IDS 705 Final Report

Côte d'Ivoire Biomass Density Estimation Project

A Comparative Study of Random Forests and CNN in Estimating Biomass and the Relative Importance of Spectral Bands, Cloud Cover, Latitude, and Longitude in the Modeling Process

Team Members:
Lorna Maria Aine
Nick Carroll
Heather Qiu

Team Number: 3

Abstract

This project's goal is to to improve the efficiency of biomass estimation using machine learning techniques in areas experiencing deforestation. Biomass has traditionally been measured manually by field experts, which is costly, time-consuming, and difficult to scale. This project explores remote sensing techniques using satellite images. Specifically, this analysis compares the performance of Random Forest models against ResNet Convolution Neural Networks in predicting Above Ground Biomass Density (AGBD) of cocoa plantations in Côte d'Ivoire.

The results of this project suggests that these models are ineffective at predicting biomass from Sentinel-2 satellite imagery. While other studies have successfully implemented both of these methods for predicting biomass, the referenced work appears limited to higher resolution imagery (i.e. UAV and high resolution imagery). Additionally, this analysis compares the importance of the image data (i.e., spectral bands, cloud cover, and location), which suggests that the Infrared and vegetation red wavelengths are most important in predicting biomass. However, due to the inconsistency in feature importance across models, and multicollinearity between bands, this analysis cannot be considered a causal effect.

1. Introduction

Côte d'Ivoire has experienced significant deforestation since its independence in 1960, where 80% of its forests have vanished. Forests are essential to the ecosystem, and promote sustainable cocoa farming practices in cocoa growing regions. To combat deforestation, the government and private sector have prioritized the planting of shade trees, with the goal of restoring 20% of previously lost forests by 2030. However, field verification of reforestation efforts is both costly and difficult, due to intensity of labor, time, and accessibility. As a result, remote sensing is a potential option to more effectively monitor reforestation efforts and forest degradation.

The 2023 Africa Biomass Challenge is an ongoing data science competition hosted by Zindi (Blaser, W), with the goal to develop a machine learning method which accurately estimates biomass in Côte d'Ivoire. The challenge provides Global Ecosystem Dynamics Investigation (GEDI) LiDAR, Sentinel-2, and ground truth biomass data for this analysis. Ultimately, the challenge hopes to develop a cost-effective and efficient method to measure reforestation.

Recently, there has been substantial research using machine learning techniques to estimate biomass. Random Forest (RF) and Convolutional Neural Networks (CNN) are well-established techniques with strong performance in above-ground biomass estimation (Tamiminia et al., 2008). However, unlike the data in this challenge, these studies primarily use high resolution imagery in the visible light bands (Castro et al., 2020). This analysis compares the performance of these techniques in biomass estimation on the Africa Biomass Challenge dataset. Furthermore, it compares the relative model performance across the different data channels (i.e. spectral bands, cloud cover, location) included in the dataset. The ResNet models were originally developed and trained on ImageNet classification dataset (He et al., 2015), but recent studies have successfully implemented them to estimate above-ground biomass (Astola et al., 2021).

2. Background

2.1 The Project background

Côte d'Ivoire, a key player in the cocoa industry, is currently experiencing a significant reduction in its forest cover, which poses a severe threat to both biodiversity and cocoa production. In response to this, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH has developed a joint mission to facilitate monitoring deforestation and support reforestation efforts in collaboration with Bureau National d'Études Techniques et de Développement (BNETD). Stemming from this initiative, the Africa Biomass Challenge's goal is to develop effective remote monitoring techniques. The data science competition, which commenced in September 2021, is divided into two phases. The first phase developed a high-quality open-source dataset using the GEDI LiDAR system and Sentinel-2 satellite imagery, along with a collection of raw images of biomass (ground truth) from 94 locations in the towns of Guiglo, Abengourou, and Aboisso. The second phase uses this dataset to train machine learning models to accurately predict biomass in Côte d'Ivoire.

Satellite imagery and LiDAR-based remote sensing technologies could potentially monitor forest cover and change, but their accuracy is dependent on high-quality data for calibration and validation. Unfortunately, gathering accurate, field-measurements is time-consuming, laborious, and difficult. Other open-source options, like BIOMASS, are limited in terms of scale and accuracy because they are not dynamic enough to infer tree types or unknown vegetation. This project was developed to analyze the ability to predict biomass from Sentinel-2 and GEDI data.

Traditional methods for detecting biomass in forested areas include field sampling and remote sensing. Manual measurements in the field are time-consuming, expensive, and prone to human error. Remote sensing estimates biomass using satellite images, but calibration requires ground truth data, which can be difficult to obtain in some areas. To overcome these limitations, researchers have been increasingly exploring machine learning and deep learning techniques previously used in computer vision, natural language processing, and medical imaging in recent years. They have been shown to outperform traditional methods in a variety of ways, including increased efficiency, adaptability to various ecosystems, increased accuracy, and real-time monitoring capabilities (Chen et al., 2021, 1-3).

Biomass estimation is limited to the number of spectral bands available in the satellite imagery data; however, recent access to Sentinel-2 data provides 12 spectral bands, including Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, Vegetation Red Edge 2, Vegetation Red Edge 3, Near Infrared NIR, Narrow NIR, Water Vapor, SWIR 1, SWIR 2, (Hoersch, 2013) which could provide more information about the vegetation structure and health. Access to these additional spectral bands provides hope for increased accuracy in biomass estimation from satellite imagery.

2.2 Related works

In recent years, several studies have investigated the use of machine learning algorithms for biomass estimation. Random forests were shown effective at predicting biomass using remote sensing data by Esteban et al. (2019); however, based on the image size noted, the Airborne Laser Scanning (ALS) data seems to have substantially different resolution than the Sentinel-2 data. Additionally, ResNet Convolutional Neural Networks (CNNs) were shown effective at predicting biomass (Varela et al. 2022) using high-spatiotemporal-resolution UAV imagery.

