Evaluating Wildfire Impact Through Spectral Indices and Machine Learning: A Case Study of the Alessandropoli Fire of 2023 – EARTH OBSERVATION EXAM 2024 (Module B)

Matteo Gobbi Frattini – 10581031

ABSTRACT

Wildfires pose significant environmental challenges, requiring effective monitoring and management strategies. This study employs Earth Observation data, specifically Sentinel-2 imagery, to assess wildfire impacts using spectral indices and machine learning techniques. The analysis focuses on the Alessandropoli Fire of 2023, where the Normalized Burn Ratio (NBR) and its variant, NBR+, were utilized within Google Earth Engine (GEE) to evaluate burn severity. Additionally, Random Forest (RF) and Support Vector Machine (SVM) classifiers were developed to identify burnt areas using only post-fire imagery, a method that proves beneficial when pre-fire images are unavailable or compromised by cloud cover.

The classifiers demonstrated high accuracy when applied to the same fire they were trained on. When tested on different fires, particularly those in different regions or years, the classifiers exhibited a decline in precision and recall. However, improved performance was observed when the classifiers were applied to fires within the same geographical area and wildfire season as the training data.

Keywords: Wildfire analysis, Sentinel-2, Google Earth Engine, Random Forest, Support Vector Machine, Burned Area, NBR

1. INTRODUCTION

Wildfires are increasingly recognized as a critical environmental issue, contributing to substantial ecological damage, loss of biodiversity, and significant economic costs. The frequency and intensity of wildfires have been aggravated by climate change, making the need for effective monitoring and management more pressing than ever. Accurate assessment of burn areas and the severity of fires is essential for ecological recovery, resource allocation, and long-term environmental planning.

Remote sensing technologies, particularly those involving multispectral satellite imagery, have become invaluable tools in wildfire monitoring. Among these, the Sentinel-2 mission provides high-resolution multispectral data that is particularly well-suited for detecting and analysing the impacts of wildfires through various spectral indices. The Normalized Burn Ratio (NBR) and its derivative, the differenced NBR (dNBR), are widely used to quantify burn severity by comparing pre- and post-fire images. However, acquiring high-quality imagery for both pre-fire and post-fire conditions can be challenging due to factors like cloud cover or other atmospheric interferences. While a lack of post-fire imagery can halt the analysis altogether, the study of burn severity can still proceed even if pre-fire images are compromised, by employing alternative methods that rely solely on post-fire data.

To address these challenges, this study explores the use of machine learning techniques to classify burnt areas using only post-fire imagery. By focusing on the post-fire environment, the developed methodology offers a flexible approach that can be applied even in the absence of reliable pre-fire data. The study employs Google Earth Engine (GEE) to streamline the processing of Sentinel-2 imagery and to implement machine learning classifiers, specifically Random Forest (RF) and Support Vector Machine (SVM), for burn area identification.

The Alessandropoli Fire of 2023 serves as the primary case study for developing and testing these methods. Additionally, the generalizability of the classifiers is evaluated by applying them to other fires in different regions and time periods. The results from this analysis provide insights into the applicability and limitations of using post-fire imagery for wildfire assessment, with implications for broader environmental monitoring practices.

2. MATERIALS AND METHODS

Study Area

The primary focus of this study is the Alessandropoli Fire of 2023, located in northeastern Greece. This fire was selected due to its significance, being one of the largest wildfires ever recorded in Europe (Euronews.com, 2023), and the availability of high-quality post-fire satellite imagery. In addition to the Alessandropoli Fire, the study also analysed several other wildfire events to assess not only the generalizability of the machine learning classifiers but also the robustness and adaptability of the workflow code template developed for this analysis. These additional fires include the Rhodes Fire of 2023, another significant event in Greece, and two fires that occurred in Alberta, Canada, during the 2019 wildfire season.

Data Collection

This study utilized Sentinel-2 multispectral imagery, provided by the European Space Agency (ESA), as the primary source of data for analysing the wildfires. Sentinel-2 imagery was employed to capture both pre-fire and post-fire conditions, allowing for a detailed assessment of burn severity through the calculation of the Normalized Burn Ratio (NBR) and its variants.

