

Machine Learning With Medium Resolution Satellite Imagery To Map the Biomes In South America

Calum McMeekin



MInf Project (Part 2) Report
Master of Informatics
School of Informatics
University of Edinburgh

2022

Abstract

Knowing the area that each of the biomes in South America covers is crucial to monitoring land use changes, illegal deforestation, and the planet's climate. Current mapping techniques have two major limitations: (i) they are expensive and time consuming to produce with only two maps being released in the past 15 years and (ii) fieldwork is only carried out in specific areas leading to assumptive classifications everywhere else. In this project, a database of over 67,000 freely available Landsat 8 satellite images is created and then modern machine learning and image processing algorithms are applied. The project presents the first research that compares the effectiveness of different Neural Networks, ConvLSTMs and ResNets, in order to classify the three largest biomes in South America using visible light and vegetation indexes as inputs. The work shows a ResNet model using visible light inputs achieved an excellent macro F1 score of 92.0%, 9.5% greater than the ConvLSTM. The research found that the land each biome covers is changing at a rate much faster than current maps are being produced. However, it is possible to frequently, accurately, and cheaply produce maps of the three largest biomes in South America using the techniques presented here. This project has provided the framework for a system to be built that allows open access to the maps our ResNet model is capable of producing. A system such as this would enable conservation charities to make use of a powerful tool to monitor land use changes across one of the largest continents in the world, while focusing the attention of fieldwork to areas of intense change.

Research Ethics Approval

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

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)

Acknowledgements

The following work achieved in this paper could not have been completed without the continued support from my supervisor Sohan Seth. His professional guidance steered me towards discoveries I didn't think were possible, and his constant enthusiasm gave me drive to continue pushing myself beyond what I thought I was capable of.

I would also like to thank both my parents and my grandparents for their love and encouragement from my first day of school, and my friends, for not shaming me when I ask silly questions.

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Summary of Previous Work	2
1.3	Objectives	3
2	Background	4
2.1	South American Biomes	4
2.2	Neural Network (NN)	5
2.2.1	Convolutional Neural Network (CNN)	5
2.2.2	Recurrent Neural Network (RNN)	5
2.2.3	Long Short Term Memory (LSTM)	6
2.2.4	Convolutional LSTM (ConvLSTM)	7
2.3	Short Wave Infrared (SWIR) Imaging	8
2.4	Previous Work	8
3	Data Collection	9
3.1	Landsat 8 Image Collection	9
3.2	Landsat 8 Image Preprocessing	10
3.3	Annotated Sources	11
4	Methodology	12
4.1	Train/Test Split	12
4.2	Bayesian Optimization	13
4.3	Batch Normalisation	15
4.4	SNDVI	15
4.5	ConvLSTM	16
4.6	ResNet	17
4.7	Dropout	17
5	Experiments	19
5.1	Bayesian Optimization vs Manual Search	19
5.1.1	Approach	19
5.2	SNDVI vs RGB using a ConvLSTM	20
5.2.1	Models and Data Used	20
5.3	Spatial vs Spatial-Temporal	20
5.3.1	Models and Data Used	20

6 Results	22
6.1 Bayesian Optimization vs Manual Search	22
6.2 SNDVI vs RGB using ConvLSTM	23
6.3 Spatial vs Spatial-Temporal	24
7 Exploratory Analysis	25
7.1 Mapping Capability	25
7.2 Frequency of Change	26
7.3 Detecting Change on a Large Scale	28
7.4 Epistemic Uncertainty	28
7.5 Determining Classification	29
8 Discussion	31
9 Conclusion	33
10 Future Work	34
Bibliography	35
A ResNet	39
A.1 Hyperparameter tuning	39
B Mapping Capabilities	40
B.1 Detecting Change on a Large Scale	41

Chapter 1

Introduction

Problem Statement: To assess if it is possible to accurately map the main biomes in South America using modern machine learning and remote sensing techniques.

1.1 Motivation

A biome is a collection of ecosystems that are close together and share similar fauna and flora. The Amazon, Caatinga and Cerrado are three of the four largest biomes that make up South America, covering an area over 9,460,037km². Knowing the area each of these biomes cover is crucial to monitoring land use changes, illegal deforestation, and the planets climate [1]. Current mapping techniques are expensive and time consuming with only two maps being produced, one in 2004 and one in 2019 [2]. These maps are produced using fieldwork in the areas where biomes border each other [2]. As a consequence, the areas within each biome where the fauna and flora have changed are not captured in the maps currently produced, leading to inaccuracies.

There are many reasons the fauna and flora change within a biome and not just on the borders. Both illegal and legal deforestation run rampant throughout South America with the Amazon being subject to losing 3,000km² of tropical rainforest every year [3]. The rate at which these biomes change is increasing with just under 10,000km² of tropical rainforest being cut down in 2019. An increase in droughts has also led to wildfires wreaking havoc across Brazil with 16% of the Amazon's remaining forest expected to burn by 2050 [4]. This increased rate of land use change means that maps are not only required to be produced more frequently, but they must also include the areas within the borders of the biomes so that these internal changes do not go undocumented.

Recent groundbreaking developments in deep learning have led to the ability to accurately classify the different eco-systems that make up the biomes in South America [5][6]. By combining the machine learning and image processing methods used in these approaches with freely accessible medium resolution Landsat 8 satellite imagery it is hoped that **this project will show for the first time it is possible to inexpensively map, not just eco-systems, but the three largest biomes in South America.**

1.2 Summary of Previous Work

There is currently only one official source of biome maps in South America, the Brazilian Institute of Geography and Statistics (IBGE) [2]. Given IBGE relies on field sampling to create its maps they are limited in the extent to how much land they can accurately cover. As a result the studies are only carried out in places where neighbouring biomes meet and the areas inside the borders of each biome are given a blanket classification. The most recent map produced by IBGE in 2019 is at a resolution of 1,250m per pixel.

MapBiomas is a collaboration of NGOs, universities and technology companies that provides updated maps of the current land uses in Brazil on a yearly basis [7]. They achieve this impressive service by using a combination of machine learning with remote sensing and expert consultancy to cover large areas of land while using the maps produced by IBGE as a baseline. This publication does not update the biome boundaries themselves, however, it does show it is possible to accurately map large areas of land using machine learning and remote sensing.

Time series data in combination with machine learning has recently been able to provide accurate classifications of land use over time. Barreto et. al. found that ConvLSTMs can accurately classify areas of deforestation in Indonesia across a four year period [8]. Duku et. al. have also found that a ConvLSTM was able to predict the amount of rainfall across regions in Africa by inputting historic rainfall and vegetation data [9].

Our previous project has already made great strides in showing it is possible to map the biomes in Brazil using remote sensing with machine learning [10]. The results showed that both the Normalized Difference Vegetation Index (NDVI) and the visible light spectrum were able to correctly classify the three biomes with a respective 71.4% and 84.3% accuracy. This attempt achieved the best result using transfer learning with a ResNet18 architecture, leading to concerns that had the model been trained solely on the dataset it would have performed even better. Furthermore, the input images were provided by Landsat 7, yet the more recent Landsat 8 includes a wider variety of bands that may provide a more accurate indicator than NDVI.

1.3 Objectives

This project wishes to show that it is possible to map the largest biomes in South America using machine learning and freely available Landsat 8 satellite imagery. The project will then move on to gain a deeper understanding of what parts of the input images are influencing the machine learning models decisions and if there is any uncertainty present in the models predictions.

In light of these goals a set of objectives can be lined out as follows:

1. Gather a database of Landsat 8 satellite images for the three biomes that covers a period of five years.
 - (a) Apply preprocessing methods used in our previous project to this data, removing any corrupted images and organising them into a normalised format that the machine learning models can later handle, [section 3.2](#).
 - (b) Separate this database of Landsat 8 images into training, validation and testing datasets that can be used by the Neural Networks, [section 4.1](#).
2. Evaluate the Neural Network techniques to understand the potential for mapping the Amazon, Cerrado, and Caatinga
 - (a) Discern if Bayesian Optimization is capable of finding a better set of hyperparameters for the machine learning models used in this project compared to our previous manual search methods, [section 6.1](#).
 - (b) Combine the new SWIR bands of Landsat 8 with NDVI to create a new input, SNDVI. The performance of this indicator will then be evaluated and compared to the visible spectrum as a potential input to the models, [section 6.2](#).
 - (c) Evaluate the effectiveness of using a spatial-temporal model such as ConvLSTM to a spatial model such as a ResNet for biome classification. Both models will be trained from scratch on the database, [section 6.3](#).
3. Explore the maps produced by the best performing Neural Network
 - (a) Examine how capable the model is at creating accurate maps in areas that undergo several different land use changes, [section 7.1](#).
 - (b) See what benefits are to be had by being able to frequently create maps showing the different biomes within an area, [section 7.2](#).
 - (c) Determine the uncertainty of the predictions made by the best performing model by inspecting any epistemic uncertainty that may be present, [section 7.4](#).
 - (d) Explore what parts of the input images are influencing the decision made by the model to assign the image to a specific biome. This will be achieved using Grad-CAM, an algorithm that is capable of producing heatmaps on the input images according to how important certain areas are perceived by the model [11], [section 7.5](#).

Chapter 2

Background

2.1 South American Biomes

When several ecosystems are situated geographically close to each other while sharing similar fauna and flora then they will be classed under an umbrella term - biome. This term is useful for when you are describing large sections of land that all react in the same way to changes in climate or land-use [12].

According to the Brazilian Institute of Geography and Statistics (IBGE) there are six biomes that make up the majority of Brazil [2]. From these biomes the Pantanal and Pampa biomes are too small to feature any further on in this paper [13]. The remaining four biomes are situated on or above the equator, covering the majority of land mass in the northern and central areas of Brazil. However, it is well known that these biomes are not restricted to Brazil and span the majority of the land in northern and central South America [13].

Each biome changes shape at different rates. Currently the fastest changing biome is the Amazon, due to deforestation and savannization (the conversion of a tropical rainforest ecosystem to a savanna ecosystem via forest fires) it is shrinking at a rate of 3,000km² each year [3][14]. Vegetation models have seen that the area covered by tropical forest biomes (i.e. the Amazon) will decrease while the area of savanna biomes (Cerrado) will increase. These models have predicted that during the period of 2020-2029 the tropical forests in the Amazon will have decreased by 3%, during 2050-2059 they will decrease by 9% and a further 18% in 2090-2099 [14].

