



Explainable Machine Learning for the Prediction and Assessment of Complex Drought Impacts

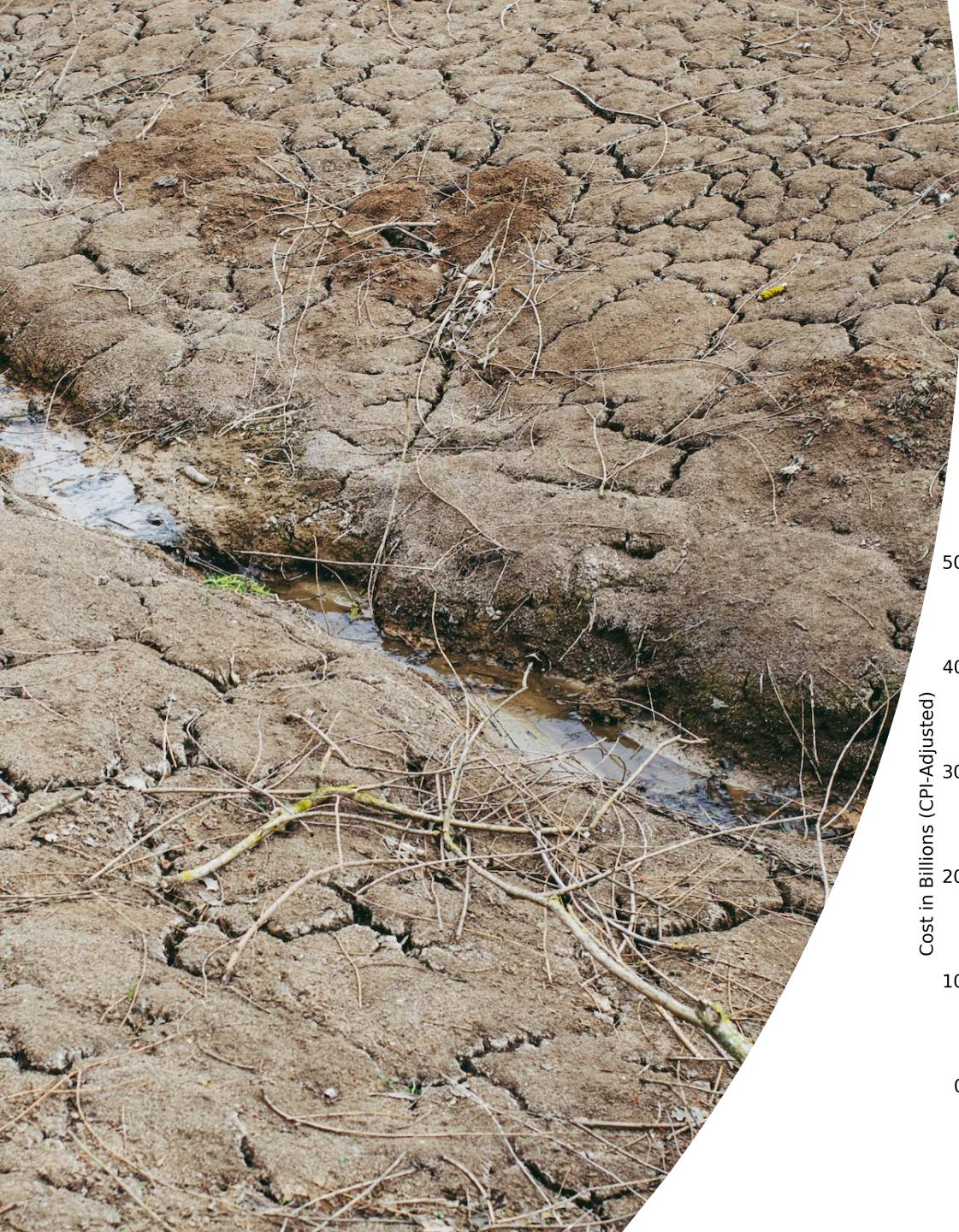


Beichen Zhang^{1,3}, Fatima Abu Salem², Michael Hayes¹, Kelly Helm Smith^{1,3} and Tsegaye Tadesse^{1,3}

