Automatic Detection of Burned Forests in Northern California

Ilan Valencius, Jose Cuevas, Alexander Ronan, Ian Dulin

5-13-21

1 Introduction

In the state of California, in the western United States, wildfires present an increasing risk to both natural and artificial landscapes. During 1972–2018, California experienced a fivefold increase in annual burned area, mainly due to more than an eightfold increase in summer forest-fire extent (Williams et al., 2019). This trend culminated in a burned area of 4.2 million acres in 2020; the result of a prolonged fire season that featured five of the ten largest wildfires in California's history (CAL FIRE, 2020). While the loss of natural habitat is almost innumerable, we can better constrain the toll the fires took on the people of California. Over the course of the year, there were 31 fatalities, and thousands of other displaced. Over 10,000 structures were damaged or destroyed, and the economic losses, in USD, are estimated to be in the hundreds of billions (CAL FIRE, 2020). Unfortunately, contemporary observations indicate that these trends could continue and wildfires in California could increase in magnitude in the coming years. Since the early 1970s, warm-season days have warmed by approximately 1.4 ^{o}C as part of a centennial warming trend, and California continues to experience prolonged-periods of drought-like conditions (Williams et al., 2019).

One of the most important mitigators for the effects of wildfires in California is the appropriate application of various land-management strategies aimed at reducing the risk and severity of fire. Much of the state consists of dry trees, grass, and shrubs. Additional factors, such as the rise of eco-dominant invasive grasses that have a greater propensity to burn, have created a veritable "tinderbox" over much of the state, year in and year out. This is compounded by an increasing percentage of fire genesis coming from anthropogenic causes, such as campfires, downed power lines, arson, and, on occasion, gender reveal parties (Morales & Waller, 2020). Indeed, the increasing interaction between humans and nature has resulted in nearly 85% of wildfires in California occurring as a direct result of people (CAL FIRE, 2020).

Unfortunately, there exist several barriers to proper management, including a lack of resources, permitting, conservation efforts, and public health concerns. Due to these demands, it is therefore important to have efficient means of prioritizing where such mitigating-efforts should be applied. Historically, factors such as existing vegetation, land-cover, climatological data, and existing infrastructure have been used to inform these decisions, and data sources such as satellite imagery have been critical in the evaluation of areas, both pre- and post-burn. Seeking a greater understanding of ecology and land-use as they relate to burn potential and severity is vital in discerning patterns of wildfires, which may aid future efforts toward mitigation. Ultimately, greater access and ease to this information could prove to be a significant advancement in how we coexist with wildfires.

2 Research Questions

This context raises several questions about how geospatial imaging techniques can assist stakeholders in land management. Employing imaging methods to detect burned forests could improve reactive management approaches. If imaging techniques can be used to detect burned forest area, then to what extent is it practical to use them? Are specific programming environments more or less ideal for automatic detection? Is it feasible to employ these methods at a large geographic scale, and if so, what geospatial patterns or trends can be observed from the Northern California wildfires, and what are their implications, in terms of management and impact on local communities? We anticipate that if MATLAB is an appropriate programming environment for this analysis, then it will be able to use hyperspectral classification to automatically detect post-burn forests, which can be verified within MATLAB based on reported fire perimeters.

To address these management concerns, we designed a two-pronged analytical pathway in MATLAB: our algorithm was designed to handle both image classification and comparison. The image classifier is intended to identify burned spaces from spatial data, while image comparison verifies the output based on a known reference image. The MATLAB environment allows us to load image-based data in from a variety of sources. Because this information is stored in the form of arrays, it is possible to perform numerical transformations on image files using existing functions, which is useful in both aspects of our analytical pipeline.

We pooled data from NASA and USGS' LANDSAT-8 data repository as well as the California Department of Forestry and Fire Protection (CAL FIRE). This availed us with both satellite image data and known boundaries of California wildfires on an annual scale.

