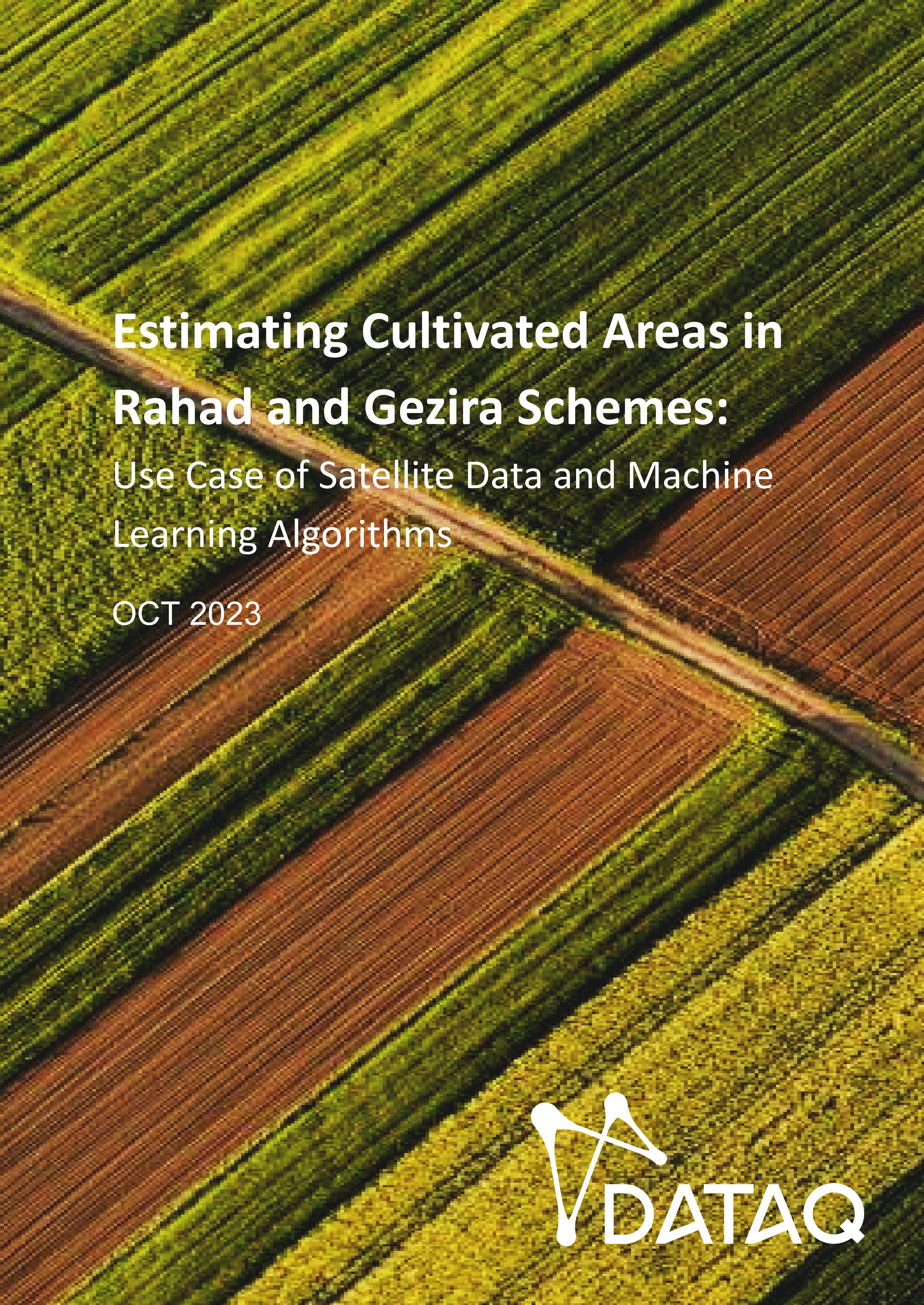

Estimating Cultivated Areas in Rahad and Gezira Schemes:

Use Case of Satellite Data and Machine Learning Algorithms

OCT 2023

DATAQ

Executive summary:

