CS230

Predicting California Wildfire Risk with Deep Neural Networks

Maria Jose Lozano Palacio
Department of Engineering Physics
Stanford University
majo@stanford.edu

Ian MacFarlane
Department of Electrical Engineering
Stanford University
ianpmac@stanford.edu

Abstract

We investigate the use of Deep Neural Networks to identify seasonal fire risk through satellite imagery of the California countryside. Using the pretrained ImageNet models MobileNetV2 and ResNetV2, we attempt to classify at-risk terrain informed by historical images in areas that experienced a wildfire during that fire season. We demonstrate initial struggles with effective identification due to high problem bias, and apply several techniques to reduce bias including learning rate optimization, addition of channel features, and cost weighting. We also posit several options for further work into a successful end-to-end California wildfire risk identification system.

1 Introduction

Over the past 4 years, the state of California has spent an average of $1 billion per year from its emergency fund to suppress wildfires [1]. Total damage estimates by third parties often place the cost of recent wildfire seasons as on the order of $100 billion per year [2]. Catching wildfires early and mitigating fire risk ahead of time can be hugely valuable as a cost- and life-saving measure. However, preventative fire measures require an accurate and easily-updated estimate of fire risk. Currently, such estimates are based on the the National Fire Danger Rating System (NFDRS), a sophisticated model designed in the 1970s incorporating humidity, fuel moisture, temperature, windspeed, precipitation, and local geography [3]. While newer sources of data like satellite imagery have been incorporated into parameter estimation for the NFDRS (e.g. using the greenness of an image to estimate live fuel moisture), end-to-end models using machine learning techniques to directly estimate fire risk from satellite imagery have only started to appear over the last few years. With our algorithm, we hope to show the viability of AI-based techniques to estimate fire risk in California.

The input to our algorithm is a high-resolution satellite image of the California countryside. We then use a deep neural network to output a predicted label for whether the imaged geography is at risk of wildfire for that season.

2 Related work

National Fire Danger Rating System: As described above, the current system publicly referenced on the National Park Service, the National Wildfire Coordination Group, and the US Forest Service websites is the NFDRS. The NFDRS relies mainly on locally observable metrics, such as humidity, precipitation, local geography, and dead/live fuel moisture.

Estimating Wildfire Risk with Satellite Imagery: Several papers [6][7] published in the last few years have applied machine learning techniques like neural networks and logistic regression

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

to estimate fire risk in India and Turkey. With the higher resolution of our Landsat imagery (in addition to the California focus), we hope to be able to achieve competitive performance and learn if higher-resolution imagery can yield more effective inference in this task.

Detecting Wildfire Smoke / Wildfire Spread with Satellite Imagery: Other papers have used deep neural networks to predict wildfire smoke [8] or spread [9] in real time using satellite imagery. While they aren't the exact same task as ours, and focus more heavily on real-time detection instead of risk forecasting, we've attached these papers since they use similar techniques and source data.

3 Dataset and Features

Our dataset is a amalgamation of two separate sources of data: the NSGS Landsat 7 and 8 imagery for the state of California over the last 10 years and CalFire's historical fire map. We generated a labelled dataset using these two sources by cross-referencing geographic boundaries of the historical fires with the Landsat imagery and labelling images of areas where a wildfire had occurred up to 6 months in the future (a common estimate for the length of a wildfire season) as "at risk."

The California fire map dataset was originally acquired from The California Department of Forestry and Fire Protection's Fire and Resource Assessment Program (FRAP). It contains information about fire names, fire perimeters, and alarm dates, that date as early as the 1910's. Given that we would like to cross-reference this dataset with the Landsat satellite imagery, we have chosen to focus on the fire data collected over the past 10 years. Since the California fire map dataset was originally represented as geographic information system (GIS) data, and not as a standard .csv or .txt format, we used an intermediate GIS application called QGIS to reproject the data and convert it to a .csv format. Using this platform, we were able to extract important fire information such as date, latitude, longitude, and area. With this information, we managed to approximate features like fire radius to better match satellite images to these data points. More on this approach can be found in the following paragraph.

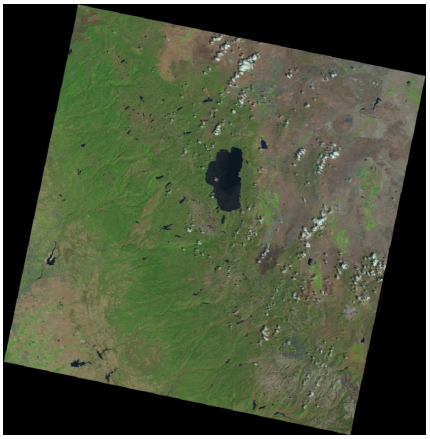
The original Landsat 7/8 dataset as downloaded from USGS EarthExplorer is comprised of roughly 11000 unlabeled high-resolution satellite images taken since January 1, 2012. However, given the size and resolution of the images (especially compared to the size of the most wildfires), we split each image into a 10x10 grid of sub-images, yielding a functional dataset of 11 million images. Images are then given a label directly after slicing. This is done with a simple radius estimate comparison between the fire dataset and imagery dataset. For each image, the lat/long center of the image is compared with the lat/long center of every fire in California over the past 10 years. If the centers are within the radius of the fire (estimated from fire area) plus the radius of the image and the fire is within 6 months (in the future direction) of when the image was taken, then the image is labeled as "at risk." We split this according to a 9:0.5:0.5 ratio into training, validation, and test datasets.

