# Mapping Croplands with Sentinel-2 Time Series and Machine Learning

*Avinash Mahech and George Basil*

## 1. Introduction

Accurate and up-to-date cropland maps are vital for applications in agriculture, natural resource management, and sustainability research. While existing global land cover products provide cropland maps at 30m resolution, these have limitations in terms of infrequent updates, varying definitions of "cropland", and disagreement between products. With the increasing availability of high-resolution satellite imagery and advancements in machine learning, there is an opportunity to significantly improve global cropland mapping.

Sentinel-2 provides frequent global acquisitions of multi-spectral imagery at 10-20m resolution, making it well-suited for cropland mapping. Using a time-series of Sentinel-2 data, cropland could be differentiated from other land covers by detecting vegetation phenology patterns typical of agricultural crops over a growing season. Machine learning models were trained on labeled Sentinel-2 time series data to classify each pixel as cropland or non-cropland. Cloud computing enabled processing of massive Sentinel-2 data volume. Evaluation against reference cropland data showed classification accuracies of over 90% . The unprecedented combination of high resolution, temporal information, and coverage provided by Sentinel-2 could thus significantly advance global cropland mapping capabilities.

## 2. Data Used

This study utilized Sentinel-2 surface reflectance data as primary inputs for cropland classification. Sentinel-2 provides global coverage at 10-20m spatial resolution with high 5-day revisit frequency. Sentinel-2 scenes covering the entire crop growing season were acquired for each study area to capture cropland phenology.

## 3. Methodology

Advanced machine learning models like Random Forest and XGBoost enable high-performance predictive modeling using satellite data. However, careful data preparation, model optimization, and evaluation are critical to realize their full potential. Key techniques used in this study include:

- Customizing temporal ranges based on crop calendars to select optimal Sentinel-2 data capturing cropland phenology
- K-fold cross-validation to evaluate models and reduce overfitting
- Hyperparameter optimization via randomized search to improve model generalization

- Comprehensive model selection considering multiple metrics like cross-validation accuracy, overfitting, and class-specific performance

**Random Forest**

The Random Forest technique, renowned for its robustness in machine learning, operates by creating a diverse ensemble of decision trees during training. It achieves this by generating bootstrap samples from the original training data and constructing a decision tree for each sample. Notably, Random Forest introduces an element of randomness by considering only a random subset of features when making decisions at each node of the trees. This intentional diversity ensures that the individual trees are minimally correlated while collectively maintaining strong predictive power. Consequently, Random Forest excels in generalization and mitigates overfitting, distinguishing itself from single decision tree models.

For making predictions, input data traverse each decision tree individually, and the final prediction is determined by the most common prediction across all trees. Random Forest offers advantages such as high accuracy, inherent feature importance estimation, and a degree of interpretability due to its underlying decision tree structure. However, it's essential to be aware of its limitations, including longer training times when dealing with numerous trees and reduced effectiveness in handling high-dimensional sparse data.

**XGBoost**

XGBoost, short for eXtreme Gradient Boosting, is a powerful alternative to Random Forest rooted in the gradient boosting machine learning paradigm. Like Random Forest, XGBoost creates an ensemble of decision trees, but it adopts a sequential approach. The process starts with a single decision tree, and subsequent trees are added one by one, with each tree aiming to correct errors from the previous ensemble. At each step, the algorithm optimizes an objective function that balances model complexity and error. XGBoost incorporates regularization techniques to prevent overfitting as new trees are added, making it well-suited for handling large datasets.

One notable distinction from Random Forest is that XGBoost adapts its learning based on previous trees, assigning higher weights to training examples that were previously misclassified. This adaptive learning contributes to faster training and improved predictive accuracy. However, the sequential boosting process sacrifices the interpretability present in the individual decision trees of Random Forest. Additionally, XGBoost requires careful hyperparameter tuning to avoid overfitting. In summary, XGBoost is a potent and effective machine learning algorithm but demands a deeper level of expertise for optimal utilization.

