Technical achievement and summary of the Geo-AI Challenge cropland extent mapping 2023

Pengyu Hao[1], Mohammad Alasawedah[2], Stella Ofori-Ampofo[3], Julius Maina[4], Vita Lorenzo[5], Alexandre Nobajas Ganau[5], Zhongxin Chen[1]

[1] Digital FAO and Agro-Informatics Division, Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, 00153, Rome, Italy, [2] Eurac Research, Institute for Earth Observation, Viale Druso 1, 39100, Bolzano, Italy, [3] Technical University of Munich, Arcistrasse 21, 80333, Munich. Germany, [4] KCA University, Thika Rd, Nairobi, Kenya, [5] Geospatial Analysis and Programme Delivery Section (GPS), Research and Trend Analysis Branch (RAB), United Nation Office on Drugs and Crime (UNODC), Vienna International Centre, Wagramer Strasse 5, Vienna, Austria

Corresponding author: Pengyu Hao, pengyu.hao@fao.org

Cropland extent serves as a critical determinant for advancing various Sustainable Development Goals (SDGs), and satellite-derived data has been widely used to generate cropland extent maps. To further address the global mission of high-resolution cropland extent mapping, the Food and Agriculture Organization (FAO) and the United Nations Office on Drugs and Crime (UNODC) have proposed "cropland extent mapping" as a focal theme for the 2023 Geo-AI Challenge. Organized by the International Telecommunication Union (ITU) in collaboration with the Zindi platform, tasks of this challenge are divided into two components: (1) annual cropland extent mapping in Sudan and Iran, and (2) temporal cropland mapping in Afghanistan, with pre-provided training samples across all three test regions. Throughout the challenge duration, a total of 74 participating teams submitted their solutions, with the top five teams selected based on classification accuracy of cropland extent maps, innovative methodology, and effectiveness of oral presentation. Several key scientific questions were addressed during the challenge, including optimal classification feature selection, comparative analysis of diverse Machine Learning (ML) models, and fine-tuning of ML algorithms. Importantly, all datasets, scripts, and technical reports resulting from the challenge are openly accessible to the public domain, thereby fostering collaborative advancements within the agricultural remote sensing community.

Keywords: Afghanistan, cropland, Geo-AI Challenge, Iran, machine learning

1. INTRODUCTION

The spatial distribution of global cropland extent serves as the foundation for various analysis in agricultural applications, including crop growth monitor and water management, natural resource utilization, environmental assessment, public health initiatives, and sustainability evaluations, and the role of the agricultural sector and its management is pivotal for fostering the 2030 Sustainable Development Agenda since farmland is the basis for human survival and development [**1**, **2**]. Furthermore, information on the state and changes in agricultural land contributes to the monitoring of the achievement of several of the 17 Sustainable Development Goals (SDGs) and many of their 169 targets [**3**]. Monitoring the cropland extent is directly linked to SDG 2: “Zero Hunger”, but it is easy to understand its indirect contributions to Goal 6 (Clean Water and Sanitation), Goal 12 (Responsible Consumption and Production), Goal 13 (Climate Action), Goal 15 (Life on Land), and Goal 17 (Global Partnership for the Goals), as well as to income-based SDGs such as Goal 1 (No Poverty) and Goal 8 (Decent Work and Economic Growth) [**4**].

Less immediate relationships can be found with Goal 16: “peace, justice, and strong institutions”, but they are evident when it comes to measuring the extent of illicit crop cultivation, such as in the framework of the Illicit Crop Monitoring Programme (ICMP)

© International Telecommunication Union, 2025
Some rights reserved.
This work is available under the CC BY-NC-ND 3.0 IGO license: https://creativecommons.org/licenses/by-nc-nd/3.0/igo/.
More information regarding the license and suggested citation, additional permissions and disclaimers is available at:
https://www.itu.int/en/journal/j-fet/Pages/default.aspx

of the United Nation Office on Drugs and Crime (UNODC).

With several new earth observation plans and others to be implemented soon, more satellite imagery with increasing spatial and temporal resolution is available [5, 6]. However, satellite data processing and land surface information extraction go far beyond the satellite's raw image acquisition and provision. Machine learning and artificial intelligence are promising to improve the accuracy and robustness of land cover classification with satellite images.

