# *An Interface for Data Curation and Mapping of Irrigated Areas Using Active Learning*

Kin Wai NG[1], Jack Marquez[1], Gabriel Laboy[1], Endalkachew Kebede[2], Holly Michael[2], Kyle Davis[2], and Michela Taufer[1]
[1]University of Tennessee, Knoxville, USA; [2]University of Delaware, Newark, USA

## Importance of Irrigation Mapping

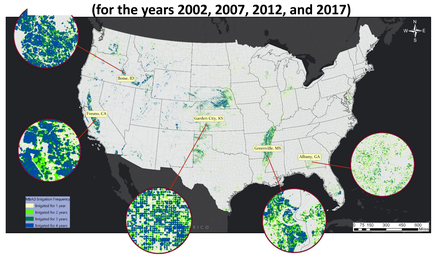
**Current Use:** Irrigation maps support water management decisions by helping scientists identify regions of high water demand and areas where water resources need careful monitoring

**Problem: Current irrigation mapping is limited**

- Large-scale datasets have lower resolutions
- Large-scale datasets are compiled over multi-year intervals

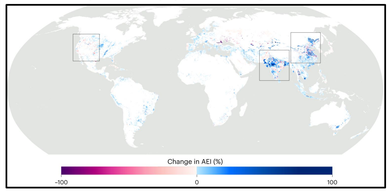
**Effect: Less accurate data for small scale and more difficult to analyze trends**

MIrAD map of irrigated agriculture for the United States [1] (for the years 2002, 2007, 2012, and 2017)



250m Resolution; 5-year intervals; Stops at 2017

Global map of area equipped with irrigation in 2015 [2]




5km Resolution; 5-year intervals; Stops at 2015

## Active Learning Workflow for Irrigation Mapping

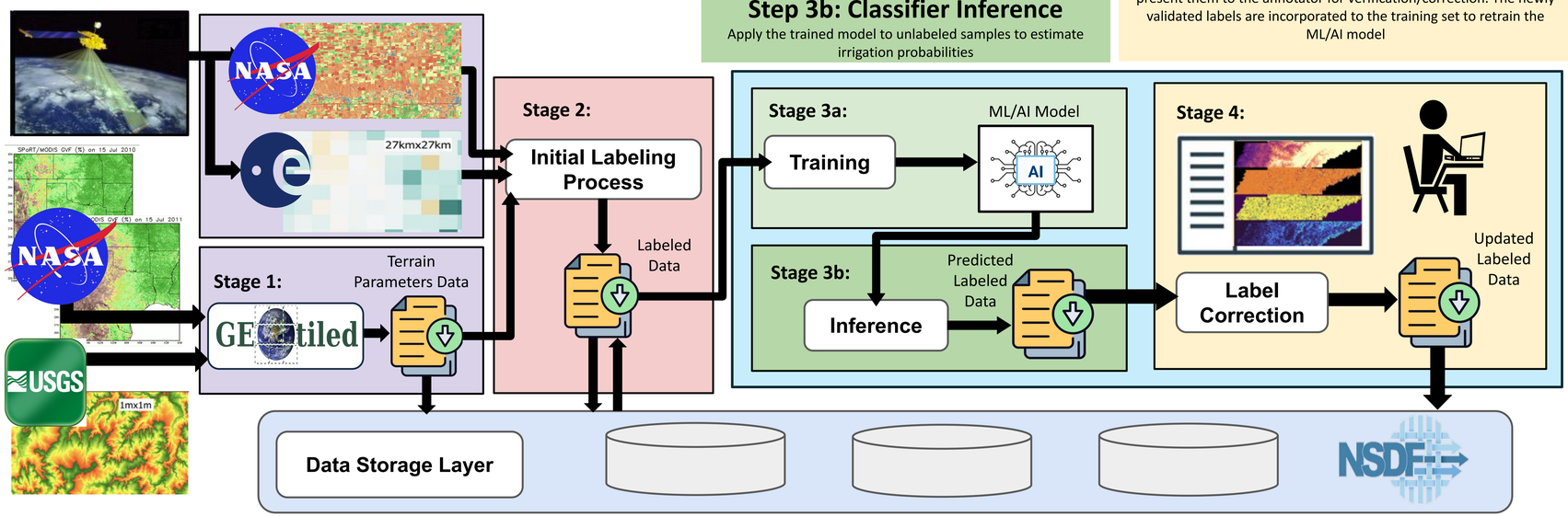
**Step 1: Data Preparation**
Download multi-source terrain, climate, and vegetation datasets (USGS, NASA) and preprocess them into AI-ready inputs

