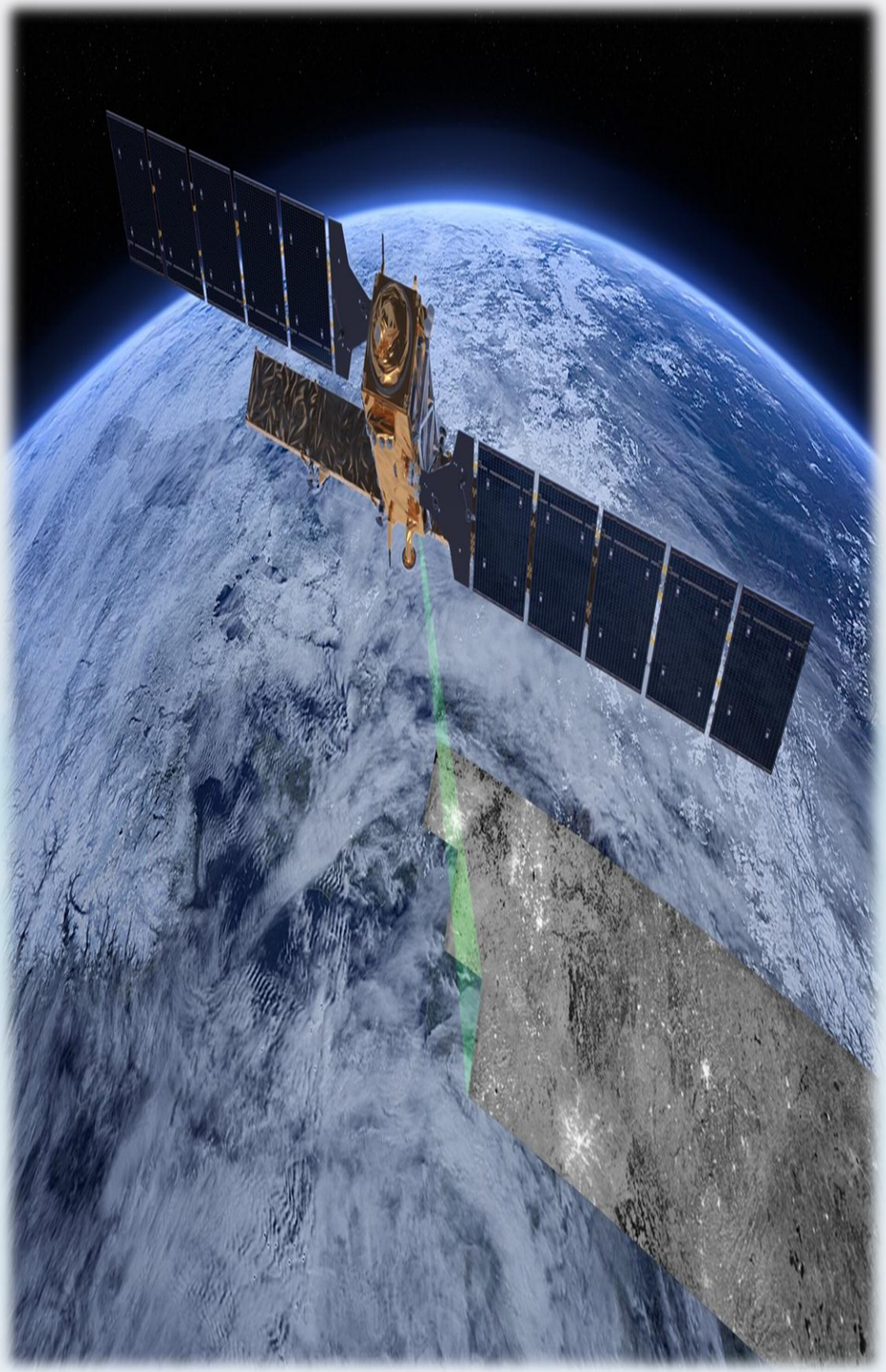
CROPLAND MAPPING Using Multi-Sensor Time Series Data and Gradient Boosting Models

Maryam Salehi, PhD in Remote Sensing



Introduction

Cropland mapping and monitoring play crucial roles in effective agricultural management, land-use planning, and ensuring food security. Earth observation sensors like Sentinel-1, Sentinel-2, and Landsat-8 provide valuable data, but each has limitations. In the GEO-AI Challenge by ITU, we extract bands and spectral indices from the time series data of these sensors to enhance cropland mapping accuracy.