We had to address a few non-idealities of our dataset, one example being the obfuscation of some images by cloud cover. We were able to automatically filter the images to limit the presence of cloudy data, but some examples did make it through the automatic filtering (more discussion in Results). We also had to confront the imbalance of our two classes, since the not-at-risk class was heavily overrepresented (by a 10:1 ratio). More details are presented in the results, but we partially mitigated this through a combination of data augmentation, class weighting, and undersampling of the not-at-risk class.

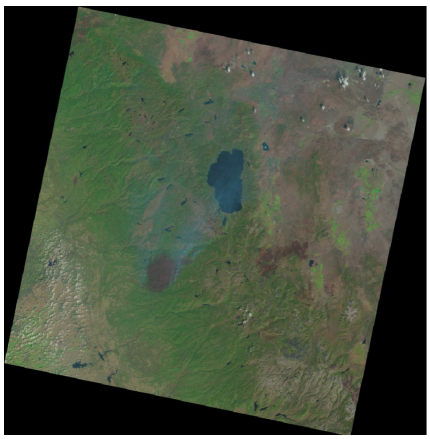
4 Methods

Due to their ability to easily identify feature maps, convolutional neural networks (CNNs) were used for our fire risk estimation model. While landscape identification may be different from object detection, we hoped that some of the internal features of object detection (such as edge groupings or color patterns) could capture topological or greenery-based information from the satellite imagery. We limited ourselves to using residual network models, which use a technique called skip connections. This technique offers an alternate shortcut by skipping training from a few layers and connecting directly to the output, helping the vanishing gradient problem. Skipping connections also ensures that the higher layers of the model learn as well as the lower layers.

For our models, we turned to two pretrained CNN implementations optimized on ImageNet, ResNet50 (mentioned in class) and MobileNetV2. While ResNet50 is based on a residual structure that has 50



(a) Satellite photo centered on South Lake Tahoe one month before the 2021 Caldor Fire (grassy lower left)



(b) Satellite photo centered on South Lake Tahoe during the 2021 Caldor Fire (smoke and burn region clearly visible)

Figure 1: Two dataset images

layers [10], MobileNetV2 is based on an inverted residual structure where the input and output of the residual block are thin bottleneck layers opposite to traditional residual models [11]. Additionally, MobileNetV2 uses lightweight depthwise convolutions to filter features in the intermediate expansion layer.

In order to best match the data to both of these models, we limited our image band data to RGB channels. Having picked testable models and given that our dataset is structured to make this a fairly end-to-end logistic regression problem, we then use the cross entropy loss:

$$L(\hat{y}, y, m) = -\frac{1}{m}(ylog(\hat{y}) + (1-y)log(1-\hat{y}))$$

In order to judge the effectiveness of our "fire risk classifier."

5 Experiments/Results/Discussion

We chose to keep a relatively small learning rate ranging from 0.001-0.00001 (varying by learning stage and trainable portion of the pretrained CNNs), in order to best preserve the previous knowledge of the two CNNs (since high learning rate increases the risk of deviating from pretrained weights). In addition to tuning the learning rate, we chose to use small batches of size 32. We viewed this batch size to be large enough to avoid most of the noise of stochastic GD (which mattered given that our models converged quickly), while still being small enough to train in reasonable time steps.

Upon initial training, our model quickly converged to a high training accuracy (95%). Unfortunately, due to the data imbalance mentioned earlier, this was because the model had learned to classify almost every input as negative. We then compensated for this by weighting the cross-entropy cost of positive examples higher and undersampling the negative examples. After this modification, we still encountered a high-bias performance barrier visible in Figure 3.

In response to the high bias demonstrated by our initial results in Figure 3, we also introduced the depth of the spectral data / number of channels (variable from 1 to 9) as a hyperparameter. The rationale for this change was that introducing new features – possibly indicating moisture or other indirectly measurable parameters that affect fire risk – would allow the model to better distinguish between at-risk and and not-at-risk images. However, neither ResNet50V2 nor MobileNetV2 were initially built to work with more than 3 input channels, so we added a translation layer convolving the 9-channel input to 3 channels before feeding it into the model. However, we found this simple form of adding more features failed to improve the accuracy significantly.

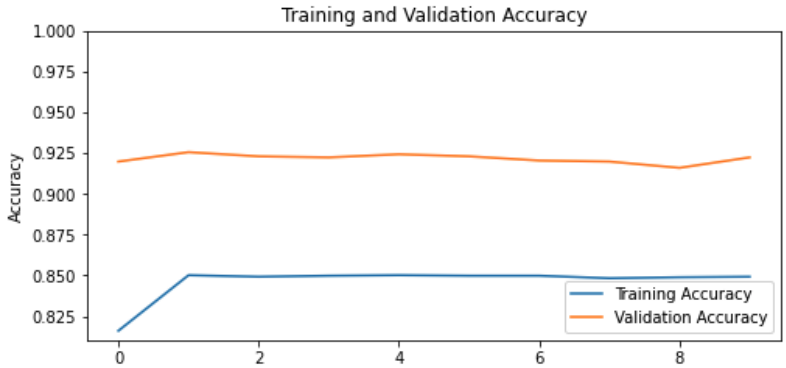


Figure 2: Initial training results had high bias and converged quickly, prompting the introduction of more spectral data per image (features).