**Step 2: Initial Data Labeling**
Specify irrigation-related variables and threshold-based heuristics to produce an initial, proxy-labeled dataset

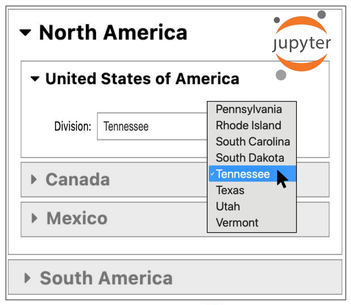
**Step 3a: AI/ML Model Training**
Train a model on the labeled dataset, using user-selected variables to predict irrigated versus non-irrigated locations

**Step 3b: Classifier Inference**
Apply the trained model to unlabeled samples to estimate irrigation probabilities

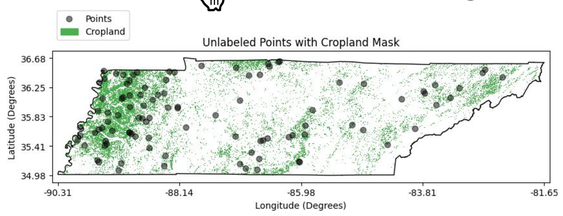
**Step 4: Query Uncertain Data Points and Update Labeled Set**
Extract data points with the highest prediction uncertainty and present them to the annotator for verification/correction. The newly validated labels are incorporated to the training set to retrain the ML/AI model




## Region Selection

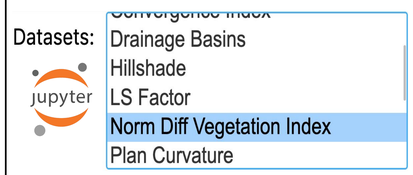



- Select region and year of interest
- Download cropland maps
- Sample data points for labeling

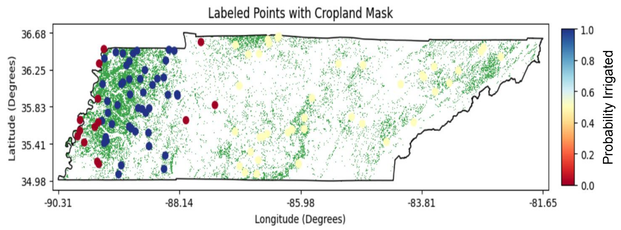



## Initial Labeling Process

- Select target variables from dropdown menu
- Set threshold values for each variable to determine the likelihood of irrigated land

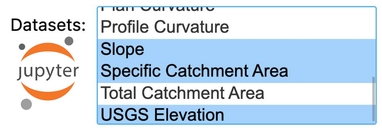



- Visualize resulting labels and adjust thresholds if needed

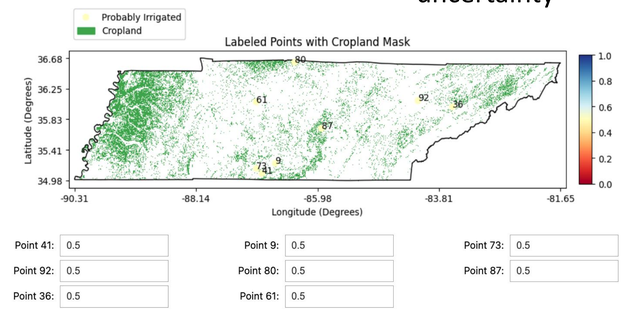



## Interactive Label Feedback

- Select input features for model training
- Specify number of active learning iterations




- Review and manually relabel points with high uncertainty




## Future Work

- Conduct broader evaluation across diverse regions to understand how well the interface improves labeling speed, ease of use, and productivity
- Compare active learning against traditional labeling to quantify how much effort is saved while maintaining/improving model accuracy
- Benchmark results against existing irrigation maps

## References

[1] Pervez MS, Brown JF. Mapping Irrigated Lands at 250-m Scale by Merging MODIS Data and National Agricultural Statistics. Remote Sensing. (2010). https://doi.org/10.3390/rs2102388
[2] Mehta, P., Siebert, S., Kummu, M. *et al.* Half of twenty-first century global irrigation expansion has been in water-stressed regions. *Nat Water* 2, 254–261 (2024). https://doi.org/10.1038/s44221-024-00206-9

THE UNIVERSITY OF TENNESSEE KNOXVILLE

***Acknowledgments:*** This work was supported by the National Science Foundation (NSF) under grant numbers #2331152, #2334945, #2223704, #2138811, #2103845 and #2449103.