In addition to Sentinel-2 data, the Global Forest Loss due to Fire dataset, developed by the Global Land Analysis & Discovery (GLAD) group, was used to supplement the analysis (Tyukavina, A., Potapov, P., Hansen, M.C., Pickens, A., Stehman, S., Turubanova, S., Parker, D., Zalles, V., Lima, A., Kommareddy, I., Song, X-P, Wang, L. and Harris, N. 2022). This dataset provided crucial information on the extent and distribution of forest loss due to fire, which was integral for supporting the burned area assessments derived from Sentinel-2 imagery.

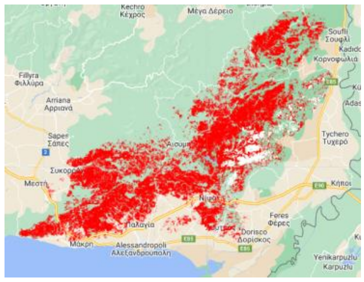

Figure 1. Global Forest Loss due to Fire.

To accurately determine the fire duration and to identify the optimal time interval for capturing post-fire imagery, a custom script was developed. This script was designed to manually analyse the changes in the delta NBR (dNBR) across various time intervals following the fire event. By systematically evaluating the dNBR over different periods, the script allowed for the selection of the most representative "after" image, ensuring that the analysis captured the full extent of the fire's impact while minimizing the influence of post-fire recovery or other disturbances.

This approach ensured that the selected post-fire images were not only timely but also reflective of the maximum burn severity, thereby improving the accuracy of the overall assessment.

Image Preprocessing

To enhance the accuracy of the wildfire analysis, several preprocessing steps were applied to the Sentinel-2 imagery:

Cloud Masking: The QA60 band was used to identify and mask clouds and cloud shadows, ensuring only clear pixels were included in the analysis.

Water Masking: The Normalized Difference Water Index (NDWI) was calculated using the green (B3) and near-infrared (B8) bands to mask water bodies, preventing them from being misclassified as burned areas.

Mosaic Creation: For both pre-fire and post-fire periods, mosaics of the imagery were created to provide a continuous and comprehensive view of the area before and after the fire.

This preprocessing ensured that the analysis was focused on land surfaces where burn scars would be visible.

Index Calculation

The assessment of burn severity and its ecological consequences demanded the quantification of key spectral indices.

To this end, the Normalized Burn Ratio (NBR) and the differenced NBR (dNBR) were crucial in detecting the extent of fire-induced alterations (UN-SPIDER, 2023).

$$NBR = \frac{NIR - SWIR}{NIR + SWIR} \quad (1)$$

$$dNBR = NBR_{before} - NBR_{after} \quad (2)$$

In addition to the standard NBR, an alternative index, NBR+, was utilized (Alcaras, E.; Costantino, D.; Guastaferro, F.; Parente, C.; Pepe, M.; 2022). This index offers enhanced sensitivity to burn severity under certain conditions.

$$NBR+= \frac{SWIR - NIR - Green - Blue}{SWIR + NIR + Green + Blue} \quad (3)$$

$$dNBR+= NBR +_{before} - NBR +_{after} \quad (4)$$

Both indices were computed for pre- and post-fire images, and their differenced forms were used to highlight the areas impacted by the fire.

Threshold Application

Thresholds for both dNBR and dNBR+ were manually selected to delineate the burned areas effectively. These thresholds were chosen based on careful visual inspection and comparison with reference data, ensuring that the identified burn areas accurately reflected the extent of fire-induced changes. This manual approach allowed for a tailored analysis that accounted for the specific characteristics of each fire event (Storey, E.; West, K.; Stow, D.; 2021)

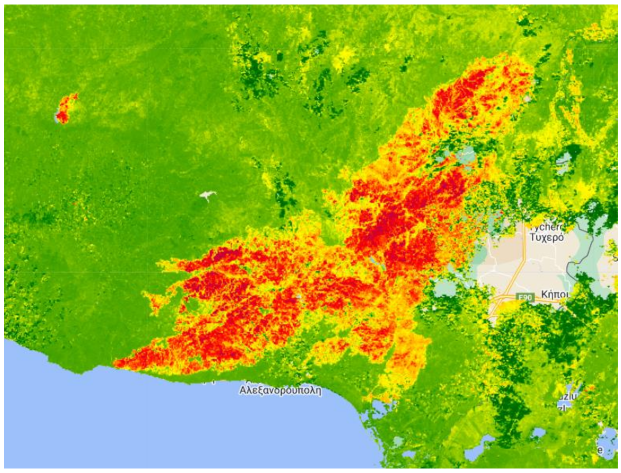

Figure 2. dNBR, Alessandropoli Fire.