Most promising, CNN's were shown effective at predicting above-ground biomass in the Democratic Republic of the Congo (DRC) (Tappayuthpijarn & Vindevogel, 2001, 6-9) using Sentinel satellite imagery. Furthermore, this study suggested longer-wavelength bands should provide more accurate predictions of biomass.

When comparing model performance, the Challenge provided a baseline model with performance of 52.9 MKg/ha RMSE. Nevertheless, two reasonably related studies were able to show very high performance for both Random Forest and ResNet models at predicting biomass. In a study predicting biomass of shrub willow, the Random Forest model performed with a 1.73 Mg/ha RMSE, and the Convolutional Neural Network performed with a 2.69 Mg/ha RMSE (Tamiminia et al., 2008). Another study predicting biomass of aeroponic cultivation using ResNet-50 model performed with a 0.0466 g/(g·day) RMSE (Åström et al., 2023). While these studies suggest that both models are suited for predicting biomass, both studies used much higher resolution cameras than the Sentinel-2 satellite, and the study on aeroponic cultivation used images which faced the flora instead of overhead imagery.

Based on these effective models, Random Forests were compared against ResNet CNNs for estimating biomass using the Cote d'Ivoire dataset. Furthermore, building off the analysis of Tappayuthpijarn et al., the relative importance of spectral bands, cloud cover, and location were analyzed.

3. Data

3.1 Data Description

As previously mentioned, the first phase of the Africa Biomass Challenge developed a high-quality open-source reference dataset using the GEDI LiDAR system and Sentinel-2 satellite imagery. The challenge dataset was formed by combining Sentinel-2 images with corresponding GEDI data. The ground truth held-out dataset was collected from field measurements and annotated using the BIOMASS package in R with expert supervision.

The reference data was split into train, validation and test sets and the ground data kept as a held out dataset (*see Figure 1*). The training set consists of over 25,000 observations, while the validation and test sets each consist of over 5,000 observations. The held-out test data set consists of over 90 observations.

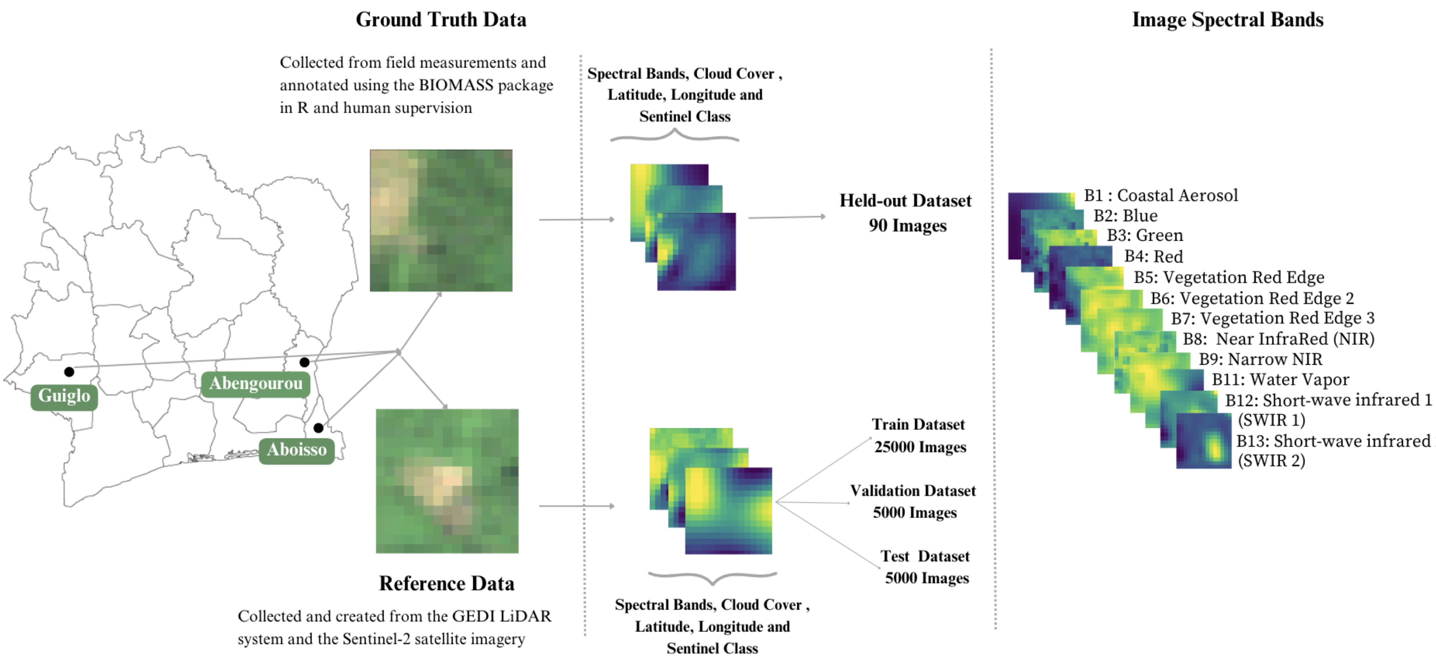

Figure 1: Data used in the project

The dataset includes Above-Ground Biomass Density (a measure of the amount of living plant material present above the ground,, measured in kilograms per hectare (MKg/ha)), cloud cover (a value of relative cloud cover on each pixel) the images (consisting of 12 bands: Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, Vegetation Red Edge 2, Vegetation Red Edge 3, Near InfraRed (NIR), Narrow NIR, Water Vapor, SWIR 1, SWIR 2), and location data (latitude and longitude for each pixel). Finally, the Sentinel-2 classification data provides information about the land cover type for each pixel. This could include categories such as forest, grassland, cropland, or urban areas.

