AN ACCESSIBLE LOW COST TOOL FOR CITIZEN SCIENTISTS: USING REMOTE SENSING TECHNIQUES TO PREDICT FIRE DAMAGE PROPENSITY

Vedant Janapaty[1]
[1]Silver Creek High School, San Jose, CA, United States

Abstract – **Over the last few years, devastating fires in California have destroyed habitats, released toxic fumes, and caused extreme heat. Researchers expect wildfire incidents to increase by over 77% by the end of the decade; the frequency of wildfires with a burned area range greater than 25,000 acres will increase by 50%. The fires of 2020 burned 4.5 million acres, straining fire agencies and local communities, which could be alleviated through preemptive measures. The goal of this project is to (1) establish a correlation between land conditions prior to fire, using satellite indices, and the extent of fire damage and (2) build a tool for preemptive work. Wildfire statistics, such as latitude, longitude, and acres burned were obtained from CalFire's annual red books. In order to analyze land conditions, twelve parameters were identified and obtained from NASA and USGS Earth Data. Python code was written to process CSV files and classify the indices into bands for a machine learning model (ML). ML was first trained on data from 100 California wildfires of varying intensities over the last 19 years. ML was then used to predict fire intensities for 227 locations selected across California, based on satellite indices in August 2020. The ML model has a precision of 95.3% and area under Curve at a statistically significant value of 99.3%. This easily-accessible and high-precision model offers a prediction window to citizen scientists and ecological nonprofits for preventative measures. Some measures include removal of invasive species and the restoration of native plants.**

NOMENCLATURE

NDVI	Normalized Difference Vegetation Index
EVI	Enhanced Vegetation Index
NBR	Normalized Burn Ratio
LAI	Leaf Area Index
FPAR	Fraction of Photosynthetically Active Radiation
DT	Day Temperature
NT	Night Temperature
VO	Vegetation Opacity
ST	Surface Temperature
SM	Soil Moisture
SW	Surface Wetness
RW	Rootzone Wetness

I. INTRODUCTION

Every year, California experiences devastating wildfires. The fires of 2019 burned 253,321 acres, while the fires of 2020 destroyed 4.5 million acres of land, with financial losses in the order of billion of dollars. Wildfires have cause environmental impact, destroy habitats, and release toxic fumes into the atmosphere.

Sensors aboard Earth-orbiting satellites measure electromagnetic radiation reflected from the Earth at different wavelengths, known as bands. The band information can be analyzed using mathematical formulas for predictive analysis done by machine learning algorithms. The vegetation in a region is the primary fuel for spreading a wildfire. Removal of invasive species and reintroducing native species of California's chaparral is a strategy employed by volunteers, conservationists, and citizen scientists prior to a fire[13]. These strategies could be employed in regions identified as high risk by remote sensing technologies, thus potentially reducing or completely avoiding a fire altogether.[4]

The goal of this project is to establish a correlation between land conditions prior to a fire with the potential extent of fire damage, in the event of fire. Based on this information, preemptive steps can be taken by citizen scientists or fire agencies to protect the region. Current mathematical fire models to predict the rate of spread and intensity include the Huygen's principle and the Rothermel's surface fire spread model. [9] Both require specialized equipment available to firefighters, and do not take the vegetation and surface temperature into consideration. [10]

Other big data and statistics-based models such as Phoenix RapidFire and Fire Regime Operations Simulation Tool (FROST) are available solely for fire agencies. Another project, multi-tier GOLIAT, uses drones to provide fire fighters with infrared and visible-spectrum images, for fire dynamics and prediction. An average drone with thermal imaging costs around $40,000 [12].

However, a simple, low cost, fast tool that would enable citizen scientists and volunteers to preemptively change land conditions (in their neighborhoods) or manage resources based on fire propensity is not currently available or accessible.

II. METHODS

The overall goal of this project is to create a predictive measure of the extent of fire damage using satellite data indices. California wildfire statistics, such as latitude, longitude and acres burnt were obtained from CalFire's annual Red books (Table 1). Satellite data indices were obtained from the United States Geological Survey (USGS) and NASA Earth Data. In order to analyze the land conditions, twelve parameters were identified, studied, and obtained from datasets in CSV file format.