Introduction

While previous research papers on remote sensing and cropland mapping have focused on models such as Random Forest, Support Vector Machine, k-Nearest Neighbor, Naive Bayes, and Neural Networks, our study aims to fill the gap by exploring the potential efficacy of gradient boosting models like XGBoost or CATBoost in the context of cropland mapping.

A Lack of Gradient Boosting Models

Wu et al. (2023)

RF, SVM, KNN, NB - Pixel/Object-Based Classification of Semi-Arid Grassland in Northern China.

Trivedi et al. (2023)

RF - Cropland Mapping in Tropical Smallholder Systems.

Yi et al. (2020)

RF - Crop Classification Using Multi-Temporal Sentinel-2 Data in China.

Teluguntla et al. (2018)

RF - Cropland Extent Product of Australia and China using RF on Google Earth Engine.

Belgiu and Csillik (2017)

RF - Sentinel-2 Cropland Mapping using Pixel-Based and Object-Based Time-Weighted Dynamic Time Warping Analysis.

Li et al. (2022)

RF, SVM - Cropland Map from Satellite Images Time Series in Indian.

Bofana et al. (2020)

RF, SVM - Cropland Classification Methods in the Zambezi River Basin.

Shelestov et al. (2017)

RF - Classification of Multi-Temporal Imagery for Crop Mapping.

Jia et al. (2019)

RF, SVM, Neural Networks - Automated Monitoring Cropland Using Remote Sensing Data.

Lingwal et al. (2023)

RF, SVM, Neural Networks - Landcover Semantic Segmentation for Cropland Mapping.

Latif et al. (2023)

RF, SVM - Mapping Cropland Extent in Pakistan.

Despite the prevalence of Random Forest, Support Vector Machine, and Neural Network models, the absence of gradient boosting models prompts exploration into their potential efficacy in remote sensing applications.

Study Area

The study encompasses three regions:

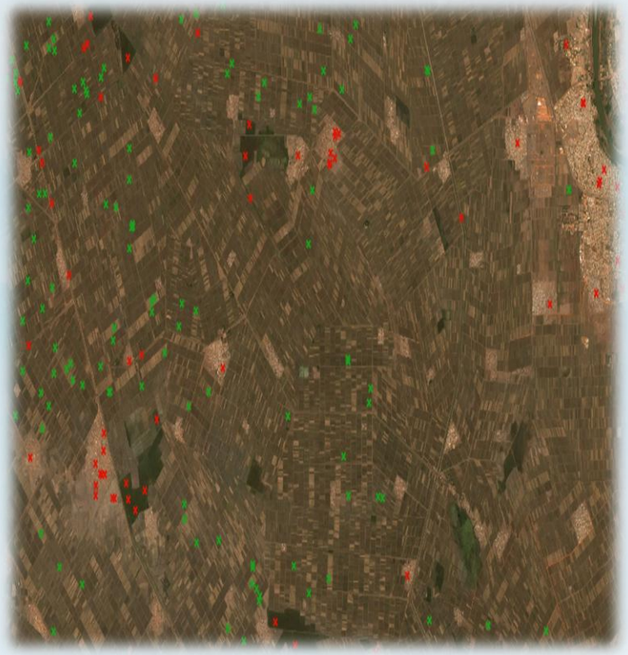
Sudan

Afghanistan

Iran

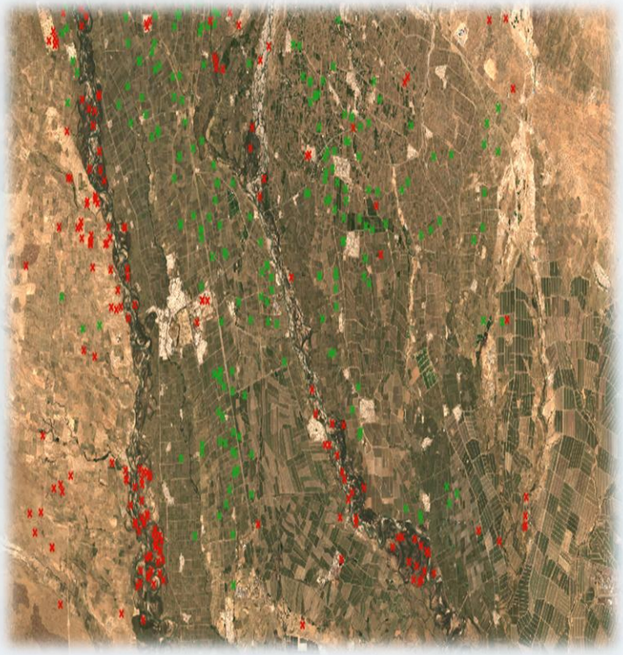
Study Area

Ground truth samples were collected during specific time frames, from July 2019 to June 2020 for Sudan and Iran, and in April 2022 for Afghanistan.



Sudan

Afghanistan



Iran

Methodology

1

Data Preprocessing

Acquiring and preprocessing data from Sentinel-1, Sentinel-2, and Landsat-8, including cloud cover removal of Sentinel-2 SR (surface reflectance) HARMONIZED and Landsat-8 SR images, selecting VH and VV bands, and calculating the mean values of VH for both ascending and descending orbits, as well as the mean value of VV for the combined ascending and descending orbits of Sentinel-1 GRD images.

```
s2_collection = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED')
bands = ['B1', 'B2', 'B3', 'B4','B5', 'B6', 'B7', 'B8','B8A', 'B9', 'B11', 'B12']
```

```
s1_collection = ee.ImageCollection('COPERNICUS/S1_GRD')
bands = ['vhAsc', 'vhDesc', 'VVAscDesc']
```

```
l8_collection = ee.ImageCollection('LANDSAT/LC08/C01/T1_SR')
bands = ['B1', 'B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B10', 'B11']
```