3 Methods

3.1 Data Sources

1. Landsat 8 L1TP hyperspectral data obtained on 2019-01-02.

   Doi: https://doi.org/10.5066/P975CC9B. The L1TP dataset has been both precision and terrain corrected. Landsat 8 data is one of the most widely available and commonly used hyperspectral data and therefore should provide a general use-case for MATLAB hyperspectral image classification. Bands 1-11 were used which refer to the following wavelengths (nm) [440, 480, 560, 660, 870, 1610, 2200, 590, 1370, 10900, 12010].

2. Perimeter data of major California files from 2015 to 2019. Obtained from CA state GiS geoportal. A culmination of various wildfire datasets, this data was used to determine areas of CA which have been previously burned.

3.2 Image Classification

We employed functions availed by MATLAB's Image Processing Toolbox in order to design our classification pipeline. As discussed, MATLAB represents image data in rectangular or cuboid data formats, either as an array to represent the space the image occupies, or a three-dimensional array for spatial coverage and some other variable, such as light or hyperspectral color bands. Consequently, hyperspectral data is managed as a hyperspectral data cube, which can further be analyzed to output spectral plots, which are in turn interpreted based on the relevant scientific questions.

We were able to subsample, synthesize, and colorize the data provided by LANDSAT and CAL FIRE and ingest that data into this pipeline. Hyperspectral analysis of this image should have resulted in a .png file readable by MATLAB to compare to a known reference image.

3.3 Classification Pipeline

Roughly following the MATLAB hyperspectral pipeline outlined in Fig. 1.

1. Compile LANDSAT 8 L1TP data into a composite raster containing all bands.

2. Crop this raster to region of interest (roi) (fire perimeters inside lat,long bounding box [38, -124, 40, -122], See Fig. 2)

3. Load both rasters as hypercubes.

4. Extract endmembers from roi hypercube (Fig 3).

5. Classify total study area using endmembers from roi (Fig 3).

3.4 Image Comparison

he image comparison section utilizes MATLAB's image processing toolbox to simplify and binarize the image classification output and a chosen reference image to best compare the two. The reference image is a .png file from the California State Geoportal that depicts polygonal shapes that represent the location of known wildfires (Fig. 4.b). The reference image was then altered to mirror the image classification output's dimensions and subsequently binarized (Fig. 4.e). The final reference image used for comparison depicts wildfires as 1 (white) and non-wildfires as 0 (black). The area of the wildfire was found using the bwarea function, and the perimeter was found using the bwperim function. Depending on how the reference image was generated, connectivity parameters of the perimeter could be changed to enlarge the perimeter to include wildfires on the edges that may have been cut out.

The resulting image classification could not be compared unless it was simplified and binarized because it is a .png file in RGB colors (Fig. 4.a). While it is possible in MATLAB to isolate the red values (which mark the locations of wildfires), the binarize function, which needs to be used to make it comparable to the reference image, requires the image not to be shown in RGB. As such, the rgb2gray function was used to convert the RGB colors to their respective gray values (Fig. 4.c). All hues of gray that were not wildfires were removed, and the image was binarized (Fig. 4.d). The altered image classification file depicts wildfires as 1 (white) and non-wildfires as 0 (black).

By binarizing an image, MATLAB is able to recognize individual shapes in an image, as well as perform calculations such as area and perimeter. To calculate the overlap between the reference image and the image classification output, the two images were subtracted from each other since they are in the same dimensions and binarized. We can then compare the resulting area from the original area of the reference image to determine percent overlap between the two.

3.5 Sources of Uncertainty

There are various sources of uncertainty in this study. Initially, the wildfire perimeters used in this study were from a four year range, 2015-2019. This leads to the possibility that forest cover could have returned and thus the spectral signature of "burned" area is less representative.

Other issues arise with the LANDSAT 8 L1TP data. This study did not correct for cloud cover or for pixel quality. This leads to the potential for a less accurate classification, as well as issues binarizing pixel.

