A 12 band archive imagery at a resolution of 10m was also acquired. To assist in labeling, Pleaides satellite imagery at a resolution of 0.5m was also acquired, mainly to be used for label validation because poor color correction in our drone imagery and inconsistent mangrove canopy coloring caused inconsistent manual labels. Since acquired drone imagery was of poor quality, we only have complete satellite imagery of the survey area. Thus, the Jamaica expedition and the data acquired serves as a practical example to test methods where only satellite imagery is available, notably classical machine learning methods.

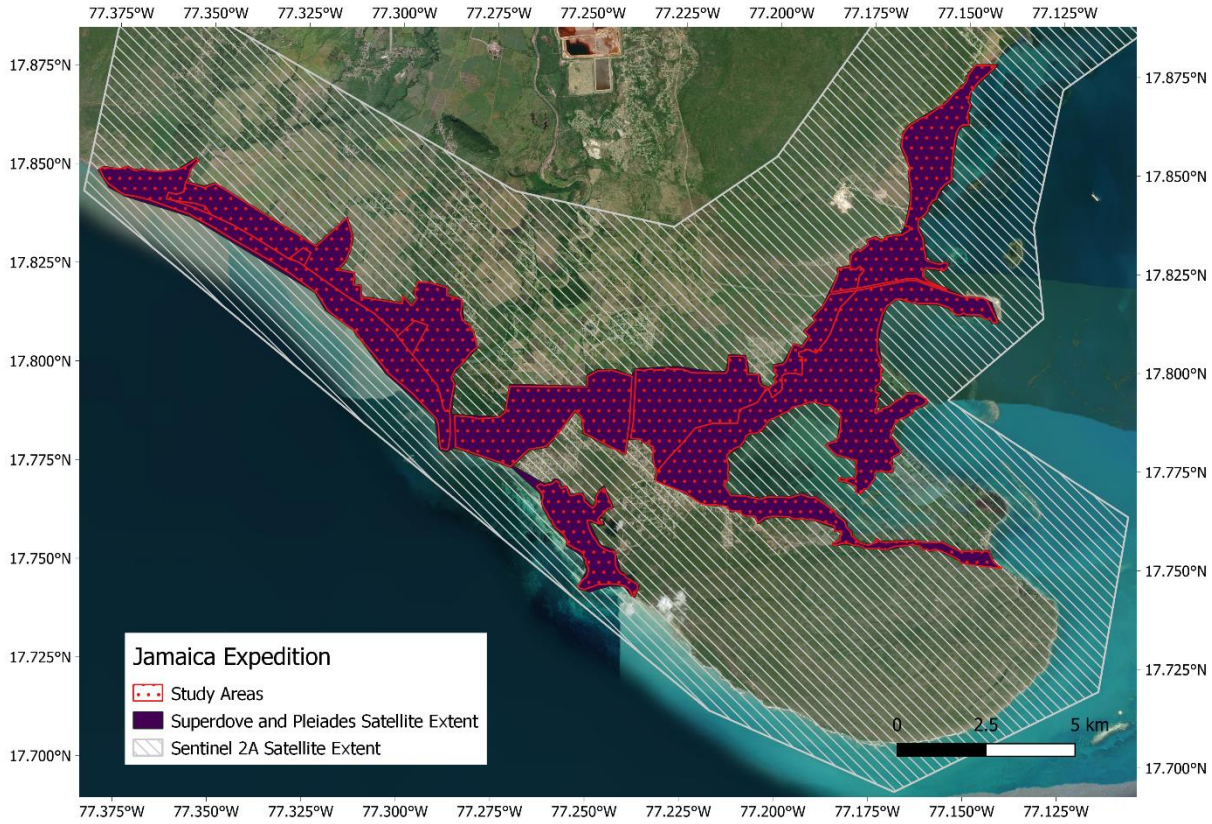

Figure 18 - *Map highlighting satellite imagery coverage of areas surveyed during the Jamaica Expedition*

5. Data Pipeline

In the previous chapter we introduced two practical examples in which imagery was gathered for mangrove extent classification. Before the data from the examples are tested, we must discuss how this data is processed for our proposed machine learning architectures. In this section, we outline different steps of the data processing pipeline for the classical model for satellite image classification and of the hybrid model for the classification of paired satellite and drone imagery.

5.1. Image Processing

5.1.1. Drone Imagery

To train our hybrid model, processing of raw UAV imagery is required. First, images are first color calibrated in Adobe Lightroom using reference images of color calibration cards taken at the start of every flight. Next, these color corrected images can be imported into Agisoft Metashape, a photogrammetry software used to generate 3D models from a set of oriented cameras. Since each image taken by the P4P contains gyroscopic and location data, the camera orientation can be determined for each image. Structure from Motion algorithms are then used to take overlapping sections of individual images and their locations to create a contiguous 3-dimensional image over the entire survey area. In the case of our Jamaica imagery, ground control points (GCPs) are also imported at this stage, which help correct and align the placement of our final images. Also, since some of our Mexico imagery contained reflections from the sun, or ‘sunspots’ into the camera lens, we manually masked these sunspots using the ‘Patch’ tool in Metashape. These images are then aligned with Metashape, and any images taken that are too off nadir during turns are removed. From this alignment, a sparse point cloud is generated, and a dense point cloud can be generated directly afterwards. From this dense point cloud, any error or outlier points can be removed manually to prevent any inaccuracies in our final images. Lastly, a digital elevation model (DEM) is created, which is then used as a reference for the output orthomosaic, or our final image. All processing steps are completed at the highest quality available to ensure that intermediate point clouds and final orthomosaics are precise. Orthomosaics for the drone imagery of both the Mexico and Jamaica expeditions had a spatial resolution of 3 cm. These final orthomosaics, are then ready to be hand annotated and utilized with our ML algorithms.

5.1.2. Satellite Imagery

Since our satellite images sources provide images that are already processed and orthorectified, there is little extra processing that needs to be done. Therefore, much of the processing done on our satellite imagery is feature engineering, where raw data values are transformed to help augment and improve performance of our ML algorithms. Using the raster calculator tool in QGIS and the near infrared band (NIR), NDVI can be calculated to serve as an additional feature for our ML algorithms using the following equation:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

Similarly, NDWI can be calculated in the same way with the following equation:

$$NDVI = \frac{(Green - NIR)}{(Green + NIR)}$$

NDVI and NDWI are calculated for our Planetscope Dove and Superdove Imagery, and our Sentinel 2A Imagery. Because the Pleiades imagery is only used for assistance in labeling of our Jamaica survey area, this feature generation is not needed.