Biome	Area (km ²)	Ecosystems	Seasonal
Amazon	6,700,000	- Tropical forest - Blackwater forests - Whitewater forests - Brazilian highlands	Partly
Cerrado	1,910,037	- Forest savanna - Savanna wetlands	Yes
Mata Atlantic	1,315,460	- Broad-leaf tropical forests - Tropical grasslands - Subtropical grasslands - Mangrove forests	Yes
Caatinga	850,000	- Xeric shrubland - Thorn forest	Yes

Table 2.1: Breakdown of the four largest biomes that make up Brazil

2.2 Neural Network (NN)

Neural networks (NNs) are a network of nodes that aim to mimic the process of the brain. The nodes are usually sectioned into layers where nodes in each layer take a set of inputs, apply a mathematical operation and then pass on the output to the next layer of nodes [15]. Weights are applied to the inputs of each node, determining how much influence each particular output from the previous layer should have at this current moment. When a NN is trained on a set of data, it is these weights that are tweaked during each training iteration in order to minimise the difference between what was expected to be output by the final layer, and what was actually output [15].

2.2.1 Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a deep neural network that has been tailored towards the processing of images. The convolutional layer that was introduced as part of the CNN is what makes the network so adept at processing images. By applying a kernel over the image, it simultaneously reduces the size of the input matrix while distilling it down, holding only onto the information that is useful [16][17].

A more in-depth and detailed description of CNN's can be found in the previous part of this project [10].

2.2.2 Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN) is a form of neural network that is best suited towards temporal data given they use information from previous inputs as part of the weights for future outputs. Instead of conventional backpropagation that other NNs use, RNNs make use of backpropagation through time (BPTT) to determine how the gradients of the weights will be updated. As the parameters are shared across layers – unlike standard NNs - the errors at each time step are summed during BPTT [18][19]. Strict

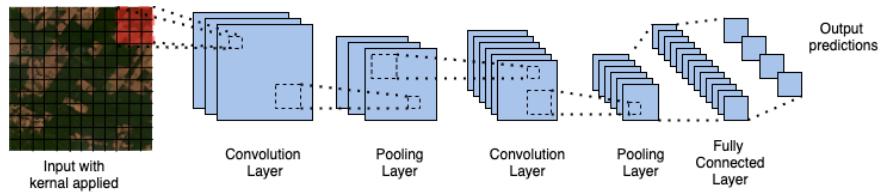


Figure 2.1: Convolution Neural Network with two convolution layers, two pooling layers and one fully connected layer

regularization is required in an RNN model due to vanishing gradients and exploding gradients, two problems that are likely to occur during BPTT. This is because it is not only the neurons that directly contributed to the output but also neurons from back in time that need updating, so the chances of the weights exponentially increasing (exploding) or decreasing (vanishing) becomes more likely [19][20]. Vanishing gradients is where the gradient becomes so small it might as well be 0. This is a problem as it means the network is no longer updating and adapting the hidden units to give better predictions during training [20]. Exploding gradients on the other hand, is where the gradient has become so large it might even be classed as *Nan* and the same problem to vanishing gradients occurs, where the network is no longer updating [21]. Exploding gradients can be solved by using a technique called ‘Gradient Clipping’. This technique involves setting a range of values with a maximum and a minimum that the gradient is not allowed to exceed and can even help to train the network faster [21]. In order to solve the vanishing gradient problem a new type of RNN was created called a Long Short Term Memory model. This network avoids the vanishing gradient by removing the weight from the BPTT equation. Instead a new input is stored and used in its place, this input is called a ‘forget gate’ [22].

2.2.3 Long Short Term Memory (LSTM)

The Long Short Term Memory (LSTM) model is a type of Recurrent Neural Network (RNN) and is the most widely used model for time series forecasting. As the input is temporal, it is sequentially processed at each hidden unit and LSTM cell as it passes through the layers [23]. The LSTM cell is made of three gates that together are what make the ‘memory’ component of the LSTM and separate it from other RNNs [23]:

- **Input Gate:** Decides what information should be updated in the current state.
- **Output Gate:** Responsible for deciding how much the current hidden state (that contains information relative to the previous inputs) should be changed at each time step.
- **Forget Gate:** Decides how much information from the previous cell is relevant and if it should be passed on to the current LSTM cell or should be ‘forgotten’.

2.2.4 Convolutional LSTM (ConvLSTM)

A Convolutional LSTM (ConvLSTM) model is a recurrent neural network for spatio-temporal prediction. To achieve this it uses recurrent layers from the LSTM (extracting temporal features) but instead of doing internal matrix multiplications at each layer a convolutional operation is used (extracting the spatial features)[24].

Study	Study Summary	Input Data Type	Spatial Resolution (metres per pixel)	Bands Used
The impact of deforestation on rainfall in Africa: a data-driven assessment	Feeding a ConvLSTM model climate and vegetation time-series data for regions in Africa to predict rainfall and simulate experiments for how geographical areas would react to certain rainfall and vegetation changes. The model was able to confidently show that deforestation was related to both local and remote rainfall.	Tree Cover	250	Spectrometer
		Leaf Area Index	500	Spectrometer
Spatial-Spectral Feature Extraction via Deep ConvLSTM Neural Networks for Hyperspectral Image Classification	To apply a ConvLSTM to 3D hyperspectral images to see if it is better at feature extraction and classification when compared to applying LSTM's to 2D versions of the same data. The LSTM was able to improve the quality of the chosen spatial-spectral features and in turn improve the classification performance.	Indian Pines Airborne Dataset	20	R, G, B, Infrared
		University of Pavia Dataset	1.3	103 bands in range 0.43 to 0.86µm
Framework for Deforestation Prediction Using Satellite Images and Neural Networks	Compares the predictions of different networks (including variations of a ConvLSTM and a GANS-LSTM) would create different predictions for deforestation that can be used to estimate total area loss. The model with the most ConvLSTM layers was found to have the best precision when predicting degradation and deforestation. It was found that combining a Conv2D layer after a ConvLSTM layer decreased the overall performance of the model.	Satellite Imagery	15	R, G, B

Table 2.2: Different applications of Convolutional LSTM's in satellite images, in order of appearance [9][25][8]

2.3 Short Wave Infrared (SWIR) Imaging

SWIR is a new band of electromagnetic radiation that was added as part of Landsat 8. It detects light between the ranges of $1.57\text{-}2.29\mu\text{m}$ as opposed to standard Near-Infrared (NIR) that detects light between the ranges of $0.85\mu\text{m}$ and $0.88\mu\text{m}$. Mid-wave and Long-wave infrared both use the light that is emitted from the object itself whereas short-wave infrared uses the light that is reflected from the object, just like visible light. The band of radiation that SWIR detects is the only range of the electromagnetic spectrum that can pass through clouds. Given stars and background radiance are constantly emitting SWIR, it is still effective at night time, allowing for 24 hour coverage [6].

2.4 Previous Work

Mapping and segmenting the different biomes in South America has been tried several times before. The techniques to do this range from using machine learning, as is similar to the work reported here, to more manual techniques where samples are taken from different sites across the continent. However, there is yet to be a platform which provides frequent mapping of the three largest biomes in South America.

- The closest existing tool to what is proposed in this paper has been developed by MapBiomas [7]. MapBiomas is a collaboration of NGOs, universities and technology companies that has been established to help study the environment within Brazil. They currently have a website that is updated yearly and is mainly focused on analysing the changes in land use throughout Brazil meaning the biome boundaries are not updated with every year, unlike the other features.
- The previous part of this project [10] was looking at how viable it would be to train a machine learning model to classify the different biomes in South America. It was found that by using simple RGB images alone a ResNet18 was able to classify the Amazon, Cerrado and Caatinga biomes with a macro F1 score of 0.843, showing that such a tool would have promising accuracy.
- SWIR 2 has been used by Beuno et. al. [6] to map the land use and seasonal changes of the Cerrado biome with great success. They believe the SWIR band performed so well given it has a higher absorption rate in water and vegetation when compared to other infrared bands.
- The Brazilian Institute of Geography and Statistics (IBGE) have been manually mapping biomes in Brazil since 2004 [2]. These maps currently provide the most accurate way to view the boundaries between the biomes that are at the center of this paper. However, these maps are created using manual samples so take a long time to produce. Since the first map was released to the public in 2004 it was not for another 15 years that an updated map was released.
- Techniques that use remote sensing to monitor land changes in biomes or ecosystems have been tried in places like the Amazon [26], Cerrado [26][27][28], and Caatinga [28]. The approaches and purposes vary but each study is looking at monitoring the temporal changes within images over a long period of time.

Chapter 3

Data Collection

3.1 Landsat 8 Image Collection

Landsat 8 was launched on February 11, 2013 and until 2021 it was the most recent Landsat satellite in orbit. It takes the satellite 16 days to capture every part of the planet meaning there are 22 images for a given area throughout the entire year. The satellite captures nine spectral bands, one more than its predecessor - Landsat 7. These bands include the visible wavelengths of light as well as the near-infrared, short-wave infrared (SWIR), panchromatic and cirrus bands. The main improvement of Landsat 8 over Landsat 7 was the inclusion of the SWIR band, which, as mentioned in [section 2.3](#), can pass through clouds. This means that regardless of how cloudy an image may appear to be in the visible light bands, there will still be one band that has successfully captured the ground surface. It is hoped that this will help to provide more detailed images for places such as the Amazon that have a high percentage of cloud cover throughout the year. The resolution of an image captured by Landsat 8 is 30 metres per pixel. Although this is only a medium resolution it is detailed enough to capture the patterns within the large biomes that are the focus of this project.

As a ConvLSTM requires temporal data, images had to be extracted for the same area over a sequential time period, unlike our previous project that was only making use of spatial models such as ResNet [10]. Therefore, images were captured from the year 2015 through to the year 2019. Another difference to the first part of our project arises here as seasonality was not factored into the extraction of images. Previously, it was found using NDVI that the difference between the biomes was most distinct during the summer season, and so for that reason only images taken during summer were used. However, it was also concluded that a model trained during only one season would not be effective for providing year round classifications of the different biomes and so although it may impact the overall accuracy of the model, it would be better generalised to different seasons.

A ResNet is also used in this project and so data from just one year was used as the input to this model. However, in order to make sure the ConvLSTM was not receiving four times the training data compared to the ResNet, additional images were extracted from 2019 so that the training set size of the ResNet would have the same number of

images in it as the training set for the ConvLSTM.

Each image downloaded should have a shape of 51×51 pixels, those that do not are discussed in [section 3.2](#). Given the resolution of the images is 30 meters per pixel, each image covers an area of 1.5km^2

3.2 Landsat 8 Image Preprocessing

Each image is downloaded from the Google Earth Engine (GEE) platform, which provides a comprehensive catalog of all Landsat 8 images ever taken. This dataset is free for academic use and so was openly embraced by this project. Each image obtained from GEE has been atmospherically corrected for surface reflectance by GEE, thus providing a clearer image. GEE apply the LaSRC algorithm that uses the metadata in each image such as “cloud masks”, that record how much aerosols are present, and “high aerosol”, which is the impact of the aerosols on surface reflectance, alongside external input data such as water vapor to calculate the surface reflectance of the image [29].

