Integrating Machine Learning and Citizen Science in CS-FLARE for Real-Time Wildfire Risk Assessment

Student Researchers: Leena Dudi, Sanaa Mulay, Anjali Singh, Ashvin Tiwari, Amogh Thodati, Vincent Villarreal, Zane Zacharia, Arnold Zhang, and Charlotte Zhou
Mentors: Dr. Russanne Low, Cassie Soeffing, Peder Nelson, James Ervin, and Nikita Agrawal
July 2024

Abstract

Wildfires pose an increasingly significant threat to natural ecosystems, private property, and public health. Precautionary methods can be taken to mitigate the possibility of a large-scale wildfire. Our research aims to inform citizens of wildfire risk factors on their properties by developing a mobile application, CS-FLARE (Citizen Science: Fire Likelihood and Risk Evaluation), that analyzes wildfire risk on a local community level based on a culmination of factors that compose a typical wildfire. CS-FLARE uses YOLOv8, an image segmentation model to identify the presence of flammable materials in the four cardinal directions at each Area of Interest (AOI) in the GLOBE Observer database. We also utilize Google Teachable Machine to train a machine learning model to classify the downward photo at each AOI by ground moisture. The app implements a U-Net model trained with satellite imagery, temperature, precipitation, elevation, and slope datasets from Earth Map to provide users with accurate predictions validated by existing wildfire risk assessment datasets. Utilizing our identified fire risk factors and satellite datasets, we have created an algorithm that provides citizens with a quantified wildfire risk rating, while offering comprehensive insights of surrounding risks that can help prevent future wildfires. CS-FLARE's user-friendly application interface, created on Flutter (Google's multi-platform app development tool), presents these insights by proposing potential actions to minimize wildfire risk in the pictured area. We hope to integrate this app into local fire departments and environmental agencies so that citizens have a trusted tool to better understand the risks and impacts that wildfires can have on the area around them.

Keywords: GLOBE Observer, Wildfires, Machine Learning

1. <u>Introduction</u>

Wildfires have become an increasingly prevalent and destructive force, burning over 12.6 million acres of land globally since 2022. While we often associate wildfires with rural and forested areas, it's crucial to recognize the severe damage they can inflict on urban and suburban communities. The consequences of wildfires extend beyond the loss of life and property, affecting air quality, damaging infrastructure, and destroying power grids (*5 Ways Wildfires Can Affect Cities*, 2023). Additionally, vital resources such as water supply and agriculture are jeopardized due to the changes in erosion patterns and sediment concentrations, leading to long-term environmental impacts (Smith et al., 2010). Many studies have shown that low-income communities face a much grimmer outcome due to issues like inadequate insurance, the high cost of fire protection, and a lack of prevention and clean-up resources, pushing the communities deeper into poverty. In fact, over 103,900 homes have been burned by wildfires in the past 18 years, highlighting the need for effective mitigation strategies (Barrett, 2024). While it is true that wildfires are natural disasters and therefore are often unpredictable, it is also true that the public can take steps to protect themselves and their communities from these cataclysms and alleviate this risk. The National Park Service, referencing federal data, emphasized that nearly 85% of wildfires are caused by human activities, emphasizing the critical need for citizen participation in wildfire prevention (*Wildfire Causes and Evaluations*, n.d.).

A study conducted in 2022 used remote sensing and geographic information science to map forest fire risk in Gachsaran, Iran. They were able to use GIS-based decision analysis and analytical network processes to examine criteria including soil moisture and land surface temperature to identify areas with a high fire risk (Feizizadeh et al., 2022). Despite the significant measures already taken by experts to mitigate fire risks on a larger scale, there is a growing need to empower citizens to better protect themselves. By utilizing citizen science, we aim to involve the general public more actively in wildfire risk management for their own homes and suburbs. GLOBE Observer is a citizen science program that allows individuals to collect and submit environmental data to support scientific research. While the GLOBE Observer application's land cover feature allows citizens to upload images to a central database, it is not able to return the user information about the location's fire risk.