In addition, we created a distance from water for our Jamaica dataset, which GMW used to great effectiveness (Bunting, et al., 2018). Since mangroves ecosystems are generally concentrated along coastlines, we created a distance from water feature to help our machine learning algorithms better capture the distribution of mangroves along coastlines. To generate this distance from water feature, we first had to generate a water mask. We utilized our Sentinel 2A imagery to create this feature, as our Planetscope Superdove imagery prioritizes survey area coverage does not have full coverage for water pixels that border mangrove pixels. Using the previously calculated NDWI feature using visual inspection, all NDWI values over 0.5 were classified as water and given the value of 1, with all other pixels given the value of 0 to create a

binary mask. This binary mask is then inputted into the Proximity (Raster Distance) tool in QGIS to create our final distance to water feature. While this feature was added to our Jamaica dataset and should be considered for future datasets, we were unable to add distance to water to our Mexico dataset due to practical considerations of overhauling a previously generated dataset. An example illustration of the water binary mask and distance to water feature are shown in Figure 19.

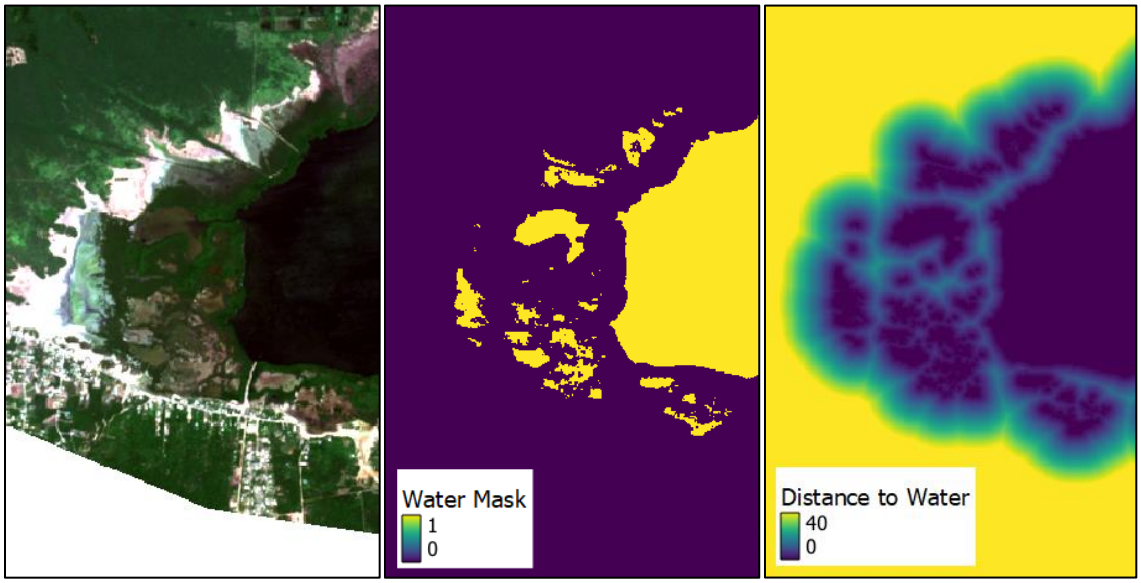

Figure 19 - *Sentinel 2A Water Features; Pane 1: Sentinel 2A Source Image, Pane 2: Example Water Mask Image, Pane 2: Example Distance to Water Image,*

5.2. Image Labeling

Once imagery is processed, manual image labeling is done over the processed drone imagery to serve as ground truth for machine learning model training and testing. Labeling is done with QGIS, with trained labelers creating polygons covering mangrove and nonmangrove areas, which are then saved in Shapefile format. Labels are then reviewed by mangrove experts and are cross referenced to high-resolution drone imagery (Hsu, et al., 2020). In the case of the Jamaica dataset, labels are cross referenced to high-resolution Pleiades satellite imagery. This

process is completed for our Mexico and Jamaica dataset and serves as a ground truth for our respective classification models.

6. Results

In this section, we discuss the performance of both the Hybrid model and the classical model, presented originally in Section 3.3 and Section 3.1, respectively, over their respective datasets while comparing both models to baselines. Visualizations of chosen methods and their baselines from both datasets are provided as well.

6.1. Classification Based on Low-resolution Data

6.1.1. Low Resolution – Training and Baselines

Once labels are complete, data can be formatted to serve as direct inputs to train our ML models. For our classical models, where only satellite imagery is used, calculated NDVI and NDWI layers are stacked as additional bands on top of the original Planetscope-Superdove imagery. However, since the resolution of the distance to water feature is at the resolution of the Sentinel 2A imagery, we must first align this feature to be the same resolution as the Planetscope-Superdove raster. This is done using the QGIS “Align Rasters” tool, where with a bilinear interpolation, the distance to water feature can be aligned and stacked as another feature on top of the Superdove raster.

In addition, since labels in our Jamaica dataset do not have full coverage over the survey area such that not every pixel in the image has a corresponding label, the process for creating labels ready for input to the classical model is slightly different. First, a mask must be created indicating which pixels have valid labels. This is done by combining the mangrove and

nonmangrove polygons, and then rasterizing this combined mask polygon at the resolution of the Planetscope-Superdove data, where pixels that have a geographical center inside the mask polygon are given a 1, and all other pixels are given a 0. Then, using the indices of this mask, we can then isolate all valid pixels from the previously processed raster and labels into an array, like the format used with our satellite baseline for the Mexico dataset.

Since classical models require much less processing power when compared to deep learning models, each model was trained on the Jamaica dataset directly on a personal computer taking ~2 minutes for each model to train. All classical models were trained over CPU using Scikit-learn (Pedregosa, et al., 2012). In total we trained **decision tree, random forest, gradient boosted trees, and ERT** models, as described in Section 2.1, for testing over our Jamaica dataset. We utilized the most recent GMW labels from 2020, rasterized to the same resolution as our Jamaica dataset, as a baseline result to compare against the classical models trained on the Jamaica dataset.

6.1.2. Low Resolution – Performance

With the classical models trained on the Jamaica dataset, we can verify their performance over our hand-labeled Jamaica dataset ground truth and compare these models to our baseline for these low-resolution methods, the 2020 Global Mangrove Watch (GMW) labels. We can measure the performance of our low-resolution classical models using a 5-fold cross validation, where for each fold a different 20% of the data is used as validation data, and 80% is used as training data. This split between training data and validation can ensure that for each model trained, enough training samples are provided to train our machine learning models. We tracked intersection over union (IoU), also known as the Jaccard score, and accuracy over all valid pixels in the Jamaica dataset. The equations for IoU and Accuracy are defined below, where true

positives are mangroves classified as mangroves, and false negatives are nonmangrove classified as nonmangrove.

$$IoU = \frac{True\ Positives}{True\ Positives + False\ Negatives + False\ Positives}$$

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Negatives + False\ Positives}$$

From this cross validation utilizing our hand-labeled ground truth, we have the performance over all classical models, **Extremely Randomized Trees, Random Forest, Decision Trees, and Gradient Boosted Trees**, listed in Table 2.

