Mapping Agricultural Plastics in California with End-to-End CV Pipeline

Harvard John A. Paulson School of Engineering and Applied Sciences

April Chang, Amelia Li, Rebecca Qiu, Peter Wu, Evan Arnold

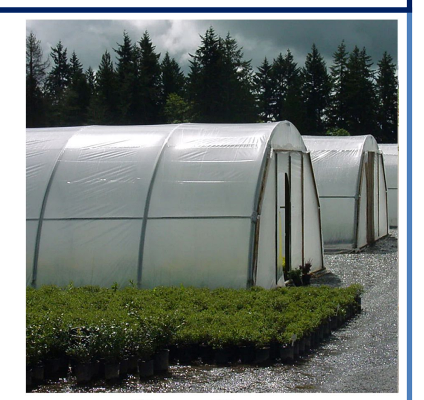
AC297r Capstone: The Nature Conservancy | **Project Partners**: Darcy Bradley, Kirk Klausmeyer | **Project Mentors**: Chase Van Amburg, Usha Bhalla

Background

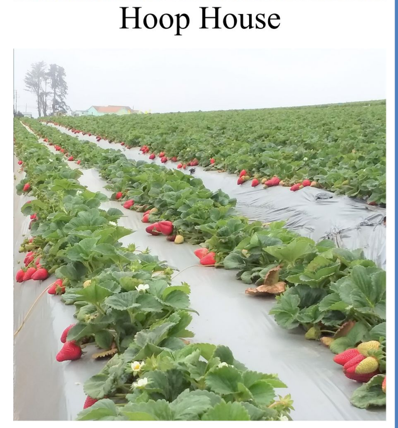
The problem we are addressing is the **rampant use of single-use plastics** such as mulch films and hoop houses in California's agricultural sector.

The problem is of importance because California's extensive agricultural activities have both localized and global repercussions. The state's unmonitored use of agricultural plastics contributes to environmental degradation, impacts human health, and raises social justice concerns. As a major player in the U.S. and worldwide agricultural markets, California's practices can set a precedent for other regions. The unchecked plastic usage is not only a direct threat to the state's ecosystems but also exacerbates global issues like climate change.

Moreover, the lack of data on plastic use limits our ability to understand its full impact, hindering effective policy-making and conservation efforts. Therefore, addressing this problem can have far-reaching implications for environmental sustainability, social equity, and global climate goals.



Hoop House



Plastic Mulch

Aims

- **Environmental Protection:** Monitor and map agricultural plastic use in California to address its environmental impact.
- **Mapping:** Identify and quantify agricultural plastics across California using satellite imagery and machine learning.
- **Policy Impact:** Establish a baseline for policy implications and assess community impacts of plastic pollution.

Methods

Data Collection and Processing

- **Labelled Data Acquisition**: Gathered labelled data from diverse Californian locations, including Oxnard, Santa Maria, Watsonville, and Mendocino counties.
- **Remote Sensing Data**: Utilized satellite imagery, primarily from SENTINEL-2, to provide a broad and detailed view of the agricultural landscapes.