12 unique satellite parameters were selected from Landsat, MODIS, and SMAP satellites (Table 2).

TABLE I
SAMPLE DATA FROM CALFIRE'S ANNUAL RED BOOKS

Fire Name	Acres	Start	Latitude	Longitude
Creek 5	4276	12/23/2020	32.7157	-117.1611
Airport	1087	12/23/2020	33.9806	-117.3755
Bond	6686	12/2/2020	36.7783	-119.4179
Laura 2	2800	11/17/2020	40.5394	-120.712
Silverado	12466	10/26/2020	33.7175	-117.8311
Coleman	1500	10/18/2020	36.3136	-121.3542
Snow	6254	9/17/2020	33.9533	-117.3961
Slink	26759	8/29/2020	37.9219	-118.9529
Moc	2857	8/20/2020	38.0297	-119.9741
SQF Lightning Complex	174178	8/19/2020	36.1342	-118.8597
LNU Lightning Complex	363220	8/17/2020	39.1041	-122.2654

TABLE II
PRODUCTS USED FOR OBTAINING PARAMETERS

Parameters	Product	Satellite
ST, SW, RW	Level 1, Soil Moisture Active Passive (SMAP)	Sentinel
LAI, FPAR	Terra MODIS	Sentinel
NBR	Landsat ARD CONUS	Landsat
NDVI, EVI	Terra MODIS	Sentinel
DT, NT	Aqua MODIS	Sentinel

- *Day, Night Temperature and Surface Temperature* - directly influences fire ignition [3]. Normalized Burn Ratio - uses both near infrared and shortwave infrared wavelengths to study healthy vegetation and model future burn severity to better predict areas of potential high severity [2] [6].
- *Leaf Area Index (LAI)* - indicates crop development and growth, atmospheric circulation, biomass accumulation etc. [7].
- *Fraction of Absorbed Photosynthetically Active Radiation (FPAR)* - the fraction of the solar radiation absorbed by live leaves for photosynthesis activity.
- *Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI)* - drought indices that are used to predict fire danger [1].
- *Soil Moisture and Surface Wetness* - critical hydrological variables that represent the wetness state of the available fuel [11].
- *Rootzone Wetness* - critical for wildfire danger forecasting, can be used to prescribe optimal soil conditions to prevent fires and also predict the spread of invasive plants that exploit root zone soil moisture [5].
- *Vegetation Opacity* - measures the water content in the vegetation.

TABLE III
SAMPLE TEST BAND CLASSIFICATION FOR AREA BURNT, EVI, SURFACE TEMPERATURE

Bands	Acres Burnt	EVI	Surface Temperature (K)
Band 1	< 30,000	< 0.2	< 300.01
Band 2	< 60,000	< 0.25	< 304.93
Band 3	< 100,000	< 0.3	< 309.86
Band 4	< 150,000	< 0.35	< 314.78
Band 5	> 30,000	> 0.35	> 314.78

Over the course of 2002 to 2020, 101 fires were processed and 227 locations were analyzed. Python code was written to process individual CSV files, identify anomalies, calculate indices based on satellite band information, classify all parameters into test-bands for processing by Auto ML and ultimately generate a file to be fed to the ML tool. The Python code analyzed all available data (data is unavailable when cloud cover is too high or if the fires occurred prior to the satellite being launched). After studying the distribution of values for each of the twelve parameters, each parameter was divided into five sub-categories, called test bands (Table 3), based on magnitude. Google's machine learning tool was used to process and train data for 101 California fires, and generate a model for analyzing 227 locations all across California and predict their fire propensity.

