GEO-AI CHALLENGE BY ITU
CROP MAPPING IN AFGHANISTAN, IRAN AND SUDAN

Submission for geoly

ABSTRACT

The 2023 GEO-AI Challenge for Crop Mapping seeks cost-effective and pracitical solutions to map cropland extents in Afghanistan, Iran and Sudan. We present our light-weight approach that relies on classical machine learning applied to Sentinel-2 data. Our approach is minimalistic; it relies on the provided ground truth and does not include additional satellite data sources.

Index Terms— crop mapping, machine learning, time series, satellite imagery.

1. INTRODUCTION

"Remote sensing is pivotal for agricultural monitoring, enabling the analysis of crop phenology and health through a series of spatial and temporal observations. This knowledge provides a global perspective on agricultural lands and contributes to food security efforts. With the wealth of satellite data available from both historical and ongoing missions, machine-learning techniques have emerged as a reliable solution for a variety of tasks, including crop mapping, yield estimation, crop condition assessment, and nutrient analysis.

At the regional/global level, remote sensing data has been used to produce crop maps such as GLAD [1], EU World Cover, and World Cereal [2]. These products are often outdated, and inconsistent and exhibit different nomenclature for describing crops. For instance, by analysing several high-resolution cropland layers, [3] deduced that nearly half of the uncertainty across the layers stems from landuse definitions. Another source of error is the lack of local data for validation. [4] confirms that low consensus between several crop maps analysed over Africa and points out that although researchers seek a one-for-all transferable solution, high intra-class variance can hamper their accuracy.

In crop mapping, several studies have employed different sensors (optical, RADAR or their combination) to distinguish crop and non-crop areas [5, 6]. By integrating time series data, crop phenological states are captured to facilitate their discrimination from other landuse. Domain-knowledge features such as spectral indices obtained by band difference and ratio operations are documented as informative sources of information which can be used to improve spectral bands or replace them entirely [7]. In terms of methodological advances, temporal convolution, recurrent neural networks and transformers techniques have been used [7, 5]. In some cases, the gain in accuracy is often a few units from classical baselines.

In this paper, we present our solution to the 2023 GEO-AI Challenge for Crop Mapping by International Telecommunication Union (ITU). The challenge solicits accurate and cost-effective classification models for mapping cropland extent in three regions (Afghanistan, Iran and Sudan) using remote sensing data. A critical requirement is the practicality of the solution thus we rely on light-weight models that can be deployed to open-access systems like Google Earth Engine.

The remainder of the paper is structured as follows:

- Section 2 describes the dataset provided and satellite data extraction process
- Section 3 details the data processing strategies and our experimental setup
- In Section 4, we evaluate the performance of our solution for the three different regions and compare our findings with additional feature engineering strategies.

2. DATASET

In this challenge, participants are given 500 samples (per region) of crop and non-cropland classes. Participants are allowed to use any freely accessible satellite data in the study regions and collect additional reference data for crop and non-crop if required. It is important to note that the reference data collected over Sudan and Iran is within one year (2019-2020). For Afghanistan, data was collected in April 2022.

2.1. Data downloading

Relying on Google Earth Engine[8], time series arrays of shape $X \in R^{T \times C}$ is retrieved for each point in our train and test set. T denotes the number of time stamps and C is the number of spectral channels. Data is collected for Sentinel-2 bands (only 10 and 20m) in the visible, near and shortwave infrared range. We manually remove certain dates in the collection because filtering using cloud-cover properties does not provide spatial insights on cloud-affected areas. It is worth noting that several images may be required to completely cover a region hence, spatial mosaics were performed

on footprints acquired on the same day or a day apart. Without this operation, points that fall in overlapped footprints will have additional time sequences. Our data downloading pipeline, feature extraction and model are available here .