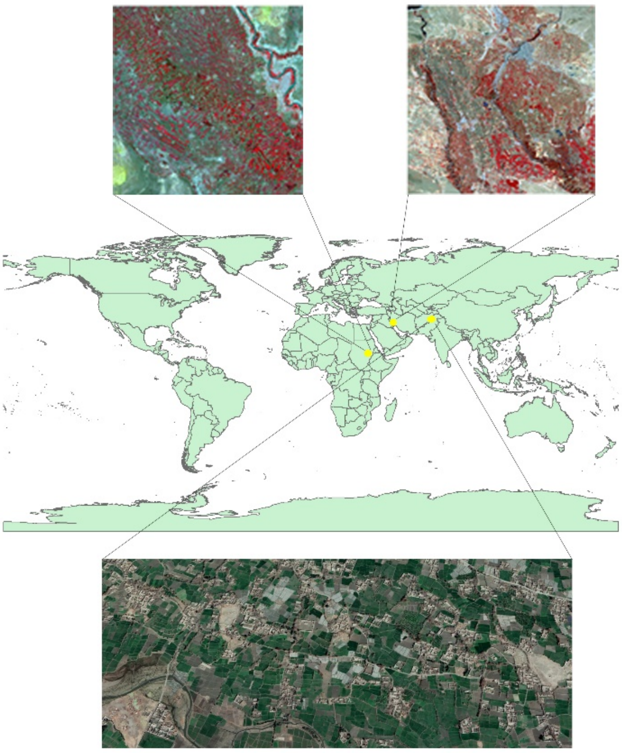

Figure 1 – Locations of the test regions of Geo-AI Challenge 2023

Timely and accurate crop maps are essential for various applications in agriculture, as well as other relevant research fields, such as natural resources, environment, peace, security, health, and sustainability. Cropland extent maps are the basic products for practical agricultural applications. Numerous algorithms have been proposed for cropland mapping using satellite observations, and several freely available land cover products provide global cropland extent maps at a 30m resolution, such as WorldCover, Globaland30, and GFSAD30 [7, 8, 9]. However, the data has several limitations: (1) the data is not updated annually, limiting their usefulness for monitoring changes over time; (2) each product has its definition of "cropland", which differs from FAOSATA's definition of "6620 cropland" or "6621 arable lands"; and (3) the existing global cropland masks have a significant disagreement.

To address these challenges and advance the mission of global high-resolution cropland extent mapping using remote sensing data, the 2023 Geo-AI Challenge of cropland extent mapping section has been proposed by FAO and UNDOC. The challenge aimed at developing accurate and cost-effective classification models for cropland extent mapping with machine learning techniques. There were 74 submissions, and the top 20 teams with high cropland mapping accuracy were selected for the second-round evaluation, and then the top 5 teams were rewarded based on all performances including mapping accuracy, novelty of technical report, and performance of oral presentation. This paper is organized as topics and data provided by FAO and UNODC, solutions of the three participants, and a summary.

2. TASKS AND TRAINING

The tasks of the Geo-AI Challenge 2023 could be divided into two parts: (1) annual cropland extent mapping in two test regions located in Sudan and Iran using satellite data collected between July 2019 and June 2020; (2) temporal cropland extent mapping in Afghanistan using satellite data collected in April 2022. Fig. 1 shows the locations of the test regions and false-color composited satellite images.

2.1 Task and data in Afghanistan

Since 2002, UNODC has provided statistically-sound information on the area under poppy cultivation in Afghanistan. The estimates are based on visual interpretation of VHR commercial satellite images. Images are taken of selected sample units of a 5x5 km grid that is superimposed on potential agricultural land in the country. For the proposed challenge, UNODC provided training and test data produced from the analysis of VHR satellite imagery (Pleiades at 0.5m spatial resolution) acquired over the Nangarhar Province in April 2022, and the model is used to identify cropland extent in the same period.

2.2 Task and data in Sudan and Iran

The tasks for Iran and Sudan are using the satellite image time series collected during July 2019 and June 2020 to identify cropland extent at a 10-meter spatial resolution. There are two test regions, both with a 1-degree by 1-degree extent. 500 training samples were provided for each test region, and the participants could add training samples by themselves. Basically, we encouraged the participants to use a small training sample dataset, so that the total number of training samples in each test region should be smaller than 1000.

 © International Telecommunication Union, 2025

3. SOLUTIONS

3.1 Solution from Mohammad

3.1.1 Materials

Google Earth Engine (GEE) has been widely used for crop mapping to ensure food security and manage natural resources [**10**, **11**]. In this solution, GEE was utilized to extract different satellite data, as described in the following sections, over three different study areas: Afghanistan, Sudan, and Iran. Monthly mean values of the data were computed using GEE over different time periods for each study area.

For Afghanistan, ground truth samples were collected for one month, April 2022. However, we extracted monthly mean values for the past three months (January, February, and March) to enable near-real-time assessment, as a one-month interval would have limited cropland mapping accuracy. For Iran and Sudan, monthly mean values were extracted between July 2019 and July 2020 to cover the time frame of the collected samples.

Sentinel-2 Data: Sentinel-2 satellites, operated by the European Space Agency (ESA), provide high temporal and spatial resolution optical images, with a spatial resolution of 10 m, 20 m, and 60 m spatial resolution, and revisit cycle every 5 days [**12**]. The optical image from Sentinel-2 consists of 13 bands covering the spectrum range from visible to short-wave, including the near-infrared range. Table 1 shows the extracted 10 bands from Harmonized Sentinel-2 MSI, level-2A collection on GEE and used in the solution. The extracted bands have been used to compute vegetation indices, like the Normalized Difference Vegetation Index (NDVI) and Green Normalized Difference Vegetation Index (GNDVI).