- This report examines the impact of conflict on areas sown in Sudan, specifically focusing on the Gezira state and its two major irrigation schemes, the Gezira Scheme and the Rahad Scheme. The study utilizes satellite imagery and machine learning classifiers to compare cultivated areas in 2023 with previous years.
- The background section provides an overview of the history of conflict in Sudan, highlighting the recent war between SAF and RSF that broke out on April 15, 2023. This sets the context for understanding the potential impact on agricultural activities in the region.
- The scope of this study focuses on the Gezira state, which encompasses two major irrigation schemes: the Gezira Scheme and the Rahad Scheme. These schemes play a crucial role in agricultural production in Sudan.
- The research methodology comprises three key phases: Data Generation, Model Training and Optimization, and Evaluation and Metrics. Sentinel-2 satellite imagery is utilized for data generation, while an XGBoost classifier is employed to identify whether an area is sown or not. The model's performance is evaluated using various metrics such as True Positive Rate (TPR), True Negative Rate (TNR), False Negative Rate (FNR), False Positive Rate (FPR), accuracy, and F1 score. (Should add more info about that+confidence intervals)
- The spatial and temporal resolutions of satellite imagery pose constraints on accurately identifying sown areas. These limitations should be considered when interpreting the results.
- The key findings reveal that only 37% of agricultural land in the Gezira Scheme was sown in 2023 compared to previous years' percentages of 47%, 41%, 53%, and 35% from 2022, 2021, 2020, and 2019 respectively. Similarly, in the Rahad Scheme, only 44% of agricultural land was sown compared to percentages of 53%, 77%, 69%, and 62% from previous years.
- In conclusion, this report highlights a significant decrease in sowing activities within both irrigation schemes due to ongoing conflict in Sudan. Future directions for research include improving data quality for more accurate analysis, developing models for rainfed areas to capture a broader picture of agricultural activity, as well as creating models that can identify specific crops grown within these regions.
- Overall, this study sheds light on how conflict impacts agricultural practices in Sudan's Gezira state. It provides valuable insights for policymakers and stakeholders involved in addressing food security challenges amidst ongoing conflicts.

Introduction:

Sudan, a country with a history of conflict, has experienced significant challenges recently, particularly in its agricultural sector. The ability to accurately monitor and assess the impact of conflict on areas sown in Sudan is of paramount importance for understanding the dynamics of these conflicts and devising effective strategies for recovery and development. This research paper aims to examine and compare the impact of conflict on agricultural areas in Sudan between the current year and previous years, utilizing satellite imagery and machine learning classifiers. Limited research exists that integrates satellite imagery and machine learning classifiers to provide a comprehensive analysis of the impact on areas sown. Modern scholars have explored the potential of remote sensing and computational methods for conflict analysis, but few studies have specifically focused on Sudan, particularly in a comparative context.

The timeliness of addressing this topic is crucial for multiple reasons. Firstly, Sudan continues to grapple with ongoing conflicts, necessitating a comprehensive understanding of the impact on agricultural areas. Secondly, advancements in satellite imagery and machine learning techniques have significantly improved our ability to monitor and analyze complex phenomena at various scales, making it an opportune moment to apply these tools to conflict-affected regions. Lastly, as the international community increasingly recognizes the importance of sustainable development and post-conflict reconstruction, this research will contribute to the existing body of knowledge and provide practical insights for policymakers, humanitarian organizations, and development practitioners.

To conduct this research, a comprehensive collection of satellite imagery will be acquired for multiple years, encompassing both conflict and non-conflict periods. These images will be processed and analyzed using machine learning classifiers to detect and classify areas sown in Sudan, particularly in Gezira State. The choice of machine learning classifiers was made due to their ability to handle complex and high-dimensional data, enabling the identification of patterns and changes over time.

Problem Statement:

The ongoing political conflict in Sudan has substantially affected the summer agricultural season. While anecdotal reports suggest a significant reduction in the sown area for the current season compared to previous years, there is a lack of quantifiable data to corroborate these claims. This study employs machine learning classifiers applied to Sentinel-2 satellite data to investigate and quantify the impact of conflict on areas sown in Sudan. By analyzing satellite imagery and utilizing advanced computational techniques, our research aims to provide a precise measurement of the agricultural land affected by conflict and uncover patterns, trends, and use cases that can offer insights into the impact of this conflict on food security and livelihoods in Sudan. We anticipate that these insights will be invaluable for policymakers, aid organizations, as well as other local and regional stakeholders in guiding their decision-making processes.

Objectives:

The aim of this paper is to:

- Develop a binary land cover classifier specifically designed to differentiate cultivated croplands.
- Utilize the predicted land cover classes (cultivated vs. uncultivated) to accurately estimate the area sown in the Gezira Scheme and Rahad Schemes for the current year, allowing for a comparison with previous years.
- Explore the use cases for the developed model and methodology, identifying how it can be applied beyond this study to address similar challenges in agricultural monitoring.

Background and Literature Review:

This literature review aims to explore and analyze the existing research on the impact of conflict on areas sown in Gezira State, focusing on the utilization of satellite imagery as a tool to detect changes in agricultural land use and compare them to previous years. By examining relevant studies, this review seeks to identify key findings, methodologies, limitations, and gaps in knowledge, while also highlighting the significance of satellite imagery in understanding agricultural dynamics in conflict-affected regions. It provides valuable insights into the consequences of conflict on food security and livelihoods.

1. Historical Context of Conflict in Sudan

Sudan's political instability has had an ongoing detrimental effect on agricultural productivity and food security. The various forms of conflict in the country's different regions have led to food shortages, agricultural productivity decline, and lowered contributions to GDP growth (Alhelo et al., 2023).