Table 1. Data summary for each region

Region	Time window	Input shape
Afghan.	April 2022	$X \in R^{6\times10}$
Iran	July 2019 - June 2020	$X \in R^{55\times10}$
Sudan	July 2019 - June 2020	$X \in R^{51\times10}$

3. EXPERIMENTAL SETTING

For each region, separate models are created to predict cropland. We experiment with tree-based models (random forest and gradient boosting), k-nearest neighbour and support vector classifier (SVC). These models are applied to either raw spectral, spectral indices or their combination and tuned using optuna [9]; a hyper-parameter optimization framework which employs Bayesian optimization. Each experiment setup was run with 50 trials to find the best set of hyper-parameters.

3.1. Feature Engineering

Several spectral indices attenuating vegetation condition and biomass, water and impervious surfaces presence were derived from original spectral bands. The total number of spectral indices derived is 15.

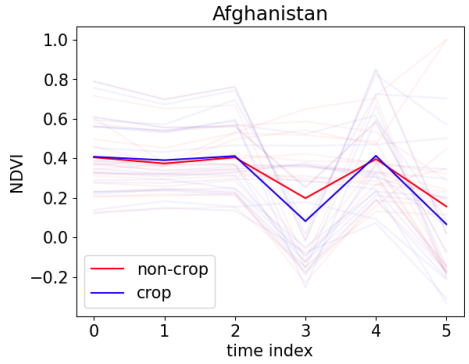

Fig. 1. NDVI profile for random samples of crop and non-crop areas in Afghanistan. Mean of samples shown with vivid colours

1. List of spectral indices :

(a) *Normalized Difference Vegetation Index (NDVI)*:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \tag{1}$$

(b) *NDVI Red-Edge (NDVI-RE)*:

$$\text{NDVI-RE} = \frac{\text{NIR} - \text{Red Edge}}{\text{NIR} + \text{Red Edge}} \tag{2}$$

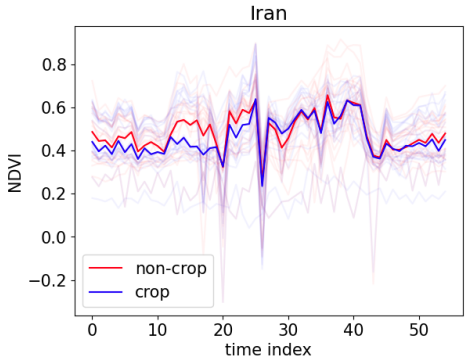

Fig. 2. NDVI profile for random samples of crop and non-crop areas in Iran. Mean of samples shown with vivid colours

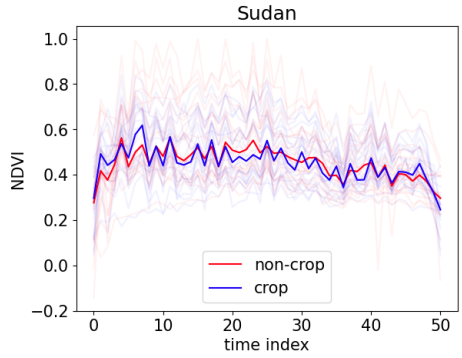

Fig. 3. NDVI profile for random samples of crop and non-crop areas in Sudan. Mean of samples shown with vivid colours

(c) *Normalized Difference Water Index (NDWI)*:

$$\text{NDWI} = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}} \tag{3}$$

(d) *Enhanced Vegetation Index (EVI)*:

$$\text{EVI} = 2.5 \times \frac{\text{NIR} - \text{Red}}{(\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue}) + 1} \tag{4}$$

(e) *Normalized Difference Built Index (NDBI)*:

$$\text{NDBI} = \frac{\text{SWIR} - \text{NIR}}{\text{SWIR} + \text{NIR}} \tag{5}$$

(f) *Soil Adjusted Vegetation Index (SAVI)*:

$$\text{SAVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + 0.5} \times (1 + 0.5) \tag{6}$$

(g) *Browning Reflectance Index (BRI)*:

$$\text{BRI} = \frac{1/\text{Green} - 1/\text{Red Edge}}{\text{NIR}} \tag{7}$$

(h) *Chlorophy Index (CI) - Red Edge*:

$$\text{CI} = \frac{\text{NIR}}{\text{Red Edge}} - 1 \tag{8}$$

(i) *Land Surface Water Index (LSWI)*:

$$\text{LSWI} = \frac{\text{NIR} - \text{SWIR}}{\text{NIR} + \text{SWIR}} \tag{9}$$