As stated in the Africa Biomass Challenge - Dataset Preparation document, prior to distribution, the data was pre-processed to remove "noisy" data, by selecting Sentinel-2 images below a maximum cloud threshold of 20 and ensuring all images have at least 50% valid pixels. Furthermore, data was removed if the biomass variable was negative or if the geolocation data was irrelevant. Finally, a large number of GEDI variables were removed (Jiang & Becker, 2022).

3.2 Data Limitations

The images are 15 x 15 pixels in size, which is very small, and when viewed visually, appear blurry. Each pixel is expected to cover an approximate area of 10 meters by 10 meters. This is expected to make learning difficult for CNN methods, because the low resolution will limit the shapes that the CNN can learn, and the small size will reduce the pooling and dimensionality reduction layers available before the layer is too small. *(See figure 2 for an example of the images in the dataset)*. Furthermore, the output data has a very skewed distribution, which means the data must be balanced prior to training or the models will have difficulty predicting the data in the tail. Below shows the distribution of the target variables *(See figure 3)*.

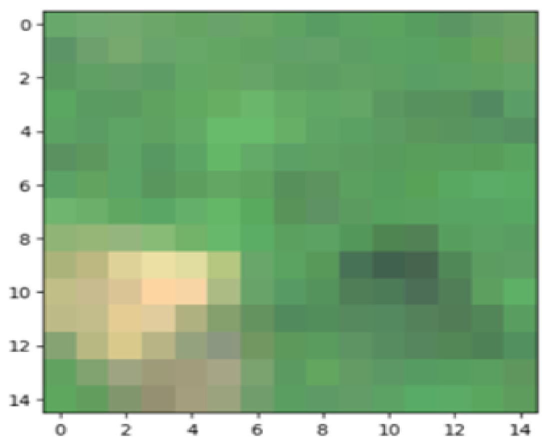

Figure 2. Example Image from Sentinel-2 Data

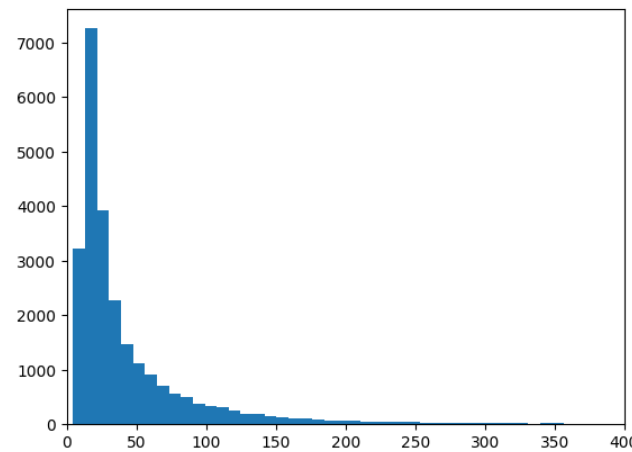

Figure 3. Distribution of Target Values

Finally, the held-out dataset was collected using a different method, includes substantially more cloud cover in the images, and has substantially less observations than the primary dataset. These factors tend to reduce generalization performance to a held-out dataset, but it is in fact this type of field verified data that is truly expected to be estimated.