Using our AI models and data from the GLOBE Observer database, we were able to create an application that can analyze user-inputted images to detect possible fire risk factors such as brush, trees, grass dryness, and leaf litter. These factors are then combined to return a risk rating along with strategies for the user to minimize the location's fire risk. By integrating machine learning and satellite data, CS-FLARE provides users with real-time information about fire risk, empowering them to take proactive measures to protect themselves and mitigate the devastating impacts of wildfires. Utilizing advanced models like YOLOv8 for image segmentation, Google Teachable Machine for image classification model training, and a U-Net model for comprehensive risk predictions, CS-FLARE enhances the accuracy of wildfire risk assessments. This approach fosters community engagement in wildfire prevention efforts, making our app a valuable tool for citizens.

2. **Methods**

2.2. NESW Image Segmentation Model

GLOBE Observer Land Cover Observations taken in the north, south, east, and west directions were compiled into a database of X images depicting typical land cover conditions across the United States.

2.2.1 NESW Image Preprocessing

In order to train the models behind CS-FLARE we first needed to collect citizen science-sourced images. We downloaded 979 images taken in the north, east, south, and west directions from the GLOBE Observer's Landcover Database from the NASA SEES 2024 AOI (Area of Interest) pictures. Areas of brush, trees, grass, and leaf litter were manually labeled using polygonal segmentation on Roboflow, an open-source dataset management platform. Using Roboflow's built in tools for datasets, we then resized all images to a standard dimension of 640x640 pixels and globally applied a histogram equalization based contrast filter to enhance feature extraction.

The dataset was augmented using a series of image filters: rotation, crop, exposure, blur, and noise. The image augmentation process was conducted to ensure that the model is robust to a variety of environmental conditions and to mitigate data availability constraints, bringing the total image count up to 1802 images. The dataset was split into training, validation, and testing sets in a 70-15-15 ratio.

2.2.2 Training our YOLOv8 model for Vegetation Segmentation

We trained the YOLOv8 small segmentation model ('yolov8s-seg.pt') for vegetation segmentation in Google Colab. Our dataset of 1802 labeled images was imported into the Python notebook using Roboflow's API, allowing the data to be exported in the YOLOv8 dataset format most suitable for our model. The model was trained for 100 epochs with a batch size of 16, using YOLOv8's pretrained weights as a baseline for transfer learning. We used the AdamW optimizer in our training to ensure the best performance, which started with an initial learning rate of 0.00125 and a patience of 50 to prevent overfitting. Training was conducted on a Tesla T4 GPU with CUDA support in Google Colab.

2.3. Grass Moisture Evaluation Model

Using GLOBE Land Cover observations taken in the downward direction, a model using Google Teachable Machine was trained to recognize ground moisture in the pictures as either a healthy, slightly dry, or very dry.

2.3.1 Downward Image Preprocessing

Similar to the training process of the NESW model, citizen science-sourced images from GLOBE observers were used to train this grass moisture evaluation model. A dataset of strictly downward images was cleaned and removed of duplicates leaving 102 images. This dataset was labeled and classified into 3 groups: Healthy (30 images), Slightly Dry (36 images), Very Dry (36 images). This dataset, despite being small, was sufficient to bring this model high accuracy.

2.3.2 Training our Downward Image Grass Moisture Evaluation

Within Google Teachable Machine, the labeled dataset was split into 3 classes based on grass dryness. The model was trained for 50 epochs with a batch size of 16 and a learning rate of 0.001. It was then exported as a TensorFlow Lite model to allow for utilization in mobile applications.

2.4. Satellite Data Gathering

Daymet V4 Daily Surface Weather and Climatological satellite was used to obtain daily temperature, water vapor pressure, and incident shortwave radiation flux density. The Terraclimate satellite dataset was used to gather precipitation accumulation. Additionally, we obtained fire mask labels from the MODIS sensor aboard NASA's Terra satellite. These labels serve as ground truth data for training our machine learning model, categorizing areas into different fire risk levels. These datasets were collected through Google Earth Engine (GEE) APIs. The dataset size of 500 images were used for training the U-Net image segmentation model. Prior to training the model, necessary preprocessing steps, such as normalization and filtering, were done in Google Colab.

