Machine Learning Approaches to Wetland Mapping in Google Earth Engine

Abigail Stone

CS 701, Fall 2019

Department of Computer Science

Introduction

Google Earth Engine is a cloud-based geographic information system that allows for rapid processing of large remotely sensed geospatial datasets. The Earth Engine API, available in both Python and JavaScript, includes several out-of-the-box image processing tools and allows for extensive customization of analytic algorithms. Machine learning classifiers are becoming more and more prevalent in remote sensing research, especially for environmental monitoring.

Wetlands are an incredibly important natural resource, though they are frequently under-mapped or improperly classified because of their dynamic nature. Wetlands have a very particular topographic signature, easily recognizable to the human eye on topographic datasets and some optical imagery.

The goal of this project is to investigate the accuracy of a few different machine learning techniques in generating wetland classifications based on remotely sensed datasets.

Methods

Google Earth Engine provides direct access to several national and worldwide geospatial datasets; other datasets were imported from the Vermont Open Geodata Portal and Vermont LiDAR (Light Detection and Ranging) program.The following datasets were used as input to the classifiers:

- Digital Elevetion Model derived from LiDAR elevation data from the Vermont Open Geodata Portal (2017, hydro-flattened)
- Topographic Position Index (TPI) derived from LiDAR data (above)
- Optical imagery from the National Agricultural Inventory Program (2014)
- Normalized Difference Vegetation Index (NDVI) derived from the red and infrared bands of NAIP imagery (above)

Existing vector polygon wetland delineations from the State of Vermont's Agency of Natural Resources were rasterized and then used to define sampling regions to train the classifier and as a basis for comparison for the classifier outputs.

Figure 1. Infrared false-color image derived from NAIP

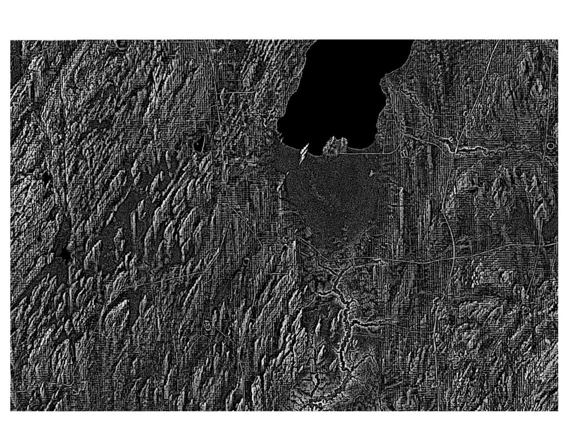

Figure 2. Topographic Position Index derived from LiDAR

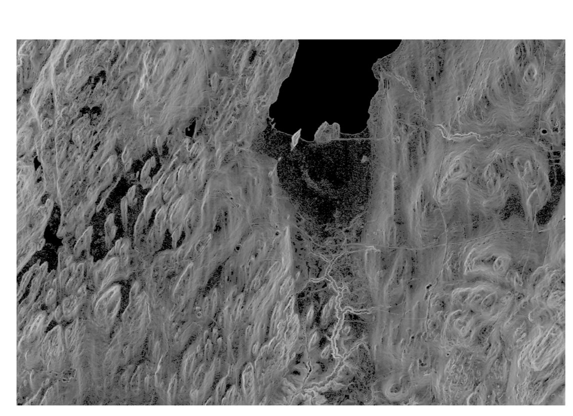

Figure 3. LiDAR-derived elevation model

Google Earth Engine handles rasterized data as Images and ImageCollections. Since the classifiers are more easily customizable with Image inputs, each index was treated as a single band and concatenated into one multi-band input Image object.

Two different classifiers (Support Vector Machines and Random Forests) were evaluated on a region in the Mississquoi River Basin in Franklin County, Vermont. This region has some of the most recent and thorough wetland mapping available from the State of Vermont, and has a high diversity of wetland types.

