Using Medium Resolution Satellite Imagery With Machine Learning To Monitor The Expansion And Collapse Of Biomes Over Time

Calum McMeekin

Master of Informatics

School of Informatics
University of Edinburgh
2022

Abstract

This paper proposes a first ever attempt to separate the three largest biomes in South America; Amazon, Cerrado and Caatinga using one machine learning model. A dataset of 50,000 medium resolution Landsat 7 images gathered from the sub-equatorial summer (due to the better separation of the biomes given the seasonal properties of the Cerrado biome) were used for both establishing a baseline and the application of state-of-the-art (SOTA) algorithms. The baseline was created using the Normalized Difference Vegetation Index (NDVI) as the only feature, with the best classifier (a Random Forest classifier) achieving a macro-f1 score of 68.4%. Five state-of-the-art models were then applied, two of which were the Tile2Vec algorithm (one trained from scratch on the dataset, another applied with the pre-trained network) that builds upon the spatial similarity hypothesis used by more famous models such as Word2Vec. The remaining models were a ResNet18, AlexNet and SwAV. The ResNet18 accompanied with a multinomial logistic regression classifier achieved the highest result with a macro f1-score of 84.3%. Exploratory analysis on the clusters created by the best model revealed that certain eco-systems such as rivers were being associated with specific biomes (such as the Amazon) when they are not necessarily an eco-system unique to that biome. However, SwAV's inability to improve over ResNet18 suggests that there may be too many overlapping or too little independent clusters to use them as a point of separation. To improve on the accuracy of the best performing model it is recommended that future work introduces additional features such as the shortwave infrared channel (SWIR 2) that Landsat 8 has, or exploiting the spatiotemporal properties of the data as well as analysing the performance of the model on temporal data.

Acknowledgements

The following work done in this paper could not have been achieved without the continued support from my supervisor Seth Sohan. His constant enthusiasm gave me drive to continue pushing myself beyond what I thought was capable and has taught me lessons that will last a lifetime. I would also like to thank my parents for their love and encouragement from my first day of school, and my friends, for not shaming me when I ask silly questions.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Calum McMeekin*)

Table of Contents

Chapter 1

Introduction

1.1 Motivation

Brazil alone has 530 million hectares of native vegetation, 349 of which belongs to the Amazon, 92 to the savannas of the Cerrado and 89 to Caatinga and other smaller biomes [1]. Manually mapping each of these biomes is a long and lengthy task that would be impossible to frequently repeat, given current methods (such as those by IBGE) which update once every 15 years [2]. During this time period the shape of each biome undergoes drastic changes, whether it is a part of the accelerating deforestation that is ongoing in the Amazon [3] or it is through the reforestation of the Cerrado [4], the maps quickly become out-of-date and, furthermore, they do not accurately represent the area of each biome.

Since 1990 the rate at which deforestation occurs in Brazil is only getting worse with every passing year [3]. Fighting this illegal and natural deforestation is a challenge due to the sheer scale at which it occurs and the massive land area it occurs in. This makes it hard for conservationists to pinpoint exactly where the deforestation is ongoing amongst the 530 million hectares of native vegetation Brazil has [1]. Current remote sensing techniques are limited in their coverage and rely on law enforcement and education to prevent deforestation in those areas that are not monitored [1].

Using machine learning in combination with Landsat 7 satellite images to identify where the largest biomes in South America are and are not, would allow for the area each biome covers to be updated as quickly as new images can be obtained to provide up-to-date maps on the coverage each of these biomes currently has. On top of this, experts will be able to identify changes in these maps over time, providing a first of its kind way to easily monitor the entirety of South America for deforestation as anomalies

appear in the areas previously classed as a certain biome.

1.2 Summary of Previous Work

In 2004 the Brazilian Institute of Geography and Statistics (IBGE) released a map of the different biomes in South America and their boundaries. The map was very accurate for the time of release with a scale of 1:5,000,000 (the ratio of distance on the map to the corresponding distance on the ground) but it was not until 2019 that an updated map was released, a full 15 years later. This map boasted an impressive improvement on the scale by making it 20 times more accurate (1:250,000) [2]. Both maps were curated through extensive manual field sampling but the more accurate boundaries in the 2019 map were achieved through improvements in bibliographic revision and inter-institutional contacts. These methods are evidently very time consuming and require a lot of resources in order to map the 851 million hectares of Brazil. If it was possible to accurately map each of the biomes using satellite imagery then these maps could be updated at a rate equal to the time taken for the satellite to make a full pass over all the areas being monitored - roughly 16 days with Landsat 7.

Recent groundbreaking developments in deep learning have led to the ability to accurately classify wildfires [5], deforestation [6] and the different eco-systems that make up the biomes in South America [7] by only using true colour images. Despite this promising result showing a distinction between the biomes, it is yet to be applied to Brazil in its entirety.

1.3 Objectives

A cheap model that updates regularly would be the ideal solution to the classification of the biomes in South America. This would ideally replace the need for manual field-sampling in areas where machine learning can say - with confidence - that it belongs to a particular biome. By detecting where one biome is and is not it would be possible to then identify destructive changes to the biome as they would change the appearance it takes, this would then focus the attention of anyone wishing to monitor the biome to these areas of interest as opposed to the time consuming task of manually scanning satellite images of the entire biome.

Therefore, this work hypothesises that by using the true colour difference alone, an advanced machine learning algorithm will be able to form distinct representations

for each biome from medium resolution Landsat 7 imagery. This task can then be separated into the following stages:

- Determine what preprocessing will be applied to each satellite image such that it represents the desired features, i.e. removing clouds, detecting if a certain season has better contrast between the biomes or if normalisation is required.
- Create a baseline from NDVI data as it has previously shown success in separating the Cerrado and Caatinga biomes in South America [8].
- Apply state-of-the-art algorithms to the RGB Landsat 7 dataset by using transfer learning as well as training a Tile2Vec model on the data itself. It is hypothesised the Tile2Vec model that is trained on the dataset of 40,000 images will give the best accuracy.
- Evaluate the results of the best performing model and analyse how the number of false positives can be reduced through means such as introducing a probability threshold.
- Inspect the possibility of contrastive cluster learning to exploit the numerous eco-systems that make up each biome.
- Finally, the embedded images will be analysed to see how finer categories of images are represented. This will allow categories such as farmland, rivers, mountains and other defining eco-systems to be associated with a specific biome, giving rise to the ability for an unsupervised method to be created based on the results of the semi-supervised method.

Chapter 2

Background

2.1 Biome

A biome is a geographical region containing diverse combinations of fauna and flora. The difference between biomes depends on their temperature, humidity and how fertile the soil is. The main biomes in South America (where this thesis is focused) are the Amazon, Caatinga, Cerrado, Atlantic Forest, Pampas and the Pantanal. Each of these biomes are made up of unique combinations of smaller ecoregions. A set of 30 (1km x 1km) images for each biome can be found in Appendix A.

For the classification of each biome, exact co-ordinates of that biome must be located. Currently no database contains the location of every biome on the planet and where its boundaries are exactly. However, smaller databases do contain such data. For example the MapBiomas [9] project contains labelled data for the types of land use (including information about the biome type) all layered on top of Landsat 8 imagery. This data is open source and freely available making it perfect for training a model.

There is also the PREDICTS database [10] which contains 1,783 latitude and longitude co-ordinates for South America where at each site it provides a brief description of the biodiversity including the biome type. Satellite imagery can then be obtained for each site - using Landsat 8 - where the biome within the image can be labelled.

As the appearance of a biome will vary depending on the time of year all images are taken from within the same season.

2.2 Convolutional-Neural-Network

A convolutional-neural-network (CNN) is a sub-class of neural network that has been designed to work well for image classification. Its combination of *convolutional* and *pooling layers* reduces the number of fully connected layers in the model, decreasing the chances that training data will be overfitted. A CNN will change the weights and biases such that they represent the aspect and objects in the image that are most important. Famous examples of the CNN architecture are AlexNet by Alex Krizhevsky [11] and VGGNet by Karen Simonyan and Andrew Zisserman [12].

CNN's are very effective at reducing the input image into a more manageable form without losing features that are essential for an accurate prediction. To achieve this the *convolutional layer* will apply a kernel to its input. This will both reduce its dimensions as well as picking out features such as edges and colours that will be useful for classifying. This output will then be passed on to the *pooling layer* [13][14].

The *pooling layer* reduces the size of the already convolved feature but this time it will focus on extracting features that are rotational and positional invariant. This additional reduction in size greatly increases the processing speed as the size of the input data is now a fraction of what it was to begin with. The pooling layer may also compute what is called *max pooling* or *average pooling*. *Max pooling* is when the maximum value from each cluster of the input is used to represent it while *average pooling* takes the average value of the cluster.

The complexity of the input image will determine the number of *convolutional* and *pooling layers*, with more layers extracting more features from the input image but at the cost of processing speed.

The final layer will then take the image that has been outputted by the *convolutional* and *pooling layers* and flatten it into vector form. This vector is fed into a fully connected layer - much like a multi-layer perceptron - in order to classify the image [13].

2.3 Normalized Difference Vegetation Index (NDVI)

NDVI is a technique used to determine the density of green on a per pixel basis. It is calculated by normalising the near-infrared (NIR) light with respect to the visible red (RED) light (Equation 2.1) [15]. Healthy vegetation will absorb more visible red light as part of a stronger photosynthesis cycle whilst reflecting more near-infrared light

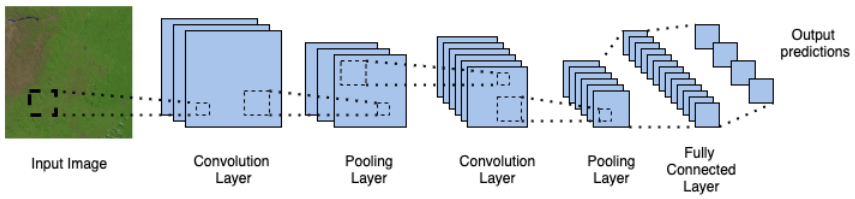


Figure 2.1: Convolution Neural Network with two convolution layers, two pooling layers and one fully connected layer

[15]. Meanwhile unhealthy vegetation will reflect more visible light and absorbs more near-infrared light [15]. These two contrasting reactions to visible light are what make NDVI an effective tool for analysing how much vegetation a patch of land contains, as well as how healthy this vegetation is. An NDVI result for each pixel is always between -1 and 1. A result close to -1 would represent water, a result close to 0 would represent no vegetation and a result close to 1 would represent strong vegetation.

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{2.1}$$

2.4 Previous Work

Similar projects that segment biomes using satellite imagery have been done before but often this is done using a supervised data set, a vast training set or as a binary biome classification. A summary of such projects follows:

- MapBiomas (section 2.4.1) is a project that runs an annual classification of the different land cover and land uses for Brazil using 30m resolution Landsat imagery. The output, although accurate and freely available has taken a team of just under 200 people to collect, process and output. This has required lots of funding as well as partnerships with large companies making it arduous to repeat.

- Similar to MapBiomas, the Brazilian Institute of Geography and Statistics (IBGE) [16] have released maps of the different biomes within brazil. However the maps are released very infrequently with only two editions having been released, one in 2004 and the other in 2019. The scale of these maps also varies greatly; 1:5,000,000 and 1:250,000 respectively.

