Identifying Wildfire Burn Severity Risks in the WUI Landscape Using High Resolution Remote Sensing Images and GIS Data

Minho Kim[1], Keats Hua[2], Marta Gonzalez[2]

[1]Dept. of Landscape Architecture & Environmental Planning, University of California, Berkeley
[2]Dept. of Civil & Environmental Engineering, University of California, Berkeley

1. Introduction

Wildfires have devastated California more frequently over the past years due to climate change [1], as the risk of wildfires is only likely to grow considering a 30 to 50% predicted increase in wildfires by 2080 [2]. With the ongoing exacerbation of climate change effects, our communities, particularly in the Wildland Urban Interface (WUI), are at risk of being devastated by higher severity and more frequent wildfires over time. The WUI is commonly defined as the line, area, or zone where structures and other human development meet or intermingle with undeveloped wildland or vegetative fuels [3]. The importance of wildfire mitigation in these areas is first to ensure the safety of the community; however, this goal is hampered by the complexity and heterogeneity of WUI landscapes as well as the potential for the cause of catastrophic wildfire damage [4]. Uncontrolled and highly severe wildfires can result in significant and long-lasting impacts on human society and natural ecosystems. Further, wildfires also disrupt and weaken infrastructures, invoking damages ranging from short-term impediments to first responders to long-term impacts such repairment or replacement of damaged physical components [5].

To combat fires and protect our communities, there is an urgent need to understand the wildfire-causing factors towards the development of effective and sustainable wildfire mitigation. The emergence of remote sensing and GIS data has provided wildfire mitigation using data-driven approaches; however, there is a disconnect between our understanding of landscape patterns and wildfire burns, especially at a high resolution feasible of providing community-level information for the inhabitants, firefighters, and higher-level decision-makers that can ultimately save our properties and lives [6]. Stemming from these limitations, we aim to develop a data-driven, machine learning approach to assess wildfire burn severity with high resolution remote sensing data and generate an understanding of burn severity with different landscape-related variables. We frame this machine learning approach as a multi-label classification problem (given the multi-class burn severity data) and we select input data based on the major proponents of wildfires. Local community members, firefighting groups, and policy decision makers can use the results from our study to help understand potential wildfire hazards in the WUI and formulate effective and sustainable mitigation plans.

2. Related Works

The majority of past studies on burn severity prediction only focused on one or a small number of wildfires [7-9, 10, 11]. While these studies looked at the relationship of factors related to vegetation, terrain, and weather with burn severity, the results were local-scale and not generalizable. Hence, there is a pressing need to be able to characterize landscape features by considering a larger, more

comprehensive set of fires. Understanding common trends and patterns in the landscape features with respect to burn severity will help to link wildfire hazards and risks present in our landscapes.

With the growth of larger data volumes, machine learning models and data-driven approaches are needed. Previous studies have investigated burn severity using machine learning models such as decision trees [7], support vector machine [8], and random forest (RF) [9]. Many studies indicated that RF achieved superior prediction accuracy of burn severity and demonstrated lower sensitivity to different combinations of input variables [8-11]. However, these studies used coarse resolution satellite images such as Landsat (30 m) or MODIS (500 m) [7-9]. At coarser resolutions, there is a higher level of commission and omission errors (mostly above 10% for both error types) that could lead to inaccurate and unreliable detection of landscape features on the ground. Instead, higher spatial resolution (<10 m) is required to view the landscape features in sufficient detail to interpret the heterogeneity of the vegetation fuels and topographic patterns [13]. Interpreting this heterogeneity is crucial since the spatial distribution of the fuel can help determine the distribution of combustible features on the landscape and local wind flows, thereby informing driving factors of fire behavior. [13]. Higher spatial resolution of features, especially vegetation fuels, has been shown to be significant for the performance of predictive wildfire models [14], since fire breaks, vegetation structure, spatial distribution of fuels, and the heterogeneity of the features can be better examined [15].

Recent works on data-driven, machine learning-based approaches of wildfire burn severity prediction continue to rely on RF [9, 11, 16], deep neural networks with five hidden layers [17], and U-Net [18-20]. However, all the reviewed studies had one or more of the aforementioned limitations, such as focusing on one or a small number of wildfires or used coarse spatial resolution images (> 30 m) [9, 11, 16, 17]. Some studies used deep learning (ie. U-Net model) to integrate data from 24 wildfires, but only used satellite images as input to train the model, rather than considering a variety of features related to vegetation, terrain, and weather.

In light of the limitations from previous literature, we present a data-driven framework using RF to classify soil burn severity from 37 different wildfires in California over a 4-year span. We use high resolution, 4-band Sentinel-2 imagery at 10 m resolution, 1/3 arc-second digital elevation models (DEM), and the newly presented land cover map by the European Space Agency (ESA). Our main contribution in this study is the assessment of feature importance with respect to soil burn severity classification using a comprehensive wildfire burn severity dataset with high resolution, open-source data acquired via Google Earth Engine. We use soil burn severity data as our ground truth labels for wildfire burn severity. Further, we extend our feature importance analysis and add averaged features from image patches into our training dataset, as a possible way to improve classification performance. We also present k-means as a possible approach to characterize data and to inform potential training features.

The study is organized in the following order. Section 3 will introduce the study area in California's WUI and the soil burn severity data from the selected 37 wildfires. Then, the input datasets and extracted features used in our training dataset will be presented. Section 4 outlines preprocessing steps and random sampling methodologies proposed to overcome the large data volumes in our data-driven pipeline. Section 5 displays our two-step data processing approach based on K-means and machine learning. Section 6 provides the results and discussion from our

data analysis. Here, we discuss data characterization and feature importance with respect to soil burn severity. Lastly, section 7 provides the conclusion and summative remarks on the study.

3. Study Area and Data

3.1 Study Area

For our analysis, we focused on areas nearby the WUI in California that experienced wildfires from 2018 to 2021. Soil burn severity datasets were obtained from post-response field surveys conducted by the Burned Area Emergency Response (BAER) team, affiliated with the USDA. Soil burn severity is influenced by the fire on ground surface characteristics, including char depth, organic matter loss, altered color and structure, and reduced infiltration [12, 13]. During post-fire assessments, there has been an intentional effort to use the term "soil burn severity" to differentiate post-fire soil properties from fire effects on vegetation (such as tree mortality) and/or general fire effects on long-term ecosystem health [14]. Soil burn severity is categorized into four different classes: low, moderately-low, moderately-high, and high. For this study, we acquired soil burn severity scars from 37 different previous wildfires, and their area of interests (AOI) are displayed in Figure 1. The soil burn severity datasets are provided at 20 m and we resampled them to 10 m to use as ground truth labels in our study.