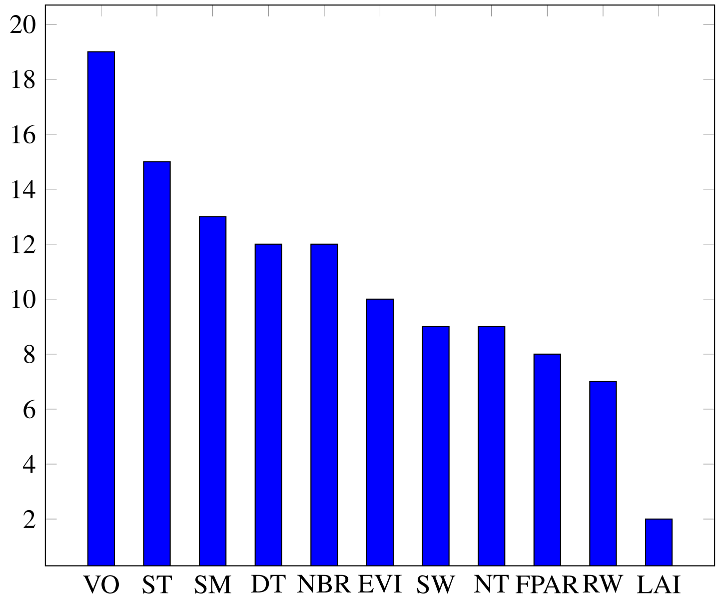

Fig. 1. ML Model Feature Importance Chart

III. RESULTS

Of the 12 parameters, the ML model found vegetation opacity to be of most importance, followed by surface temperature and soil moisture (Figure 1). Day and night temperature, NBR, EVI, surface wetness, FPAR and root zone wetness were found to be of equal importance. The Leaf Area Index was of least importance. The ML model performed at 95.3% precision. The performance of the ML model was measured by Area Under Curve (AUC-ROC) and was found to be 0.993. (Figure 2) The confusion matrix summarized the success of the classification model (Table 4).

In summary, the results from the ML model also concurred with existing research on invasive versus native species - invasive species have lower vegetation opacity and soil moisture compared to native species.

The ML model predicted that 27 locations in California had a probability of greater than 90% of having a fire impacting over 100,000 acres, while 14 locations had a probability of greater than 90% probability for having a fire burning 150,000 acres.

Heat maps were generated with R using the data from the ML model with all predictions (Figure 3).

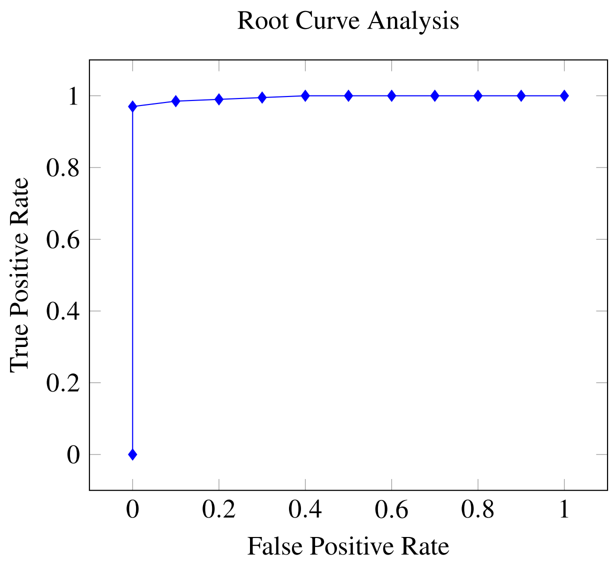

Fig. 2. Root Curve Analysis for Machine Learning

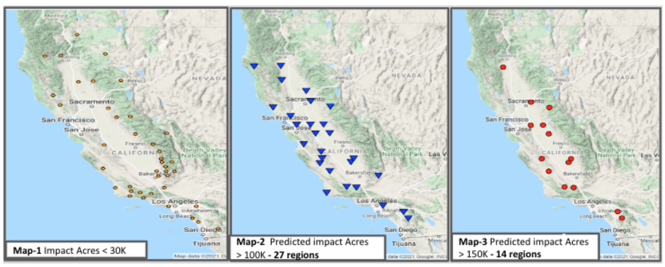

Fig. 3. Heat Maps Generated Using R

TABLE IV

ML Model's Confusion Matrix: True Labels Versus Predicted Labels

True Labels	1	2	3	4	5
1	100%	-	-	-	-
2	5 %	91 %	-	-	5%
3	-	-	94 %	-	6 %
4	-	-	-	100%	-
5	5%	-	-	-	95 %

