A Novel Integrated Machine Learning Approach Utilizing Radar and Satellite Imagery for Selective Logging Remote Sensing Detection and Accompanying AI-Logging Map-Generating Webtool

Background

USAID estimates illegal logging to be a $150 billion industry, destroying the world's forests. More than half of all tropical deforestation is illegal, and contributes to the 1.5 gigatons of carbon released from deforestation annually (WWF). However, developing countries struggle without the funding or human resources to monitor their vast expanse of forests through forest patrol. The advent of machine learning allows for a remote sensing solution able to monitor the large region of forestry at low costs. At the mass quantities of selective logging occuring, forests are left with significant reductions in tropical biomass, growth of weeds/ poor quality - low diversity trees, loss in biodiversity, and are more susceptible to forest fires and soil erosion.

Selective Logging

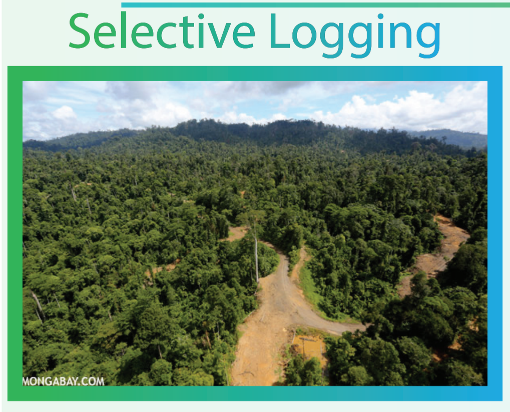


Image Credited To Tahreer Photography / Getty Images

Clear Cut Logging

Image Credited To Mongabay Photos by Rhett A. Butler

In the world's humid tropics, home to vast majority of forestry, persistent cloud cover often hinders the acquisition of clear optical satellite imagery. However, radar imagery overcomes this limitation by penetrating cloud cover, presenting an untapped opportunity for monitoring these regions.

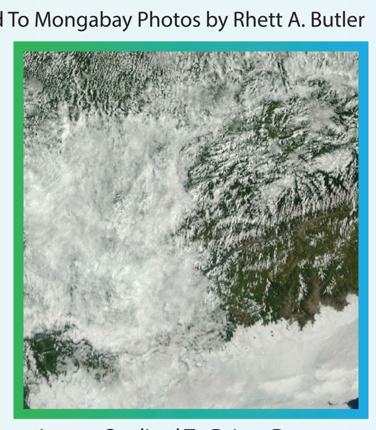


Image Credited To DrivenData

- Nearly 50-90% of tropical timber is illegally logged
- Illegal logging is estimated to be a $150 billion industry
- 1.5 gigatons of carbon result annually from deforestation CO_2

Research Question

How can an integration of optical satellite and radar (SAR) sensory data be used to improve logging detection models performance and accuracy in classifying selective logging and in addition create a interactive tool for forest protection agencies to identify logging occurences?

Data Acquisition

Sentinel 1 and 2 imagery was obtained through Google Earth Engine. In addition, to classify data points as logged / stable, the open source GFC annual forest map was used. The dataset selected location as Jamari National Park, which has sustainable forest management practice (selective logging) permits yearly.