Table 2 - *Model performance of the Classical ERT model and other baseline models tested on the Jamaica Dataset*

Model	***Mean IOU***	***Mean Accuracy***
Extremely Randomized Trees	**0.873**	**0.934**
Random Forest	0.865	0.927
Decision Trees	0.829	0.901
Gradient Boosted Trees	0.865	0.927
2020 Global Mangrove Watch (GMW) Labels	0.742	0.823

For all these models, performance is strong, with ERT performing slightly better than all other baseline models with a mean accuracy of 0.934 and a mean IoU of 0.873, with all methods having considerably higher performance than our GMW baseline. Full mangrove extent measurements for the Jamaica survey area as shown in Figure 20.

Figure 20 - *Map showing mangrove extent classified from the classical ERT model of the Jamaica dataset*

To get a better understanding of how classical models perform on higher resolution datasets compared to GMW, we can visualize and qualitatively inspect the results of the ERT model at a more detailed view. Below in Figure 21, we plot a visualization of the 2020 GMW labels over 2022 input Planetscope-Superdove Imagery. From the map, we can see that GMW offers a poor resolution, being unable to precisely capture the edges of the mangrove ecosystem. Additionally, since 2020 is the most recent release of GMW labels, mangrove extent measurements for many areas are outdated. This is present in several areas above 17.805 degrees north, as certain are classified as mangrove despite these regions having dead mangroves or no vegetation.

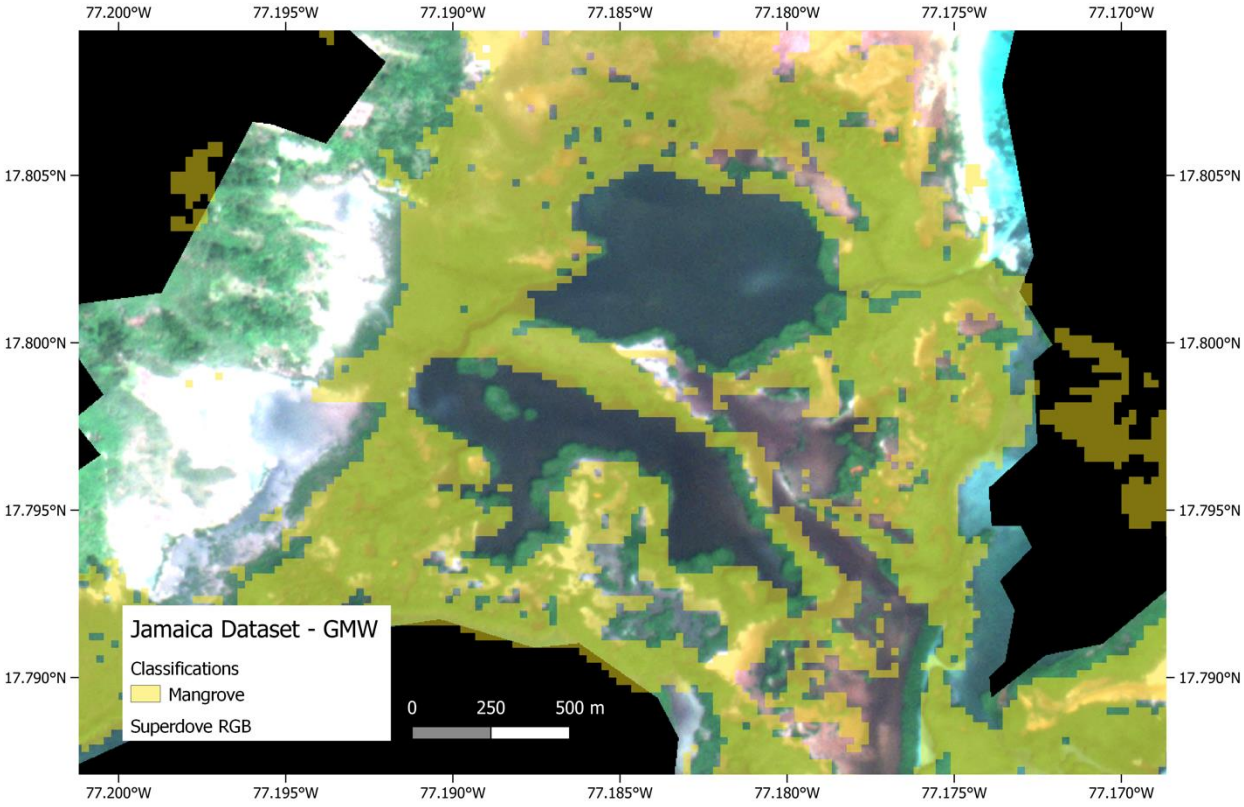

Figure 21 - *Map with a detailed view of GMW 2020 classifications of the Jamaica dataset laid over Planetscope Superdove Imagery*

These issues that are present within the lower resolution and outdated GMW labels are almost completely alleviated when newer, higher resolution Planetscope-Superdove imagery is

used to train an ERT model, as obvious nonmangrove areas are properly classified and the edges mangrove extent is precisely captured at a resolution of 3m, as shown in Figure 22.

