# GEO-AI Challenge for Cropland Mapping by ITU

Author: Maryam Salehi, PhD in Remote Sensing

Introduction

The GEO-AI Challenge for Cropland Mapping by ITU represents a significant endeavor in the field of remote sensing and machine learning. Remote sensing allows us to harness the power of satellite imagery to monitor and analyze Earth's surface, while machine learning techniques enable us to extract valuable insights from this vast and complex data. In this report, we provide an exploration of the code and methodology employed to predict cropland areas using satellite imagery, demonstrating the synergy between these two fields.

Code Overview

The provided code predicts cropland areas using the following workflow:
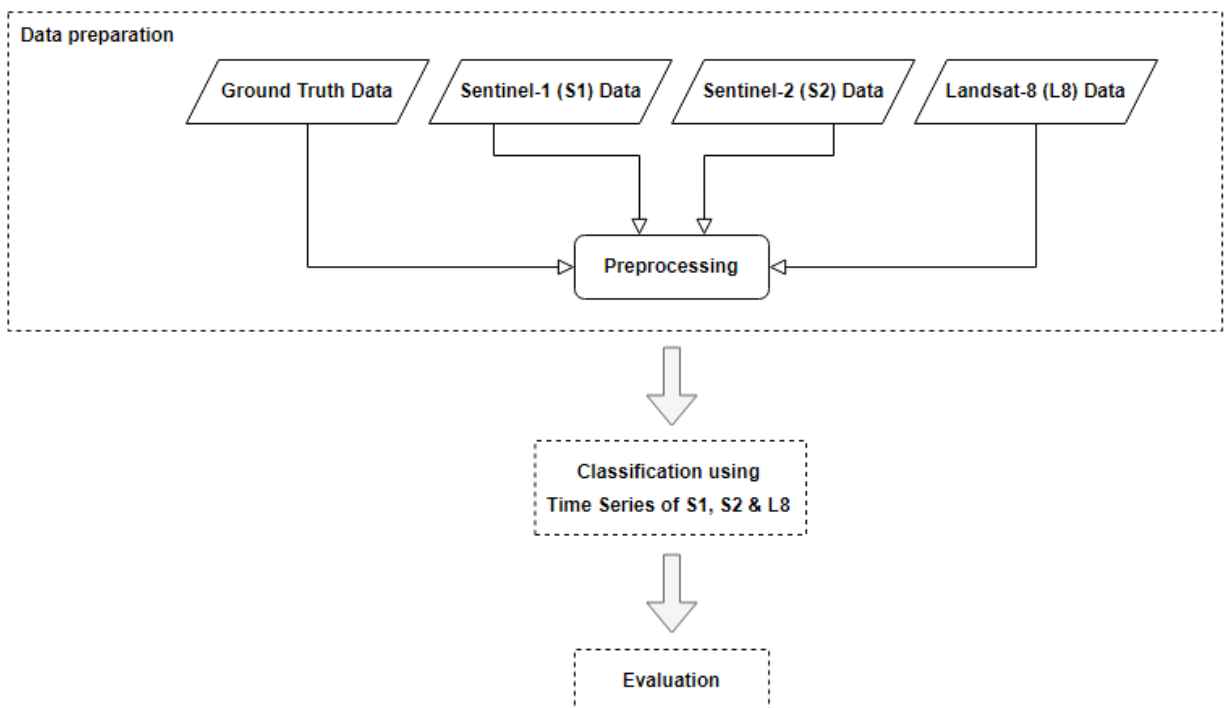
The code processes satellite data from multiple sensors (Sentinel-1, Sentinel-2, and Landsat-8), engineers features using a wide range of remote sensing indices, trains gradient boosting models, and ensembles predictions to generate a robust cropland classification model. The systematic data processing and modeling approach demonstrates an effective machine learning workflow.

The sources of information for remote sensing indices can be found at the following links:

https://www.indexdatabase.de/db/is.php?sensor_id=168

https://www.indexdatabase.de/db/is.php?sensor_id=96

The main steps in the code are:

1. Import packages and data

- Import packages like Pandas, NumPy, CatBoost, etc.

- Import training and test data for 3 countries - Iran, Sudan, Afghanistan. Data includes Sentinel-1, Sentinel-2, and Landsat-8 images.

2. Define functions

- vegetation_index: Calculates 181 spectral indices for a given month's data.

- vegetation_index_month: Calculates same indices for a single month's data.

3. Preprocess data

- For each country data, calculate spectral indices for each month.

- Merge indices with main bands to create feature dataset.

- Do same for test data.

4. Add Landsat-8 derived indices

- For each month, calculate indices like NDVI, EVI, etc from Landsat bands.

- Merge with Sentinel dataset.

5. Model and predict

- Split data into train-test sets.

- Fit the best model on training data.

   Note: Various models, such as XGBoost and AdaBoost, were trained separately for each country, and the CatBoost Classifier demonstrated the highest level of accuracy. The models were assessed based on accuracy and log loss.

- Predict on test data and calculate accuracy.

- Predict on final test data and generate submission file.

6. Ensemble

- Load prediction files for each country.

- Concatenate them to create final prediction file.

The model achieved a score of 0.9288 on the public leaderboard (#6 on the private leaderboard), demonstrating its strong performance in the competition.

Hardware resources: Google colab

Code Execution Instructions:

To execute the code, please follow these instructions:

1. Upload the 'Data_preparation.ipynb' notebook to Google Colab and run all cells to obtain the data files (train_data_iran_2019_2020_month_0to100.csv, …). These files contain processed data essential for subsequent modeling steps. Please note that these data files will be saved in the directory where the notebooks are located.

2. Upload the 'Cropland_mapping_Final.ipynb' notebook to Google Colab and run all cells to obtain the results from the best models.