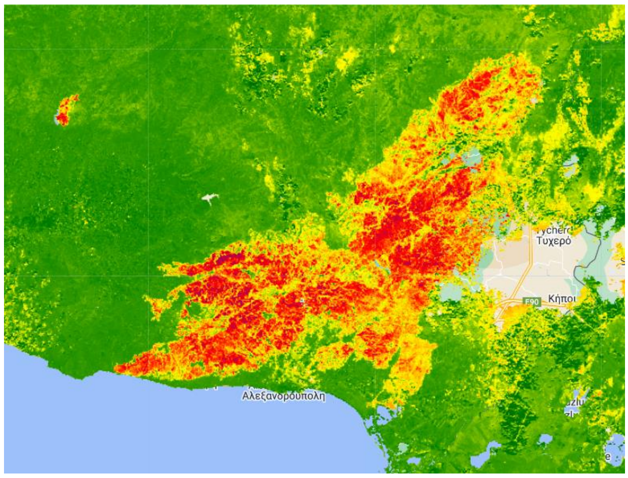

Figure 3. dNBR+, Alessandropoli Fire.

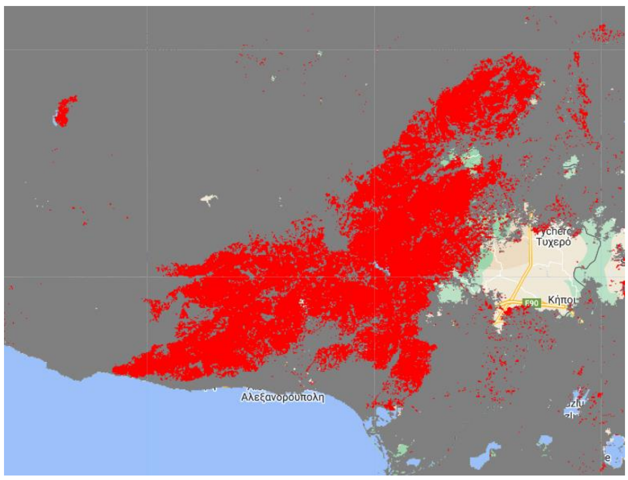

Figure 4. dNBR threshold classification, Alessandropoli Fire.

Machine Learning Classification

To enhance the burn area assessment, two machine learning classifiers, Random Forest (RF) and Support Vector Machine (SVM), were employed. These classifiers were trained using post-fire imagery, with the manually classified burn areas from dNBR serving as the ground truth for training. The choice to use dNBR as the ground truth, rather than the GLAD Global Forest Loss dataset, was made because the GLAD dataset tends to underestimate the extent of fires. This underestimation occurs because GLAD only considers pixels that were classified as forest and that are no longer classified as forest at the end of the year due to fire damage. In contrast, dNBR captures all burned areas, including those in regions that were not classified as "forest" proper, thus providing a more comprehensive representation of the fire's impact.

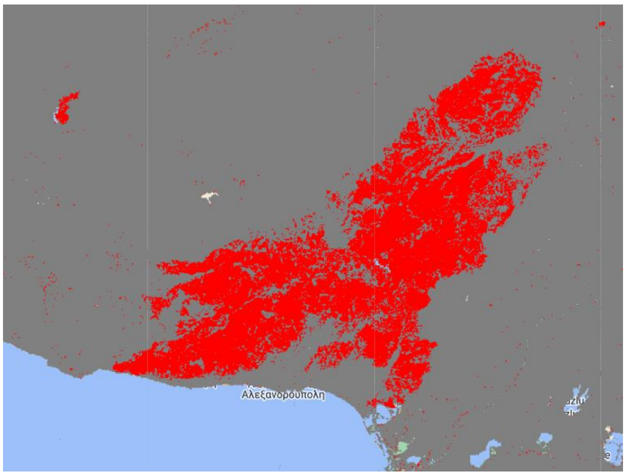

Figure 5. Random Forest classification, Alessandropoli Fire.