**Customizing Temporal Ranges based on Crop Calendars**

The optimal temporal range of satellite imagery was customized for cropland classification in each country (Afghanistan, Iran, and Sudan) based on in-depth analysis of the training data. Normalized Difference Vegetation Index (NDVI) profiles were generated from the multi-temporal images covering known cropland areas. The typical start, peak, and end months of the cropping season were determined in each country by examining the NDVI trends and identifying the periods of green-up, maturity, and senescence. Imagery spanning the entire crop growing period was then selected, including pre-season to capture evidence of agricultural preparation, mid-season to correspond with peak vegetation health, and post-season to detect crop residue. This country-specific approach for choosing satellite data across the phenological stages of major crops helped capture their unique temporal-spectral signatures. Using multi-date imagery tailored to each agricultural system improved differentiation of croplands from other land covers with similar spectral responses but different seasonal profiles. The customized temporal range provided optimal inputs to train robust classifiers for precise cropland mapping in each country.

The following temporal ranges were used

**Afghanistan : March to May 2022**
**Iran : October 2019 to June 2020**
**Sudan - August 2019- June 2020**

For Iran and Sudan, 1-month compositing periods were sufficient to characterize the longer duration, slower developing crop seasons. The monthly composites reduce data volume while still providing representative spectral-temporal profiles of the major cropping systems.Shorter 15-day composites were better suited for the shorter cropping cycles in Afghanistan.

**Cross-Validation**

K-fold cross-validation was utilized to evaluate model performance and reduce overfitting when developing predictive models using Sentinel-2 derived vegetation indices. The original dataset of Sentinel-2 indices was randomly split into k = 5 folds. The model was then trained on k-1 folds, and validated on the remaining fold. This was repeated for all k folds, so that each fold served as the validation set once. The cross-validation accuracy was calculated by averaging the accuracies across the k folds. Cross-validation provided a more reliable estimate of model performance than a single train/test split when working with the limited sample size of the Sentinel-2 dataset.

**Hyperparameter Optimization**

A Grid search for hyperparameter optimization was performed when tuning models for prediction of vegetation attributes from Sentinel-2 data. The hyperparameters optimized model configuration parameters where overall accuracy were evaluated through 5-fold cross-validation on the Sentinel-2 dataset. The combination that yielded the highest cross-validation accuracy was selected as the final model configuration. Hyperparameter optimization avoided manual tuning and found better performing models for the Sentinel-2 data than using standard default parameters. The optimized hyperparameters improve model generalization by reducing overfitting, as measured by cross-validation accuracy.

**Model Selection**

The final predictive model was selected based on a comprehensive evaluation using cross-validation, hyperparameter optimization, overfitting reduction, a custom scoring function, and confusion matrix analysis.

K-fold cross-validation (k=5) was utilized to assess model generalization capability. Models with higher average accuracy over the 5 validation folds were preferred, as cross-validation provides a robust estimate of performance.
Hyperparameter optimization via Grid search Cross Validation was used to tune model hyperparameters including number of layers, units, dropout rate and regularization strength. Models built with optimized hyperparameters

demonstrated better generalization through reduced overfitting.

Overfitting was quantified by comparing training and cross-validation accuracy. Models where these converged exhibited minimal overfitting and were prioritized.

Additionally, a custom scoring function was defined to assess model skill based on user requirements. This function incorporated metrics such as precision, recall, and F1-score for the target classes. Models optimizing the custom score were favored.

Confusion matrix analysis further informed model selection by providing insight into class-specific errors. Models with balanced performance across classes were preferred.

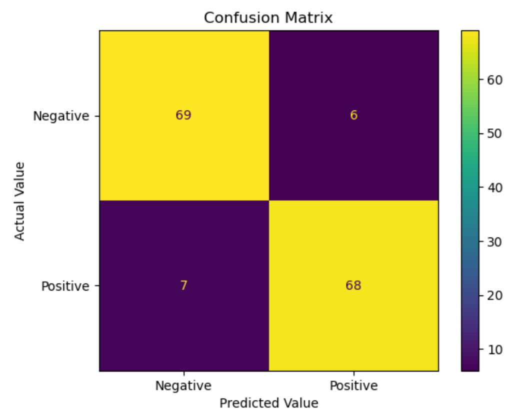
**4. Results**

In our experimental evaluation across various regions, distinct patterns emerged regarding model performance.