For each location a maximum of one image is extracted for each year. Given Landsat 8 records roughly 22 images of each location for every year, the image with the lowest percentage of cloud cover between the 1st of January and the 31st of December is chosen. However, if the percentage of cloud cover is less than 10% then no image is used for that location during that year. This decision was made as it is common for large sections of the Amazon to be under cloud cover for the majority of the year. If images with more than 10% cloud cover were included in the dataset then it is quite likely that these clouds would be used by the models as a feature of the Amazon. The amount of cloud cover present in an image is calculated by GEE and is held in the metadata of each image so a simple filter is applied to all images collected for that area.

No normalisation is applied after images have been downloaded from the Google Earth Engine. Instead, a BatchNorm layer is used in the machine learning models as manually normalising each image proved to be too computationally expensive given the large size of the dataset. BatchNorm will have the same effect, standardising the inputs to a network where needed to provide stable training and is explained in greater detail in [section 4.3](#).

Not all images downloaded from GEE share the same shape. To account for this and standardise the shape of each image an algorithm is applied that recursively appends or deletes, columns or rows of an image till it has the standard shape of 51×51 . When adding a new row or column to the image the mean value of the image is used to populate the new vector. This is to prevent the added vector from having any effect on the overall value of the image, an effect that would happen if a vector of zeros was added, for example. Finally, some images contain NaN values in places, the reason for this is unknown but could likely be the result of a sensor recording a value on its limit. To account for this rare occurrence each image has any NaN values that are present replaced by the mean value of the image.

Depending on the dataset being created, different bands within the image are required. For instance, as well as the spatial and spatial-temporal experiments (where images

are either taken from one year or from multiple years) there was also an experiment to see if SNDVI would obtain a higher accuracy than RGB data. Therefore, in total there were four datasets created, one SNDVI and one RGB dataset for each of the spatial and spatial-temporal experiments. To create the RGB dataset, only the second (Blue), third (Green), and fourth (Red) bands of each Landsat 8 image were used. To create the SNDVI dataset, the fourth (Red), fifth (NIR), sixth (SWIR 1) and seventh (SWIR 2) bands of each image are used. An example of the RGB and SNDVI equivalents of the same image is shown in [Figure 3.1](#).

In summary, preprocessing resulted in all the images having a consistent shape of 51×51 pixels and no image has more than 10% cloud cover. Normalisation was not applied to the dataset, but instead a BatchNorm layer would be used during training (objective 1 (a)).

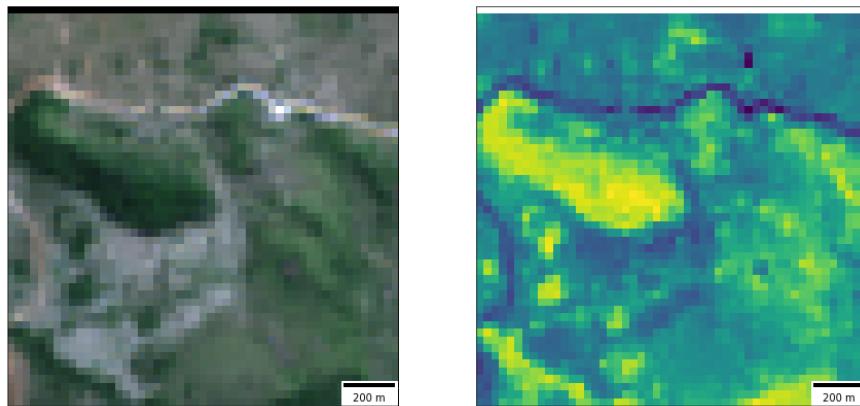


Figure 3.1: An image taken from the Cerrado biome during 2018 with the RGB version on the left (normalising for better viewing purposes) and the SNDVI version on the right.

3.3 Annotated Sources

Similarly to the first part of this project the maps produced by IBGE (as described in [section 2.4](#)) are used as the source for gathering ground truth data. However, unlike our previous project [10] that used the higher scale (1:1,000,000) map that was developed by MapBiomas [7] this project uses the map directly produced by IBGE and has a scale of 1:250,000, making it four times more accurate than that produced by MapBiomas. Additionally, the updated map from IBGE was published in 2019 making it more relevant to the time period that images used in this project are extracted from. The shapefiles for each of the biomes were obtained from ArcGIS. From [Figure 4.2](#) it can be seen that the new biome boundaries are no longer separated by a gap. This gap was previously a result of the MapBiomas shapefile having the biomes overlapping in sections, however, the new IBGE map does not suffer the same drawback and so the boundary of one biome touches that of the other biomes.

Chapter 4

Methodology

4.1 Train/Test Split

Two different training and test sets had to be created. One that captured temporal data and so spanned multiple years but as a consequence covered less area, and another which was used for spatial models so data was only extracted from one year, but, as a result covered more ground.

For the temporal dataset, 10,913 unique locations were chosen across all three biomes. Five images were taken for each location, one from each of the years 2015 through to 2019, inclusive. The training set was then created by leaving out the year 2019 as well as one quadrant from each of the three biomes. This meant that as well as having a spatial separation there was also a temporal separation between the biomes, minimising any possible data leakage from influencing the training and evaluation. For the spatial dataset, 35,340 images of unique locations were taken between the 1st of January and 31st of December for the year 2019. This data was then split spatially by removing one quadrant from each of the biomes to create a testing set and training set. The ratio of training set size to testing set is roughly 70/30 for both the spatial and temporal sets.

By using the more accurate IBGE map, a greater area for each biome is covered, providing a larger breadth of differing terrains to be associated with each biome. [Figure 4.1](#) depicts how each biome has grown in size by roughly 1.17 times. It is also visible from this figure that the Amazon biome covers an area just over five times greater than the smallest biome (Caatinga). Despite this greater surface area the same number of images have been taken from each biome to avoid the Amazon biome having a bias during training over the smaller biomes. By taking images from randomly selected locations in each biome a uniform coverage is achieved, ensuring that despite the same number of images being taken from the Caatinga biome and the larger Amazon biome, a representative depiction of each is still captured in the training data.

The number of images in each biome are roughly equal for both the temporal dataset and the spatial dataset, as can be seen in [Figure 4.3](#). This figure also shows that both datasets roughly share an equal number of images, preventing training set size from having an influence during training of the machine learning models. The number of

images are also evenly distributed across all quadrants, in spite of the Amazon having a larger percentage of cloud cover throughout the year, a quality that led to there being less images in the Amazon biome during the first part of this project [10].

In summary, a dataset had been created with a 70/30 split into training and test sets, with each biome being given equal representation to avoid bias during training or evaluation (objective 1 (b)).

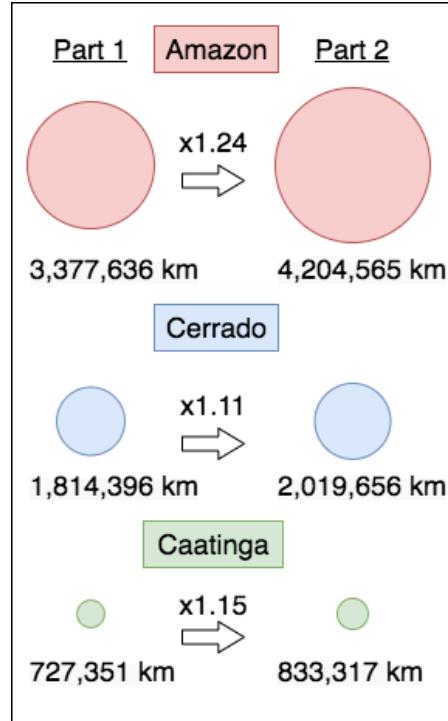


Figure 4.1: Comparison of biome area and how it has increased from our previous project [10] that used the MapBiomas shapefiles, to this project, that uses the more accurate 2019 IBGE shapefiles.

4.2 Bayesian Optimization

A hyperparameter is a parameter in the neural network that is set prior to training, and effects how the network learns. Hyperparameter tuning has previously been achieved by manually searching through a reduced set of possible combinations. This was both time consuming and often did not cover all combinations, meaning there was a high likelihood that the best set of hyperparameters was completely missed. To automate the process of searching through different hyperparameters Bayesian Optimisation creates a probabilistic model of the learning algorithm by using a Gaussian Process [30].

The Gaussian process defines a mean and covariance of the objective function (the learning algorithm). The mean and covariance are then fed into an acquisition function which will select the next set of hyperparameters such that they have the highest probability of improving the model [30]. The results of these parameters are then fed

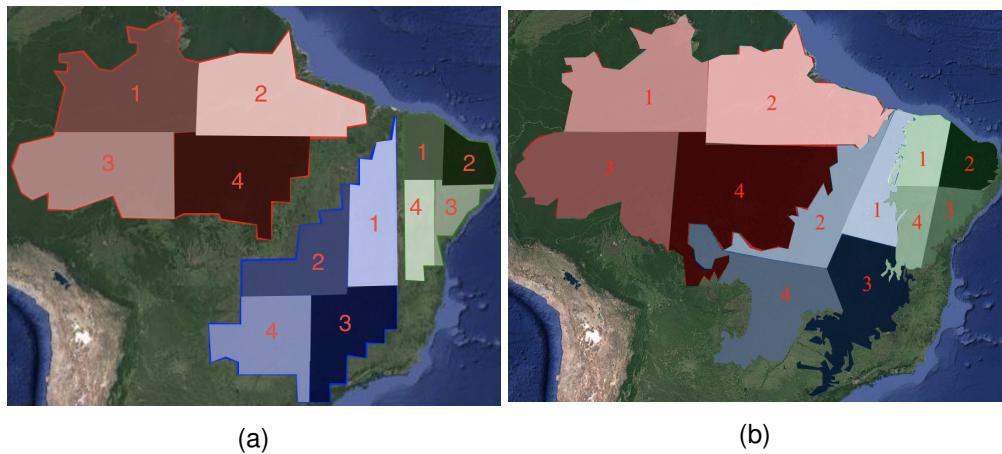


Figure 4.2: (a) The quadrants used in our previous project [10], based off the 2004 IBGE and MapBiomes shapefiles, (b) The biomes and quadrants used in this project based on the 2019 IBGE map. Each polygon represents the area of each class/biome. Red background is Amazon, blue background is Cerrado and the green background is Caatinga (with the label of each quadrant overlaid and labelled in red).