2.5. App Framework

In order to make CS-FLARE more reachable to everyone, the app was coded in Flutter using the Dart language, allowing for integration across mobile operating systems such as iOS and Android. This made it the perfect software to build CS-FLARE on in order to allow as many people the opportunity to use the app no matter what operating system they are using. To first plan out the creation of CS-FLARE, Figma, an extremely customizable wireframing tool, was used. This helped create three main pages for the app in order to keep the process as streamlined as possible. In addition to this though, behind the user side of CS-FLARE, Tensorflow Lite was used to integrate the AI models within the app.

3. Results

In the following section, we analyze the performance metrics of the machine learning models deployed in CS-FLARE.

3.1 NESW Image Segmentation Model

The vegetation segmentation model achieved a mean average precision (mAP50) value of 0.439 on the validation dataset. The class-specific performance showed that the model performed best on the "grass" class with an mAP50 of 0.743, while the 'leaf litter' class exhibited the lowest performance with an mAP50 of 0.0571. The performance of the model is shown in Figure 1, which shows the confusion matrix of the model.

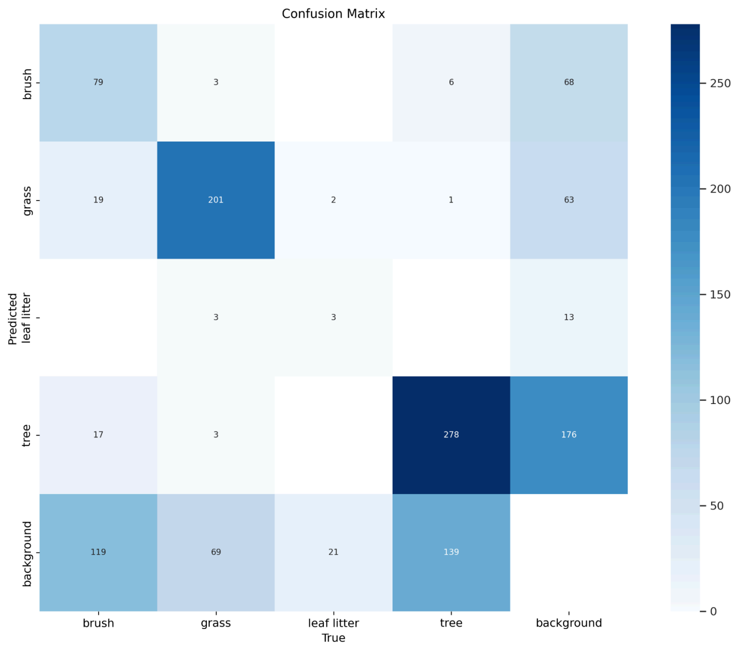


Figure 1: Confusion matrix of YOLOv8 vegetation segmentation model, depicting true versus predicted class labels for the validation dataset. The diagonal values represent correct classifications, while off-diagonal values indicate misclassifications.

Figure 2: Segmentation masks generated by our YOLOv8 model. Vegetation masks are color-coded by class, with each color representing a different vegetation type: trees are yellow, brush is red, grass is pink, and leaf litter is orange.

3.2 Grass Moisture Evaluation Model

The Grass Moisture Classification Model performed at around .94 accuracy at the 49th epoch. The performance of the model is shown in Figure 3, which depicts the accuracy per epoch of this model.

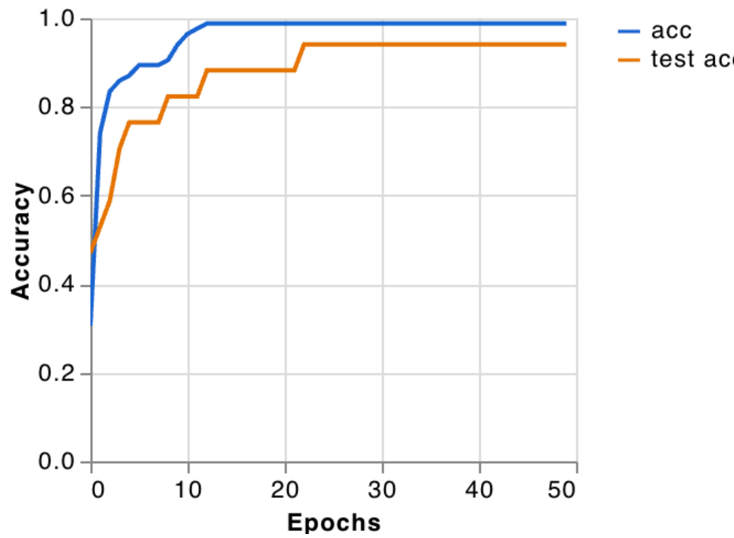


Figure 3: Accuracy per Epoch graph of model depicting models accuracy and test accuracy over 50 epochs.

