Plant Species and Pathogen Identification with Reflectance Spectra, UAV Imagery and Machine Learning

Mark Ryan
Level 4 Project, MPhys Physics and Astronomy
Supervisor: Dr Anthony M. Brown
Department of Physics, Durham University
Submitted: 17/04/24

This paper presents an investigation into the ability to discriminate between different plant species and identify pathogens within them using reflectance spectra for a sample of 798 mangroves from Suriname, UAV imagery and machine learning. Three species are present in the dataset, Red, White and Black, with the Red and Black mangroves being subdivided depending on their health status. A statistical approach and artificial neural networks (ANNs) are used to successfully identify the discriminating wavelengths between the species and trial-and-error methods are used to determine the reflectance boundary which categorises each species at each wavelength. The ANNs are then used to find the wavelengths which can distinguish between healthy and stressed mangroves within the Red and Black distributions. In each case, multiple approaches are produced to ensure the results are accessible regardless of budget, as many of the most discriminating wavelengths are in the near-infrared region meaning the necessary spectral cameras are expensive. Potential problems which arise when applying the model in practice are then considered and the effects of a 5% drop in the ambient light level is modelled, allowing uncertainties to be quantified on the results. This concludes the final results of the investigation, with the Red mangroves having the highest efficiency and purity of 78.3 ± 2.3% and 83.5 ± 1.3% respectively, for species classification and the Red healthy mangroves having the highest for the health status determination with 77.8 ± 2.8% and 96.7 ± 2.0% respectively. These values are reasonable and suggest that using the model to survey mangroves remotely will be an improvement, however, large misclassifications bring into question the scalability of the model, suggesting that further work is required with larger samples to create a more robust method.

Contents

1. Introduction

1.1. Mangroves

Mangroves are among the most important and productive ecosystems on the planet [1]. Typically, they can be found in the intertidal zone between the sea and land [2], along sheltered and shallow watered coastlines. Living in such an active zone means that mangroves must be robust and highly adaptable [3] so that they can survive through the daily disruptions caused by changes in tide level, tidal temperature, salinity, and oxygen content of the soil, to name but a few factors. To deal with the extreme environments in which they live, mangroves have developed many evolutionary traits which allow them to thrive where many ecosystems could not.

One of the most distinctive of these adaptations is the large root systems consisting of aerial stilt like roots and lateral roots extending sideways away from the stem [1]. These help to stabilise the mangroves [2] and are one of the reasons why mangroves are considered to be the best species for mitigating the effects of powerful tidal waves [4]. With coastal human inhabitation being strongly encouraged to be behind dense mangrove vegetation [4], the preservation and protection of these ecosystems is imperative and must be taken seriously. Mangroves having the ability to reduce the energy of waves can not only protect humans from the immediate threats of tsunamis, but also from the long-lasting future devastation of coastline erosion [5].

The aerial roots also create a habitat for both marine and terrestrial plants and animals [2] meaning mangroves are pivotal for coastal fisheries, providing breeding grounds and nurseries for fish, shrimp and many other marine species [6]. This allows many countries around the world, and particularly small villages, to survive as it provides a food source and income path for locals whose livelihood is reliant upon the ability to fish and farm the marine life [7].

Furthermore, mangroves have great potential in helping tackle global climate change. Blue carbon ecosystems (tidal and coastal ecosystems which undergo carbon sequestration) are extremely efficient carbon sinks and despite only mitigating 0.42% of global CO_2 emissions from fossil fuels in 2014, carbon sequestration from mangroves has been seen to negate over 1% of national carbon emissions in many countries [8]. The long-term impact of maintaining mangroves will help add to the carbon budget of these countries [9] and thus, will aid in the global efforts of meeting the terms of the Paris Agreement to tackle climate change [10].

Found in approximately 112 different countries and territories, mangroves follow a circumtropical distribution [11] and estimations in their global coverage sit at around eighteen million hectares [12].

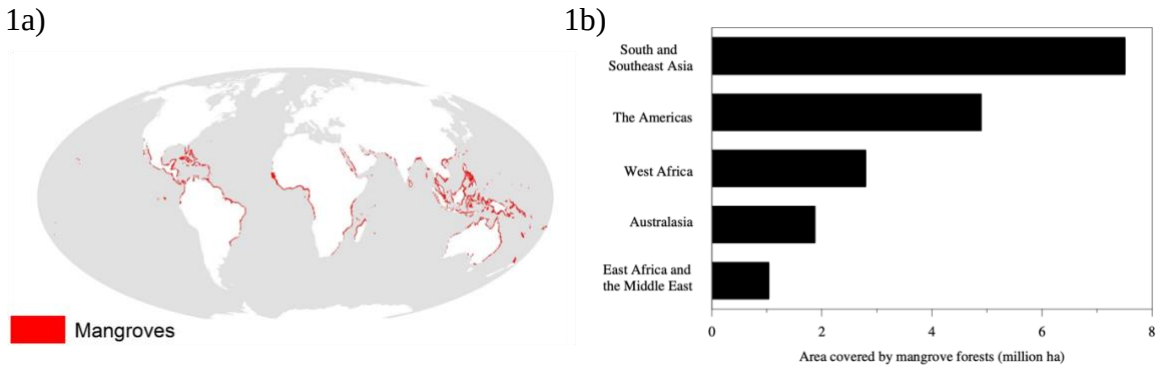

Figure 1: Global distributions of mangroves. Both charts show their extensive coverage around the world.

a) Map of Earth with red areas indicating populations of mangrove forests [13].

b) Bar chart detailing the area covered by mangrove forests across the globe [14].

This vast array of mangrove locations means that they hold a great responsibility in helping the planet flourish and most notably, battle climate change. With the global coverage area of mangroves decreasing by 1.04 million hectares between 1990 and 2020 [14], it is important that any methods of preserving and protecting mangrove ecosystems are used accordingly.

Currently, mangrove forests are being destroyed or left degraded at a rate that is 3-4 times faster than that of terrestrial forest types [15]. Based on current estimates, all forests globally would disappear within the next 200 years as between 2000 and 2012, the deforestation rate was calculated to be approximately 2×10^5 km^2 per year [16]. With the mangrove loss rate being much higher than this, methods of optimising the farming and maintenance of mangrove trees are critical if the ecosystems are to be saved.

Deforestation can greatly reduce the aforementioned carbon sequestration benefits and could result in releasing more carbon into the atmosphere than is being captured by the mangroves if the destruction rate becomes too high and more existing plants are destroyed before new ones replace them. Between 2000 and 2015, the loss of captured carbon by mangroves into the atmosphere due to mangrove habitat loss was estimated to be up to 122 million tonnes [17]. Based on the fact this is a 15-year total estimate, preventing this loss of stored carbon could help to extend the projected 2020 – 2045 global 1.5 ºC CO_2 emission budget of 400 billion tonnes by almost 1% over the 25-year period [18]. This is quite a large impact on the global scale and highlights why the conservation and growth of mangrove ecosystems is important for humanity and the future of the planet.

Furthermore, people in less developed countries like Suriname would suffer great consequences, losing food supplies, income sources and protection against the tides if mangrove ecosystems are not protected. With the majority of people in Suriname living in the coastal zone [12], much of the daily life of the inhabitants is governed by the quality of the marine life which mangroves help enhance. Exports from the fishing industry generate 2.2% of Suriname's GDP [19] which emphasises the importance of preserving this industry.

Moreover, the maintenance and growth of mangrove ecosystems can aid in Suriname's battle against the impact of sea level rise, an effect which the country is described as being highly vulnerable to [20]. It is estimated that almost a tenth of the Suriname population would be affected by just a one-metre rise in sea level, with this fraction rising to approximately half of the population for a five-metre rise [21]. Due to the robust nature of their roots and the ability to reduce wave energy from the tides, mangroves are expected to be crucial in Suriname's battle against the rising sea level [20]. While coastal erosion is responsible for the loss of mangroves, coastlines upon which mangroves live will suffer reduced erosion effects from the tides which will in turn extend their lifetimes. In theory, this should allow time for proper restoration and preservation methods to be implemented to counteract the effects of rising sea levels, which will be extremely cost-effective for society [22]. This combined with the global efforts to tackle global warming could help save coastal dependent civilisations like those in Suriname.

Suriname is home to approximately 1.6 - 2% of the world's mangroves [23] despite being the smallest country in South America [24], with this population spreading across three species of the tree: Rhizophora mangle (Red), Languncularia racemosa (White) and Avicennia germinans (Black). Each of these species have their own specific traits which give them each unique benefits to both the surrounding locals and the wildlife around. Red mangroves are particularly useful and are widely exploited commercially due to their dense wood being perfect for firewood, charcoal and even conversion to fabrics like rayon [25]. This species is typically found closest to the incoming tide and can be identified by its stilt roots which can take the initial impact of the tide. In general, Black mangroves are found in slightly higher elevations than Red mangroves and have a horizontal root system [26] which means that they are more susceptible to a rise in sea level compared to the Red mangroves as there is less chance of them having access to oxygen since they may not be tall enough to protrude through the water. White mangroves are located in even higher elevations and are more upland with no aerial root systems. The different species therefore require each other to help provide the optimal conditions for each of them to grow in. Despite this general separation of mangroves depending on species, classification between the species is not so simple with most areas consisting of a mix depending on the conditions of the environment and individual mangroves themselves [27]. Consequently, manual methods of species identification typically require trained professionals and can be very time-consuming with subjective opinions having to be given as each tree is individually surveyed [28]. This includes manually monitoring the health, age and species of each mangrove which could be inefficient given the lack of ease when it comes to determining features like health and age based on a visual assessment of the tree. For example, a recent decline in health may not be visible on the outside of the trees and so the recorded information for some of the mangroves being healthy could be incorrect. If this was the case, the error would likely not become apparent until a future survey which may not be until months later. Furthermore, with the mangrove coverage in even a small country like Suriname being so large, the possibility of missing (or equally, double counting) a section of trees is highly possible and would lead to inaccuracies in the data. With mangroves being pivotal in Suriname's battle

against the rising sea level [20], inaccuracies in surveys of their health conditions could have huge consequences. Tying into this, these surveys would take a long time to complete given the number of mangroves that would need to be assessed, meaning that the surveys would not be able to be completed regularly and so the data may not be reliable as it would become outdated. Combining these points suggests that the current methods of monitoring the mangroves is not efficient and could create a downfall in Suriname's attempts to preserve their mangroves, ultimately putting the country at a disadvantage in the battle against the rising sea level.

These current methods do not provide a quick and sustainable method for mangroves to be easily maintained which is a problem not only for the coastal villages who rely on the mangroves for both protection and general livelihood but also globally. If the regeneration of mangroves does not meet or exceed the thousands of hectares of mangroves that are lost per year [14] then, without restoration projects, the planet would lose more of the trees than are regrowing. Less mangroves would mean higher sea-levels which in turn would likely destroy more mangrove forests, as the roots may become submerged and not have access to oxygen. Moreover, a reduced number of mangroves could result in the sequestrated carbon being released back into the atmosphere. Throughout the entire year, mangroves shed and replenish their leaves, with the nutrients from the decomposition of the leaf litter being used to regenerate and grow the mangrove leaves [29]. If the number of mangroves is reduced, then there is less chance for the shed leaves to be recycled and so the carbon within them could be released back into the atmosphere and negate the benefits of the mangroves capturing atmospheric carbon. Different mangroves would suffer from different effects at different rates, for example, Black mangroves would be affected by sea-level rise before Red due to their horizontal root structure [26]. Meanwhile, Red mangroves are more effected by deforestation and natural mortality due to their inability to regenerate leaves on their larger or older branches compared to White and Black who can quickly produce new leaves and roots [30]. Each species can also have different stresses induced depending on changes in the physical and chemical sediment [26], meaning there truly is no 'one size fits all' approach to monitoring the forests. As a result of this, an improved method is required for mangrove maintenance which can easily determine the species of the mangroves and assess their health.

1.2. Aims and Motivation

As previously mentioned, the current methods of looking after and maintaining mangroves in less economically developed countries (LEDCs) are rather primitive, with those wishing to do so having to manually travel from tree to tree checking each of the mangroves individually. Due to the clear unsustainability of this approach, particularly when scaling it up to national or global scales, this investigation aims to explore the potential for a remote sensing approach to classifying the mangrove species and assessing their individual health status. With this approach, the idea will be to make use of UAVs (Unmanned Aerial Vehicles) in order to speed up the inspection process and remove the manual travelling through the mangrove forests. This will also allow more frequent surveys of the mangroves to take place due to their ease. As a

result, the ability to maintain the mangrove forests will be much easier since there will be a constant supply of up-to-date information about each of the mangrove trees.