Table 1 – Sentinel-2 bands

Band name	Description
B2	Blue
B3	Green
B4	Red
B5	Red Edge 1
B6	Red Edge 2
B7	Red Edge 3
B8	NIR
B8A	Red Edge 4
B11	SWIR 1
B12	SWIR 2

Sentinel-1 Data: The Sentinel-1 mission, operated by ESA, offers high spatial resolution Synthetic Aperture Radar (SAR) data up to 10 m with a high revisit frequency [**13**]. SAR data is not affected by cloud contamination, unlike optical data, due to its ability to penetrate clouds [**14**, **15**]. That's why it has been used in combination with Sentinel-2 data. Sentinel-1 SAR Ground Range Detected (GRD) collection was utilized due to its provision of calibrated and ortho-corrected products. Two polarization bands, VV (vertical transmit/vertical receive) and VH (horizontal transmit/horizontal receive), were extracted for each scene. These two bands were used to generate more features, such as the Radar Vegetation Index (RVI).

SMAP Level-4 Data: The Soil Moisture Active Passive (SMAP) mission, operated by NASA, provides Level-4 soil moisture products with a resolution of 9 km. These products include soil moisture information related to the soil's layers, as well as additional data such as Leaf Area Index (LAI) [**16**]. Despite its coarser resolution compared to Sentinel products, SMAP data has been utilized to account for spatial variation in the surrounding area. Table 2 shows the selected bands from the SMAP Level-4 (L4) collection.

3.1.2 Algorithms

This study focused on two approaches of cropland mapping to gain insight of their respective performance:

1. Weighted ensemble tree-based models. This approach utilized two models, namely Light Gradient Boosting Machine (LightGBM) [**17**], and Categorical Boosting (CatBoost) [**18**], chosen for their optimal computational speed and their ability to handle raw data without normalization [**19**, **20**].
2. Time Series Classification (TSC) utilized by using InceptionTime architecture, an ensemble of convolutional neural network models, due to its scalability and time efficiency [**21**]. This approach will help to exploit the temporal dimension of the time series.

3.1.3 Results and discussion

The best performing approach was the weighted ensemble tree-based models, achieving accuracy scores of 0.857, 0.971, and 0.961 for Afghanistan, Sudan, and Iran, respectively. In comparison, the TSC method achieved 0.614, 0.614, and 0.598 in the same regions. The results suggest that longer time series spans, as seen in Iran and Susan, improve the accuracy of cropland mapping. Conversely, Afghanistan achieved the lowest accuracy, likely due to its shorter time span of only four months.

Fig. 2 shows the importance of the feature for the ensemble tree-based models for Afghanistan. The VH polarization band in February was the significant feature, indicating that Sentinel-1 SAR data is particularly useful for crop mapping in winter due to its insensitivity to atmospheric conditions.

© International Telecommunication Union, 2025

Table 2 – SMAP Level-4 bands

Band name	Description
sm_surface	Top layer soil moisture (0-5 cm)
sm_rootzone	Root zone soil moisture (0-100 cm)
sm_surface_wetness	Top layer soil wetness (0-5 cm)
sm_rootzone_wetness	Root zone soil wetness (0-100 cm)
surface_temp	Mean land surface temperature
land_evapotranspiration_flux	Evapotranspiration from land
vegetation_greenness_fraction	Vegetation greenness
leaf_area_index	Vegetation leaf area index

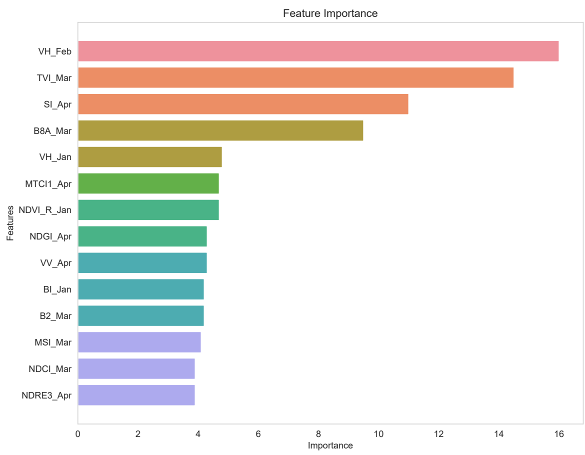

Figure 2 – Feature importance for Afghanistan

Fig. 3 shows the top features of Sudan. It is worth noting that the most significant impact is observed from Sentinel-2 bands during the summer months, particularly in June and July, when the cloud cover is less frequent, and thus optical data is less affected. Fig. 4 illustrates the features' importance for Iran. It shows that the LAI and evapotranspiration have a high impact on cropland mapping in Iran.