3.3 Satellite Imagery Segmentation Model

The satellite imagery U-Net model achieved an accuracy of 0.48 after training for 50 epochs. Training the model further caused overfit due to the small dataset size. The model performs better for differentiating between no fire and fire risk. It does not perform well for differentiating amongst low, nominal, and high fire risks. However, for a community level prediction, fire vs. no fire is the most important distinction. It provides actionable information, especially if there is imminent fire risk.

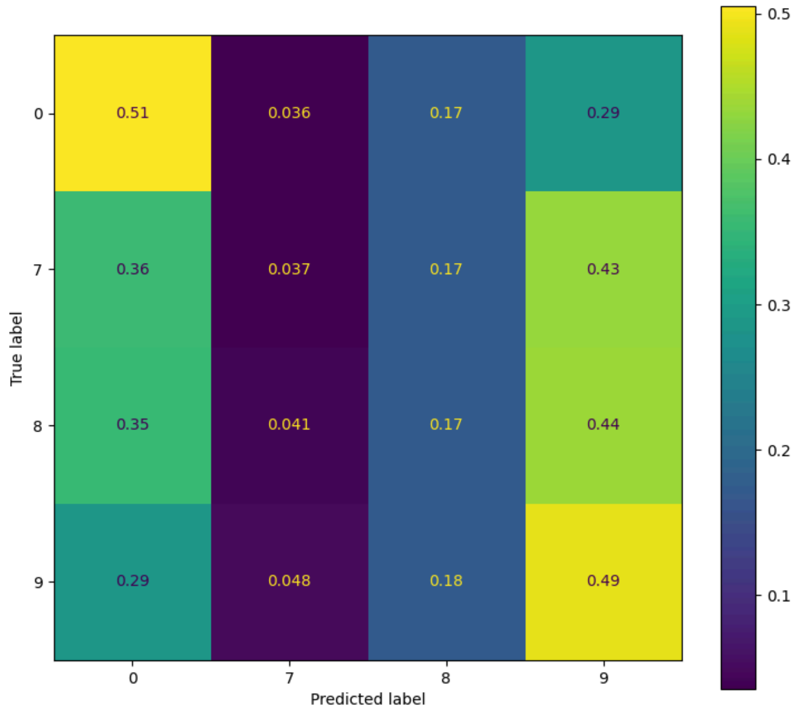


Figure 4: Satellite imagery confusion matrix for fire risk levels 0 - no risk, 7 - low risk, 8 - nominal risk, and 9 - high risk at 50 epochs.

Discussion

The constraints of the datasets used in the study suggest areas of improvement for future study. Though the dataset was manually filtered, many of the GLOBE Observer images varied in quality and orientation, contributing to image distortion and misrepresentation in the final dataset. Given that the NESW image dataset for vegetation segmentation was manually labeled and filtered, we were only able to sample from a base dataset of 979 images, which could contribute to overfitting on the training data. The relatively low performance of the vegetation segmentation model on the “leaf litter” class is likely due to the underrepresentation of leaf litter labels in the training data. In the future, we look to obtain more data, particularly of images with leaf litter, to help our model better identify vegetation classes more accurately. In addition to this, the downward imaging model was only trained on 102 images. Although it still showed impressive results, this limited sample size may have restricted the model's ability to generalize across different environmental conditions. Therefore, increasing the number of images the model is trained on is one future area of improvement. For the satellite imagery, the Daymet and MODIS image resolutions are 1km. This can only facilitate predictions for a wide area rather than at specific spots in a neighborhood. Therefore, these are suitable for community level predictions. In the future, a combined model can be built utilizing all three datasets, which will allow for a tiered wildfire risk assessment.