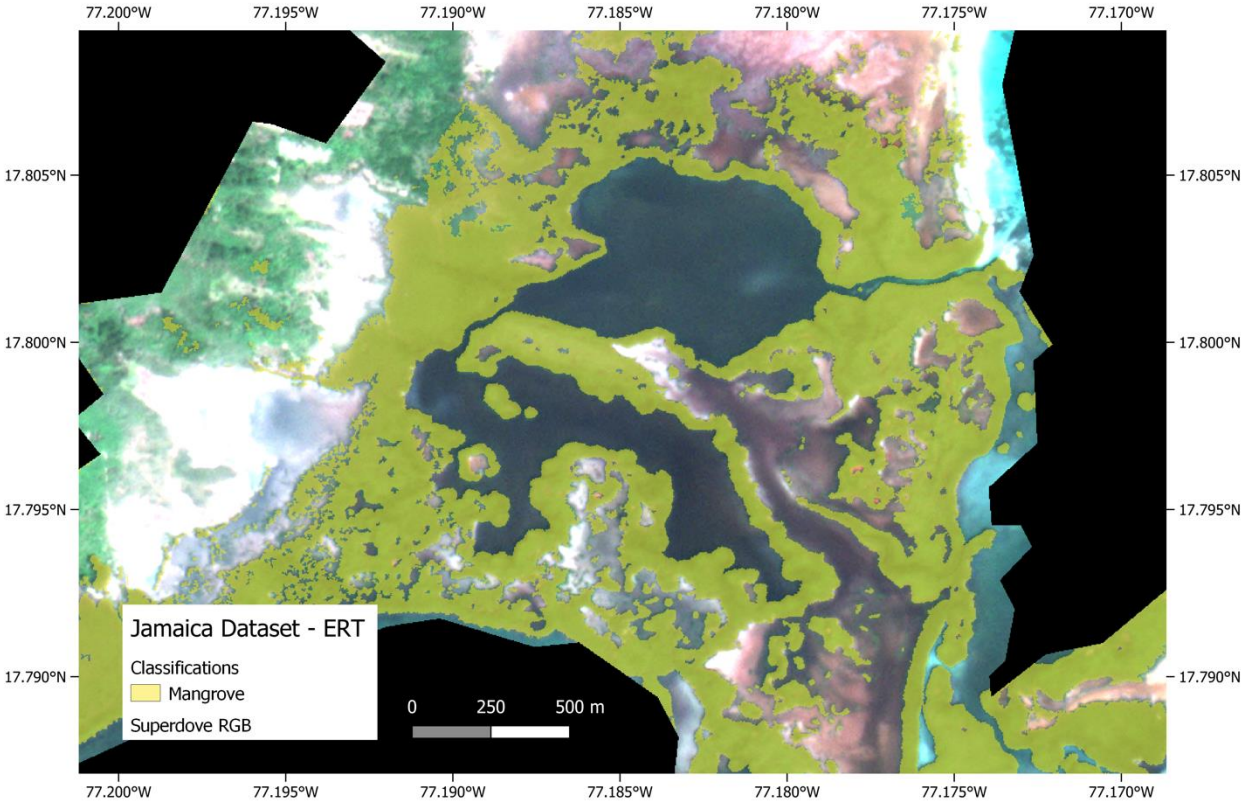

Figure 22 - *Map with a detailed view of ERT classifications of the Jamaica dataset laid over Planetscope Superdove Imagery*

Also, to show the effects of engineered features added to our Jamaica dataset, we can plot the Shapley values of each individual feature in a bar plot, which shows each feature's importance on model output (Winter, 2002). In our application, higher Shapley values for a feature indicate that the feature contributes greatly to a model's decision to classifying a pixel as mangrove or nonmangrove. From the plot shown in Figure 23, we can see that distance to water plays a large role in each classification, with high Shapley values for both the mangrove and nonmangrove class. NDVI and NDWI also have high Shapley values, indicating that generated

features can have a large impact on model output when compared to raw values in the base Superdove imagery.

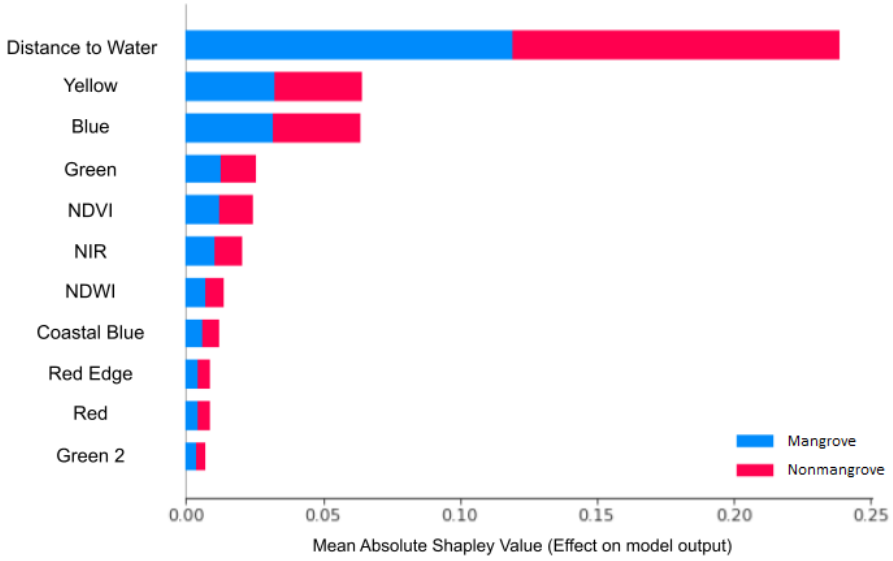

Figure 23 - *Shapley values of input features of the Jamaica dataset*

From these results, we can then conclude that when proper feature engineering is done, using classical models such as ERT is both a quantitatively and qualitatively effective way to classify mangrove extent when using a dataset that contains only satellite imagery.

6.1.3. Classical Model Recommendations

Based on our experimental results discussed above, if complete high-resolution drone imagery over the entire survey area is not attainable, utilizing low-resolution hyperspectral satellite imagery to train a classical model can be used to generate mangrove extent measurements over the complete survey area. This scenario can occur if mangrove areas are more contiguous and very large in size or if significant challenges are faced during the drone survey, which can be common in many areas around the world. In short, such a classical models trained on satellite imagery can be used when it is impractical to get high quality drone imagery. Although creating models on satellite imagery may be lower resolution than models that use

higher resolution imagery, such as drone imagery, classical models still provide an accurate way to automatically measure mangrove extent.

6.2. Classification Based on Low- and High-Resolution Data

6.2.1. Low- and High-Resolution – Training and Baselines

Next, we can test our Hybrid model, which takes advantage of both low- and high-resolution image data, first by formatting our data. For the hybrid model, data formatting and model training has key differences compared to implementation of the classical models. For our Mexico dataset, 3cm drone imagery can be tiled into 85 x 85 pixel sized square images using the "gdal_retile.py" Python tool, approximating the geographical size of a single Planetscope-dove pixel. All tiles containing no data values are then removed. Next, to ensure that each drone image tile has a corresponding Planetscope-dove pixel for the same geographical area, using the bounds of each image tile, the corresponding Planetscope-dove image is clipped and averaged. This pixel is then saved as another image and is renamed to pair it with the original tile.

Next, individual label files are combined and all images that contain both the mangrove and nonmangrove class are removed. For each tile and pixel pair, they are then segmented into folders depending on whether the center of the pair is inside of the mangrove or nonmangrove polygons. All other pairs are treated as unlabeled and discarded. These classified tile-pixel pairs then serve as the input dataset to our Hybrid model.

For evaluating the performance of our hybrid model, we then formatted our Mexico dataset to be compatible with our baseline methods. To compare the hybrid model against more basic methods, we created two other alternative classification results. The first baseline result uses the same image dataset as the Hybrid model, but only containing the drone image tiles and

corresponding labels. We then train an EfficientNet-b0 CNN only on this imagery. The second baseline result consists of GMW 2019 labels rasterized at the resolution of Planetscope-Dove, like the baseline used for the Jamaica classical models.