Although not an error, parts of this study were done in QGIS (such as merging LANDSAT 8 rasters for import). This does not impact results, however it does highlight the fact that this analysis cannot truly be done in MATLAB *alone*.

3.6 Member Roles

- Ilan Valencius - Steps 1-3 of the classification pipeline. Helped with MATLAB function research for classification as well as figure generation.

- Jose Cuevas - Steps 4-5 of the classification pipeline and provided additional support for the literature review.
- Alexander Ronan - Image comparison code and image comparison methods and results in the Paper. Assisted in Formatting of figures and literature review. Worked together with Ian on Image Comparison.
- Ian Dulin - Researched background information on wildfires in California and developed code to compare classification and reference images with Alex.

4 Results

4.1 Image Classification

Our image classification pipeline produced the desired formats of output files, though the data contents of these products were not as expected (Fig. 3). These results demonstrate that while MATLAB is capable of performing the general schema of image classification, its use is limited by the way the data is manipulated within this environment as well as the nature of our data sources. There are a small number of hyperspectral bands available for analysis, and MATLAB internally calculates the number (Fig. 3). This means that MATLAB is seeking a broad spatial range for a very specific subset of values, which causes runtime to scale up quickly. This analysis is further limited by the output data: a hypercube is a read-only data structure, which means there is no way to add or remove hyperspectral bands for analysis or no way to crop the area being analyzed. An ideal use case for geospatial information in MATLAB is a small, dense data source for investigative purposes, rather than an exploratory use on a geographically broad area (Fig. 5). Classifying aerial photographs of suspected burn regions would be a productive application of our classification schema.

4.2 Image Comparison

Our image comparison method produced desirable results. Even though the model's input was a "fake" classification with no representation to actual wildfires, it was able to compare both of the images. Past work has shown that spatial methods can be used to predict what proportion forested areas are at increased risk of wildfires (Chuvieco & Congalton, 1989). When the location of wildfires from the classification is compared to the reference image, it was found that 7.31% of the reference image's area was filled in by the classification. It was also found that 14.66% of the classification data points fell within the bounds of the reference image (Fig. 6). The Image comparison model, if needed, can alter the bounds of the reference image to account for referencing errors that may have cut off the wildfire border. The connective parameter, which dictates the grouping of pixels at the perimeter, can be changed from 4 to 8, to make the perimeter 2 pixels wide instead of just 1 (Fig. 7). While the change wasn't necessary in this experiment, it could be used in future studies if needed. If the classification was accurate, the image comparison model would be able to calculate the overall overlap between the classification and the reference image. While MATLAB is able to distinguish the different shapes, it is unable to perform isolated analysis on each shape within the reference image. This puts MATLAB at odds with ArcGIS, which researchers have used to monitor the complexities of wildfires in Alaskan boreal forests (Kasischke et al., 1993).

5 Conclusion

This study demonstrated that image classification and verification are indeed *possible* in MATLAB but are perhaps not ideal. MATLAB has difficulty correctly parsing spectra with relatively few numbers of bands (¡100) as seen in the difference between Fig. 3 and Fig. 5. Combined with the lack of other intuitive tools, such as compatibility with shapefiles, MATLAB is not as feasible for large scale classification as other programs such as ArcGIS or QGIS. The ability to not specify the area of classification is also not very useful. MATLAB can also be used to work with previously determined spectra through the ECOSTRESS spectra library compiled by NASA. For this reason we conclude that MATLAB image processing does have a use case, but one that is not applicable to this study: small, localized areas which you are investigating (i.e. you are unaware of what comprises the image).

Classification verification is also possible, however it is rather unintuitive and falls short of other programs. The verification process outlined in this study is essentially comparing the similarity of two .png images, not at all tailored for hyperspectral analysis. Other processes, like the dzetsaka classification tool in QGIS automatically generate, interpret, and georeference classification accuracy. Due to this, we conclude that verification MATLAB is unnecessarily complicated and not tailored for large scale projects.