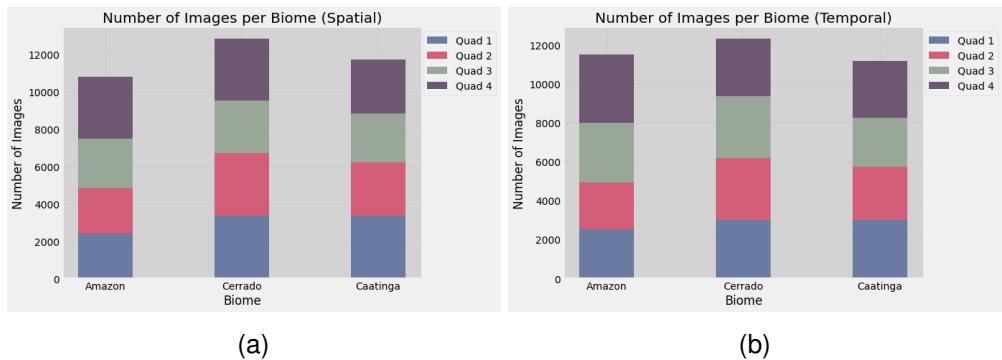


Figure 4.3: The displacement of images in each quadrant for each biome in both the (a) spatial dataset and (b) temporal dataset.

back into the Gaussian Process which adjusts its mean and covariance, and will repeat the process until the mean and covariance converge.

The acquisition function used in this report is Expected Improvement (EI). EI will determine what the next best parameters are by selecting the set that have the highest expected improvement over the current best combination [30]. This set of hyperparameters will either be selected from the model of the objective function that has the highest uncertainty or where the mean of the objective function is greater than the current best result.

In this report Bayesian Optimization evaluated the performance of each model by taking the mean Area Under the Receiver Operating Characteristic Curve (AUC ROC) after 2-fold cross validation. AUC ROC is the distribution of the true positive rate against the false positive rate. A good model will have a high true positive rate and a low false positive rate which will lead to an AUC of 1. A score closer to 0 indicates that the model is failing to accurately predict the majority of each class.

4.3 Batch Normalisation

During training on large datasets, the inputs to the network are often processed sequentially, in what are called ‘batches’. During the training of a batch, weights are adjusted so that the model will be closer to the target value during the next iteration. However, if the inputs are not normalised then the inputs in the next batch will be different to the previous batch and so the network will be constantly chasing a moving target, a phenomenon called internal covariate shift [31].

Batch normalisation applies normalisation for each training batch, giving a standardised input to the network and creating a more stable learning process. This not only means that the initial weights matter less but also allows for higher learning rates to be used (speeding up the training process) given the network is now more stable [31].

4.4 SNDVI

Given the progress made in our previous project there are two baselines coming into this project that are hoping to be improved upon. The first is the introduction of temporal data, given land use changes with time it is hypothesised that by taking images of the same place from multiple moments in time, trends will be detected that are unique to each biome and this will improve the current classification performance. The second baseline is the introduction of the short wave infrared band that Landsat 8 is now able to capture - SWIR.

SWIR can pass through clouds, and so regardless of how much cloud cover may be present in the image, a complete representation of the ground below will be captured in this band. Given the high humidity of the Amazon, it is often shrouded in clouds for the majority of the year. The results from our previous project found that one of the main features the ResNet18 was using to classify if an image belonged to the Amazon or not, was the presence of clouds [10]. This is obviously not a desired outcome given that clouds can exist anywhere in the world and are not a unique feature to the Amazon. By completely disregarding the clouds from an image it is hypothesised that this will force any machine learning model to focus purely on extracting features from the image that represent the Amazon.

A combination of SWIR and the Normalized Difference Vegetation Index (NDVI) has recently been found to be more effective at separating the different biomes when compared to NDVI being used by itself [32]. NDVI was shown to be effective at separating the biomes in our previous project [10] so it is hoped that the added level of detail will further improve upon this performance. Kumar et. al. [32] claim that SWIR’s ability to measure moisture levels allows for the internal structure of the leaves to be analysed, such as the moisture or water content, that NDVI fails to record on its own. To integrate the SWIR band into the NDVI formula Kumar et. al. use a combination of the SWIR, NIR, and Red bands as shown in [Equation 4.1](#). Given the Amazon is a rainforest it is predicted that there will be a higher moisture content in the leaves of the trees when compared to the leaves in the drier Cerrado, and Caatinga biomes. Therefore, it is hoped SNDVI will find a greater level of separation between the three biomes.

$$SWIR = \frac{(SWIR + NIR - RED)}{(SWIR + NIR + RED)} \quad (4.1)$$

4.5 ConvLSTM

The ConvLSTM was implemented using the deep learning python library, Keras [33]. This library allows for many different types of neural network layers to be combined together without the cumbersome task of personally needing to build the architecture behind each layer. This meant that many different architectures could be tried and tested with a quick turn-around. The different architectures included changing the types of layers, the hyperparameters of the layers, the loss function used or the optimiser used.

The optimiser is responsible for allowing the network to adjust the values of the weights such that the difference between the predicted output and the actual output is minimised. Typical approaches for this use Stochastic Gradient Descent (SGD). SGD attempts to minimize the error in the network by randomly selecting one weight after each iteration and calculating its derivative. If the gradient is negative then the model is learning and getting closer to the best possible arrangement of the weights while if the gradient is positive then the model is moving further away from the best possible arrangement [34]. The optimiser used in junction with the ConvLSTM was the Adam optimiser. Adam builds upon SGD by changing the learning rate during training. Making the learning rate dynamic allows the model to train faster at the beginning where large changes to the weights can help close the gap to the best arrangement of these weights. Once nearing the best arrangement a smaller learning rate then allows for precise adjustments to each weight, minimising the error as much as possible [35].

The error and loss that is used by the optimiser to change the values of the weights is calculated using categorical crossentropy. The ‘categorical’ part of the loss functions name indicates that each input can only belong to one output class. The loss is calculated by multiplying the log of the probability that the input belongs to one of the output classes (y_i where i the current class) by the ground truth target class probability (t_i which will be 1 or 0). This value is summed for all classes, indicating how close the model is to accurately predicting the data, [Equation 4.2](#).

$$Loss = - \sum_{i=1} t_i * \log y_i \quad (4.2)$$

This loss function is teamed up with the Softmax activation function, which outputs the probability of each input belonging to each of the possible classes.

The hyperparameter tuning for the ConvLSTM consisted of applying Bayesian Optimization to three independent variables, (1) the number of ConvLSTM layers in the network, (2) the amount of hidden units in each ConvLSTM layer, and, (3) the dropout percentage that is applied after each ConvLSTM layer. The best model resulted in having 3 ConvLSTM layers, each with 128 hidden units and a dropout of 30% after each ConvLSTM layer. A visualisation of the final architecture is provided in [Figure 4.4](#).

4.6 ResNet

The ResNet model was also implemented using Keras, although additional ResNet blocks had to be created as they were not included as part of the Keras library. Similarly to the ConvLSTM, different numbers of layers and dropout percentages were tested alongside optimisers and the inclusion of pooling layers.

The responsibility of the pooling layer in a convolutional neural network is to reduce the amount of information about the input that is passed through the network (an operation called downsampling). They are usually applied after a convolutional layer so that only the features of the input that are deemed most important by the convolutional layer are moved on to the next layer. Pooling layers achieve this downsampling by applying a smaller matrix mask over the output from the convolutional layer, similar to that of the kernel that was mentioned in subsection 2.2.1 [16]. There are different types of pooling that can be applied, average pooling and max pooling. In average pooling the average value of the cells underneath the matrix is stored moving forward, meanwhile with max pooling it is the maximum of the values under the matrix that is kept [16]. Given both of these pooling layers are applying different operations to the inputs the only way to tell which one is most effective is by trying them both and selecting the one that performs the best, which in the case of this dataset was the max pooling layer after the first convolutional layer and an average pooling layer after the second convolutional layer.

In the previous project [10] no activation function was included and instead the embedded representation of the input vectors were fed into a multinomial linear regression model that would classify the outputs. However, this time a softmax layer was added as the final layer, removing the need for the additional classifier as little benefit was seen from its' presence.

4.7 Dropout

It has been mentioned in both the methodology behind the ConvLSTM (section 4.5) and the ResNet (section 4.6) that dropout was included. It was also found during tweaking of hyperparameters that the inclusion of dropout gave the best performance for the ConvLSTM and the ResNet.

Dropout is a regularization technique that prevents a model from overfitting to the training data by randomly removing certain units from each iteration of training [36]. This forces the model to try and learn from the data with a different representation of the inputs each time. It has been seen that by removing random units from each layer the layers output becomes sparse, and in turn forces the model into learning a sparse representation of the data [36]. The dropout percentage is what has been tweaked during hyperparameter tuning and can be expressed as the percentage of units that are randomly removed from the previous layer during each training cycle. Of course the dropout layer doesn't actually delete hidden units from the previous layer but instead simply blocks the output from the previous layer from getting to the next. Thus, a dropout layer with a dropout percentage of 20% after a ResNet layer will randomly

block 20% of the hidden units outputs by setting them to zero before they reach the next layer.