For the Mexico dataset, all models were trained on Google Colab. Premium instances with high RAM configurations were used as the large drone dataset quickly overloads the RAM of standard Colab instances. Therefore, we were able to upload datasets directly to the cloud instance. All CNN-based models were implemented using Keras and trained using the TensorFlow backend (Chollet, 2015). For our hybrid model, the RMSprop optimizer was used with a binary cross entropy loss. In total, the hybrid model trained for ~14.5 hours, with the best performing network during the 30 epochs training chosen as the primary model. Our CNN baseline models were trained in a similar fashion, utilizing Colab cloud resources for training. In total we trained the hybrid model and two EfficientNet-b0 CNNs, which served as our baseline CNN methods for the Mexico dataset, classifying images of width 85, the size of hybrid tiles, and a width of 256. The 2019 GMW labels were also used as a baseline for the Mexico dataset.

6.2.2. Low- and High-Resolution – Performance

Now, we can compare the Hybrid model to our Mexico dataset ground truth and our baselines which consist of a standard Efficentnet-b0 CNN, trained only using high-resolution drone data, and 2019 GMW labels. With the Hybrid model and all relevant baselines trained on the Mexico dataset, we measured the mean IoU and accuracy for each model over all 5 folds. A selection of classifications over an orthomosaic approximately 0.5 km² area in the Mexico Dataset are shown below in Figure 23.

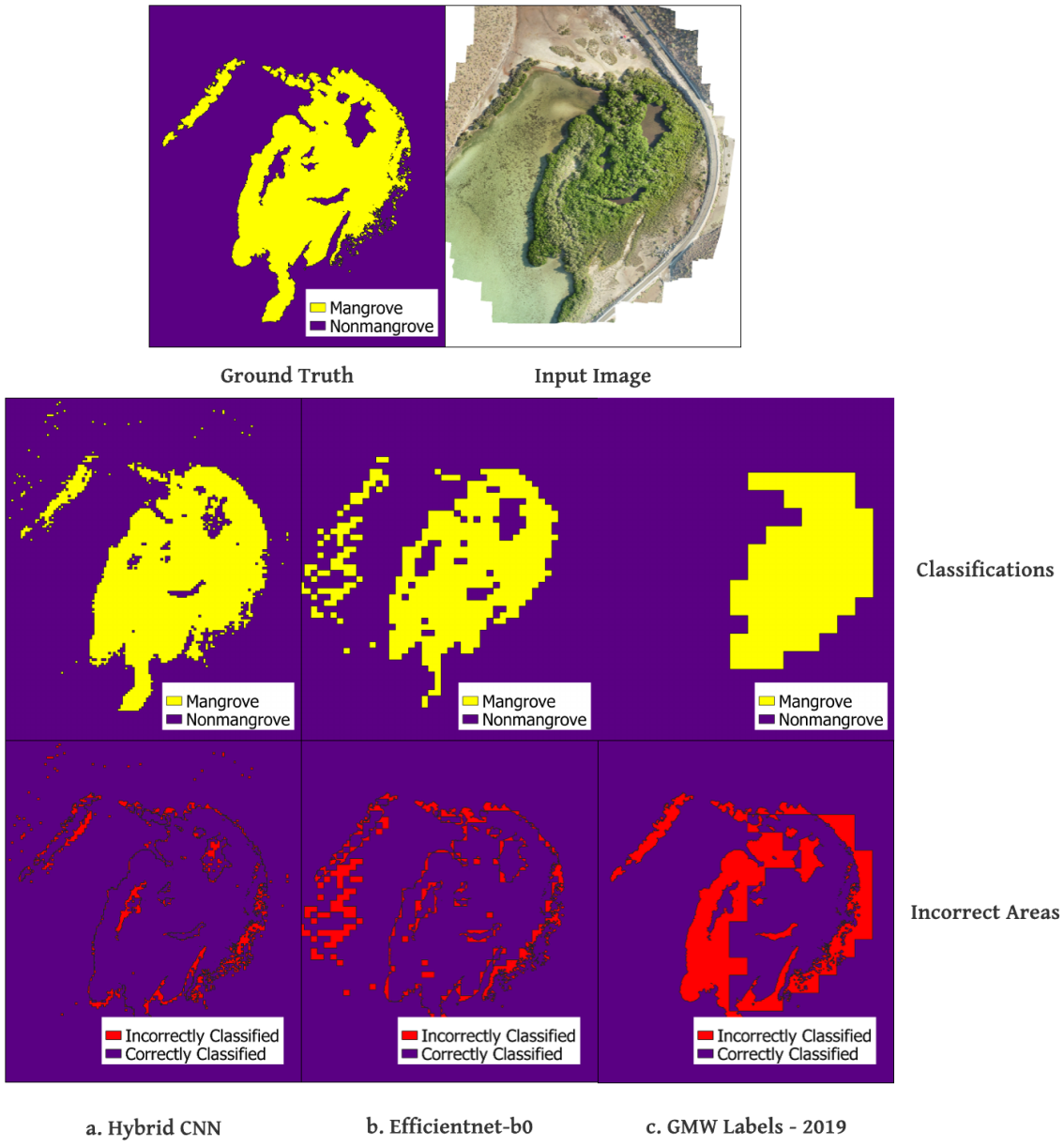

Figure 24 - *Images highlighting performance of Hybrid model and baselines over a test image*

The result of the cross validation shows that the Hybrid model has the highest performance over the Mexico dataset when compared to the CNN baseline and GMW labels, having the highest mean IoU and mean accuracy of 0.953 and 0.967 respectively. From this, we can see that fusing of low-resolution (LR) satellite imagery and high-resolution (HR) drone

imagery can be effective at increasing the performance of mangrove classification. In addition, since the Hybrid model classifies at the same resolution as the input Planetscope-Dove imagery, it has a classification spatial resolution at 3m², higher than the 10m² GMW Labels, while still having higher accuracy and IoU than either of these models. A summary of these performance metrics is outlined below in Table *3*.

Table 3 - *Model performance of Hybrid Model and other baseline models tested on the Mexico Dataset*

Model	***Mean IOU***	***Mean Accuracy***	***Resolution***
Hybrid Model (LR + HR)	**0.953**	**0.967**	3m²
EfficientNet-b0 CNN (HR)	0.898	0.954	3m²
2020 GMW Labels	0.621	0.719	10m²

We can also qualitatively assess the Hybrid model, seeing improvements over our EfficientNet-b0 baseline network. Comparing the Hybrid model directly to the baseline CNN model and hand-labeled ground truth, we can see that the Hybrid model performs much better than the baseline CNN when predicting areas with ambiguous textures. For example, when classifying the same image as in Figure 10, in Figure 25, we can see that the Hybrid model classifies textured water areas correctly as water that the standard CNN model originally classified as mangrove incorrectly, when comparing both to the hand-labeled ground truth. This can likely be attributed to satellite features being used in the model, as the NDWI feature and lack of detailed image textures Planetscope-Dove imagery are likely helping the hybrid CNN both form bias against these types of features, while still increasing variance as to not overtrain on these features.