Conclusion

In this study, we introduce CS-FLARE (Citizen Science: Fire Likelihood and Risk Evaluation), a novel approach to citizen science and machine learning in a mobile application format to contribute to wildfire risk reduction. We highlight a three-pronged approach to wildfire risk reduction on a citizen level, utilizing data from the GLOBE Observer Land Cover Database and Google Earth Engine to compile a framework for comprehensive real-time wildfire risk assessment.

This study underscores the untapped potential of citizen science in earth science research. By empowering citizens to conduct risk evaluations at the property level, CS-FLARE facilitates greater public involvement in safeguarding homes and local environments. Furthermore, citizen science initiatives like CS-FLARE can bridge the gap between the public and scientific communities, encouraging collaboration and shared responsibility in addressing environmental challenges. The data collected through such platforms can inform local policies, contribute to large-scale environmental monitoring, and enable the development of tailored, community-specific strategies for disaster prevention and response.

In the future, we hope to work with local government agencies to expand on the app's capabilities and begin distributing CS-FLARE to citizens for additional testing and use. By doing so, we hope to enhance the accuracy of wildfire risk assessments, improve public awareness and preparedness, and ultimately contribute to the reduction of wildfire-related damages on a global scale.

Data Availability

These data were obtained from the GLOBE Program. GLOBE Observer data were obtained from NASA and the GLOBE Program and are freely available for use in research, publications, and commercial applications. GLOBE Observer data analyzed in this project are publicly available at globe.gov/globe-data (accessed 31 July 2024). The Python code to read, analyze, and visualize GLOBE data for this article and the analyzed datasets are available on github.com/IGES-Geospatial (accessed 31 July 2024). Dashboard access to Mosquito Habitat Mapper and Land Cover data is available at geospatial.strategies.org (accessed on 31 July 2024).

Acknowledgements

The authors would like to acknowledge the support of the 2024 Earth System Explorers Research Team, NASA STEM Enhancement in the Earth Sciences (SEES) Virtual High School Internship program. Special thanks to Dr. Rusty Low, Cassie Soeffing, Peder Nelson, Andrew Clark, James Ervin, and Nikita Agrawal for guidance when preparing this work.

The NASA Earth Science Education Collaborative leads Earth Explorers through an award to the Institute for Global Environmental Strategies, Arlington, VA (NASA Award NNX6AE28A). The SEES High School Summer Intern Program is led by the Texas Space Grant Consortium at the University of Texas at Austin (NASA Award NNX16AB89A).

The authors have no competing interests to declare.

Author Contributions:
Data Annotation: Ashvin Tiwari, Charlotte Zhou, Sanaa Mulay, Arnold Zhang, Anjali Singh; Machine Learning Models: Charlotte Zhou, Amogh Thodati, Anjali Singh; App Development: Leena Dudi; Writing-Original Draft: Ashvin Tiwari, Charlotte Zhou, Sanaa Mulay, Arnold Zhang, Amogh Thodati, Anjali Singh, Vincent Villarreal, Zane Zacharia; Writing—review and editing: Ashvin Tiwari, Charlotte Zhou, Sanaa Mulay, Arnold Zhang, Amogh Thodati, Anjali Singh, Zane Zacharia.

References

Barrett, K. (n.d.). Wildfires destroy thousands of structures each year. Headwaters Economics. Retrieved July 31, 2024, from https://headwaterseconomics.org/natural-hazards/structures-destroyed-by-wildfire/#:~:text=Since%202005%2C%20wildfires%20have%20destroyed,%2C%20businesses%2C%20and%20other%20structures

Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision.