4. Experiments

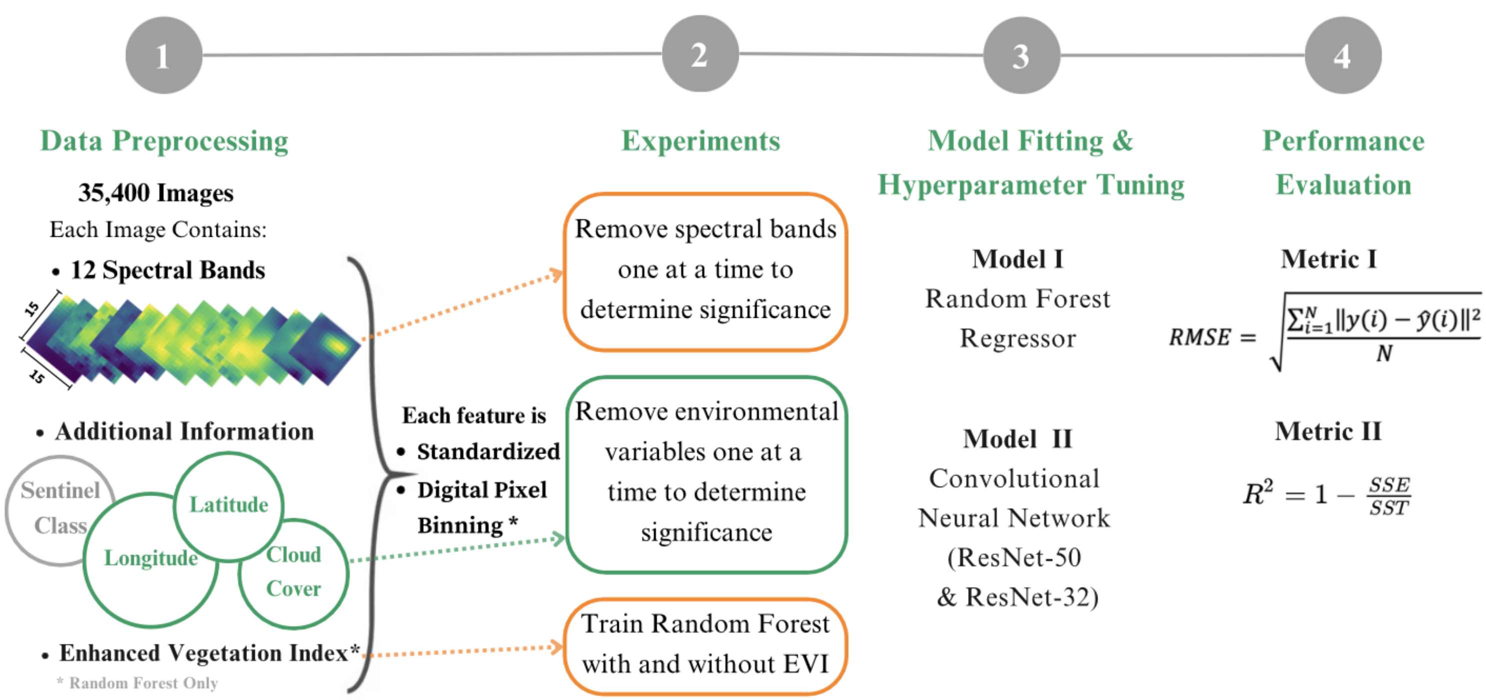

Figure 4: Modeling Process Map

4.1 Data Preprocesssing

The Challenge's data collection phase already included some data validation (i.e. removing invalid pixels and missing data). Nevertheless, the data was reviewed to ensure there was no missing data or errors during uploading and reading of it. Then, feature engineering was completed, adding a feature entitled the Enhanced Vegetation Index (EVI), which estimates vegetative indices. The EVI is calculated using reflectance values from red, blue, and near-infrared (NIR) light wavelengths *(Appendix 1),* and it is intended to account for factors such as atmospheric interference and soil background reflectance, which can affect the accuracy of biomass estimation. The Normalized Difference Vegetation Index (NDVI) is a similar index, but EVI has been shown to be more robust to atmospheric changes and better at distinguishing vegetation from bare soil in dense canopies. (Matsushita et al., 2007)

Following feature engineering, each feature was scaled to ensure comparable magnitudes in hopes of creating a well-formed loss space. Each feature was transformed to a space where its mean was zero and variance was one. Finally, feature engineering continued by creating three-bin histograms for each feature. This "pixel binning" is a technique which has been shown to capture image information while reducing issues of dimensionality. (Y et al., 2015)

4.2 Baseline Models

The Challenge provided its own baseline models, a linear regression using lasso regularization, which had quoted performance of 52.9 MKg/ha RMSE, and a CNN with 255.8 MKg/ha RMSE.

4.3 The Random Forest and ResNet Models

4.3.1 Random Forest Models

Multiple Random Forest models were compared with various feature engineering techniques. The first approach employed pixel binning and the Enhanced Vegetation Index (EVI) and explored the correlation between different combinations of spectral bands and environmental factors, such as cloud cover, latitude and longitude of the locations. Previous studies have demonstrated the effectiveness of both of these engineered features on the performance of Random Forests (pixel binning: (Gurung et al., 2016), vegetation index: (Tappayuthpijarn & Vindevogel, 2021)). The second approach used the features provided by the output of the hidden layers of a pre-trained Resnet-50 model from Hugging Face (Zuppichini, n.d.) to train the Random Forest. Hyperparameter tuning of the Random Forest adjusted the number and the maximum depth of trees to maximize performance.

4.3.2 ResNet Models

In addition to the Random Forest model fine-tuned on the pre-trained ResNet-50 model, two ResNet models were compared. The first was a fine-tuned regressor model of the pre-trained ResNet-50, while the second was a custom-configured ResNet-32 model. The pre-trained ResNet-50 models were trained on ImageNet. It (with both the fine-tuned regressor head and the fine-tuned Random Forest head) only used the RGB bands of the data, because the ImageNet dataset that they were configured and pre-trained on only included RGB data. Conversely, the custom-configured ResNet-32 was configured to train on all channels of the data, including the spectral bands, cloud cover, and location data. The ResNet-32 architecture was chosen over the deeper ResNet-50 to increase regularization without reducing the image to a degree of too much loss.

5. Results

5.1 Evaluation Metrics

The performance metric for this project is root mean squared error (RMSE), as established at the 2023 African Biomass Challenge; however, R-squared is the primary measure shown in the charts below for clarity. The associated RMSE of each model is reported in tables in the appendix along with the R-squared. For reference, the RMSE on the test dataset when compared to the mean value of the test dataset is 65.18 MKg/ha.

Each Random Forest model was run three times with the same hyperparameters and features to account for noise in the data. The average RMSE and standard deviations were calculated and reported in the appendix.

5.2 Overall Model Results

Below is a chart (*Figure 5*) comparing the overall model performance for each model architecture. The chart shows that Zindi's baseline reported performance is best (although this was not reproducible by this team), and that the pre-trained model performed worst. While the ResNet-32 model performed worse than the baseline model, it performed substantially better than the Random Forest model. Nevertheless, all models accounted for only a small portion of the error in the test data. In particular, the pre-trained ResNet-50 model performed very poorly with negative R-squared.