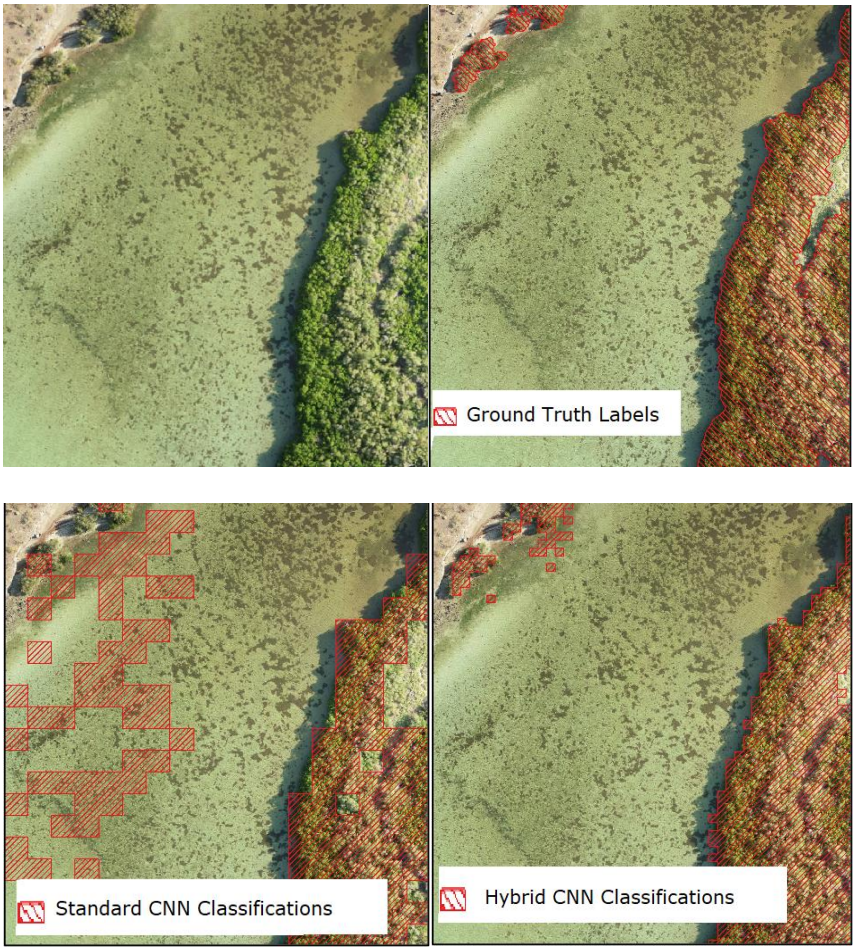

Figure 25 - *Maps highlighting performance of Hybrid model over texture areas; Pane 1: Example input drone image area, Pane 2: Hand-labeled mangrove ground truth, Pane 3: EfficientNet CNN Classifications, Pane 4: Hybrid Model Classifications*

Lastly, the hybrid model qualitatively performs better when classifying drone imagery with poor lighting when compared to standard CNN models, as shown in Figure 26 below. Although there are still some misclassifications of mangrove areas as nonmangrove, there is a noticeable increase in performance when classifying these poorly lit areas, likely due to the fact that the satellite features used in the hybrid model are relatively invariant to lighting changes.

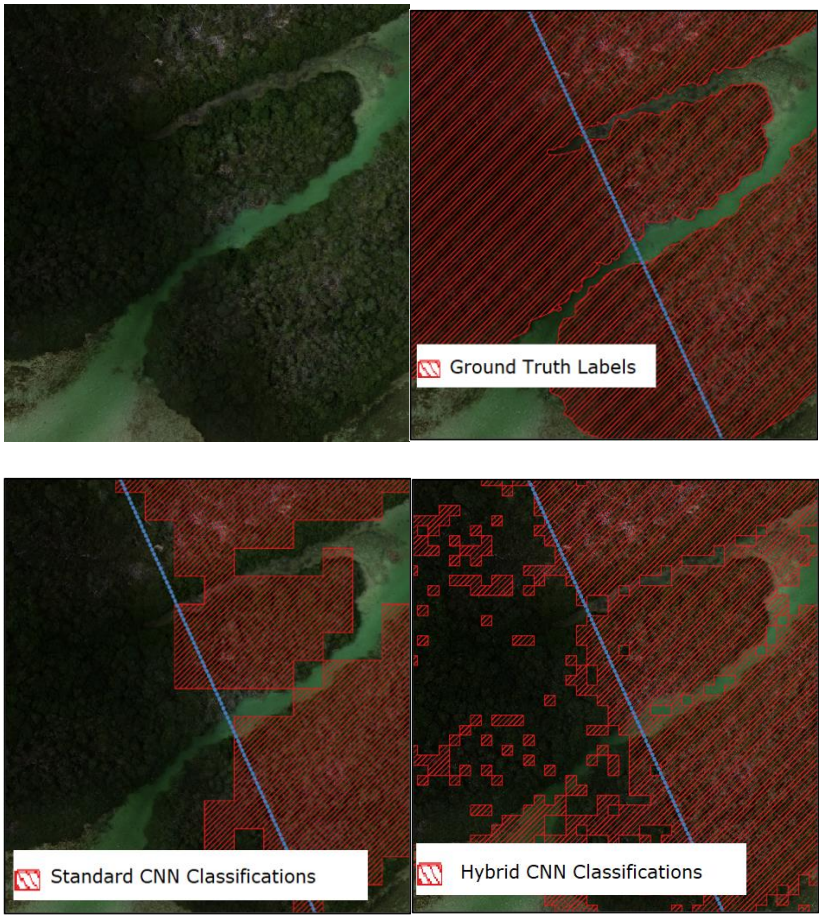

Figure 26 - *Maps highlighting performance of Hybrid model with poor lighting; Pane 1: Example input drone image area, Pane 2: Hand-labeled mangrove ground truth, Pane 3: EfficientNet CNN Classifications, Pane 4: Hybrid Model Classifications*

6.2.3. Hybrid Model Recommendations

Based on the analysis in the previous section, we recommend that Hybrid models be used if complete drone imagery over the entire survey area is acquired. This scenario can occur when survey areas are easy to access, or the resources are available to record quality drone imagery over a large survey area. These favorable conditions are not guaranteed, but the Mexico dataset and survey serves as an example of such a survey area. As seen from our results on the Mexico dataset, using both drone and satellite imagery shows a significant improvement in mangrove extent classification when compared to using satellite or drone imagery alone.