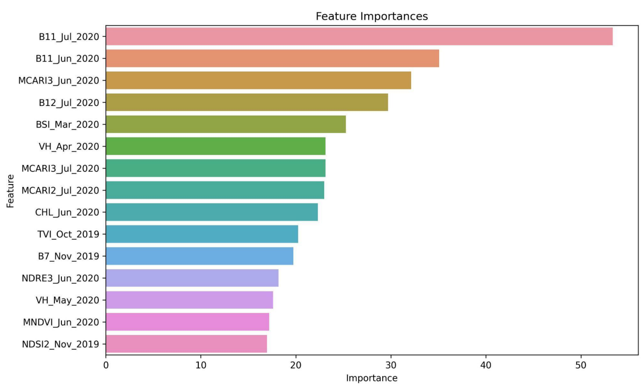

Figure 3 – Feature importance for Sudan

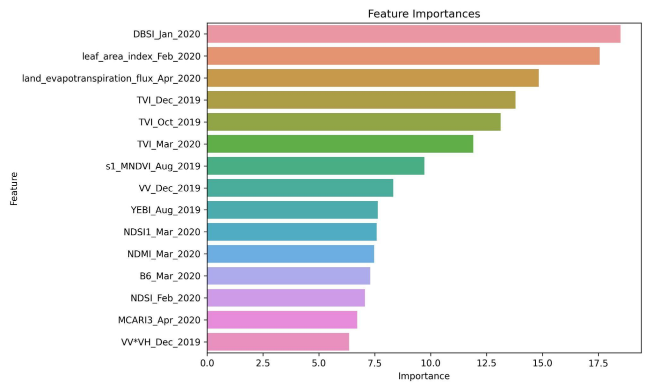

Figure 4 – Feature importance for Iran

3.2 Solution from Stella

Remote sensing data serves as an essential input for the generation of large-scale crop maps. Prominent examples include products such as EU World Cover [**7**], Global Land Analysis and Discovery crop expansion products [**22**], and World Cereal crop maps [**23**]. However, the production of these maps is either irregular (resulting in varying temporal coverage), outdated or utilize distinct nomenclature for classifying cropland, leading to significant uncertainty [**24**]. An analysis of regional-level crop maps conducted by [**25**] on numerous crop maps revealed a lack of consensus, primarily stemming from intra-class variance. This underscores the challenge of developing a singular comprehensive product that accurately represents the diverse definitions and cropping seasonality across different landscapes. In light of these complexities, our approach to mapping cropland in the three regions is predicated on recognizing these nuances, treating each region as an independent entity.

3.2.1 Materials

Only 1000 ground labels (500 each of crop and non-crop areas) are provided per region. Initially, a workflow is devised, leveraging the Google Earth Engine API to extract Sentinel-2 surface reflectance time series for each label [**26**]. All spectral bands, except the aerosol and water vapor bands, are utilized, resulting in a total of 10 bands.

© International Telecommunication Union, 2025

Out of these, 15 spectral indices, mostly encompassing vegetation greenness, biomass, and canopy water content, are derived. For simplicity, the names of the spectral indices are provided below. An interested user can consult the GitHub repository or the index database (https://indexdatabase.de) for the formulas used to derive these indices.

- normalized Difference Vegetation Index (NDVI)
- NDVI red-edge
- normalized difference water index
- enhanced vegetation index
- normalized difference built index
- soil adjusted vegetation index
- browning reflectance index
- chlorophyll index
- land surface water index
- normalized difference pond index
- water ratio index
- plant senescence reflectance index
- difference vegetation index
- ratio vegetation index
- visual atmospheric resistance index

A limitation of optical remote sensing is the presence of clouds that often hinder the usability of satellite images. In this workflow, the identification and removal of cloudy observations are conducted manually by scrutinizing satellite images independently for each region. This manual approach is chosen over filtering by cloudy pixel percentage since the latter cannot provide insights into the location of clouds within an image. Alternatively, the Scene Classification (SCL) layer can be employed to locally mask clouds in a fully automated workflow. However, this substitution may lead to samples with uneven time series lengths, necessitating temporal interpolations to ensure uniformity, a typical requirement of machine learning models. The final step in our data extraction workflow involves spatial mosaicking. Here, footprints acquired on the same day are merged to prevent points that intersect overlapping areas from containing additional and redundant temporal information.

It is also important to recognize the conditions imposed on extracting satellite data for the ground labels. In Iran and Sudan, ground labels were surveyed between July 2019 - June 2020, and only in April 2022 for Afghanistan. This means that, while there is enough information to distinguish the temporal characteristics of crops and non-crops in Iran and Sudan, this information is limited in Afghanistan. Table 3 shows the input shape in the dimension (time × number of channels).