- Several studies have been conducted to see if deep learning techniques could be applied to segment and classify different subject areas in an attempt to reduce the amount of human intervention required (see section 2.4.2). Although the results were very accurate it required a vast number of training samples which often are not available.

- PREDICTS database and other studies (section 2.4.3) use sampling to help globally analyse biomes and biodiversity. These studies provide invaluable data for mapping the biomes of the world, however, it faces the common issue that progress is slow, expensive and requires a lot of human intervention.

- NDVI has been successfully used to separate different biomes for a long time (section 2.4.4), proving to be a reliable method for finding unique characteristics between different geological areas. NDVI has been shown to not be the best performing technique for this application, with newer techniques outperforming it considerably. However, these newer methods make use of bands that are not available with Landsat 7 and therefore not suitable for this study.

2.4.1 MapBiomas and IBGE

Both IBGE [16] and the MapBiomas [9] project form their maps by collaborating with many different organisations and professions. In fact MapBiomas obtained their map by using the IBGE 2004's map of the biome boundaries. These large partnerships that are required to generate the detailed maps results in a procedure that is very costly to run and not scalable for a worldwide application. However, the results are all open source and can be freely downloaded, providing invaluable labelled data which can be used to train a more general machine learning model for the entirety of South America.

2.4.2 Deep Learning Techniques

Several approaches aimed to reduce the large amounts of human intervention that are required to identify a subject from a satellite image by applying various deep learning techniques. Such attempts have proven that it is possible to detect forests [17], deforestation [6], wildfires [5] and even the different eco-systems that make up the biomes of South America [7] with an overall accuracy that is above 90%. The majority of the time these techniques rely on having large amounts of high quality training data. As data of this standard is often hard to come by, it is preferable to use techniques that rely

less on having substantial amounts of it, for example the method attempted by Jean et al. [18].

2.4.3 PREDICTS Database and Sampling Techniques

Possibly the most extensive database of biome classification is the PREDICTS [10] database which contains measurements (taken through sampling) for 13 out of the 14 biomes over 208 ecoregions.

Similar methods of sampling have lead to very accurate studies into land cover changes over time [19] or a combination of sampling along side manual analysis of satellite imagery has helped map the world's biomes [20]. Although these methods all produce accurate results, the data is slow to process and expensive to obtain, making it infeasible to accurately classify entire countries or continents.

2.4.4 NDVI

In Northern Brazil NDVI has been used by Batista et al. to detect seasonality between the Cerrado and Caatinga biomes in 1997 [21]. This study was also able to detect land changes between the two biomes indicating that NDVI is a valid technique for monitoring a time series of events. In 2017, Erasmi et al. again showed NDVI to be an effective tool for separating the Caatinga and Cerrado biomes [8] by commenting on the contrasting amounts of vegetation present in each biome. More recent techniques have shown that NDVI is not the best solution for the separation of these biomes. Bueno, Inacio T., et al. [22] found that for object-based change detection in the Cerrado biome NDVI was only half as good as the leading technique, using shortwave infrared light - a band that is now available with Landsat 8.

Chapter 3

Data Collection

3.1 Terminology

The following satellite imagery terms are used throughout the rest of this paper and so have been summarised here for easier reading.

- Bands: Also known as channels or layers, each satellite image is made up of several bands where each band measures the intensity of a different part of the electromagnetic spectrum. By combining different bands you obtain different representations of the image, for example, combining the bands one (blue), two (green), and three (red) of a Landsat 7 satellite image will produce a visible light image.

- Spatial Resolution: The amount of geographical area a single pixel in the image represents. Low resolution images have a spatial resolution of 60m/pixel, medium resolution images have a resolution of 10-30m/pixel and high to very high resolution images have a spatial resolution of 30cm-5m/pixel [23].

- Scan Line Corrector (SLC): An instrument that is used on board satellites to prevent the forward motion of the satellite from capturing some areas twice while missing other areas completely.

- Landsat: An ongoing project by NASA/U.S. Geological Survey to constantly capture images of the Earth's land surfaces. Since 1972 several Landsat satellites have been launched into space, with each one providing more information than the last.

- Top of Atmosphere (TOA) Reflectance: The ratio of radiation reflected by a surface to the incident solar radiation of the same surface. In doing so it accounts for the spectral band differences that give different values for solar irradiance.

3.2 Landsat 7 Image Summary

Landsat 7 has been running since 1999 and takes 16 days to make a full orbit. This resulted in an abundance of time stamped images, representing the change in the planet over the following 18 years till it was retired. Each Landsat 7 image is made up of 8 bands; blue, green, red, near-infrared, thermal, mid-wave infrared and panchromatic. Except for thermal and panchromatic each band has a spatial resolution of 30 metres. The size of each tile (image) needs to be large enough such that the machine learning algorithm is still able to extract useful information during downstream processing but small enough that neighbour tiles are still semantically similar [18]. For this paper, satellite images of size 50×50 pixels were used. An example 50×50 pixel tile for each biome can be seen in Figure 3.1. This figure demonstrates the subtle visual differences between the Caatinga and Cerrado biomes with the Amazon biome appearing to be a haze of pure vegetation - as is to be expected for a dense rainforest.

The images were taken during the summer of 2002 (specifically from early October to late March as the area of interest is in the southern hemisphere). This season was chosen as it was seen to give the best contrast between each of the biomes, as is discussed later on in section 4.2.

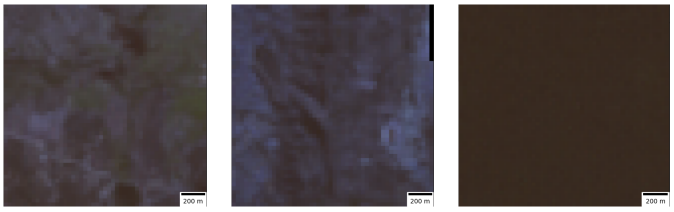


Figure 3.1: 1.5km x 1.5km tiles for each biome. Left: Caatinga, center: Cerrado, right: Amazon

3.3 Landsat 7 Image Preprocessing

As can be seen in Figure 3.1 and Figure 4.2, the standard Landsat 7 RGB images are very dark. This is not consistent across all images, with some images being brighter

than others. To try and prevent this from being an influential feature on later classification, each Landsat 7 image was initially normalised. This was achieved by taking the max and min values for each band, for each image, then scaling the range of the original band - that was between [min, max] - to be between [0, 1], by following Equation 3.1. In the code this is accomplished by making use of the Earth Engine library function *ee.Image.unitScale* to produce the results that can be seen in Figure 4.2.

$$Output\_band = \frac{(Input\_band - min)}{(max - min)} \tag{3.1}$$

However, this was not beneficial as each image was now scaled between different values and the desired standardised image set was not achieved. An alternative idea would be to normalise on the image set as a whole, but this would not change anything as some images contained null pixels (that have 0 value) and others clouds (that are white and have value 255), leading to the max and min values being set to the limits (0 and 255) and leaving all images unaffected. In light of this it was decided that Landsat 7 images would be calibrated for tier 1 TOA reflectance instead. Planetary Top of Atmosphere reflectance (p_λ) can be computed for each image by combining the mathematical constant π with the spectral radiance at the sensor's aperture (L_λ), the mean solar spectral irradiance ($ESUN_\lambda$) and the solar zenith angle (θ_λ) in the way Equation 3.2 demonstrates.

$$p_\lambda = \frac{\pi * L_\lambda}{ESUN_\lambda * \theta_\lambda} \tag{3.2}$$

To prevent cloud cover from influencing the classification of any models a mask was applied to the clouds in each image, if this mask covered more than 10% of the image then the image was removed. As can be seen in 4.1, after removing images that had clouds in them the Amazon biome only represented 23% of the entire dataset while Cerrado represented 46%. This was to be expected as the Amazon is made mostly of tropical forest that is notorious for having large amounts of persistent cloud cover [24][25].

It was also found that certain images were corrupt (i.e. they did not give an accurate representation of the land they were photographing) and so these had to be removed. The first attempt to filter corrupt images was to inspect their histograms, unfortunately this did not prove possible as the difference between the histograms of corrupt images and satisfactory images was not distinct enough (as can be seen in Appendix A.1 and so any method of separating them would likely result in lots of satisfactory images incorrectly being removed. Instead the images for each quadrant were first embedded into

a ResNet18 model and then k-means was applied to the two principle components that expressed the majority of the variance in the biome. After manually inspecting the 5 images closest to the center of each cluster all images that belonged to the corrupt cluster were removed from the dataset. Although this method proved to be time consuming it minimised the amount of satisfactory images that were unnecessarily removed from the dataset.

There were also certain images that had been corrupted in other ways due to external issues and so a final layer of preprocessing was required. This consisted of two steps:

1. The size of the image would be reset to be equal to the standard $x \times y \times z$ format of $50 \times 50 \times 3$ that was most common across the images. Every image would have a Z dimension of 3 but often the number of x and y dimensions would be greater than or less than the desired shape ($50 \times 50 \times 3$). To achieve this either an x or y row/column would be dropped or added. New rows/columns were copies of the neighbouring row/column in the image to prevent it from changing the representation of the image too much.

2. Finally, any NaN values that were present in the image would be replaced by the mean value of the image. These NaN values possibly occurred from previous preprocessing steps that tried to perform a mathematical operation that was not possible.

3.4 Annotated Sources

The MapBiomas project (as described in section 2.4.1) is used as the source of ground truth data. The project was created to establish the different land covers and land uses of Brazil on an annual basis, and, as a by-product creates boundaries for the Cerrado, Caatinga and Amazon biomes. MapBiomas was able to construct the boundaries for each biome by using the official Map of Biomes that is produced by the Brazilian IBGE (that has a scale of 1:5,000,000) [16] and adapting it using the vegetation map of Brazil (1:250,000) to create a 1:1,000,000 scale map containing the limits of each biome [9]. For this project the boundaries created by MapBiomas are extracted and overlapping sections removed. An example of these boundaries can be found in Figure 4.1. As these maps were created using data gathered in 2004 the images representing each biome would ideally have been taken from 2004 to keep these boundaries as accurate

as possible, however, due to the scan line corrector on Landsat 7 failing after May 31st, 2003, [26] it is no longer possible to extract images without gaps in the image. Attempts were made to try and fill the gaps, however this proved only effective for the central band of each image, leaving the gaps remaining in the extremities of each image. As the boundaries of each biome are unlikely to change drastically over the course of 2 years, images were taken from the summer of 2001/2002 as this is the closest summer where the SLC was still working.

By using these geometries as ground truth data it is possible to extract a tile from within these geometries and say with confidence that it is a member of that geometry. The only time this is not the case is where the boundaries of these biomes overlap, therefore to avoid this instance a tile is never taken from an overlapping area. The Amazon, Cerrado and Caatinga biomes were selected as the only classes to be used given they are the three largest biomes and therefore have the greatest quantity of training data available. In light of this gaps can be seen in between each biome in Figure 4.1 due to either another biome being present or the area consisting of overlapping biomes so has been removed completely to avoid confusion.