A destructive war broke out in Khartoum on April 15th, 2023, between the Sudanese Armed Forces (SAF) and the Rapid Support Forces (RSF), bringing to an end a three-year alliance in the Transitional Military Council (TMC).

Quantifying the immediate effects of an ongoing armed conflict is challenging due to the disruption of traditional data-gathering efforts and procedures. In the absence of conventional survey data, remote sensing tools and remote data collection methods hold significant potential to fill critical data gaps in fragile states (Dabalen et al., 2016; Hoogeveen and Pape, 2020) to inform policymakers, the private sector, and humanitarian aid workers of the potential impact of the conflict on agriculture, food value chains, and food insecurity.

There is a noticeable gap in the existing literature regarding studies that quantitatively assess the impact of the ongoing war in Sudan on the agricultural sector. While there have been numerous studies examining the broader socio-economic consequences of the conflict, there is a lack of specific research focusing on how this conflict has affected agriculture.

2. Study Area

Sudan's Gezira region (i.e. herein defined as Lat. 140 34' 24.5" N to Lat. 130 33' 49.94" N and Long. 340 34' 12.56E to Long. 330 29' 44.68"E) was selected for the study because of its large-scale dependence on both food crops and cash crops, diversity in the mode of production, availability of synoptic seasonally-matched satellite imagery, and the relative abundance of rainfall data for the study area.

Gezira State, located in central Sudan, holds immense significance in terms of agriculture and food security for both the country and the region. Known as the "breadbasket of Sudan" Gezira State has a long-standing history of agricultural excellence and plays a crucial role in sustaining the nation's food production.

The Gezira Scheme, established during the colonial era in the early 20th century, further enhances the state's agricultural significance. The scheme is one of the largest irrigation schemes in the region, comprising 880,000 hectares (ha) [10], and utilizes water from the Blue Nile to irrigate vast stretches of land. The scheme consumes 35% of the Nile's water in Sudan and produces half of the country's agricultural output. The scheme has transformed Gezira into a major agricultural hub, contributing significantly to Sudan's overall food production.

The Gezira state also holds another large scheme, the Rahad scheme, which is one of the major irrigation schemes in Sudan with a command area of 126,000 hectares (ha). It extends between latitudes 13° 31′ and 14° 25′ N, and longitudes 33° 31′ and 34° 32′ E. The area is 130 km long and its width ranges from 15 to 25 km over the flat alluvial plains of central Sudan. Topography is about 400 m above mean sea level, and slopes gently downwards from east to west, and from south to north. The only outcrops in the area are the El Fau mountains, at the foot of which the scheme headquarters are located [13]. One of the key features of the Rahad scheme is its extensive network of canals and channels that distribute water to farms across a vast area. The availability of water throughout the year has not only increased crop yields but also enabled farmers to diversify their production. Moreover, the Rahad scheme has had a positive impact on rural livelihoods by creating employment opportunities for local communities. The increased agricultural productivity has led to economic growth and improved living standards for many people living in the region.

Sudan is one of Africa's largest cotton producers, and Gezira accounts for a significant portion of this output. In addition to cotton, Gezira State cultivates various other crops such as sorghum, wheat, maize, groundnuts, sesame seeds, and vegetables [11]. These crops are vital not only for domestic consumption but also contribute to Sudan's export market. Furthermore, Gezira State's agricultural productivity plays a pivotal role in meeting the food demands of both Sudanese citizens and those in surrounding regions.

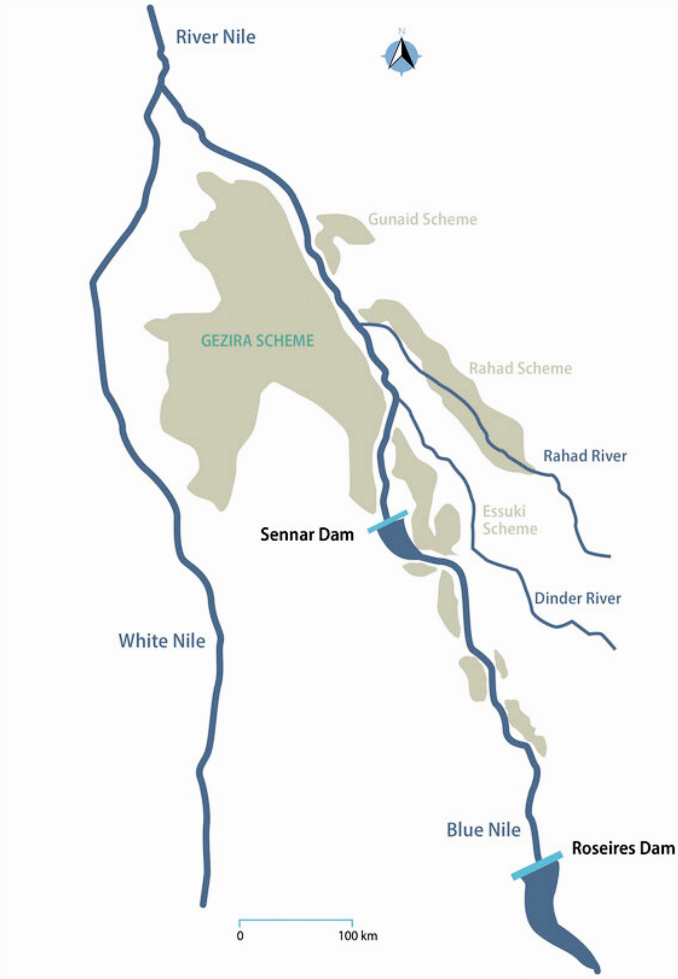

Figure 1. The Gezira and Rahad Schemes. Source: [**14**]