Sentinel 1 Radar Imagery

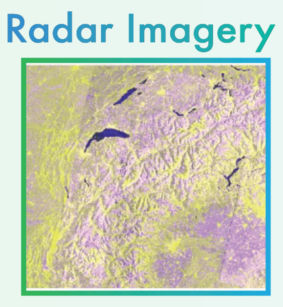


Image credited to Google Earth Engine

Sentinel 2 Optical Imagery

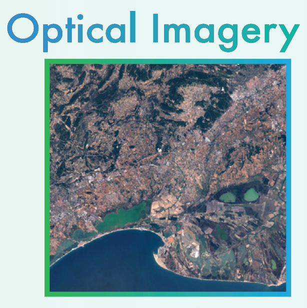


Image credited to Google Earth Engine

Global Forest Cover Map

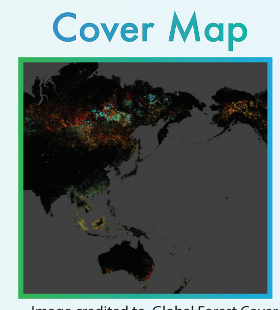


Image credited to Global Forest Cover

Data Location

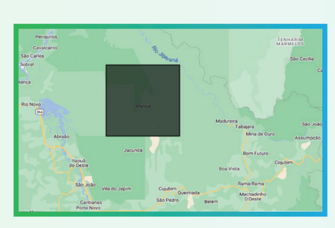


Image credited to Google Earth Engine

The data sets comprimised the following 12 band values from Sentinel 1 and Sentinel 2. Sentinel: VH and VV bands Sentinel 2: B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12 bands for both Jan/Dec. In all, 24 band values were used for the combined Sentinel 1 and Sentinel 2 data set.

Selected Sentinel 2 Bands

Sentinel 2 Features

	Resolution	Wavelength	Description
B2	10m	443.9nm (S2A) / 442.3nm (S2B)	Blue
B3	10m	496.6nm (S2A) / 492.1nm (S2B)	Green
B4	10m	560nm (S2A) / 559nm (S2B)	Red
B5	20m	664.5nm (S2A) / 665nm (S2B)	Red Edge 1
B6	20m	703.9nm (S2A) / 703.8nm (S2B)	Red Edge 2
B7	20m	740.2nm (S2A) / 739.1nm (S2B)	Red Edge 3
B8	10m	782.5nm (S2A) / 779.7nm (S2B)	NIR(Near Infrared)
B8A	20m	835.1nm (S2A) / 833nm (S2B)	Red Edge 4
B11	20m	864.8nm (S2A) / 864nm (S2B)	SWIR 1
B12	20m	945nm (S2A) / 943.2nm (S2B)	SWIR 2

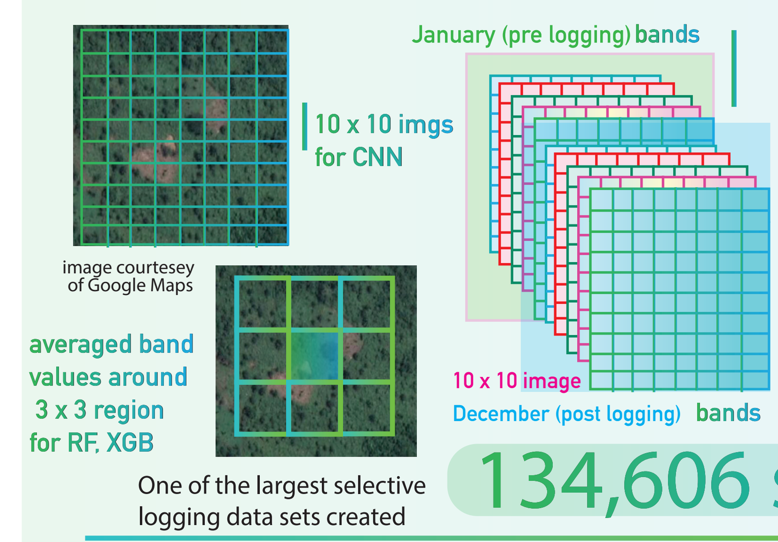
Selected Sentinel 1 Bands

Sentinel 1 Features

	Resolution
VH	10m
VV	10m

Data Processing

Using GFC map as reference, logged pixels and a subset of the stable forest pixels were identified to create a balanced dataset for unbiased model. For RF/XGB, averaged values of all bands were taken around a 3x3 region. For the CNN, raw 10x10 images were used. The December and January bands were then merged for model to learn the difference in band values before/after logging. For the combined Sentinel 1/2 dataset, the Sentinel 1 + 2 imagery was merged through concatenation.