¹ University of Nebraska-Lincoln, USA

² American University of Beirut, Lebanon

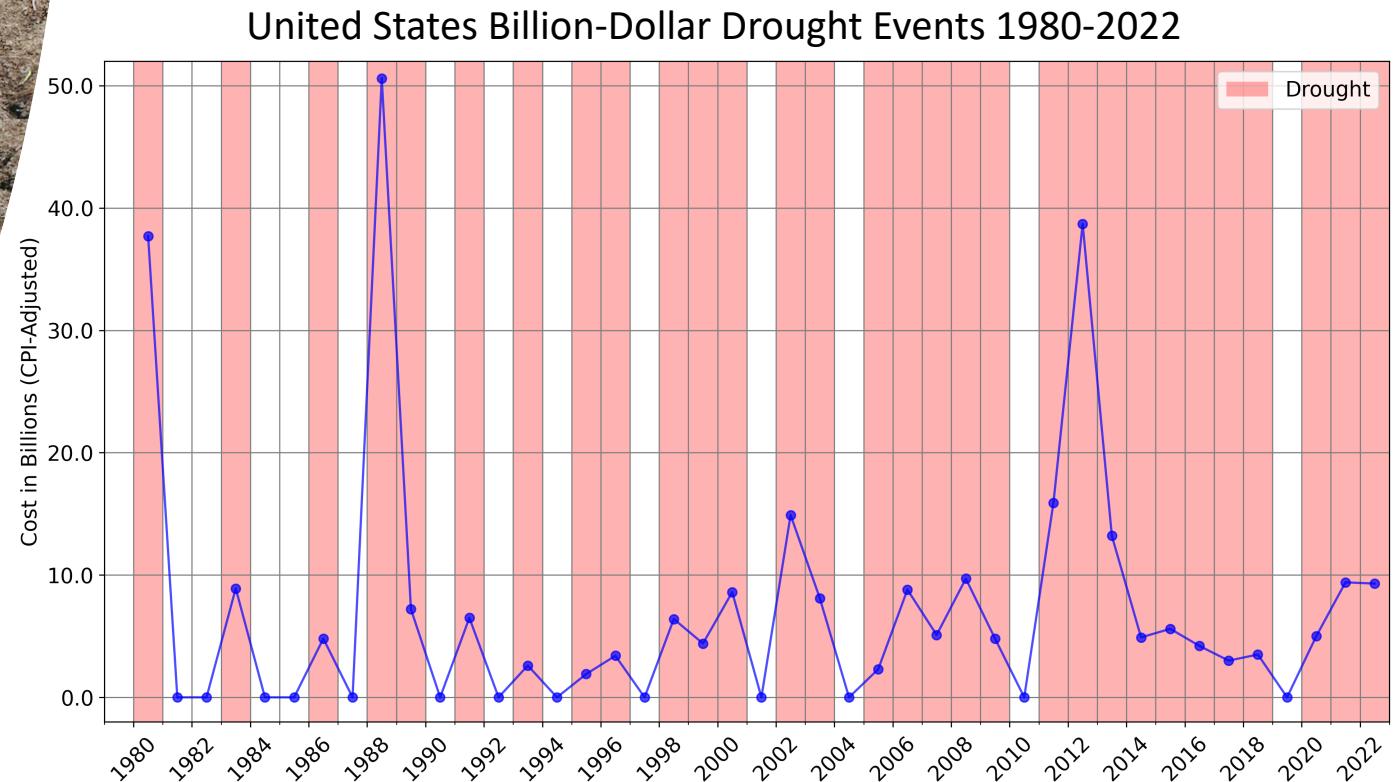
³ National Drought Mitigation Center, USA