For the machine learning algorithms, specific parameters were chosen to optimize performance. The SVM classifier was implemented with a linear kernel, as a study on SVM for burned area assessment demonstrated that the linear kernel was the best-performing option in this context (Petropoulos, G.P.; Kontoes, C.; Keramitsoglou, I.; 2011). The Random Forest classifier was used with default parameters, but with the number of trees set to 100, which is a common choice that balances computational efficiency with classification accuracy.
The classifiers were trained on a subset of pixels within the area of interest (AOI) and were subsequently applied to the entire AOI to generate detailed burn area maps. This process not only provided a more refined classification but also helped to fill the gaps in the dNBR maps caused by masking issues in the "before" images. These gaps often arise due to cloud cover or other obstructions in the pre-fire imagery, and the use of machine learning allowed for a more continuous and accurate representation of the burned areas.

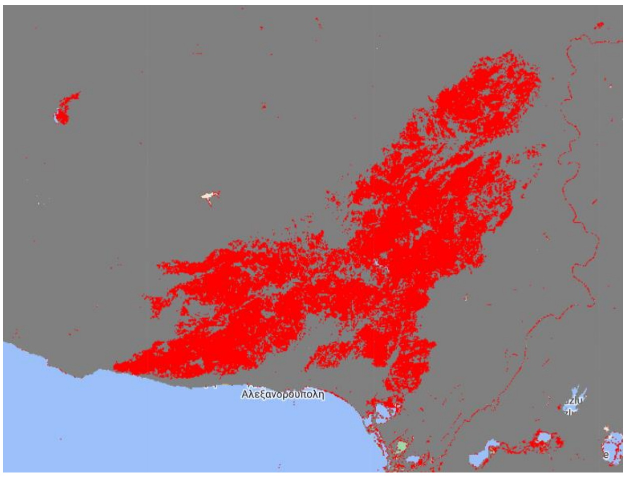

Figure 6. SVM classification, Alessandropoli Fire.

Model Evaluation
The performance of the RF and SVM classifiers was evaluated using key metrics such as accuracy, precision, and recall. These metrics provided insight into the effectiveness of the models in identifying burned areas, particularly when applied to different regions and fire events

3. RESULTS DISCUSSION

This section presents the findings from applying various burn area estimation methods across different wildfire events. The discussion begins with an analysis of the Alessandropoli and Rhodes fires, which occurred in close geographical proximity and within a short temporal span. Both fires were analysed using the same workflow and the same Random Forest (RF) classifier, trained on the Alessandropoli fire. It is important to note that, while the RF classifier could be exported for application to other areas, the Support Vector Machine (SVM) classifier could not be exported using Google Earth Engine (GEE) as of now.
The methods are then applied to wildfires in other regions, specifically two fires in Alberta (Canada), which differ in geography and vegetation from the first two fires. The results from these different environments are compared to assess the performance of the workflow and models under varying conditions.

Alessandropoli and Rhodes Fires
The burned area estimates for the Alessandropoli and Rhodes fires, as derived from various methods, are summarized in Table 1. The dNBR and dNBR+ methods provided higher estimates of burned area compared to the GLAD dataset, which significantly underestimated the total affected area, particularly for the Alessandropoli Fire. This discrepancy is likely due to GLAD's focus on forested areas only.