Methodology

2

Spectral Index Extraction

Deriving various spectral indices such as NDVI, EVI, and SAVI from the optical datasets to capture different aspects of vegetation growth and health.

```
MI = (b8a - b11) / (b8a + b11+ eps)
ARVI_ =  (b08 - (2 * b04) + b02) / (b08 + (2 * b04) - b02+ eps)
SIPI = (b08 - b02) / (b08 - b04+ eps)
RENDVI = (b06 - b05) / (b06 + b05+ eps)
MRESR = (b06 - b01) / (b05 - b01+ eps)
RYI = b03 / (b02+ eps)
NDYI = (b03 - b02) / (b03 + b02+ eps)
DYI = b03 - b02
ACI = b08 * (b04 + b03)
AVI_ = (b08 * (1 - b04) * (b08 - b04))
SI = ((1 - b02) * (1 - b03) * (1 - b04))
BSI = ((b11 + b04) - (b08 + b02)) / ((b11 + b04) + (b08 + b02)+ eps)
L = 0.33; SAVI = ((b08 - b04) / (b08 + b04 + L)) * (1 + L)
FIDET = b12 / (b8a * b09+ eps)
MTCI_ = (b06 - b05) / (b05 - b04+ eps)
NPCRI = (b04 - b02) / (b04 + b02+ eps)
S2REP = 705 + 35 * ((((b07 + b04) / 2) - b05) / (b06 - b05+ eps))
CCCI = ((b08 - b05) / (b08 + b05+ eps)) / ((b08 - b04+ eps) / (b08 + b04+ eps))
MCARI = ((b05 - b04) - 0.2 * (b05 - b03)) * (b05 / (b04+ eps))
TCARI = 3 * ((b05 - b04) - 0.2 * (b05 - b03) * (b05 / (b04+ eps)))
PVI = (b08 - 0.3 * b04 - 0.5) / ((1 + 0.3 * 2) ** (1 / 2.0))
EVI = 2.5 * (b08 - b04) / (b08 + 6 * b04 - 7.5 * b02 + 1)
EVI2 = 2.4 * (b08 - b04) / (b08 + b04 + 1)
```

```
  NDVI = ((b08 - b04) / (b08 + b04+ eps))
  NDVI2 = ((b12 - b08) / (b12 + b08+ eps))
  MNDVI = (b08 - b04) / (b08 + b04 - 2 * b02+ eps)
  BAI = 1 / ((0.1 - b04) ** 2 + (0.06 - b08) ** 2)
  NDSI = (b03 - b11) / (b03 + b11+ eps)
  MRENDVI = (b06 - b05) / (b06 + b05 - 2 * b01+ eps)
  NDVIre = (b08 - b05) / (b08 + b05+ eps)
  CIre = ((b08 / (b05+ eps)) - 1)
  NDMI = (b08 - b11) / (b08 + b11+ eps)
  TNDVI = [(x) ** (1 / 2.0) for x in ((b08 - b04) / (b08 + b04) + 0.5)]
  SWIRmin = 0.378; SWIRmax = 0.397; SWIIRmin = 0.027; NDVIc = (b08 - b04) / (b08 + b04) * (1.0 - (b12 - SWIIRmin) /
(SWIRmax - SWIRmin))
  VDVI = (2 * b03 - b04 - b02) / (2 * b03 + b04 + b02+ eps)
  TVI = (120 * (b06 - b03) - 200 * (b04 - b03)) / 2
  EXG = 2 * b03 - b04 - b02
  PSRI = (b04 - b02) / (b06+ eps)
  RDVI = [(i - j) / (i + j) ** (1 / 2.0) for i, j in zip(b08, b04)]
  RATIO1 = b01 / (b03+ eps)
  RATIO2 = b01 / (b05+ eps)
  RATIO3 = b11 / (b12+ eps)
  RATIO4 = b05 / (b04+ eps)
  NDWI = (b03 - b08) / (b03 + b08+ eps)
  NDRE = (b09 - b05) / (b09 + b05+ eps)
  NDRE_ = (b05 - b04) / (b05 + b04+ eps)
  NDRE7 = (b07 - b04) / (b07 + b04+ eps)
  MSI = b11 / (b8a+ eps)
  RECI = (b08 / (b04+ eps)) - 1
```

Methodology

3

Applying Gradient Boosting Models

Training models using multi-temporal bands and indices as input features. Notable models used were XGBoost, AdaBoost, and the highly accurate CatBoost Classifier.

Results

The methodology was applied to the three study regions. Using **Sentinel-1 (S1)** data alone provided decent accuracy for cropland classification, with accuracy scores ranging from 83-95% across the three study regions. However, S1 had higher log loss scores, indicating less certainty in predictions.

Study Area	Accuracy & Loss (using S1)
Sudan	0.92, 2.88
Afghanistan	0.83, 6.25
Iran	0.95, 1.68

Results (contd.)

Using only **Sentinel-2 (S2)** data improved accuracy and reduced log loss compared to using S1 alone, with accuracy ranging from 85% to 99% and a lower log loss. This improvement is likely attributed to the additional spectral bands and indices from S2, which offer more detailed information about plant properties such as chlorophyll content, moisture levels, leaf area, and more.

Study Area	Accuracy & Loss (using S2)
Sudan	0.99, 0.24
Afghanistan	0.85, 5.29
Iran	0.96, 1.2

Results (contd.)

Fusing **S1 and S2** data further boosted accuracy for Afghanistan to 87%, while maintaining high accuracy for Sudan and Iran. Log loss was also reduced for Iran. This demonstrated the value of combining optical and radar data.

Study Area	Accuracy & Loss (using S1 and S2)
Sudan	0.99, 0.24
Afghanistan	0.87, 4.57
Iran	0.97, 0.96

Results (contd.)

Adding **Landsat-8** to the **S1+S2** fusion did not yield further improvement. The likely reasons that Landsat-8 did not enhance results are as follows:

- Landsat-8 possesses a coarser spatial resolution (30m) in comparison to Sentinel-2 (10m), thus failing to contribute additional spatial detail.
- Its spectral bands are similar to those provided by Sentinel-2, leading to redundancy in spectral information.