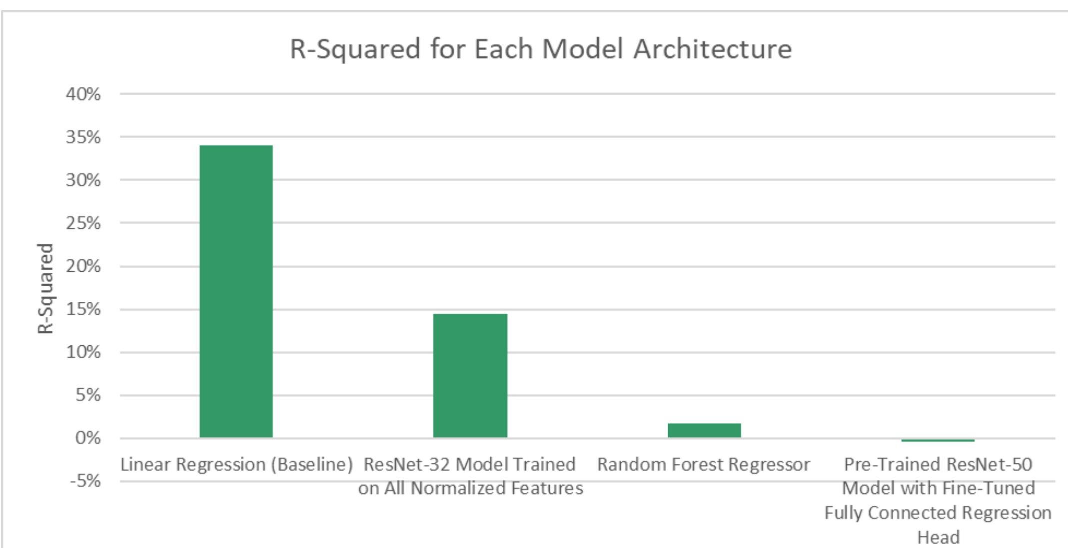

Figure 5: R-Squared for Different Model Architecture

5.3 Random Forest Feature Engineering Results

As previously mentioned, Random Forest models were trained on three different feature spaces: the original standardized features with pixel binning, the data with the additional EVI feature, and the feature space of the pre-trained ResNet-50 model. Below is a chart showing the R-squared for each model. The chart (*Figure 6*) shows that the Random Forest model with the best performance is the model without any feature engineering. Furthermore, it shows that the Random Forest trained on the pre-trained ResNet-50 feature space performed extremely poorly with negative R-squared.

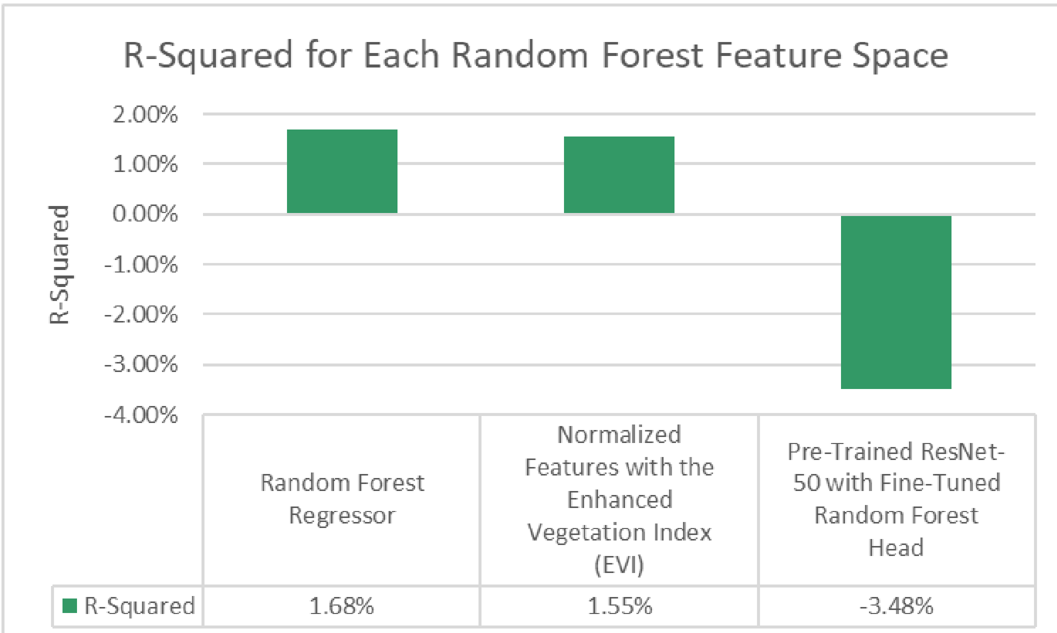

Figure 6: R-Squared of RF Regressor with Different Feature Engineering

5.4 Impact of Spectral Information, Cloud, and Location Data

Following model training, experiments were performed on each model to understand the relative importance of each spectral band, along with the other channels of data (cloud, location, etc.). These experiments were performed in two ways on the Random Forests: first, feature importance was measured by calculating the average cumulative impurity decrease (the amount of nodes that become pure when applying the feature) of a feature; second, as each band was removed, a Random Forest was trained on the data excluding that band. The feature importance measure shows how important a spectral band is for a single model, while retraining the Random Forest on the data excluding the band shows how well the model can perform without the data (due to correlation in the data). Below is a chart (*Figure 7*) of feature importance which suggests that EVI, near infrared, and red wavelengths are the most important features to the model.