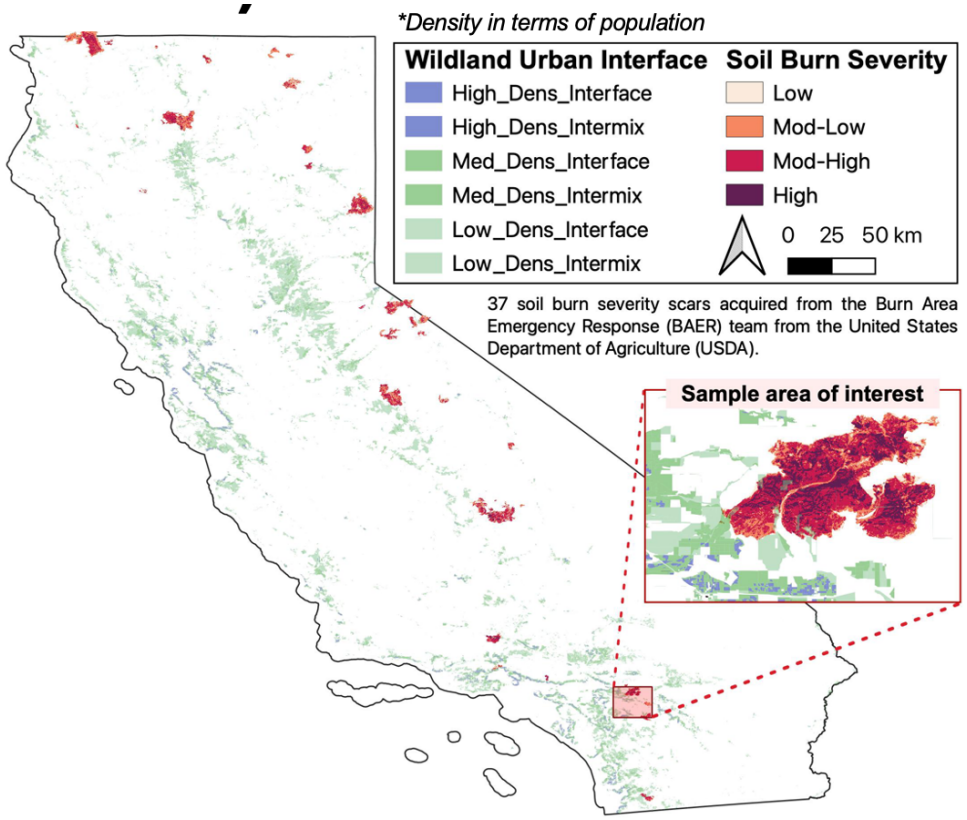

Figure 1. Study areas showing the 37 soil burn severity AOIs used for this study.

The 37 AOIs were deemed suitable as reliable ground truth data because BAER uses satellite image data and performs field surveys as part of their after-response to wildfire occurrences. The intention of limiting the datasets to these AOIs was to focus on the areas that are directly linked with wildfire burns and connect landscape patterns with the burn severity. The proportion of pixels in each soil burn severity class is shown by yearly interval in Figure 2.

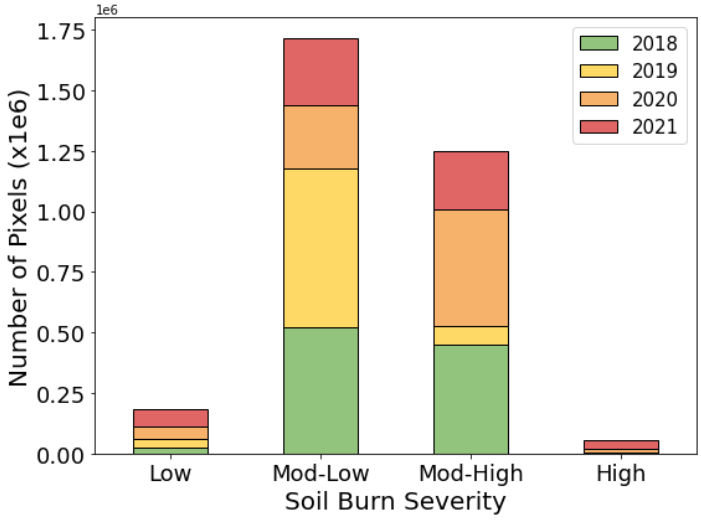

Figure 2. Cumulative graph of soil burn severity class by pixel count for each year.

3.2 Input Data

When acquiring our input datasets, one important constraint was to address the resolution trade-off with the remote sensing imagery. Planetscope offers images at 3 m resolution, but is only available from 2018, which hampers our ability to acquire time-sensitive acquisitions with respect to the AOIs. NAIP imagery offers 1 m resolution but is limited in geographic coverage and is only provided yearly. As a result, we aim to utilize Sentinel-2 images at 10 m resolution (5-day intervals) for our analysis. Another important constraint for data collection was data parsing to match the input features' time stamps with respect to the AOIs. Since we are interested in linking landscapes to burn severity, we filtered for datasets up to one month prior to the starting date of each AOI's wildfire. One month was considered to be appropriate to evaluate changes in vegetation and climatic dynamics with respective potential wildfire behavior. We also filtered for the exact geographic location, ensuring to maintain the same projection and spatial extents for all input features.

We collected 10 different data features to create our input dataset. First, we collected 4-band Sentinel-2 imagery (red, green, blue, NIR bands). Using Sentinel-2 imagery, we computed the normalized difference vegetation index (NDVI) and soil adjusted vegetation index (SAVI) to characterize vegetation cover and bare soils, respectively, and their formulas are shown below:

$$NDVI = \frac{NIR-RED}{NIR+RED} \tag{1}$$

$$SAVI = \frac{NIR - RED}{NIR + RED + L}(1 + L) \tag{2}$$

The NIR and red bands are used to compute both vegetation indices, exploiting the stark difference in reflectance spectra at these wavelengths for vegetation. The "L" factor in SAVI is used to adjust the index for bare soil content in the image scene and is calculated by the amount of vegetation cover in the scene. In general, an "L" factor value of 0.5 is used for moderate green cover and was assumed for all SAVI calculations in this study. In addition, DEMs were acquired from the 3DEP dataset as 1/3 arcsecond resolution. The DEM was used for elevation measurements and was used to calculate slope and aspect. The slope and aspect are essential factors in wildfire events since they characterize the different patterns in the terrain, thus influencing wind effects and rate of spread. The land use land cover map was used as a proxy to the WUI and to interpret the highly heterogeneous landscape. Detailed specifications of all input features in the training dataset are shown in Table 1 and examples are shown in Figure 3. The ground truth data (soil burn severity from BAER) is highlighted with a yellow background.

Table 1. Acquired datasets, features, and detailed specifications.

Training Datasets	Data Source	Analyzed Feature	Time Frame	Spatial Resolution
DEM	USGS	Elevation, Slope, Aspect	2013	1/3 arcsecond (~10 m)
Satellite Imagery	Sentinel-2	R, G, B, NIR, NDVI, SAVI	2018-2021	10 m
Land Use Land Cover Map	ESA	Land cover classes	2018-2021	10 m
Soil Burn Severity*	BAER	Burn Severity	2018-2021	10 m

**Ground truth labels*

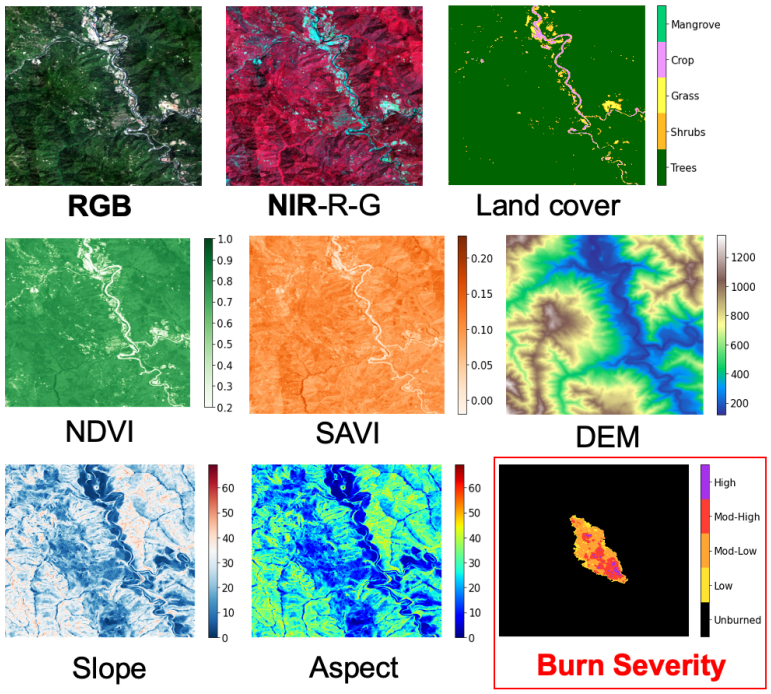

Figure 3. Images of a sample AOI shown for the 10 extracted features used in the input dataset.