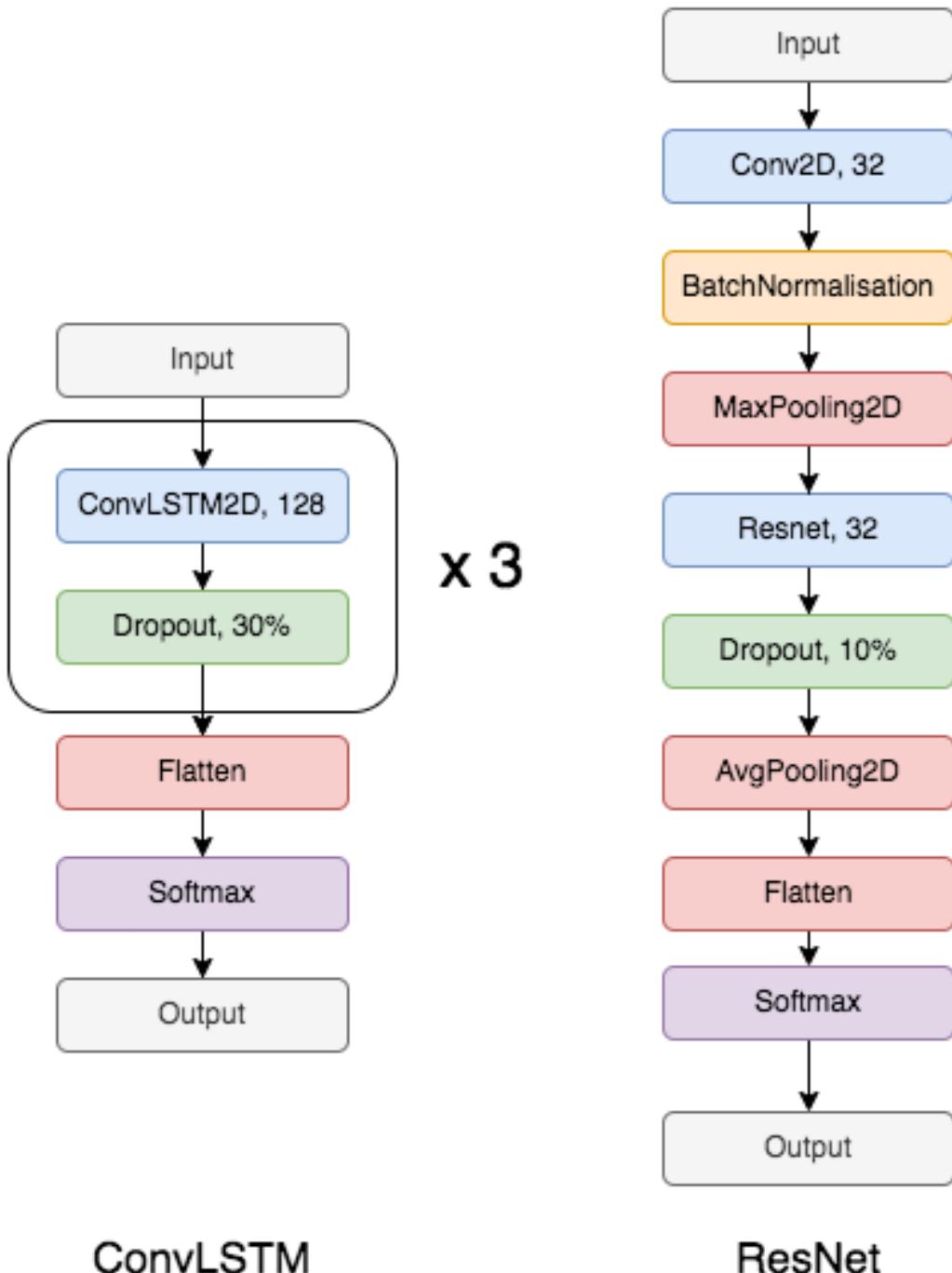


Figure 4.4: The architectures of the best performing models, with the ConvLSTM on the left and the ResNet on the right

Chapter 5

Experiments

The following chapter documents how each experiment was conducted in order to meet objectives 2 (a), (b), and (c). The first experiment establishes the superior method of hyperparameter tuning and the second experiment compares whether it is most effective to use the SNDVI indicator or the visible light bands for each input image. The last experiment compares the two different models, ResNet and ConvLSTM, while using the most effective input that was determined by the previous experiment.

5.1 Bayesian Optimization vs Manual Search

Aim: Does Bayesian Optimization improve performance over manual grid search?

5.1.1 Approach

Rather than blindly assume that Bayesian Optimization would find the best set of learning parameters for the chosen model, the same, manual approach that was applied in the previous part of this project was used as a benchmark [10].

This manual approach is similar to that of a grid search, only not all combinations are tried. Instead, the parameters are rotated through one by one with the best performing value for each parameter being used moving forward. When one parameter is being evaluated, all other parameters remain constant. This means that if there were three hyperparameters that were being evaluated, each of which could take on three different values then there would be a total of 27 ($3 \times 3 \times 3$) combinations, but by testing them individually before moving forward there are only 9 ($3 + 3 + 3$) combinations. However, this can be reduced even further by evaluating each hyperparameter ordinally, where if the performance of the model begins to decrease as the value of the hyperparameter increases then evaluation of that hyperparameter will cease.

Of course the manual approach is limited and does not cover all possible combinations, and as a result it is likely to miss the best combination, but it does take a fraction of the time when compared to an exhaustive search. By using Bayesian Optimization it is

predicted that the model will be able to cover a larger search space in similar time and find a better arrangement of hyperparameters.

Bayesian Optimization was applied to a ResNet model, and the hyperparameters in question were the number of ResNet layers (either 1, 2 or 3), the number of hidden units each hidden layer had (where the number of hidden units in each layer was exactly double the number of the previous layer) and the dropout percentage (10%, 20% or 30%). The model was both trained and evaluated on the RGB dataset with a cross validation of 2 where AUC ROC was used for evaluating the performance.

5.2 SNDVI vs RGB using a ConvLSTM

Aim: To see if using SNDVI data provides a better accuracy over the RGB equivalent for the same location.

5.2.1 Models and Data Used

The training set consisted of 6,528 unique locations. As each location has four images representing it, one from each of the years 2015, 2016, 2017 and 2018 there was a total of 20,335 images in the training set. A validation set was also created, unlike the testing set that is spatially and temporally separated from the training set, the validation set is simply a random percentage of the original training dataset that was created using the scikit-learn *train_test_split* function [37]. This provided a validation set size of 5,176 unique images, that were used in between each epoch of training to determine if the model was overfitting, and a final test size of 9,063 images. If the model's performance failed to improve on the validation set after 4 epochs of training then it was deemed sufficient evidence the model was now overfitting and so training would be cancelled with the version of the model from 4 epochs prior being returned. This form of early stopping helps to prevent the model from training more than is necessary.

Two different datasets were used - both identical sizes to what was just described. One consisted of SNDVI data, the seventh band in a Landsat 8 image, and the other consisted of RGB images - the 2nd, 3rd, and, 4th bands in a Landsat 8 image. The same architecture of ConvLSTM was applied to both datasets and is described in [section 4.5](#).

5.3 Spatial vs Spatial-Temporal

Aim: To see if there is a benefit to including the temporal changes a biome has over exclusively analysing the spatial difference between biomes.

5.3.1 Models and Data Used

Two different models were compared. One was a ConvLSTM, a spatial-temporal model, and the other a ResNet, a temporal model. The ConvLSTM is identical to the one used in the previous experiment ([section 5.2](#)), containing three ConvLSTM layers with 32 hidden units and a dropout of 30% after each layer. The ResNet used the

hyperparameters that were found by Bayesian Search, one ResNet layer with 32 hidden units, follow by a dropout layer with a dropout of 10%. Both models were trained using early stopping with a batch size of 100.

Given the input data to each model is different - one model requiring images of the same location over multiple time periods and the other needing only an image of an area from one time period - two different datasets were used. Although the models are both trained and evaluated on different datasets the data within each dataset remains the same, Landsat 8 images of locations throughout South America. On top of this, both datasets share similar size, and ratio of classes, as is shown in [Figure 4.3](#). The Landsat 8 images only contain the RGB bands as the SNDVI band was shown not to achieve a high accuracy in the previous experiment ([section 5.2](#)).

The training set size for the ResNet is 20,782 images with a validation set size of 5,196 images and test set size of 7,267 images, while the training set size for the ConvLSTM is 20,355 images, a validation set size of 5,176 images and a test set size of 9,063 images.

Chapter 6

Results

6.1 Bayesian Optimization vs Manual Search

The Bayesian Optimization found that the best hyperparameters for the ResNet were 1 ResNet layer, 32 hidden units and a dropout percentage of 10%. This combination of hyperparameters scored a mean AUC ROC over the 2 fold cross validation of 97.4% ([Table 6.1](#)). This combination was not included in the manual search as 2 ResNet layers were found to be better than 1 and so subsequent parameter tuning only used 2 ResNet layers.

The next best hyperparameter combination found by the Bayesian search was 2 ResNet layers with 32 and 64 hidden units and a dropout of 10%, which is the same as the best values found by the manual search, achieving an AUC ROC of 97.3%, just 0.1% less than the best combination found by the Bayesian search, [Table 6.1](#).

Moving forward Bayesian Optimization was used as it was found that the manual search can miss the most optimal parameters (objective 2 (a)).

Model ID	Layers	Hidden Units	Dropout Percentage	AUC ROC
Manual Search	2	32, 64	0.1	0.973
Bayesian Optimization	1	32	0.1	0.974

Table 6.1: Performance of models on test set using best hyperparameters found by Bayesian optimization and a manual search.

6.2 SNDVI vs RGB using ConvLSTM

Given the much lower accuracy of SNDVI when compared to RGB ([Table 6.2](#)), it is highly likely the SNDVI band does not have sufficient separation between the different biomes, to allow distinct decision boundaries between each of the classes to be made. A dummy classifier was included to act as a baseline that the SNDVI and RGB data could be compared against, allowing an informed decision to be made in regards to how much of an influence both data types had over the models performance.

Macro metrics were used to evaluate the performance of each of the different features used as the classes do not share an exact equal number of data points, as can be seen from [Figure 4.3](#). Macro average accounts for this by treating the score of the model on each individual class equally, regardless if one class made up 90% of the dataset and the other 10%.

Therefore, SNDVI was not found to be a good indicator, with the ConvLSTM performing 29.8% more accurate on the RGB bands (objective 2 (b)).

	Dummy	ConvLSTM
Dataset	RGB	SNDVI
Accuracy	0.426	0.543
Macro Precision	0.142	0.645
Macro Recall	0.333	0.567
Macro F1 Score	0.199	0.554
		0.841
	<td>0.842</td>	0.842
		0.833
		0.825

Table 6.2: Performance of a Dummy classifier and a ConvLSTM on RGB and SNDVI datasets

6.3 Spatial vs Spatial-Temporal

The performance of both the spatial ResNet and the spatial-temporal ConvLSTM was evaluated on the test set using the RGB equivalent for both types of input data. The early stopping criterion was activated for both models, meaning neither ran for the full set of epochs. The ConvLSTM was cut short after 8 epochs and the ResNet after 26. The learning rate remained constant throughout training of both models at 0.001 and the loss that was evaluated using categorical crossentropy reached an all time low of 0.249 for the ResNet and 0.377 for the ConvLSTM.

The F1 score is the main point of comparison between the two models as it is the harmonic mean of both the precision and recall metrics. This therefore means it takes into account both the ratio of how many relevant images were correctly classified out of all images classified as that class, and the ratio of how many images were correctly classified as their class out of all images belonging to that class. The accuracy metric, although useful does not give a fair representation of the networks true performance. This is because the class distribution is slightly imbalanced and accuracy does not account for this, it is simply how many images were correctly classified over the total number of images.

There was no benefit seen from using the spatial-temporal ConvLSTM, as the ResNet achieved a 9.5% better macro F1 score (objective 2 (c)).

	Accuracy	Macro F1 Score	Macro Precision	Macro Recall
ConvLSTM	0.841	0.825	0.842	0.833
ResNet	0.920	0.920	0.921	0.920
ResNet18 + MLR	0.851	0.843	0.834	0.873

Table 6.3: Performance of the ResNet ConvLSTM on the RGB test set. The ResNet18 and Multinomial Linear Regression model combination that achieved the best performance from our previous project is also included for reference [10].

Chapter 7

Exploratory Analysis

7.1 Mapping Capability

The main aim of this project is to show it is possible to create maps of the three largest biomes in South America using machine learning and satellite imagery. In order to evaluate the capability of the best performing ResNet at achieving this goal, a $45 \times 45\text{km}$ patch inside the Amazon was selected. This area was chosen as it contained multiple different land types, such as forested areas, deforested areas and farm land that is often found in the Caatinga biome. Roughly 121, $1.5 \times 1.5\text{km}$ images were extracted from 2016 in a grid so as to ensure the maximum coverage of the area. The images were then fed through a classifier and the results can be seen in [Figure 7.1](#).