Drought

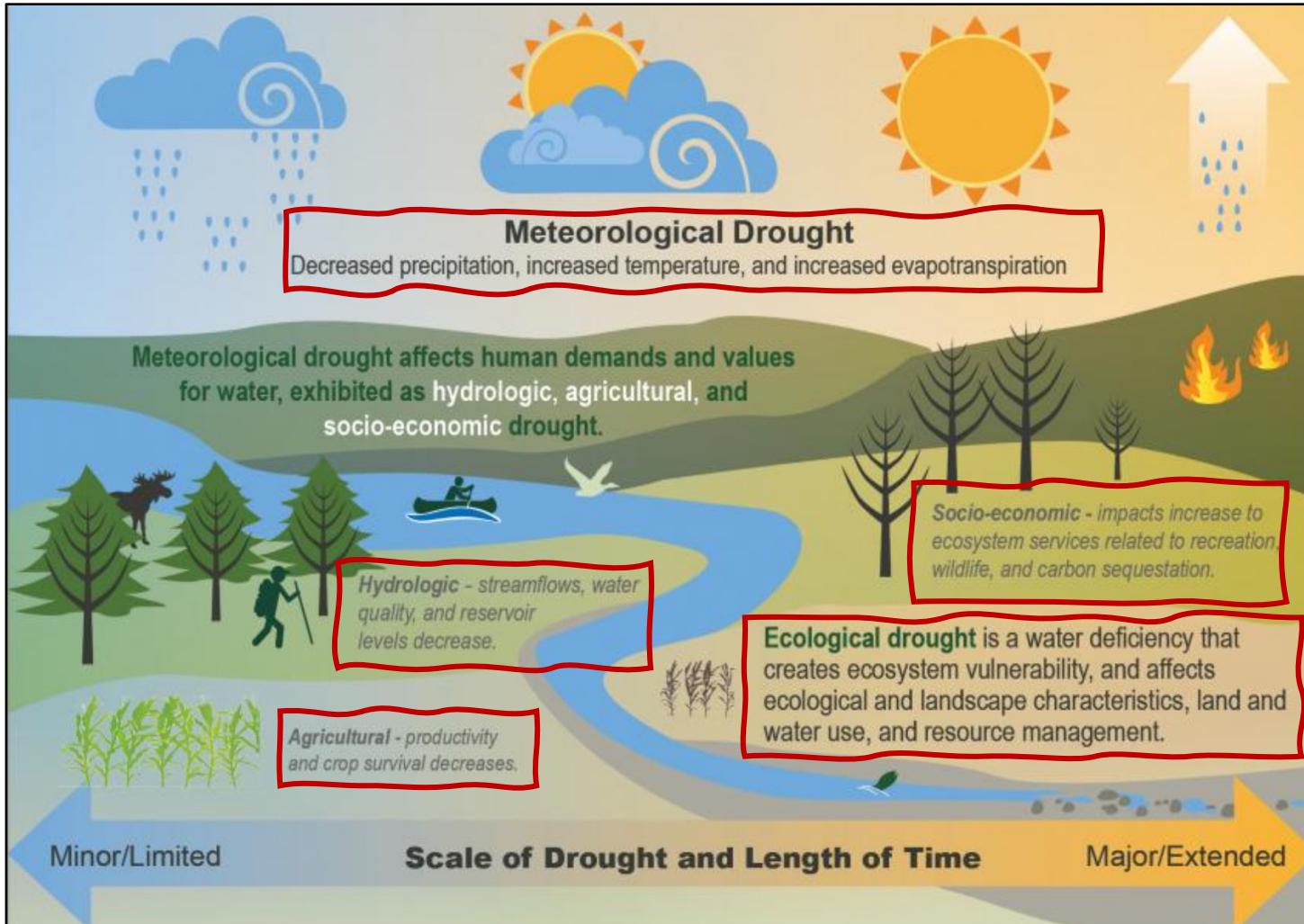
A lack of precipitation

A very common, but costly natural disaster



Data source: NCEI, <https://www.ncei.noaa.gov/access/billions/time-series/US>

Drought Characteristics



However, unlike other natural disasters, drought evolves slowly and is often not visible or tangible.

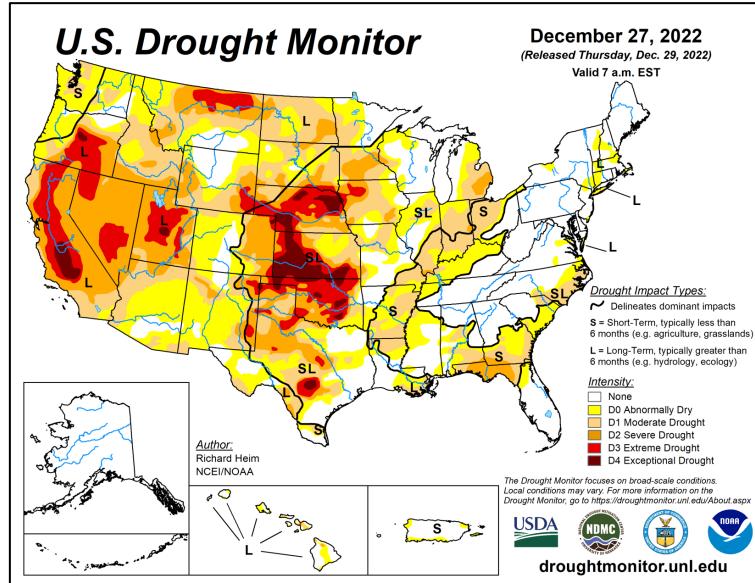
How to define/characterize drought events?

Onset
Frequency
Severity
Intensity
Duration
etc...