It is important when developing a new approach that the accuracy of the data is not overly compromised for speed. Achieving the greatest efficiencies – percentage of a specific species (or health status) correctly categorised as itself – and purities – percentage of correct classifications of a specific species (or health status) out of the total number of those classified as that species (or health status) – was at the forefront of this investigation and the many steps taken to meet this demand are detailed throughout this paper. Optimisation of these accuracy metrics was done against the number of measurements required to make each classification. To make the product as effective as possible, it is important that the number of measurements required is kept to a minimum as this would make the measurement taking process faster, allowing for more frequent surveys and less wasted time.

1.3. Reflectance Spectra

Analysis of reflectance spectra from mangrove leaves is one method which can be used to study the health and conditions of the crops. Spectral characteristics differ depending on whether the mangroves are healthy or if they are suffering from physiological stress or disease [31], thus removing the subjective nature in assessing the tree's individual health. Hyperspectral data has been used to successfully quantify characteristics of a variety of different types of trees and has the potential to discriminate between different species of mangroves [32]. This is important as different species have different needs and uses, like Black mangroves being more useful than Red mangroves when it comes to stabilising the shorelines, for example [33].

Using a hyperspectral imager, like an ASD FieldSpec 4 spectrograph, it is possible to obtain a reflectance spectrum for mangroves. This spectrograph allows non-destructive measurements and is designed to take measurements whilst in the field [34]. Typically, they measure from optical wavelengths to the near infrared domain meaning there is a wide range of wavelengths (350 – 2500 nm) for which characteristics can be determined from [31]. For this data to be useful, the reflectance values measured by the spectrograph must be normalised relative to the ambient light when the measurement was taken. To maintain consistency this would usually be done by measuring the reflectance from a standardised white Spectralon panel with nominal reflectance 99% [35] before each set of measurements so that said set of readings can be calibrated by the ambient light in that area at the time of measurement. The data obtained could then be analysed using a variety of data analysis techniques to group different samples based on common characteristics in their reflectance spectra. These groups could then be compared to the metadata of the visual assessments of each mangrove plant to figure out which physical features are responsible for the characteristics in the spectral data. Manually, this would take a lot of time for the vast numbers of mangroves in Suriname, even after the key characteristics are found, however it would remove one of the key issues - misdiagnosing the health of the mangroves would now be less likely as it would no longer be a subjective opinion. The process

could be sped up by seeking a combination between the spectral approach and remote sensing by using UAVs equipped with spectral cameras [36]. Spectral cameras measure the reflected light at specific wavelengths. Choosing these specific wavelengths requires finding those which most discriminate between the different mangrove species and health statuses, whilst also maintaining a high spectral resolution to ensure the correct wavelengths from the correct trees are indeed being measured. This paper aims to find the best choice of these wavelengths, providing options for those with financial freedom, like large Non-Governmental Organisations (NGOs), and those with a limited budget, like an average mangrove farmer for example.

1.4. Overview of this Paper and the Desired Working Model

The goal of this project was achieved by finding defining characteristics of the reflectance spectra which can discriminate between the different physical features of mangroves based on a sample of 798 mangroves taken from five different locations in Suriname. The sample included three different species of mangrove (Red, Black and White) and relevant metadata describing a visual assessment of each tree's physical features was included. Reflectance data was taken using an ASD FieldSpec 4 spectrograph every nanometre from 350 – 2500 nm.

A manual method of data analysis was conducted on the reflectance data to find a range of reflectance values at a specific wavelength which corresponds to the mangroves being of a certain species. Once these wavelengths and the respective reflectance ranges were determined, the sample data set was passed through each devised range and an efficiency and purity was calculated. Once these values were determined, the same process was done but instead of categorising the different species, the goal was to find a range of reflectance values at a specific wavelength which corresponds to the mangroves being either 'healthy' or 'stressed'. To close, an assessment is then made as to whether the efficiencies and purities obtained are suitable and whether they are good enough for the investment to be made into using UAVs to help monitor the mangroves in Suriname.

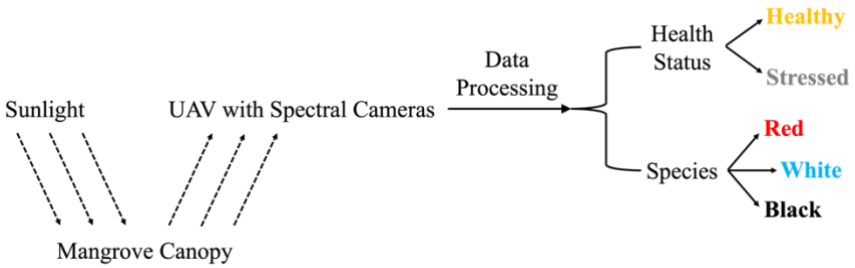

Figure 2: A simple diagram of the process that should be possible with the work in this report. By detecting the reflected light from the mangroves at the discriminating wavelengths found in this investigation, the health status and species of each mangrove tree can be deduced remotely.

2. The Dataset

To begin this investigation, the data for the sample of 798 mangroves in Suriname was plotted in three graphs separated by the species of the mangroves.

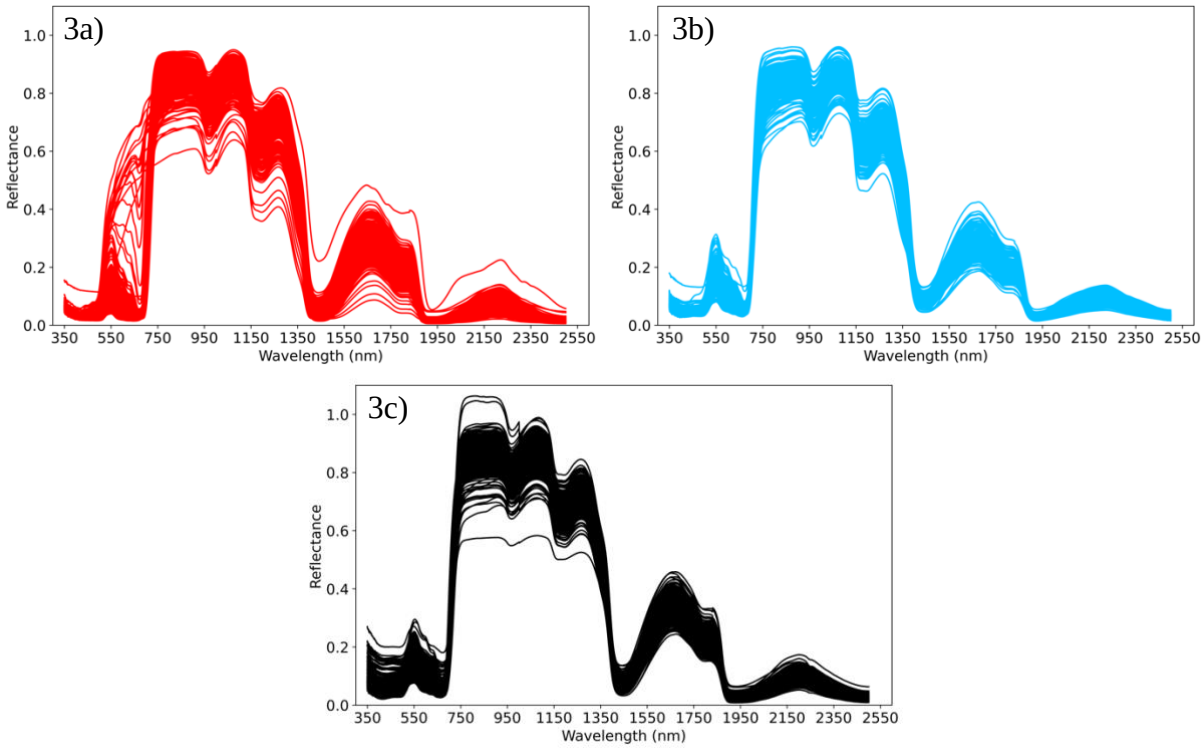

Figure 3: Reflectance spectra for each of the 798 mangroves split across three species plotted every nanometre from 350 – 2500 nm. Reflectance on the y-axes quantifies the fraction of incoming light that has been reflected off the mangrove leaf, normalised to the ambient light level at the time of each measurement. Each spectrum is comprised of 2151 data points across the wavelength range. Some slight differences can be observed between the spectra of the three species, but more in-depth analysis is required in order to identify clear discrepancies.

a) Reflectance spectra for the sample of 250 Red mangroves.
b) Reflectance spectra for the sample of 218 White mangroves.
c) Reflectance spectra for the sample of 330 Black mangroves.

The ASD FieldSpec 4 spectrograph has a varying resolution which depends on the wavelength it is measuring, with the trend being that there is a decrease in the wavelength precision as the wavelength increases. Furthermore, since the reflectance values are already normalised to the ambient light level at the time of their measurement, no uncertainty on the measurements is given, nor is the data available to calculate one on each reading. For these reasons, the decision was made to chunk the data into a number of different 'bins' depending on the wavelength and take the mean reflectance readings across the number that were in each bin. This would mean that the wavelength uncertainty for each bin, α_λ, was calculated via,

$$\alpha_\lambda = \pm \frac{Bin\ size}{2} \ , \tag{1}$$

and the uncertainty in the average reflectance in each bin, α_{Ref}, would be equal to the standard error [37] within in each bin,

$$\alpha_{Ref} = \frac{\sigma_{N-1}}{\sqrt{N}}, \tag{2}$$

where, N is the number of values within the current bin and σ_{N-1} is the standard deviation of the reflectance within the current bin. Decisions of bin sizes were made based on the precision of the spectrograph which, according to the EnMAP Field Guides Technical Report [34], grows from 3 nm at 700nm to 10 nm at 1400 – 2100 nm, with no details on the resolution before, after or in between these wavelengths. As a result of this, some extrapolation and assumptions had to be made to estimate the spectral resolution for all wavelengths from 350 – 2500 nm.

Table 1: Bin sizes and their respective uncertainties at each wavelength.

Wavelength Range (nm)	Bin Size	Uncertainty (± nm)
350 – 1050	3	1.5
1051 – 2100	10	5
2101 – 2485	12	6
2486 – 2500	15	7.5

The data was binned according to the categories in Table 1 and each bin was labelled using its median wavelength. This reduced the number of data points for each mangrove from 2151 to 372, with each of the data points now having uncertainties for its wavelength and reflectance.

3. Species Classification

3.1. Statistical Approach

3.1.1. Selection Criteria

In order to find characteristic wavelengths which discriminate between the different species, the variance of the reflectance at each wavelength were looked into. Lower variances in the reflectance at a specific wavelength for a particular species will indicate a tight range which is common to said species and so, investigating the reflectance in these areas could reveal a key feature of the reflectance spectrum which can be used to identify mangroves of that species.

As well as the variance, the p-values of each species against the other two species were used to infer whether there was statistical significance in there being a difference in the species in question against the other two. The p-values were taken at the 5% level and were calculated using independent t-tests which compare the means of two independent groups, i.e. the species in question and the other two combined as background [38]. Since the p-values were taken at the 5% level, wavelengths where the p-values are below 0.05 indicated areas where there is statistical significance suggesting that the species in question differs from the other two.

To determine a level of variance that was considered low enough to be of significance, the variance of the three species were compared. As seen in Figure 4, the variance in the Red mangroves has a large spike in the 500 – 750 nm range, reaching values more than four times the variances for the other two species. It would be expected that all three species would yield

similar variance values at each wavelength since the reflectance values are expected to be normally distributed within each bin for the mangrove samples. Considering the desire for a low variance when searching for the discriminating wavelengths, the desired variance in the queries were set to be less than 0.005 which approximately concurs with the maxima found in the White and Black mangroves. With this, there was now a set of criteria which narrowed down the number of potential wavelengths which were discriminating amongst the species.

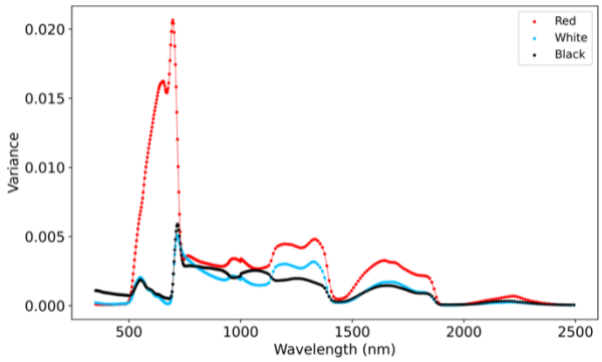

Figure 4: Variance of the reflectance within each species at each wavelength. Similar patterns can be seen across the three species other than a large spike in variance for the Red mangroves between approximately 500 - 750 nm which differs from the patterns observed in variance of the reflectance for the White and Black mangroves.

The wavelengths which fit into the p-value and variance criteria were then calculated for each species, and Figure 5 was developed to distinguish between these wavelengths. It can be seen clearly that though this method did in fact narrow down the potential discriminating wavelength bins, there were still hundreds for each species which were up for consideration and so a new approach needed to be developed. The full variance and p-values can be found plotted in Appendix A to help further interpret Figure 5.