3. Satellite Imagery and its Applications in Agricultural Monitoring

Satellite imagery refers to the acquisition of images of the Earth's surface using satellites orbiting the planet. These satellites capture data in various wavelengths across the electromagnetic spectrum, including visible, infrared, and microwave.

Satellite imagery plays a crucial role in monitoring land use and agricultural activities due to its ability to provide large-scale and accurate data; it provides a synoptic view of extensive agricultural areas, enabling the monitoring of land use and crop conditions over large regions, and provides reliable and unbiased data, eliminating subjective biases associated with ground-based surveys. The images offer consistent information across time and space, enabling precise analysis of land use changes and agricultural practices.

Satellite imagery may have limitations in its spatial and temporal resolution; lower spatial resolution can hinder the identification of small-scale features and crop-specific characteristics. Additionally, limited temporal resolution may result in missed or delayed observations of short-duration events. Cloud cover and atmospheric conditions can obstruct satellite imagery, reducing data availability. The presence of clouds can limit the frequency and regularity of image acquisitions, potentially affecting the accuracy and reliability of agricultural monitoring. Analyzing and processing large volumes of satellite imagery data can be computationally intensive and time-consuming, specialized skills and tools are required to extract meaningful information from the raw satellite data.

4. Methodological Approaches for Assessing Areas Sown Using Satellite Imagery

Assessing areas sown using satellite imagery involves several methodological approaches that can provide accurate and reliable results. These approaches typically include image acquisition, preprocessing, classification techniques, and validation.

Sentinel-2 is a newer satellite program compared to other programs such as Landsat, it has many advantages including its high spatial resolution, multispectral bands, and frequent revisit times, Sentinel-2 data has immense potential for accurate and timely land cover and land use mapping. The literature also shows that the use of Sentinel-2 data produces high accuracies (>80%) with machine-learning classifiers such as support vector machine (SVM) and Random forest (RF) [1]. However, there are some challenges associated with using Sentinel-2 data, such as atmospheric effects and data preprocessing [6].

Before analyzing satellite imagery for sown areas, it is crucial to preprocess the images to enhance their quality and remove any noise or distortions. This may involve radiometric calibration, atmospheric correction, and geometric correction [2,3]. Once the preprocessed images are available, various classification techniques can be applied to identify areas that have been sown. The most commonly used method is supervised classification, which involves training a classifier using labeled samples of sown and non-sown areas. The classifier then assigns class labels to pixels in the image based on their spectral characteristics. Classification algorithms such as Random Forest (RF), Support Vector Machines (SVM), and Artificial Neural Networks were studied extensively, RF was compared with classification trees, which are also known as decision trees, by Rodriguez-Galiano et al.(2012) who found that RF produced a high accuracy of 92%, thereby outperforming classification trees, compared to SVM and Xgboost [4].

A novel approach for multi-temporal land cover classification using sequential recurrent encoders was proposed in [5]. The authors leverage the temporal information inherent in remote sensing time series data to improve land cover classification accuracy. The sequential recurrent encoders capture the temporal dependencies and patterns in the data, thereby enhancing the classification performance. The study demonstrates the effectiveness of the proposed method using Sentinel-2 data and compares it with other traditional classifiers. The results indicate that the sequential recurrent encoders outperform conventional approaches, highlighting the potential of temporal information for land cover classification. Another powerful and popular machine learning algorithm used for classification tasks is Extreme Gradient Boosting (XGBoost) [7], its ability to handle imbalanced data, capture non-linear relationships, provide feature importance, and scale well makes it a suitable choice for classifying whether the land is sown or not. However, it is important to note that there might not be a single algorithm that universally performs best across all scenarios, as it depends on the specific characteristics of the study area and the objectives of the classification task.

After applying classification techniques, it is essential to validate the accuracy of the results obtained. Validation can be done through ground truth data collection or reference data from other sources such as field surveys or existing agricultural databases. Several validation metrics like overall accuracy, producer's accuracy, user's accuracy, and kappa coefficient [9] can be calculated to assess the reliability of the classified sown areas. In [8], results of the Kappa coefficient show that all the classifiers have a similar accuracy level with minor variation, but the RF algorithm has the highest accuracy of 0.89 and the MD algorithm (parametric classifier) has the lowest accuracy of 0.82. Overall, combining these methodological approaches helps in accurately assessing areas sown using satellite imagery and provides valuable information for agricultural monitoring, land management, and decision-making processes.