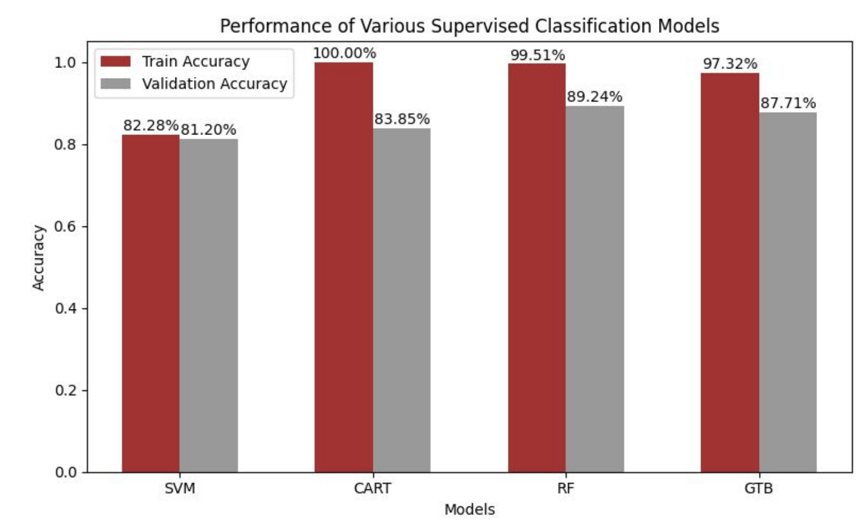
Model Training Implementation

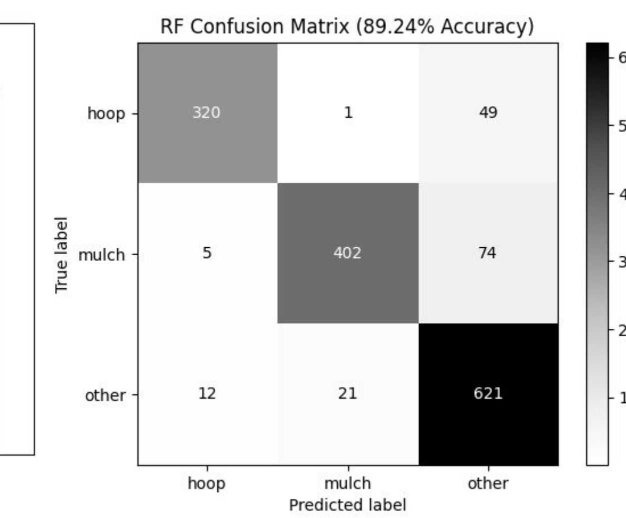
- **Model Training and Testing**: The collected data were used to train the SVM, CART models, and Random Forest via Google Earth Engine. Rigorous testing was conducted to evaluate model performance and optimize parameters.
- **Feature Engineering:** Experimented with different image bands and geographical features
- **Mapping and Monitoring**: Implemented the trained models to map the distribution and intensity of agricultural plastic use. Included temporal analysis to observe changes over time.
- **Validation and Refinement**: Continuously validated the model outputs with ground-truth data and refined the algorithms based on feedback and observed performance discrepancies.

Results

Initial Model Selection

The bar chart on the right displays the train and validation accuracies for four different machine learning models: Support Vector Machine (SVM), Classification and Regression Tree (CART), Random Forest (RF), and Gradient Tree Boosting (GTB). Notably, the Random Forest model outperforms the others in terms of both training and validation accuracies, making it our preferred choice for the task. Additionally, the Random Forest's decision trees can be serialized and retrieved with the Google Earth Engine API, offering a valuable advantage for model deployment and future use.



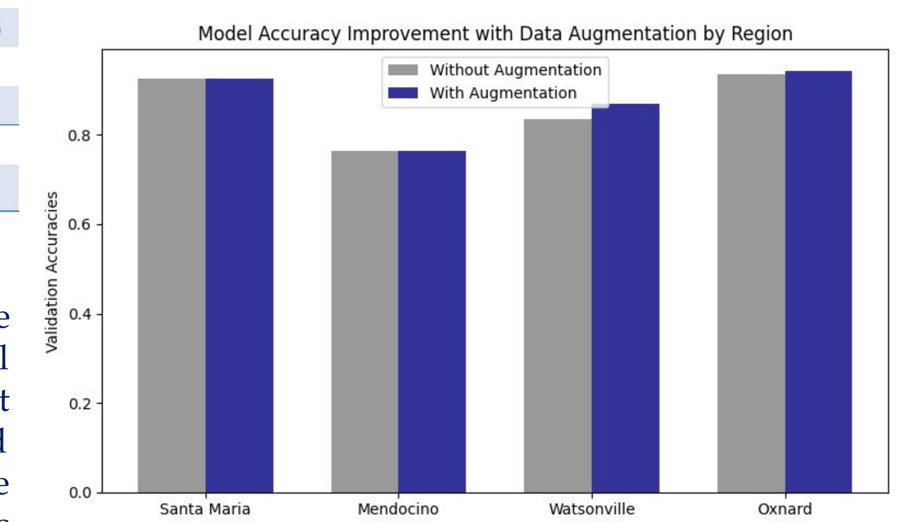


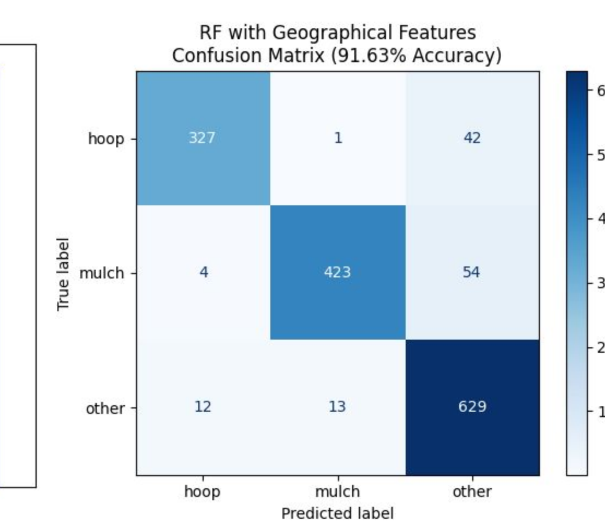
Feature Engineering for Agricultural Plastic Detection

