Mapping South America's Biomes with Machine Learning

Calum McMeekin

The University of Edinburgh, Department of Informatics, Masters Project

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

1. Introduction

A biome is a collection of ecosystems that are close together and share similar fauna and flora. The Amazon, Caatinga and Cerrado are the three of the four largest biomes that make up South America covering an area over 9,460,037 km^2. **Knowing the area each of these biomes cover is crucial to monitoring land use changes, illegal deforestation, and the planets climate**. Current mapping techniques are expensive and time consuming with only two maps being produced, one in 2004 and one in 2019 [1]. 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 [2]. The results of this project have shown that for the first time **exceedingly accurate maps can be produced with great speed and low resources** when using medium resolution RGB satellite imagery with a ResNet architecture.

2. Data

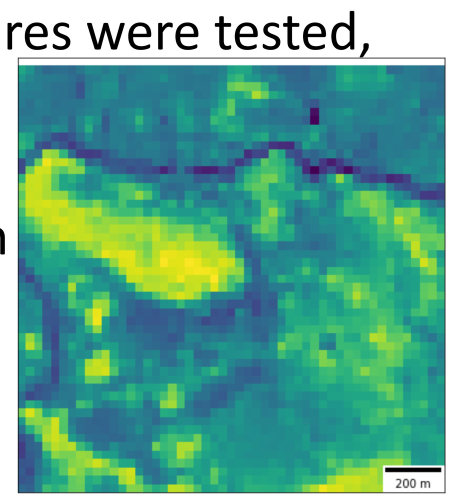
- **Source:** Landsat 8 Top-Of-Atmosphere corrected via Google Earth Engine
- **Bands**: Red, Green, Blue, Near-Infrared (NIR), Short-Wave Infrared (SWIR)
- **Resolution:** 30m per pixel
- **Area of Image:** 1.5km^2
- **Datasets:** Temporal Dataset, Spatial Dataset
- **Num. Of Images**: 34,594, 33,245
- **Years Collected From**: 2015, 2016, 2017, 2018, 2019

3. Approaches Evaluated

Two different models were evaluated, a temporal model (ConvLSTM) where the aim was to capture seasonality within each biome, and a spatial model. For each model two different sets of input features were tested, one using the Red, Green, and Blue (RGB) bands, and another using the Short Wave Natural Vegetation Index (SNDVI)