Methodology:

This research methodology comprises a structured approach that sequentially encompasses three key phases: Data Generation, Model Training and Optimization, and Evaluation and Metrics. These phases collectively form the foundation for our comprehensive analysis of the impact of conflict on agricultural land use in the Gaziera and Rahad Schemes.

In the upcoming sections, we will detail the specific processes and techniques employed in each phase. The Data Generation phase lays the groundwork by ensuring data accuracy and relevance, followed by Model Training and Optimization, which focuses on refining the binary land cover classifier. The Evaluation and Metrics phase assesses the methodology's performance and its ability to provide valuable insights.

1. Data:

In this section, we discuss the key steps in our data collection and processing. We begin with the acquisition and preprocessing of Sentinel-2 satellite imagery data, highlighting its suitability for accurate land cover analysis. We then detail our image selection process, focusing on the Gezira summer season. Next, we explain our label acquisition approach, which involves manual labels generation from historical satellite imagery. Finally, we address model evaluation, emphasizing measures taken in the sample selection to ensure reliable and generalizable predictions.

Data Source and Preprocessing:

In our efforts to acquire the most suitable satellite imagery data, we carefully consider the selection of the source and the preprocessing techniques applied. We opted for Sentinel-2 data primarily due to its numerous advantages, making it an ideal choice for our study. The accessibility of Sentinel-2 imagery, coupled with its 10-meter spatial resolution and a revisit interval of just 5 days, facilitates timely and accurate land cover analysis. Our imagery was obtained in the Level-1C top-of-atmosphere (TOA) reflectance processing format, which means it had undergone radiometric and geometric corrections but not atmospheric correction [17].

Each Sentinel-2 image consists of visible, near-infrared, and shortwave infrared sensors, encompassing 13 spectral bands. To standardize our dataset and enhance analysis, we upsampled all bands to a uniform 10-meter resolution using bi-linear interpolation. Finally, we used the Fraction of Green Vegetation Cover Index (FCover) as our 14th band. The FCover index quantifies the spatial extent of vegetation within our study area, which is important information that can improve the accuracy of the classifier [4].

Dates of Samples Selection:

Temporal and spatial considerations play a vital role in the selection of imagery for our land cover classification. Given our focus on the Al Gezira summer season, our temporal window is limited to the months between June and September. In order to accurately depict how cultivated areas change throughout the season, we decided to choose one image from each month. Then, we chose the images from each month that had the fewest clouds present. This procedure was repeated from 2019 through 2023. This selection process results in four Sentinel-2 scenes per year, acquired specifically from the Tile 33VXG. These images were sourced from the Copernicus Open Access Hub Sentinel-2A satellite on September 1, 2023 [18].

Label Acquisition:

The foundation of our supervised learning approach rests on the quality and diversity of our training data labels. Labels for remote sensing data can be collected directly from field surveys or through automatic derivation from existing land cover maps. However, our resources limit our ability to collect data through surveys, and the use of existing land cover maps limits the validity of our estimates due to error propagation [19].

Thus, we opted for the approach of manual label generation from historical satellite imagery. This method involves establishing sample points or polygons based on the manual interpretation of these images. We generate the labels by meticulously reviewing imagery in the Sentinel Hub EO Browser [20] in the first week of October for each of the years 2019 to 2023. Utilizing the clear visual distinction between cultivated and uncultivated fields at the beginning of the harvest season, we generate polygons across the entire state, representing both classes and covering a 20-square-kilometer area for each class annually. These resulting polygons serve as pixel-wise masks for the previously acquired full-scene imagery and are then processed and converted to tabular format for classification.

In Figure 1, we show the spatial distribution of the collected labels from all years across the Gaziera State and specifically our study area; the Gaziera and Rahad Schemes. This spatial diversity is critical to ensuring our collected samples are good proxies for the target areas without overfitting our classifier. Additionally, Table 1 describes the processed labeled data for each year, showing the area and number of pixels (rows) for each class per year.

Year	Label	Area KM2	Number of Pixels
2019	Cultivated	17.71	176,988
	Uncultivated	23.69	217,728
2020	Cultivated	19.89	198,650
	Uncultivated	19.89	189,473

2021	Cultivated	11.17	111,656
	Uncultivated	11.92	119,122
2022	Cultivated	22.55	225,254
	Uncultivated	20.40	104,887
2023	Cultivated	16.33	158,362
	Uncultivated	14.69	140,654
Total	-	178.24	1,642,774

Table 1: Area and Pixel Distribution by Class and Year for Processed Labeled Data