It is expected that a biome will gradually change from one to the other and so to account for this grace period a new “Inconclusive” output class was included. Given the final layer in the ResNet model is a Softmax layer, the output is a probability distribution of each input image belonging to each possible class. If the class with the highest likelihood of the input belonging to it is below a certain threshold then the image will be labelled as the inconclusive class. The chosen threshold for this exploratory analysis was based on our previous project which found 80% to be an appropriate threshold [10]. For example, if an input image was output with a probability distribution of [0.6, 0.3, 0.1] then there is a 60% likelihood the image belongs to the Amazon, a 30% likelihood the image belongs to the Cerrado, and a 10% likelihood it belongs to the Caatinga biome. Yet, seeing as no biome has a probability greater than the threshold of 80%, the output class will be Inconclusive.

In the classifications for 2016 [Figure 7.1](#) (a) it appears, as expected, the majority of images are being classed as the Amazon - the correct biome for this area according to IBGE’s 2019 map. However, a few images have been classed as either Caatinga or Inconclusive. Upon inspection of these images ([Figure B.1](#)) those that are classed as Caatinga all contain fields, a very different landscape to the images that have been classed as the Amazon where the majority of the image is trees. Meanwhile images classed as Inconclusive do not resemble the images classed as the Amazon but share a closer resemblance with images classed as Caatinga. When compared to MapBiomas’ 2016 deforestation map, one of the images classed as Caatinga falls precisely in an area

of deforestation.

This shows that land use changes are being detected by the ResNet model and by producing maps at a faster rate it is possible to detect these changes (objective 3 (a)).

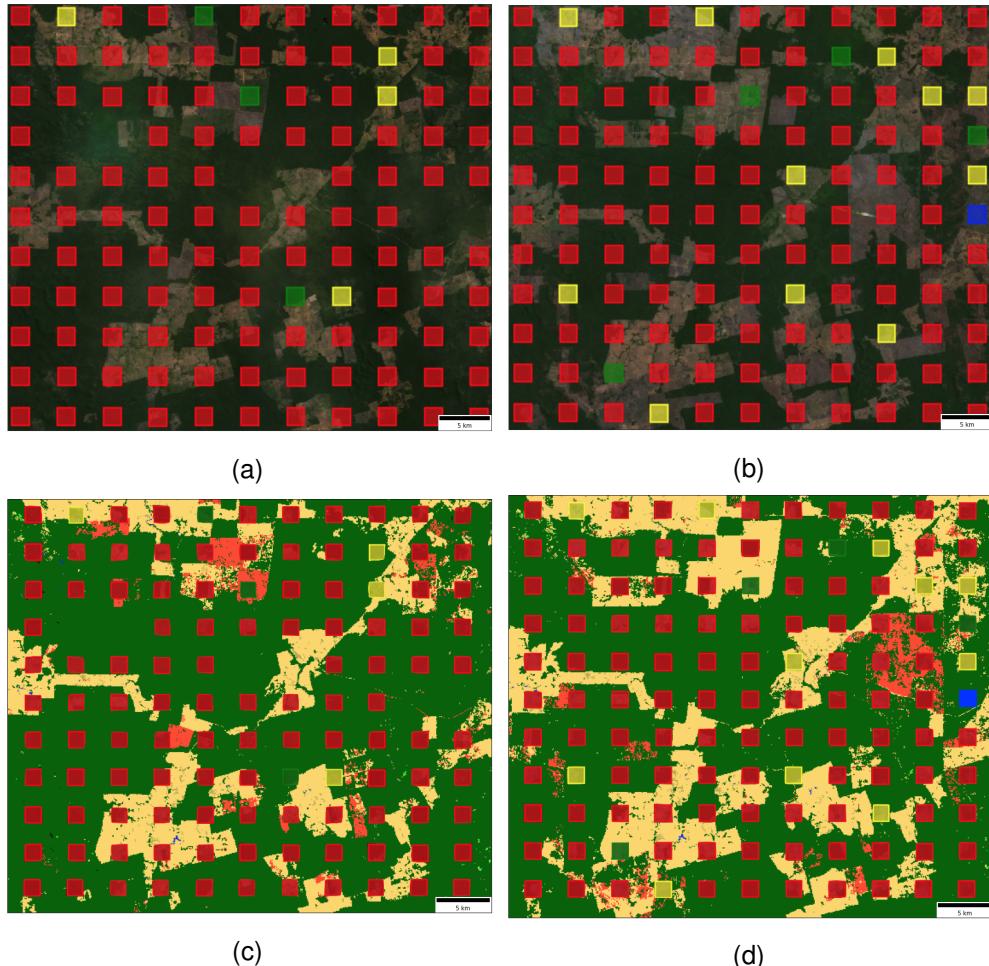


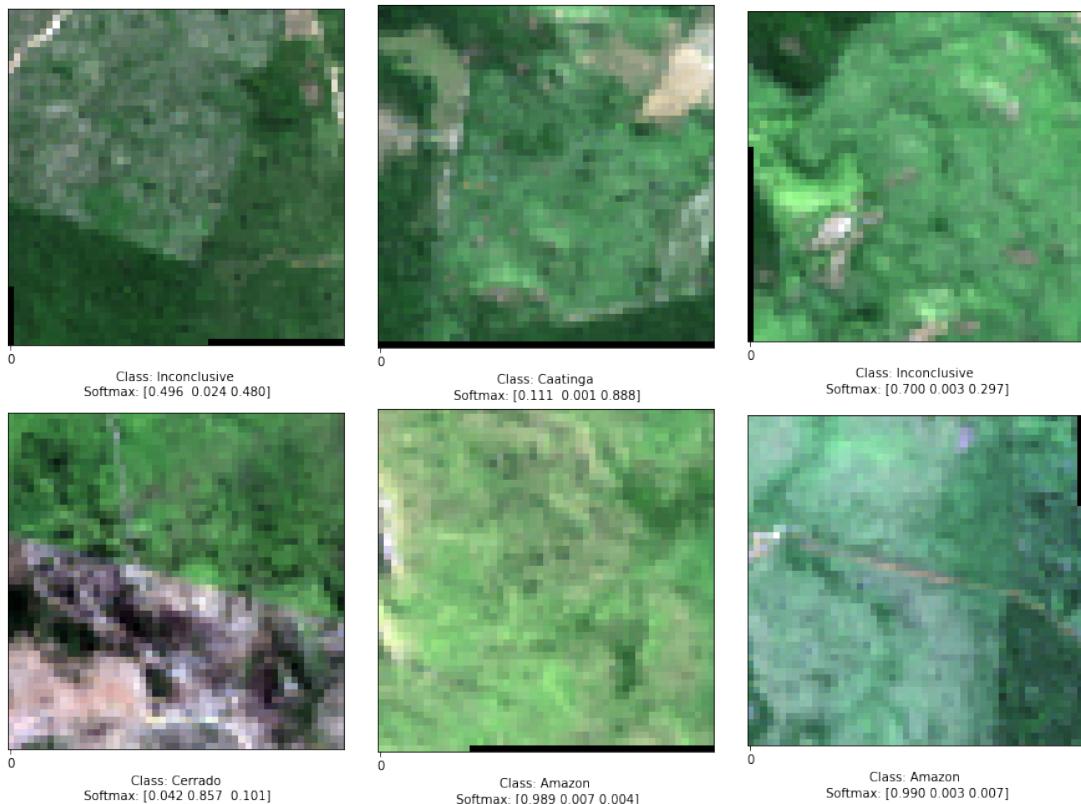
Figure 7.1: ResNet classifications for 121 images taken from a 45km^2 area in (a) 2016 and (b) 2020. Images classed as belonging to the Amazon are red, Cerrado are blue, Caatinga are green and the Inconclusive class as yellow. The MapBiomas deforestation maps for, (c) 2016, and (d) 2019 have also been included with the image classifications overlaid on top. These images were taken directly from the MapBiomas toolkit with areas where their machine learning algorithms have detected deforestation have been included in red, forested areas in green, and non-forested areas in yellow [7].

7.2 Frequency of Change

Once it was established the model was capable of producing maps it was necessary to see if producing maps at rate greater than that currently in place by IBGE (every 15 years) would have any benefit. A further 121, $1.5 \times 1.5\text{km}$ images were extracted from the same $45 \times 45\text{km}$ area inside the Amazon biome used in section 7.1, only this time they were taken from 2020.

From [Figure 7.1](#) it can be seen that the land has undergone further deforestation (based on the most recently available MapBiomass deforestation map that was released in 2019). In these four years 7.8% of the images changed from being previously classed as the Amazon or Caatinga to now being classed as Inconclusive. Upon inspection these are all areas where deforestation has taken place. By using the MapBiomass deforestation map as reference, the high number of images no longer being classed on the right hand side could be a result of land use changes. However, upon closer inspection, the contents of these three images appeared to resemble those most often found in the Amazon, [Figure 7.2](#). The Softmax has been included below these images and it should be noted that the classification of Caatinga (top center of [Figure 7.2](#)) is rather high. It is also worth noting that in areas where it appears there is no longer forest, certain images are still being classed as the Amazon. Upon closer inspection of these tiles (the bottom middle and right hand images in [Figure 7.2](#)) it is obvious that trees do in fact appear in this area.

The increased rate of map production allows for land use changes, such as deforestation, to be documented that current mapping techniques miss (objective 3 (b)).



[Figure 7.2](#): Images that have been classed as Inconclusive, Caatinga, Cerrado and the Amazon from the 2020 mapping capability analysis in [Figure 7.1](#). Their softmax output has also been included.

7.3 Detecting Change on a Large Scale

The model has been shown to effectively and accurately identify changes to land use at a small scale for the most part. It is still unclear if the best performing ResNet will be useful at easily distinguishing large scale changes in a biome over time. To determine whether it was capable of this or not, 1000 images from a large area (roughly 100km^2) were randomly extracted from both 2016 and 2021 and passed through the same classification process as the images taken from the smaller 100km^2 patch.

The area in [Figure 7.3](#) that is highlighted by the white circle indicates a large section of land that has undergone noticeable deforestation (a figure highlighting just the deforestation can be found in the [Appendix B.2](#). The ResNet model has identified this change, as a large proportion of the images taken from within this circle have changed class from Amazon in 2016 to the three other possible classes in 2021.

