REMOTE SENSING-BASED CROP CLASSIFICATION FOR SUSTAINABLE AGRICULTURE IN SUDAN

*Abrar MohamedElmugadad*** *Ammar Nasr** *Samer Mohamed***

**AmunData* ***DataQ* ****University of Khartoum*

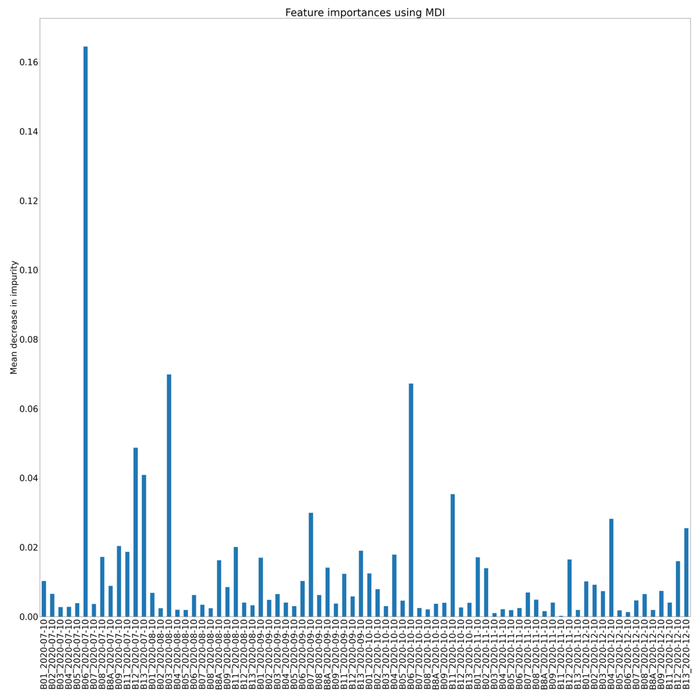
BACKGROUND

Crop have been monitered and their attributes evaluated for many years in order to monitor food security, direct the optimum use of the land, and support agricultural policy. High-quality crop mapping has become a requirement for most nations and is a hot topic in the policy, economics, and land management disciplines because to its significance in national and international economics, trade, and food security.

Currently, the majority of countries use publicly available satellite imagery to categorize crop kinds and evaluate biophysical characteristics due to its affordability, comprehensive perspective, and multitemporal monitoring. The capacity of optical and/or remote sensing techniques to describe the structure and biochemical properties of plants has made them an essential tool for crop information retrieval.

OBJECTIVE

- 85% minimum overall target accuracy
- National in size, with a final spatial resolution of 10 m.
- Operational program, with a largely automated workflow.



ABSTRACT

This study seeks to calculate the area that specific crops cover in a given area of interest. The authors' focus is on obtaining the required level of classification accuracy to support governmental authorities in managing food security requirements and strategic planning.

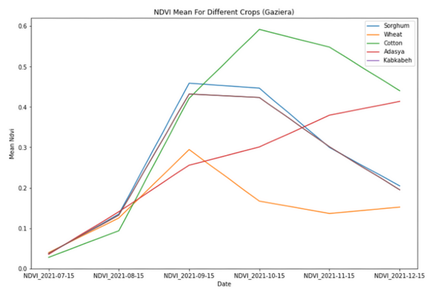
In order to create a multi-class classification pipeline that goes from image acquisition all the way through inference, optical satellite imagery is combined with ensemble machine learning models like Random forest, which have shown to be a good fit for such cases.

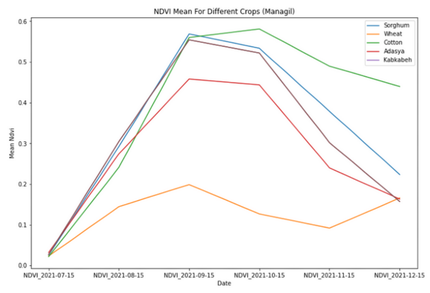
In this study, the Near Infrared, Red, Green, and Blue spectral bands of the Sentinel-2 satellite are stacked for classification. Overall, Random Forest's categorization accuracy using Sentinel-2 imagery is 89.22%.

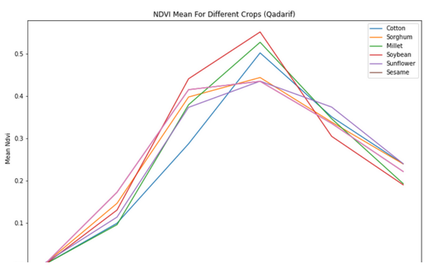
A 2.4 KM2 random sample size and 385 samples are used to generalize this accuracy to the state of El Gaziera, producing a 95% confidence interval and a 5% margin of error.

