Update on agroforestry detection

09-2023 : 12-2023

* Established a project [github repository](https://github.com/GeospatialCentroid/Agroforestry) for maintaining code base and tracking of development milestones.
* Developed a presence absence sampling methodology for collection training and validation data for the modeling process.
  + Image 1
* Created and tested an ensemble modeling methodology with includes.
  + Imagery segmentation classification of NAIP imagery
  + Pixel based classification of NAIP imagery
  + Figure 1
* Developed a testing methodology to compare how currently modeling efforts compared against the existing 2016 classification of agroforestry in Antelope county Nebraska )
  + Image 2
* Established an iterative parameter testing process for model tuning.
  + See gatheredData\_2023-12-19.csv
  + Figure 2
* Generate a series of development priorities based on review meeting with project partners in 12/2023
  + Increase the input parameters to the models to include some GLCM texture-based features.
  + Utilize a 10% canopy coverage mask as defined by the NLCD canopy cover layer to exclude areas already accounted for by the forest land inventory.
  + Include some census-based reference layers to exclude built environments.
  + Evaluate applying color balancing to NAIP imagery when apply models to areas outside of the region in which they were trained.
  + Evaluate using center pivot reference data set to help exclude these regions from the modeling process.

An aerial view of a city

Description automatically generated

Image 1 : Example of the presence absence sampling structure using NAIP imagery. Red points are absence. Green points are presence.

A diagram of a company

Description automatically generated

Figure 1: Simplified version of the model methodology. NAIP is the sole parameter input for both the pixel and cluster-based models. The random forest models are trained independently. The models are combined so that area where both outputs defined tree cover because the final representation. The ensemble model can be validated against an independent testing split of the presence absence data or when possible, the USDA reference layer.

A map of different colored squares

Description automatically generated

Image 3 : A confusion matix based comparison of the initial draft model outputs and the existing usda reference layer. These initial results show that the USDA reference model is prediciting more tree cover. The current draft model is misclassifying some center pivot agricultural areas as tree cover.

True negative : Neith models defined tree cover

False positive : Current draft model defined tree cover but USDA reference did not

False Negative : Current draft model did not defined tree cover but USDA reference did

True positive : Both models defined tree cover

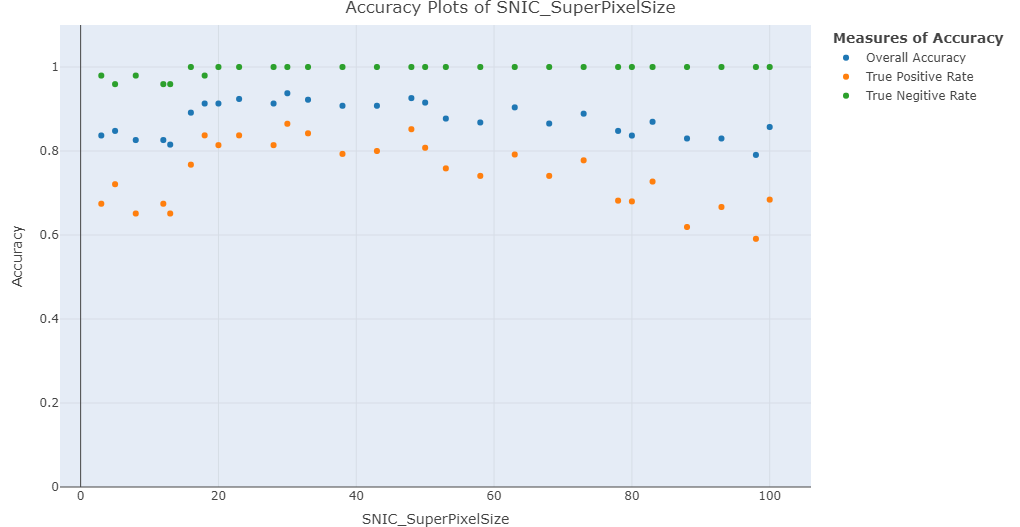


Figure 2 : Results of iteratively change the Super Pixel Size parameter of the cluster model. This parameter sets the initial distance between the clusters starting at 3m and moving to 100m. As this value increase the size of the clusters is expected to increase. While there is variability presence the general trend is peaking near 30 meters then gradually decreasing. The ‘True positive rate’ (how offer we correctly classify a presence point as a tree) is the most sensitive parameter for evaluation.