Table 3 – Time window and input shape per region

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

3.2.2 Algorithms

Separate classical machine learning models namely Random Forest (RF), extreme Gradient Boosting (XGBoost), Support Vector Classifier (SVC) and k-Nearest Neighbor (KNN) are trained for each region. For each model, we experiment using either spectral bands only (SB), spectral indices only, or their combination in a 5-fold cross-validation. A hyperparameter search for each experiment setup is achieved using Optuna to find the best set of parameters that maximizes the overall accuracy [**27**].

$$accuracy = \frac{\text{number of correct predictions}}{\text{total number of predictions}} \tag{1}$$

As advancements in AI continue to progress rapidly, research endeavors comparing classical machine learning methodologies like random forest with deep neural networks frequently reveal competitive accuracies [**28**, **29**]. Although our ultimate submission leans towards classical machine learning, we delve deeper into investigating the advantages of Deep Learning (DL) models when contrasted with classical baseline approaches.

3.2.3 Results and discussion

This study undertook a comparative analysis of the performance of four classical baseline models across various combinations of regions and features. Table 4 presents the accuracy metrics for the four classical machine learning models.

In Afghanistan, the best performances can be achieved using RF applied to spectral bands, XGBoost on spectral bands and indices altogether or KNN on indices or spectral bands enriched with indices. RF is preferred here since it uses fewer features to reach an accuracy of about 82%. Despite the limited amount of time steps for samples in Afghanistan, the models still achieved acceptable accuracies.

In both Iran and Sudan, achieving an accuracy exceeding 94% is feasible, contingent upon the specific configuration. Integrating spectral bands in conjunction with indices frequently produced outcomes on par with those derived solely from indices. This implies that augmenting features does not invariably enhance model efficacy.

© International Telecommunication Union, 2025

Table 4 – Performance comparison of different models and feature combinations for each region

Model	Spectral bands	Indices	SB + Indices
Afghanistan			
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
Iran			
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
Sudan			
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

Additionally, the superiority of employing indices over spectral bands was apparent in a performance boost ranging from 1% to 4%, except for the Support Vector Classifier (SVC) in Sudan, where the enhancement was notably substantial, at approximately 20%.

Impact of increased time steps: The high performance of Iran and Sudan can be attributed to the high number of time steps available to aid the classification of crops and non-crops. To improve the accuracy in Afghanistan, the phenology for selected croplands was examined beyond the limited time window (1 – 30 April). Fig. 5 suggests that crops were in season prior to and after the month of April.

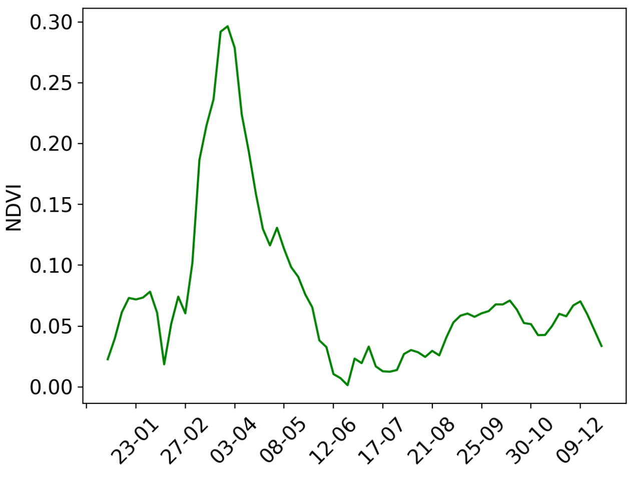

Figure 5 – Annual NDVI profile of a cultivated area in Afghanistan

The time window is then adjusted by one month (March – May) to triple the number of time steps. Table 5 shows that, in effect, this increases the mapping accuracy by 2% on spectral bands and nearly 5% on indices (with respect to RF).

Table 5 – Performance comparison on short and long time steps using random forest

Time	Spectral bands	Indices	SB + Indices
April	0.824	0.790	0.810
March-May	0.844	0.838	0.826

Benefiting from deep learning models: Although it is acknowledged that classical machine learning methods do not effectively utilize temporal order, two deep learning models, namely TempCNN [**29**] and LSTM [**30**], are applied using implementations from [**28**]. TempCNN applies 1D convolutions across dimensions to learn spectro-temporal features. LSTMs, on the other hand, leverage recurrent connections to capture sequential dependencies through memory cells. In addition, a Multilayer Perceptron (MLP) is used as a deep learning baseline. MLPs are comparable to RF in that they do not regard temporal order. In a similar way to the classical ML models, each model is tuned with Optuna under the same cross-validation configuration.

Table 6 – Comparing accuracies for random forest and deep learning models on limited time steps for Afghanistan

	RF	MLP	TempCNN	LSTM
Spectral bands	0.824	0.784	0.796	0.808
Indices	0.790	0.774	0.862	0.858

Table 6 compares the accuracy for random forest and deep learning models when considering spectral bands or indices. On spectral bands, DL RF outperformed all DL models with a margin between 2-4%; however, TempCNN and LSTM applied to indices surpassed RF's performance by about 4%.