6 Code Source Statement

1. GitHub: here
2. Presentation: here

7 References

- CAL FIRE. (2021). Perimeter data of major California files from 2015 to 2019 [Perimeter data]. Retrieved from https://frap.fire.ca.gov/mapping/gis-data/.
- Chuvieco, E. & Congalton, R.G. (1989). Application of Remote Sensing and Geographic Information Systems to Forest Fire Hazard Mapping. *Remote Sensing of Environment*, 29(2), 147-159. DOI: https://doi.org/10.1016/0034-4257(89)90023-0
- Kasischke, Eric S., Nancy H. F. French, Peter Harrell, Norman L. Christensen, Susan L. Ustin, and Donald Barry. (1993) "Monitoring of Wildfires in Boreal Forests Using Large Area AVHRR NDVI Composite Image Data," *Remote Sensing of Environment* 45, no. 1: 61–71. https://doi.org/10.1016/0034-4257(93)90082-9.
- Morales, C., & Waller, A. (2020, Sept. 7). A Gender-Reveal Celebration Is Blamed for a Wildfire. It Isn't the First Time. *The New York Times.*

  https://www.nytimes.com/2020/09/07/us/gender-reveal-party-wildfire.html
- USGS. (2019). Landsat 8 L1TP hyperspectral data [Hyperspectral data files retrieved on 2019-01-02]. Retrieved from https://doi.org/10.5066/P975CC9B.
- Williams, A. P., Abatzoglou, J. T.,Gershunov, A., Guzman-Morales, J.,Bishop, D. A., Balch, J. K., & Lettenmaier, D. P. (2019). Observed impacts of anthropogenic climate change on wildfire in California. *Earth's Future*,7, 892–910.

  DOI: https://doi.org/10.1029/2019EF001210

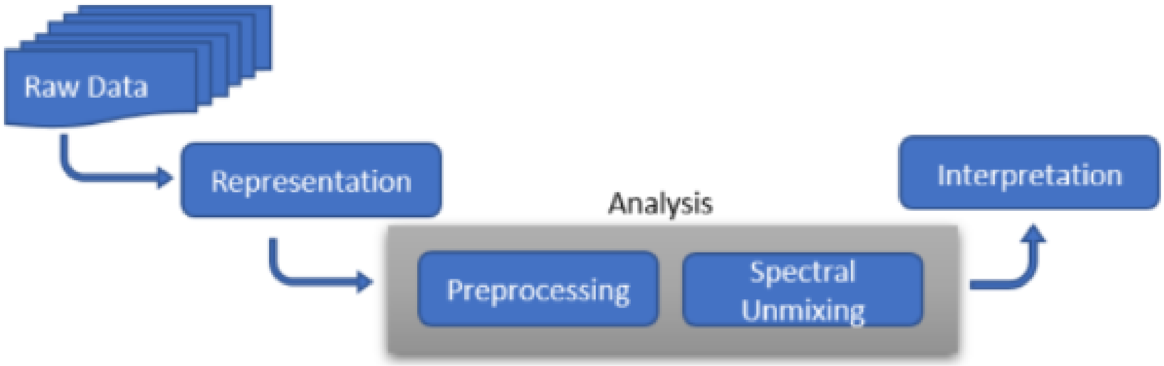

Figure 1: Generalized pathway to classify hyperspectral data in the MATLAB environment.