4. Data Preprocessing

(1) Input Data Preprocessing

Preprocessing is imperative due to the usage of multi-scaled remote sensing data. For the Sentinel-2 images, all images were scaled and set to surface reflectance products with a range of 0 to 1. The NDVI product was thresholded to values greater than 0.2 to filter noise and to include significant amounts of vegetation. The LULC map was filtered for only vegetation and terrain-related classes. All images were resampled using nearest neighbor interpolation to 10 m resolution. Then, all ten of the input features are stacked together and normalized by their minimum and maximum values. In this step, minmax normalization is crucial to match the different scales of the input data and to assist with efficient model training.

(2) Random Sampling Methods

Random sampling was imperative due to the large volume of pixels in the input dataset and because of the uneven distribution of pixels across different soil burn severity classes and the various wildfire AOIs, as shown in Figure 4.

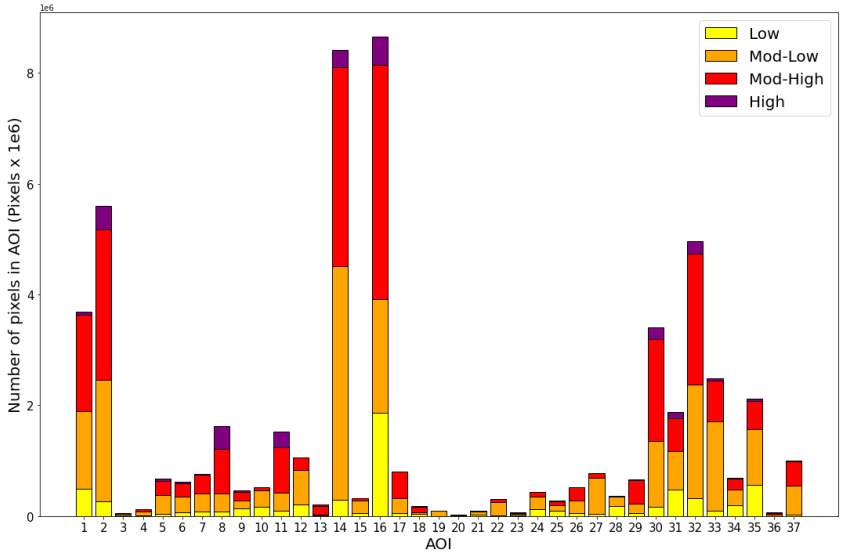

Figure 4. Cumulative graph of soil burn severity class pixels for each AOI.

In this study, we utilized two random sampling methods to extract point samples from the data stacks of the input features for each AOI and for each burn severity class. First, we used a random sampling method based on the *wildfire size*, where the pixel count for sampling ($p_{AOI,class}$) was determined as the product of the total number of sampled points and the relative pixel count ($n_{AOI,class}$) of an AOI for the corresponding burn severity class, as expressed below:

$$p_{AOI,\,class} = \frac{n_{AOI,\,class}}{\sum_{i=0}^{37} n_{AOI,class}}$$
$$\text{where } n_{AOI,\,class} = p_{AOI,\,class} \times n_{sample} \tag{3}$$

In total, 10,000 points were sampled from each burn severity class, and we used the following number of samples: Low: 9979, Low-Med: 9979, Med-High: 9984, and High: 9985. The training and test datasets were created with an 80% (31941 samples) and 20% (7987 samples) split from the input features and ground truth labels. Second, we randomly sampled using a *fixed* number of point samples. In other words, we extracted 300 sample points from each AOI for each burn severity class and created a training set of 29760 points and a test set of 7440 points. A summary of the two random sampling methods is given in Table 2.

Table 2. Random sampling methods and dataset splits.

Name	Sampling Method	Dataset (# Samples)
Random	**Wildfire Size** Pixel count per class by relative size	Train (31948) = 80% Test (7987) = 20%
Fixed	**Fixed sample count** 300 pixels per class	Train (29760) = 80% Test (7440) = 20%

5. Methodology

The aim of this study was to find correlations between the landscape features and the soil burn severity of previous wildfires by using different remote sensing and GIS data. For this purpose, we used K-means to cluster the data distributions of the input features and applied machine learning model to classify soil burn severity classes and generate feature importance.

(1) Data Processing Part 1: Data Clustering Using K-means

To determine the true number of clusters in a data set, the elbow method is the most commonly used visualization tool in conjunction to k-means clustering. By varying the number of clusters(K), for each value of K, the value of WSS (Within-Cluster Sum of Square) which is the sum of squared distance between each point and the centroid in a cluster could be calculated. After plotting the WSS value and K value, a line plot with an elbow shape is often shown with an inflection point somewhere in the middle.

(2) Data Processing Part 2: Burn Severity Classification and Feature Importance

We used machine learning models to train and classify our input features to the four burn severity classes. The input training datasets were created using the two random sampling methods mentioned in Section 4 (2). For this study, we chose five machine learning algorithms, as shown in Table 3. The significance of this multi-label classification was in using feature importance metrics to gauge the influence of each factor for burn severity classification. As a result, we could rank the influence of the different features on the classification of each soil burn severity class as well as their permutation importance. An overview of the entire framework is provided in Figure 5 for the reader's convenience.

Table 3. Machine learning classifiers used for burn severity classification.

Classifier	Parameters
Multi-layer Perceptron (MLP)	2 hidden layers, 100 neurons
Naïve Bayes	-
AdaBoost Classifier	50 estimators
Gradient Boosted Classifier	50 estimators
Random Forest (RF)	100 estimators