3.3 Solution from Julius

3.3.1 Materials

Our study area spanned two continents, Africa (Sudan) and Asia (Iran and Afghanistan), as shown in Fig. 1. For Sudan and Iran, the time span was a two full-year classification while for Afghanistan there was a temporal classification using one month of data. We attempted to build one model that could generalize well across these countries/continents with varying crop types, farming practices, seasons, and using varying time spans.

The satellite imagery datasets were extracted from Sentinel-2 using the Google Earth Engine (GEE). For Afghanistan, the time span for the imagery spanned one month, April 2022, and for both Iran and Sudan from July 2019 to June 2020 for both train and validation datasets. The cropland fields and training samples collected for the test regions are shown in figures 6, 7, and 8. The Afghanistan test region presented a unique challenge and opportunity

© International Telecommunication Union, 2025

Figure 6 – Iran croplands sample

Figure 7 – Sudan croplands sample

Figure 8 – Afghanistan croplands sample

for extra data collection. This work ventured beyond provided data sources, seeking to augment the available data and broaden the scope of analysis. As such, an additional 500 training samples, well-labeled from Afghanistan's Nangarhar province, were introduced to enrich the dataset. These samples were extracted from GEE after the registration of a Google account and subsequent GEE enrollment. This tool proved instrumental in data acquisition and analysis.

3.3.2 Algorithms

To ensure the efficacy of our predictive model, we commenced with a comprehensive comparison of various machine learning algorithms. The following models were considered for evaluation: logistic regression, random forest, AdaBoost, XGBoost, HistGradientBoosting, LightGBM, and CatBoost. These models were trained and tested to gauge their performance on our dataset. The evaluation was based on accuracy, a fundamental metric for classification tasks. The results of this initial comparison are summarized in Table 7.

It is evident from the comparison that both XGBoost and LightGBM outperformed other models in terms of

Table 7 – Performance of various models before hyperparameter tuning

Model	Accuracy
XGBoost	0.884
LightGBM	0.880
HistGradientBoosting	0.874
Random Forest	0.866
CatBoost	0.864
AdaBoost	0.828
Logistic Regression	0.682

accuracy. For fast and efficient training and evaluation LightGBM was chosen as the proposed model because it is light and fast [**30**]. This final proposed model was subjected to hyperparameters' tuning to determine the optimal parameters for best performance. Optuna, an efficient tuning approach was applied [**27**]. The following were optimized: number of boosting rounds, learning rate, maximum depth of the trees and the L2 regularization strength. After getting the optimal parameters, we employed a repeated k-fold cross-validation with five folds and three repeats to ensure robust model evaluation and selection of the best hyperparameters [**32**]. This approach allows for thorough testing and validation of the model's performance.

3.3.3 Results and discussion

Log loss learning curves for evaluating results were considered to help detect where overfitting and underfitting occurred. Learning curves provide insight into the dependence of a learner's generalization performance on the training set size. This important tool can be used for model selection, to predict the effect of more training data, and to reduce the computational complexity of model training and hyperparameter tuning. The validation loss curve (Fig. 9) initially began at a value of 0.7 and experienced a decrease before it flattened around 0.3. The training loss curve similarly commenced at the same value and underwent a decrease. However, the curve did not flatten and showed a continued decrease of the log loss yet the validation log loss had plateaued. This was around iteration 200 showing the point at which overfitting started and therefore further training would be stopped.

Accuracy, representing the ratio of correctly classified samples to the total samples, served as the primary evaluation metric for assessing the performance of the LightGBM model. When assessing the performance of the LightGBM model on a validation set generated through an 80-20 split of the training dataset, it achieved an accuracy of 0.87 across different folds with low standard deviations of 0.0167. Subsequently, the same model was utilized to train on the entire training dataset and make predictions on an unseen withheld dataset, resulting in

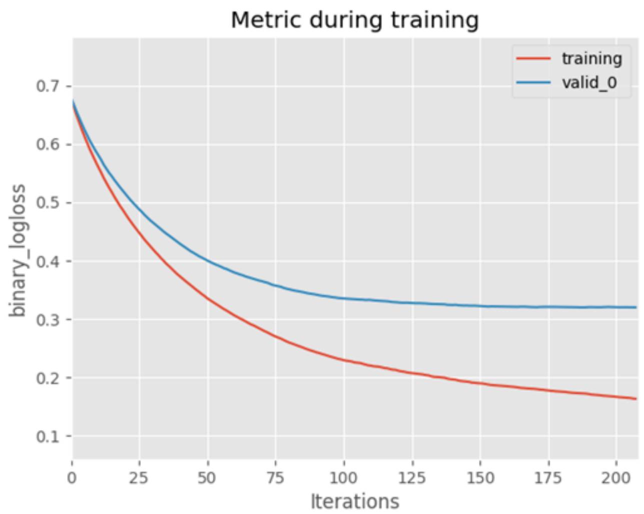

Figure 9 – Training loss curve with 10000 rounds and an early stopping of 20 rounds (80% training, 20% validation)

an accuracy score of 0.913 on the public leaderboard and 0.918 on the private leaderboard, as detailed in Table 8.