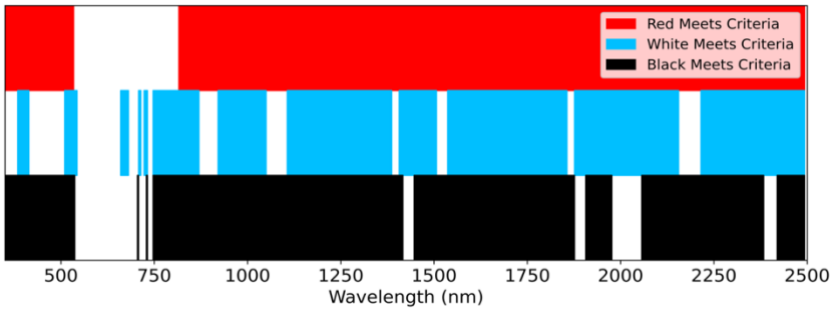

Figure 5: Distribution of wavelength bins for each species whose p-values (derived from independent t-tests) are less than 0.05 and variances are less than 0.005. Wavelength bins which meet these criteria are highlighted in the colour of the species that they are met for.

3.1.2. Indirect Discrimination

Since the specified p-value and variance criteria alone did not narrow down the number of wavelengths to be investigated, the variances, seen in Figure 4, were re-examined. This time, the spike in the Red variance between 500 – 750 nm was studied in more detail. The hypothesis causing this investigation was that a high variance may indicate outliers which are causing inaccuracies in the t-tests and causing potential characterising wavelengths to be overlooked.

Using a visual assessment of Figure 3, two wavelength bins within the 500 – 750 nm range that had large variances were chosen, 600 ± 1.5 nm and 651 ± 1.5 nm, and the reflectance data for each species was plotted.

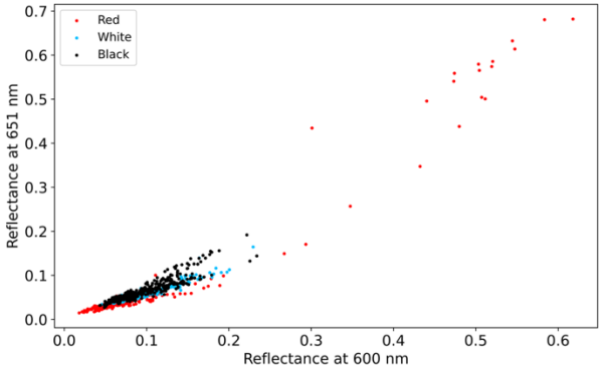

Figure 6: Reflectance spectra for the 651 ± 1.5 nm wavelength bin plotted against the 600 ± 1.5 nm for each species. A small group of reds in this range yield reflectance values much higher (up to approximately 0.7) than the maximum seen in both the White and Black (approximately 0.25).

Reflectance for each species in these two bins seemed to be relatively consistent, being confined to values below 0.3 in each bin for both the White and Black mangroves. Despite the Red mangroves following a similar trend to that of the other two species, a group of 17 Reds exceed the maximum reflectance observed within the White and Black mangroves and reach values of almost 0.7. This seemed anomalous given that it is only a small group and there was no obvious pattern in the metadata which could have caused the deviation in the reflectance values. The group had mangroves: from three different locations, eliminating the possibility that it was a problem with that location or the light at that particular site; of different health statuses, eliminating the possibility that the cause is solely due to them being stressed for example (an initial hypothesis made when this group was first discovered); and different ages, eliminating the possibility that the cause is that they are all old for example. A table summarising this information can be found in Appendix B.

As seen in Figure 7a, apart from the 500 – 750 nm range, the group of outliers follow the same trend of reflectance suggesting that the results in this range could be affected by random errors made while making the measurements. It is possible that errors could occur due to many factors, from unexpected changes in ambient light to stress on the investigators themselves while taking measurements caused by the environment. As a result, this group was classified as a set of outliers and removed from the sample. This decreased the size of the Red sample by 6.8% from 250 to 233, meaning that the sample size was still suitably large.

After removing these outliers, the variances of each of the species were again plotted to examine whether the removal of this group improved the sample. Comparing Figures 4 and 7b, a clear improvement can be seen after altering the Red sample as now the variances are much more consistent, with the maximum in Red being approximately 0.008 rather than 0.020. Also, the

patterns of the variances of the three species are more closely linked between 500 - 750 nm, with peaks and troughs being observed in similar locations.

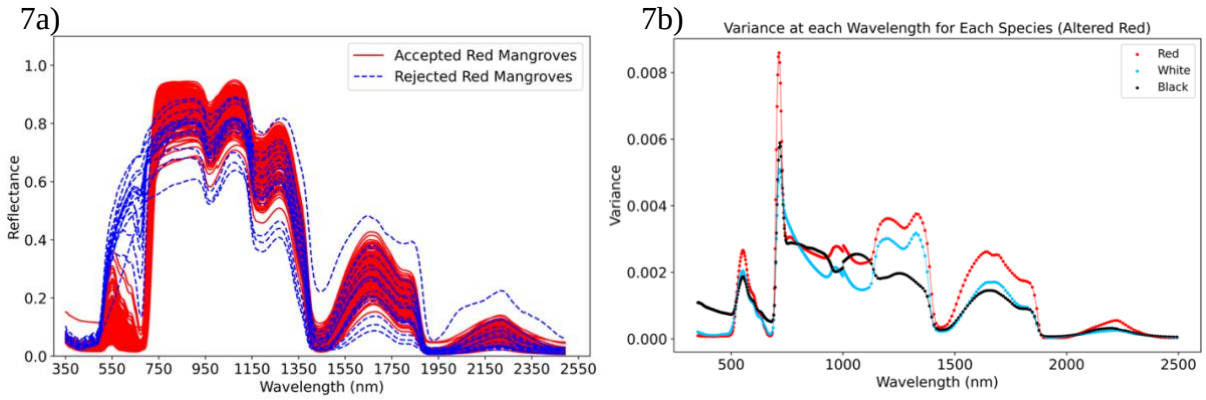

Figure 7:

a) *Reflectance spectra for the 250 Red mangroves with the 17 outliers highlighted in blue. The 17 outliers follow the same general patterns as seen for the rest of the Red mangroves apart from in the 500 – 750 nm range which suggest that the reflectance measurements for this group in this range may have been subject to random errors, explaining why they differ from the expected pattern.*

b) *Variances of the reflectance for each species at each wavelength. The Red sample has now been altered so that a group of 17 outliers has been removed and the variance of the altered Red group now follows the pattern and magnitude of the variance of the other two species much more closely.*

Using the altered group of Red mangroves, wavelength bins yielding suitable p-values and variances for each of the species were recalculated. Since the goal of the project is to develop a remote method of categorising the mangroves which improves the efficiency of the mangrove monitoring system, the least number of required measurements is desired as it means less readings are required per mangrove. As seen in Figure 8a, there are no wavelengths where the means of the three species are separated by two standard deviations meaning the normal distribution of each species will overlap within each wavelength bin making it difficult to distinguish between them. As a result, the next best method is to find combinations of wavelength bins which have the means of each species as far away from each other as possible, relative to their standard deviations. This method allows a Venn-diagram system to be set up so that each sample will fit into the criteria of up to two species (i.e. the central species and only one of the other species), with subsequent wavelengths separating these further.

With no wavelength bins having the individual means of the three species separated by the desired separation of two-sigma, the distances, d, from each species mean to each other species mean as a fraction of the standard deviation of the species was calculated for each species. The mean of the moduli of the fractional distances at each wavelength, $\bar{x}_i$, was then calculated using,

$$\bar{x}_i = \frac{|d_{i,a}|+|d_{a,i}|+|d_{i,b}|+|d_{b,i}|}{3}, \qquad d_{i,n} = \frac{\bar{r}_n - \bar{r}_i}{\sigma_i}, \tag{3}$$

where i represents the species in question, n is a placeholder for a and b which represent the two other species of the three, $\bar{r}$ is the mean reflectance of a species at each wavelength, and σ_i

is the standard deviation of the reflectance of the species at each wavelength. The mean of the moduli of the fractional distances were then plotted at each wavelength in Figure 8b.

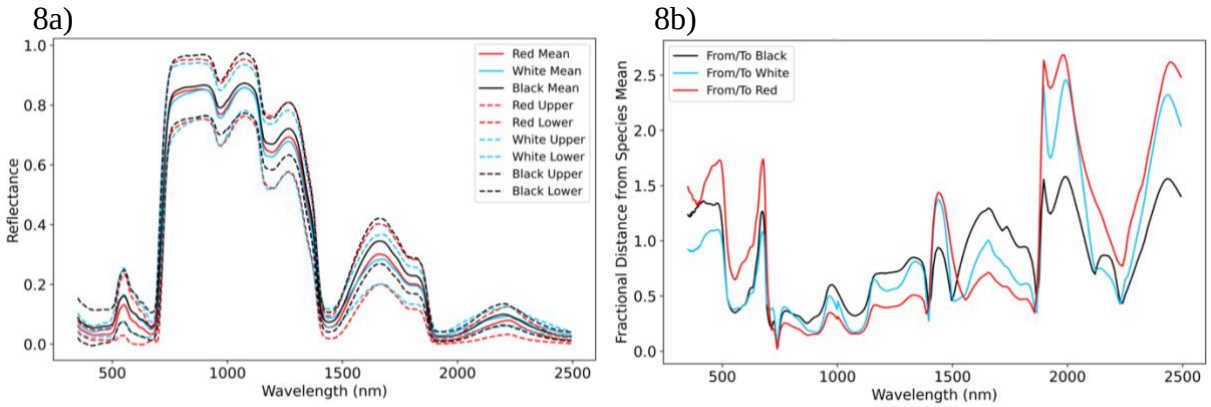

Figure 8:

a) The means and two-sigma range plotted for each species at each wavelength. The dashed lines represent the two sigma boundaries for each species, where 'upper' and 'lower' refer to the line where two standard deviations have been added or subtracted from the mean reflectance respectively. There are no areas where the three species are outside of the two-sigma range of each other.

b) Average distances from/to each species' mean as a fraction of the standard deviation of the species that each distance is taken from. Many peaks in the fractional distance between the different species can be seen through the spectra, with the largest being found for the Red mangroves at approximately 2000 nm.

Using both Figures 8a and 8b, wavelengths which yielded a high average fractional distance between each of the species were identified and combinations of those with different central species could be investigated to find their accuracy at distinguishing between the mangrove species. By observing the distribution of reflectance within these bins, boundaries were able to be set for each species through visual assessments combined with trial-and-error so that samples could be classified as a single species. Calculations could then be made to determine the efficiency and purity of each combination of wavelengths and the reflectance criteria for each.

3.1.3. Results

Whilst developing the combinations of wavelengths and reflectance values which defined each species, it was important to consider the final use case. Many of the wavelength bins which have the largest average fractional distances between species, are found in the near-infrared (NIR) region of the electromagnetic spectrum. NIR spectral cameras are very expensive when compared to their optical counterparts [39], which to large NGOs is not a problem but to an individual mangrove farmer or a scientific researcher, these cameras may not fit their budget. For this reason, a variety of species classification systems were developed: the 'best possible' results – the set of wavelength bins which produced the best results using the fewest number of wavelengths; the 'optical only' results – the best results that could be obtained on a strict budget where the possible wavelengths were restricted to being optical (up to ~700 nm); and the 'compromise' results – these often made use of a single NIR wavelength of around 1600 nm

(due to recent developments seeing lower cost NIR cameras of up to this wavelength being sold [40]) with the remainder of measurements being made in the optical regime.

Table 2: Results table for the three approaches to discriminating between the different species of mangroves developed using a statistical approach. The classification criteria can be found in each case, where 'R' represents the reflectance and the subscript is the wavelength at which the reflectance is measured.

	Best Possible	**Optical Only**	**Compromise**
Red Criteria	$R_{483 \pm 1.5\text{ nm}} < 0.041$ $R_{1976.5 \pm 5\text{ nm}} < 0.023$	$R_{483 \pm 1.5\text{ nm}} < 0.041$ $R_{402 \pm 1.5\text{ nm}} < 0.047$	$R_{483 \pm 1.5\text{ nm}} < 0.041$
Red Efficiency	83.7%	86.7%	86.7%
Red Purity	76.2%	66.2%	66.2%
White Criteria	$R_{1656.5 \pm 5\text{ nm}} < 0.34$ $R_{1976.5 \pm 5\text{ nm}} > 0.023$	$R_{483 \pm 1.5\text{ nm}} > 0.041$ $R_{402 \pm 1.5\text{ nm}} < 0.047$	$R_{483 \pm 1.5\text{ nm}} > 0.041$ $R_{1656.5 \pm 5\text{ nm}} < 0.34$
White Efficiency	86.7%	60.1%	72.5%
White Purity	68.5%	48.3%	51.5%
Black Criteria	$R_{483 \pm 1.5\text{ nm}} > 0.041$ $R_{1656.5 \pm 5\text{ nm}} > 0.34$	$R_{483 \pm 1.5\text{ nm}} > 0.041$ $R_{402 \pm 1.5\text{ nm}} > 0.047$	$R_{483 \pm 1.5\text{ nm}} > 0.041$ $R_{1656.5 \pm 5\text{ nm}} > 0.34$
Black Efficiency	45.5%	47.9%	45.5%
Black Purity	88.8%	77.1%	88.8%

As seen in Table 2, developing criteria based on a Venn-diagram approach, where multiple wavelengths could be used to define each species, caused a great improvement in the purities achieved in all cases, with the Black mangroves obtaining the highest purity and White the lowest in each approach.

