

Project Title: Classifying Burn Scars Using Machine Learning & Multi-Spectral Data

Participant name: John Miller

Date: July 31, 2024

Abstract:

This project aims to combine multi-spectral satellite data with machine learning to find and quantify damage caused by wildfires within a study area in southern Labrador/Quebec.

1. Introduction

As the planet's climate changes, extreme weather events have become more common. A hotter and drier climate has led to an increase of wildfires, with Canada seeing a record number of 16.5 million hectares burnt in 2023. It's important to be able to quantify the amount of damage caused by wildfires. It would be impossible to manually measure the extent of damage in a large country like Canada, so there needs to find a way to automate it. This project will look at how to quantify wildfire damage using a combination of machine learning with python and multi-spectral satellite data.

2. Background

Objects absorb and reflect radiation from the sun in different ways. This makes it possible to detect different objects on the Earth's surface by analyzing their spectral signatures. Healthy vegetation strongly reflects near infrared light and absorbs shortwave infrared light, while burned areas are the opposite. This knowledge can be used to determine what areas have been damaged by wildfires. A multi-spectral satellite, such as Sentinel-2, carries an optical sensor that can sample several spectral bands. Sentinel-2 collects 13 bands. The near infrared and shortwave infrared bands collected by Sentinel-2 can be used for analysis.

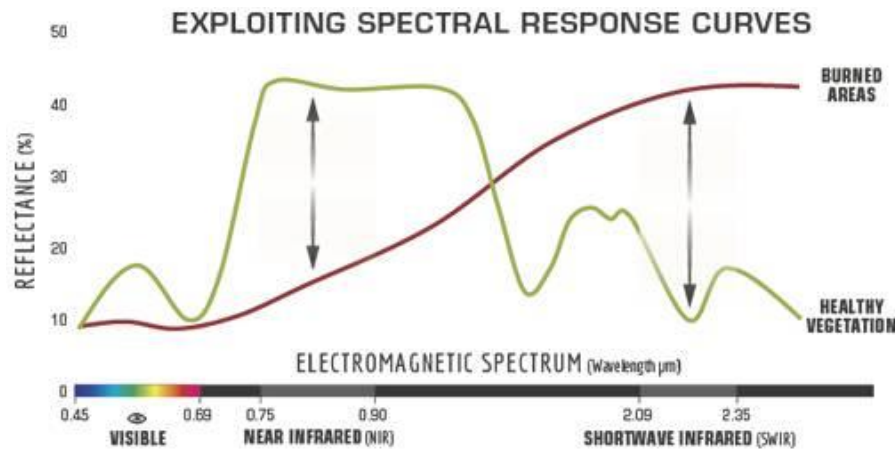


Figure 1 - Reflectance Curve of Healthy Vegetation vs Burned Areas (Source: US Forest Service)

3. Analysis

First, a study area had to be chosen. The Canadian Wildland Fire Information System contains data of wildfires that occurred across Canada in polygon format. The study area, located in southern Labrador/Quebec, was chosen based on a fire that occurred in June 2023 at 58.933°W 52.0869°N. Additional polygon data of previous fires was also retrieved and clipped to the study area. This polygon data will be used to compare results and help create training data. Sentinel-2 satellite data was downloaded from the Copernicus Browser. A dataset with low cloud coverage dated September 22, 2023, was found, approximately 3 months after the fire occurred.

The two bands used for analysis are band 8A, a near infrared band, and band 12, a shortwave infrared band. A scene classification layer from the dataset will also be used. These bands have a 20 m spatial resolution, making the dataset 5490x5490 in size. The python library Rasterio will be used to read in the satellite data as an array. Bands 8A and 12 will be used to create a Normalized Burn Ratio (NBR). The NBR is an index that is used to highlight burnt areas, with burnt areas having lower values. One issue with the NBR is that water will also return lower values. This is because water has a low spectral reflectance. The scene classification layer classifies water pixels as 6, making it possible to create a water mask.