Table 8 – Performance of the LightGBM model (80% training, 20% validation)

Metric	Accuracy
Cross-validation	0.870
Public leaderboard	0.913
Private leaderboard	0.918

4. SUMMARY

The 2023 Geo-AI Challenge of cropland extent mapping has achieved good outputs; participants have contributed to excellent advancements for global cropland mapping, enabling a more precise and comprehensive understanding of agricultural landscapes worldwide. Thanks to ITU for organization this and the Zindi platform for the operation of the challenge (https://zindi.africa/competitions/geo-ai-challenge-for-cropland-mapping-with-satellite-imagery). The cropland maps, scripts, and technical reports for all the prized submissions are open to the public on GitHub (https://github.com/ITU-GeoAI-Challenge).

REFERENCES

[1] P. Hao, F. Löw, and C. Biradar, "Annual Cropland Mapping Using Reference Landsat Time Series – A Case Study in Central Asia," *Remote Sensing*, vol. 10, no. 12, p. 2057, 2018. [Online]. Available: http://www.mdpi.com/2072-4292/10/12/2057.

[2] R. Yang et al., "Stabilizing unstable cropland towards win-win sustainable development goals," Environmental Impact Assessment Review, vol. 105, p. 107395, 2024/03/01/ 2024, doi: https://doi.org/10.1016/j.eiar.2023.107395.

[3] United-Nations. "Sustainable Development Goals (SDGs)." https://www.un.org/sustainabledevelopment/ (accessed Feb 22, 2024).

[4] A. K. Whitcraft, I. Becker-Reshef, C. O. Justice, L. Gifford, A. Kavvada, and I. Jarvis, "No pixel left behind: Toward integrating Earth Observations for agriculture into the United Nations Sustainable Development Goals framework," Remote Sensing of Environment, vol. 235, p. 111470, 2019/12/15/ 2019, doi: https://doi.org/10.1016/j.rse.2019.111470.

[5] USGS. "Landsat 8 (L8) Data Users Handbook." http://landsat.usgs.gov/l8handbook_section1.php

[6] ESA. "Sentinel-2 for Agriculture." http://www.esa-sen2agri.org/SitePages/sentinel2.aspx

[7] D. Zanaga et al. "ESA WorldCover 10 m 2021 v200." https://zenodo.org/records/7254221 (accessed December, 7, 2023)

[8] J. Chen et al., "Global land cover mapping at 30m resolution: A POK-based operational approach," ISPRS Journal of Photogrammetry & Remote Sensing, vol. 103, pp. 7-27, 2015.

[9] W. USGS. "Global Food-and-Water Security-support Analysis Data (GFSAD)." https://www.usgs.gov/centers/western-geographic-science-center/science/global-food-and-water-security-support-analysis (accessed Dec, 12, 2023).

[10] L. Liu et al., "Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine," Remote Sensing of Environment, vol. 239, p. 111624, 2020/03/15/ 2020, doi: https://doi.org/10.1016/j.rse.2019.111624.

[11] H. Tian, N. Huang, Z. Niu, Y. Qin, J. Pei, and J. Wang, "Mapping Winter Crops in China with Multi-Source Satellite Imagery and Phenology-Based Algorithm," Remote Sensing, vol. 11, no. 7, p. 820, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/7/820.

[12] K. Segl, L. Guanter, F. Gascon, T. Kuester, C. Rogass, and C. Mielke, "S2eteS: An End-to-End Modeling Tool for the Simulation of Sentinel-2 Image Products," Ieee Transactions on Geoscience and Remote Sensing, vol. 53, no. 10, pp. 5560-5571, Oct 2015, doi: 10.1109/tgrs.2015.2424992.

[13] R. Torres et al., "GMES Sentinel-1 mission," Remote Sensing of Environment, vol. 120, pp. 9-24, 2012/05/15/ 2012, doi: https://doi.org/10.1016/j.rse.2011.05.028.

 © International Telecommunication Union, 2025

[14] A. Shelestov et al., "Cloud Approach to Automated Crop Classification Using Sentinel-1 Imagery," IEEE Transactions on Big Data, vol. 6, no. 3, pp. 572-582, 2020, doi: 10.1109/TBDATA.2019.2940237.

[15] L. Orlíková and J. Horák, "Land cover classification using sentinel-1 SAR data," in 2019 international conference on military technologies (ICMT), 2019: IEEE, pp. 1-5.

[16] R. H. Reichle et al., "Version 4 of the SMAP level-4 soil moisture algorithm and data product," Journal of Advances in Modeling Earth Systems, vol. 11, no. 10, pp. 3106-3130, 2019.

[17] R. H. Reichle et al., "Version 4 of the SMAP level-4 soil moisture algorithm and data product," Journal of Advances in Modeling Earth Systems, vol. 11, no. 10, pp. 3106-3130, 2019.