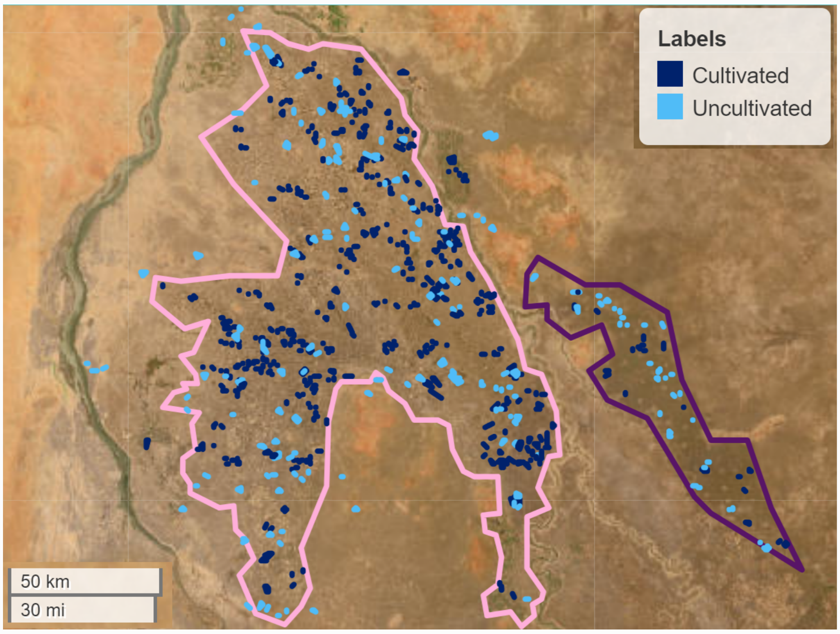

Figure 2: Spatial Distribution of Collected Labels in the Gezira and Rahad Schemes

Evaluation Considerations in Sample Selection:

Robust model evaluation is essential to ensuring the reliability of geospatial predictions and actions based on these predictions. Nevertheless, evaluating models in the context of geospatial modeling through conventional machine-learning techniques often has issues related to the strong clustering of samples or discrepancies between the training and prediction data [21]. These limitations in the evaluation process can potentially result in models that overfit the training data, thus producing overly optimistic results.

To address these limitations, our methodology takes deliberate measures when collecting samples and conducting evaluations. We ensure that the collected samples are representative both in terms of location and time and serve as accurate proxies for the real-world use cases of the model. This is achieved by disseminating the samples across all of the Gazira State over the course of five years.

Moreover, the model is trained on only the data from 2019 and 2020, which is further divided into 70% training and 30% testing splits. The data from the years 2021, 2022, and 2023, in addition to the 30% test split, is used to evaluate the performance of the model. This diversity helps in preserving the independence and identically distributed (IID) assumptions necessary for rigorous model assessment. These precautions collectively enhance the reliability and generalizability of our geospatial predictions.

3. Classifier:

In this section, we discuss our choice of the XGBoost classifier for cropland classification in satellite data. We outline the data preparation steps, training process, and hyperparameter tuning.

Classifier Selection and Rationale:

Due to its great compatibility for the challenges provided by the satellite imaging domain, XGBoost has been selected as the classifier of choice for the task of classifying cultivated croplands in satellite data [16]. XGBoost is the optimal model for this job thanks to a number of vital attributes. It offers robustness to noise, supports ensemble learning and gradient boosting, parallelization for faster training, interpretable results, and strong community support for remote sensing tasks.

Data Preparation and Preprocessing

Extensive data preparation was undertaken to streamline our dataset while preserving its representativeness. This involved reducing the original 6,800,168 samples to 219,324, transforming integer labels to strings, reordering columns alphabetically, strategic data sampling, addressing class imbalance, and narrowing the dataset to 2019-2020 for training and 2021,2022 and 2023 for testing. Label encoding and meta-column pruning were applied to ensure efficient XGBoost classifier training and evaluation.

Training the XGBoost Classifier:

The dataset is then split into training and testing sets using a 70-30 split, where 70% of the data is used for training the classifier, and 30% is reserved for testing purposes. We initialize the XGBoost classifier with carefully selected hyperparameters, including parameters related to tree depth, learning rate, and regularization, which have been fine-tuned through optimization techniques such as Hyperopt and RandomizedSearchCV [15,16]. These hyperparameters play a crucial role in the model's performance and generalization capability.

With the classifier initialized, we proceed to train it on the training data, leveraging the gradient boosting framework. The XGBoost algorithm constructs an ensemble of decision trees, iteratively refining the model's predictions. This ensemble approach enhances prediction accuracy and mitigates overfitting, a critical consideration in the complex landscape of remote sensing tasks.

Hyperparameter Tuning

In our pursuit of accurately classifying cultivated and uncultivated croplands in Sudan's satellite imagery, hyperparameter tuning was performed alongside 5-fold cross-validation using optimization techniques including Hyperopt and RandomizedSearchCV.