Chapter 4

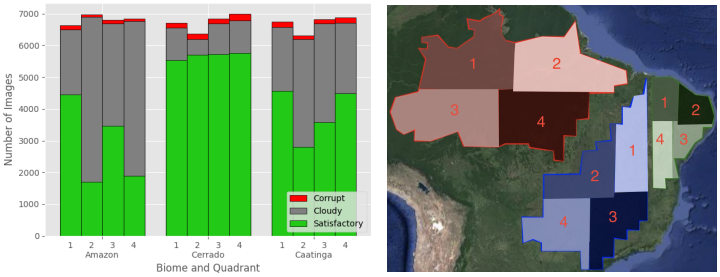
Methodology

4.1 Train/Validation Split

In order to prevent the training data from overlapping with the testing data the training and testing datasets are split both geographically and by quantity. Each biome is split into quadrants that each represent roughly 25% of the total area of the biome (Figure 4.1). Training points are then randomly selected from 3 of the quadrants while the testing points are randomly selected from the remaining quadrant. Removing geographical regions from the training area and using them for testing gives an impression of the predictive capability of the model on unseen areas. This method also prevents the training and testing data coming from the same areas and running the risk of data leakage, giving an artificially high performance.

The randomness in image extraction was achieved by using the Google Earth Engine function *ee.FeatureCollection.randomPoints*. All the images in each quadrant were extracted sequentially with the random seed set to the number of that image, for example the 10th image that was extracted in a quadrant would have been extracted with a random seed of 10. This allowed the images to be extracted using multithreading (a necessary feature given the lengthy extraction time) and still keep the dataset creation repeatable.

The ratio of training set size to testing set size remains consistent at 75/25 for both the geographical area covered and the number of data points used. This means that the training dataset covers an area three times larger than that of the testing dataset, allowing the trained model to gain an accurate representation of the data whilst leaving enough separate unseen data to measure its performance.



(a) (b)

Figure 4.1: (a) Number of images each biome has per quadrant (b) each polygon represents the area of each class/biome. Red background is Amazon, blue background is Cerrado and the green background is Caatinga (with the label of each quadrant overlaid and labelled in red).

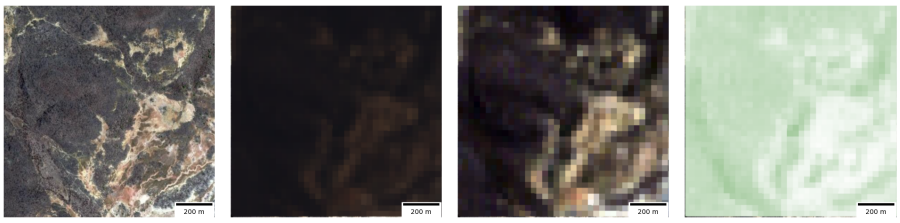
4.2 Baseline

In order to evaluate if using a vector representation of the data - such as Tile2Vec - improves the accuracy of the biome classification, a baseline must first be established. The baseline that is being used is the Normalized Difference Vegetation Index (NDVI). This was chosen for the baseline given its success in previous studies at separating the biomes - as was discussed in section 2.4.4. By applying NDVI to each tile, a representation of how much vegetation each pixel represents can be obtained. As the Amazon is known to have more vegetation than the Cerrado biome, and the Cerrado biome more than Caatinga - it is hoped that NDVI would be able to exploit this and find a good separation between the biomes.

To calculate the NDVI value of each tile the near-infrared (NIR) and red (RED) bands are normalised, as shown in Equation 2.1. To give a visual representation of what this represented, a colour palette was applied to each tile after each pixel's NDVI score was calculated for that tile. The palette is a spectrum from white (no vegetation) to green (vegetation) where the intensity of each pixel represents the amount of vegetation present. This can be seen in Figure 4.2.

A multinomial logistic regression, decision tree and Random Forest were trained on the training split and then their performance evaluated on the test split. Further details on these classifiers can be found later in section 4.7.

The macro-average was used for precision, recall and f1-score as it prevents classes that appear more often from having a bias over the classification performance simply because it occurs more frequently. The macro of each metric is achieved by individu-



(a) (b) (c) (d)

Figure 4.2: Comparison of the different types of preprocessing applied to a tile. (a) ultra-high resolution image from Google Earth for reference, (b) RGB Landsat 7 TOA image, (c) Landsat 7 TOA image after it has been normalised, (d) Landsat 7 TOA image after NDVI and a colour palette have been applied.

ally computing the evaluation metric (precision, recall or f1-score) for each class and then averaging the result. This was necessary given the class imbalance that became present when images containing clouds were removed (Figure 4.1). The precision recall and f1-score are evaluation metric based upon the number of true positives, false positives, true negatives and false negatives that the classifier has made.

4.3 Tile2Vec

Tile2Vec is an unsupervised machine learning algorithm which was developed to allow large scale machine learning tasks to be applied to satellite imagery or images taken from a bird's eye view. This came as a response to the majority of recent machine learning success coming from the use of supervised algorithms that had access to large sets of labelled data. [18]

In order to achieve this, Tile2Vec takes inspiration from similar unsupervised natural language processing algorithms such as Word2Vec. It assumes that features in images that are close to each other will be similar and share semantics while features that are further away in the image will be dissimilar and thus have different semantics. Also, like Word2Vec, Tile2Vec is built on top of a convolutional-neural-network (CNN). This is a sub-class of neural network that focuses on identifying local dependencies.

4.3.1 Triplets

To select features close to each other the entire image is divided into tiles. The size of the tiles can be changed depending on the resolution of the image. These tiles are then divided into groups of three which is what the CNN is trained on.

The triplet consists of an anchor tile, a neighbour tile and a distant tile. The anchor and neighbour tiles are near to each other in the image (with the area between them being called the neighbourhood) while the distant tile is further away. The distant tile is chosen such that its distance from the anchor is equal to that of the neighbours plus a predefined margin. This prevents the distant tile being so far away from the anchor that it loses any significance [18].

Furthermore, the likelihood that a tile shares a neighbourhood with another distant tile is negligible, allowing us to pair an anchor and neighbour tile with any distant tile. This drastically increases the number of unique triplets the network trains on from $O(N)$ to $O(N^2)$.

Jean et al., show that the results of Tile2Vec when compared to other unsupervised algorithms in combination with a classifier was consistently the most accurate algorithm for both 1,000 and 10,000 labelled images[18, 1].

4.3.2 Trained Model

To create the training set, the same training and test quadrants that were chosen when creating the baseline were chosen again. Two Tile2Vec algorithms were used, both having a slightly different architecture and also being trained on different datasets.

The change made to the architecture of the first Tile2Vec algorithm was related to the triplet sampling technique used. As was said in greater detail in the previous section, the original Tile2Vec triplets are created by taking two tiles that are close to each other creating an anchor and neighbour tile that will be considered a part of the same class and then a 3rd tile that is considered a distant tile. The purpose behind this is that tiles which occur close to each other will share similar semantics and belong to the same class while tiles that are further from the anchor tile will have different semantics and belong to a different class. In the original Tile2Vec algorithm these triplets are formed from the same satellite image, as the classes they were using were small and each image often showcased a variety of classes. That is not the case here. As the size of each class in this application is 3,388,067km^2 for the Amazon, 1,811,916km^2 for the Cerrado and 727,906km^2 for the Caatinga, there are very few images where

one tile - that only covers 78km^2 - will contain more than one class. To address this issue and give the Tile2Vec algorithm training on this dataset a better chance at finding distinctions between the classes, the scale of the triplet sampling was changed. Instead of taking an anchor and neighbour tile from the same image they were taken from the same quadrant while a distant tile would be taken from any quadrant from a different biome. It was also the case that each neighbour and distant tile could only belong to one anchor tile in order to prevent the same tile occurring in multiple triplets and having a greater influence over other tiles. This new approach to the triplet sampling changes the unsupervised Tile2Vec algorithm into a semi-supervised model as it is using the knowledge that certain images belong to one class while other images do not. For this reason it is expected that the trained semi-supervised Tile2Vec model will find a strong separation between the classes.

The Adam learning rule was used as it is well known to quickly converge to a local minimum, a quality that was favoured in this paper given the lack of resources to train for more than a hundred epochs [27]. The Adam Learning rule computes the running averages for both the first and second moments of gradients and then controls their decay through two parameters (β_1 and β_2) [27]. In light of this, the Adam learning rule converges faster to a local minimum compared to other methods such as Stochastic Gradient Descent, Root Mean Square Propagation (RMSProp) and Adaptive Gradient Algorithm (AdaGrad) [27]. Furthermore, it was used by the original Tile2Vec paper during training of their model giving further confidence behind choosing it as the optimizer [18].

4.3.3 Pre-trained Model

As well as the Tile2Vec model that was trained on the dataset, a pre-trained Tile2Vec model was used. The model was pre-trained on the NAIP dataset, a dataset consiting of 300k tiles (or 100k triplets) where each tile has a resolution of 50×50 pixels (or 30-by-30m) and consists of 66 classes, each a different type of land cover. This difference in resolution means that every image in the NAIP dataset covers an area equivalent to one pixel of an image from the dataset presented in this paper. The trained Tile2Vec model was also trained on images with 4 channels, in order to make it compatible with the dataset used in this paper where each image has three channels, a fourth channel of 0's was added to each image before being embedded by the pre-trained Tile2Vec model.

4.4 ResNet18

The third SOTA algorithm used was a ResNet18 [28]. A ResNet is a deep CNN that was created to address the vanishing gradient problem many deep neural networks experience. The vanishing gradient occurs in deep CNN architectures where the lower layers fail to get the attention they need during training to allow them to adapt to the inputs. This is a result of the training technique backpropogation, where the error is passed back from the output, through each layer, to the input with each layer adjusting its weights to account for this error. However, deep networks tend to find that as the error moves back through the network it moves closer and closer to zero, until eventually it no longer has any influence on the lower layers resulting in their weights not updating to match the error of the network and so the network as a whole will fail to become any more accurate [29]. The problem which has been coined the term "vanishing gradient" - as the gradient appears to seemingly disappear as it moves backwards through the network - was seemingly solved with the release of ResNet though the use of an additional "shortcut connection" between layers [28]. This shortcut connection is used to add the input of a layer to its output and as a result aims to negate any negative effect a stacked layer can have on the size of the weights. The progress made by He et. al. [28] allows deeper networks to perform better than shallower networks as the increased number of layers are able to extract more information from the input data without the vanishing gradient problem having as much of an effect. Despite the ResNet implementation the vanishing gradient problem is still an issue for very deep networks and so for this paper the ResNet18, giving the benefits of a deep network without the drawback of vanishing gradients.

The ResNet18 used in this paper has (as the name suggests) 18 layers, one 7x7 convolutional layer, 16 3x3 convolutional and one fully connected linear layer. The model has been pre-trained on the ImageNet database ([30]) and scored a top-5 accuracy of 10.92 %[31]. This is a database designed for visual object recognition software and consists of more than 1.2 million labelled natural images across 1,000 categories [30].

As the pre-trained Tile2Vec model is not a supervised method like ResNet18 and AlexNet it does not have a Softmax layer. As a result the output from the ResNet18 and AlexNet model had to be taken from the average pool layer, the layer before the Softmax function is applied. The softmax function takes the embedded vector from the previous layer and converts it to a vector of length equal to the number of classes, where each value in this vector is the probability that the input vector belongs to that

class. The z-dimension of these outputs is 512 for both the Tile2Vec models and the ResNet model.