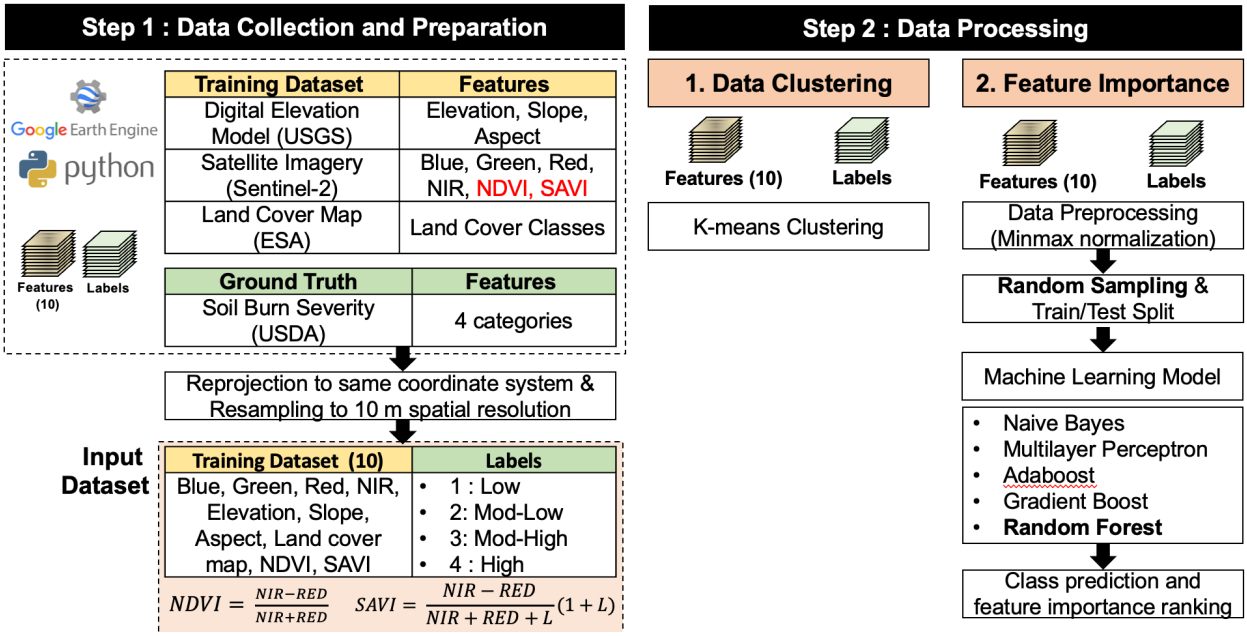

Figure 5. Overview of the data collection and processing framework for our data-driven wildfire burn severity characterization and soil burn severity classification machine learning model.

6. Results and Discussion

(1) Method Part 1: Data Clustering Using K-means

The WSS value is the largest when K=1 and starts to decrease as K value increases, after the inflection point the graph reaches a plateau which is almost parallel to the x-axis. The K value corresponding to the inflection point is the optimal K value or the optimal number of clusters. From the following elbow method graph generated from our sample dataset with size of 39927 data points and 10 input features, the inflection point with K value of 2 shows the most dramatic change and thus we choose K=2 as our optimal number of clusters. There are also drawbacks to using the elbow method in terms of determining the "elbow". It cannot always be unambiguously identified, and sometimes there could be no elbow or several elbows as we try different sizes of sample datasets.

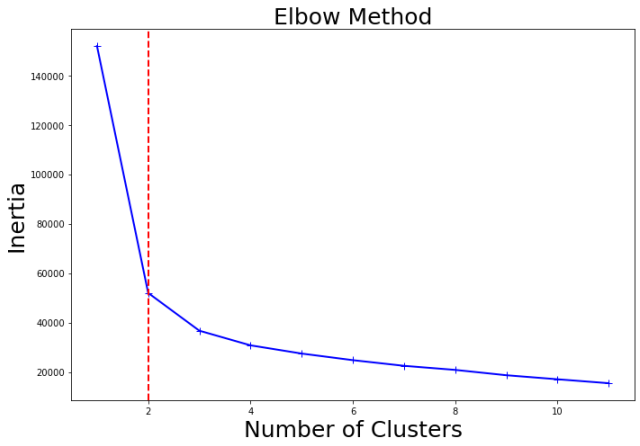

Figure 6. Elbow method plot.

In our analysis, the silhouette score is used to evaluate the quality of clusters created using clustering algorithms of K-Means, and to generate results of showing how well samples are clustered with other samples that are similar to each other. The Silhouette score is calculated for each sample of different clusters based on two parameters, the mean intra-cluster distance a(i) and mean nearest-cluster distance b(i). Using the following formula, the value of the Silhouette score could be determined varying from -1 to 1.

$$s(i) = \frac{b(i) - a(i)}{Larger\ of\ b(i)\ and\ a(i)} \tag{4}$$

The closer the score draws near to 1, the cluster is more dense and well-separated than other clusters. If the score is near 0, it represents overlapping clusters with samples very close to the decision boundary of the neighboring clusters. A negative score from range [-1, 0] indicates that the samples might have been assigned to the wrong clusters. With the meaning of scores being clarified, the following result table shows the silhouette scores for each case of a different number of clusters. When K=2 (2 clusters), the value of silhouette scores is approximately 0.53 which is the highest among other scenarios. Starting after K=7, the silhouette score is normalized around the value of 0.3 which indicates a differentiated relationship between all the clusters.

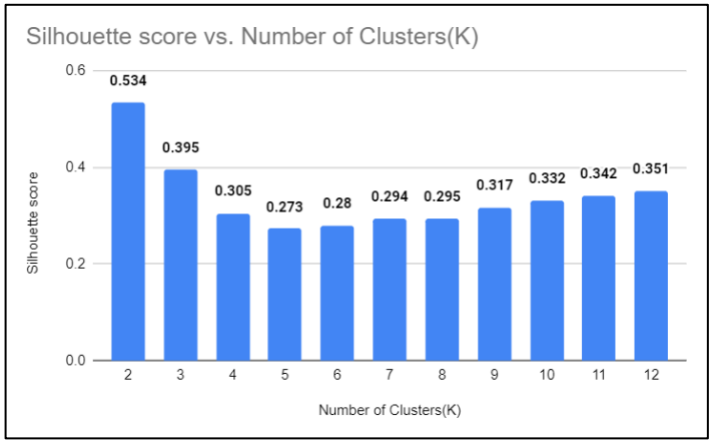

Figure 7. Comparison of silhouette scores for each cluster categorization from K-means.

The following silhouette plots show the result of silhouette analysis which is used to select the most optimal value of the K (number of clusters) in K-means clustering. The selection of optimal value is based on three criteria, the presence of clusters with below-average silhouette scores, the level of fluctuations in the size of silhouette plots, and the uniformity of thickness of the silhouette plots. According to the criteria, a detailed analysis could be performed for each silhouette plot. For K=2, the plot is the most optimal because the two clusters have no below-average silhouette scores, nearly no fluctuation in size among clusters, and the thickness of these two silhouette plots are uniform. Plots with K=3, 4, 5, 6, 7 are suboptimal comparing to K= 2 because the fluctuations in the size of silhouette plots and the thickness of the plots varies significantly. Therefore, we can select the most optimal of K, the number of clusters, as 2.

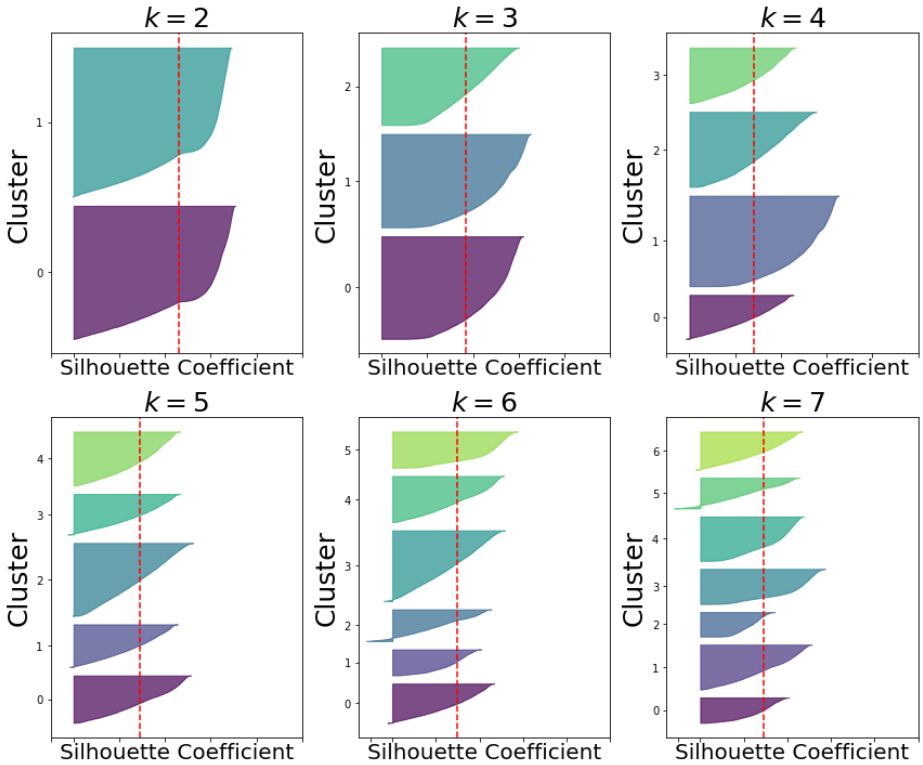

Figure 8. Silhouette analysis for a various number of clusters.

