# Predicting Wildfire Risk Based On Land Cover Classification and Past Wildfire Data in California

Atharva Kulkarni, Akshada Guruvayur, Ananya Chakravarthi, Cristina Marculescu
Mentors: Andrew Liu, Benjamin Herschman

## Abstract

Wildfires have the capacity to destroy forests, homes, and the health of ecosystems. As a result of rising global temperatures, wildfires have experienced increased frequency and severity, leading to devastating outcomes throughout sections of North America. All of these factors have led to an increase in demand for tools that can accurately and efficiently identify potential risk areas, particularly after the devastating California wildfires of the 2010s and 2020s. Our project aims to analyze pictures taken by NASA's GLOBE Observer app to identify land cover types in order to classify their potential contribution to wildfire risk in a given region. Additionally, we also looked at past wildfires in California from 200-2018 as another factor for future wildfire risk. We created our model using Python, the FEMA Wildfire Risk Database and the GLOBE Observer database. The two features of land cover type and historical wildfires are paired with the FEMA wildfire risk database to determine the wildfire risk. We used the random forest algorithm to ensure that each decision tree in our algorithm contributes to the final prediction, with the most frequent risk level chosen as the output. This approach ensures robustness and accuracy by combining the insights from multiple trees.The model uses these features to make predictions about wildfire risk, assigning a risk level of high, moderate, or low based on the patterns it learned during training. In the end, we found that forests were least susceptible to wildfire spread while Urban areas presented a major threat to wildfire risk. Scientists can use our algorithm to analyze the risk of wildfires, as its classification of various land cover variables can be a determinant of which preventive measures should be taken to protect wildlife, vegetation, and communities in the midst of wildfires. By identifying the different types of wildfire risks found in a location, our framework can not only serve as a baseline for determining which of the risks are most relevant for a given region, but can also provide scientists with real-time, ad-hoc data, which could potentially assist in more accurate visualization of wildfire risks when satellites may not accurately capture land cover imagery.

In the late 2010s and early 2020s, California experienced a series of devastating wildfires that decimated homes, forests, and habitats. Tracking and predicting wildfires has become a focus for many scientists, hoping to mitigate the effects of these natural disasters.[1] Additionally, as global temperatures continue to rise due to climate change, severe weather events such as these become more common and more destructive; our team hopes to take a new approach to this problem.

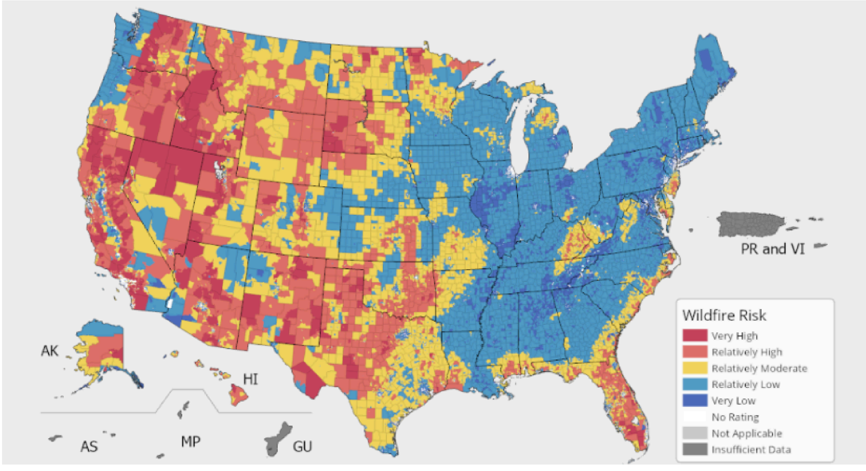
**Objective: To find correlation between land cover types and wild fire risk**

- Identifying which types of land cover are most susceptible to wildfires can improve risk assessment and management strategies.
- Better understanding helps in prioritizing resources for prevention and mitigation efforts.
- Informs the development of targeted mitigation strategies, such as controlled burns or vegetation management, to reduce fire risk.

## Methodology

1. Data from past California wildfires to predict fire risk 2000-2018 with randomly selected GLOBE Observer points
2. Determine 2 input variables.
   a. Whether there is is or isn't a fire in that area from 2000-2018
   b. type of land cover index rating each location was given to make our algorithm adaptable
      i. Forest, grassland/shrubland, urban, and body of water - index obtained from the FEMA Fire Risk Map and location classified using GLOBE Observer images
3. Python algorithm analyzes what, if any, correlation there is between certain land cover types and the risk of fire.