4.5 AlexNet

The final SOTA algorithm that is applied is AlexNet. AlexNet was one of the first deep CNN networks to have success, owing that to its ability to use multiple GPU's during training [11]. The model also made use of the *dropout* technique. Dropout is used to help increase the generalisation (how well it performs on unseen data) of a model by randomly removing certain hidden and input units during training [32]. By using dropout a different subset of the network will be trained during each epoch resulting in the final network being a combination of an exponential number of smaller networks [32]. This helps the network cope with some input features missing as each hidden unit relies less on other hidden units. The previously mentioned ResNet18 model does not make use of dropout during training and so AlexNet has been included in this paper to see if the use of dropout makes it perform better on unseen data.

The architecture of AlexNet that is used in this paper has 8 layers, 5 convolutional and 3 fully-connected. The network has also been pre-trained on the same ImageNet dataset that ResNet18 was trained on, and achieved a top-5 error of 20.91% [31]. Much like ResNet18 the output of AlexNet was taken from the average pool layer to avoid the Softmax function. The main difference between AlexNet and the other SOTA algorithms is that the dimension of its outputs is 4096. This makes AlexNet's output vector shape eight times greater than that of the other models.

4.6 SwAV

Swapping Assignments Between Views (SwAV) is a new self-supervised machine learning approach developed as part of Facebook Research by Caron et al. [33]. SwAV does not make use of auxiliary momentum networks like other contrastive methods such as SimCLRv2 by Google's Brain Team [34]. Instead, it learns by taking an input image, applying random transformations to it to gain multiple representations of the input and then swapping them such that each transformation has a different representation vector attached to it [33]. Clusters representations for the dataset are then learned while keeping the cluster representations for each image consistent across these larger

clusters [33]. This approach is able to then get better performance on smaller datasets due to the transformations on each image giving more data for the network to train on.

The SwAV model used in this paper was pre-trained on the ImageNet database and is built on top of a ResNet50 architecture with an output projection size of 1000. This method has been included in this paper in the hope that it would find representations for the smaller clusters that make up each individual biome (i.e. it would be able to associate mountains with the Cerrado biome and semi-arid eco-systems with the Caatinga biome).

4.7 Classifiers

4.7.1 Decision Tree

In a decision tree features are represented by nodes with branches between each feature representing the decisions that are made by the classifier. These decisions can be represented as 'if-else' statements given there is always at least 2 options the decision tree will split based on the current nodes conditon [35]. The leafs of the tree will then represent the class to which the decisions made up to that point belong [35]. The deeper and wider a decision tree gets the more it can separate the data allowing for very complex decision boundaries to be created between the classes. The one drawback to this is that it then becomes very likely that the tree will overfit to the training data and so preventative measures are used during training to avoid this such as *stopping* - one example of which is where a tree is no longer trained once a node reaches a certain depth [35]. Tree pruning is another method used when stopping is not enough to prevent the model from overfitting to the training data. There are two types of pruning, pre and post. *Pre-pruning* takes place during training and could consist of a node not being added to the tree if it does not in some way improve the decision boundary in a significant way. *Post-pruning*, on the other hand, will first have a model trained to a considerable depth and then remove branches that do not significantly improve the accuracy of the decision tree [35].

The depth of each decision tree in this paper was determined by first splitting the testing dataset into a testing and validation set (using an 80:20 split). Many trees were then trained on the training set, all with a different depth from 1 to 40 and the model which performed the best on the validation set would be used on the testing set.

4.7.2 Random Forest

A Random Forest is a collection of decision trees that are each trained on a random subset of the full training dataset, that is the same size as the full dataset but may have some vectors repeated (a technique called *bagging*) [36]. When the Random Forest receives a new input vector each decision tree will cast a vote on what class it believes the new input vector belongs to, the class with the majority of votes will be the prediction made by the Random Forest classifier [36]. In a Random Forest each tree is not pruned, instead this allows for many complex decision boundaries that individually have a high generalisation error but thanks to the voting system that Random Forests use, the decision boundary is balanced out to give a representation of the best decision boundary [36].

The random element of the Random Forest comes from each tree being trained on a randomly selected subset of the data, therefore, to ensure that each experiment could be repeated all Random Forests were trained with a random seed of 0 and the optimal max depth determined using the same technique as the decision tree. The optimal number of estimators (number of decision trees that are in each Random Forest) was determined by training multiple Random Forests with increasing numbers of estimators and choosing the number that performed the best on the validation data.

4.7.3 Multinomial Logistic Regression

Multinomial logistic regression (MLR) is a version of logistic regression that is used to predict more than two classes. Logistic regression, although similar to linear regression it does not rely on the data being linearly separable. Logistic regression achieves this by using a sigmoid curve to separate the data, where if a new input vector lies in the upper half of the sigmoid curve it belongs to one class, while if it lies on the lower half of the sigmoid it belongs to the other class [37]. To turn logistic regression into a multinomial problem a different weight vector is created for each class and then fed into a Softmax function which will return the probability of the input vector belonging to each of the possible output classes [37]. The class with the highest probability is chosen as the prediction.

The MLR model in this paper used L2 regularisation (also referred to as ridge regression) which prevents the model from overfitting to the data while minimising the error that occurs. This is achieved by making the weights less sparse and penalising weights that become too large [38].

Chapter 5

Experiments

5.1 Establishing Baseline Performance

5.1.1 Aim

To establish whether seasons have an impact on the distinction between the Amazon, Cerrado and Caatinga biomes and then using this knowledge to establish a baseline using NDVI.

5.1.2 Seasonal Influence

Before pursuing NDVI any further it was important to see if it would at all be possible for NDVI to separate each of the biomes and if the season that data is extracted from has an impact on the representation of each biome.

This was achieved by sampling 1000 random points for each biome from the training split (discussed in section 4.1) for both summer and winter and then plotting a histogram of each biome to analyse the distribution of NDVI scores. Figure 5.1 shows that during summer there was a clear distinction between each of the three biomes while in winter the Cerrado and Caatinga biomes are overlapping. This reinforces the fact that the Cerrado biome is seasonal and so it would be sensible to only extract data during summer to get a better representation of the data. The number of components to represent each biome was selected by plotting the log-likelihood for each model using 1 to 20 components and then taking the elbow (point where the log likelihood begins to diminish as the number of components increases), which turned out to be 2 components for each biome.

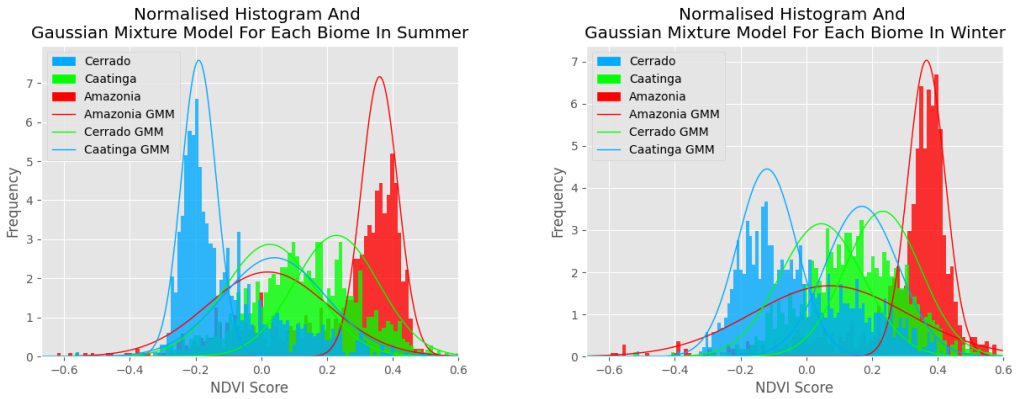


Figure 5.1: NDVI histograms for each biome for summer (left) and winter (right). A Gaussian Mixture Model (GMM) has also been overlaid for each biome with 2 components.

5.1.3 Training and Classifying

A training dataset of 40,691 images, a validation set of 2,300 images and a test set of 9,200 images were used in this experiment with the validation and test set images both coming from the same quadrant. Three models were trained on the training set, a Random Forest, a multinomial logistic regression model and a decision tree.

In order to compute the optimal depth of the decision tree, multiple depths were used from 1 to 40 and the depth with the best accuracy on the validation set was used (see appendix B.1). The best depth turned out to be 11. The multinomial logistic regression was used with the *saga* solver and the Random Forest was initialised with a random seed of 0, depth of 8, 16 estimators and ran for 10 epochs where the results were averaged and standard deviation computed. All three models were obtained from the scikit learn library [39].

5.2 Trained and Transfer Learning SOTA Models

5.2.1 Aim

Applying SOTA algorithms to RGB images of the Amazon, Cerrado and Caatinga will create a distinct representation of the three biomes that is better than that achieved by NDVI.

5.2.2 Experiments with Transfer Learning

To obtain what class each embedded vector belonged to, a classifier was trained on the embedded vectors of the training set and then its performance was recorded on the embedded vectors of the test set. The inclusion of the external classifier allowed for more experimenting to be made on what classifier maximised the results of the network. The same classifiers used in the baseline were applied to the embedded vector representation of the training set, multinomial logistic regression, decision tree and a Random Forest. Like in section 5.1.3 the depth of the decision tree was determined by separating the training dataset into a training and validation dataset (an 80:20 split) and then plotting the accuracy of the model on the validation set against the depth of the tree and taking the global maximum to be the depth (Figure B.1). In order to help the multinomial logistic regression model to converge, the sklearn standard scaler was applied to the embedded representation of the vectors [39]. This subtracted the mean vector embedding from each embedded vector and divided by the standard deviation of the dataset.

For the trained Tile2Vec model the Adam learning rule was used with a learning rate of 0.001, a beta 1 coefficient of 0.5 and a beta 2 coefficient of 0.999. These hyperparameters were chosen as they were used by the original Tile2Vec paper and were found to optimise the model [18].

Chapter 6

Results

6.1 Baseline

From Figure 5.1 we see that the biomes are indeed seasonal. We also see from the Guassian Mixture Model (GMM) that each biome has 1 component for the main cluster of NDVI samples relevant to that biome as well as another seemingly shared component across all three biomes at around 0 - such as clouds (as is seen in Figure 7.2).

By only using data gathered in the summer months and training two different classifiers with these images we see that the Random Forest had the best performance, with a macro f1-score of 0.684, 0.003 more than the multinomial logistic regression model. The more complex decision boundary of the Random Forest allows it to create a better separation between the overlapping points compared to what the multinomial logistic regression model is capable of. This complex decision boundary is reflected in the macro precision and macro recall of the Random Forest and the MLR model. The macro precision for the Random Forest is 0.014 greater than that of the MLR model, indicating that of the predicted values more of them are relevant, yet the higher recall of MLR shows it is able to correctly predict a greater number of the images. From Figure B.1 it makes sense that as the depth of the decision tree and Random Forest increases the accuracy will decrease as the models begin to overfit to the training data. From Figure 5.1 it is obvious the biomes have significant overlap and so increasing the complexity of the decision boundary during training time will create a decision boundary that is extremely overfitted to the overlapping points in the training data. By keeping the tree on the shallower end it prevents the idiosyncrasies of the training data from having an effect on future predictions with unseen data.

