Deep Learning and Remote Sensing for Coastal Resilience: Scikit-Learn Model Performance

Sofia Nieto[1], Patrick Gray[2]

[1]Department of Computer Science, Duke University, Durham, North Carolina

[2] Duke Marine Lab, Nicholas School of the Environment, Duke University, Durham, North Carolina

Background

Remote sensing facilitates the study and management of coastal ecosystems. Satellite imagery and unoccupied aerial systems (drones) provide inexpensive, on-demand remote sensing capabilities with modular sensors that can assess a wide variety of variables. These tools generate large amounts of data. However, the expertise and time required to analyze this data, currently limits the speed and capacity at which this data can be used to generate ecological insights. The purpose of this project is to leverage machine learning to develop a methodology for rapid and high-resolution monitoring of the coastal ecosystems.

Research Questions

Landcover

- Are Scikit-learn models able to accurately classify landcover?
- How much training data is need to classify landcover?

Fractional Canopy Cover

- Are Scikit-learn models able to accurately predict fractional canopy cover?
- Does adding land cover data into the training data improve the accuracy of the models?

Methods

Study Site

- North Carolina coastline

Fractional Canopy Cover Input Tests

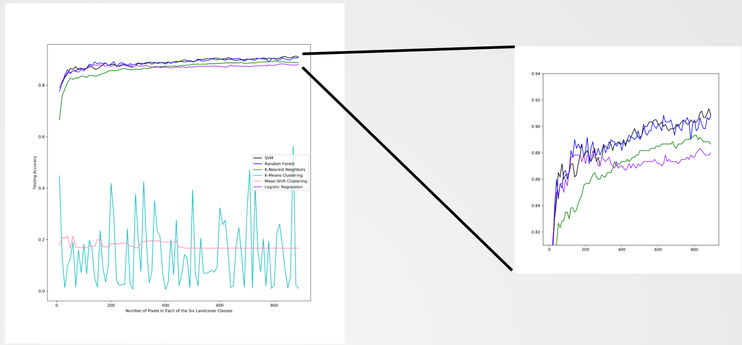
- Only pixel data
- Pixel data and landcover data added as one hot encoding
- Pixel data and landcover data added as class value (0 = water, 1 = developed, 2 = forest, 3 = cultivated, 4 = barren, 5 = wetland)

Models Tested

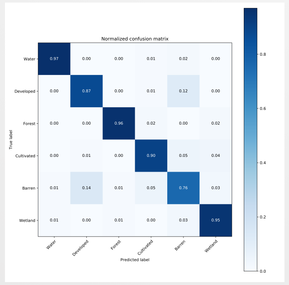
Landcover	Fractional Canopy Cover
• Random Forest Classifier • Support Vector Machine (SVM) • Logistic Regression • K-Nearest Neighbors (KNN) • K-Means Clustering • Mean-Shift Clustering	• Linear Regression • Random Forest Regression

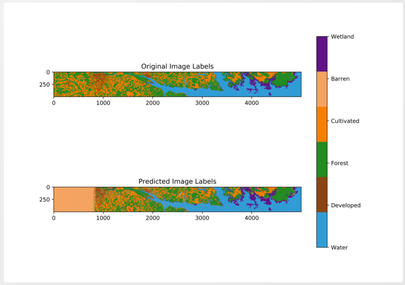
Results

Landcover Classification

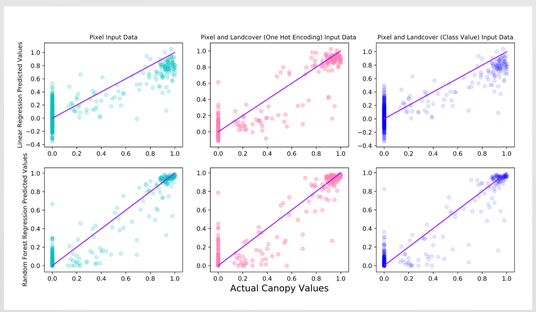


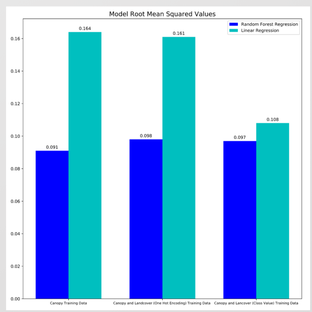
Performance of Best Landcover Model (SVM)





Fractional Canopy Cover Classification





Discussion

Apart from the clustering algorithm, the Sci-Kit Learn models are able to classify landcover with relatively high accuracy. The SVM, Random Forest Classifier, Logistic Regression, and KNN models reached their peak accuracy at around 600 pixels per class. Additional training data did not significantly increase or decrease their accuracy. The Mean-Shift Clustering and K-Means Clustering algorithm appear independent of the size of the training data.

The Sci-Kit Learn models are unable to predict the fractional canopy cover percentages accurately. This is likely partially due to inaccurate training data. Although adding the landcover data to the input data appears to have increased the canopy cover predictions at extreme values (0% coverage and 100% coverage), this decreases the accuracy of the models overall.

Next Steps

- Investigate methods to improve or new sources of fractional canopy cover input data
- Test if these models are generalizable to different areas