As we have obtained the cluster labels and categorized the sample data points into two clusters, the following plot displays the spatial distribution of the clusters from all the sample data points on a choropleth map in scale of California. The cluster distribution could help us to visualize the locations of each cluster and quickly identify any spatial patterns that the clusters might have. Since each cluster represents areas that have similar characteristics or features, mapping them could allow us to determine what area with similar features will tend to have similar locations. The sample cluster points distribution area on the map shows two sets of tracts with the same color differentiate clearly from left to right. While the map could only give us very limited information on the types of area each cluster represents, to complement the geo-visualization of the clusters, we can explore the statistical properties of the cluster map.

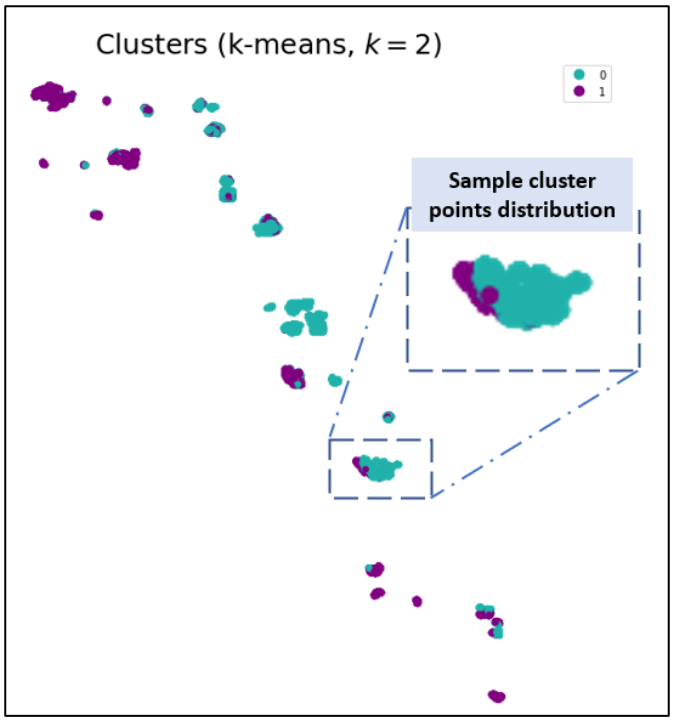

Figure 9. Sample data points distribution for Cluster 0 and Cluster 1

The further interpretate the cluster characteristics, we generated the following two plots to visualize and determine what observations are part of each cluster and what their characteristics are. The radar plots displays the mean value of the 10 features for each cluster and the sample data, and it shows that while some attributes such as the aspect, red, blue, and green display largely the same value for each cluster, others paint a much more divided picture especially in Elevation (DEM) and Slope. The distribution plot displays the value distribution of 10 features for each cluster, and it shows similar results as the radar plots. We could observe 2 clear clusters on the plots for feature NDVI, SAVI, Elevation (DEM), and Slope, while the cluster value distribution for other features are superposed to a great extent. From all the statistical observations from plots, we could describe each cluster by its relative feature distribution. Cluster 0 depicts the area where it has relatively high elevation, low volume in vegetation, and flatter surfaces, and cluster 1 depicts areas that are in the opposite where it has low elevation, high volume in vegetation, and more steep surfaces. For features of red, blue, green, Near-infrared (NIR) bands, the value for Cluster 0 is slightly less than the value of Cluster 1. While two clusters have similar pattern in aspects, Cluster 1 has greater value in LULC which depicts areas in cluster 1 has more existence of human-related residence.

Taken altogether, these plots allow us to start delving into the multidimensional complexity of each cluster and the types of areas behind them, which helps us to better corelate the finding of feature importance from machine learning. In conclusion, the clustering method reduces the complexity of categorizing datasets and allow us to describe complex and multifaceted data by creating groups of observations with distinct and internally consistent distributional/descriptive characteristics.