$$\text{NBR} = \frac{\text{NIR}-\text{SWIR}}{\text{NIR}+\text{SWIR}}$$

Equation 1 - How to Calculate the Normalized Burn Ratio

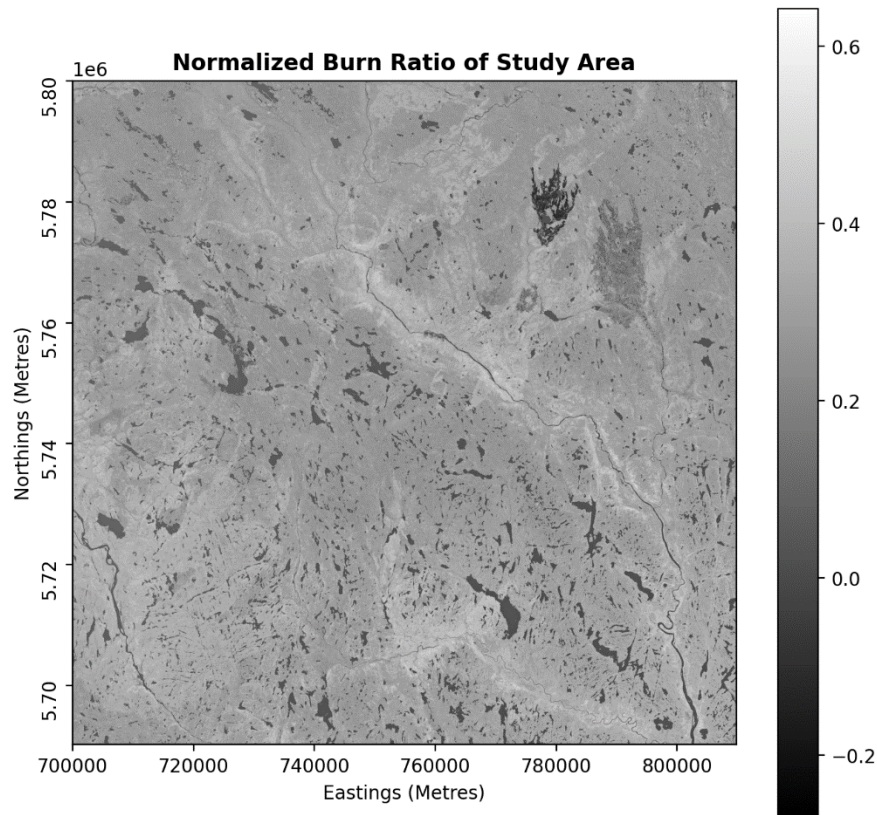


Figure 2 - Normalized Burn Ratio of Study Area

Classifications are performed on the NBR. Four different classifications were used for this study – unsupervised classification (K-Means Clustering), and three supervised classifications (Support Vector Machines (C-Support Vector Classification), Random Forest, and Decision Trees). Unsupervised was chosen to see how well it worked without training data. Previous studies have tested SVM and Random Forest and how they apply to burn scar classifications. Decision Trees were chosen as a fourth classification method for comparison. The classifications classified pixels as either 0 (Untouched) or 1 (Burn Scar). The water mask was then applied to convert the water pixels (which were classified as 1) to 0. The scene classification layer misses some water pixels, leaving the resulting classified raster datasets with noise. The scikit-image

library can be used to remove pixel clusters of a minimum size. A minimum size of 300 pixels was chosen for this analysis, although 600 pixels was also used during testing.

For the size of the study area, a high-performance computer was not necessary, however, the code was run on Siku to see how it ran. If a larger area was studied, such as the entirety of Labrador or Canada, then a HPC would be necessary to speed up the process. The library `rioxarray` (a Rasterio extension of `xarray`) would have been tested to see how it handles running raster datasets in parallel.

4. Results

The resulting classification raster datasets all returned similar values, except the unsupervised. The SVM, Random Forest, and Decision Tree classifications all returned 17,000-18,000 hectares (Ha) burnt, but these numbers were on the higher end. In other tests, the returned Ha could be as low as 14,500 Ha. Random Forest and Decision tree always returned the same value. SVM often returned slightly more Ha burnt than RF and DT, however, when testing was done with a minimum size of 600 pixels per cluster, the RF and DT often returned more burnt area than the SVM, suggesting SVM pixel clusters were not as connected as the RF/DT. The unsupervised always returned a much lower value of 11,076.6 Ha.

The supervised and unsupervised classifications performed better in different ways. The supervised classifications returned more water pixels as burnt, but they did a better job picking up more actual burnt area. This was evident in areas where wildfires occurred 10+ years ago and regrowth would have started, as it was possible to train the classifier to include these areas as burnt.