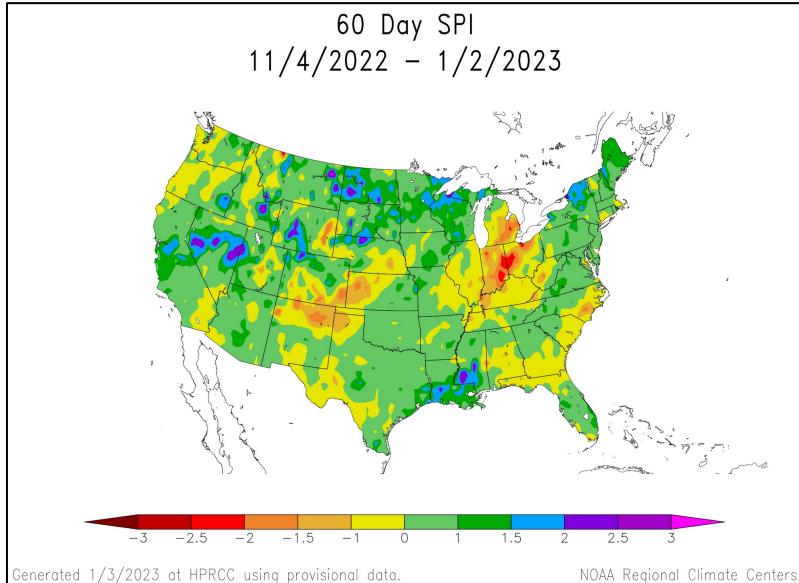
Drought Indicators/
Indices

Drought Monitoring and More

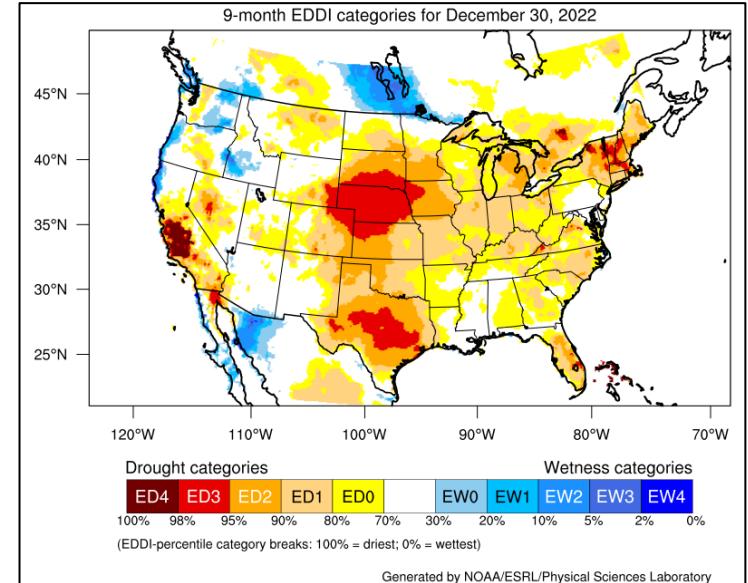
USDM



Standardized Precipitation Index



Evaporative Demand Drought Index



Over 100 drought indicators have been proposed or used.

Drought
Monitoring &
Early Warning



Vulnerability &
Risk Assessment

Policy &
Planning

Drought
Impacts?

Drought Impacts

How to identify the multi-dimensional drought impacts on the economy, society, and ecosystems?



Agriculture



Energy



Business & Industry



Tourism & Recreation



Wildfire



Plants & Wildlife



**Relief, Response &
Restrictions**



Society & Public Health



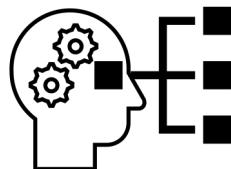
Water Supply & Quality

Why is it Challenging to Quantify Drought Impacts?

The **data scarcity** of various drought impacts on different societal sectors with good quality and high spatial and temporal resolutions.

The **complexity** and **non-linearity** of the relationships between various types of impacts and the existing drought indicators/indices.

How to Use ML to Help Address these Challenges?