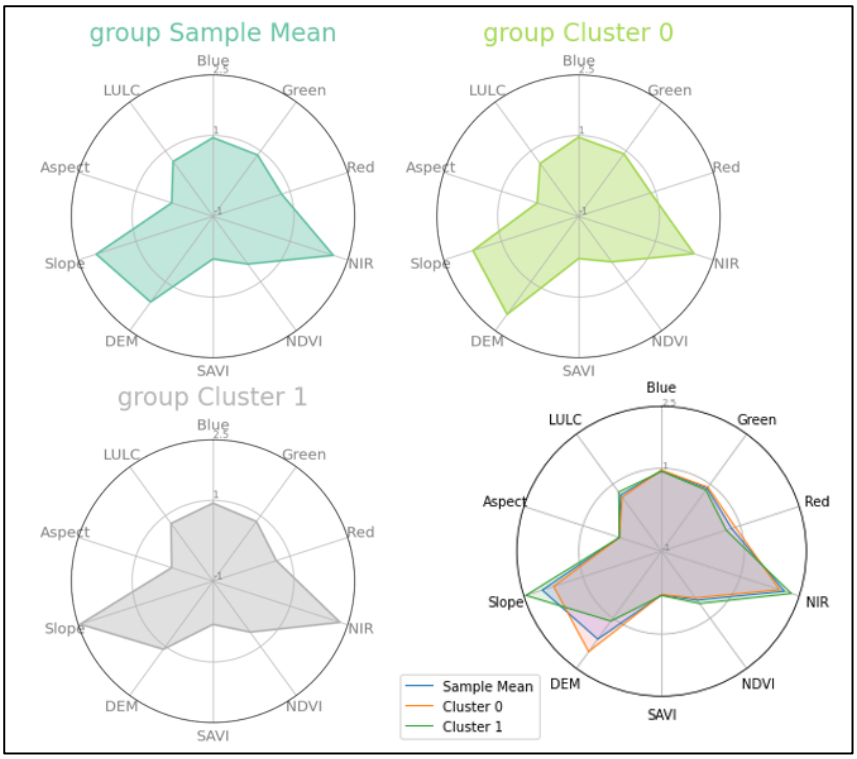

Figure 10. Mean values of each cluster and all sample data for 10 features

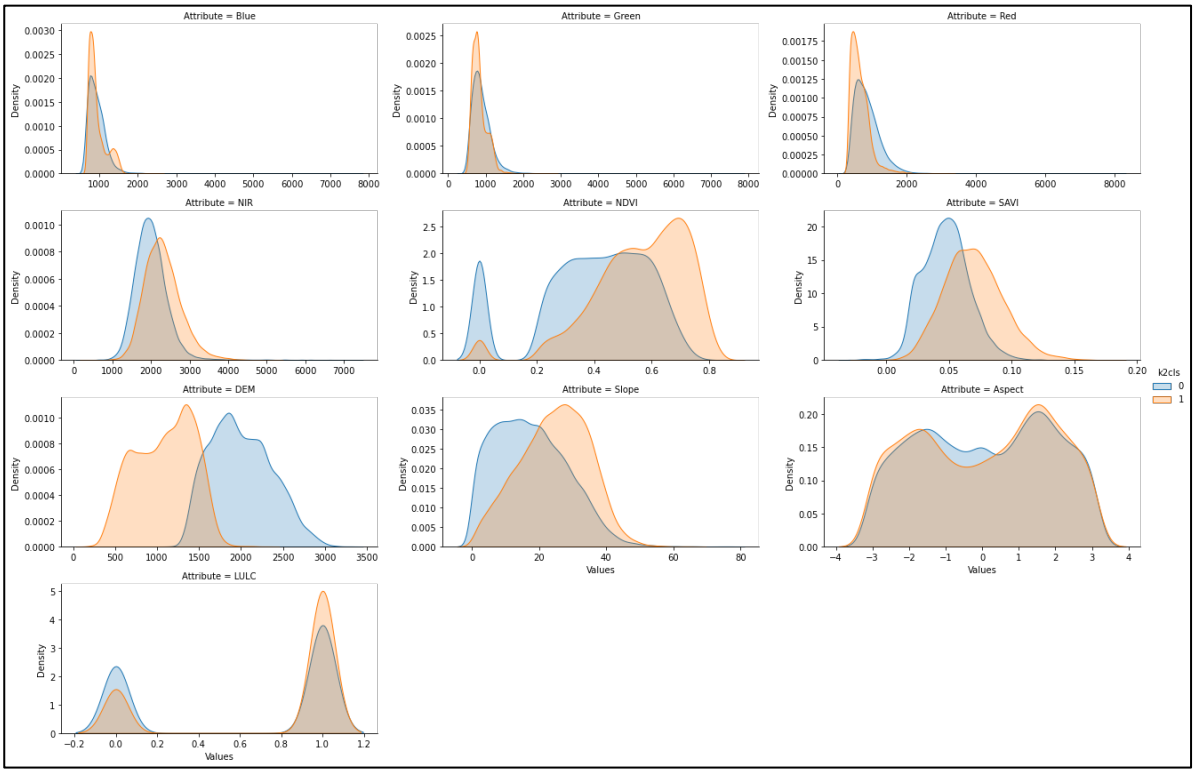

Figure 11. Distribution of each cluster's values for 10 features

(2) Data Processing Part 2: Burn Severity Classification and Feature Importance

The soil burn severity classification and accuracy results for each machine learning model is presented in Table 4. The RF model returned the best classification performance, reaching up to 59.13% in overall accuracy using the random sampling method. Increasing the number of estimators in RF helped increase the accuracy to an extent but tended to plateau after 100 estimators. Furthermore, increasing the number of samples past 10,000 sample points per burn severity class showed negligible changes in the accuracy.

Table 4. Accuracy results from soil burn severity classification of the five models.

Models	Random	Fixed
MLP	56.00%	49.23%
Naïve Bayes	40.20%	35.71%
Adaboost	47.49%	42.58%
Gradient Boost	53.59%	49.01%
Random Forest	**59.13%**	**58.00%**

A detailed inspection at the F1-scores for RF in Table 5 shows good classification performance for each burn severity class. The low severity class and high severity class recorded the highest accuracy values (approximately 0.75 F1-score), whereas the "moderate" severity classes were relatively low. The difficulty of classifying the moderately burned regions has always been an issue, as previous studies have also produced ambiguous results. The results therefore indicate that more improvements are necessary in the mixed, moderate-level classes in order to further improve the classification accuracy.

Table 5. Detailed classification accuracy results for RF per burn severity class.

Severity	F1-Score	Number of Pixels
Low	0.64	2039
Mod-Low	0.49	1906
Mod-High	0.47	1973
High	0.75	2069

We also conducted feature importance using mean decrease in impurity (MDI) or Gini impurity. The change in impurity of the split between the decision tree estimators (at each branch and leaf within each tree estimator). Here, the impurity value is measured as how often a randomly chosen element from the set would be incorrectly labeled if labels from the subset were used at random. This impurity value ranges from 0 to 1 and optimal splits are achieved at higher impurity values.

$$G = \sum_{i=1}^{c} p(i) * (1 - p(i)) \tag{5}$$

In more detail, the impurity value, "*G*", can be expressed by the following formula, where *C* denotes the burn severity classes and "*p(i)*" is the probability of selecting a point with class "*i*". The feature importance values in Figure 12 are the average of the impurity in the entire algorithm.

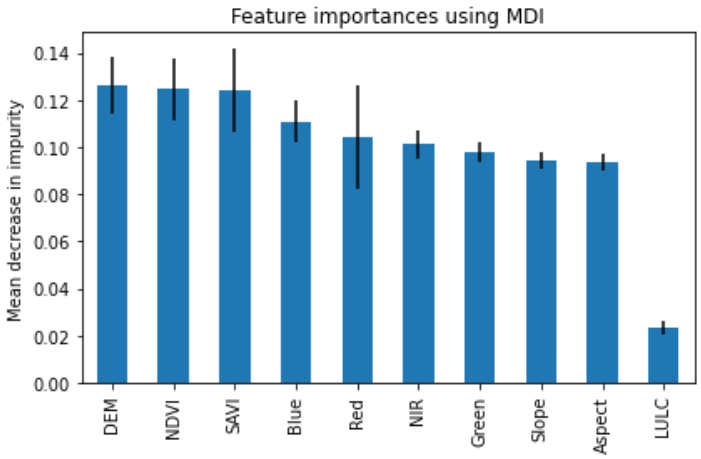

Figure 12. Feature importance by mean decrease in impurity (MDI).

DEM, NDVI, and SAVI were found to be the most influential based on MDI. The higher influence of the vegetation indices (NDVI and SAVI) was expected, given the close relationship between vegetation fuels and wildfires. Interestingly, the DEM was found to have the greatest influence, underscoring the fact that the terrain is a potentially significant factor. However, DEM by-products, namely slope and aspect, recorded low influence levels. One explanation is possibly due to the use of point samples, since there may not have been sufficient contextual data for the model to learn the patterns in the topography. The inclusion of neighboring pixels add context which may be crucial for features like aspect, and slope. Further, although the LULC map was ranked the lowest in terms of feature importance, but this may have been due to the categorical nature of the data.

Determining feature importance solely by using MDI can be biased, especially for features with high cardinality (randomness) like LULC. In particular, purely random factors or non-predictive features may be found to be highly important due to model overfitting. To remove this bias, we conducted permutation importance on the features as presented in Figure 13. This analysis overcomes the limitations from MDI-based bias. Input features are shuffled randomly and refitted to the model to re-estimate the feature's importance. Higher changes in mean accuracy values indicate that the model is more dependent on the features. In our analysis, the vegetation indices were the most influential, followed by the DEM. Again, the slope and aspect were found to have the lowest delta value in permutation importance. While these two by-products of the DEM scored considerably low in ranking, the fact that the DEM shows high delta values still poses an encouraging sign that the topography needs more contextual information to accurately represent the complexity. In contrast to the MDI result above, LULC demonstrated higher permutation importance in between the red and NIR bands. The similar level of influence may be attributed to the vegetation-related classes thresholded in the LULC.