IV. CONCLUSION

Previous studies have found surface temperature has an impact on igniting a fire. Soil moisture and vegetation opacity are critical hydrological variables representing the wetness of the fuel. NBR has been used to study healthy vegetation and model future burn severity. NDVI and EVI have been studied as drought variables. Root zone wetness can be used to prescribe optimal soil conditions and the spread of invasive species [8]. LAI and FPAR are parameters that have been used to study the health of vegetation.

This novel model combines these 12 satellite indices to offer a tool that is easily accessible and has a high precision. The prediction window could give valuable information to citizen scientists and ecological nonprofits to take preemptive measures. Pro-active vegetation management can be undertaken for high propensity areas in winter and spring - native rehabilitation, shrub management, reducing invasive species, controlled fires, building pathways, etc. Some measures include removal of invasive species and planting native plants.

V. FUTURE WORK

In the future, more variables such as elevation and wind speed will be added. Developing an app from which citizen scientists can access a heat map for restoration purposes is a next step. This vision for this app is that it can eventually process more fires and update itself with the latest satellite vegetation and temperature parameters.

In addition, considering the effect of man-made fuels on forest fires is part of the upcoming investigations. Furthermore, the analysis of more precise locations will be developed to pinpoint the exact location of interest. The ML model will also be cross-referenced to field studies to validate band values.

References

[1] Alireza Farahmand et al. "Introducing spatially distributed fire danger from earth observations (FDEO) using satellite-based data in the contiguous United States". In: *Remote Sensing* 12.8 (2020), p. 1252.

[2] Mohamed Wafik Hamoud. "Land Use Land Cover (LULC) and Burn Severity (dNBR) changes analysis in Latakia, Syria visualized on Web maps." In: ().

[3] Juliet Jiang, Monika Narain, and Isha Shah. "Fighting Fire with Fire: A Matrix-Based Approach for HRB vs. Logging Allocation and Fire Spread Minimization". In: *Triangle* (2021).

[4] Jay Kerby and Lisa Feldkamp. "Fighting Fire with Native Plants". In: ().

[5] Matthew R Levi et al. "Rating fire danger from the ground up". In: *Eos* 100 (2019).

[6] Sparkle L Malone et al. "Modeling relationships among 217 fires using remote sensing of burn severity in southern pine forests". In: *Remote Sensing* 3.9 (2011), pp. 2005–2028.

[7] Roya Mourad et al. "Assessment of leaf area index models using harmonized landsat and sentinel-2 surface reflectance data over a semi-arid irrigated landscape". In: *Remote Sensing* 12.19 (2020), p. 3121.

[8] Wanshu Nie et al. "Towards effective drought monitoring in the Middle East and North Africa (MENA) region: implications from assimilating leaf area index and soil moisture into the Noah-MP land surface model for Morocco". In: *Hydrology and Earth System Sciences* 26.9 (2022), pp. 2365–2386.

[9] Richard C Rothermel. *A mathematical model for predicting fire spread in wildland fuels*. Vol. 115. Intermountain Forest & Range Experiment Station, Forest Service, US …, 1972.

[10] AL Sullivan. "A review of wildland fire spread modelling, 1990-present 3: Mathematical analogues and simulation models". In: *arXiv preprint arXiv:0706.4130* (2007).

[11] Jaison Thomas Ambadan et al. "Satellite-observed soil moisture as an indicator of wildfire risk". In: *Remote Sensing* 12.10 (2020), p. 1543.

[12] Zhu Zhongming et al. "New tools to help "predict" forest fires". In: (2020).

[13] Kristin Zouhar, Jane Kapler Smith, and Steve Sutherland. "Effects of fire on nonnative invasive plants and invasibility of wildland ecosystems". In: *Wildland fire in ecosystems: Fire and nonnative invasive plants* 6 (2008), pp. 7–45.