Method	Estimate (hectares)	
	Alessandropoli	Rhodes
GLAD	45180	5324
dNBR	77958	13376
dNBR+	82733	11523
RF	74694	13970
SVM	73306	
Declared	81000	13500

Table 1. Alessandropoli and Rhodes burned area comparison. (Euronews.com and Greekreporter.com, 2023)

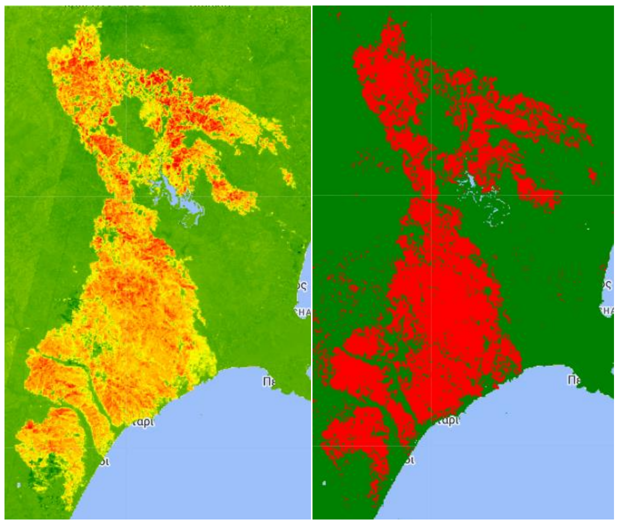

Figure 7. dNBR (left) and Random Forest classification (right), Rhodes Fire.

The Random Forest (RF) classifier produced estimates that were consistent with the dNBR and dNBR+ results, with the RF classifier slightly underestimating the burned area in comparison to the declared official figures. The Support Vector Machine (SVM) classifier also provided estimates in a similar range, though with slight variations.

For the Alessandropoli Fire, the RF classifier achieved a validation accuracy of 0.96, and the SVM classifier showed a similar performance, with a validation accuracy of 0.95. Both confusion matrixes are shown in figure 8.
When applied to the Rhodes Fire, which occurred about a month later in a different part of the Aegean region, the RF classifier—despite being trained on the Alessandropoli data—performed reasonably well. The RF classifier achieved an accuracy of 0.96, with a precision of 0.89 and a recall of 0.95.

```
Validation Error Matrix RF:
[[855,16],[23,190]]
  0: [855,16]
  1: [23,190]

Validation Error Matrix SVM:
[[854,17],[36,177]]
  0: [854,17]
  1: [36,177]
```

Figure 8. Alessandropoli Model Confusion Matrixes.

Alberta Region
The general workflow, which included dNBR and dNBR+ classifications, was applied to two different 2019 wildfires in the Alberta region to assess its effectiveness across multiple fire events within the same geographic area. The results from these analyses are summarized in Table 2. The dNBR and dNBR+ methods provided consistent estimates of the burned areas, demonstrating their robustness in identifying fire-affected regions even across different fires within the Alberta region.

Method	Estimate (hectares)	
	Manning	Slave Lake
GLAD	24482	128348
dNBR	41487	199152
dNBR+	45082	211678
RF	37300	120577
Declared	52868	211869

Table 2. Alberta Fires burned area comparison. (Wikipedia, 2024)

While the exported ML model maintained a relatively high overall accuracy, achieving 0.90 for the Slave Lake fire and 0.92 for the Manning fire, the recall dropped significantly in both cases, to 0.47 and 0.40, respectively. This drop in recall indicates that the model failed to correctly identify a substantial number of burned areas, resulting in a large number of false negatives. Consequently, this led to a severe underestimation of the total burned area in both fires.

Using the same approach applied to the Alessandropoli Fire, new Random Forest (RF) and Support Vector Machine (SVM) models were trained on the Slave Lake fire data. Both models achieved a validation accuracy of 0.93. The confusion matrices for these models, presented in figure 10, provide a detailed view of their classification results.
This approach enabled classification based solely on the post-fire imagery, effectively addressing gaps in the dNBR classification caused by missing or compromised data in the pre-fire image, as visible in figure 11.
The RF model quantified the area of forest loss at 196252 hectares, while the SVM model estimated it at 235995 hectares. These estimates are much closer to the actual reported burned area, demonstrating the improved accuracy of the newly trained models in capturing the extent of the fire's impact.