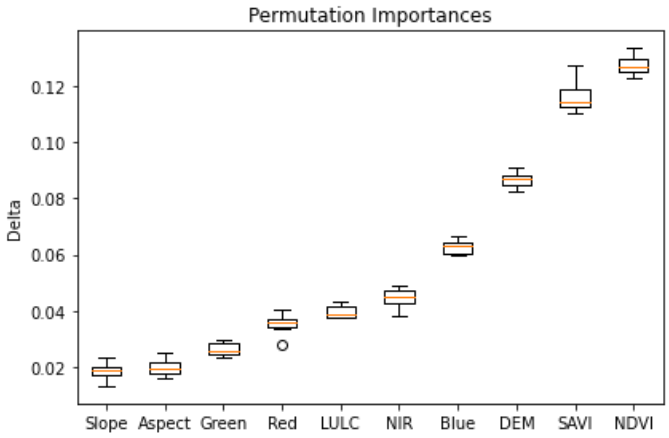

Figure 13. Permutation importation plot for each feature.

To expand the number of features in our training dataset, we computed the average values of the original input features from neighboring pixels in a patch. Since high resolution satellite images may contain salt-and-pepper noise and the fine-grained features may not be fully characterized by single pixels, patches provide additional information that could be more representative of the landscape feature's reflectance patterns. An example is shown in Figure 14 for a 3x3 pixel patch. The average values were calculated for all features except LULC and aspect because LULC contains categorical data and aspect scales in terms of cycles (radians). Hence, a total of 18 features were used for the new training dataset. The same five models were trained using datasets with averages computed for 10x10, 20x20, 30x30, 40x40, and 50x50 pixel patches and the classification results are presented in Table 6.

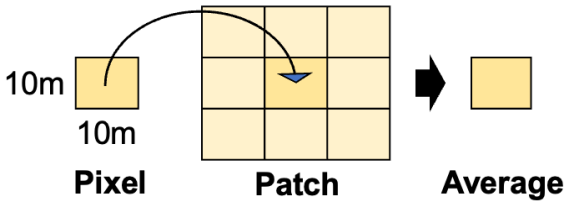

Figure 14. Feature importance by mean decrease in impurity (MDI).

Table 6. Detailed classification accuracy results for RF per burn severity class.

Models	Pixel	p=10	p=20	p=30	p=40	p=50
MLP	56.00%	57.48%	**58.08%**	55.75%	57.60%	57.48%
Naïve Bayes	40.20%	**41.15%**	40.68%	40.11%	40.68%	40.64%
Adaboost	47.49%	**48.82%**	48.13%	47.88%	47.82%	48.01%
Gradient Boost	53.59%	**55.13%**	54.44%	54.03%	54.33%	54.38%
Random Forest	59.13%	60.42%	61.32%	61.83%	62.19%	**62.90%**

All models experienced an increase in accuracy with the inclusion of the patch-based averages. This increase could be the influence of additional features; nevertheless, the addition of contextual data does help to improve the soil burn severity classification. The RF model benefits the most from the additional features, continuing to increase in accuracy for larger patches. However, this result may have been caused by overfitting and future works will require a cross-validation or

holdout set sampling to assess the robustness of the models. The other models (ensemble based and Naïve Bayes) plateau after a 10x10 pixel patch, suggesting that average values computed from exceedingly large patches may not provide indicative values of the original feature. In this light, the addition of the averaged values was not very helpful when comparing feature importance. Most of the original features and the averaged values presented a similar level of influence, as shown in the feature importance by MDI in Figure 15. Interestingly, LULC yielded a high permutation importance value, suggesting that the land cover classes could be useful when training a large number of features. Given the high influence of DEM, to uncover the influence of different topographic patterns, two approaches can be taken for future research. First, additional features measuring different topographic patterns or spatial metrics can be added to the input dataset. Different metrics such as the topographic position index, roughness index, landform classes, etc. Second, full patch samples can be inputted into the model (instead of computing a single pixel value), which may require a deeper model (ie. Convolutional neural network) to be able to learn the higher level of complexity and heterogeneity in the larger-sized input data.

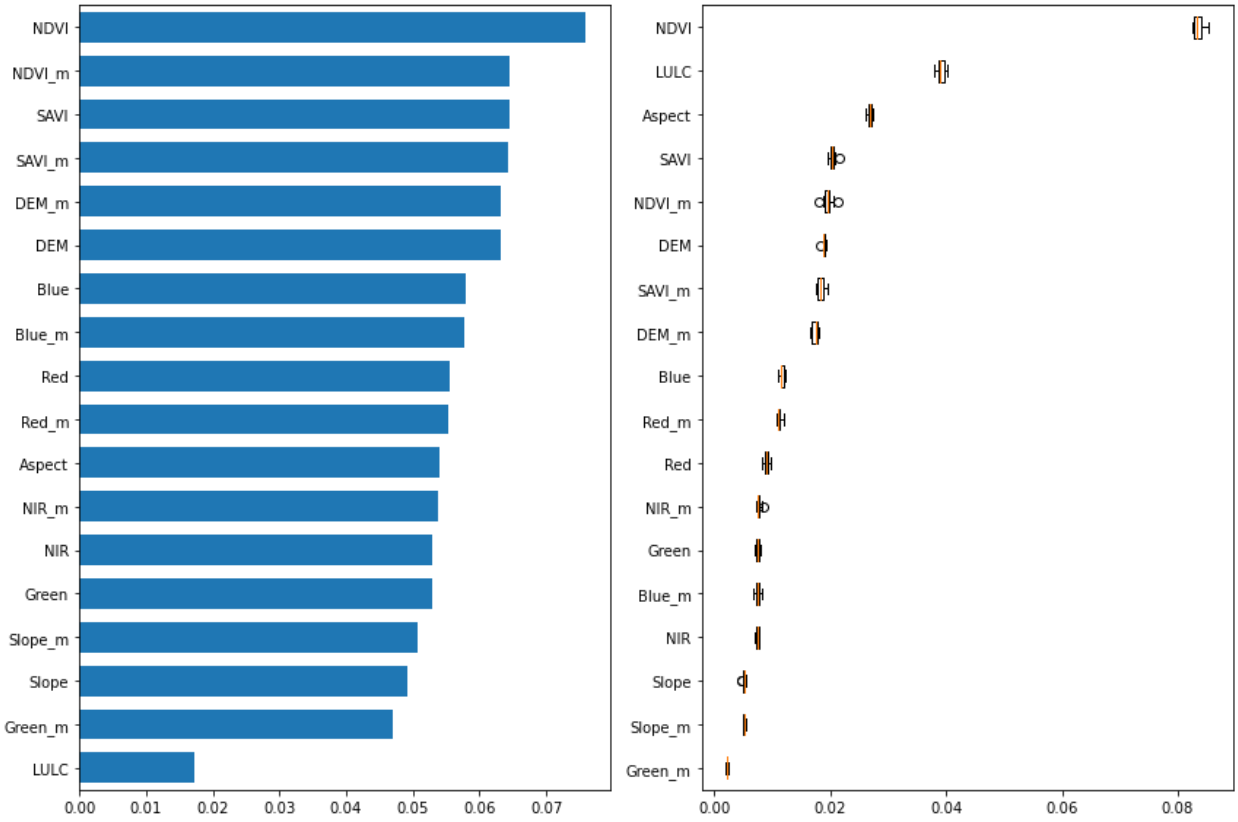

Figure 15. Feature importance by MDI and permutation importance.