5 ways wildfires can affect cities - even thousands of miles away. (2023, June 14). World Economic Forum. Retrieved July 31, 2024, from https://www.weforum.org/agenda/2023/06/wildfires-affect-cities-air-quality-water-shortage-floods/#:~:text=In%20recent%20years%2C%20wildfires%20have,short%20and%20the%20long%20term

Smith, H. G., Sheridan, G. J., Lane, P. N.J., Nyman, P., & Haydon, S. (2010, November 4). Wildfire effects on water quality in forest catchments: A review with implications for water supply (V. Andréassian & A. Bardossy, Eds.). https://doi.org/10.1016/j.jhydrol.2010.10.043

Wildfire causes and evaluations. (n.d.). National Park Service. Retrieved July 31, 2024, from https://www.nps.gov/articles/wildfire-causes-and-evaluation.htm#:~:text=Nearly%2085%20percent*%20of%20wildland,and%20intentional%20acts%20of%20arson.&text=Lightning%20is%20one%20of%20the%20two%20natural%20causes%20of%20fires

Bakhtiar Feizizadeh, Davoud Omarzadeh, Vahid Mohammadnejad, Hoda Khallaghi, Ayyoob Sharifi & Bahaoldein Golmohmadzadeh Karkarg (2023) An integrated approach of artificial intelligence and geoinformation techniques applied to forest fire risk modeling in Gachsaran, Iran, Journal of Environmental Planning and Management, 66:6, 1369-1391, DOI: 10.1080/09640568.2022.2027747

Agrawal N, Nelson PV, Low RD. A Novel Approach for Predicting Large Wildfires Using Machine Learning towards Environmental Justice via Environmental Remote Sensing and Atmospheric Reanalysis Data across the United States. *Remote Sensing*. 2023; 15(23):5501. https://doi.org/10.3390/rs15235501

Chuvieco E, Yebra M, Martino S, Thonicke K, Gómez-Giménez M, San-Miguel J, Oom D, Velea R, Mouillot F, Molina JR, et al. Towards an Integrated Approach to Wildfire Risk Assessment: When, Where, What and How May the Landscapes Burn. Fire. 2023; 6(5):215. https://doi.org/10.3390/fire6050215

Mhawej M, Faour G, Adjizian-Gerard J. Wildfire Likelihood's Elements: A Literature Review. Challenges. 2015; 6(2):282-293. https://doi.org/10.3390/challe6020282

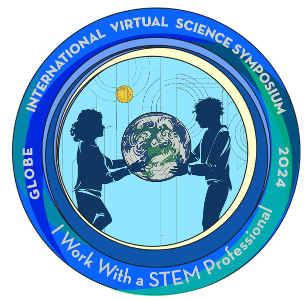
Integrating Machine Learning and Citizen Science in CS-FLARE for Real-Time Wildfire Risk Assessment

Students: Leena Dudi, Sanaa Mulay, Anjali Singh, Ashvin Tiwari, Amogh Thodati, Vincent Villarreal, Zane Zacharia, Arnold Zhang, and Charlotte Zhou
Mentors: Dr. Russanne Low, Cassie Soeffing, Peder Nelson, James Ervin, and Nikita Agrawal
Stem Enhancement in Earth Science (SEES) Summer High School Intern Program
United States of America
July 31, 2024

IVSS Badges

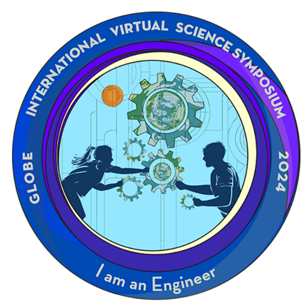
I Work With a STEM Professional:

This badge applies to us and our report as we collaborated with Peder Nelson, a Senior Instructor and Faculty Research Assistant at Oregon State University. We also worked with Andrew Clark, systems engineer and environmental scientist with IGES. Mr. Nelson and Mr. Clark helped refine our API models and aided us in integrating machine-learning technologies into our application.



I Am an Engineer:

This badge applies to us and our report because we used a large quantity of student-generated data to train a machine-learning model. This dataset was made of hundreds of GLOBE Observer photos in the four cardinal directions and downwards. We used YOLOv8 for semantic segmentation of the images and then ran this data through Google Teachable Machine to train a model that can identify levels of wildfire risk on the CS-FLARE application.



I Am a Data Scientist:

Our report includes citizen science imagery from the GLOBE Observer database and Satellite imagery from Google Earth Engine. We used this data to train a machine-learning model to evaluate wildfire risk by cross-referencing new imagery and satellite data. In our report, we discuss the limitations of our data and how it can be improved.