6.3. Discussion

Although both the hybrid model and classical methods tested have high performance on their respective datasets, we have provided little direct comparison between the two methods. This is mainly since we do not have a dataset that has high quality drone imagery paired with higher resolution hyperspectral satellite imagery such as Planetscope-Superdove, since Superdove was not available for the 2019 Mexico dataset as shown in Table 1, and Drone imagery was incomplete and of lower quality for our Jamaica dataset. Therefore, we suspect that implementations of the hybrid model over the Jamaica dataset would results in mediocre results over an incomplete dataset, but these suspicions can be confirmed in future work. Therefore, if such a dataset becomes available, we would like to thoroughly compare classical methods and hybrid models. Our Jamaica dataset has drone imagery, which is used to create labels, but as previously discussed, the drone imagery has lower quality and is not over the entire survey area.

7. Conclusion

7.1. General Recommendations

To conclude, we recommend that, if possible, drone image acquisition surveys should be attempted for every expedition to generated accurate label data in the worst case, or a complete drone image dataset in the best case. Even a small amount of drone imagery can serve as a reference if labeling of satellite imagery is necessary. In addition, we recommend that satellite imagery be acquired over the entire survey area for every expedition. In many cases global hyperspectral satellite image sources such as Sentinel 2A imagery are free, with Planetscope-

Superdove being low cost, meaning that satellite data sources can be used as a data source for virtually all mangrove extent surveys.

7.2. Future Work

Also, Since the creation of our Hybrid Model, many new and advanced ML methods were published. These could be added to our exploration of models used to classify mangrove extent and compared to the Hybrid and Classical models. Segmentation algorithms such as UNet (Ronneberger, Fischer, & Brox, 2015) or DeeplabV3 (Chen, Papandreou, Schroff, & Adam, 2017) can classify imagery in a pixelwise manner, compared to a standard CNN which classifies whole tiles. Replacing tile based CNNs in the Hybrid model with a segmentation algorithm could allow of mangrove classifications to be the spatial resolution of drone imagery instead of the resolution of input satellite imagery.

Lastly, since we proved that additional generated features such as distance to water play a large effect in mangrove classification, we would like to implement this feature when implementing hybrid models on future datasets. Since a large overhaul of the data generation process is needed to implement these features, they were not included in this study, but these features could bring future improvements to the Hybrid model. Lastly, since Planetscope-Superdove satellites were not available during the Mexico survey, and with the Jamaica dataset having non-ideal drone imagery for the Hybrid model, we would like to test the Hybrid model with Planetscope-Superdove instead of Planetscope-Dove imagery in a future dataset, as we showed that the additional bands not available in Dove can play a large role in mangrove classification.

7.3. Summary

In this study, we implemented machine learning methods to create mangrove extent maps from image datasets of two types of mangrove survey scenarios: datasets created from surveys that have full low- and high-resolution imagery coverage, and datasets created from surveys that only have full low-resolution imagery coverage. For the case of full low-resolution coverage, we implemented several classical machine learning models that utilized 3m spatial resolution hyperspectral Planetscope-Superdove satellite imagery and extensive feature engineering. Our highest performing classical model, ERT, achieved an accuracy and IOU of 0.934 and 0.873, respectively, over our Jamaica dataset, having a superior performance, resolution, and a potential for higher frequency than Global Mangrove Watch labels. For the case of full low- and high-resolution imagery, we implemented a hybrid model that utilizes both satellite and drone imagery to effectively fuse 3cm high-resolution drone imagery with 3m low-resolution Planetscope-Dove satellite imagery. These hybrid models had the best performance of all models tested over our Mexico dataset, achieving an accuracy and IOU of 0.967 and 0.953, respectively. We established that Hybrid models were exceptional at providing mangrove extent classifications when both satellite and drone imagery is available, outperforming both standard CNNs and GMW labels while providing classifications at a resolution of 3m. Thus, satellite imagery can be used alone, or in tandem with drone imagery to allow for high performance mangrove extent classifications that can be generated frequently and accurately. Given the high variability in mangrove survey conditions, flexibility in data from methods to monitor mangrove extent outlined in this study can ensure that policymakers always have the recent and accurate information to establish mangroves as a nature-based solution our climate crisis.

REFERENCES

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon, 4*, e00938.

Airbus Intelligence. (n.d.). *Pléiades | Very High-Resolution (50cm) Satellites*. Retrieved from Airbus Intelligence: https://www.intelligence-airbusds.com/imagery/constellation/pleiades/

Alongi, D. M. (2002). *Present state and future of the world's mangrove forests.* Cambridge University Press.

Atkinson, P. M., & Tatnall, A. R. (1997). Introduction Neural networks in remote sensing. *International Journal of Remote Sensing, 18*, 699-709. doi:10.1080/014311697218700

Bai, T., Li, D., Sun, K., Chen, Y., & Li, W. (2016). Cloud Detection for High-Resolution Satellite Imagery Using Machine Learning and Multi-Feature Fusion. *Remote Sensing, 8*. doi:10.3390/rs8090715

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24-31. doi:https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bhatia, K., Vecchi, G., Murakami, H., Underwood, S., & Kossin, J. (2018). Projected response of tropical cyclone intensity and intensification in a global climate model. *Journal of Climate, 31*, 8281–8303.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32. doi:10.1023/A:1010933404324

Brenning, A. (2009). Benchmarking classifiers to optimally integrate terrain analysis and hyperspectral remote sensing in automatic rock glacier detection. *Remote Sensing of Environment, 113*, 239-247. doi:https://doi.org/10.1016/j.rse.2008.09.005

Bunting, P., Rosenqvist, A., Lucas, R. M., Rebelo, L.-M., Hilarides, L., Thomas, N., . . . Finlayson, C. M. (2018). The Global Mangrove Watch—A New 2010 Global Baseline of Mangrove Extent. *Remote Sensing, 10*. doi:10.3390/rs10101669

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. *CoRR, abs/1706.05587*. Retrieved from http://arxiv.org/abs/1706.05587

Chollet, F. (2015). Keras. *Keras*.