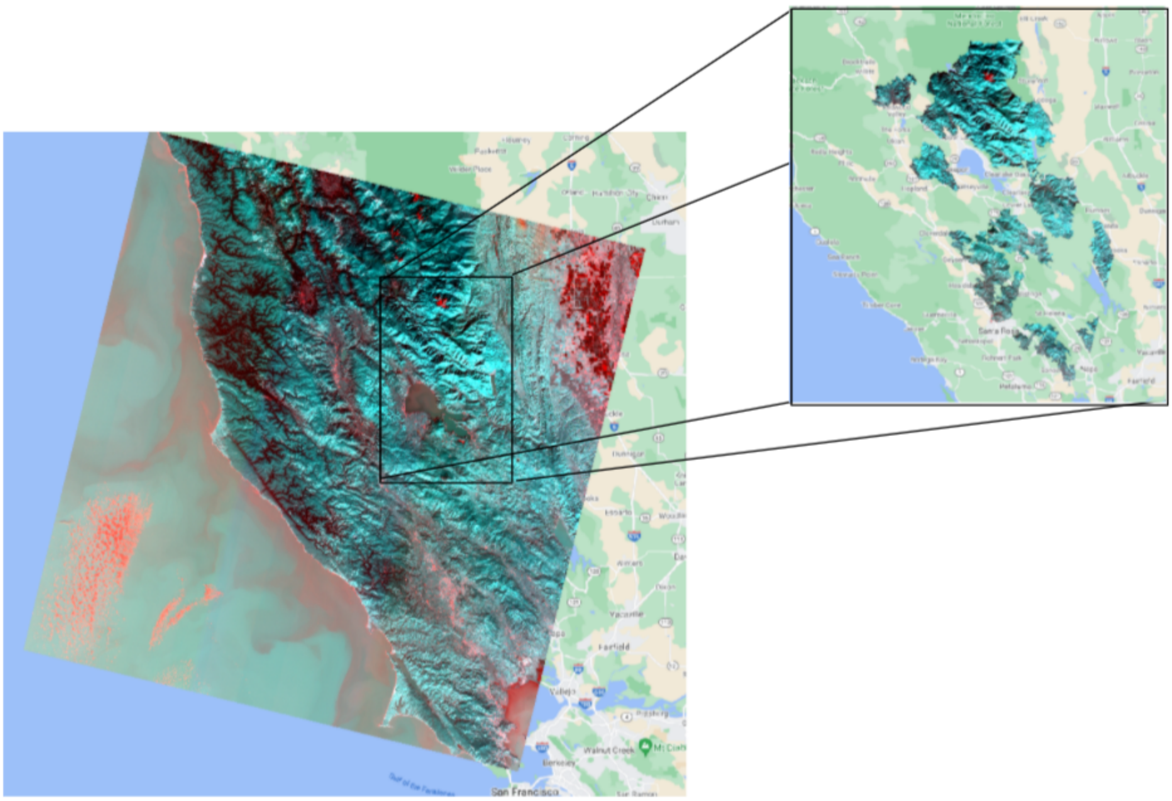

Figure 2: Landsat 8 L1TP colorized data clipped to roi (fire perimeters). This merged raster contained the 11 bands used to perform the classification.