[18] G. Ke et al., "Lightgbm: A highly efficient gradient boosting decision tree," Advances in neural information processing systems, vol. 30, 2017.

[19] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," Advances in neural information processing systems, vol. 31, 2018.

[20] D. A. McCarty, H. W. Kim, and H. K. Lee, "Evaluation of light gradient boosted machine learning technique in large scale land use and land cover classification," Environments, vol. 7, no. 10, p. 84, 2020.

[21] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," the Journal of machine Learning research, vol. 12, pp. 2825-2830, 2011.

[22] H. Ismail Fawaz et al., "Inceptiontime: Finding alexnet for time series classification," Data Mining and Knowledge Discovery, vol. 34, no. 6, pp. 1936-1962, 2020.

[23] P. Potapov et al., "Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century," Nature Food, vol. 3, no. 1, pp. 19-28, 2022/01/01 2022, doi: 10.1038/s43016-021-00429-z.

[24] K. Van Tricht et al., "WorldCereal: a dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping," Earth Syst. Sci. Data, vol. 15, no. 12, pp. 5491-5515, 2023, doi: 10.5194/essd-15-5491-2023.

[25] F. N. Tubiello et al., "Measuring the world's cropland area," Nature Food, vol. 4, no. 1, pp. 30-32, 2023/01/01 2023, doi: 10.1038/s43016-022-00667-9.

[26] H. Kerner et al., "How accurate are existing land cover maps for agriculture in Sub-Saharan Africa?," arXiv preprint arXiv:2307.02575, 2023.

[27] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," Remote Sensing of Environment, vol. 202, pp. 18-27, 2017/12/01/ 2017, doi: https://doi.org/10.1016/j.rse.2017.06.031.

[28] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019, pp. 2623-2631.

[29] M. Rußwurm, C. Pelletier, M. Zollner, S. Lefèvre, and M. Körner, "BreizhCrops: A Time Series Dataset for Crop Type Mapping. arXiv 2020," arXiv preprint arXiv:1905.11893.

[30] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," Remote Sensing, vol. 11, no. 5, p. 523, 2019.

[31] L. S.-T. Memory, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 2010.

[32] Y. Zhang, C. Zhu, and Q. Wang, "LightGBM-based model for metro passenger volume forecasting," IET Intelligent Transport Systems, vol. 14, no. 13, pp. 1815-1823, 2020.

[33] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," Statistics and Computing, vol. 21, pp. 137-146, 2011.

ACKNOWLEDGEMENT

The authors thank Ms Livia Peiser from Land & Water Division (NSL) of FAO and Saied Pirasteh from University of Waterloo for providing ground truth samples, which are used for challenge submission validations. The authors also thank ITU for the operation of the challenge and the fund provided by both ITU and UNODC for awarding the challenge winners.

AUTHORS

Pengyu Hao is an IT Officer at the Food and Agriculture Organization of the United Nations. He works in remote sensing data processing, land surface data comparison and data quality control.

MOHAMMAD ALASAWEDAH is an Earth observation and climate change researcher, currently working within a research group focused on advanced computing for Earth observation.

STELLA OFORI-AMPOFO is a PhD student at the Chair of Data Science in Earth Observation, Technical University of Munich, Germany. She specializes in applying cutting-edge artificial intelligence techniques for food supply forecasting using remote sensing data. Her research endeavors are to deepen our understanding of agricultural systems, refine methods for crop yield estimation, and enable the development of early warning systems to mitigate food shortages. Driven to make meaningful impacts, she actively collaborates with interdisciplinary teams of researchers, policymakers, and stakeholders to translate research findings into actionable insights and solutions.

JULIUS MAINA is an aspiring data scientist currently undertaking a Master of Science in Data Analytics program at KCA University in Kenya. Prior to this master's degree, he pursued a Bachelor of Education with specialization in mathematics and economics from the University of Nairobi. Julius has worked on multiple problem statements, among them are Natural Language Processing (NLP), computer vision and time-series problems.

VITA LORENZO is serving as a Geospatial Information Officer for the Illicit Crop Monitoring Programme (ICMP) at the Research and Trend Analysis Branch of the United Nations Office of Drugs and Crime (UNODC) in Vienna, Lorenzo has been actively involved in designing and implementing technical assistance programs for the development of national illicit crop monitoring systems.

ALEXANDRE NOBAJAS GANAU is a GIS and Remote Sensing expert serving as an Assistant Geographic Information Officer at the United Nations Office of Drugs and Crime (UNODC). He has been involved in technical assistance programs for satellite data-based cropland monitoring.

ZHONGXIN CHEN is a senior IT officer at the Digitalization and Informatics Division at United Nations Food and Agricultural Organization (FAO). Now he is leading the development and operation of FAO's Hand-in-Hand Initiative Geospatial Platform, agricultural applications of cutting-edge technologies including artificial intelligence (AI) and geospatial IT. He is Co-Chair of the Geo-AI Working Group of UN Open GIS.

© International Telecommunication Union, 2025