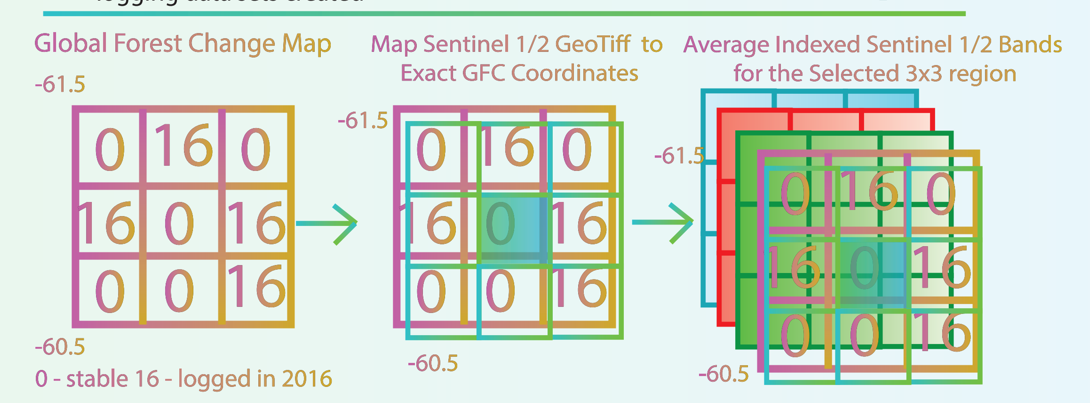


$$RVI = \frac{4\sigma VH}{\sigma VV + \sigma VH}$$

$$NDVI = \frac{NIR - RED}{NIR + RED} = \frac{B8 - B4}{B8 + B4}$$

Calculated NDVI & RVI (vegetation) indices

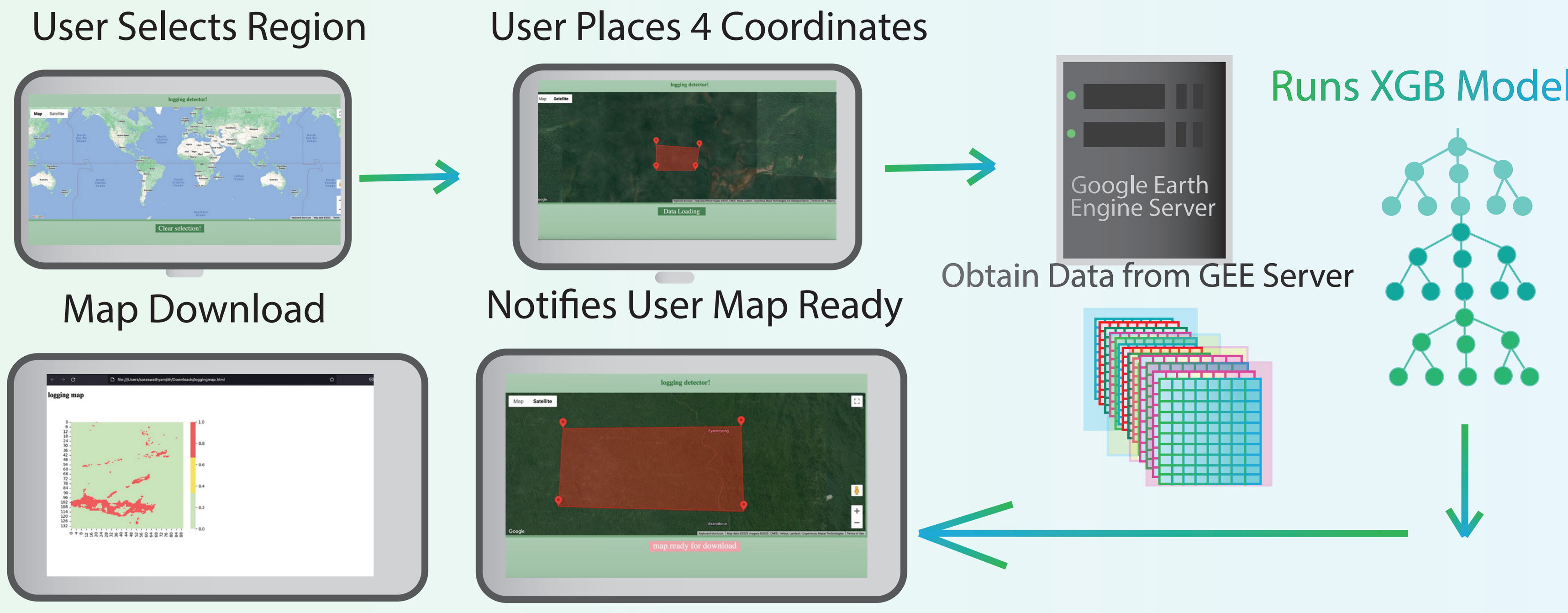
134,606 samples



classifcation	VVdec	VHdec	VVjan	VHjan	B2dec	B3dec	B4dec	B5dec	B6dec	B7dec	B8dec
1.0	1733.111111111110	1512.111111111110	1285.5	1458.8333333333300	2799.222222222220	3569.3888888888900	3470.5	4001.722222222220	2189.3333333333300	1288.888888888890	1143.888888888890
1.0	1761.3333333333300	1603.388888888890	1484.0555555555600	1603.3333333333300	2481.3333333333300	3005.222222222220	2851.4444444444400	3347.1666666666700	2510.6666666666700	1567.7777777777800	1077.388888888890
1.0	2098.5555555555600	1926.2777777777800	1752.3333333333300	1839.6666666666700	2720.1111111111100	3304.1111111111100	3252.777777777780	3640.222222222220	2521.3333333333300	1588.7777777777800	1020.8888888888900
1.0	970.777777777780	813.0	545.777777777780	869.777777777780	2085.777777777780	2667.8888888888900	2517.1111111111100	3103.6666666666700	1877.3333333333300	773.444444444450	980.111111111110
1.0	927.555555555560	723.0	422.888888888890	743.111111111110	2385.5555555555600	3253.4444444444400	3062.8888888888900	3725.6666666666700	1579.2222222222200	536.666666666670	1029.0
1.0	962.888888888890	723.444444444450	454.77777777777800	733.0	2328.777777777780	3208.777777777780	3022.0	3649.4444444444400	1611.0	561.666666666670	961.0
1.0	1096.6666666666700	926.888888888890	754.777777777780	1067.6666666666700	1882.0	2437.777777777780	2488.8888888888900	2925.4444444444400	2083.8888888888900	1082.3333333333300	949.611111111110
1.0	1483.888888888890	1140.111111111110	1005.5555555555600	1167.6666666666700	1762.0	2252.0	2372.5555555555600	2749.222222222220	2630.3333333333300	1451.2222222222200	976.944444444450
1.0	1469.111111111110	1300.0	1152.111111111110	1265.7777777777800	1893.2222222222200	2350.6666666666700	2568.3333333333300	2706.5555555555600	2477.3333333333300	1504.5555555555600	893.944444444450

Example slices of data points (CSV file- comma separated file) of averaged data values for Random Forest and XGB models data.