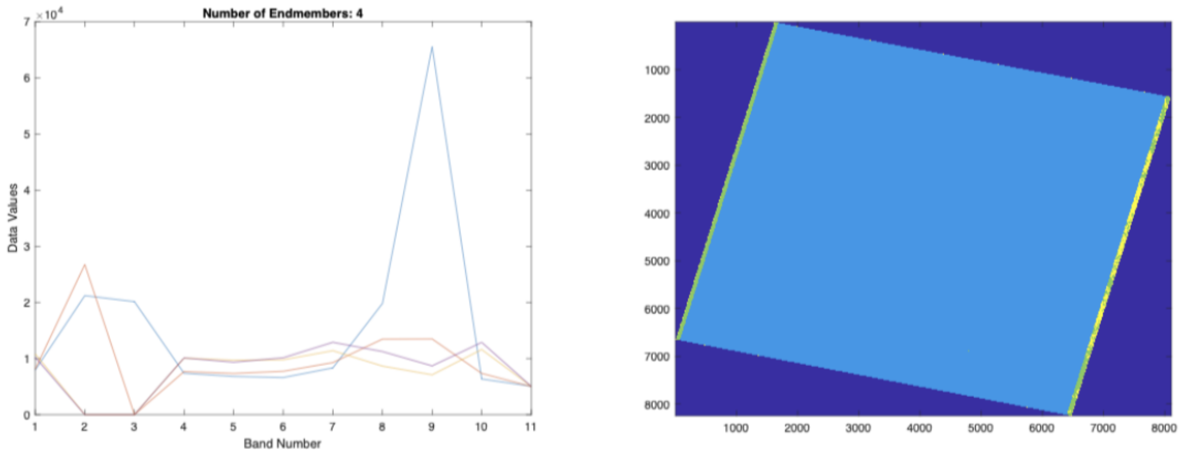

Figure 3: Plot of the 4 endmembers determined for our region of interest. Every endmember is prevalent in a certain band (blue and band 9 for instance). Along with this is the classification map of California determined using the 4 determined endmembers.

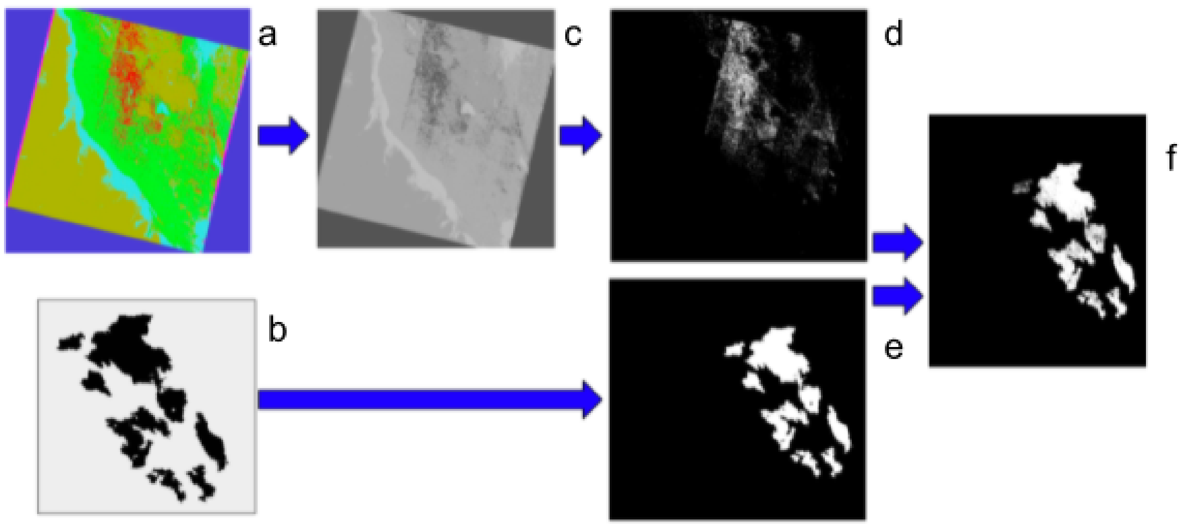

Figure 4: Image comparison methodology. Each subfigure is as follows: a) Fake classification, b) Reference image of burned areas, c) Grayscale of fake classification, d) Isolated and binarized "wildfire" classification, e) Binarized reference image, f) Pixel overlap between binarized reference and classification.