$$SNDVI = \frac{SWIR + NIR - Red}{SWIR + NIR + Red}$$

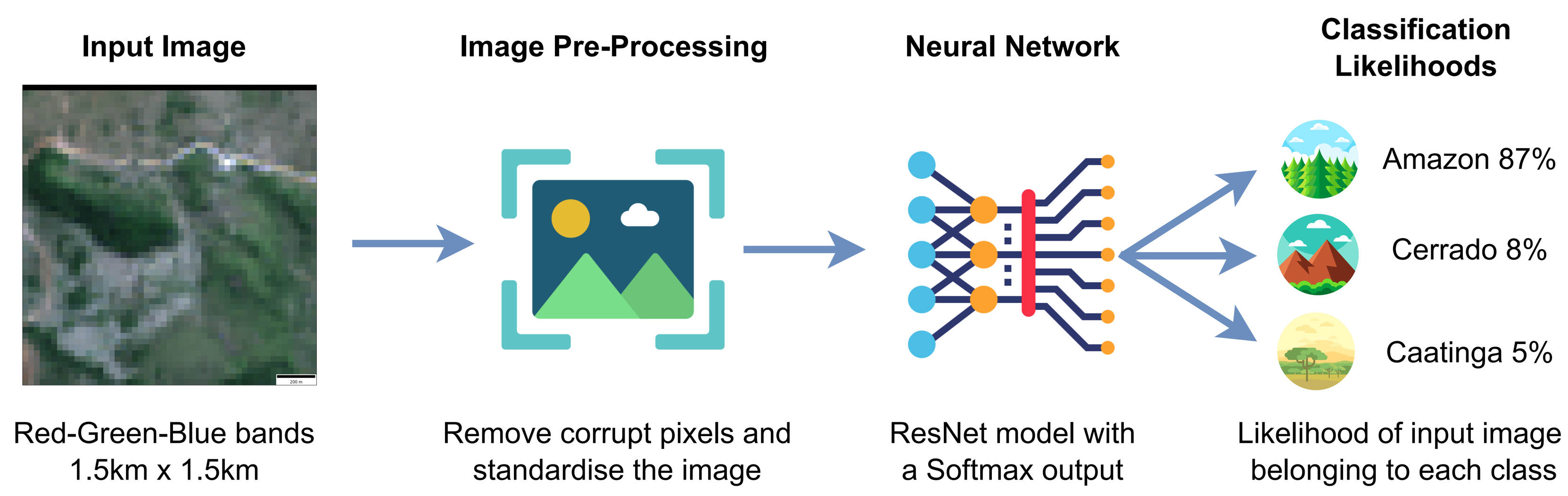
RGB



SNDVI

4. Final Pipeline

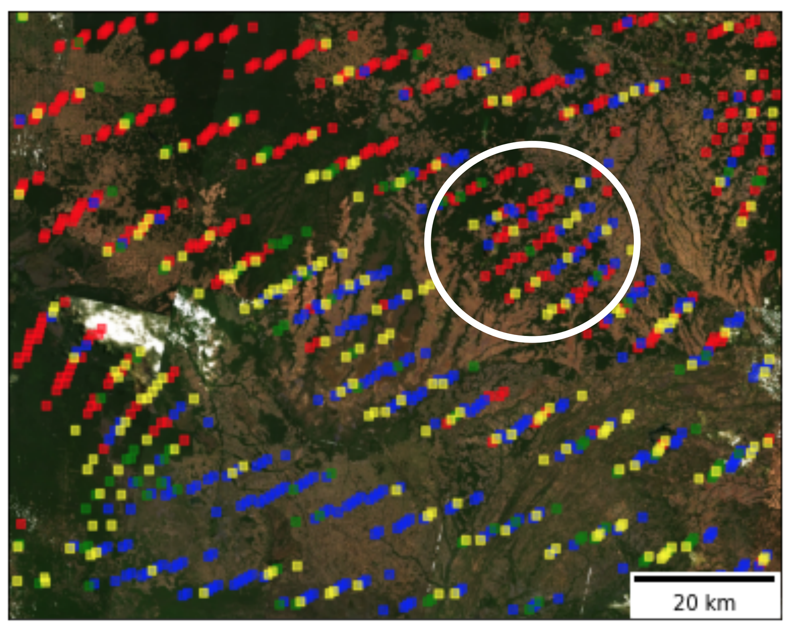
Each input image is fed through minor pre-processing, before being fed into the ResNet model. The model then outputs three percentages, indicating how likely it believes the input image belongs to each of the three biomes. If the model is less than 80% sure the input image belongs to any of the Biomes, then the predicted output will be a new class, which simply states the model is not sure which biome the image belongs to.



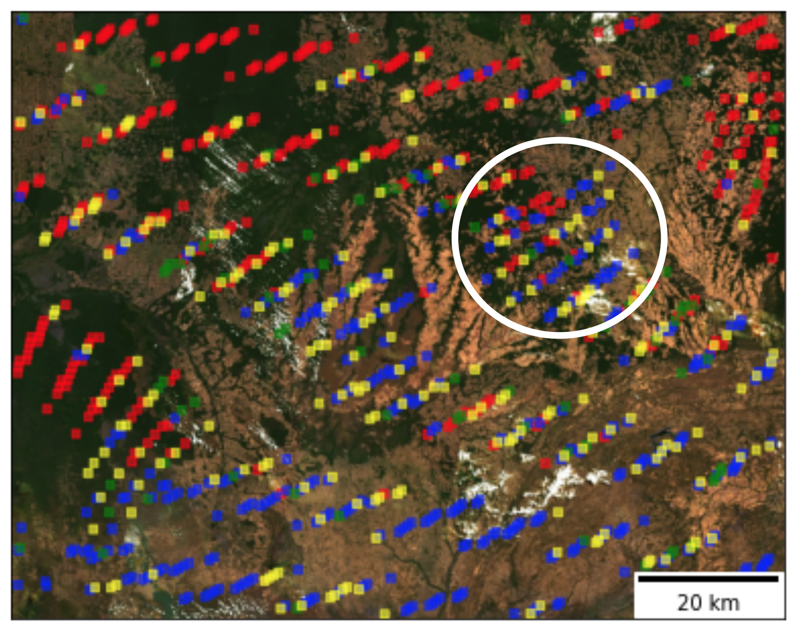
Once the class of the input image has been determined a mask is applied over the entire image. This can then be placed onto Google Earth Engine or any mapping software that supports latitude and Longitude shape files

5. Overall Evaluation

The machine learning models were all evaluated on data taken from a geographical region which the model had never seen before. This helped evaluate real world performance as closely as possible. The best model and feature combination shown in the Final Pipeline managed to achieve an overall macro f1-score of 92%, **proving that machine learning can accurately identify the different biomes using freely available satellite imagery**. Explorative analysis of the model shows that it is capable of accurately mapping an area. By monitoring its performance over an area of deforestation, visible changes in the land use become apparent.



2016



2021

Amazon | Cerrado | Caatinga | Inconclusive | Area Of Notable Change

6. Conclusions

This work has shown that by combining the freely accessible Landsat 8 satellite imagery with the latest machine learning techniques it is possible to monitor the changing land use of the largest biomes in South America. With this achievement the previously expensive task of taking in-person sampling to estimate the biomes covering 9,460,037 km^2 can now be done:

1. Quickly
2. Affordably
3. At an increased resolution
4. With a high accuracy (92%)

With the high affordability of this solution it also makes it perfect for Conservation Charities to use it for monitoring:

- Deforestation
- Reforestation
- Desertification
- Wildfires
- Planets Climate
- Habitat Health

7. References

[1] IBGE (Instituto Brasileiro de Geografia e Estatistica). Mapa de biomas do Brasil. primeira aproximacao, 2004

[2] A. K. Neves, T. S. Korting, L. M. G. Fonseca, C. D. Girolamo Neto, D. Witich, 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