Logging Detector Webtool

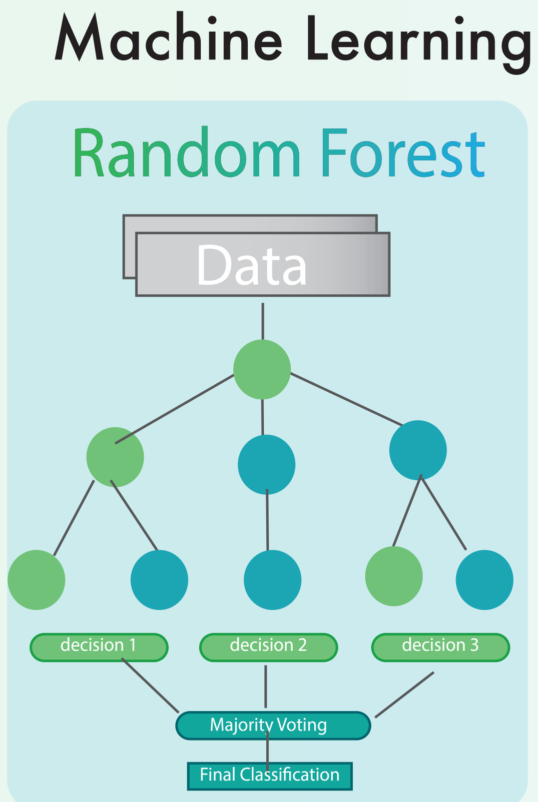


Models

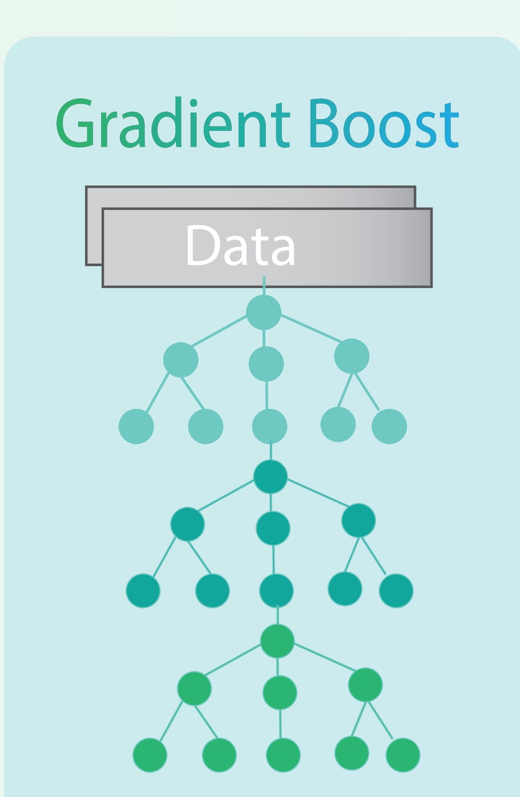
The various models explored are CNN (U- Net), Random Forest, Gradient Boosted Trees. The models are built using python libraries and trained and tested on the Sentinel 1, Sentinel 2, and Sentinel 1 and 2 datasets.