	Multinomial Logistic Regression	Decision Tree	Random Forest
Accuracy	0.714	0.707	0.714
Macro-Precision	0.684	0.691	0.698
Macro-Recall	0.677	0.665	0.670
F1-Score	0.681	0.678	0.684

Table 6.1: Performance of multinomial logistic regression model, decision tree with a depth of 11 and a Random Forest on the NDVI test set

6.2 SOTA Algorithms

The final performance of each classifier was measured by how it performed on the unseen test set, which in this case was the same quadrant as was used in the baseline experiments. To achieve this, the test set vectors were embedded to each model and then the classifier would predict what class they belong to after having trained on the full training set. As there was a class imbalance present - due to certain biomes having more corrupt and/or cloudy images dropped during preprocessing - the macro evaluation metrics were used. The number of images used for each class and in each quadrant were labelled as 'satisfactory' and can be found in Figure 4.1.

		Acc.	Macro Precision	Macro Recall	Macro F1-Score	Depth	Est.
RF	Tile2Vec	0.798	0.823	0.780	0.788	23	200
	Tile2Vec †	0.824	0.859	0.802	0.815	17	200
	ResNet18	0.836	0.872	0.800	0.823	20	32
	AlexNet	0.8340	0.8696	0.8073	0.8241	28	100
	SwAV	0.764	0.812	0.714	0.737	29	64
DT	Tile2Vec	0.773	0.793	0.753	0.758	8	-
	Tile2Vec †	0.806	0.833	0.783	0.796	4	-
	ResNet18	0.810	0.821	0.781	0.796	7	-
	AlexNet	0.819	0.844	0.797	0.808	6	-
	SwAV	0.726	0.732	0.675	0.690	7	-
MLR	Tile2Vec	0.801	0.825	0.781	0.789	-	-
	Tile2Vec †	0.826	0.846	0.812	0.817	-	-
	ResNet18	**0.851**	**0.873**	**0.834**	**0.843**	-	-
	AlexNet	0.840	0.865	0.826	0.832	-	-
	SwAV	0.815	0.838	0.787	0.800	-	-

Table 6.2: Performance of SOTA algorithms with different classifiers on the RGB test set; where Tile2Vec † is the pre-trained Tile2Vec model, Est. is the number of estimators the Random Forest had and Acc. is the accuracy of the model. The model that achieved the best performance for each individual evaluation metric is highlighted in bold

Chapter 7

Exploratory Analysis

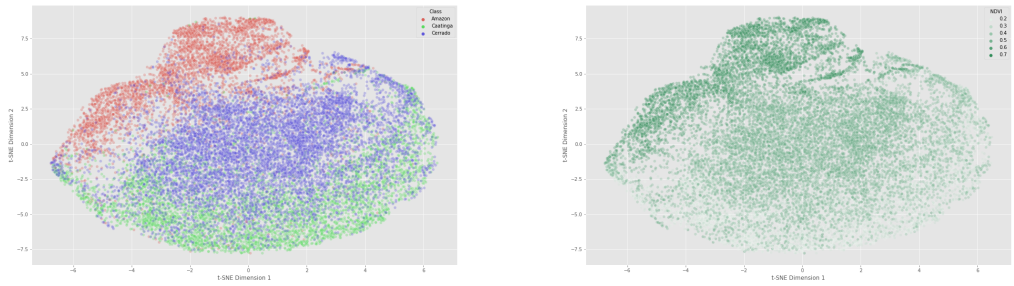
7.1 Viewing Representations

Now that it is understood that the SOTA algorithms all outperform the baseline a better idea for what characteristic the SOTA algorithms are using to separate the data was investigated. This was first achieved by plotting a t-SNE graph of the embedded vectors of the best performing model (ResNet-18, table 6.2). One plot had the class of each vector and the other had the NDVI score applied as a hue (Figure 7.1). From these two plots we see that NDVI scales from the highest values in the top left (where the Amazon biome is) to the bottom right (where the Caatinga biome is). Looking back to the NDVI values demonstrated in Figure 5.1 we see that this t-SNE plot does look as we would expect, identifying NDVI as a characteristic used by the ResNet18 model to seperate the three clusters.

However, the SOTA algorithms all outperformed the baseline and so they must be using other characteristics to help distinguish one biome from another. To get an idea of what these could possibly be, k-means was applied to the vector representation of the data and the 5 images closest to each center were viewed. The value of K was determined by plotting the sum of squared distances for each cluster of multiple k-means models that each had a different value of K from 1 to 14. This was repeated on the training set of quadrants for each biome. The results are found in Figures 7.2 and D.1, where each row represents a separate cluster.

From this we are able to tell that certain clusters are shared across each biome, for example, cluster 1 for the Amazon, 0 for the Cerrado and 4 for Caatinga all represent images that have clouds in them. We are also able to see that the Amazon is made of clusters that show varying degrees of extensive vegetation, except for cluster 4 which

represents bodies of water. This cluster has likely come from the large amounts of water that the Amazon drainage basin stores [40]. Despite large bodies of water being a distinguishing characteristic of the Amazon it is not the case that if an image contains water then it must belong to the Amazon and therefore this cluster should instead be given its own class alongside the current Amazon, Cerrado and Caatinga classes. Unlike the Amazon, we see that the Cerrado class has been made up of fields (cluster 4), mountainous regions (clusters 1 and 2) and semi-dense forests (cluster 3). These are all expressive features of the Cerrado biome which is made up of an Atlantic forest in the east (cluster 3), dense grassland (cluster 3) as well as sparse shurbs and trees that are typical of a savanna ecoregion (clusters 1 and 2) [41]. Finally from the clustering we are able to determine that the defining characteristics of the Caatinga biome, namely its hot and dry climate that make up the semi-arid ecosystems, have been captured in clusters such as 0, 1, 2 and 3 that each show a varying degree of vegetation [19].



(a) (b)

Figure 7.1: Two t-SNE dimensions for the ResNet18 RGB training set. Image (a) shows the different classes associated with each embedded vector and RGB image (b) shows the NDVI score of each RGB image

7.1.1 Reducing False Positives

In order to understand why false positives were occurring the predictions made by the best model (ResNet18) were evaluated. This was achieved by plotting the locations where images had been correctly classified as the current biome (true positives) as well as plotting the locations where the image had been incorrectly classified as one of the two other possible biomes. The results of this can be found for each of the three test quadrants in Figure 7.3. The slight elliptical pattern that can be seen in image (a) in Figure 7.3 is due to the Earth engines random algorithm, which after roughly 2000

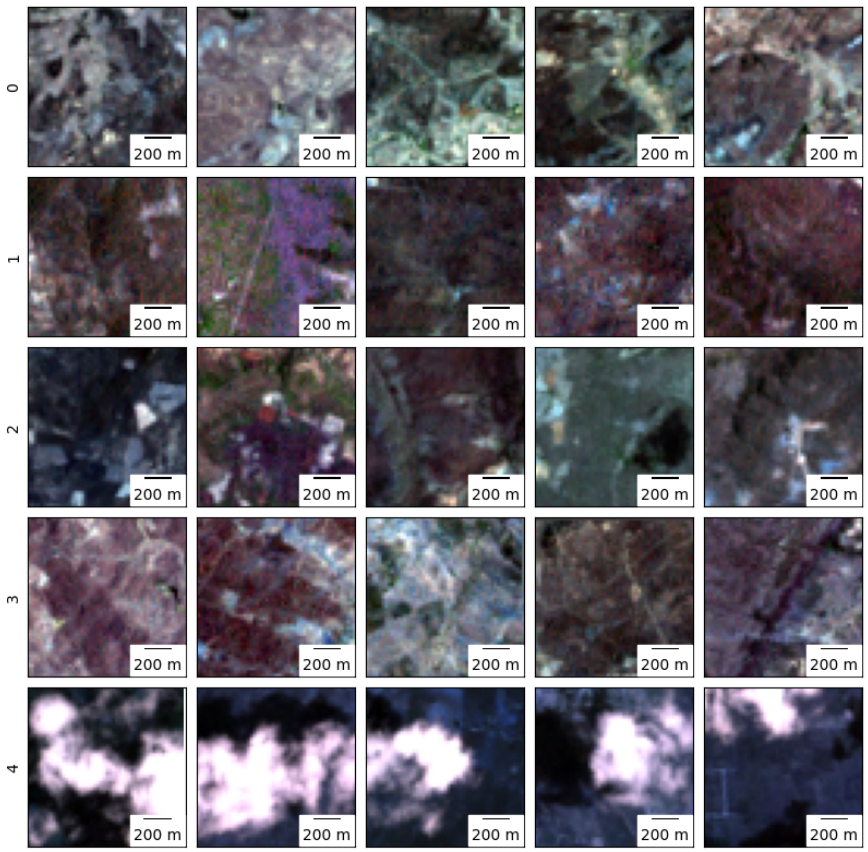
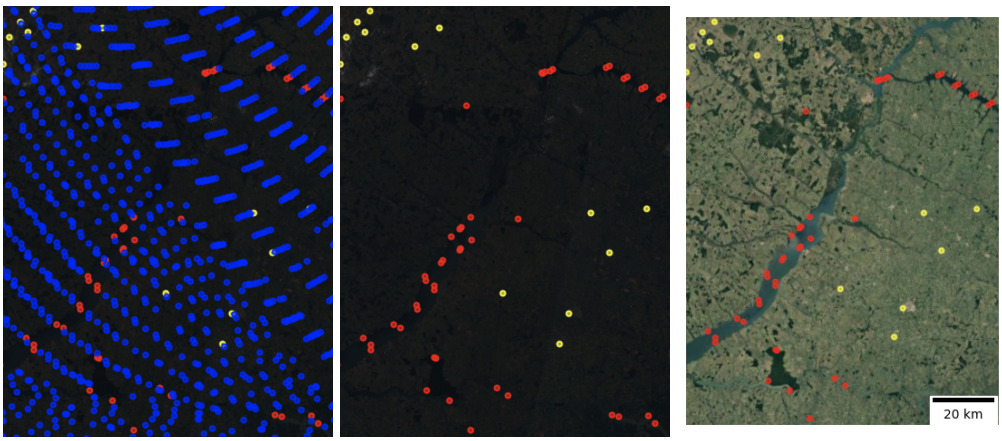


Figure 7.2: Each row represents the five images closest to the center of a cluster for the Caatinga biome, with each row representing a different cluster. Each image has been normalised to make them brighter and easier for analysis.

images had been extracted, it appeared to begin forming a pattern. However, as enough images were still being extracted to cover the overwhelming majority of the quadrant the pattern created would not have had a bias towards the results. From Section 7.3 we get an idea of where the miss classifications are occurring, for example, we see that the river has been mostly predicted as belonging to the Amazon. As was discussed in section 7.1 this could be amended by creating an additional class that contains bodies of water. By doing the same analysis for the Amazon and Caatinga it was found that a large amount of false positives in the Amazon were a result of a section not being forested (D.2), and for Caatinga a large amount of area had been classed as Cerrado due to it having semi-dense forest (D.2). Figures 7.3 and D.2 have had the false positives overlaid on Google Earth high resolution satellite imagery to make for easier analysis given the darkness of the Landsat 7 imagery when viewed in a PDF.