As expected, the 'best possible' approach achieved the best results overall, managing to distinguish between the different species to a reasonable degree, yielding high efficiencies for both Red and White mangroves and high purities for all three species. The efficiency for the Black mangroves was poor in all cases, with the highest result being drawn from the 'optical only' approach which had less than half of the Black mangroves fit into the boundaries and perhaps means that the criterion for this species is too strict and should be reviewed.

Furthermore, despite the large improvement in the purities, when scaled up to the eighteen million hectares of mangroves around the world [12], even the highest purity achieved of 88.8% would result in over two million hectares of mangroves being mislabelled. This highlights the importance of finding the perfect wavelength bins and the ideal boundaries between species. It also suggests that further work should be done to improve the efficiencies and purities.

In the 'best possible' approach, several samples for each species (23 Red, 8 White and 49 Black) did not fit into any of the categories which would be a problem on large scales as in the Black case, this would mean approximately 15% of the trees are missing from each survey. One potential solution on smaller scales, would be to first survey all of the Black mangroves manually and then use the UAV method on the rest. This would mean that the poor efficiency of the boundaries categorising the Black mangroves would not affect the monitoring process. Furthermore, by doing this one would increase the purity in both the Red and White mangrove classification using the UAV method as the 16% and 23% of Black mangroves that were characterised as Red and White respectively would be removed from the sample.

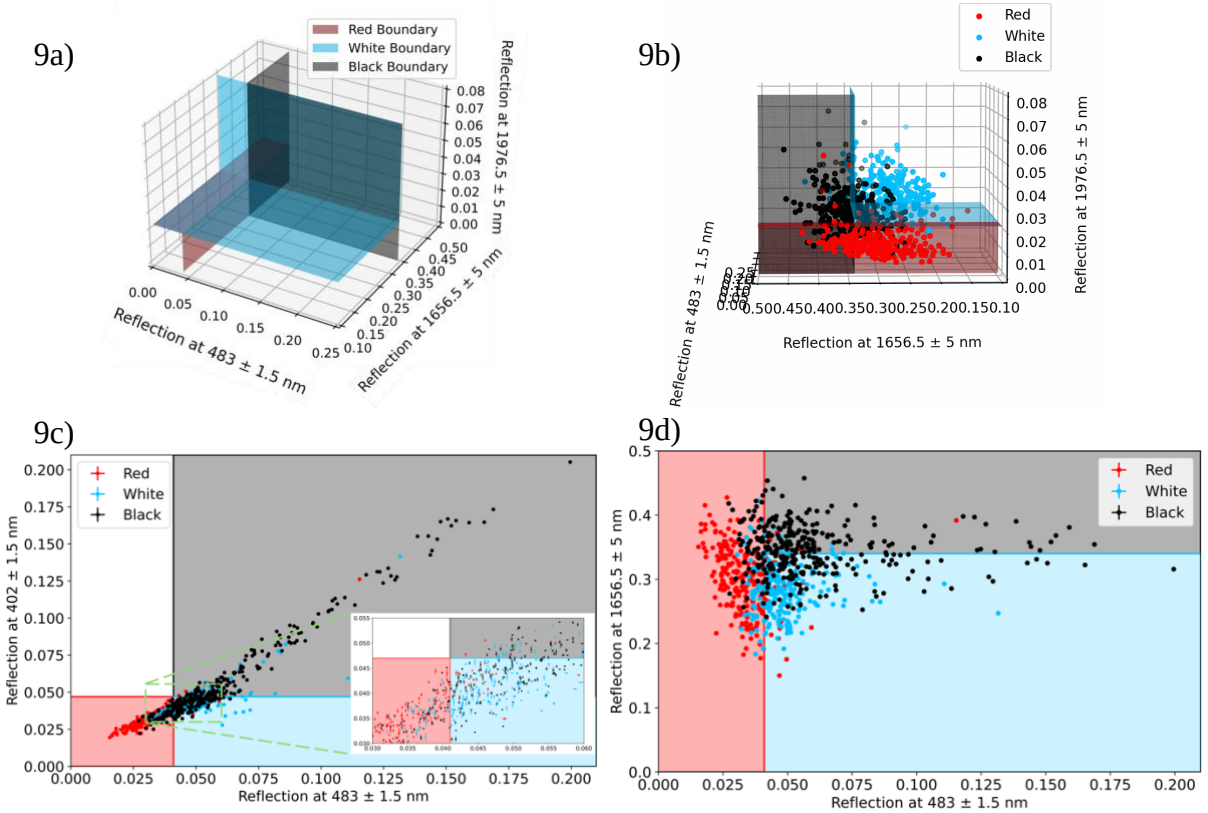

Figure 9: Distributions and species boundaries for each of the three approaches determined via a statistical approach. The equations for the boundaries can be found in Table 2. Vertical and horizontal error bars are plotted on each point. In all cases, much of the distribution is clustered around the boundary intersection meaning slight shifts in measurements could alter the accuracy of the results obtained.

- *a) 'Best possible' (boundaries only): 483 ± 1.5 nm, 1656.5 ± 5 nm and 1976.5 ± 5 nm.*
- *b) 'Best possible': 483 ± 1.5 nm, 1656.5 ± 5 nm and 1976.5 ± 5 nm. See Appendix C for a link to a video tour of this plot.*
- *c) 'Optical only': 483 ± 1.5 nm and 402 ± 1.5 nm. The zoomed in portion gives an insight into the misclassification problem.*
- *d) 'Compromise': 483 ± 1.5 nm and 402 ± 1.5 nm.*

The results obtained by the statistical approach provide a quick method at distinguishing between the three mangrove species at the expense of increased cost and decreased accuracy, when compared to the current manual methods. In terms of the practicality of these results, without the financial freedom to afford NIR spectral cameras with good spectral resolution,

these results may not provide much use. The 'optical only' results are poor when it comes to categorising the White and Black mangroves, meaning that this statistical approach does not suffice when trying to keep the results accessible. As a result, a new approach to finding the discriminating wavelengths was attempted which made use of machine learning (ML) methods.

3.2. Machine Learning

3.2.1. Neural Networks

Humans are limited when it comes to the patterns they can identify and the number of parameters they can visualise, however ML algorithms have the ability to analyse a far greater parameter space than humans, boasting very high performance when it comes to data analysis [41]. There are many different ML algorithms, each with different strengths and weaknesses, making some more suited to certain situations than others and for this investigation in particular, neural networks (artificial neural networks to be specific) were the algorithm of choice. This choice was made as there are many papers which use artificial neural networks (ANNs) for species classification problems, for example 'Relationship between Hyperspectral Measurements and Mangrove Leaf Nitrogen Concentrations' by Zhang et al [42] makes use of an ANN to separate Red and Black mangroves, thus providing lots of avenues to draw comparisons from as an attempt is made to develop a solution to the classification problem.

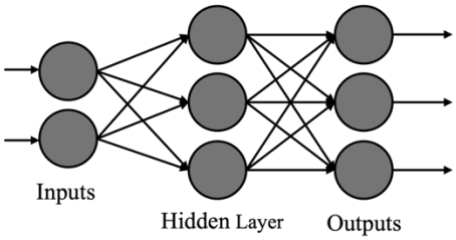

Figure 10: Diagram of a simple artificial neural network (ANN) containing an input layer consisting of two neurons, a single hidden layer consisting of three neurons and an output layer consisting of three neurons.

ANNs imitate the human brain by processing data through a chain of interconnected neurons [43]. These neurons are separated into several different layers, with the neurons in each layer being individually connected to each of the neurons in subsequent layers (as seen in Figure 10). This creates a feedforward path to the output in the network where the inputs are fed into the input layer before passing through the hidden layers. Each hidden layer has a Rectified Linear Unit (ReLU) activation function – the most successful and widely used activation function [44] – applied to it which introduces non-linearity to the network, allowing it to learn. This continues through the hidden layers until it reaches the outputs [45]. The output of the neural network is then compared to the known, desired output and an accuracy calculation is made using the cross-entropy loss to measure the difference (error) between the two [43]. A gradient descent algorithm called 'Adaptive Moment Estimations' (also known as 'Adam') was then used to calculate and update the weights and biases in the direction which should minimise the loss of the network, thus improving its performance. This algorithm was chosen due to the high speed at which it can converge on the optimal solution [46] and this process is essentially what allows

the neural network to learn with each iteration. The ANN then continues in this process and gradually adjusts its parameters to get closer to the desired output until it has completed all of its epochs. The size of the adaptations and adjustments made to the network are determined by the learning rate (LR) [43] which makes the optimisation of this particular parameter critical.

3.2.2. Hyperparameter Optimisation

Hyperparameters are the parameters in an ANN which define the model [41]. They include the number of hidden layers, neurons in each layer, epochs, as well as the learning rate and the split of data between the training and testing groups. These must be manually set as unlike the model parameters (like weights and biases), the ANN does not tune them as it progresses. As a result, methods of tuning these hyperparameters and finding their optimal values had to be developed. This is a time-consuming process [41] and as a result of this investigation requiring many different ANNs to individually find the best wavelength to distinguish a particular species from the other two (and then later discriminate between 'healthy' and 'stressed' mangroves), some compromises had to be made in the interest of time and computational power available.

To make the hyperparameter optimisation in this investigation efficient and applicable in all scenarios required to find the best wavelengths to discriminate between the mangrove features, the network architecture needed to be flexible so that it was easily optimised for each task. This meant some trialling and testing of hyperparameters before running the ANN and a series of loops were set up to run the dataset through different combinations of hyperparameter values. The accuracy, efficiency and purity of each combination were then compared so that the optimal solution could be found in each case which helped to avoid any overfitting which may have occurred if the model was not flexible.

As stated earlier, the LR is an important hyperparameter which needs optimising [41]. A simple approach to find the best LR was chosen by which the LR was set to its maximum value, one, for the first trial and then the LR for the next trial would be taken to be a tenth of this value [47]. In doing this, the entire dataset was passed through the ANN five times, each time with a different LR, allowing results to be generated for LRs of 1, 0.1, 0.01, 0.001 and 0.0001. Five different options were chosen in an attempt to draw a compromise between the performance, as the more LRs tested the more chance of finding the optimal value, and the run time for the algorithm. These values also cover the majority of the optimal range for the LR found in a paper by Ghorbanian et al when using ANNs to map mangroves using remote sensing [48].

For each LR, the data would be processed through an ANN which could have up to three layers, with each layer containing either one, two, four or eight neurons and each of the 84 combinations were individually trialled. When choosing these options, it was important to try to cover as many combinations as possible whilst also considering the time the algorithm would take to test each one. The decision to set the maximum number of hidden layers to three was drawn from a combination of literature comparisons, for example Zhang et al used a single

hidden layer in an ANN to distinguish between Red and Black mangroves [42] and Ghorbanian et al mapped mangroves using an ANN with four layers [48], along with the trend that an increased number of layers increases execution time of the ANN [41]. In terms of the numbers of neurons in these layers, the decision to use powers of two from one to eight was made in order to balance model complexity with computational time as more neurons results in longer training and recalling time [49], since more weights and biases need updating within each iteration. This range of neurons attempted to avoid overfitting, which was important since there was just a single input into the network, and kept towards the lower range of those found in Zhang et al's [42], Nezami et al's [50] and Ghorbanian et al's [48] investigations into similar classification tasks, which all started their range of neuron numbers between six and ten. By looping through options for the hyperparameter values and combinations, the results achieved by each model were tracked and the combination of LR and numbers of hidden layers and neurons which gave the highest accuracy could be easily selected out of the 420 total options.

Once loops for the LR, number of layers and number of neurons were set up, the optimal number of epochs was found using a trial-and-error system, where the evolution of the loss function across the epochs was assessed to find the number of epochs required to cause this function to plateau. Once the loss function reaches a plateau, having more epochs would spend valuable computational time on negligible improvements. Using Figure 11a, the decision was made to set the number of epochs to be 100 for all variations of the ANN, ensuring that the loss function was examined at each run to check that each variation followed the pattern of plateauing before 100 epochs (and this was true in all cases).