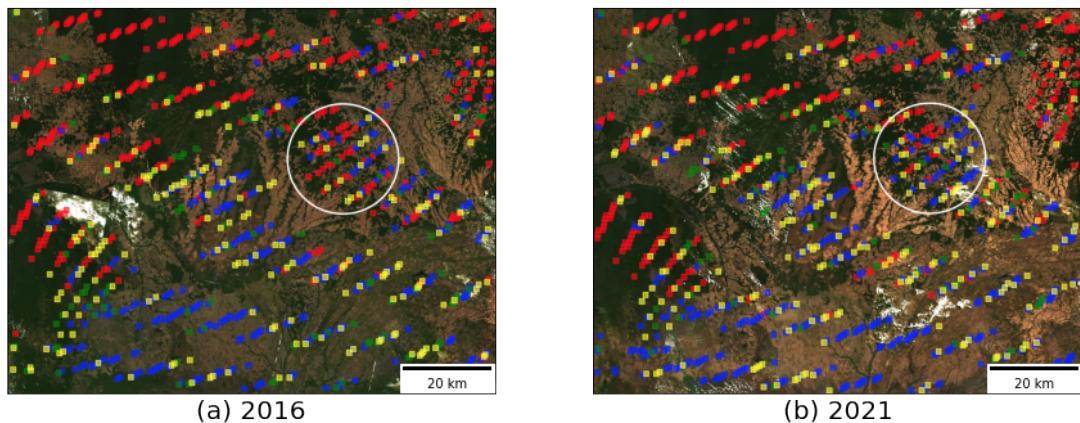


Figure 7.3: ResNet classifications for 1000 images taken from a 100km^2 area in (a) 2016 and (b) 2021. Images classed as belonging to the Amazon are red, Cerrado are blue, Caatinga are green and the Inconclusive class as yellow. An area of significant change has been highlighted in a white circle.

7.4 Epistemic Uncertainty

It is possible that the classification of the Caatinga biome in [Figure 7.2](#) is a false positive. The assumption behind this being a false positive is that the image does not resemble the typical Caatinga eco-systems, a few examples of which are shown in [Figure B.1](#).

Epistemic uncertainty is the error coming from the inadequacy of the model on the input data [38]. I.e., how confident is the model that this input belongs to the predicted class. It can be calculated by taking the output from a model where the penultimate layer is a Dropout layer that is kept active during inference. According to Yarin and Zoubin including Dropout during inference results in many different subsets of the model creating their own output from the given input, and as a result is mathematically equivalent to Bayesian Inference [39]. The mean and standard deviation of the output Softmax over x amount of predictions can then be likened to how certain the model is of its decision.

The Softmax predictions for the image classed as Caatinga were evaluated 1000 times and the mean value for the image being classed as the Amazon was $14.74\% \pm 9.50\%$ while the mean value for the image being classed as Caatinga was $85.19\% \pm 9.49\%$. From these results it appears that the model is not confident in the classification of Caatinga given that the standard deviation of the Softmax percentage extends below the 80% confidence threshold that was set.

Therefore, a better way of evaluating the confidence threshold is required, in order to catch cases like this where the model is uncertain with the outcome (objective 3 (c)).

7.5 Determining Classification

In order to better understand what features in the input images influence the models output classification three images that were previously explored as part of [section 7.1](#) and [section 7.2](#) were selected for further analysis. These images were chosen given they had multiple different land uses inside the frame, hopefully making it easier to highlight which land types were being used as part of each classification.

Neural Networks are often black boxes, where an input is passed into the network and at the other end a classification is output. What goes on in-between these two steps is usually far too complicated for a human to interpret and understand. In order to better understand the judgements that a network makes as part of its classification a technique called Gradient Weighted Class Activation Mapping (Grad-CAM) has been proposed with great success [11]. This technique uses the gradients that are passed into the final convolutional layer (although in this case it is a ResNet layer) in order to find out what regions of the input image are being heavily weighted - and thus are the most important. The final convolutional layer is used as it is the layer with the highest level of both semantic and spatial information that the network has extracted from the input image. With this information a heatmap can then be overlaid on the input image, indicating which regions of the input image are the most important (those in red, i.e., the hottest areas) and which regions are used the least (those in blue, i.e., the coldest areas). These heatmaps are unique to the class being analysed, as the network will use different parts of the image in order to make its judgement on what class the input image belongs to.

The heatmap that this method overlays onto the input image can be seen in [Figure 7.4](#). The Grad-CAM for each of the classes has been placed horizontally to the input tif image with the classification of the input tif image placed as the label for each row. For the Amazon Grad-CAM, in each of the three input images it is mostly focusing on areas where there is an absence of trees. This is the inverse of what was previously assumed to be the main feature associated with the Amazon, i.e., instead of focusing on how many trees are in the image it is focusing on how little trees there are in the image. Either way the outcome appears to be the same with images that have a majority concentration of trees being classed as the Amazon. The Grad-CAM for the Cerrado biome seems to focus on areas that both the Amazon and Caatinga Grad-CAM's also focus on, only the Cerrado Grad-CAM is focusing with a higher intensity. It is unclear exactly what land types the Cerrado biome is looking for but it could possibly be looking for a balance of several land types that on their own resemble the Amazon or Caatinga.

In the Grad-CAMs for the final biome, Caatinga, the hottest parts of the images are focused on what appears to either be darker patches of vegetation or the lightest part of the image. This also makes sense, given that the Caatinga biome consists mostly of either a xeric shrubland eco-system (that appears bright in satellite imagery) or thorn forests, which could possibly be the darker patch of vegetation on the right hand side of the image that has been classed as Caatinga.

The use of Grad-CAM has highlighted that the absence of trees influence Amazon predictions, the presence of xeric shrublands and thorn forests influence Caatinga predictions, yet it is still unclear what determines a Cerrado prediction (objective 3 (d)).

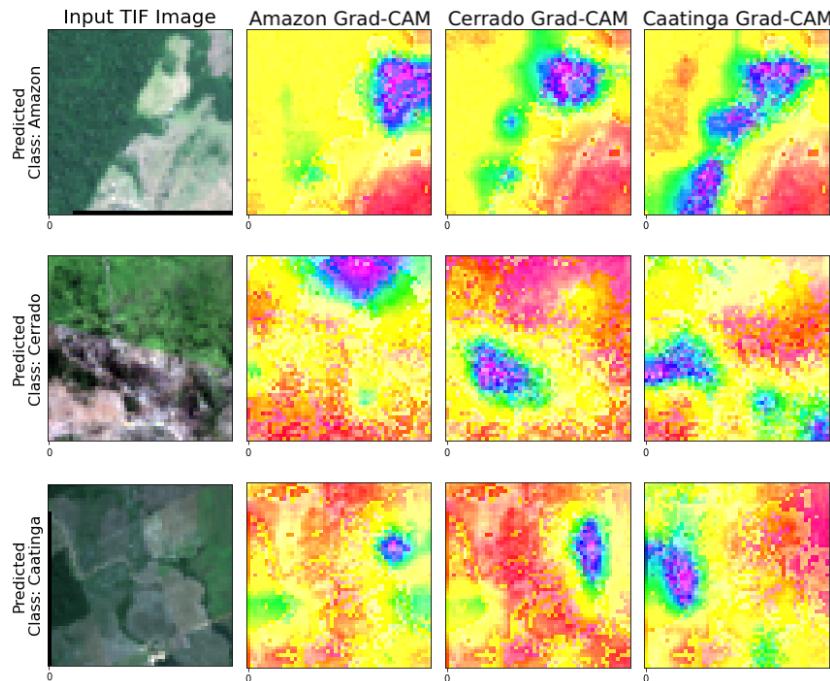


Figure 7.4: Grad-CAM being applied to three input images, each of which was classed as a different biome. Areas in red highlight the parts of the input image that were used the most during classification whilst areas highlighted in blue highlight the areas that were used the least during classification.

Chapter 8

Discussion

The ResNet model trained on the RGB dataset produced by far the most accurate classifications on the test set. This model was 7.9% more accurate than the ConvLSTM which in turn was 1% less accurate than the ResNet18 and Multinomial Linear Regression combination that was the best performing model in our previous project [6.3](#). Despite only having 2 ResNet layers this simple ResNet model has managed to improve the performance over the ResNet18 by 6.9% (the macro F1 score from [Table 6.3](#)), suggesting that the ResNet18 was overfitting to the data and the network was too complex for the task.

The new ResNet was established by using Bayesian Optimization to find the most optimal parameters. This was a different approach to the manual search used in our previous project. However, the Bayesian Optimization took four times longer than the manual search which still managed to find the second best set of hyperparameters (only 0.1% lower mean AUC over two fold cross validation [6.1](#)). Given there was no time constraint for this project, the improved search space tactics of Bayesian Optimization were used to train the ConvLSTM.

The ConvLSTM did not perform as well as the ResNet model on the RGB dataset. This indicates that although using the temporal aspects of the land is capable of achieving an adequate performance, a far better performance can be ascertained from using the spatial data and a state of the art machine learning algorithm. The temporal images were spread over the course of 5 years, with each image taken from the part of the year that had the least amount of cloud cover for that image's location. Perhaps if these images were instead taken from set intervals through just one year it would have allowed the model to find better patterns between the classes. However, such a method would drastically reduce the training set size as the months during winter were found to have extensive cloud cover across large sections of the Amazon, which would potentially lead to other issues such as class imbalances or clouds being associated with the Amazon.

The SNDVI dataset would not be phased by such an issue given it can pass through clouds, yet it unfortunately performed poorly on the data, achieving an accuracy 29.8% less than the same model when applied to the RGB dataset. This inability to find a separation between the classes could likely arise from each image only having one

dimension, although when NDVI was applied in our last project it achieved a much better performance (68.4% when used with a Random Forest [10]) even though it had the same number of dimensions. However, too many variables have changed between the NDVI results from our previous project and the SNDVI results in this project for them to be a fair comparison. The NDVI results were obtained from just one season in 2002 using Landsat 7. SNDVI on the other hand was collected across the full time period of January 2015 to December 2019. On top of this the NDVI results were obtained from a Random Forest but this was not possible with the SNDVI images that needed to be fed through a temporal model such as the ConvLSTM.

In our previous project a desired next step was to investigate how well the best performing model would be at detecting change over multiple years. The exploration achieved in this project has shown that the ResNet model can accurately help to identify areas of large change when viewed from a far ([Figure 7.3](#)) and also up close ([Figure 7.1](#)). This exploration has proven that there are two different use cases for this model; (1) the model can be used to identify large scale change, and (2) to identify change over a small scale where illegal deforestation or natural degradation of a biome can be easily identified. The first use case shows that this tool does indeed make it possible to continuously map the area of each biome as wildfires, deforestation, savannization or other land use changes force the flora to change. The second use case demonstrates that the model allows for continuous monitoring of a specific area of interest.

The exploratory analysis to view how capable the ResNet is at creating the maps highlighted that the current blanket classification approach the IBGE maps use is not sufficient. This project found that the areas within each biome where the land use changed and thus the flora - and likely the fauna that were dependent on this flora - changed with it. This meant the eco-systems that led to the area being classed as belonging to its specific biome were no longer the same and so the biome classification had changed with them.

The current mapping techniques used by IBGE consist of field-sampling the fauna and flora in an area in order to make the classification. This is a lengthy and expensive process but by focusing field sampling to areas where the model is inconsistent with its predictions (an example is shown in [Figure 7.3](#)) this could help reduce the amount of field sampling that is carried out.