7. Conclusion and Future Work

The surge in wildfire occurrence and severity in recent years underscores the importance of understanding our current landscapes to prepare against future wildfire risks and hazards. As communities traverse deeper into the WUI, understanding the complex environment and the relationship with potential wildfire risks is essential for our survival, our properties, and for safeguarding our environment against catastrophic damage. Only by interpreting the landscape

heterogeneity at sufficiently high resolution will we be able to impose impactful wildfire mitigation and decision-making for sustainable and safe communities.

In this study, we compared high resolution remote sensing data as independent variables to compare with the burn severity data using methods. In more detail, we characterized the topography (eg. basic statistics of elevation profiles, slope variation) and vegetation (eg. NDVI distribution, fractional vegetation cover, canopy cover) for the AOIs to investigate potential patterns in the landscape with previous wildfire burn severity. Our RF model demonstrated strong classification accuracy for low and high burn severity classes, but still struggled with moderate-level classes. We conducted feature importance analyses and determined that vegetation indices, namely NDVI and SAVI, were most influential in predicting burn severity classes. We also noted the potential of using DEMs and topographic information to interpret the burn severity classes. Lastly, we conducted K-means clustering to identify key groups and patterns in the data distributions of the input features.

Ultimately, this preliminary analysis is an attempt to interpret the landscape heterogeneity by looking at patterns and contextual information. For our future works, we plan on including more input features such as downscaled weather and climate data, texture data from satellite imagery, effective vegetation height, and canopy data. One notable data source is evapotranspiration which was discovered to have a significant influence in burn severity estimation [21]. We also look to conduct additional sampling strategies that are more representative of the complexity of the wildfire burn severity. For instance, we plan on using patch-based samples to include neighboring contextual information instead of averaged pixel values. In this case, we aim to use the full image patches as training data and adopt convolutional neural networks, similar to previous deep learning studies [18-20]. Given our results with Sentinel-2 (10 m resolution), we also plan to utilize high resolution Planetscope imagery (3 m resolution) for the same burn severity dataset to investigate the influence of higher spatial resolution on our prediction. Wildfire burn severity prediction has not yet been conducted with Planetscope images nor for a high resolution image dataset (<10 m) with an input dataset as diverse and burn severity dataset as comprehensive as the one introduced in the current study. Lastly, we aim to use our findings and determine optimal features and key patterns in the landscape that can inform fire spread simulations and modeling techniques.

Authorship Contribution Statement

- **Minho Kim**: Conceptualization, literature review, original manuscript, methodology (machine learning), coding (Google Earth Engine data pipeline, k-means, machine learning, data visualization, mapping), editing and review
- **Keats Hua**: Literature review, BAER burn severity data, methodology (k-means), coding (k-means result visualization), editing and review
- **Marta Gonzalez**: Supervision, editing and review

References

1. NERC. (n.d.). Retrieved October 6, 2021, from https://nerc.ukri.org/research/partnerships/ride/lwec/report-cards/.
2. Cheng, L., K. M. McDonald, R. P. Angle, and H. S. Sandhu (1998), Forest fire enhanced photochemical air pollution: A case study, Atmos. Environ., 32(4), 673–681.
3. Martinuzzi, S., Stewart, S. I., Helmers, D. P., Mockrin, M. H., Hammer, R. B., & Radeloff, V. C. (1970, January 1). The 2010 wildland-urban interface of the conterminous United States. Retrieved October 6, 2021, from https://www.nrs.fs.fed.us/pubs/48642.
4. Scasta, J. D., Weir, J. R., & Stambaugh, M. C. (2016). Droughts and wildfires in western U.S. rangelands. Rangelands, 38(4), 197–203. https://doi.org/10.1016/j.rala.2016.06.003
5. What is the wui? U.S. Fire Administration. (2021, July 9). Retrieved October 6, 2021, from https://www.usfa.fema.gov/wui/what-is-the-wui.html.
6. Northern Research Station News releases. Most California Fires Occur in Area of Wildland-urban Interface with Less Fuel and More People. (n.d.). Retrieved October 6, 2021, from https://www.nrs.fs.fed.us/news/release/wui-interface-intermix.
7. Sá, A. C. L., Pereira, J. M. C., Vasconcelos, M. J. P., Silva, J. M. N., Ribeiro, N., & Awasse, A. (2003). Assessing the feasibility of sub-pixel burned area mapping in miombo woodlands of northern Mozambique using MODIS imagery. International Journal of Remote Sensing, 24(8), 1783-1796.
8. Hultquist, C., Chen, G., & Zhao, K. (2014). A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests. *Remote sensing letters*, *5*(8), 723-732.
9. Holden, Z. A., Morgan, P., & Evans, J. S. (2009). A predictive model of burn severity based on 20-year satellite-inferred burn severity data in a large southwestern US wilderness area. Forest Ecology and Management, 258(11), 2399-2406.
10. Lydersen, J. M., Collins, B. M., Brooks, M. L., Matchett, J. R., Shive, K. L., Povak, N. A., ... & Smith, D. F. (2017). Evidence of fuels management and fire weather influencing fire severity in an extreme fire event. *Ecological Applications*, *27*(7), 2013-2030.
11. Kane, V. R., Cansler, C. A., Povak, N. A., Kane, J. T., McGaughey, R. J., Lutz, J. A., ... & North, M. P. (2015). Mixed severity fire effects within the Rim fire: relative importance of local climate, fire weather, topography, and forest structure. Forest Ecology and Management, 358, 62-79.
12. Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. Environmental Reviews, 28(4), 478-505.
13. Ruscalleda-Alvarez, J., Moro, D., & Van Dongen, R. (2021). A multi-scale assessment of fire scar mapping in the Great Victoria Desert of Western Australia. International Journal of Wildland Fire.
14. Taneja, R., Hilton, J., Wallace, L., Reinke, K., & Jones, S. (2021). Effect of fuel spatial resolution on predictive wildfire models. International journal of wildland fire, 30(10), 776-789.

15. Atchley, A. L., Linn, R., Jonko, A., Hoffman, C., Hyman, J. D., Pimont, F., ... & Middleton, R. S. (2021). Effects of fuel spatial distribution on wildland fire behaviour. International Journal of Wildland Fire, 30(3), 179-189.
16. Malik, A.; Rao, M.R.; Puppala, N.; Koouri, P.; Thota, V.A.K.; Liu, Q.; Chiao, S.; Gao, J. Data-Driven Wildfire Risk Prediction in Northern California. Atmosphere 2021, 12, 109. https://doi.org/10.3390/atmos12010109
17. Langford, Z., Kumar, J., & Hoffman, F. (2018, November). Wildfire mapping in Interior Alaska using deep neural networks on imbalanced datasets. In 2018 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 770-778). IEEE.
18. Farasin, A., Colomba, L., & Garza, P. (2020). Double-step u-net: A deep learning-based approach for the estimation of wildfire damage severity through sentinel-2 satellite data. Applied Sciences, 10(12), 4332.
19. Monaco, S., Pasini, A., Apiletti, D., Colomba, L., Garza, P., & Baralis, E. (2020, December). Improving Wildfire Severity Classification of Deep Learning U-Nets from Satellite Images. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 5786-5788). IEEE.
20. Monaco, S., Greco, S., Farasin, A., Colomba, L., Apiletti, D., Garza, P., ... & Baralis, E. (2021). Attention to Fires: Multi-Channel Deep Learning Models for Wildfire Severity Prediction. Applied Sciences, 11(22), 11060.
21. Quintano, C., Fernández-Manso, A., & Roberts, D. A. (2020). Enhanced burn severity estimation using fine resolution ET and MESMA fraction images with machine learning algorithm. *Remote Sensing of Environment*, *244*, 111815.