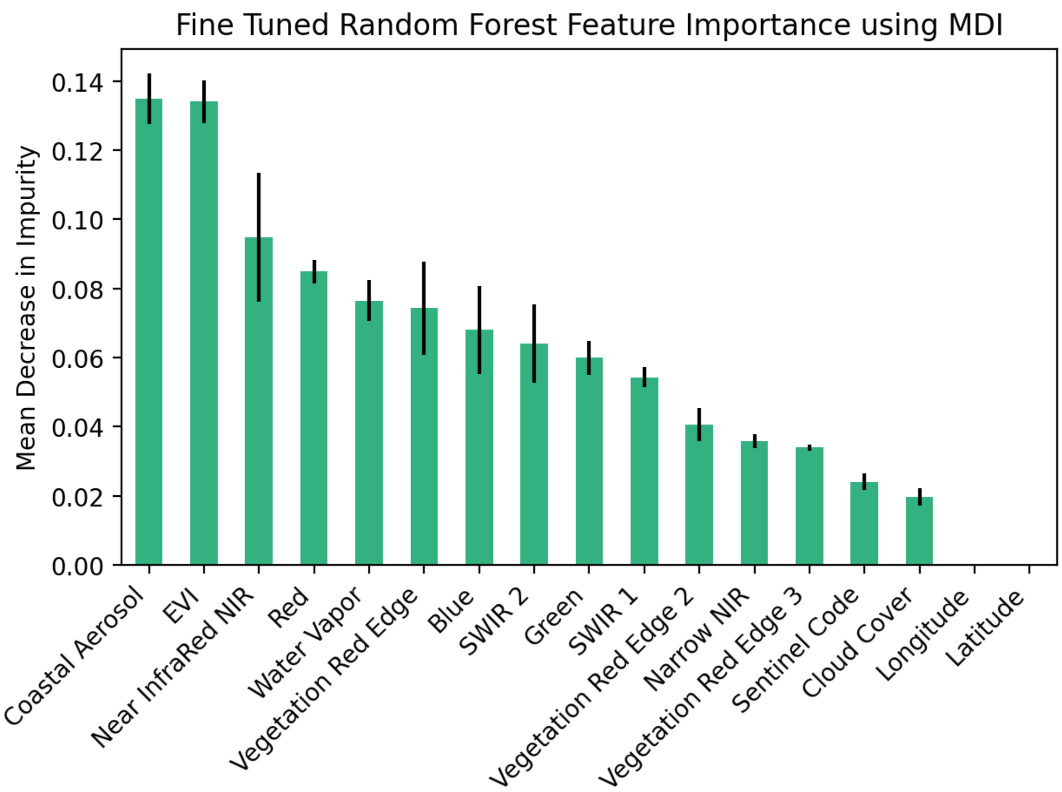

Figure 7: Random Forest Feature Importance

Additionally, below is a chart (*Figure 8*) showing the RMSE and R-Squared performance of the Random Forest models trained on the data with missing bands. This chart shows that the RMSE did not change substantially as bands were removed, suggesting a large degree of correlation (or mutual information) in the bands. Nevertheless, the chart shows more noticeable differences in the R-Squared. In particular, the RGB bands alone performed worst. This chart also suggests that red and infrared are the most important wavelengths in predicting biomass.

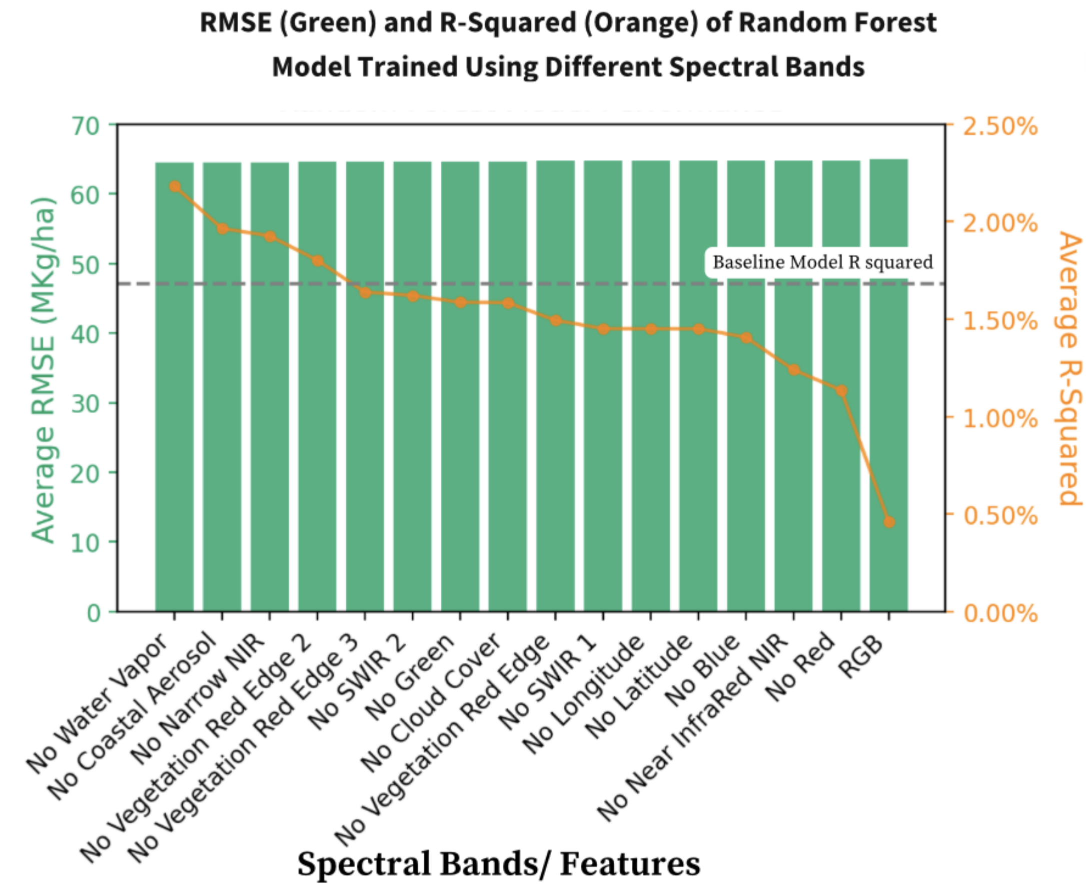

Figure 8: Random Forest Feature Removal Comparison

Similar to measuring feature importance for the Random Forest models, after the ResNet-32 model was trained, experiments were completed by masking each input channel to understand its relative importance. Below is a chart (*Figure 9*) showing the RMSE of the model with each band masked, one-by-one. The masked bands which have the highest RMSE suggest that those bands are most critical to the model's performance. While it generally does not agree with the feature importance of the Random Forest model, it does also suggest that near infrared and vegetation red wavelengths are likely important to predicting biomass.