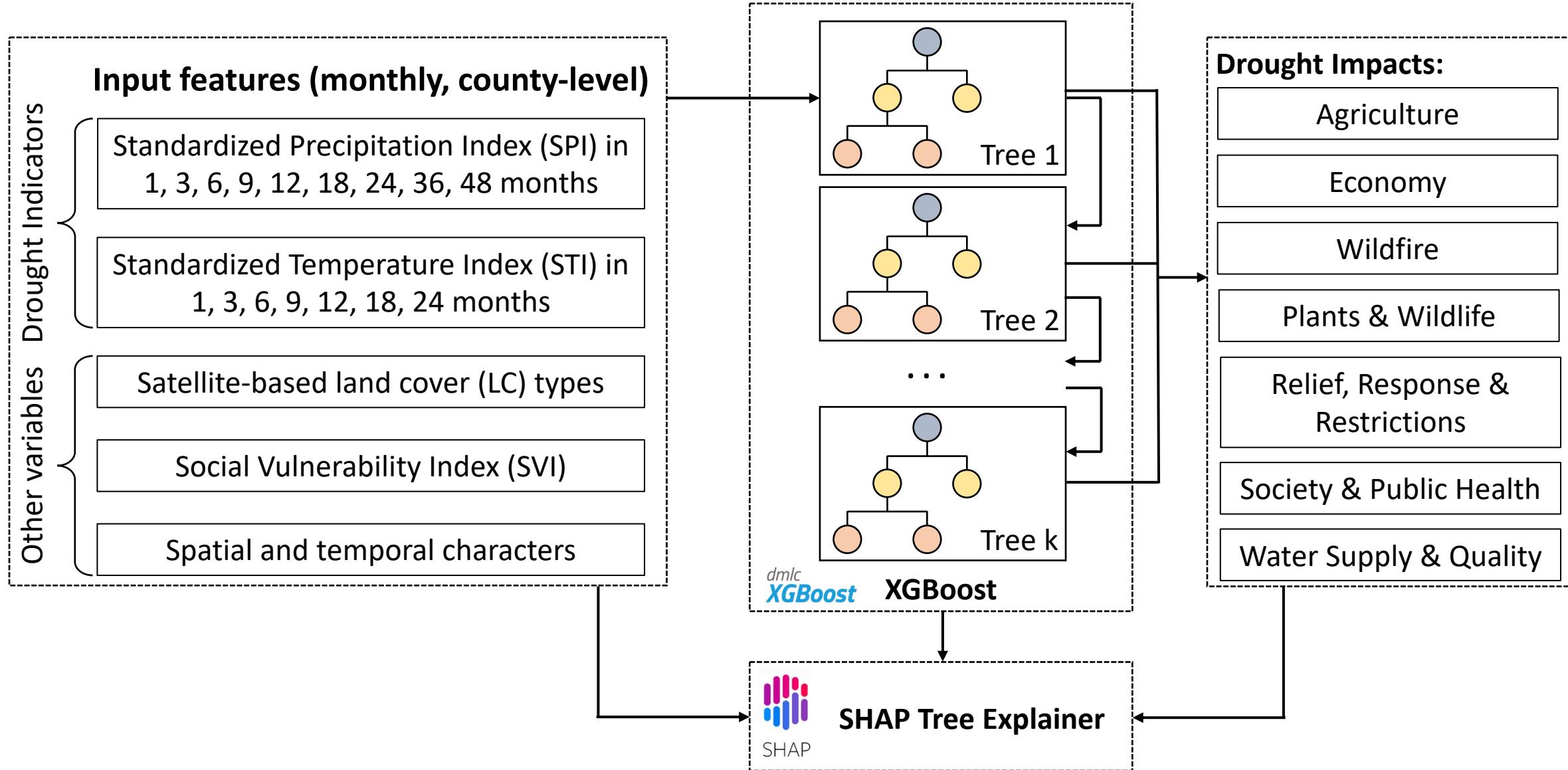
Goal of this research project:

Build a **robust and interpretable** machine learning (ML) pipeline to predict and analyze the complex drought impacts.