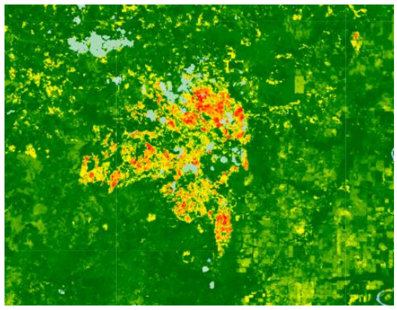

Figure 9. dNBR, Manning Fire.

```
Validation Error Matrix RF:
[[829,26],[46,150]]
  0: [829,26]
  1: [46,150]

Validation Error Matrix SVM:
[[805,50],[27,169]]
  0: [805,50]
  1: [27,169]
```

Figure 10. Slave Lake Model Confusion Matrixes.

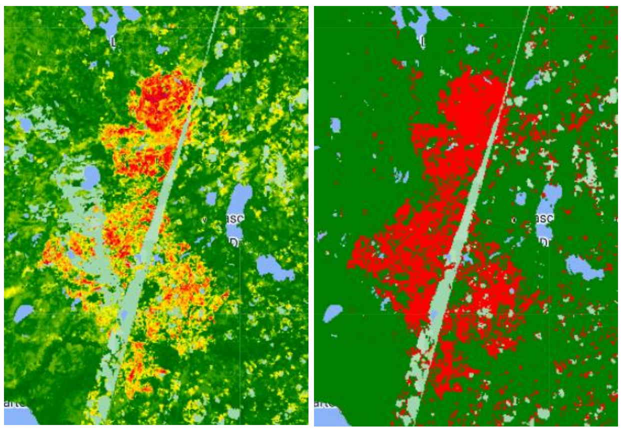

Figure 11. dNBR (left) and Random Forest classification (right), Slave Lake Fire.

4. CONCLUSIONS

This study evaluates the effectiveness of various methods for estimating burned areas across different wildfire events using multispectral satellite imagery and machine learning techniques. The use of dNBR and dNBR+ indices proved to be consistent and reliable in assessing burned areas across diverse geographical contexts, such as the Aegean Sea region and the Alberta fires.

However, when applying machine learning models like Random Forest (RF) and Support Vector Machine (SVM) that were initially trained on one specific region, the study found that their performance significantly declined when applied to different regions, particularly in terms of recall. This resulted in a notable underestimation of burned areas in regions like Alberta.

To address this limitation, the study trained new RF and SVM models on local data from the Slave Lake Fire, which achieved validation accuracies of 0.93. These models successfully filled the gaps left by missing pre-fire imagery, providing a more comprehensive classification of burned areas based solely on post-fire images.

The findings underscore the need for localized training data or more adaptable machine learning models to accurately map burned areas across different environmental conditions. Future research should explore the development of such models and the integration of additional data sources to enhance the accuracy and applicability of wildfire assessments.

REFERENCES

All the code scripts developed for this research are available at the following GitHub repository.

Euronews.com, Wildfire in northeastern Greece is the biggest the EU has ever recorded. 30/08/2023

Tyukavina, A., Potapov, P., Hansen, M.C., Pickens, A., Stehman, S., Turubanova, S., Parker, D., Zalles, V., Lima, A., Kommareddy, I., Song, X-P, Wang, L. and Harris, N. (2022) Global trends of forest loss due to fire, 2001-2019. Frontiers in Remote Sensing

UN-SPIDER, 2023. Office for outer space affairs un-spider knowledge portal. normalized burn ratio (nbr). Online.

Storey, Emanuel & West, Krista & Stow, Douglas. (2021). Utility and optimization of LANDSAT-derived burned area maps for southern California. International Journal of Remote Sensing. 42. 486-505. 10.1080/01431161.2020.1809741.

Alcaras, E.; Costantino, D.; Guastaferro, F.; Parente, C.; Pepe, M. Normalized Burn Ratio Plus (NBR+): A New Index for Sentinel-2 Imagery. Remote Sens. 2022, 14, 1727

George P. Petropoulos, Charalambos Kontoes, Iphigenia Keramitsoglou. Burnt area delineation from a uni-temporal perspective based on Landsat TM imagery classification using Support Vector Machines. *International Journal of Applied Earth Observation and Geoinformation*, Volume 13, Issue 1, 2011

Greekreporter.com, Fires on Rhodes, Greece Burned 13,500 Hectares of Forest. 01/08/2023

Wikipedia, 2019 Alberta Wildfires, 2024