**Objective:** The goal of this project is to classify individual trees to their taxonomic species from hyperspectral imagery. The project is inspired by the 2020 data science competition: IDTReeS, [Integrating Data science with Trees and Remote Sensing](https://idtrees.org/competition/).

**Introduction:** Understanding the number, size, and species of individual trees in forests is crucial to mitigating the effects of climate change, managing invasive species, and monitoring shifting land use on natural systems and human society. However, collecting data on individual trees in the field is expensive and time consuming, which limits the scales at which this crucial data is collected. Remotely sensed imagery from satellites, airplanes, and drones provide the potential to observe ecosystems at much larger scales than is possible using field data collection methods alone. IDTReeS investigates combining large scale survey efforts with remotely sensed imagery to scale and improve long-term forest conversation and climate change mitigation efforts.

**Related Work:**

Why hyperspectral data?

Hyperspectral images contain multiple (typically between 64 and 256) continuous narrow bands, providing significant levels of detail, which allow for the distinction of fine spectral variations among tree species. This has resulted in the extensive use of hyperspectral imagery for tree species classification.

*General Approach from Literature*

1. Reduce Dimensionality of Data (using PCA for example)
2. Run a multi-class object classifier on bounding box info

**Explaining the Data:**

*On Remote Sensing Data:*

The competition provides three primary data sources: remote sensing, field data, and individual tree crowns. The following project will only consider remote sensing geospatial datasets, specifically passive sensing systems. Passive systems measure the amount of reflectance at different wavelengths for ground-detected objects. The remote sensing datasets are generated by the NEON Airborne Observation Platform (AOP), and distributed in RGB and Hyperspectral formats at 100 cm2 and 1 m2 spatial resolutions, respectively. Data is stored as raster files, which means an image or array of pixels, whereby each pixel is stored as a vector of numbers. An RGB image is stored as a 3-band raster (3-element vector pixels). Each band represents the reflectance at different points in the electromagnetic spectrum corresponding to red, green, and blue wavelengths, respectively. Hyperspectral data consist of reflectance information from a much wider electromagnetic spectrum (380-2510 nanometers), totaling to 426 wavelength bands.

*On Individual Tree Crown (ITC) Delineations:*

Individual tree crown (ITC) delineates are generated by IDTReeS research group. Each delineation is a 2-D rectangular bounding box defining the maximum tree crown extent in an image, and is provided in vector format as ESRI shape files.

*Location of the Tree Crown Data*

The data consist of three NEON ecoclimatic sites in Eastern United States. In other words, each site is characterized by distinctive environmental, geographic, and vegetative properties. The sites are:

* **Ordway-Swisher Biological Station, Florida (OSBS):** The region contains mixed forests of hardwood and conifers, mostly dominated by pine trees.
* **Talladega National Forest, Alabama (TALL):** Forests made of mixed hardwood and conifers (mostly pine) in the Ozarks complex.
* **Mountain Lake Biological Station, Virginia (MLBS):** The region is mainly made of hardwood forests in the Appalachians and Cumberland Plateau.

*Training Data:* Contain tree crown delineations with tree class (i.e., taxonomic species) labels.

*Testing Data:* Tree clown delineations without tree class labels.

**Method:**

*The Objective:* Determine the probability that each ITC belongs to a species class. Two options: pixel-level or object-level classification.

* If we apply a pixel-level classification models must be upscaled to the crown.
* Narrow the classes we test for, and just add an ‘Other’ class for those that don’t fit.
* Test our model by the classification of species on bounding boxes of unknown species identity.

*Preparing the Data:* Clip the HSI raster files to single ITC with the corresponding species label.

**Results:**

*Performance Metrics:* F1 Scores (sklearn.metrics.f1\_score); Average Cross-Entropy Loss (sklearn.metrics.log-loss); Confusion Matrix (sklearn.metrics.confusion\_matrix).

**Conclusion:**

**References:**

*Neon Data References*

National Ecological Observatory Network. 2020. Data Product DP1.30010.001, High-resolution orthorectified camera imagery. Provisional data downloaded from http://data.neonscience.org on March 4, 2020. Battelle, Boulder, CO, USA NEON. 2020.

National Ecological Observatory Network. 2020. Data Product DP1.30003.001, Discrete return LiDAR point cloud. Provisional data downloaded from http://data.neonscience.org on March 4, 2020. Battelle, Boulder, CO, USA NEON. 2020.

National Ecological Observatory Network. 2020. Data Product DP1.10098.001, Woody plant vegetation structure. Provisional data downloaded from http://data.neonscience.org on March 4, 2020. Battelle, Boulder, CO, USA NEON. 2020.

National Ecological Observatory Network. 2020. Data Product DP3.30015.001, Ecosystem structure. Provisional data downloaded from http://data.neonscience.org on March 4, 2020. Battelle, Boulder, CO, USA NEON. 2020.