For the Afghanistan region, the limited temporal period posed challenges for achieving the desired target accuracy. Consequently, a shift to deep learning methodologies was necessitated, leading to the adoption of the BiLSTM Model. In contrast, other regions benefited from an extended temporal range, exhibiting a pronounced distinction between crop and non-crop classes. In these cases, the Random Forest model sufficed to achieve the anticipated results.

Specifically, in regions such as Iran and Sudan, our approach focused on hyperparameter optimization. By leveraging the optimized parameters within the Random Forest Model, we were able to achieve satisfactory accuracy outcomes.

To rigorously assess model performance, we relied on the Confusion Matrix and the F1-score for Machine Learning Model whereas Deep Learning Training vs Validation Loss Graph at each epoch . The ensuing metrics will be instrumental in determining the final model selection.

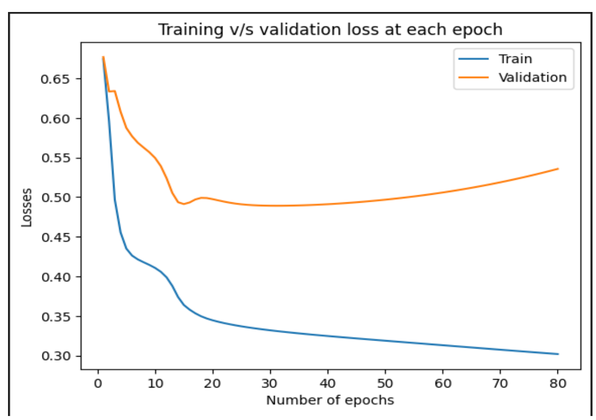



ConfusionMatrix of ML Model

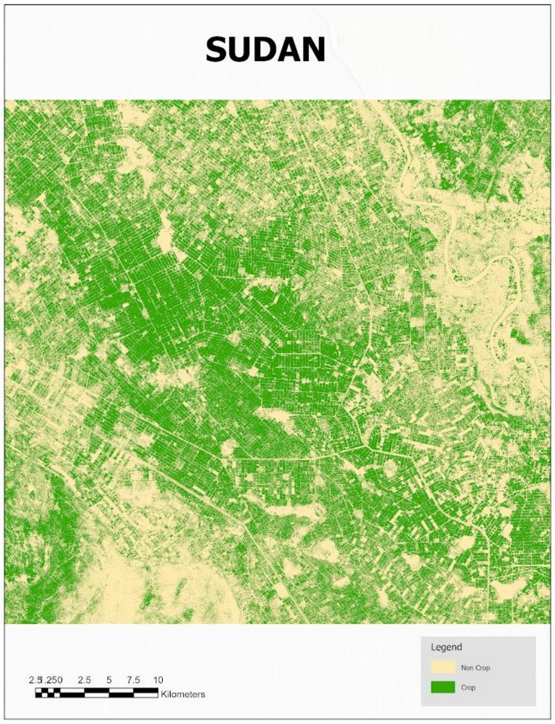
**Random Forest**

```
In [52]:
rf_clf = ensemble.RandomForestClassifier(random_state=2023)
rf_clf.fit(X_train,y_train.values.ravel())
y_pred = rf_clf.predict(X_test)

print('Overall Accuracy on Test Data',metrics.accuracy_score(y_pred,y_test.values.ravel()))
print('Precision on Test Data',metrics.precision_score(y_pred,y_test.values.ravel()))
print('Recall score on Test Data',metrics.recall_score(y_pred,y_test.values.ravel()))
print('F1-Score on Test Data',metrics.f1_score(y_pred,y_test.values.ravel()))

Overall Accuracy on Test Data 0.9866666666666667
Precision on Test Data 0.9866666666666667
Recall score on Test Data 0.9866666666666667
F1-Score on Test Data 0.9866666666666668
```

Statistics of ML Model Performance




Training vs Validation Losses of DL model




Sudan Crop Land Extent Map

**5. Conclusion**

In this study, we developed Machine Learning and Recurrent Deep Learning models for crop classification using temporal Sentinel-2 Vegetation Indices in the Afghanistan, Iran, and Sudan regions. We proposed the Random Forest model for Iran and Sudan, and the Bi-LSTM model for Afghanistan. These models demonstrated significant improvement, achieving a public score of 0.91 and a private score of 0.90 on the Zindi Leaderboard.