The systematic process of hyperparameter optimization was conducted within the context of 5-fold cross-validation. We explored key hyperparameters such as learning_rate, max_depth, min_child_weight, subsample, colsample_bytree, and n_estimators. Optimal parameter combinations were selected based on cross-validated performance, fine-tuning the XGBoost classifier for improved accuracy.

After the training, and hyperparameter tuning, we assessed the classifier's performance on testing data, providing crucial insights into its accuracy, precision, recall, and overall classification performance. These metrics are pivotal for evaluating the model's suitability and guiding further refinements.

3. Model Metrics and Evaluation:

In the Evaluation and Metrics section, we delve into the rigorous assessment of our methodology. Here, we detail the process through which we determine the sown area in the targeted regions and calculate confidence intervals. The subsequent sections will provide a comprehensive overview of these vital components, ensuring a robust evaluation of our methodology's performance and the accuracy of our findings.

We commence by presenting the key metrics for the classifier we've developed. These metrics are essential for evaluating the accuracy and effectiveness of our binary land cover classifier. Following the classifier metrics, we describe our sampling strategy and our calculations for the confidence levels and intervals for the areas sown.

Classifier Metrics:

Add a small paragraph about generalization

At the foundational level of our analysis, we evaluate our classifier's performance using key metrics, including true positive rates (TPR) and true negative rates (TNR). These metrics provide valuable insights into the model's ability to classify "cultivated" and "uncultivated" regions accurately. It's important to note that we consider TPR and TNR as deterministic aspects of the classifier's performance in our assessments.

One standout feature of our work is the emphasis on generalization. To test the model's capability to handle entirely new, previously unseen satellite imagery, we utilize previously unobserved test data, particularly for the years 2021, 2022, and 2023. This approach can be likened to a real-world stress test where the model encounters imagery it has never seen before. This analysis is crucial for determining the model's flexibility, a critical aspect for real-world applications, particularly in the context of cultivated and uncultivated land classification.

Furthermore, we've taken a meticulous approach by breaking down TPR and TNR values for each year separately. This unique feature of our analysis allows us to observe minor trends and variations in the model's performance over time. By doing so, we gain a deeper understanding of how well the model adapts to changing yearly conditions. We can pinpoint the years in which the model excels and those in which it encounters challenges. This level of granularity in our evaluation enables us to assess whether the model's performance remains consistent or exhibits variations when exposed to changing environmental factors. It's a critical step in evaluating the model's real-world applicability, especially in the domain of cultivated and uncultivated land classification.

.....

Sampling Strategy and Confidence Intervals:

We will take a random sample of N Pixels from each of the target areas. We will use the classifier to output a count of the predicted true (cultivated) pixels for each sampled population. This of course will have false positives and false negatives which will skew the generated count, resulting in a random variable for the areas cultivated. We will then use the ratio of cultivated vs. uncultivated pixels to calculate the areas cultivated in each targeted area and year.

In Appendix 1 we show that for a sample size *N* and a known true positive rate *tpr*[1] and true negative rate *TNR*, the lower CI_L and upper confidence CI_U intervals for the number of pixels cultivated *X* can be calculated as follows;

$$CI_L(X) = \max C \cdot (1 - tpr) + \max C \cdot z \cdot \sqrt{\frac{tpr \cdot (1 - tpr)}{\max C}} - \left(minC \cdot (1 - tnr) + minC \cdot z \cdot \sqrt{\frac{tnr \cdot (1 - tnr)}{minC}} \right)$$

[1] We are assuming that the TPR and the TNR are certain for our classifier.

$$CL_U(X) = (N - minC) \cdot (1 - tnr) + (N - \min C) \cdot z \cdot \sqrt{\frac{tnr \cdot (1 - tnr)}{N - \min C}} - \left((N - \max C) \cdot (1 - tpr) + (N - \max C) \cdot z \cdot \sqrt{\frac{tpr \cdot (1 - tpr)}{(N - \max C)}} \right)$$

Where *z* is the standard score, *maxP* is the largest number of cultivated pixels cultivated in the sampled pixels, and *minP* is the smallest number of cultivated pixels in the sampled pixels.

For 95% tpr and TNR and 2000 samples, the CL and CL is…

Results and discussion:

Intro:

Classifier Results:

Table showing (**TPR, FPR, TNR, FNR, Accuracy, F1 Score**)

Year	%TPR	%FPR	%TNR	%FNR	%ACC	%F1
2019	99.99	0.03	99.97	0.01	99.98	99.98
2020	99.72	0.25	99.75	0.28	99.74	99.73
2021	90.01	28.37	71.63	9.99	80.52	79.15
2022	99.83	1.65	98.35	0.17	99.36	98.98
2023	97.82	14.00	86.00	2.18	92.26	91.27
Mean	97.47	8.86	91.14	2.53	94.37	93.82
Var	14.55	122.23	122.23	14.55	56.30	64.24

Table 2: Classifier's Performance Metrics (2019-2023)

Discussion of tables:

Figure 3: The Gazira Scheme Inference Sample (2023)