Test Set Confusion Matrix (normalized to number of examples)	Label Not-at-risk	Label At-risk
Prediction Not-at-risk	0.732	0.242
Prediction At-risk	0.025	0.001

Figure 3: Confusion Matrix of the better model (MobileNetV2), after adding higher spectral data

At this point, we diagnosed more specific performance issues with this implementation using Accuracy, Precision, and Recall, each given by:

$$A = \frac{TP + TN}{TP + FP + TN + FN}, P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}$$

As can be seen from the Confusion Matrix and the APC scores, the model still struggles significantly to identify positive classifications.

6 Conclusion/Future Work

As mentioned earlier, one of the factors contributing to our models insufficient ability to identify positive example is the disproportionate imbalance between classes. As a potential solution, we propose that future work focus on collecting more positive examples while also investigating the phenomena underlying those positive examples. For example, fires caused by some phenomena like lightning strikes may be most heavily represented as areas that are frequently cloudy. Since we initially filtered cloudy data, this information would have been missed. Generally speaking, having a fairly represented "at risk" class (both in example count and in the underlying features) will help the model learn better. Another thing to consider would be adopting a different approach when assigning labels. Currently, labelling is done by cross-referencing the fire location with the date of the fire. While this approach sounds reasonable, it doesn't take into account the cause of the fire. Many fires are indeed caused by factors characteristic of local geography, whose features can be extracted from a satellite image, but there are fires that can also be caused by human incidents. Unfortunately, in these cases it is most likely that there will be no features to extract from a satellite image, so an instance of such a fire should not be included in the "at risk" class. Another potential solution would be to consider the use of deeper spectral data with CNNs pretrained to handle more than 3-channel data. This could perhaps allow the added features to better propagate into the model and lead to the model being better able to learn features linked to positive classification.

Accuracy	Precision	Recall
0.733	0.04	0.004

Figure 4: Accuracy, Precision, and Recall for the high learning-rate, added-spectral-data MobileNetV2 model

Figure 5: False negative example where the grayscale intensity maps to 510nm (greenish) channel intensity.

We believe that implementing these changes will result in higher APC scores and a better Confusion Matrix. While we only had a couple of weeks to complete this project, we consider that this is a problem worth investing in and hope to continue putting the time into making this a better fire risk estimation model. An effective model for such purposes would not only save billions of dollars, but it could also save thousands of lives.

7 Contributions

Both team members contributed equally to each part of the project (literature search, data acquisition, preprocessing, model iteration, and this report), with the exception that Ian spoke for the presentation because Maria Jose was sick.

References

[1] California Department of Forestry and Fire Protection. 2019. https://www.fire.ca.gov/media/px5lnaaw/suppressioncostsonepage1.pdf

[2] Roman, J.; Verzoni, A.; Sutherland, S. 2020. "Greetings from the 2020 Wildfire Season," NFPAJournal, https://www.nfpa.org/News-and-Research/Publications-and-media/NFPA-Journal/2020/November-December-2020/Features/Wildfire

[3] Deeming, John E.; Burgan, Robert E.; Cohen, Jack D. 1977. "The National Fire-Danger Rating System." Gen. Tech. Rep. INT-39. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station.

[4] Wildland Fire Learning In Depth Series. 2021. National Park Services, https://www.nps.gov/articles/understanding-fire-danger.htm

[5] Jafari Goldarag, Y.; Mohammadzadeh, A.; Ardakani, A.S. 2016. "Fire Risk Assessment Using Neural Network and Logistic Regression." J Indian Soc Remote Sens 44, 885–894. https://doi.org/10.1007/s12524-016-0557-6

[6] Larsen, A.; Hanigan, I.; Reich, B.J. et al. 2021. "A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication." J Expo Sci Environ Epidemiol 31, 170–176. https://doi.org/10.1038/s41370-020-0246-y

[7] Zhang, G. et al. 2021. "Deep neural networks for global wildfire susceptibility modelling," https://www.science direct.com/science/article/pii/S1470160X21004003

[8] Ray, J. et al. 2021. “Predicting wildfire burns from big geodata using deep learning,” https://www.sciencedirect.com/ science/article/pii/S0925753521001211

[9] Chollet, F. et al. 2015. Keras. GitHub. Retrieved from https://github.com/fchollet/keras

[10] He, K.; Zhang, X.; Ren, S.; Sun, J. 2015. "Deep Residual Learning for Image Recognition," https://arxiv.org/abs/1512.03385

[11] Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. 2018. "MobileNetV2: Inverted Residuals and Linear Bottlenecks," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520, https://arxiv.org/abs/1801.04381