(a) (b) (c)

Figure 7.3: Area of high density false positives in quadrant 4 of the Cerrado biome. Each point represents the location of an image and its colour the predicted class (yellow is Caatinga, blue is Cerrado and red is the Amazon) (a) All images and their predicted class overlaid on Landsat 7 imagery (b) False positives overlaid on Landsat 7 images (c) False positives overlaid on high resolution imagery obtained from Google Maps © 2021 TerraMetrics [42]

To reduce the number of false positives and increase the confidence behind future predictions, a threshold was introduced in the classifier. The threshold meant that any image where the probability of belonging to the predicted class was not greater than the a set amount then the image would not be used. The threshold used was 0.7 and the results for the Caatinga biome can be seen in Figure 7.4. From Table 7.1 it is obvious that by introducing a threshold for the likelihood of an image belonging to the predicted class that the number of false positives has been greatly reduced. Micro f1-score was used in Table 7.1 given the overwhelming class imbalance that is present when evaluating on each individual test quadrant.

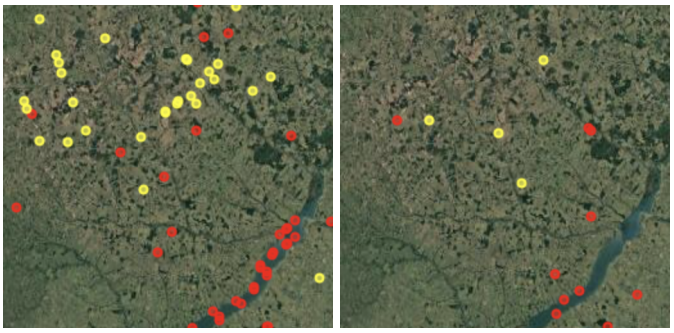
7.1.2 Predicting Unseen Areas

Two areas were chosen for future predictions, the first was the area the Amazon covers outwith Brazil, into countries such as Bolivia, Peru, Columbia and Venezuala, the second was the margin in between the boundaries of each of the biomes (as can be seen in Figure 4.1).

As would be expected, 76% of the 838 images used in the first geometry belong to the Amazon, with only 9.90% belonging to Cerrado and 1.43% to Caatinga. The model

	Micro F1-Score	
	Without Threshold	With Threshold
Amazon Test Quad	0.875	0.924
Cerrado Test Quad	0.954	0.975
Caatinga Test Quad	0.673	0.727

Table 7.1: Micro f1-score for each test quadrant with and without the introduction of a threshold probability for each predicted image



(a) (b)

Figure 7.4: Before (a) and after (b) of applying a threshold to the probability of an image belonging to a class for a section of the Cerrado biome. Satellite imagery obtained from Google Maps © 2021 TerraMetrics [42]

has managed to accurately capture the location of the Amazon rainforest as it expands outwith Brazil while very impressively capturing the section of Cerrado biome that is isolated to the South.

The margins that the second geometry made up were excluded from the training and test set due to the coarse resolution of the ground truth data giving overlapping results for each of the biomes. Now that the model has been trained on each of the biomes this geometry would provide useful analysis on whether or not the ResNet18 would be able to create a finer boundary between the biomes. Figure 7.5 (b) shows that the ResNet18 combined with a MLR classifier was able to nicely fill in the margin, providing a clear boundary between each of the biomes. It should be noted that there are certain gaps with no predictions at all. This is a direct result of no images during the given time period having less than 10% cloud cover and so were not used as part of the predictions. Out of all 4028 images used in this geometry, 23.36% were predicted

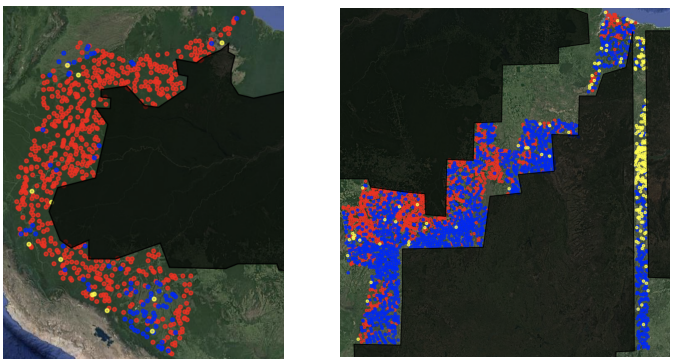


Figure 7.5: Predictions for (a) the area of the Amazon outwith of Brazil and (b) the margin between the training and test quadrants for each of the biomes. Points are coloured depending on the class they were predicted; red for Amazon, blue for Cerrado and green for Caatinga. Satellite imagery obtained from Google Maps © 2021 TerraMetrics [42]

as the Amazon, 43.00% as Cerrado and 5.56% as Caatinga, thus leaving 28.08% being dropped due to the probability threshold. These ratios make sense given the margin surrounds the Cerrado biome on all sides while only bordering the Caatinga biome for a very small portion.

Chapter 8

Discussion

Of all the models used, ResNet18 in combination with a multinomial logisitic regression classifier had the highest macro f1-score (6.2). This is likely due to it having the largest training dataset of 1.2 million labelled natural images across 1,000 categories as well as its award winning architecture. From the t-SNE in Figure 7.1 we see that ResNet18 has indeed found a clean separation for the Amazon class with surprisingly little overlap between the Cerrado and Caatinga classes, unlike what NDVI expressed in Section 5.1.

Surprisingly, the trained Tile2Vec model performed the poorest despite having been trained from scratch on the data. This could either be from a lack of training data - the dataset it was trained on was the smallest of all the training sets each SOTA model was trained on - or it could be from the changes made to the triplet sampling method that reduced the efficacy of the Tile2Vec model. There is also the fact that the pre-trained Tile2Vec model was trained on a resolution of 0.6m per pixel while the Landsat 7 images that the poorer performing Tile2Vec model used had a resolution of 30m per pixel. The trained Tile2Vec model had the deepest depth out of any of the decision trees (6.2), further emphasising the fact it has struggled to find a clean representation of the data.

The pre-trained SwAV model was the second worst performing model, with the best classifier (MLR) in combination with SwAV producing a macro f1-score of 80%. There are many possible reasons behind this but the three most interesting are; (1) that there are less overlapping clusters in the data as was originally thought, (2) that there are so many overlapping clusters that the biomes cannot be separated by using them and (3) that the model itself simply has an architecture that does not perform as well as the other SOTA algorithms used. However, it is worth stating that option (3) is the

most unlikely possibility given that SwAV achieved a higher top-1 accuracy on the ImageNet dataset compared to ResNet18 (by 4.6%)[33][28].

The multinomial logistic regression (MLR) classifier was the best performing classifier for every model as well as the NDVI dataset. Despite the more complex decision boundary that the Random Forest is capable of, the logistic regression model was still able to separate the data with higher precision and recall. This better performance by the MLR model could likely be a result of there being more explanatory variables which Random Forests tend to struggle with [43].

During the exploratory analysis it become apparent that certain false positives - where images were classed as belonging to a biome different to the one expected - were correctly identifying an eco-system that belongs to a different biome. For example in Figure D.2, where the area that is no longer the Amazon rainforest has been classed as Cerrado and Caatinga biomes. Although these data points are not a part of the Cerrado or the Caatinga biomes they are also not a part of the Amazon. This highlights an interesting point about the use of a tool that can predict the location of certain biomes where the model would be best used alongside an expert. The tool would be able to remotely classify the majority of the area each biome covers allowing experts to only need to examine the suspicious areas that stand out such as that in Figure D.2.

By including the threshold we see that the micro f1-score of each test set prediction increases significantly (Table 7.1), proving that it has greatly reduced the number of incorrect predictions and that it is a valid method for increasing confidence behind the classifications.

There is the further point that all data used in this paper was gathered between the sub-equatorial summer months of October to March. It was seen in Figure 5.1 that the winter months had a greater overlap of the Cerrado and Caatinga biomes due to the reduced amount of vegetation in the Cerrado biome at that time of year. It would be interesting to see the effect this has on the models capabilities at separating the data as it is possible the model is only effective during the summer months running the possibility that the 16-day update time proposed in Section 1.2 would not be achievable year round.

Chapter 9

Conclusion

By applying transfer learning to the data, a ResNet18 model was able to achieve accurate results (84.3% macro F1-score) on the three main biomes in Brazil; the Amazon, Cerrado and Caatinga. The transfer learning technique was even more effective than the Convolutional Neural Network that was trained on the data itself, perhaps arising from a poor CNN architecture or due to a lack of training images. By looking at the results that were incorrectly classified it was possible to increase the precision of the model by only classifying images that had a greater than 70% probability of belonging to the predicted class, after this was applied the false positives were almost completely removed. Furthermore, the classification was made using medium resolution Landsat 7 satellite imagery that is open source and free to access, making it a significantly cheaper technique compared to those currently in use. From these results it was possible to extract how the ResNet18 model was representing each biome through eco-systems such as farmland, rivers and forests which the contrastive cluster learning model SwAV was unable to use to get a better performance. This highly accurate ResNet18 model proves that it is possible for an online model to provide live updates on the area of the Amazon, Cerrado and Caatinga biomes in South America, and when used alongside an expert can help quickly identify anomalies in each biome due to deforestation or any other factor that would change the appearance of a biome.

Chapter 10

Future Work

The following is work that can be carried out for the second part of this project in order to improve upon current results, as well as expand the scope of what is currently achievable to improve on the ability in monitoring the transformation of biomes over time.

The best model in this paper (ResNet18 with an MLR classifier) was successful using RGB images alone. By including other features such as NDVI or SWIR 2 - Landsat 7's shortwave infrared channel that was shown to be a much better classifier for the Cerrado biome by Beuno et. al. [22] - it is highly likely that an even better performance can be achieved.

All data gathered and used in this paper was collected during the sub-equatorial summer months. It is likely that during the winter months when the Cerrado biome has less vegetation making it harder to separate from other biomes that the performance of the classifier worsens.

During the exploratory analysis and predicting unseen areas it was clear that there would often be an outlying point predicted as a biome that was clearly not there. In order to prevent this from happening a possible technique would be to check the class of all surrounding points and if less than a certain number share the same class as the predicted point then the false positive would be removed.

As was seen during the exploratory analysis, certain eco-systems - such as rivers - were being associated with a particular biome, despite these eco-systems not being a defining characteristic of that biome. In order to prevent this from becoming a bias and having an adverse effect on the classification of future images, additional classes should be introduced, for example a class to represent rivers and another for infrastructure are obvious choices. Figure 7.2 also highlighted that every biome was sharing

the cluster of cloudy images, and so either images with clouds in them should be completely removed, however, this runs the risk of entire areas not being covered by predictions as was seen in Figure 7.5, so a better solution may be to include cloudy images as a separate class.

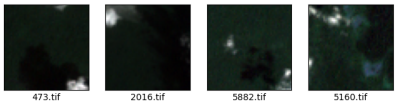
The main proposal of this paper is to use the best performing model to analyse how the classification of an area changes with time. To achieve this, data from the same location from different time periods will need to be compared, specifically the difference of each biomes coverage between the images (e.g. comparing the coverage of biomes from data gathered in 2010 to data gathered in 2015).

Ayush et. al. [44] have managed to reduce the performance gap between contrastive learning methods with supervised learning methods (a gap that was seen in this paper between SwAV and ResNet18) for remote sensing data by making use of the spatiotemporal property. By applying Ayush et. al.'s methods here it is hopeful that SwAV will outperform ResNet18.

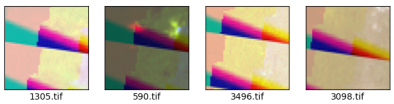
Currently the data must be manually exported and then fed into the classifier, before being overlaid onto a map. This is very time consuming and not an efficient solution for the proposed use case of monitoring the expansion and collapse of biomes on a bi-weekly basis. To improve on this, a closed pipeline should be created where a user can select the area they wish to monitor and then a timeline of how the classification of biomes has changed within that area during the timeline is returned.