Machine Learning

Random Forest

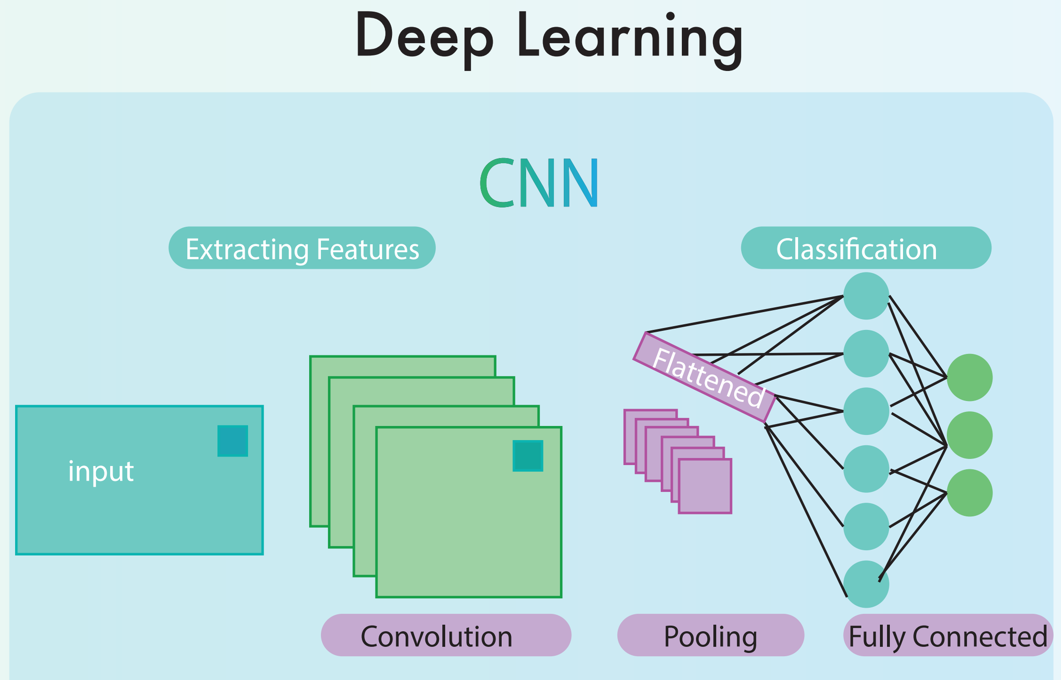


Gradient Boost

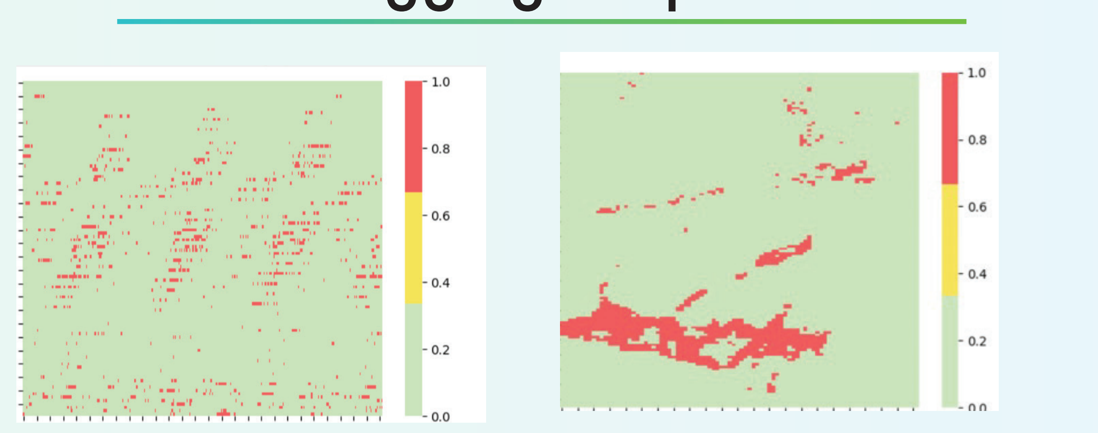


Deep Learning

CNN



Logging Maps

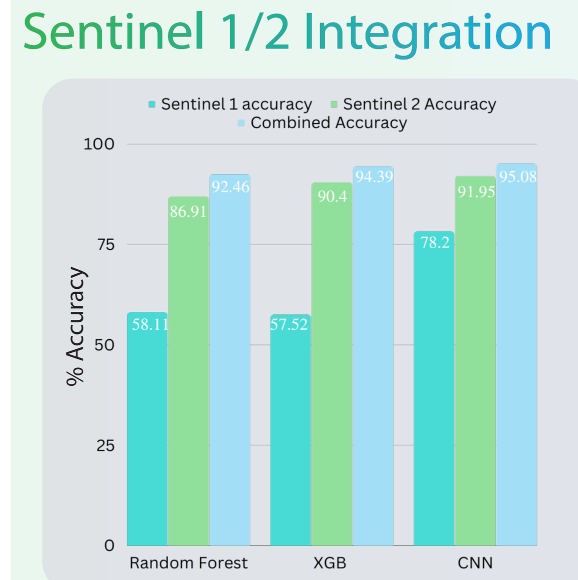


Small regions were selected for use as testing locations for logged /stable forest prediction maps. Using MATLAB and seaborn libraries, the models will be used to output prediction for each pixel and generate the maps shown above.

Results

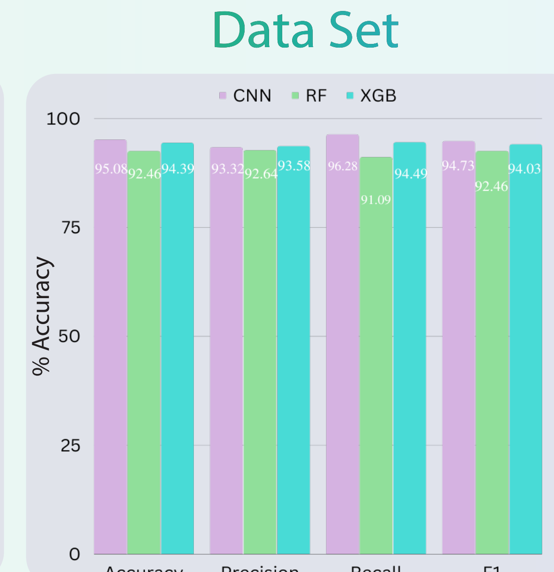
The models all improved from the integration of radar and satellite imagery, with the CNN performing best at 95.08 % accuracy and 94.73 F1. Pre existing solutions record 88% accuracy rate using only Sentinel 1, so this is a 7.08% increase.