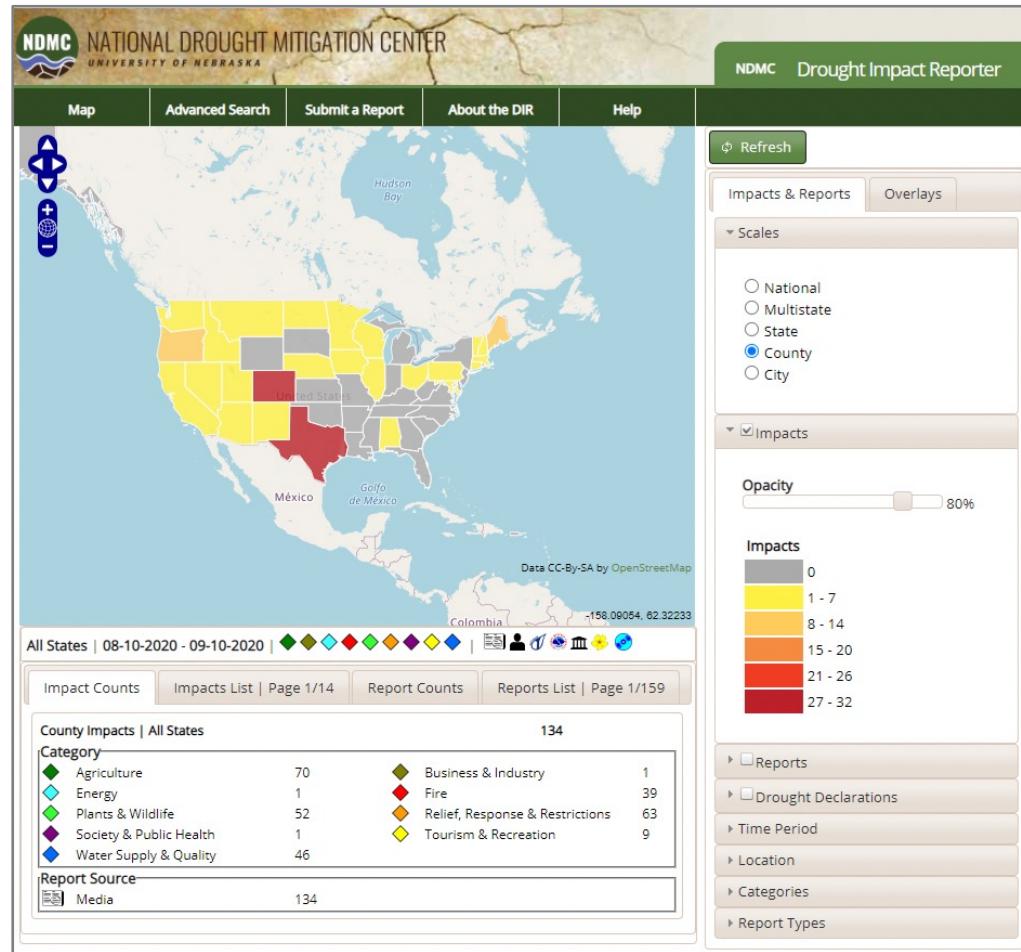
In real-world applications, the consequences of false negatives for drought impacts could be costly.



Explainable ML Pipeline: XGBoost + SHAP



Drought Impacts: Drought Impact Reporter (DIR)



- The first national comprehensive drought database.
- Multiple reporting resources.
- Nine categories of drought impacts.
- Flexible spatial resolution and available historical records (2002 – current).

Example of a record in California:

Title	Start Date	End Date	Categories	Places
Updated water allocation of 5% for California's State Water Project	12/1/20	12/31/21	Agriculture; Relief, Response & Restrictions; Water Supply & Quality	Los Angeles County, Napa County, Orange County, etc.

<https://droughtreporter.unl.edu/map/>

Case Study

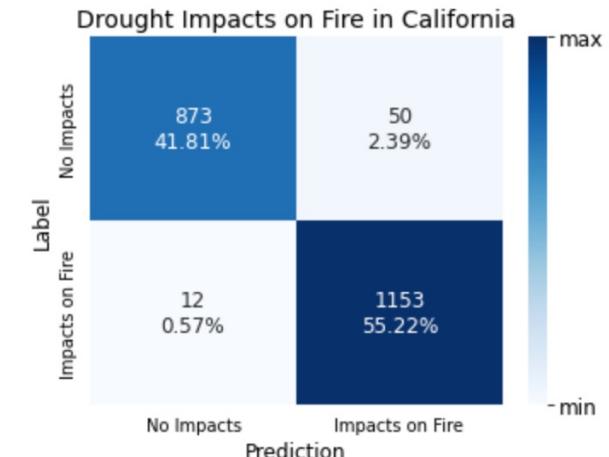
Drought Impacts Related to Wildfire in California

- The datasets are monthly at the county level from 2011 to 2020.
- The impacts are binarized into 0 and 1.
- The dimension of the input dataset: 6960 entries, 40 features.
- 70% of the data are used as the training and validation data, 30% are used as the test set.
- Nested cross-validation is used to fine-tune the model.