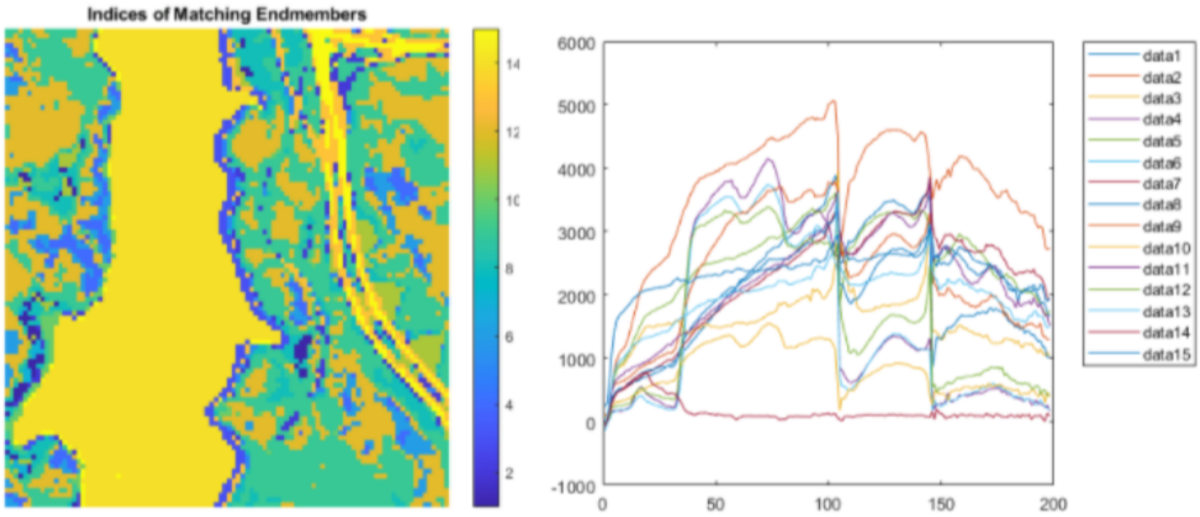

Figure 5: Reference endmember plot and classification of a small ROI of hyperspectral data with 198 bands. The larger amount of bands leads to a more robust endmember plot and thus more accurate classification.