(j) *Normalized Difference Pond Index (NDPI)*:

$$\text{NDPI} = \frac{\text{SWIR} - \text{Green}}{\text{SWIR} + \text{Green}} \tag{10}$$

(k) *Water Ratio Index (WRI)*:

$$\text{WRI} = \frac{\text{Green} + \text{Red}}{\text{NIR} + \text{SWIR}} \tag{11}$$

(l) *Plant Senescence Reflectance Index (PSRI)*:

$$\text{PSRI} = \frac{\text{Red} - \text{Blue}}{\text{Red Edge}} \tag{12}$$

(m) *Difference vegetation Index (DVI)*:

$$\text{DVI} = \text{NIR} - \text{Red} \tag{13}$$

(n) *Ratio Vegetation Index (RVI)*:

$$\text{RVI} = \frac{\text{NIR}}{\text{Red}} \tag{14}$$

(o) *Visual Atmospheric Resistance Index (VARI)*:

$$\text{VARI} = \frac{\text{Green} - \text{Red}}{\text{Green} + \text{Red} - \text{Blue}} \tag{15}$$

3.2. Evaluation metric

The leaderboard submissions are based on accuracy. The same metric is used in our experiments due to the equal number of samples per class.

$$\text{Accuracy} = \frac{\text{number of correct predictions}}{\text{total number of predictions}} \tag{16}$$

3.3. Submission

For submission, the best hyper-parameters for any selected model in each region is used to re-initialize the model.

4. RESULTS AND DISCUSSION

In each region, we compare the performance of several classical models. For Afghanistan, RF on raw bands and KNN applied to indices achieved relatively higher performances. For RF, the difference compared to other models on raw bands was about 2-5%. Combining raw bands and indices (compared to raw bands only), reduced the performance of RF. This marginally reduces the accuracy for SVC compared to using indices and with no significant improvement to KNN. The synergy was however beneficial to XGBoost applied to either raw bands or indices.

Table 2. Accuracy comparison of raw and spectral indices for Afghanistan. Results show average of 5-fold cross validation

Model	Raw	Indices	Raw+Indices
RF	0.824	0.790	0.810
XGBoost	0.796	0.788	0.824
SVC	0.766	0.800	0.788
KNN	0.806	0.826	0.828

In Iran and Sudan, combining spectral bands and indices was generally not beneficial. Where the use of indices over raw bands resulted in a better prediction, the improvements were marginal (1-4% for tree-based methods) except for Sudan which increased by 20%.

Table 3. Accuracy comparison of raw and spectral indices for Iran. Results show average of 5-fold cross validation

Model	Raw	Indices	Raw+Indices
RF	0.962	0.962	0.960
XGBoost	0.942	0.958	0.958
SVC	0.954	0.964	0.962
KNN	0.948	0.938	0.950

Table 4. Accuracy comparison of raw and spectral indices for Sudan. Results show average of 5-fold cross validation

Model	Raw	Indices	Raw+Indices
RF	0.942	0.984	0.982
XGBoost	0.938	0.970	0.966
SVC	0.788	0.986	0.986
KNN	0.958	0.950	0.972

4.1. Improving accuracy for Afghanistan

Of the three regions, Afghanistan exhibits the lowest average accuracy. In the following section, we explore various strategies to improve this score.

4.1.1. Impact of increased time steps

The time window restriction for Afghanistan is very narrow. Only 6 timesteps were possible within the data collection period (April). Assuming crops were in season before and after this period, we advance the time window to a month before (March) and after (May) data collection.

Table 5. Short versus long time sequences comparison - Random Forest

Time	Raw	Indices	Raw+Indices
April only (6 steps)	0.824	0.790	0.810
March-May (18 steps)	0.844	0.838	0.826

As shown in Table 5, this extension resulted in 2% and 5% increase when considering RF applied to raw bands and indices respectively. In our final submission, we apply RF on indices for Sudan and Iran. For Afghanistan, RF on raw bands acquired from March-May is used.