Performance metrics:

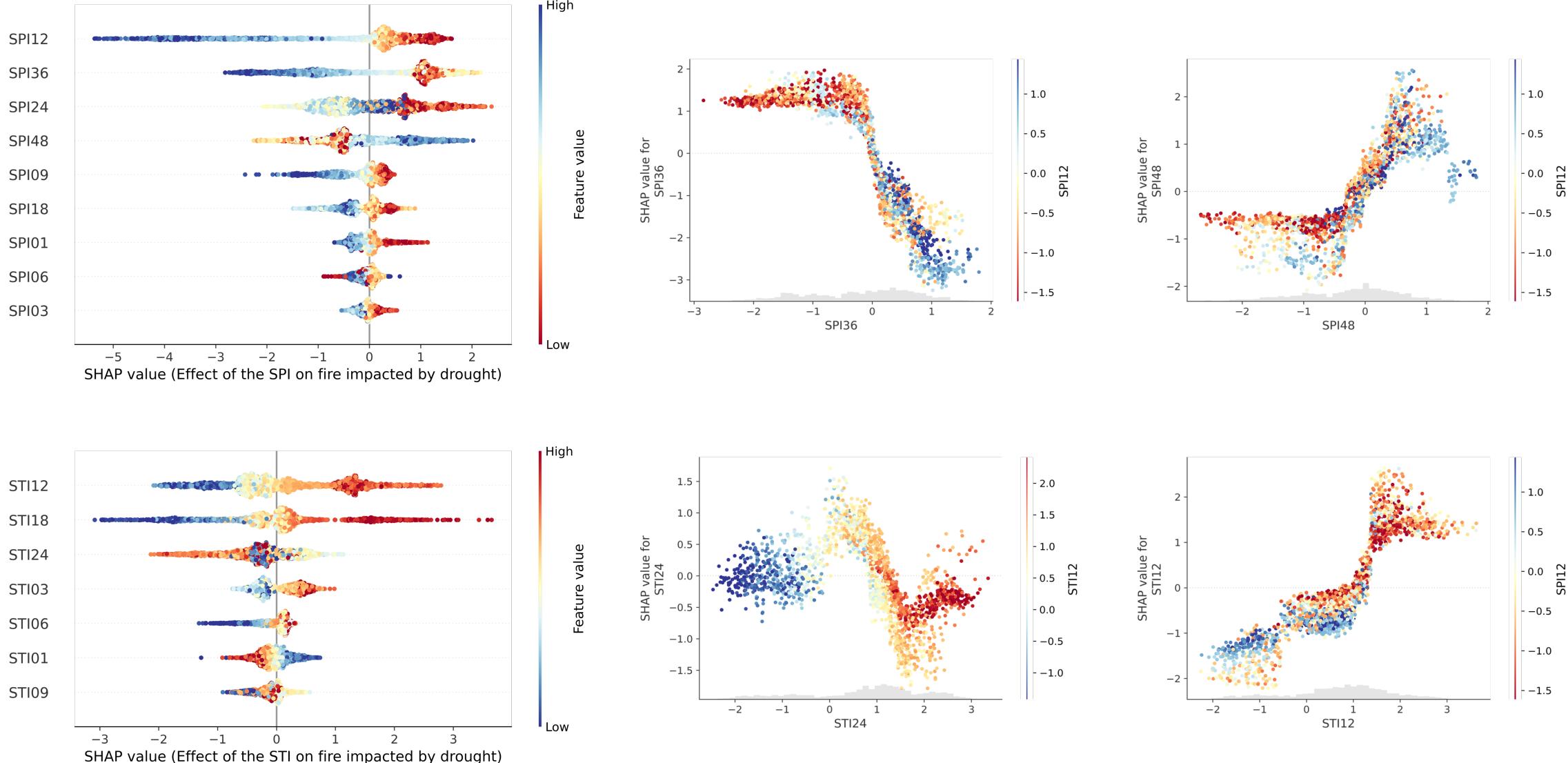
	Validation	Test
F2 score	0.97	0.98
Recall	0.99	0.99
Precision	0.91	0.96

Confusion matrix (test dataset):