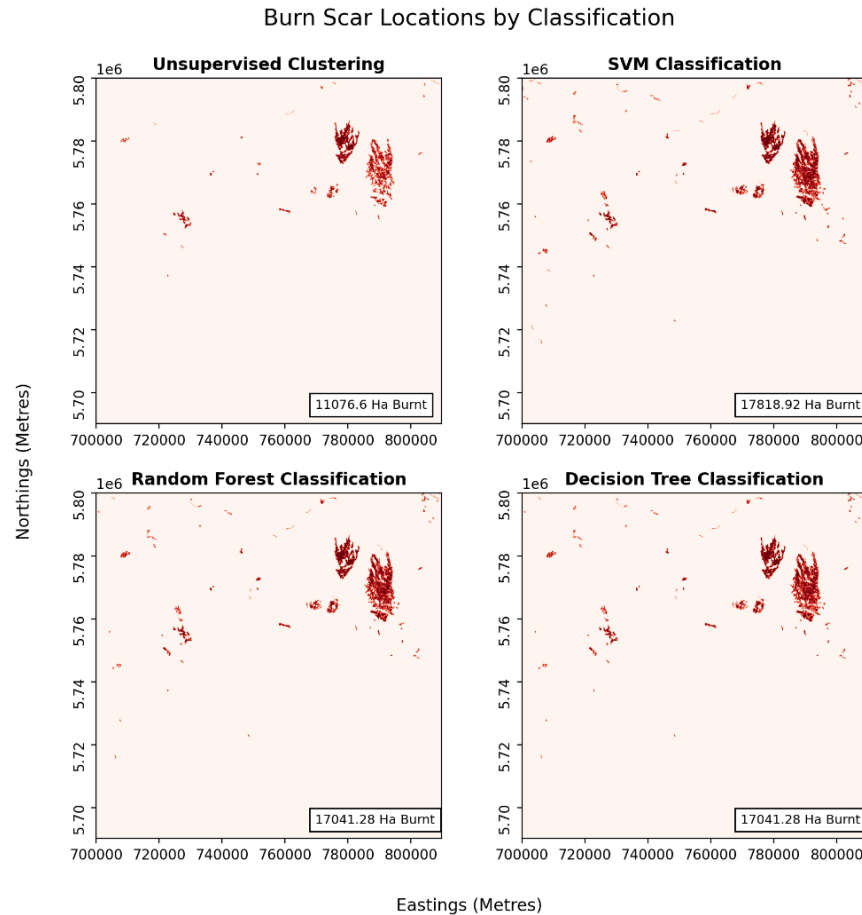


Figure 3 - Burn Scars by Classification

5. Discussion

The classifications did a pretty good job picking up burn scar locations, albeit with varying results. The unsupervised missed a lot of burnt area, but it picked up the freshest burn scar quite well. This suggests that the unsupervised classification is perfectly fine to use if one wanted to find the extent of damage in a short period, such as a year. The supervised classifications were better for overall damage throughout the years. Two issues with supervised classifications would be the water pixels and ground truth data. No ground truth data was used for this analysis – the training data was based on historical wildfire data and the NBR. There was no way to find a piece of land that appeared to be a burn scar but was dry vegetation and train those pixels untouched. Even so, that could skew results if the classifier classified a burn scar pixel as untouched based on dry vegetation. A similar issue would happen if the data was

trained to see water pixels as untouched. Manually cleaning the classified raster dataset to remove falsely classified water pixels would help with smaller datasets but would be impossible with larger study areas. It's important to note that each pixel was 20 x 20 m, covering 400 m² of land, so one pixel could cover an area that contains both burnt and untouched land.

Conclusion

The need to calculate the amount of damage done by wildfires increases as they become more frequent. This study looked at four different classifications, each with their own strengths and weaknesses. For recent fires, unsupervised seems to be the best due to simplicity and results. For damage over several years, a supervised classification is best, as it allows one to train the classifier to consider areas experiencing regrowth. Future improvements in satellite sensors could lead to 10 m resolution data, leading to better results. There are also hyperspectral scanners that cover hundreds of bands. However, hyperspectral satellites have lower spatial resolutions. Hyperspectral imaging would work better mounted on a plane or UAV, but that would not cover as large of an area like a satellite would.

References

Github link: <https://github.com/jvemiller/BurnScarISP>

Presentation: <https://arcg.is/0nPGDb0>

Polygon data: <https://cwfis.cfs.nrcan.gc.ca/home>

Sentinel-2 data: <https://browser.dataspace.copernicus.eu/>

2023 Wildfire background: <https://natural-resources.canada.ca/simply-science/canadas-record-breaking-wildfires-2023-fiery-wake-call/25303>

Normalized Burn Ratio: <https://un-spider.org/advisory-support/recommended-practices/recommended-practice-burn-severity/in-detail/normalized-burn-ratio>

Code references: https://fire.trainhub.eumetsat.int/docs/figure2345_Sentinel-2.html

https://github.com/acgeospatial/Satellite_Imagery_Python/blob/master/Clustering_KMeans-Sentinel2.ipynb

https://github.com/waleedgeo/lulc_py/blob/main/lulc_py.ipynb