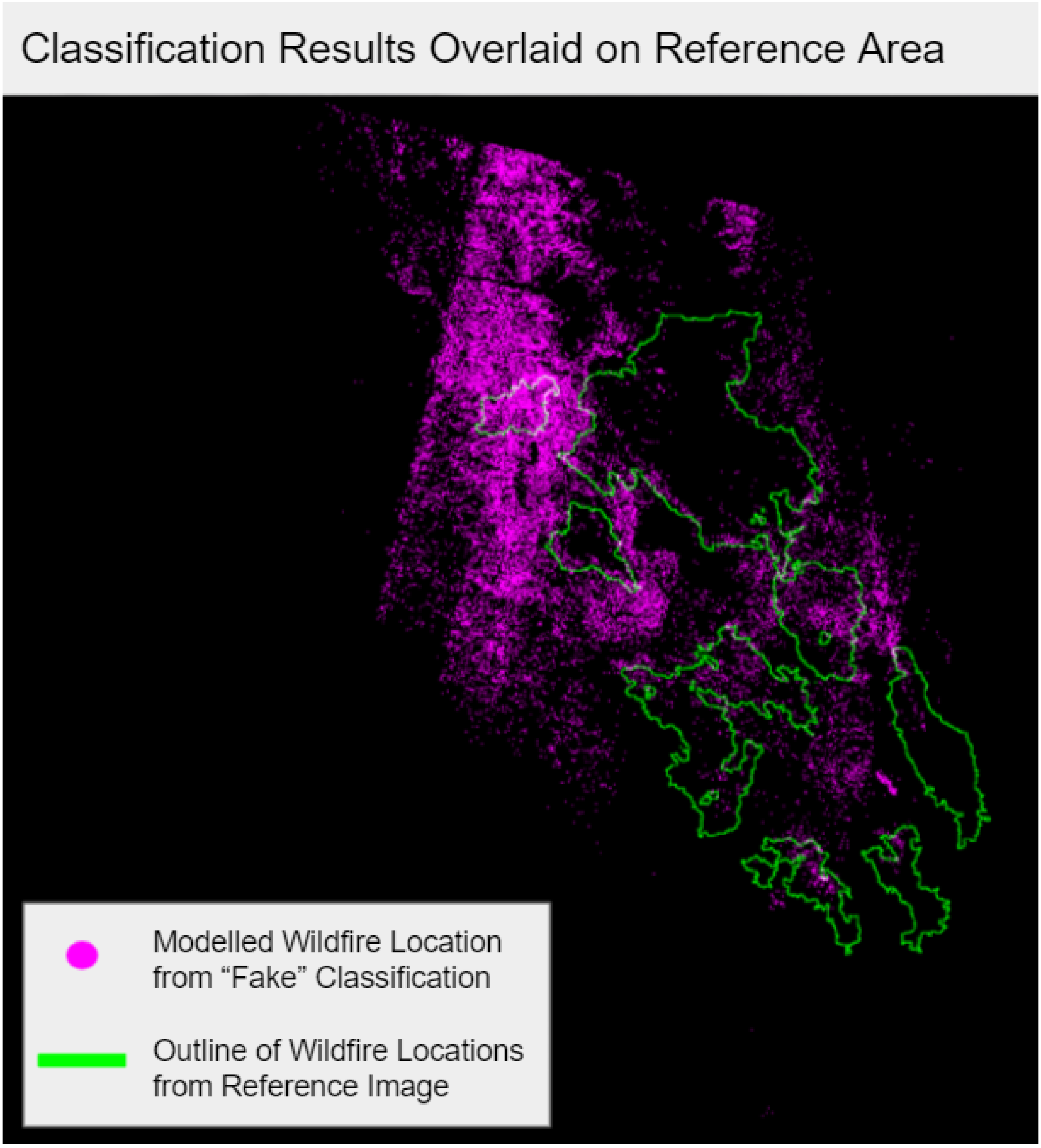

Figure 6: Classification results overlaid onto the outline of the reference image.

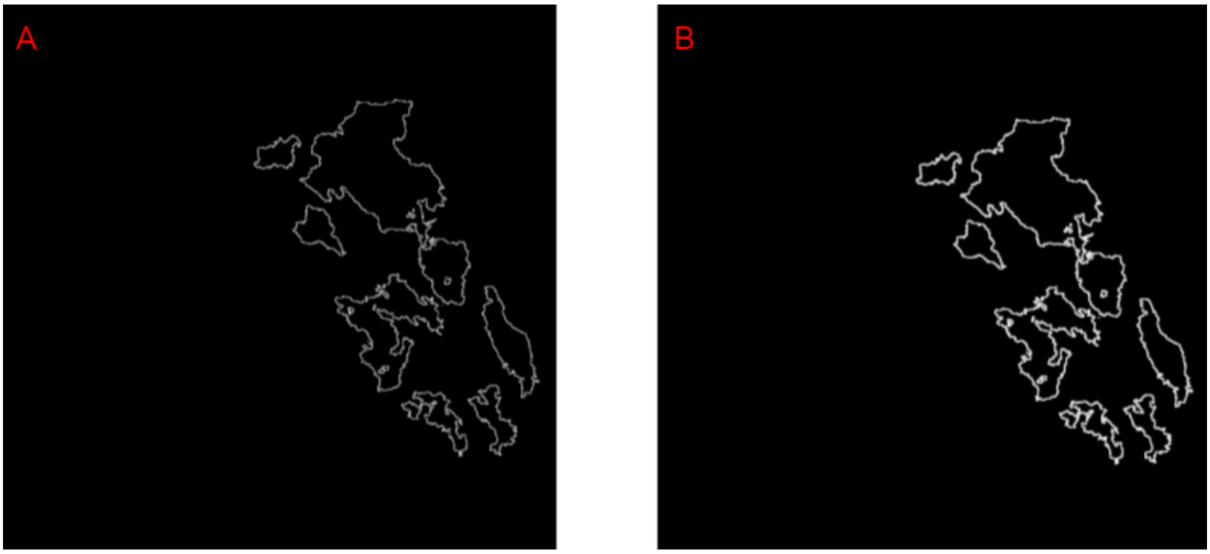

Figure 7: A) Perimeter of reference image with a width of 1 pixel and a connective parameter of 4. B) Perimeter of reference image with a width of 2 pixels and a connective parameter of 8