Das, S., & Vincent, J. R. (2009). Mangroves protected villages and reduced death toll during Indian super cyclone. *Proceedings of the National Academy of Sciences, 106*, 7357-7360. doi:10.1073/pnas.0810440106

Dash, J. P., Watt, M. S., Pearse, G. D., Heaphy, M., & Dungey, H. S. (2017). Assessing very high-resolution UAV imagery for monitoring forest health during a simulated disease outbreak. *ISPRS Journal of Photogrammetry and Remote Sensing, 131*, 1-14. doi:https://doi.org/10.1016/j.isprsjprs.2017.07.007

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 248-255). doi:10.1109/CVPR.2009.5206848

Donato, D. C., Kauffman, J. B., Murdiyarso, D., Kurnianto, S., Stidham, M., & Kanninen, M. (2011). Mangroves among the most carbon-rich forests in the tropics. *Nature Geoscience, 4*, 293–297. doi:10.1038/ngeo1123

Gao, B.-c. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*, 257-266. doi:https://doi.org/10.1016/S0034-4257(96)00067-3

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment, 32*, 2627-2636. doi:https://doi.org/10.1016/S1352-2310(97)00447-0

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*, 3–42. doi:10.1007/s10994-006-6226-1

Goldberg, L., Lagomasino, D., Thomas, N., & Fatoyinbo, T. (2020, October). Global declines in human-driven mangrove loss. *Global change biology, 26*(10), 5844-5855.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *CoRR, abs/1512.03385*. Retrieved from http://arxiv.org/abs/1512.03385

Hicks, D., Kastner, R., Schurgers, C., Hsu, A., & Aburto, O. (n.d.). Mangrove Ecosystem Detection using Mixed-Resolution Imagery with a Hybrid-Convolutional Neural Network.

Hsu, A. J., Kumagai, J., Favoretto, F., Dorian, J., Guerrero Martinez, B., & Aburto-Oropeza, O. (2020). Driven by Drones: Improving Mangrove Extent Maps Using High-Resolution Remote Sensing. *Remote Sensing, 12*. doi:10.3390/rs12233986

Hsu, A. J., Lo, E. K., Dorian, J. B., & Guerrero Martinez, B. (2019). Drone Flight Manual: UCSD Mangrove Imaging Procedure. Version 1.2.

Jian, X., Yunquan, Z., & Yue, Q. (2021). Remote Sensing Image Classification Based on Different Convolutional Neural Network Models. *2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT)*, (pp. 312-316). doi:10.1109/ISCIPT53667.2021.00069

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Ed.), *Advances in Neural Information Processing Systems. 25.* Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers, 7*, 3-10. doi:https://doi.org/10.1016/j.gsf.2015.07.003

Liu, K., Li, X., Shi, X., & Wang, S. (2008). Monitoring mangrove forest changes using remote sensing and GIS data with decision-tree learning. *Wetlands, 28*, 336–346. doi:10.1672/06-91.1

Mukherjee, N., Sutherland, W. J., Dicks, L., Hugé, J., Koedam, N., & Dahdouh-Guebas, F. (2014). *Ecosystem service valuations of mangrove ecosystems to inform decision making and future valuation exercises.* PLoS ONE.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2012). Scikit-learn: Machine Learning in Python. *CoRR, abs/1201.0490*. Retrieved from http://arxiv.org/abs/1201.0490

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution, 20*, 503-510. doi:https://doi.org/10.1016/j.tree.2005.05.011

Planet. (n.d.). *High-Resolution Imagery with Planet Satellite Tasking*. Retrieved from Planet: https://www.planet.com/products/hi-res-monitoring/

Purnamasayangsukasih, P. R., Norizah, K., Ismail, A. A., & Shamsudin, I. (2016, June). A review of uses of satellite imagery in monitoring mangrove forests. *IOP Conference Series: Earth and Environmental Science, 37*, 012034. doi:10.1088/1755-1315/37/1/012034

Richards, D. R., & Friess, D. A. (2016). Rates and drivers of mangrove deforestation in Southeast Asia, 2000–2012. *Proceedings of the National Academy of Sciences, 113*, 344–349.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR, abs/1505.04597*. Retrieved from http://arxiv.org/abs/1505.04597

SkyWatch. (n.d.). *Industry Leading Access EarthCache Pricing - SkyWatch*. Retrieved from SkyWatch: Commercial Satellite Imagery & API: https://skywatch.com/pricing/

Spalding, M. (2010). *World Atlas of Mangroves.* Routledge.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *CoRR, abs/1905.11946*. Retrieved from http://arxiv.org/abs/1905.11946

Tang, L., & Shao, G. (2015). Drone remote sensing for forestry research and practices. *Journal of Forestry Research, 26,* 791–797. doi:10.1007/s11676-015-0088-y

The European Space Agency . (n.d.). *Sentinel-2 - Missions - Resolution and Swath*. Retrieved from Sentinel Online: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath

The European Space Agency. (n.d.). *Planetscope - Earth Online*. Retrieved from Earth Online: https://earth.esa.int/eogateway/missions/planetscope

Thomas, N., Lucas, R., Bunting, P., Hardy, A., Rosenqvist, A., & Simard, M. (2017). Distribution and drivers of global mangrove forest change, 1996–2010. *PloS one, 12*, e0179302.

Tomlinson, P. B. (2016). *The Botany of Mangroves.* Cambridge University Press.

U.S. Geological Survey . (n.d.). *What is remote sensing and what is it used for?* Retrieved from https://www.usgs.gov/: https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used

Vaiphasa, C., Ongsomwang, S., Vaiphasa, T., & Skidmore, A. K. (2005). Tropical mangrove species discrimination using hyperspectral data: A laboratory study. *Estuarine, Coastal and Shelf Science, 65*, 371-379. doi:https://doi.org/10.1016/j.ecss.2005.06.014

Valderrama-Landeros, L., Flores-de-Santiago, F., Kovacs, J. M., & Flores-Verdugo, F. (2017). An assessment of commonly employed satellite-based remote sensors for mapping mangrove species in Mexico using an NDVI-based classification scheme. *Environmental Monitoring and Assessment, 190*, 23. doi:10.1007/s10661-017-6399-z

Wan, L., Zhang, H., Lin, G., & Lin, H. (2019). A small-patched convolutional neural network for mangrove mapping at species level using high-resolution remote-sensing image. *Annals of GIS, 25*, 45-55. doi:10.1080/19475683.2018.1564791

Wang, D., Wan, B., Qiu, P., Su, Y., Guo, Q., & Wu, X. (2018). Artificial Mangrove Species Mapping Using Pléiades-1: An Evaluation of Pixel-Based and Object-Based

Classifications with Selected Machine Learning Algorithms. *Remote Sensing, 10*. doi:10.3390/rs10020294

Winter, E. (2002). Chapter 53 The shapley value. Elsevier. doi:https://doi.org/10.1016/S1574-0005(02)03016-3

Yang, Y., & Newsam, S. (2010). Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 270–279). New York, NY, USA: Association for Computing Machinery. doi:10.1145/1869790.1869829

Yi, J., & Prybutok, V. R. (1996). A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental Pollution, 92*, 349-357. doi:https://doi.org/10.1016/0269-7491(95)00078-X