4.1.2. Exploring deep learning models

Although our submissions relied on classical MLs, we experimented with time-guided deep learning techniques, consider-

ing that classical ML is not designed to harness temporal information. Temporal convolution networks (TempCNN) and LSTMs are explored using implementations from [5]. Similar to the classical ML models, each model is tuned with Optuna under the same cross-validation configuration. We use multilayer perceptron (MLP) as a comparable baseline to RF since it does not regard temporal order. As seen in Table 6, all deep learning models are not beneficial on raw bands. Temporal convolutions and LSTMs can however improve the RF baseline on indices by nearly 7%.

Table 6. Comparing classical and deep learning models. Here we strictly adhere to the time window for Afghanistan (April)

Model	RF	MLP	TempCNN	LSTM
Raw	0.824	0.784	0.796	0.808
Indices	0.790	0.774	0.862	0.858

5. CONCLUSION

We address the 2023 GEO-ITU Crop Mapping challenge using lightweight machine learning techniques applied to spectral bands and spectral indices for different regions. Our findings emphasize the importance of long temporal information, achieving over 96% accuracy for Sudan and Iran. Combining spectral bands and indices showed mixed results, sometimes improving or worsening performance. In the case of Afghanistan, deep learning techniques were beneficial only when considering spectral indices. In summary, we have provided insights to the task by experimenting with several techniques and feature sets. Our findings infer that, with limited labels, and only multi-spectral information or indices, reasonable accuracy can still be achieved hence reducing data collection efforts.

6. RECOMMENDATIONS

Domain knowledge, such as cropping seasons in different regions, can streamline the data collection process. For example, understanding specific planting and harvesting periods is crucial. In Sudan and Iran, ground data collection covered a one-year period, suggesting that planting and growing occurred within this time. In Afghanistan, where labels were collected in April 2022, it's unclear whether this covers the entire planting-to-harvesting period or if cultivation began in the previous year.

7. REFERENCES

[1] Peter Potapov, Svetlana Turubanova, Matthew Hansen, Alexandra Tyukavina, Viviana Zalles, Ahmad Khan, Xiao-Peng Song, Amy Pickens, Quan Shen, and Jocelyn Cortez, "Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century," *Nature Food*, vol. 3, 12 2021.

[2] K. Van Tricht, J. Degerickx, S. Gilliams, D. Zanaga, M. Battude, A. Grosu, J. Brombacher, M. Lesiv, J. C. L. Bayas, S. Karanam, S. Fritz, I. Becker-Reshef, B. Franch, B. Mollà-Bononad, H. Boogaard, A. K. Pratihast, and Z. Szantoi, "Worldcereal: a dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping," *Earth System Science Data Discussions*, vol. 2023, pp. 1–36, 2023.

[3] Francesco Tubiello, Giulia Conchedda, Leon Casse, Pengyu Hao, Zhongxin Chen, Steffen Fritz, Douglas Muchoney, and Giorgia DeSantis, "Measuring the world's cropland area," 07 2022.

[4] Hannah Kerner, Catherine Nakalembe, Adam Yang, Ivan Zvonkov, Ryan McWeeny, Gabriel Tseng, and Inbal Becker-Reshef, "How accurate are existing land cover maps for agriculture in sub-saharan africa?," 2023.

[5] Marc Rußwurm, Charlotte Pelletier, Maximilian Zollner, Sébastien Lefèvre, and Marco Körner, "Breizhcrops: A time series dataset for crop type mapping," 2020.

[6] Stella Ofori-Ampofo, Charlotte Pelletier, and Stefan Lang, "Crop type mapping from optical and radar time series using attention-based deep learning," *Remote Sensing*, vol. 13, pp. 4668, 11 2021.

[7] Charlotte Pelletier, Geoffrey I. Webb, and François Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sensing*, vol. 11, no. 5, 2019.

[8] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, pp. 18–27, 2017, Big Remotely Sensed Data: tools, applications and experiences.

[9] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama, "Optuna: A next-generation hyperparameter optimization framework," 2019.