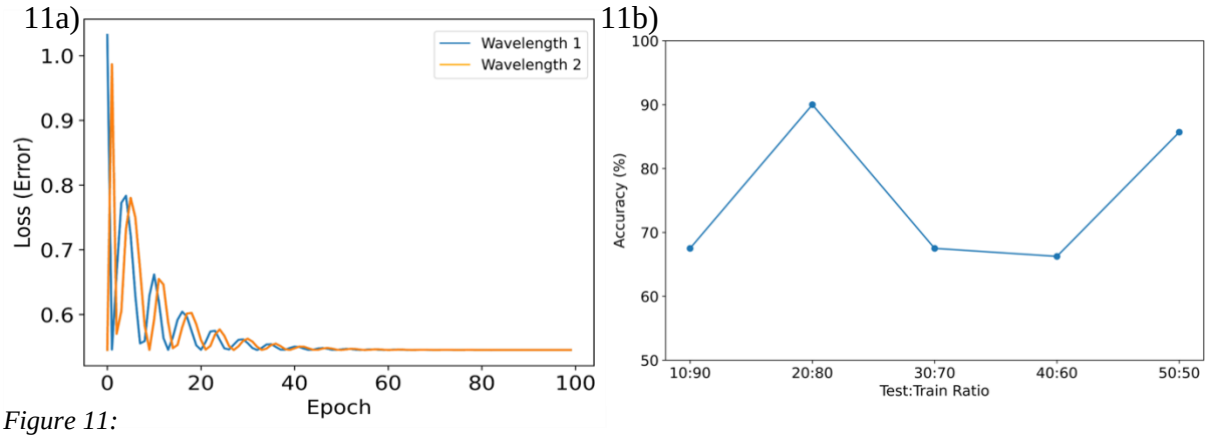

Figure 11:

a) Cross-entropy loss plotted at each epoch for an arbitrary combination of learning rates, numbers of layers and neurons. The cross-entropy loss for just two of the 372 wavelength bins (chosen at random and labelled 'Wavelength 1' and 'Wavelength 2') has been plotted, one in blue and one in orange, to make the pattern of dampening of oscillations easy to see. The error in the artificial neural network plateaus after approximately 50 epochs, with the oscillations in the loss function settling at this point.

b) Accuracies for five different test-train split ratios. A 20%:80% test:train ratio produced the highest accuracy of 90% suggesting this is the optimal split for the dataset. This means that 638 of the mangroves had their reflectance data used to train the ANN, whilst the other 160 were used for testing.

Initially the percentage of the dataset used to test the network was set to 20% as this was consistent with the testing percentage found in both Zhang et al's [42] and Nezami et al's [50] papers on similar investigations. With this initial value, the optimal LR and numbers of hidden

layers and neurons was found so that they could be used to find the optimal train-test split for the dataset available in the investigation in this paper. The ratio of testing data to training data was then varied between a 10% to a 50% testing sample in intervals of 10% for an ANN which distinguished the Red mangroves from the White and Black.

As seen in Figure 11b, the investigation into finding the optimal test-train split of the dataset supported the initial values which were drawn from literature comparisons. Consequently, the decision was made to have 160 random test samples (20% of the dataset) and 638 random training samples (80% of the dataset) for all the ANNs used throughout the investigation.

3.2.3. Results

Four ANNs were set up, one which attempted to distinguish between all three species using a single wavelength and then three others which each focused on distinguishing a different species from the other two. To avoid introducing any bias into the dataset, the 17 Red mangroves that were suspected outliers in the statistical approach were not removed as without any definitive proof confirming these as anomalous, removing them may cause the ANN to have unpredictable behaviour [42] and its results may not be reproducible.

Table 3: Results table for the three approaches to discriminating between the different species of mangroves developed using artificial neural networks. The classification criteria can be found in each case, where 'R' represents the reflectance and the subscript is the wavelength at which the reflectance is measured.

	Best Possible	**Optical Only**	**Compromise**
Red Criteria	$R_{1986.5 \pm 5\,\text{nm}} < 0.019$	$R_{483 \pm 1.5\,\text{nm}} < 0.041$	$R_{483 \pm 1.5\,\text{nm}} < 0.041$
Red Efficiency	76.0%	86.7%	86.7%
Red Purity	84.8%	66.2%	66.2%
White Criteria	$R_{1986.5 \pm 5\,\text{nm}} > 0.033$	$R_{402 \pm 1.5\,\text{nm}} < 0.045$ $R_{678 \pm 1.5\,\text{nm}} > 0.043$	$R_{483 \pm 1.5\,\text{nm}} > 0.041$ $R_{1686.5 \pm 5\,\text{nm}} < 0.31$
White Efficiency	84.9%	61.5%	61.9%
White Purity	68.3%	47.3%	60.5%
Black Criteria	$R_{1986.5 \pm 5\,\text{nm}} > 0.019$ $R_{1986.5 \pm 5\,\text{nm}} < 0.033$	$R_{504 \pm 1.5\,\text{nm}} > 0.060$	$R_{1686.5 \pm 5\,\text{nm}} > 0.31$
Black Efficiency	64.8%	63.0%	79.4%
Black Purity	70.6%	60.5%	63.7%

As seen in Table 3, using ANNs to locate characterising wavelengths equalled or improved all approaches compared to the statistical approach, with the Black efficiencies being much higher. As expected, the 'best possible' approach achieved the best results overall, with them managing to distinguish between the different species fairly well, yielding high efficiencies for both Red and White mangroves and high purities for all three species (with the Red's purity being

particularly high). This means that if a particular sample is categorised as a Red mangrove based on these criteria, there is an 84.8% chance that it is accurate which is very good when considering the time saved by using this remote sensing method. Moreover, the 'best possible' method also only uses a single wavelength to separate the species meaning that surveys of the mangrove forest can be done faster and more frequently since less measurements are required. As a result, repeat surveys can be done which would improve the accuracy of the classifications. Furthermore, the use of ANNs has made the results of this investigation more accessible. Thanks to the ease at which discriminating wavelengths can be found by the ANNs, four different wavelengths were used to close the gap to the most expensive, 'best possible', results

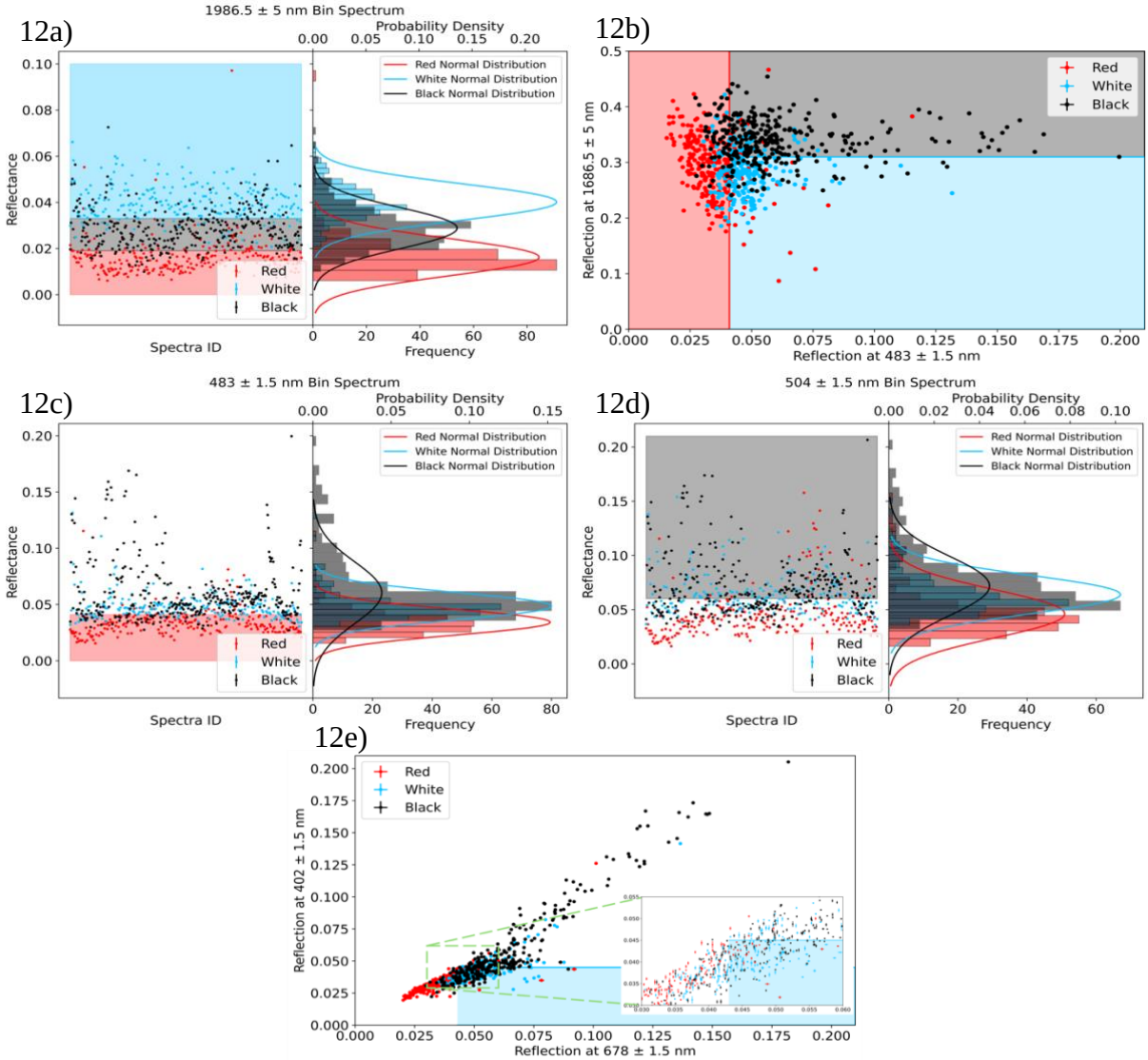

Figure 12: Distributions and species boundaries for each of the three approaches determined via a machine learning approach. The equations for the boundaries can be found in Table 3. Vertical and horizontal error bars are plotted on each point. In all cases, much of the distribution is clustered around the boundary intersection meaning slight shifts in measurements could alter the accuracy of the results obtained.

a) 'Best possible': 1986.5 ± 5 nm.
b) 'Compromise': 483 ± 1.5 nm and 1686.5 ± 5 nm.
c) 'Optical only': 483 ± 1.5 nm.
d) 'Optical only': 504 ± 1.5 nm.
e) 'Optical only': 402 ± 1.5 nm and 678 ± 1.5 nm.

so that now the more affordable, 'optical only' results do not lag so far behind. Though they are still clearly not as good at discriminating between the different species, with the White purity still falling below 50%, the majority of mangroves can be separated into their correct species, which may be useful to independent mangrove farmers who are on a tight budget.

Additionally, alongside improvements to the 'optical only' model, the 'compromise' model was now competitive with the 'best possible' approach, yielding good results in all cases. As before, this compromise made use of a single optical wavelength and a single infrared wavelength (close to the maximum wavelength of the lower cost cameras mentioned earlier [40]). This was important as now the work in this investigation was not exclusive to those on large budgets.
As previously mentioned, despite the large improvements that have come from using machine learning, the usefulness of the model is still brought into question when scaled up to the global level, suggesting that a more extensive study using a much larger sample of mangroves may be required to improve upon the results found in this investigation.

4. Health Classification

With suitable results having been achieved to classify the mangroves according to species, the development of a method to assess the trees' health statuses was attempted. As a result of the ML approach achieving results that were better than those found using the statistical approach much more easily, the same ANN model was used to discriminate between the health statuses. If successful, a complete and cohesive mangrove maintenance model would be produced as well as an insight into whether ML methods are worthwhile for this type of problem.
For this section of the investigation, the dataset was separated into the three different species so that the discriminating criteria displaying the health status within each species could be found. This decision came after an initial attempt at integrating over all species to find discriminating wavelengths for the health of the general population was unsuccessful due to no bins yielding a separation between the general spectra of the healthy and stressed mangroves.

Upon assessing the dataset, it was found that all of the White mangroves in the sample were healthy and so the trees of this species would be of no use in this part of the investigation. The Red and Black mangroves both had sufficient numbers of healthy (200 and 180 respectively) and stressed (50 and 150 respectively) mangroves and so analysis into finding the wavelengths which could distinguish between healthy and stressed mangroves within these two species was conducted. This would still allow a complete and cohesive mangrove maintenance model to be created as once the species was determined, the health status could then be found.