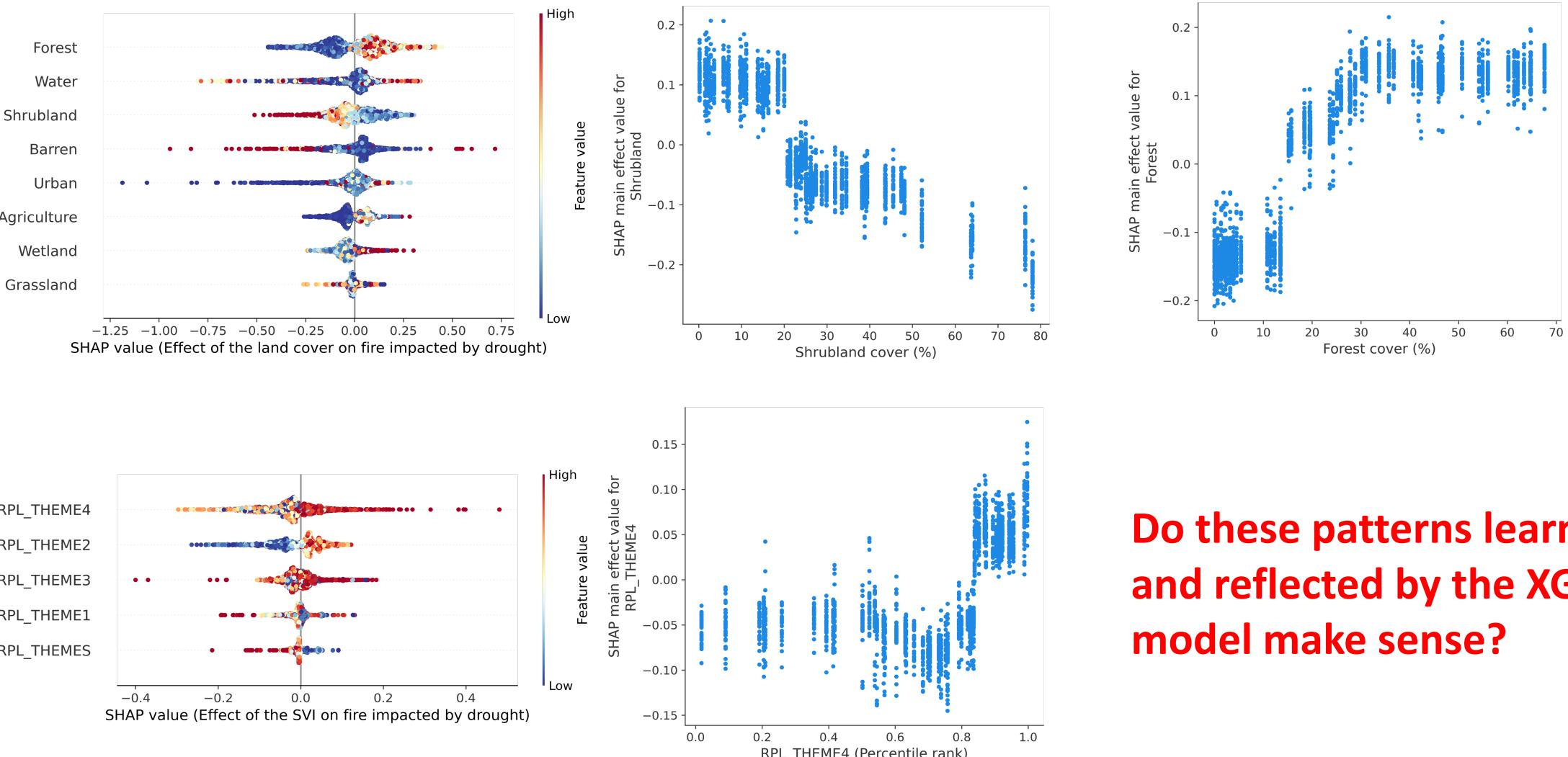
Case Study

Drought Impacts Related to Wildfire in California



Case Study

Drought Impacts Related to Wildfire in California



**Do these patterns learned
and reflected by the XGBoost
model make sense?**

Discussion

What does trustworthy AI mean to the predictive studies of the impacts of climate/natural disasters?

Climate scientists



I would like to use ML models to help answer complex questions.
Can I trust the model?

Computer scientists



I would like to use your domain expertise to validate the model explainability. Can you tell me if the interpretation of the variables is correct?

How to improve the interaction between the ML models/explainable tools (models) and the end users?

- XGBoost does achieve a very promising performance in predicting the occurrence of drought impacts in the case study.
- Overall, the model explainability based on the SHAP values is acceptable. However, further studies are suggested at the local level to better analyze the relationships among the features indicated by the model.

Takeaways

The preliminary work was submitted to NeurIPS 2020 workshop on Tackling Climate Change with Machine Learning

Paper on arXiv



The full work is under preparation and will soon be submitted to an academic journal.

Contact me if you have any further comments or questions or if you want to keep following the work.

**Contact me
Beichen Zhang:
beichen@huskers.unl.edu**



Thanks for Listening!