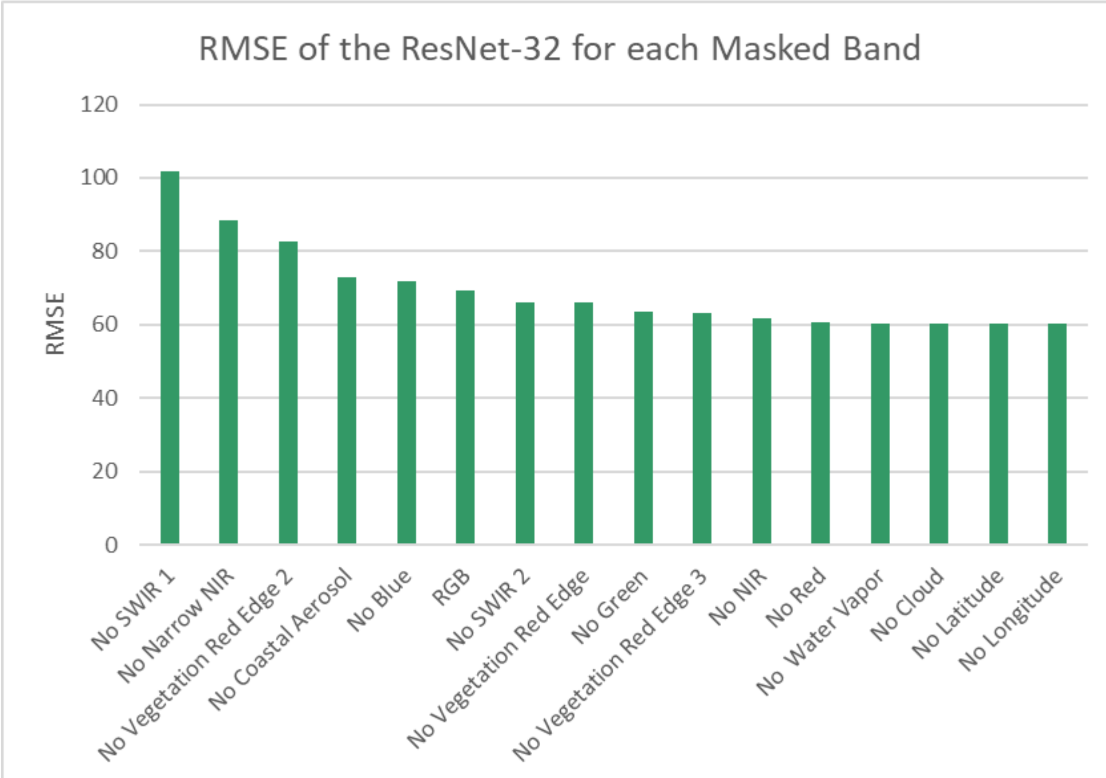

Figure 9: ResNet-32 Feature Importance Approximation

5.5 Model and Data Limitations

Both models which used the pre-trained ResNet-50 performed extremely poorly with negative R-squared values. In hindsight, this is unsurprising because, although the ResNet-50 is a state-of-the-art model, its training data (ImageNet) is drastically different from the Africa Biomass Challenge (ABC) dataset in two important ways:

1. ImageNet images are generally images of people, animals, and objects taken from the side, not overhead forest canopy; therefore, the features the pre-trained ResNet-50 learned are likely very different shapes than what is required to predict biomass from this dataset, and
2. ImageNet images are much larger with much higher resolution: images are generally 224 x 224 with each pixel capturing a spatial dimension on the order of millimeter or centimeters, whereas the ABC dataset images are 15 x 15 with each pixel covering a spatial dimension approximately 10 meters; therefore, the features space of the pre-trained ResNet-50 is in a much different order of magnitude, and the images cannot go through as many pooling and regularization layers before their information is lost.

Although the custom-configured ResNet-32 performed substantially better than the pre-trained ResNet-50, it still was unable to account for the far majority of errors in the test dataset. As previously discussed, other research has shown substantial promise using ResNet-50 to predict biomass, but this research seems to have focused on higher resolution images. The ResNet architecture does not seem well designed, or well suited, for small and low resolution images.

While experiments were completed to understand the relative importance of spectral and other data in the dataset, the underperformance of the models and large amount of relative noise in the data suggests the model is not well-suited to make inference estimates. Nevertheless, the various models suggest the infrared and vegetation red wavelengths may be most important, which aligns with previous research.

6. Conclusions

Based on this analysis, ResNet and Random Forest models appear to be unable to accurately predict biomass from Sentinel-2 satellite imagery; therefore, it is recommended that Côte d'Ivoire research alternative methods for remote sensing of reforestation. Based upon previous studies, it may be possible to accurately predict biomass with these methods from higher resolution data. Previous studies using higher resolution data (such as UAV images) suggest that image collection from UAV's are more expensive than Sentinel-2 due to the more specified use-case of the images. With this in mind, research focused on further use of LiDAR systems (like the GEDI which was used for the training dataset collection) may provide more cost-effective solutions than high resolution imagery.

Roles

Team Member	Role	Responsibilities
Lorna Maria Aine	Contributor	- Data preprocessing - Report Writing
Nick Carroll	Contributor	- ResNet Models - Report Writing
Heather Qiu	Contributor	- Random Forest Models - Report Writing

References

1. Astola, H., Seitsonen, L., Halme, E., Molinier, M., & Lönnqvist, A. (2021, June 18). Deep Neural Networks with Transfer Learning for Forest Variable Estimation Using Sentinel-2 Imagery in Boreal Forest. VTT CRIS. Retrieved April 4, 2023, from https://cris.vtt.fi/ws/portalfiles/portal/51709038/rs13122392.pdf
2. Åström, O., Hedlund, H., & Sopasakis, A. (2023, 3 21). Machine-Learning Approach to Non-Destructive Biomass and Relative Growth Rate Estimation in Aeroponic Cultivation. MDPI. Retrieved April 4, 2023, from https://www.mdpi.com/2077-0472/13/4/801
3. Blaser, W. (n.d.). Africa Biomass Challenge. Zindi. Retrieved April 17, 2023, from https://zindi.africa/competitions/africa-biomass-challenge
4. Castro, W., Junior, J. M., & Polidoro, C. (2020, August 26). Deep Learning Applied to Phenotyping of Biomass in Forages with UAV-Based RGB Imagery. NCBI. Retrieved April 4, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7506807/
5. Chen, Y., D, H., & T, H. M. (2021). Estimating Pasture Biomass Using Sentinel-2 Imagery and Machine Learning. Remote Sensing, 13(603), I - 3.