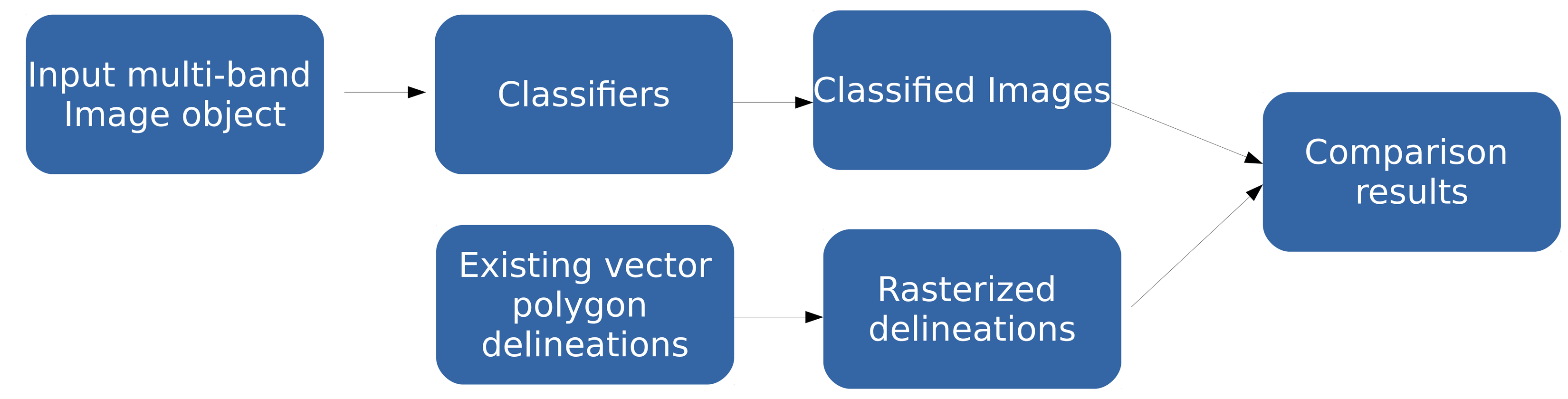

Figure 4. Workflow summary for classifications

Results

In order to quantify the results of each classifier, the rasterized outputs were reduced to raster images containing valid pixels only where wetlands were predicted to be present. These reduced images were compared via Boolean operations to the rasterized delineation polygons from the Vermont Significant Wetlands Inventory.

Figure 5. SVM classifier output

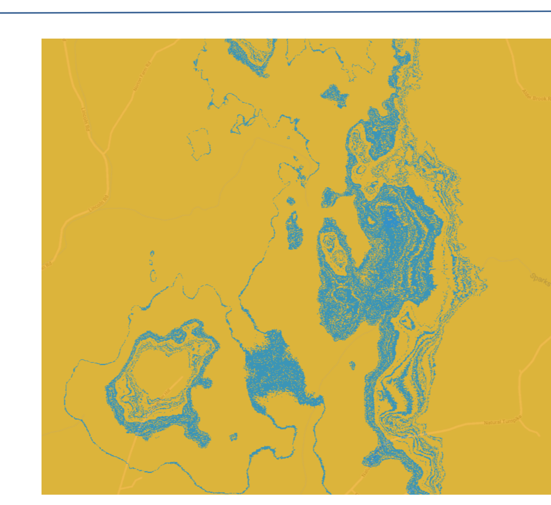

Figure 6. Random Forest classifier output

Preliminary comparisons indicate that the SVM classifier correctly classifies 76% of the known wetland pixels in the study area as wetland. The Random Forest classifier correctly classifies approximately 65% of these pixels.

The SVM classifier misclassifies an additional 10% of the total study area as wetland in areas that are known to be upland; the Random Forest classifier misclassifies an additional 7% of these pixels.

Further comparisons with smoothed versions of the rasterized output may reduce errors in areas where a very small set of pixels are classified as wetlands, or where there are a few upland pixels in an otherwise large wetland complex.

Conclusion and Future Work

Preliminary comparisons show that the SVM classifier correctly identifies more pixels than the RF classifier, however, the RF classifier has fewer misidentifications. Further comparisons with smoothed data may help identify sources of error in these classifications.

In the future, vector polygon outputs would make for easier integration with existing wetland delineation datasets. If these techniques were ever to be used to supplement current mapping, a smoothed vector polygon would need to be generated. For the purposes of this project, the raster output made for a much easier quantitative comparison

References

1. Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2(2), pp.121-167.
2. Hird, J., DeLancey, E., McDermid, G. and Kariyeva, J., 2017. Google Earth Engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. Remote Sensing, 9(12), p.1315.
3. Hsu, C.W., Chang, C.C. and Lin, C.J., 2003. A practical guide to support vector classification.
4. Huang, C., Davis, L.S. and Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. International Journal of remote sensing, 23(4), pp.725-749.
5. Lang, M., McCarty, G., Oesterling, R. and Yeo, I.Y., 2013. Topographic metrics for improved mapping of forested wetlands. Wetlands, 33(1), pp.141-155.
6. Lary, D.J., Alavi, A.H., Gandomi, A.H. and Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geoscience Frontiers, 7(1), pp.3-10.
7. Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment, 62(1), pp.77-89.
8. Veltkamp, R.C., 2001, May. Shape matching: Similarity measures and algorithms. In Proceedings International Conference on Shape Modeling and Applications (pp. 188-197). IEEE.