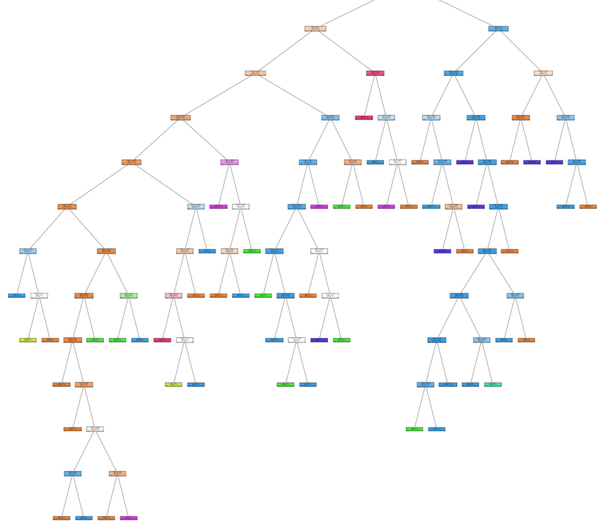
Methodology & Data

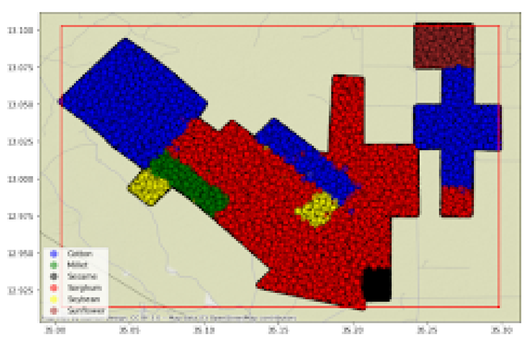
- The training data is gathered from field surveys conducted in El Gezira, EL Managil, and Qadarif, encompassing a 274 km2 area and including all forms of agriculture practiced in Sudan, including irrigated, semi-irrigated, and rainfed farming.
- The growth cycle is then established by digitizing the field boundaries and correlating the NDVI data throughout the year.
- The features are then filtered based on Feature importance based derived from the mean decrease in impurity metric for the Satellite image for the chosen cycles.
- ID for a Random Forest Model was chosen as the best model out of a range of hyper-parameters, namely the number of estimators and the depth of the tree, using k-fold cross validation on 80% of the training data.
- The 20% test set is used to assess the performance of the model based on the F1 score for each crop and the overall accuracy.

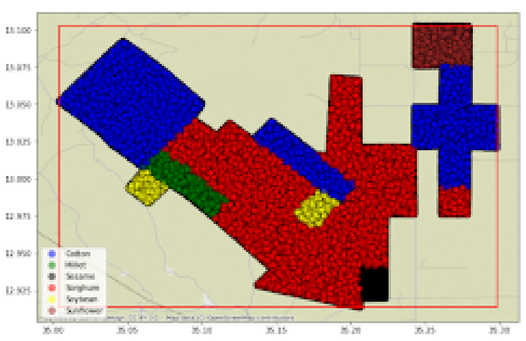


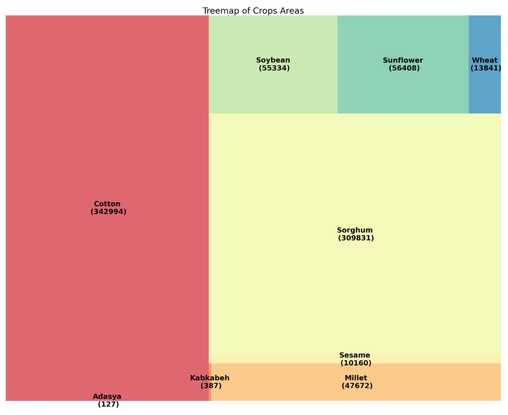






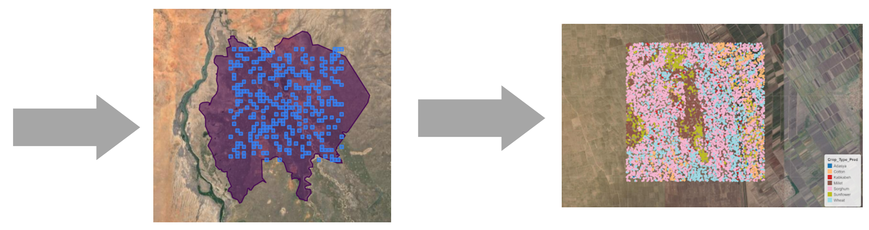






Results

	precision	recall	f1-score	support
Adasya	0.30	0.92	0.45	518
Cotton	0.90	0.81	0.85	21642
Kabkabeh	0.11	0.87	0.19	307
Millet	0.85	0.91	0.88	8306
Sesame	0.98	0.98	0.98	8598
Sorghum	0.85	0.74	0.79	17281
Soybean	0.82	0.90	0.86	8347
Sunflower	0.98	0.99	0.98	8508
Wheat	0.86	0.84	0.85	11209
accuracy			0.86	84716
macro avg	0.74	0.89	0.76	84716
weighted avg	0.88	0.86	0.87	84716



Discussion & Future Work

- Open Access Satellite Imagery is able to estimate the yield of various crops in Sudan.
- To improve the quality of the machine learning model estimates, precise field data with detailed meta descriptions are required.
- The total area grown for all crops under study in the state of El Gezira in 2022 is lower than what was declared by authorities in 2020.

1. Use an ensemble of Binary models to overcome the problem of undersampled classes.
2. Grow the training dataset by adding more samples from different states across the country.