6. Esteban, J., McRoberts, R. E., Fernandez - Landa, A., Tome, J. L., & Næsset, E. (2019). Estimating Forest Volume and Biomass and Their Changes Using Random Forests and Remotely Sensed Data. Remote Sensing, 10(11). https://doi.org/10.3390/rs11161944
7. Gurung, R. B., Lindgren, T., & Bostrom, H. (2016). Learning Decision Trees from Histogram Data Using Multiple Subsets of Bins. Florida Artificial Intelligence Research Society Conference. Retrieved April 8, 2023, from https://people.dsv.su.se/~tony/papers/FLAIRS_2016.pdf
8. He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). [1512.03385] Deep Residual Learning for Image Recognition. arXiv. Retrieved April 17, 2023, from https://arxiv.org/abs/1512.03385
9. Hoersch, B. (2013, September 1). ESA Standard Document. Sentinel Online. Retrieved April 18, 2023, from https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook
10. Jiang, Y., & Becker, A. (2022, December). ABC Dataset Report. Zindi. Retrieved April 18, 2023, from https://zindi.africa/competitions/africa-biomass-challenge/data
11. Matsushita, W, Y., J, C., Y, O., & G, Q. (2007). Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. Sensors(Basel), 7(11).
12. Tamiminia, H., Salehi, B., Mahdianpari, M., Beier, C. M., Klimkowski, D. J., & Volk, T. A. (2008, November 16). Random Forest Outperformed Convolutional Neural Networks for Shrub Willow Above Ground Biomass Estimation Using Multi-Spectral UAS Imagery. IEEE. Retrieved April 4, 2023, from https://ieeexplore.ieee.org/document/9553515
13. Tappayuthpijarn, K., & Vindevogel, B. S. (2021). High-accuracy Machine Learning Models to Estimate above Ground Biomass over Tropical Closed Evergreen Forest Areas from Satellite Data. IOP Conference Series: Earth and Environmental Science, 1006(01), 9. https://iopscience.iop.org/article/10.1088/1755-1315/1006/1/012001/pdf

14. Varela, S., Zheng, X., Njuguna, J. N., Sacks, E.J., & Allen, D.P. (2022). Deep Convolutional Neural Networks Exploit High-Spatial- and -Temporal-Resolution Aerial Imagery to Phenotype Key Traits in Miscanthus. Remote Sensing, 14(5333).
15. Wang, L., Zhou, X. D., Zhu, X. K., Dong, Z. D., & Guo, W. S. (2016, November 16). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. ScienceDirect. Retrieved March 15, 2023, from https://www.sciencedirect.com/science/article/pii/S2214514116300162
16. Yoo, Y., J, L., & J, P. (2015). Low-Light Image Enhancement Using Adaptive Digital Pixel Binning. Sensors(Basel), 15(7).
17. Zuppichini, F. (n.d.). ResNet Documentation. Hugging Face. Retrieved April 18, 2023, from https://huggingface.co/docs/transformers/model_doc/resnet

Appendix

Appendix 1: EVI calculation

$$EVI = 2.5 * [(NIR - RED) \div (NIR + 6 * RED - 7.5 * BLUE + 1)]$$

Appendix 2: Parameter spaces and the best parameter set for all model types

Model Architecture	**Best Parameter for All Bands**
Random Forest	'n_estimators': 100, 'max_depth': None, 'max_features': 1
Convolutional Neural Network	Layers: 32 (size 256), followed by 3 regularization layers (size 25), learning rate: 0.001, epochs: 7

Appendix 3: RMSE (MKg/ha) and R-Squared of RF Regressor with Different Feature Engineering

	Normalized Features with the Enhanced Vegetation Index (EVI)		**Pre-Trained ResNet-50 with Fine-Tuned Random Forest Head**	
	RMSE	R-Squared	RMSE	R-Squared
Fine Tuned Random Forest	64.67	1.55%	65.09	-3.48%

Appendix 4: RMSE (MKg/ha) and R-Squared of RF & CNN Models Trained using Different Bands

	Random Forest		**Convolutional Neural Network**	
	RMSE (Standard Deviation)	R-Squared	**RMSE**	R-Squared
RGB	65.03 (0.11)	0.46%	69.13	-12.48%
No Coastal Aerosol	64.53 (0.09)	1.97%	72.73	-24.53%
No Blue	64.72 (0.09)	1.41%	71.92	-21.74%

No Green	64.66 (0.07)	1.59%	63.36	5.49%
No Red	64.81 (0.08)	1.14%	60.53	13.76%
No Vegetation Red Edge	64.69 (0.06)	1.49%	65.94	-2.35%
No Vegetation Red Edge 2	64.59 (0.04)	1.80%	82.54	-60.36%
No Vegetation Red Edge 3	64.64 (0.05)	1.64%	63.29	5.72%
No NIR	64.77 (0.12)	1.24%	61.83	10.00%
No Narrow NIR	64.55 (0.10)	1.93%	88.26	-83.37%
No Water Vapor	64.46 (0.04)	2.18%	60.30	14.40%
No SWIR 1	64.70 (0.07)	1.45%	101.78	-143.84%
No SWIR 2	64.65 (0.13)	1.62%	66.18	-3.09%

Appendix 5: RMSE (MKg/ha) and R-Squared of Random Forest and CNN Models Trained using Different Environmental Factors

	No Cloud Cover		**No Latitude**		**No Longitude**	
	RMSE	**R-Squared**	**RMSE**	**R-Squared**	**RMSE**	**R-Squared**
Random Forest	64.66	1.58%	64.70	1.45%	64.70	1.45%
Convolutional Neural Network	60.27	14.49%	60.30	14.41%	60.25	14.56%