13a) 13b)

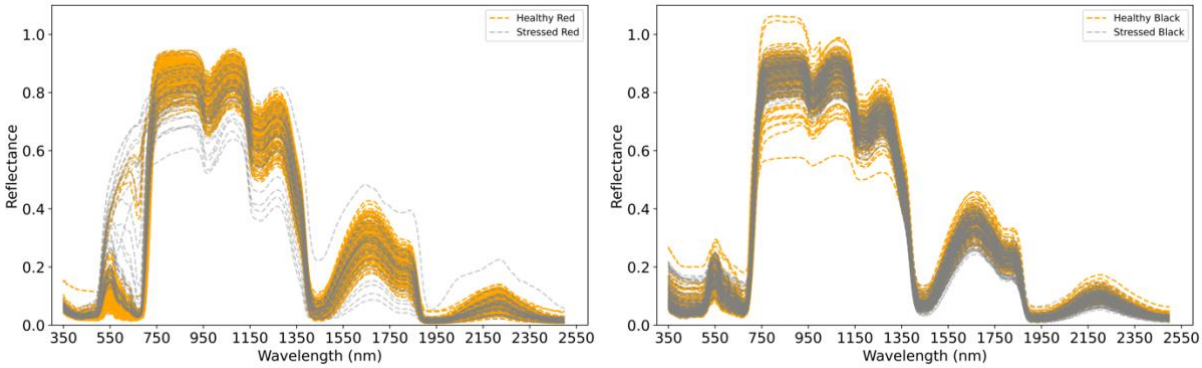

Figure 13: Reflectance spectra for the healthy (orange) and stressed (grey) mangroves split between the Red and Black mangroves plotted every nanometre from 350 – 2500 nm. Reflectance on the y-axes quantifies the fraction of incoming light that has been reflected off the mangrove leaf, normalised to the ambient light level at the time of each measurement. Each spectrum is comprised of 2151 data points across the wavelength range.

a) Reflectance spectra for the sample of 250 Red mangroves.

b) Reflectance spectra for the sample of 330 Black mangroves.

Table 4: Results table for the two approaches to determining the health status of Red and Black mangroves developed using artificial neural networks. The classification criteria can be found in each case, where 'R' represents the reflectance and the subscript is the wavelength at which the reflectance is measured.

		Best Possible	Optical Only
Red	**Healthy Criteria**	-	$R_{699 \pm 1.5\ \text{nm}} < 0.155$
	Healthy Efficiency	-	75.0%
	Healthy Purity	-	98.7%
	Stressed Criteria	-	$R_{699 \pm 1.5\ \text{nm}} > 0.155$
	Stressed Efficiency	-	96.0%
	Stressed Purity	-	49.0%
Black	**Healthy Criteria**	$R_{2493 \pm 7.5\ \text{nm}} > 0.023$	$R_{522 \pm 1.5\ \text{nm}} < 0.090$
	Healthy Efficiency	64.4%	48.9%
	Healthy Purity	62.4%	63.3%
	Stressed Criteria	$R_{2493 \pm 7.5\ \text{nm}} < 0.023$	$R_{522 \pm 1.5\ \text{nm}} > 0.090$
	Stressed Efficiency	53.3%	61.5%
	Stressed Purity	55.6%	50.8%

As seen in Table 4, the ANNs were able to find an optical wavelength (699 ± 1.5 nm) which could determine the health status of Red mangroves very well. By using the reflectance criteria set at this wavelength, only 4% of the stressed Red mangroves were misclassified, which is rather promising as it means that despite only having a 49.0% purity in the Red mangroves that are classified as stressed, one can be confident that almost all of the stressed Red mangroves

have been identified and those that have been classified as healthy are correct. This means that a manual assessment on those that have been identified as stressed can then be conducted which is a huge improvement as now the number requiring a manual assessment has dramatically decreased, saving lots of time particularly when scaled up to the global population.

Determining the health status of the Black mangroves proved to be more difficult, with the distributions between the healthy and stressed displaying very little separation as seen in Figures 14b and 14c. As a result, in both the 'best possible' and 'optical only' cases, just under half of the stressed mangroves are misclassified as healthy. Although the health status for majority of the Black mangroves is correctly identified in both the 'best possible' and 'optical only' case, the utility of this model must be questioned because a large portion of stressed mangroves being misclassified as healthy will mean that the manual assessments to find false positives within those classified as stressed will leave 46.7% and 38.5% of the stressed mangroves unidentified for the 'best possible' and 'optical only' cases respectively. Therefore, while identifying just over half of the stressed mangroves may be useful on small scales, when deploying this onto the globally, the large number of unidentified stressed mangroves could lead to mass decline in the mangrove population, thus causing substantial challenges for marine ecosystems and coastal towns who are reliant on mangroves to protect their coastlines.

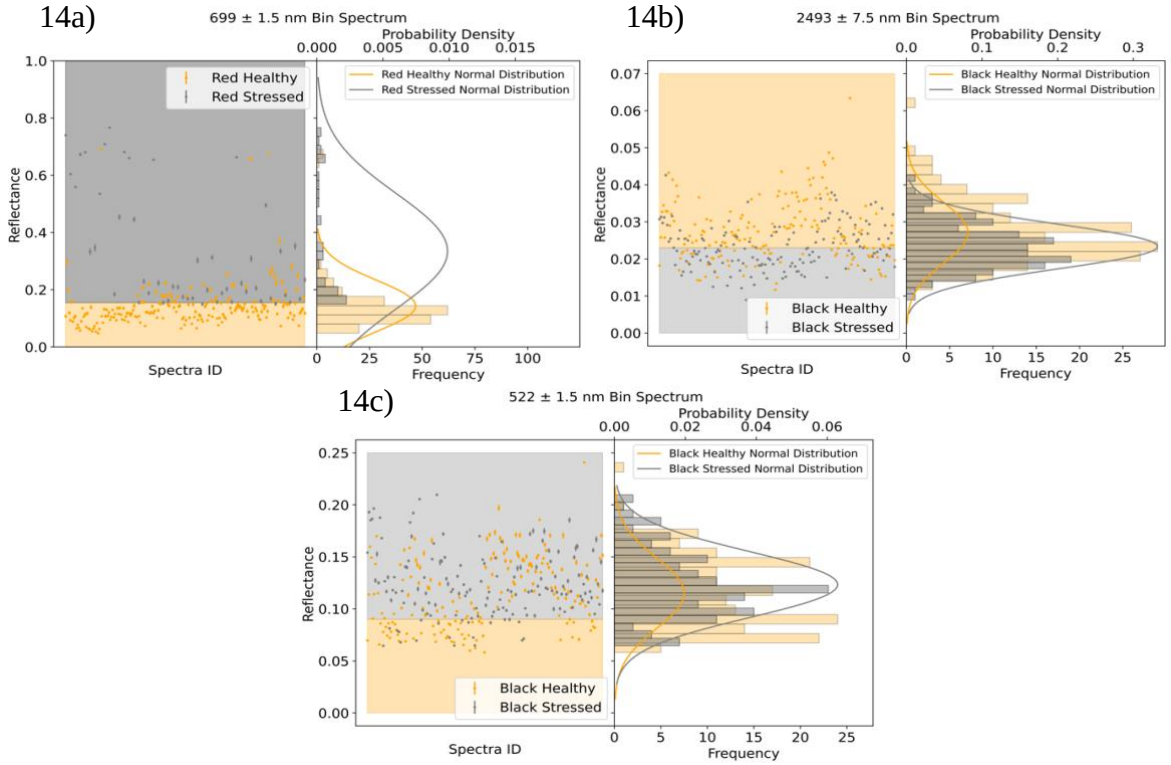

Figure 14: Distributions and species boundaries for each of the two approaches determined via a machine learning approach. The equations for the boundaries can be found in Table 4. Vertical and horizontal error bars are plotted on each point.

a) Red health status – 'Best possible' & 'Optical only': 699 ± 1.5 nm.
b) Black health status – 'Best possible': 2493 ± 7.5 nm.
c) Black health status – 'Optical only': 522 ± 1.5 nm.

5. Complete Mangrove Maintenance Model

To complete the mangrove maintenance model, some considerations must be made to consider how, if at all, the results achieved throughout this investigation will change if the discriminating wavelengths from this investigation are used in practice.

Firstly, using an UAV to take reflectance measurements would introduce errors due to atmospheric turbulence in the reflected light. Atmospheric turbulence of the sunlight on its path towards the leaf occurs for both methods, however, since the UAV takes reflectance data at an altitude while the manual spectrograph takes readings at ground level, there is an additional propagation distance for the reflected light travelling up towards the UAV where it experiences atmospheric turbulence. A balance of the perfect height for the drone flight must be struck since the Beer-Lambert Law [51] implies the higher the UAV, the more chance of atmospheric absorption, but conversely the lower it is, the longer the surveys of mangrove forests will take.

Moreover, the results in this investigation have been found using normalised reflectance data, which, unless the UAV has a built in ambient light sensor, means that reduced illumination levels at different times and the presence of clouds could cause measured reflectance values to be lower than expected. This could cause misclassifications meaning the results achieved in practice differ from those predicted in this investigation. Using the common method of modelling solar illumination with a sinewave [52] an uncertainty value for the results of this investigation was able to be quantified if the measurements are made within a certain time of midday (the time of maximum solar illumination). Assuming the sun rises at 6 a.m. and sets at 6 p.m., the variation of solar light in a day, I, was calculated using,

$$I = \sin\left(\frac{2\pi}{24}(x-6)\right) \quad , \tag{4}$$

where x was the time of day (in military time). Using this, the times at which the light level is within 5% of the maximum illumination level was able to be calculated.

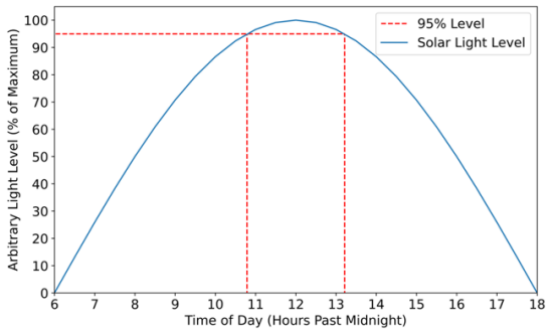

Figure 15: Ambient light level at each hour between an assumed 6 a.m. sunrise and 6 p.m. sunset. The 95% level is shown in a red dotted line to show the times when the ambient light level is within 5% of the maximum.

As seen in Figure 15, (neglecting the atmospheric absorption effects) provided the

measurements are taken between 10:48 and 13:12, the ambient light level, and so the intensity of the light incident on the leaves, is within 5% of the maximum value. Having a maximum ambient light level is equivalent to the calibrated values used in this investigation so the effect of a 5% drop in the ambient light level on the predicted efficiencies and purities can be calculated. If one assumes that measurements were taken at the exact point when the ambient light was 95% of the maximum, the reflectance values of each sample can be reduced by 5% to calculate the new efficiencies and purities. Subtracting these from the original efficiencies and purities and halving then gives uncertainties on the mean of the two efficiency/purity values.

Table 5: Final results table for the three approaches to discriminating between the different species of mangroves developed using artificial neural networks. Uncertainties have been calculated for each predicted efficiency and purity value to give an insight into the effect of a 5% drop in the ambient light level.

	Best Possible	**Optical Only**	**Compromise**
Red Criteria	$R_{1986.5 \pm 5\text{ nm}} < 0.019$	$R_{483 \pm 1.5\text{ nm}} < 0.041$	$R_{483 \pm 1.5\text{ nm}} < 0.041$
Red Efficiency	78.3 ± 2.3%	86.3 ± 0.4%	86.3 ± 0.4%
Red Purity	83.5 ± 1.3%	62.2 ± 4.0%	62.2 ± 4.0%
White Criteria	$R_{1986.5 \pm 5\text{ nm}} > 0.033$	$R_{402 \pm 1.5\text{ nm}} < 0.045$ $R_{678 \pm 1.5\text{ nm}} > 0.043$	$R_{483 \pm 1.5\text{ nm}} > 0.041$ $R_{1686.5 \pm 5\text{ nm}} < 0.31$
White Efficiency	80.3 ± 4.6%	61.4 ± 0.1%	60.3 ± 1.6%
White Purity	68.7 ± 0.4%	47.1 ± 0.2%	55.7 ± 4.8%
Black Criteria	$R_{1986.5 \pm 5\text{ nm}} > 0.019$ $R_{1986.5 \pm 5\text{ nm}} < 0.033$	$R_{504 \pm 1.5\text{ nm}} > 0.060$	$R_{1686.5 \pm 5\text{ nm}} > 0.31$
Black Efficiency	65.1 ± 0.3%	58.9 ± 4.1%	71.0 ± 8.4%
Black Purity	70.1 ± 0.5%	60.6 ± 0.1%	64.6 ± 0.9%

Tables 5 and 6 show the final results tables, which assumes the ambient light level remains within 95% of its maximum value. The results in most (but not all) cases change by less than 5% which is reasonable. Though this simple model says that these values can be achieved provided measurements are taken within one hour and twelve minutes of midday, in practice, absorption from the atmosphere or the presence of clouds will also reduce the ambient light level [51], further restricting the time-window for which the level is within 5% of the maximum.