## Data & Model



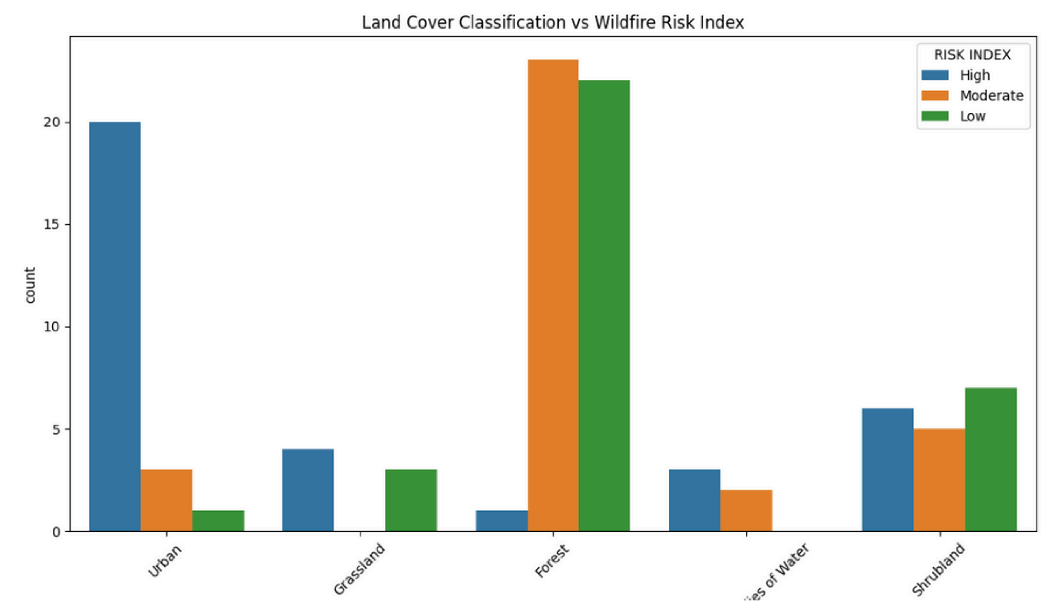

### FEMA Risk Map

We took data from this FEMA map to quantify risk and correlate it with various land cover types

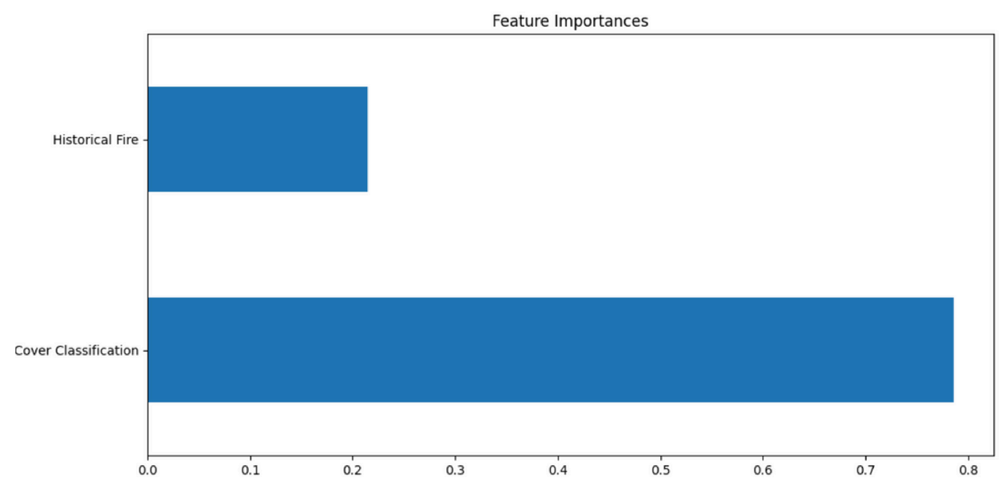
### Machine Learning Model

The model our team created looks at GLOBE Observer images as well as historic fires to determine relative probabilities of a fire at that position

## Results & Model




**Wildfires had low-moderate risk in forest and high risk rate in urban area.**




**This graph highlights the significance of cover classification in enhancing the accuracy of the ML model.**

In our project, we developed a machine learning model using random forests to predict wildfire risk. Our model processes input data, which includes the type of land cover (e.g., forest, urban, grassland) and whether there has been a historical fire in the area. The model uses these features to make predictions about wildfire risk, assigning a risk level of high, moderate, or low based on the patterns it learned during training. Each decision tree in our random forest contributes to the final prediction, with the most frequent risk level chosen as the output. This approach ensures robustness and accuracy by combining the insights from multiple trees.

Our model had a 75% accuracy rate

## Analysis

For our data analysis, we used two main approaches:

### Random Sampling

1. A sample of 100 randomly selected points was taken:
2. Each point was identified and labeled based on row number
3. The longitude and latitude were inputted into the FEMA Fire Risk Map, so we could see the fire risk at that point, and the GLOBE Observer app to see North, East, South, and West pictures at that location and classify the land cover.

### Stratified Sampling

1. Our data set was split into THREE categories (or stratas) of "very high risk of wildfire", "relatively moderate risk of wildfire" and "very low risk of wildfire"
2. A sample of 33 randomly selected points was taken from EACH STRATA: (The range inputted into the number generator varied based on the range of row values given for each strata)
3. Each point was identified and located based on row number: (if 430 was generated, row number 430 would be looked at)

## Conclusions

Based on our model, we confirmed a correlation between land cover type and fire risk, with both methods showing similar accuracy. Upon further analysis, we found out that urban areas have a high wildfire risk due to combustion which supported our research. Additionally, forests had high humidity levels which prevented wildfire spread. Using land cover classification significantly improved our model's accuracy, enhancing future fire prediction and management. Interestingly, historic fires in the area did not significantly impact the predictions, suggesting that other factors play a more crucial role in determining fire risk. However, there is a lack of data points in specific areas of California, highlighting the importance of citizen science. Engaging the public in data collection can fill these gaps, improving the model's precision and aiding in better wildfire management and response strategies.

## Future work

One way we'd like to expand our work in the future is by using a Convolutional Neural Network (CNN) model to perform these classifications and correlations. With the CNN model, less human supervision would be required, as the model would be capable of automatically extracting data on its own. This way, our work could stay up to date much faster and be more adaptable to our planet's changing climate.

## Acknowledgements

Thank you to our Earth System Explorer mentors: Dr. Russanne Low, Dr. Cassie Soeffing, Dr. Peder Nelson, Dr. Erika Podest, Mr. Andrew Clark

The material contained in this poster is based upon work supported by the National Aeronautics and Space Administration (NASA) cooperative agreements NNX16AE28A to the Institute for Global Environmental Strategies (IGES) for the NASA Earth Science Education Collaborative (NESEC) and NNX16AB89A to the University of Texas Austin for the STEM Enhancement in Earth Science (SEES). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.