It was also found that the land does change at a rate significantly faster than what the current mapping rate of 15 years is able to detect. It was noticed that actions such as deforestation have an immediate impact on the eco-systems that contribute towards the biome classification and so by automating the process of this classification it is possible to create maps that can keep up with the rapidly changing land uses.

In [section 7.4](#) it was found that despite the model having a relatively high softmax output that was over the pre-determined threshold, there can still be uncertainty in the prediction. From this it is proposed that a better heuristic for determining what class an input image is assigned to be used. This could be a combination of both the softmax, threshold, and uncertainty such that future classifications are only made when there is a low uncertainty and the model is confident the image belongs to one of the biomes.

Chapter 9

Conclusion

Our project found that **it is possible to accurately create maps of the three largest biomes in South America at a fraction of the cost and time** when compared to current methods. Classification of the Amazon, Cerrado and Caatinga was achieved with a 92% macro F1 score using a ResNet on our test ([Table 6.3](#)). The input to the model was freely available Landsat 8 visible light images. The ResNet had a 9.5% better macro F1 score than the ConvLSTM. From this it is concluded that although there may indeed be temporal trends that are unique to each biome, the better architecture of the ResNet is able to outperform them using purely spatial inputs. The maps this ResNet model is capable of creating are accurate, and able to respond to eco-systems that have changed due to actions such as deforestation. The project found that the land covered by IBGE's maps is changing at a rate that is significantly faster than that with which they are producing their maps. However, by using machine learning and remote sensing, our methods are able to create maps at a much faster rate and henceforth able to keep up with such changes to eco-systems, ensuring the most recent map produced is as true to reality as currently possible.

Chapter 10

Future Work

The ResNet trained on data taken solely from 2019 achieved a better performance than the ConvLSTM which was trained on data taken across five years. Before the ConvLSTM is completely disregarded it would be insightful to train the ConvLSTM on data taken evenly across just one year. This will possibly take better advantage of the seasons each biome goes through and thus create a better separation between the biomes.

Our project has only focused on three of the many biomes in South America. These biomes were chosen because they are the largest and had lots of training data available and gave the best chance to see if it was possible to monitor biomes from space. However, an ideal map would include all of the biomes that make up South America. The training data required to include these biomes in the model would be easy to acquire as all the biomes have been included in IBGE's 2019 map of Brazil, the source of training data for the biomes that were used in this project.

Moreover, it would be interesting to see how well the model migrates to other continents. The model has been so successful in South America due to the diverse appearances in the biomes that were chosen and so the same would likely need to be true for the new land mass. This means it is certainly likely to be effective in North America which has a similar composition of tropical rainforests and tundra but means it may be less effective in Europe which is mostly made up of different species of trees.

The conclusions presented in this project have proved it is possible for a cheap, accurate and continuous mapping tool of the main biomes in South America to be created. The next step in developing this tool would be to create a pipeline automatically connecting the data collection, image preprocessing, and the classification together. This pipeline could then be wrapped in an online website or desktop application which could be presented to numerous conservation charities and scientific bodies that are interested or concerned with the different land use changes of the Amazon, Cerrado, and Caatinga.

Bibliography

- [1] David Werth and Roni Avissar. The local and global effects of Amazon deforestation. *Journal of Geophysical Research: Atmospheres*, 107(D20):LBA–55, 2002.
- [2] IBGE (Instituto Brasileiro de Geografia e Estatística). Mapa de biomas do Brasil. Primeira aproximação. Ministério do Meio Ambiente e IBGE Brazil, 2004.
- [3] Eric A Davidson, Alessandro C de Araújo, Paulo Artaxo, Jennifer K Balch, I Foster Brown, Mercedes MC Bustamante, Michael T Coe, Ruth S DeFries, Michael Keller, Marcos Longo, et al. The Amazon basin in transition. *Nature*, 481(7381):321–328, 2012.
- [4] Paulo Brando, Marcia Macedo, Divino Silvério, Ludmila Rattis, Lucas Paolucci, Ane Alencar, Michael Coe, and Cristina Amorim. Amazon wildfires: Scenes from a foreseeable disaster. *Flora*, 268:151609, 2020.
- [5] AK Neves, TS Körting, LMG Fonseca, CD Girolamo Neto, Dennis Wittich, GAOP Costa, and Christian Heipke. Semantic segmentation of Brazilian savanna vegetation using high spatial resolution satellite data and U-Net. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*; 5, 3, 5(3):505–511, 2020.
- [6] Inacio T Bueno, Fausto W Acerbi Junior, 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.
- [7] Rosa M Azevedo T, Shimbo J. MapBiomass. <https://mapbiomas.org>, 2015. [Online; accessed 18-March-2022].
- [8] L Barreto de Melo Filho. Framework for deforestation prediction using Satellite Images and Neural Networks. Master’s thesis, Utrecht University, 2020.
- [9] Confidence Duku and Lars Hein. The impact of deforestation on rainfall in Africa: a data-driven assessment. *Environmental Research Letters*, 16(6):064044, 2021.
- [10] McMeekin C. Machine learning with medium resolution satellite imagery to detect change in South American biomes. Master’s thesis, The University of Edinburgh, 2020.

- [11] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [12] Ladislav Mucina. Biome: evolution of a crucial ecological and biogeographical concept. *New Phytologist*, 222(1):97–114, 2019.
- [13] Eduardo Mariano, Taciana F Gomes, Silvia RM Lins, Adibe L Abdalla-Filho, Amin Soltangheisi, Maria GS Araújo, Rodrigo F Almeida, Fernanda G Augusto, Luiza P Canisares, Sigleia SF Chaves, et al. Lt-Brazil: A database of leaf traits across biomes and vegetation types in Brazil. *Global Ecology and Biogeography*, 2021.
- [14] Luis F Salazar, Carlos A Nobre, and Marcos D Oyama. Climate change consequences on the biome distribution in tropical South America. *Geophysical Research Letters*, 34(9), 2007.
- [15] Chris M Bishop. Neural networks and their applications. *Review of scientific instruments*, 65(6):1803–1832, 1994.
- [16] 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.
- [17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [18] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [19] Paul J Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.
- [20] 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.
- [21] Jingzhao Zhang, Tianxing He, Suvrit Sra, and Ali Jadbabaie. Why gradient clipping accelerates training: A theoretical justification for adaptivity. *arXiv preprint arXiv:1905.11881*, 2019.
- [22] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. LSTM neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*, 2012.
- [23] Felix A Gers, Douglas Eck, and Jürgen Schmidhuber. Applying LSTM to time series predictable through time-window approaches. In *Neural Nets WIRN Vietri-01*, pages 193–200. Springer, 2002.

- [24] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.
- [25] Wen-Shuai Hu, Heng-Chao Li, Lei Pan, Wei Li, Ran Tao, and Qian Du. Spatial-spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4237–4250, 2020.
- [26] 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.
- [27] Masroor Hussain, Dongmei Chen, Angela Cheng, Hui Wei, and David Stanley. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of photogrammetry and remote sensing*, 80:91–106, 2013.
- [28] Henrique G Momm, Racha ElKadiri, and Wesley Porter. Crop-type classification for long-term modeling: An integrated remote sensing and machine learning approach. *Remote Sensing*, 12(3):449, 2020.
- [29] Sergii Skakun, Eric F Vermote, Jean-Claude Roger, Christopher O Justice, and Jeffrey G Masek. Validation of the LaSRC cloud detection algorithm for Landsat 8 images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2439–2446, 2019.
- [30] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.
- [31] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.
- [32] Saurabh Kumar, Shwetank Arya, and Kamal Jain. A SWIR-based vegetation index for change detection in land cover using multi-temporal Landsat satellite dataset. *International Journal of Information Technology*, pages 1–14, 2021.
- [33] François Chollet et al. Keras. <https://keras.io>, 2015. [Online; accessed 03-March-2022].
- [34] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT’2010*, pages 177–186. Springer, 2010.
- [35] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

- [36] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [37] 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.
- [38] Ziyi Huang, Henry Lam, and Haofeng Zhang. Quantifying epistemic uncertainty in deep learning. *arXiv preprint arXiv:2110.12122*, 2021.
- [39] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

Appendix A

ResNet

A.1 Hyperparameter tuning

Model ID	Layers	Hidden Units	Dropout	Mean	Mean	Mean
				AUC Fold 1	AUC Fold 2	AUC
1	1	32	-	0.966	0.965	0.966
2	2	32	-	0.963	0.978*	0.971
3	3	32	-	0.966	0.945	0.956
4	2	32, 64	-	0.970	0.973	0.972
5	2	32, 32	0.1	0.975*	0.970	0.973*
6	2	32, 32	0.2	0.971	0.958	0.964
7	2	32, 32	0.3	0.956	0.973	0.965

Table A.1: Manual hyperparameter tuning of the ResNet model. Tuning was applied on the RGB spatial training dataset with the results above coming from the networks performance on the spatial validation set.

Appendix B

Mapping Capabilities



Figure B.1: Images that have been classed as Inconclusive, Caatinga and the Amazon from the 2016 mapping capability analysis in [Figure 7.1](#)

B.1 Detecting Change on a Large Scale

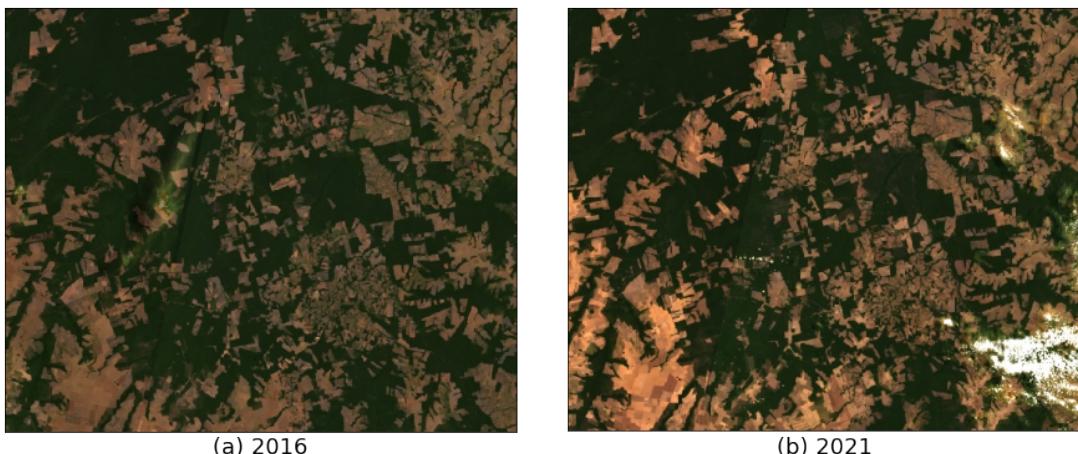


Figure B.2: Close up of an area of deforestation that is mentioned in [section 7.3](#)