One potential solution to this issue is to have an ambient light sensor on the UAV measure it before measuring each sample. The spectral measurements could then be calibrated and normalised to the ambient light level at the time of their measurement, accounting for changes in the brightness level, which would loosen the strict time conditions. Therefore, whilst using an UAV in conjunction with ambient light sensors, like reference Spectralon panels, is very expensive [53], the investment could cause large improvements and aid preservation attempts.

Table 6: Final results table for the two approaches to discriminating between the different health statuses of Red and Black mangroves developed using artificial neural networks. Uncertainties have been calculated for each predicted efficiency and purity value to give an insight into the effect of a 5% drop in the ambient light level.

		Best Possible	Optical Only
Red	**Healthy Criteria**	-	$R_{699 \pm 1.5\,\text{nm}} < 0.155$
	Healthy Efficiency	-	77.8 ± 2.8%
	Healthy Purity	-	96.7 ± 2.0%
	Stressed Criteria	-	$R_{699 \pm 1.5\,\text{nm}} > 0.155$
	Stressed Efficiency	-	89.0 ± 7.0%
	Stressed Purity	-	50.1 ± 1.1%
Black	**Healthy Black**	$R_{2493 \pm 7.5\,\text{nm}} > 0.023$	$R_{522 \pm 1.5\,\text{nm}} < 0.090$
	Healthy Efficiency	60.6 ± 3.8%	42.5 ± 6.4%
	Healthy Purity	61.9 ± 0.5%	67.0 ± 3.7%
	Stressed Black	$R_{2493 \pm 7.5\,\text{nm}} < 0.023$	$R_{522 \pm 1.5\,\text{nm}} > 0.090$
	Stressed Efficiency	55.3 ± 2.0%	71.8 ± 10.3%
	Stressed Purity	54.0 ± 1.6%	51.2 ± 0.4%

6. Conclusion

To conclude, a complete mangrove maintenance model which can discriminate between the different species of mangroves, as well as determining the health status of Red and Black mangroves has been successfully developed using ANNs to find the discriminating wavelengths. In both use cases, reasonable efficiencies and purities have been achieved (with the highest on average being found in the Red mangroves for species discrimination and the Red healthy mangroves for health status determination), with an error being quantified on these results to account for a 5% drop in the ambient light level during measurements. Overall, the results are good enough to be considered as an improvement to the manual methods of maintaining mangroves, offering an improved workflow for those undertaking this task as surveys of mangrove forests could be complete in much less time at the compromise of a drop in accuracy when compared to the manual method. The potential misclassifications from the results would cause some issues when using this method on global scales, however, in all cases the majority of the classifications are correct meaning that at the very least they can be used for quick scans of the forests. The value of this compromise is further emphasised when considering the ease at which repeat surveys can be conducted using the UAVs, as even in the optical case where the total maintenance attempt would need six spectral cameras (four to determine the species and two for its health status), the survey would be much faster than travelling tree to tree with a manual spectrograph. This is important as it would allow those

maintaining the mangroves to identify if there are decreases in a particular species and give them the chance to take the required action to preserve and regrow mangroves where necessary.

Moreover, the model developed in this investigation is accessible regardless of budget with three approaches differing in price being made in the species classification case and two in the health classification case. Many of the most discriminating wavelengths were in the NIR region but due to the high price of NIR spectral cameras [39], the most discriminating optical wavelengths were also found. This was an important step as if the maintenance of mangroves across the globe is to improve, it is important that those in the small coastal villages can contribute as that is where the mangroves' carbon sequestration abilities significantly impact national carbon budgets [8]. A 'compromise' method was also developed to bridge the gap between the NIR and the optical approaches and be a stepping stone between the two as technology develops and the cost of NIR cameras drops.

In practice, to obtain the best accuracy when using this model, regular measurement of the ambient light level during a survey is required to allow the calibration and normalisation of the measured reflectance values. This is because as the drone flies, it is not necessarily under constant illumination and the reflected light it is measuring is not subject to the same scattering and absorption effects each time. A change in the ambient light could mean that the measured reflectance is wrong and cause samples to be misclassified. Therefore, the results accounting for a 5% drop in the ambient light level from the maximum allows confidence that the results achieved in this investigation can be achieved in practice.

When comparing the findings of this investigation to existing literature, it was found that one of the wavelength bins in this investigation, 522 ± 1.5 nm, was in agreement with a discriminating wavelength found in Zhang et al's investigation in separating mangrove species and conditions [27]. A further two wavelengths from this investigation, 650 nm and 710 nm, were close to some found from the investigation in this paper, 678 ± 1.5 nm and 699 ± 1.5 nm, suggesting some spectral properties may be similar, however, the wavelengths themselves are not in agreement. Furthermore, similar to the investigation described in this paper, Zhang et al found that the Red healthy mangroves had the highest classification accuracy, but this award was shared with the White mangroves which yielded much worse accuracies as seen in Table 5, perhaps suggesting that a larger sample size is required for future investigation.

The use of ML, ANNs specifically, was superior to the manual statistical approach, producing improved results in a shorter time frame. Looking towards the future, if one had more time and was not computationally limited, advanced, automatic hyperparameter optimisation methods would be used to fine tune the ANNs [54]. This would allow for more combinations of hyperparameters to be tested, leading to large improvements in the results and the speed at which the investigation can be conducted, whilst also making the results reproducible since an algorithm will be used rather than a trial-and-error process. Time saved here could then be used to further extend the project, allowing for other ML algorithms to be researched and tested.

References

[1] S. Miththapala, *Mangroves*, Coastal Ecosystems Series, Vol 2, IUCN, 2008, p. 1
[2] I. Nagelkerken et al, *The habitat function of mangroves for terrestrial and marine fauna: A review*, Aquatic Botany, Vol 89, Issue 2, August 2008, p. 155-185, https://doi.org/10.1016/j.aquabot.2007.12.007
[3] D. M. Alongi, *Mangrove forests: Resilience, protection from tsunamis, and responses to global climate change*, Estuarine, Coastal and Shelf Science 76 (2008), p. 1-13, https://doi.org/10.1016/j.ecss.2007.08.024
[4] K. Kathiresan, N. Rajendran, *Coastal mangrove forests mitigated tsunami*, Estuarine, Coastal and Shelf Science 65 (2005), p.601-606, https://doi.org/10.1016/j.ecss.2005.06.022
[5] M. A. Othman, *Value of mangroves in coastal protection*, Hydrobiologia (1994), Vol 285, p. 277-282
[6] F. J. Manson et al, *An Evaluation of the Evidence for Linkages Between Mangroves and Fisheries: A Synthesis of the Literature and Identification of Research Directions*, Oceanography and Marine Biology 43, p.483–513
[7] Z. Adeel, R. Pomeroy, *Assessment and management of mangrove ecosystems in developing countries*, Trees (2002), Vol 16, p.235-238
[8] P. Taillardat et al, *Mangrove blue carbon strategies for climate change mitigation are most effective at the national scale*, Biology Letters (2018), Vol 14, https://doi.org/10.1098/rsbl.2018.0251
[9] J. P. Casas-Ruiz et al, *Integrating terrestrial and aquatic ecosystems to constrain estimates of land-atmosphere carbon exchange*, Nature Communications, Vol 14, 1571 (2023)
[10] Adoption of the Paris Agreement FCCC/CP/2015/10/Add.1 (UNFCCC, 2015)
[11] K. Kathiresan, B. L. Bingham, *Biology of Mangroves and Mangrove Ecosystems*, Advances In Marine Biology (2001), Vol 40, p.81-251
[12] M. Spalding et al, *World Mangrove Atlas*, International Society of Mangrove Ecosystems 1, January 1997
[13] M. Spalding et al, *World Atlas of Mangroves* (2010), v3.1, A collaborative project of ITTO, ISME, FAO, UNEP-WCMC, UNESCO-MAB, UNU-INWEH and TNC. London (UK): Earthscan, London, p. 319
[14] FAO of the United Nations 2020, *Global Forest Resources Assessment 2020: Main report*, Rome
[15] S. R. Dayal, *Deforestation and Degradation in the Mangrove Ecosystem: Implication on Environment and Livelihoods*, Land Degradation Neutrality: Achieving SDG 15 by Forest Management, Springer, 2022, p.99-116
[16] M. Bologna, G. Aquino, *Deforestation and world population sustainability: a quantitative analysis*, Sci Rep 10, 7631 (2020), https://doi.org/10.1038/s41598-020-63657-6
[17] J. Sanderman et al, *A global map of mangrove forest soil carbon at 30m spatial resolution*, Environ Res Lett. 13, 2018, 055002
[18] Z. Liu et al, *Monitoring global carbon emissions in 2021*, Nat Rev Earth Environ (2022), Vol 3, p.217-219
[19] FAO of the United Nations, Suriname, May 2019, https://www.fao.org/fishery/en/facp/sur?lang=en
[20] R. P. Nijbroek, *Mangroves, mudbanks and seawalls: whose environmental knowledge counts when adapting to sea level rise in Suriname?*, Journal of Political Ecology (2014), Vol 21(1), p.533-550
[21] S. Dasgupta et al, *The Impact of Sea Level Rise on Developing Countries: A Comparative Analysis*, World Bank Policy Research Working Paper 4136, February 2007
[22] R. Costanza et al, *The value of coastal wetlands for hurricane protection*, Ambio, Vol 37 (4), June 2008
[23] E. J. Anthony, *Assessment of peri-urban coastal protection options in Paramaribo-Wanica, Suriname*, 2015
[24] W. Veenendaal, L. Demarest, *How population size affects power-sharing: a comparison of Nigeria and Suriname*, Contemporary Politics (2021), Vol 27 (3), p. 271-291
[25] K. Ewe let al, *Different kinds of mangrove forests provide different goods and services*, Global Ecology & Biogeography Letters, Vol 7 (1), Jan 1998, p. 3-94, 10.2307/2997700
[26] S. C. Snedaker, *Mangroves and climate change in the Florida and Caribbean region: scenarios and hypotheses*, Hydrobiologia 295 (2004), p.43-49
[27] Zhang et al, *Separating Mangrove Species and Conditions Using Laboratory Hyperspectral Data: A Case Study of a Degraded Mangrove Forest of the Mexican Pacific*, Remote Sens. 2014, 6(12), 10.3390/rs61211673
[28] A. C. C. Viodor et al, *Identifying Mangrove Species using Deep Learning Model and Recording for Diversity Analysis: A Mobile Approach*, p. 1-6, 10.1109/eStream59056.2023.10134992
[29] S. Song et al, *Mangrove reforestation provides greater blue carbon benefit than afforestation for mitigating*