Appendix A

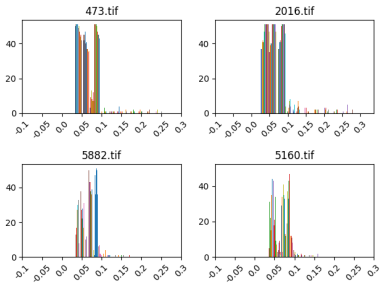
Data Preprocessing



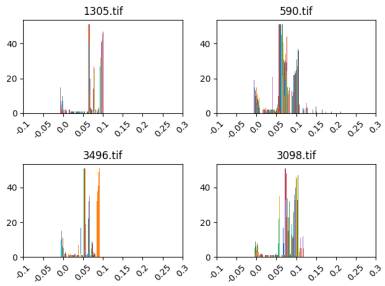
(a) Satisfactory images



(b) Corrupt images



(c) Histograms of satisfactory images

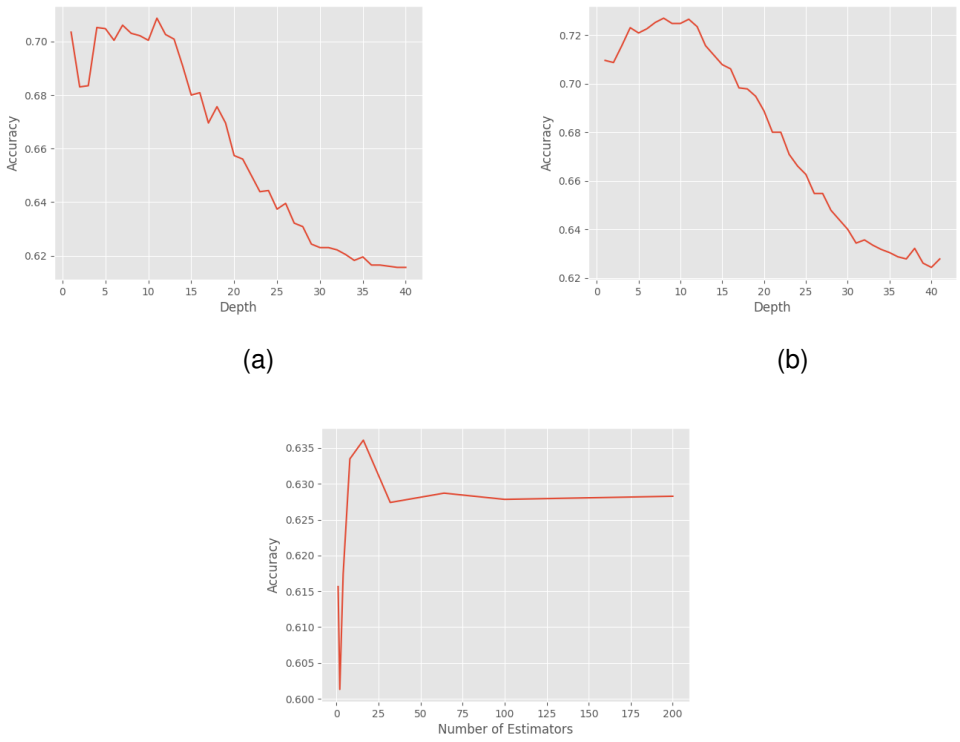


(d) Histograms of corrupt images

Figure A.1: The similarity between histograms for both satisfactory and corrupt images, making it an unfeasible method for separation.

Appendix B

Baseline Experiments



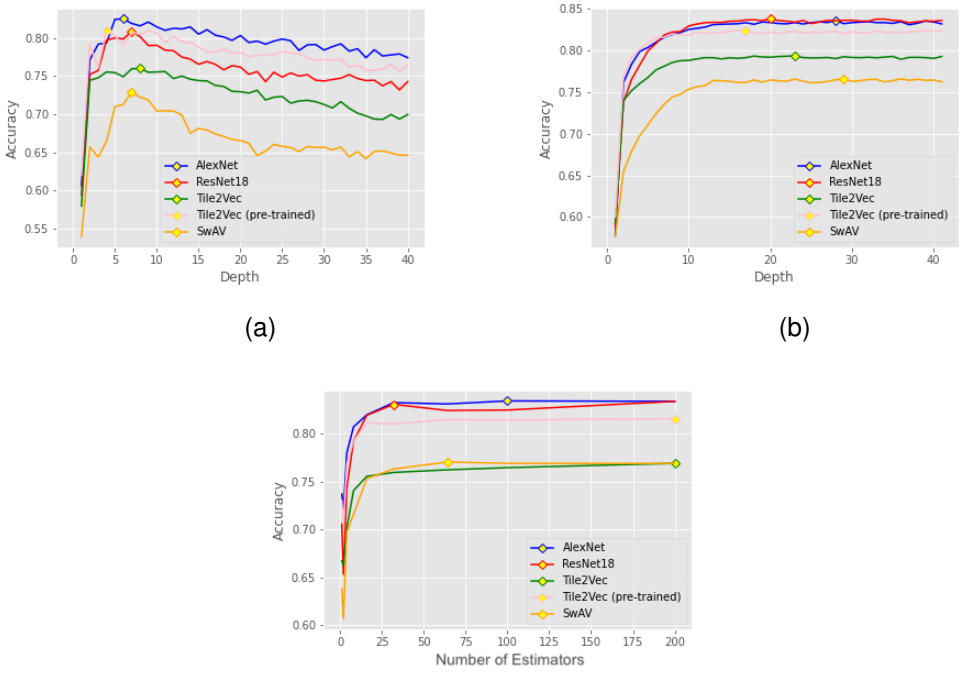
(a) (b)

(c)

Figure B.1: Accuracy of a (a) decision tree and (b) Random Forest classifier on the NDVI validation dataset as the depth increases. (c) Shows the accuracy of the Random Forest on the NDVI validation dataset as the number of estimators increases.

Appendix C

SOTA Experiments



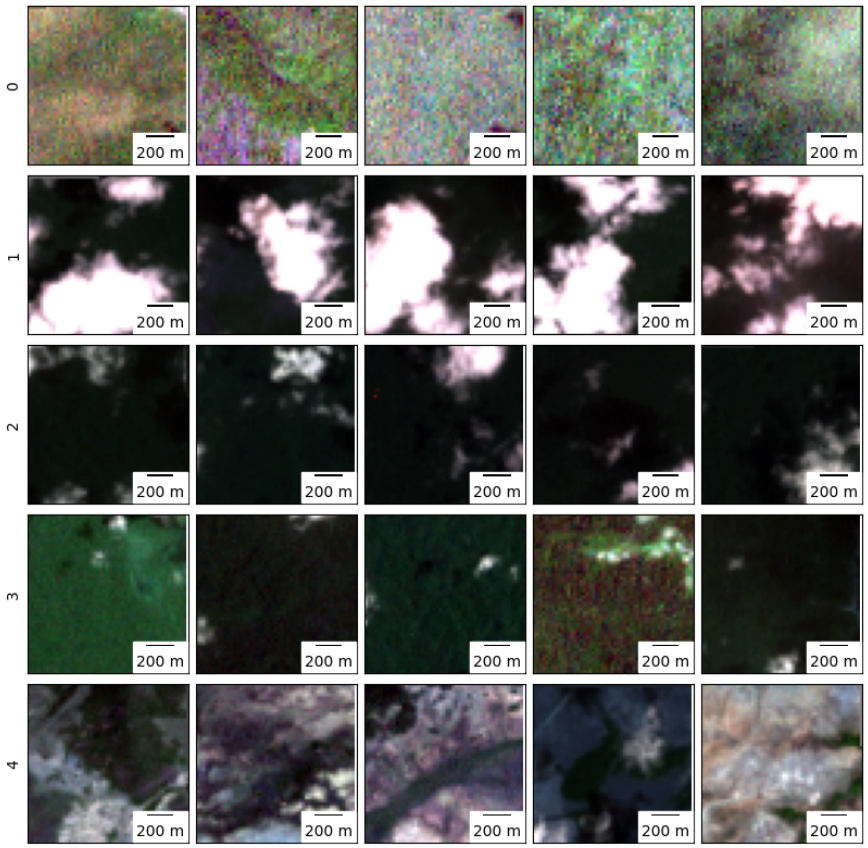
(a) (b)

(c)

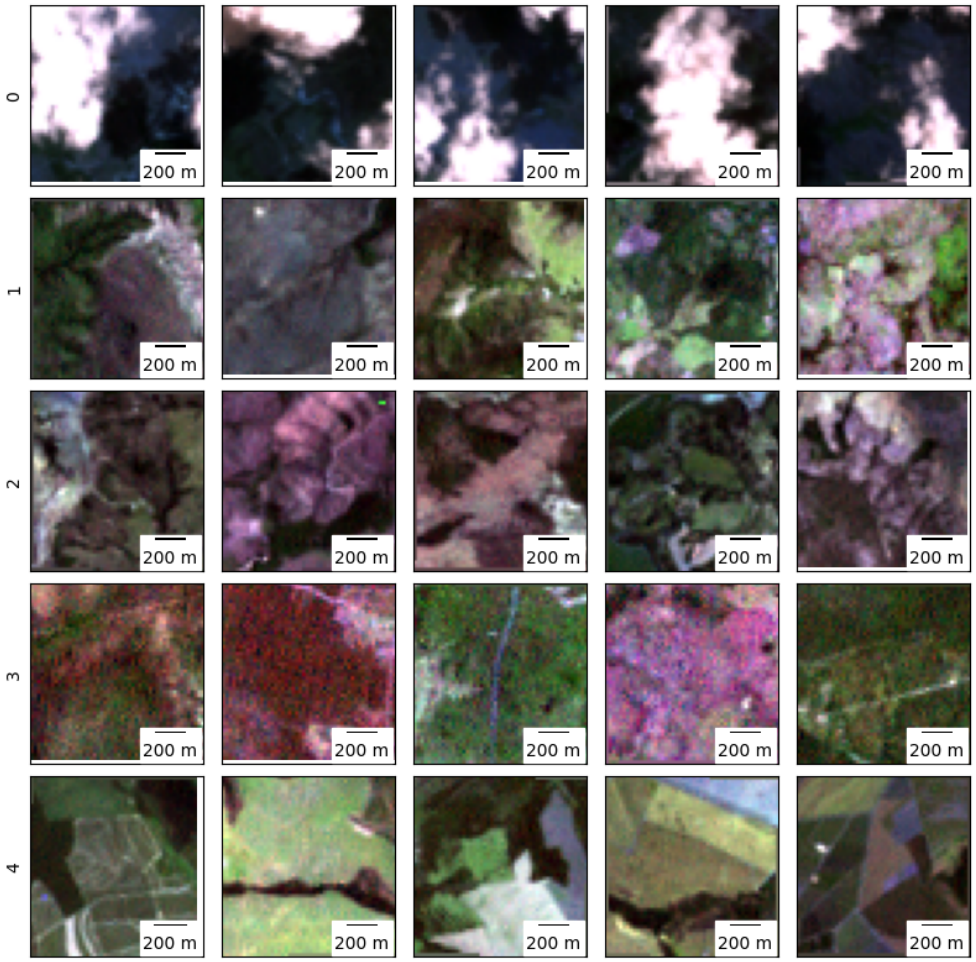
Figure C.1: Accuracy of a (a) decision tree and (b) Random Forest classifier on the RGB validation dataset for each SOTA algorithm as the depth of the decision tree increases. (c) Shows the accuracy of the Random Forest for each SOTA algorithm on the RGB validation dataset as the number of estimators increases. Optimal hyperparameters are marked with a yellow diamond.