Spectrum Features	B2	(Blue)
	B3	(Green)
	B4	(Red)
	B6	(Vegetation Red Edge-2)
	B8	(Near infrared, NIR)
	B11	(Shortwave Infrared-1, SWIR1)
	B12	(Shortwave Infrared-2, SWIR2)
Index Features	NDVI	(Normalized Difference Vegetation Index)
	NDTI	(Normalized Difference Tillage Index)
	PGI	(Plastic Greenhouse Index)
	PMLI	(Plastic-Mulched Landcover Index)
Additional Geographical Features	Distance to coastline	
	Elevation	

Feature Engineering

In addition to the Spectrum Features and Index Features encoding relevant attributes of satellite imageries, we experimented with data augmentation by calculating the distance to coastline and the elevation as additional geographical features. The rationale is grounded in the recognition that geographical and environmental factors play a pivotal role in shaping the distribution and utilization of agricultural plastics. Since our goal is to apply the model throughout agricultural lands in California, there can be a lot of variation in agricultural plastic usage due to diverse climate patterns and soil characteristics, which we try to capture with the added features.

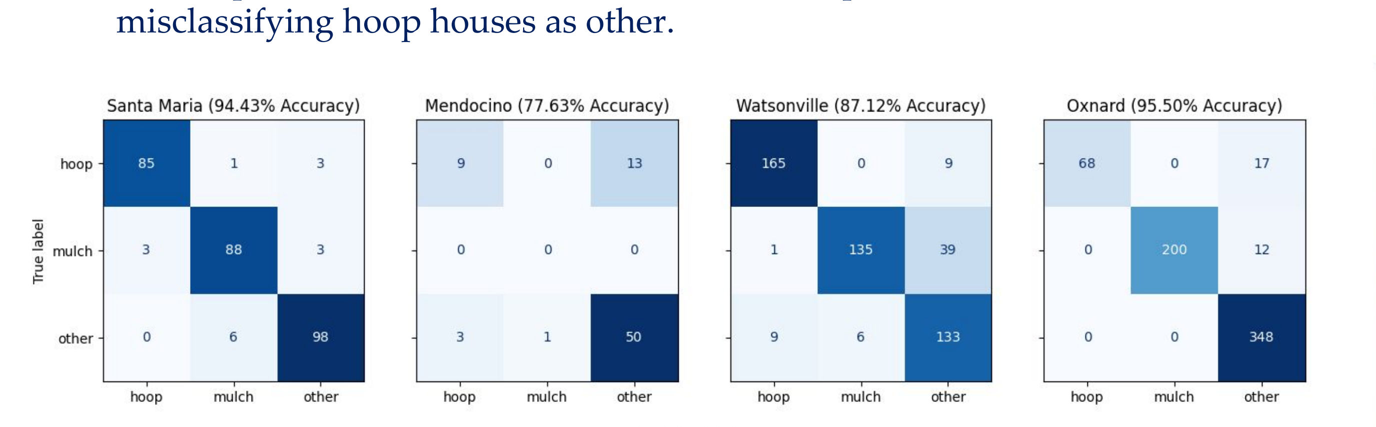


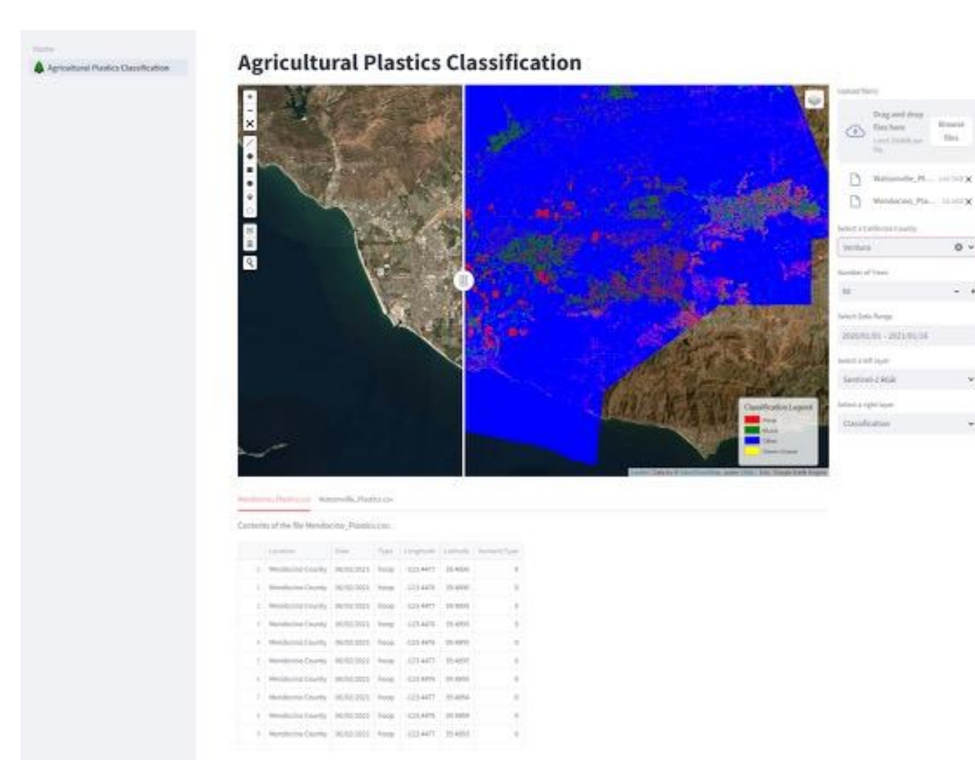


Model Performance

By incorporating the additional geographical features, the overall validation accuracy of our random forest model increases from 89.24% to 91.63%. The significant improvement with the county level accuracies in Watsonville and Oxnard also suggests a more comprehensive understanding of the regional nuances that influence plastic use and demonstrates improved accuracy in classifying agricultural plastics across diverse landscapes across California.

However, it's worth noting that in regions like Mendocino in a much different landscape of rural forested hills, the model's performance is less ideal, often misclassifying hoop houses as other.





Web Application Interface

Web Application for Visualization

We also leveraged Streamlit's rapid prototyping and interactivity capability to implement a user-friendly interface for model training and classification visualization. On the left is an image of the web application, which allows users to upload data files in csv format to train the model, then select the county that they want the model to be classified on. The content of the uploaded data is nicely presented in a table below the satellite image. The slider in the center of the map provides users with the ability to see the satellite image before and after the classification result.

Conclusion

- The implementation of a Random Forest model trained on satellite imagery augmented with geographical features has demonstrated success in the classification accuracy of agricultural plastics across different Californian regions. By integrating environmental factors such as distance to coastline and elevation into the model, we've acknowledged the impact of geographical diversity on agricultural practices and, consequently, plastic usage. The overall high accuracy of the model, particularly in regions like Santa Maria and Oxnard, underscores its effectiveness in understanding and adapting to regional nuances.