global climate change, Nature Communications,Vol 14, 756 (2023), https://doi.org/10.1038/s41467-023-36477-1
[30] S. C. Snedaker et al, *Recovery of a Mixed-Species Mangrove Forest in South Florida Following Canopy Removal*, Journal of Coastal Research 8(4), 1992, p. 919-925, ISSN 0749-0208
[31] H. H. Muhammed, *Using hyperspectral reflectance data for discrimination between healthy and diseased plants, and determination of damage-level in diseased plants*, Applied Imagery Pattern Recognition Workshop, Proceedings., Washington, DC, USA, 2002, p. 49-54, https://doi.org/10.1109/AIPR.2002.1182254
[32] L. Wang, W. P. Sousa, *Distinguishing mangrove species with laboratory measurements of hyperspectral leaf reflectance*, International Journal of Remote Sensing (2009), 30:5, p.1267-1281
[33] J. M. Carlton, *Land-building and Stabilization by Mangroves*, Environmental Conservation, 1974, Vol 1(4), p.285-294, doi:10.1017/S0376892900004926
[34] M. Danner et al, *Spectral Sampling with the ASD FieldSpec 4 – Theory, Measurement, Problems, Interpretation*, EnMAP Field Guides Technical Report, GFZ Data Services, 2015
[35] K.C. Lawrence et al, *Calibration of a Pushbroom Hyperspectral Imaging System for Agricultural Inspection*, Transactions of the ASAE (2003), Vol 46(2), p.513–521
[36] F. E. Fassnacht et al, *Review of studies on tree species classification from remotely sensed data*, Remote Sensing of Environment (2016), Vol 186, p.64-87
[37] I.G. Hughes and T.P.A. Hase, *Measurements and Their Uncertainties*, Oxford University Press:Oxford (2010)
[38] T. K. Kim, *T test as a parametric statistic*, Korean Journal of Anesthesiology (2015), Vol 68 (6), p.540-54
[39] W. Nijland et al, *Monitoring plant condition and phenology using infrared sensitive consumer grade digital cameras*, Agricultural and Forest Meteorology 184, 2014, p.98-106, doi.org/10.1016/j.agrformet.2013.09.007
[40] Edmund Optics, www.edmundoptics.co.uk/f/1460-1600nm-near-infrared-camera/12779
[41] L. Yang, A. Shami, *On hyperparameter optimization of machine learning algorithms: Theory and practice*, Neurocomputing, Vol 415, 2020, p.295-316, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.07.061
[42] C. Zhang et al, *Relationship between Hyperspectral Measurements and Mangrove Leaf Nitrogen Concentrations*, Remote Sensing, Vol 5(2), p.891-908, February 2013, doi:10.3390/rs5020891
[43] A. D. Dongare et al, *Introduction to Artificial Neural Network*, International Journal of Engineering and Innovative Technology (IJEIT), Vol 2(1), July 2012, ISSN: 2277-3754
[44] P. Ramachandran et al, *Searching for activation functions*, 2017, https://doi.org/10.48550/arXiv.1710.05941
[45] S. B. Maind, P. Wankar, *Research Paper on Basic of Artificial Neural Network*, International Journal on Recent and Innovation Trends in Computing and Communication, Vol 2(1), p.96–100, ISSN: 2321-8169
[46] C. E. Nwankpa, *Advances in Optimisation Algorithms and Techniques for Deep Learning*, Adv. Sci. Technol. Eng. Syst. J., Vol 5(5), p.563-577, 2020, doi:10.25046/aj050570
[47] T. Takase et al, *Effective neural network training with adaptive learning rate based on training loss*, Neural Networks, Vol 101, 2018, p. 68-78, ISSN 0893-6080, https://doi.org/10.1016/j.neunet.2018.01.016
[48] A. Ghorbanian et al, *Application of Artificial Neural Networks for Mangrove Mapping Using Multi-Temporal and Multi-Source Remote Sensing Imagery*, Water 2022, Vol 14 (2), 244, https://doi.org/10.3390/w14020244
[49] T. Vujičić et al, *Comparative analysis of methods for determining number of hidden neurons in artificial neural network*, Central European Conference on Information and Intelligent Systems, Vol 159, 2016
[50] Nezami et al, *Tree Species Classification of Drone Hyperspectral and RGB Imagery with Deep Learning Convolutional Neural Networks*, Remote Sensing, Vol 12(7):1070, 2020, https://doi.org/10.3390/rs12071070
[51] W. Mäntele, E. Deniz, *UV–VIS absorption spectroscopy: Lambert-Beer reloaded*, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Vol 173, 2017, p.965-968, doi.org/10.1016/j.saa.2016.09.037
[52] S. Wang et al, *New calculation methods of diurnal distribution of solar radiation and its interception by canopy over complex terrain*, Ecological Modelling 155(3),2002,p.191-204,doi:10.1016/S0304-3800(02)00122-9
[53] J. P. Arroyo-Mora et al, *Implementation of a UAV–Hyperspectral Pushbroom Imager for Ecological Monitoring*, Drones 2019, Vol 3, no. 1:12
[54] B. Bischl et al, *Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges*, WIREs Data Mining and Knowledge Discovery, Vol 13(2), e1484, 2023, https://doi.org/10.1002/widm.1484

Summary for a General Audience

Mangroves are a shrub-like plant found along coastlines. They're one of the most important ecosystems on the planet, enriching marine life around the world. Their strong root systems allow them to reduce impacts from powerful tidal waves, making them pivotal for coastlines under threat from the rising sea level. Their ability to capture and store atmospheric carbon dioxide means they can help the world tackle global warming. As a result, it is important that mangrove forests are maintained however, current methods of doing so are time-consuming meaning that deforestation may threaten the population of mangrove forests.

Different species of mangroves have different properties and require different methods to maintain them. The investigation in this paper found that by measuring the amount of particular types of light that are reflected off mangrove leaves, both the species and whether it was healthy or not could be determined to a reasonable accuracy. This means that mangrove surveys could be done remotely by flying drones fitted with cameras over the forests to measure the amount of the specific types of reflected light, improving the maintenance process. Potential problems with this method from the changing brightness of the Sun are considered to complete the investigation.

Appendix A

Further plots showing the p-values and variances at each wavelength for each species of mangrove. This can act as an aid to Figure 5, displaying the same information but in a different way.

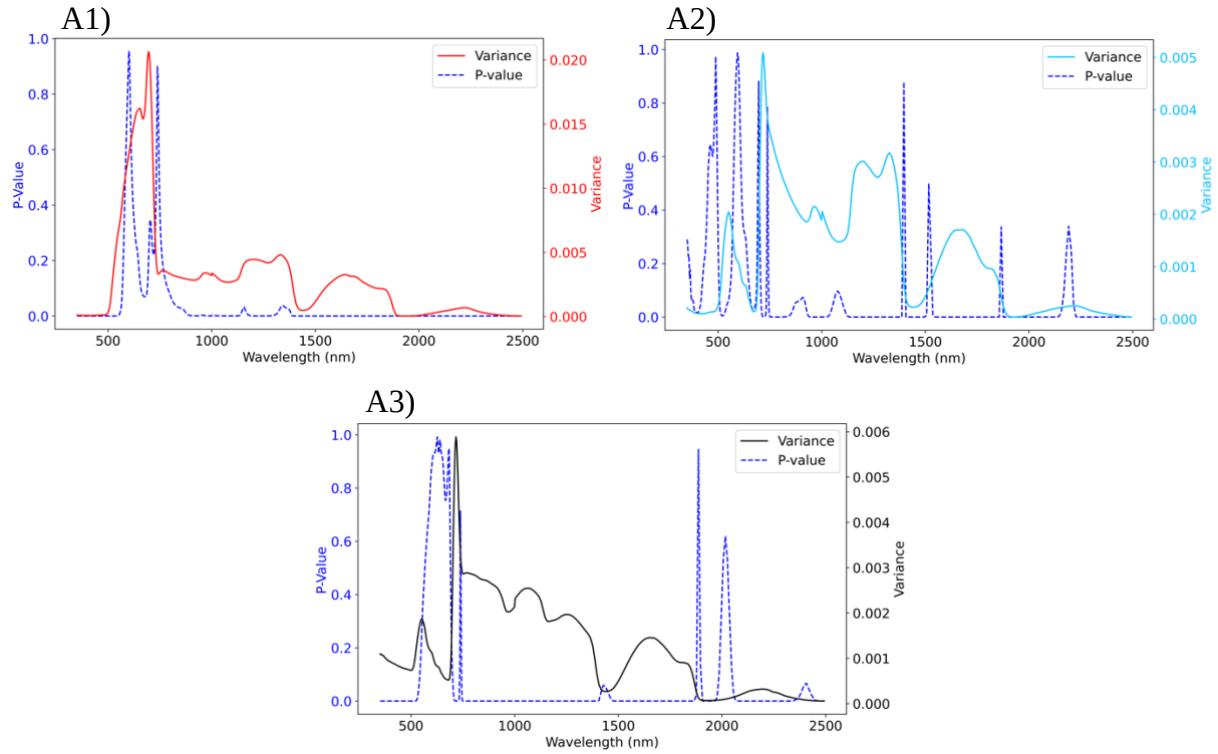

Figure A: P-values (left axes) and variance (right axes) plotted for each of the 372 wavelength bins for the sample of mangroves. P-values below 0.05 indicate that there is statistical significance in there being a difference between the species in question and the other two at that specific wavelength. Lower variance values indicate that the reflectance values within that species are closer together, suggesting that there may be a reflectance value at that wavelength which is characteristic of that species.

1) *P-values and variance for the sample of 250 Red mangroves.*
2) *P-values and variance for the sample of 218 White mangroves.*
3) *P-values and variance for the sample of 330 Black mangroves.*

Appendix B

Table B shows the lack of any clear pattern within the metadata for the group of Red mangroves, suggesting that the group of 17 Red mangroves that stray from the pattern of the other mangroves may be caused my random errors during the measurement taking process.

Table B: Metadata for the group of 17 and the remainder of the Red mangroves. No common theme is shown by the group of outliers and there are no notable differences when comparing their metadata to that for the rest of the Red mangroves. This suggests that the reason for their anomalous behave could be due to random errors in the measurement taking process.

	Locations	Health	Age
Rejected	1 x Paramaribo 3 x Weg nar see 13 x Estuary	4 x Healthy 13 x Stressed	3 x Young 14 x Mature
Accepted	24 x Paramaribo 42 x Weg nar see 67 x Estuary 100 x Pristine Site	196 x Healthy 37 x Stressed	102 x Young 131 x Mature

Appendix C

Video tour of Figure 9b. The planes showing the species boundaries can be seen cutting through the distribution of the three species in the 'best possible' approach using the 483 ± 1.5 nm, 1656.5 ± 5 nm and 1976.5 ± 5 nm wavelength bins. The criteria defining these planes can be found in Table 2.

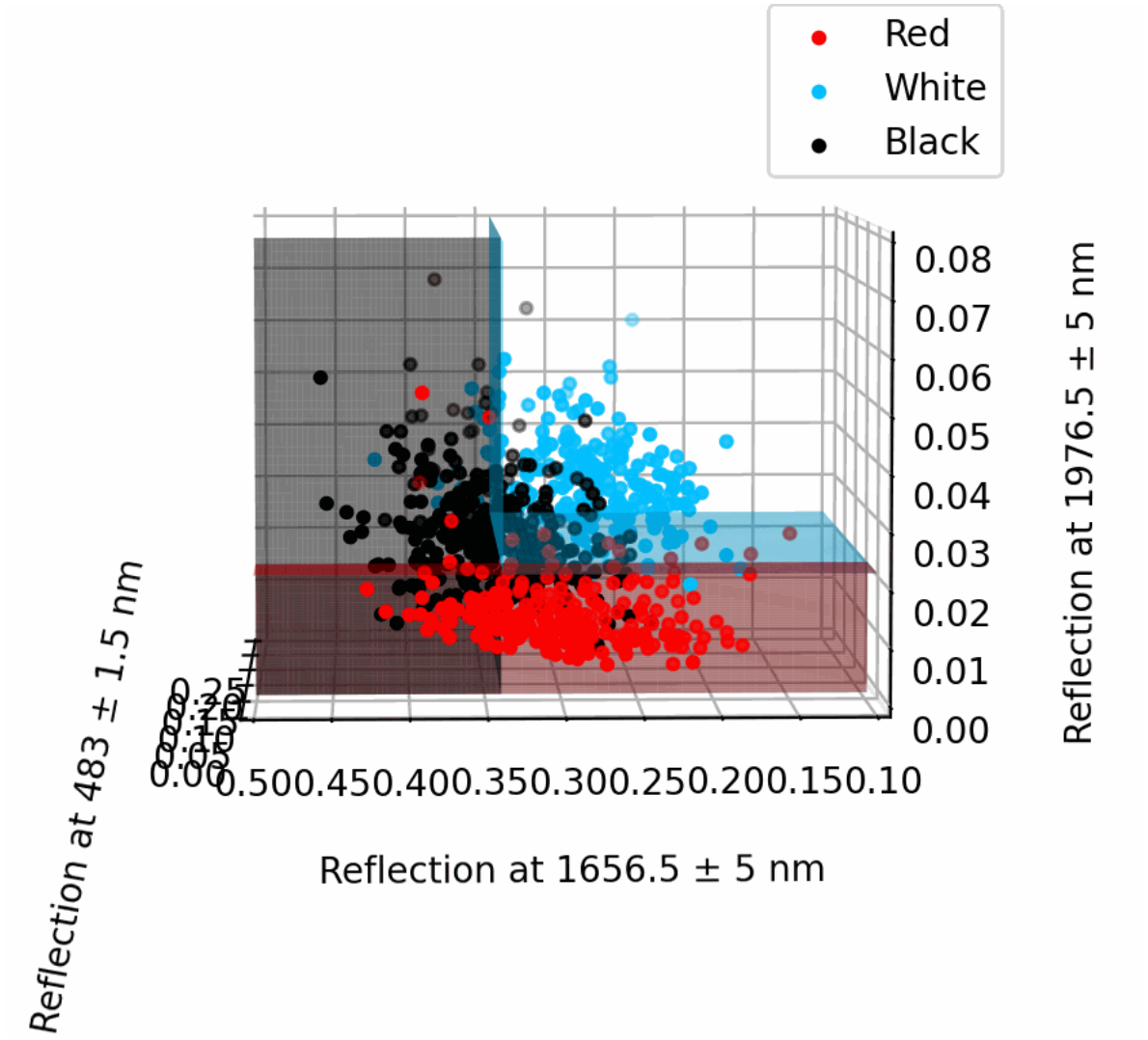

Figure C: Screenshot of the video showing the distributions and species boundaries for the 'best possible' approach using the determined 483 ± 1.5 nm, 1656.5 ± 5 nm and 1976.5 ± 5 nm wavelength bins as via a statistical approach. The equations for the boundaries can be found in Table 2. Much of the distribution is clustered around the boundary intersection meaning slight shifts in measurements could alter the accuracy results obtained. The link to the video tour of this plot can be found below.

URL:
https://youtu.be/cl525SBrMpg?si=sOl21wbPvORH5f2e