Appendix D

Evaluation



(a)



(b)

Figure D.1: Five images closest to the centers of each cluster for (a) Amazon and (b) Cerrado. Each image has been normalised to make them brighter and easier for analysis, however an issue with the normalisation process has led to certain images being corrupted

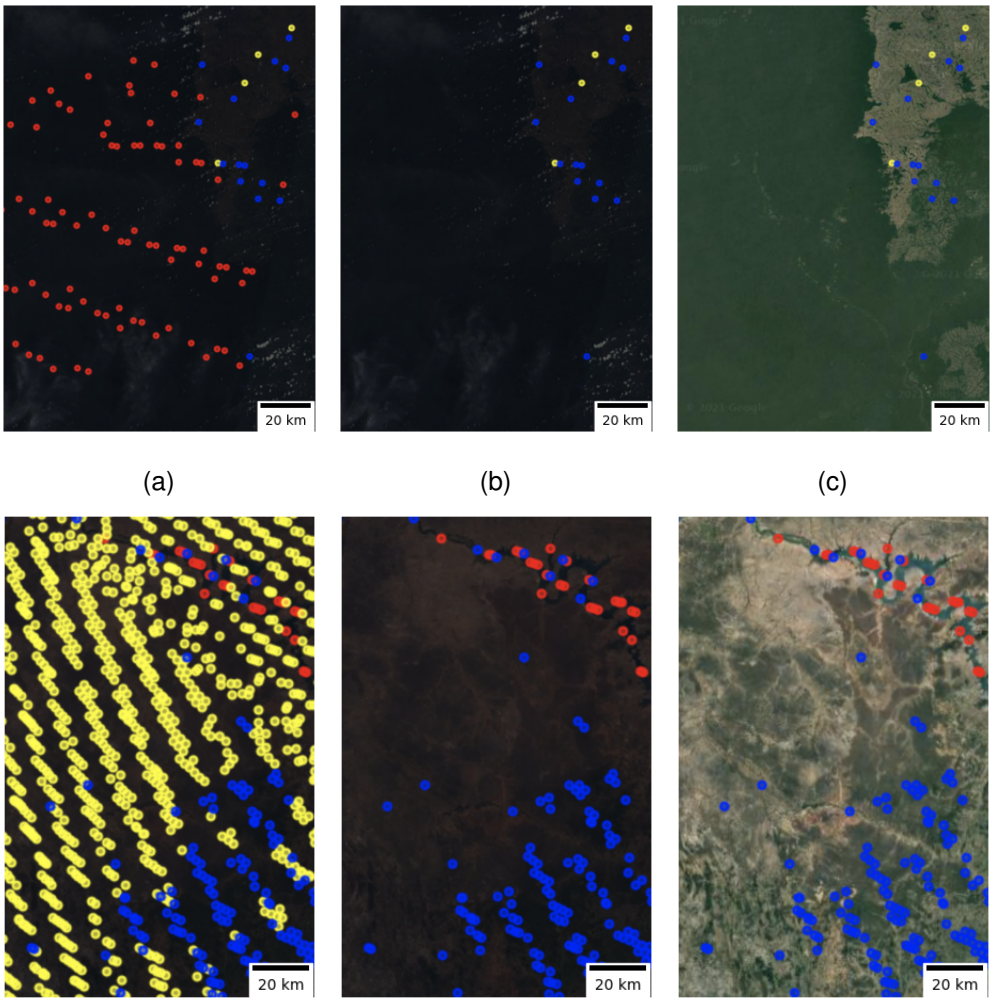


Figure D.2: Area of high density false positives in quadrant 1 of the Amazon biome (a)(b)(c) and the area of high density false positives in quadrant 3 of the Caatinga biome (d)(e)(f). Each point represents the location of an image and its colour the predicted class - yellow is Caatinga, blue is Cerrado and red is the Amazon - (a)(d) All images and their predicted class on Landsat 7 imagery (b)(e) Only false positive images on Landsat 7 images (c)(f) False positives on high resolution imagery obtained from Google Maps © 2021 TerraMetrics [42]

Bibliography

[1] JULIANO ASSUNÇÃO and CLARISSA GANDOUR. Combating illegal deforestation strengthening command and control is fundamental. Technical report, Climate Policy Intiative, February 2019.

[2] IBGE Biomas. sistema costeiro-marinho do brasil: compatível com a escala 1: 250 000. *Rio de Janeiro: IBGE*, 2019.

[3] William F Laurance. Reflections on the tropical deforestation crisis. *Biological Conservation*, 1999.

[4] Giselda Durigan, Marinez Ferreira de Siqueira, and Geraldo Antonio Daher Correa Franco. Threats to the cerrado remnants of the state of são paulo, brazil. *Scientia Agricola*, 64(4):355–363, 2007.

[5] Lisa Knopp, Marc Wieland, Michaela Rättich, and Sandro Martinis. A deep learning approach for burned area segmentation with sentinel-2 data. *Remote Sensing*, 12(15):2422, 2020.

[6] Mabel Ortega Adarme, Raul Queiroz Feitosa, Patrick Nigri Happ, Claudio Aparecido De Almeida, and Alessandra Rodrigues Gomes. Evaluation of deep learning techniques for deforestation detection in the brazilian amazon and cerrado biomes from remote sensing imagery. *Remote Sensing*, 12(6):910, 2020.

[7] A. K. Neves, T. S. Körting, L. M. G. Fonseca, C. D. Girolamo Neto, D. Wittich, G. A. O. P. Costa, and C. Heipke. Semantic segmentation of brazilian savanna vegetation using high spatial resolution satellite data and u-net. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-3-2020:505–511, 2020.

[8] Anne Schucknecht, Stefan Erasmi, Irmgard Niemeyer, and Jörg Matschullat. As-

sessing vegetation variability and trends in north-eastern brazil using avhrr and modis ndvi time series. *European Journal of Remote Sensing*, 46(1):40–59, 2013.

[9] MapBiomas. Mapbiomas. https://mapbiomas.org, 2015.

[10] Lawrence N Hudson; Tim Newbold; Sara Contu; Samantha L L Hill et al. The predicts database: a global database of how local terrestrial biodiversity responds to human impacts. *Ecology and Evolution*, 4(24):4701–4735, 2014.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[12] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[13] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Apr 2014.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[15] Nathalie Pettorelli, Sadie Ryan, Thomas Mueller, Nils Bunnefeld, Bogumila Jedrzejewsk, Mauricio Lima, and Kyrre Kausrud. The normalized difference vegetation index (ndvi): Unforeseen successes in animal ecology. *Climate Research*, 46:15–27, 01 2011.

[16] IBGE (Instituto Brasileiro de Geografia e Estatística). Mapa de biomas do brasil. primeira aproximação, 2004.

[17] Fabien H Wagner, Alber Sanchez, Yuliya Tarabalka, Rodolfo G Lotte, Matheus P Ferreira, Marcos PM Aidar, Emanuel Gloor, Oliver L Phillips, and Luiz EOC Aragao. Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, 5(4):360–375, 2019.

[18] Neal Jean, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. Tile2vec: Unsupervised representation learning for spatially distributed data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:3967–3974, 2019.

[19] René Beuchle, Rosana Cristina Grecchi, Yosio Edemir Shimabukuro, Roman Seliger, Hugh Douglas Eva, Edson Sano, and Frédéric Achard. Land cover changes in the brazilian cerrado and caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Applied Geography*, 58:116–127, 2015.

[20] F Ian Woodward, Mark R Lomas, and Colleen K Kelly. Global climate and the distribution of plant biomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1450):1465–1476, 2004.

[21] G Teixeira Batista, Y Edemir Shimabukuro, and W Thomas Lawrence. The long-term monitoring of vegetation cover in the amazonian region of northern brazil using noaa-avhrr data. *International Journal of Remote Sensing*, 18(15):3195–3210, 1997.

[22] Inacio T Bueno, Fausto W Acerbi Júnior, Eduarda MO Silveira, José M Mello, Luís MT Carvalho, Lucas R Gomide, Kieran Withey, and José Roberto S Scolforo. Object-based change detection in the cerrado biome using landsat time series. *Remote Sensing*, 11(5):570, 2019.

[23] Maria Olczak, Andris Piebalgs, and Christopher Jones. Satellite and other aerial measurements: a step change in methane emission reduction? 2020.

[24] Marc K Steininger, Compton J Tucker, John RG Townshend, Timothy J Killeen, Arthur Desch, Vivre Bell, and Peter Ersts. Tropical deforestation in the bolivian amazon. *Environmental conservation*, pages 127–134, 2001.

[25] Alber Hamersson Sanchez, Michelle Cristina A Picoli, Gilberto Camara, Pedro Ribeiro Andrade, Michel Eustaquio D Chaves, Sarah Lechler, Anderson R Soares, Rennan FB Marujo, Rolf Ezequiel O Simões, Karine R Ferreira, et al. Comparison of cloud cover detection algorithms on sentinel–2 images of the amazon tropical forest. *Remote Sensing*, 12(8):1284, 2020.

[26] James Storey, James Lacasse, Ronald Smilek, Tracy Zeiler, Pasquale Scaramuzza, Rajagopalan Rengarajan, and Michael Choate. Image impact of the landsat 7 etm+ scan line corrector failure.

[27] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[29] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116, 1998.

[30] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[31] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[32] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

[33] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*, 2020.

[34] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. Big self-supervised models are strong semi-supervised learners. *arXiv preprint arXiv:2006.10029*, 2020.

[35] Yan-Yan Song and LU Ying. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2):130, 2015.

[36] Mahesh Pal. Random forest classifier for remote sensing classification. *International journal of remote sensing*, 26(1):217–222, 2005.

[37] David G Kleinbaum, K Dietz, M Gail, Mitchel Klein, and Mitchell Klein. *Logistic regression*. Springer, 2002.

[38] Andrew Y Ng. Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78, 2004.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[40] Rodrigo Cauduro Dias De Paiva, Diogo Costa Buarque, Walter Collischonn, Marie-Paule Bonnet, Frédéric Frappart, Stephane Calmant, and Carlos André Bulhões Mendes. Large-scale hydrologic and hydrodynamic modeling of the amazon river basin. *Water Resources Research*, 49(3):1226–1243, 2013.

[41] Samuel Bridgewater, James A Ratter, and José Felipe Ribeiro. Biogeographic patterns, β-diversity and dominance in the cerrado biome of brazil. *Biodiversity & Conservation*, 13(12):2295–2317, 2004.

[42] Terrametrics 3d terrain data amp; visualization – visualizing your world.

[43] Kaitlin Kirasich, Trace Smith, and Bivin Sadler. Random forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Science Review*, 1(3):9, 2018.

[44] Kumar Ayush, Burak Uzkent, Chenlin Meng, Marshall Burke, David Lobell, and Stefano Ermon. Geography-aware self-supervised learning. *arXiv preprint arXiv:2011.09980*, 2020.