- The model excels in distinguishing between various types of plastics, with a high overall validation accuracy of 91.63%. This is indicative of the model's robustness and the potential for its application in real-world scenarios to aid in the management and reduction of agricultural plastic waste. However, areas for improvement have been identified, such as the misclassification trends observed in 'hoop' and 'other' categories in Mendocino county, suggesting a need for enriching the dataset with a broader spectrum of environmental conditions and agricultural practices. Other potential strategies could include further enhancing feature engineering or exploring more sophisticated model architectures.
- In addition, the web application that we developed facilitates the model's practical use by allowing users to upload data, train the model, and visualize classification results on a user-friendly interface, showcasing the project in an accessible and interactive manner.
- Using Google Earth Engine with machine learning, we developed a scalable, cost-effective, and technically robust solution for monitoring agricultural plastic use. Our intention is to contribute to the ongoing efforts in understanding this complex issue, which may help inform policy decisions and lay a foundation for future endeavors aimed at agricultural sustainability.

References

Xiong Y, Zhang Q, Chen X, Bao A, Zhang J, Wang Y. Large Scale Agricultural Plastic Mulch Detecting and Monitoring with Multi-Source Remote Sensing Data: A Case Study in Xinjiang, China. Remote Sensing. 2019; 11(18):2088. https://doi.org/10.3390/rs11182088

Aguilar MÁ, Jiménez-Lao R, Nemmaoui A, Aguilar FJ, Koc-San D, Tarantino E, Chourak M. Evaluation of the Consistency of Simultaneously Acquired Sentinel-2 and Landsat 8 Imagery on Plastic Covered Greenhouses. Remote Sensing. 2020; 12(12):2015. https://doi.org/10.3390/rs12122015

Sun H, Wang L, Lin R, Zhang Z, Zhang B. Mapping Plastic Greenhouses with Two-Temporal Sentinel-2 Images and 1D-CNN Deep Learning. Remote Sensing. 2021; 13(14):2820. https://doi.org/10.3390/rs13142820

Jiménez-Lao R, Aguilar FJ, Nemmaoui A, Aguilar MA. Remote Sensing of Agricultural Greenhouses and Plastic-Mulched Farmland: An Analysis of Worldwide Research. Remote Sensing. 2020; 12(16):2649. https://doi.org/10.3390/rs12162649

Acknowledgements

The Nature Conservancy

We extend our thanks to Darcy and Kirk for their expert guidance as project partners, and to our mentors, Chase and Usha, for their invaluable support. Our appreciation also goes to Yuanyuan for her advisory role and to TNC volunteer Brandee, whose data contribution was essential to our project's success.