Study Area	Accuracy & Loss (using S1, S2, and L8)
Sudan	0.99, 0.24
Afghanistan	0.86, 5.05
Iran	0.97, 0.96

Results (contd.)

Study Area	Using S1	Using S2	Using S1 and S2	Using S1, S2, and L8
Sudan	Accuracy 0.92, Loss 2.88	Accuracy 0.99, Loss 0.24	Accuracy 0.99, Loss 0.24	Accuracy 0.99, Loss 0.24
Afghanistan	Accuracy 0.83, Loss 6.25	Accuracy 0.85, Loss 5.29	Accuracy 0.87, Loss 4.57	Accuracy 0.86, Loss 5.05
Iran	Accuracy 0.95, Loss 1.68	Accuracy 0.96, Loss 1.2	Accuracy 0.97, Loss 0.96	Accuracy 0.97, Loss 0.96

Conclusion

The synergy between gradient boosting models and the integrated optical and radar data not only demonstrated improved classification accuracy but also showcased the potential for advancing land use and agriculture monitoring applications. This comprehensive approach holds promise for providing valuable insights into dynamic changes in cropland patterns, facilitating better-informed decision-making in the realm of agricultural management and resource allocation.

Wu, N.; Crusiol, L.G.T.; Liu, G.;Wuyun, D.; Han, G. Comparing Machine Learning Algorithms for Pixel/Object-Based Classifications of Semi-Arid Grassland in Northern China Using Multisource Medium Resolution Imageries. Remote Sens. 2023, 15, 750. https://doi.org/10.3390/rs15030750

Trivedi, M.B.; Marshall, M.; Estes, L.; de Bie, C.A.J.M.; Chang, L.; Nelson, A. Cropland Mapping in Tropical Smallholder Systems with Seasonally Stratified Sentinel-1 and Sentinel-2 Spectral and Textural Features. Remote Sens. 2023, 15, 3014.

Yi, Z.; Jia, L.; Chen, Q. Crop Classification Using Multi-Temporal Sentinel-2 Data in the Shiyang River Basin of China. Remote Sens. 2020, 12, 4052. https://doi.org/10.3390/rs12244052

P.G. Teluguntla, et al., A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform, ISPRS J. Photogramm. Remote Sens., 144 (2018), pp. 325-340

Belgiu, Mariana & Csillik, Ovidiu. (2017). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sensing of Environment. 204. 10.1016/j.rse.2017.10.005.

Li, Danya & Gajardo, Joaquin & Volpi, Michele & Defraeye, Thijs. (2022). Using Machine Learning to generate an open-access cropland map from satellite images time series in the Indian Himalayan Region.

Bofana, J.; Zhang, M.; Nabil, M.; Wu, B.; Tian, F.; Liu, W.; Zeng, H.; Zhang, N.; Nangombe, S.S.; Cipriano, S.A.; et al. Comparison of Different Cropland Classification Methods under Diversified Agroecological Conditions in the Zambezi River Basin. *Remote Sens.* 2020, *12*, 2096.

Shelestov A, Lavreniuk M, Kussul N, Novikov A and Skakun S (2017) Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping. Front. Earth Sci. 5:17. doi: 10.3389/feart.2017.00017

Jia, Xiaowei & Khandelwal, Ankush & Kumar, Vipin. (2019). Automated Monitoring Cropland Using Remote Sensing Data: Challenges and Opportunities for Machine Learning

Surabhi Lingwal, Komal Kumar Bhatia, Manjeet Singh; Semantic segmentation of landcover for cropland mapping and area estimation using Machine Learning techniques. *Data Intelligence* 2023; 5 (2): 370–387.

Latif, R.M.A.; He, J.; Umer, M. Mapping Cropland Extent in Pakistan Using Machine Learning Algorithms on Google Earth Engine Cloud Computing Framework. ISPRS Int. J. Geo-Inf. 2023, 12, 81.


Thank You !