Accuracy With/ Without Sentinel 1/2 Integration



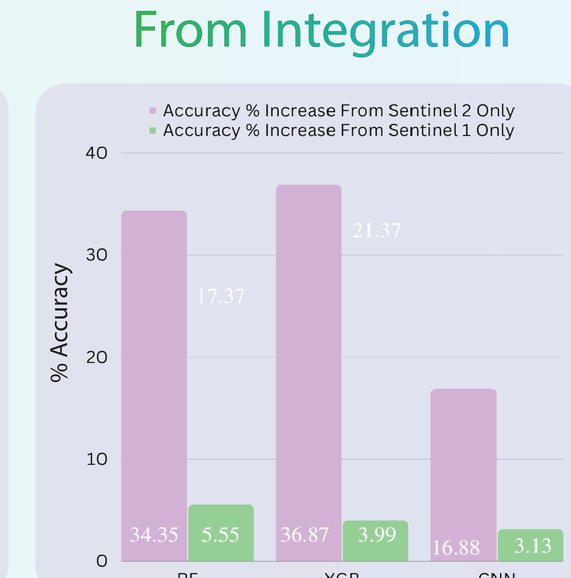
Graph depicts accuracy integrated/individual for all models evaluated

All Models on Integrated Data Set



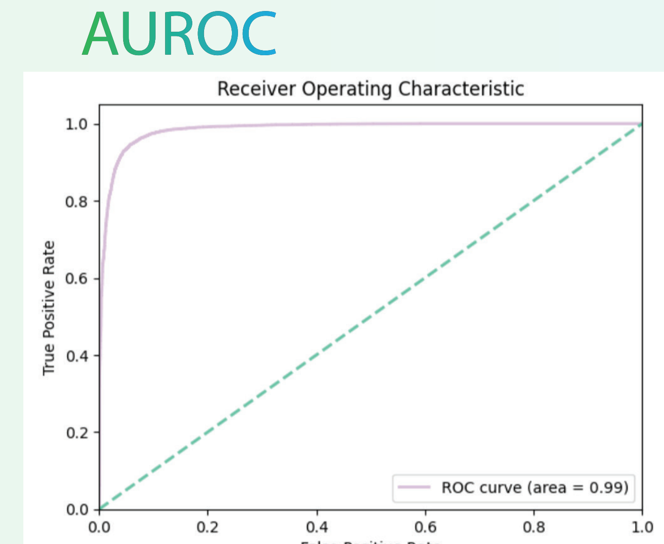
Graph depicts several metrics (Accuracy, Precision, Recall, F1) for all models evaluated

Accuracy Increase From Integration



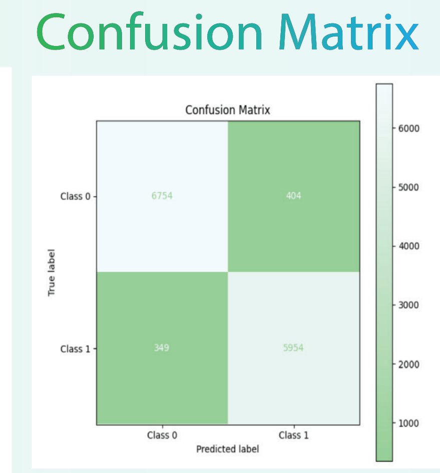
Graph depicts accuracy increase from integration for all models evaluated

AUROC



The XGB model achieved an AUROC score of 0.98537.

Confusion Matrix



In Order from Left to Right, Top to Bottom, TP, FP, FN, TN

Metrics

TP = True positives, i.e. the number of deforested areas classified as deforested.
TN = True negatives, i.e. the number of forested areas classified as forested.
FP = False positives, i.e. the number of forested areas classified as deforested.
FN = False negatives, i.e. the number of deforested areas classified as forested

Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision = TP / (TP + FP)
Recall = TP / (TP + FN)
F1 = 2 x (Precision x Recall) / (Precision + Recall)

Future Work

1. Reach out to forest protection agencies for feedback and testing of the web tool to make improvements and demonstrate functionality by contacting Amazon Trust, Rainforest Watch, and Rainforest Action Network.
2. Future Work includes hosting the website as a live website for anyone to access and use (currently limited by need for funding for cloud storage/Google Earth Engine API calls).
3. Contact Brazillian Forest Service for access to logging records and build training data classifications on more accurate logging records rather than Global Forest Cover Map

Significance

1. Integrating both optical and radar imagery for deforestation classification results in massive performance improvements (CNN - 3.13%) and 7.08% increase from existing models
2. Created a novel tool able to be used worldwide to detect logging occurrences
3. Models applicability extends beyond training location to worldwide

All images/graphs were created by the student researcher unless otherwise noted.