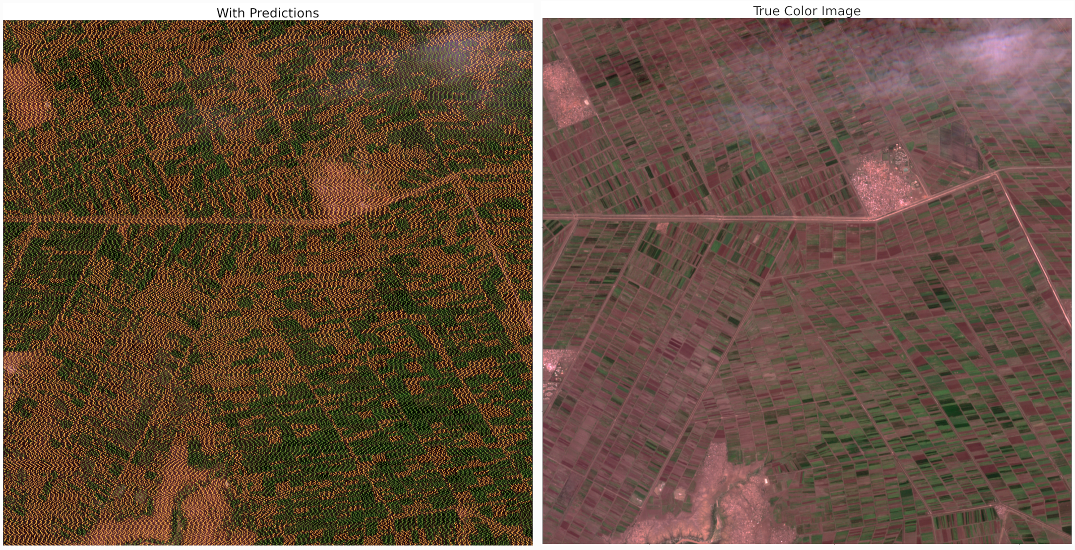

Figure 4: Rahad Scheme Inference Sample (2023)

Areas Cultivated:

How we sampled and used the classifier to predict ratios.
Add some info about the polygons
Gaziera and Rahad areas sown 2023 (glass half full)

Scheme	Area KM2	Year	% Cultivated	Total Pixels
Gezira Scheme	12285	2019	0.35	5381
		2020	0.53	5278
		2021	0.41	5346
		2022	0.47	5320
		2023	0.37	5429
Rahad Scheme	2230	2019	0.62	3278
		2020	0.69	3310
		2021	0.77	3224
		2022	0.53	3268
		2023	0.44	3336

Table 3: Percentage of Cultivated Land for Gezira and Rahad Schemes (2019-2023)

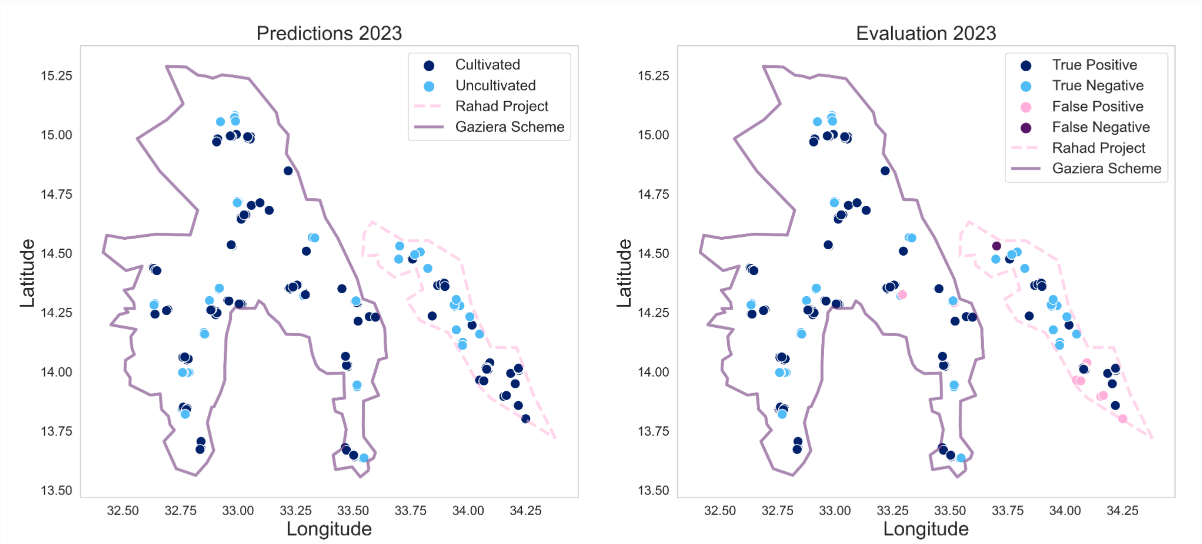

Figure 5:

Visuals showing targeted regions cultivated areas, compared with previous years.

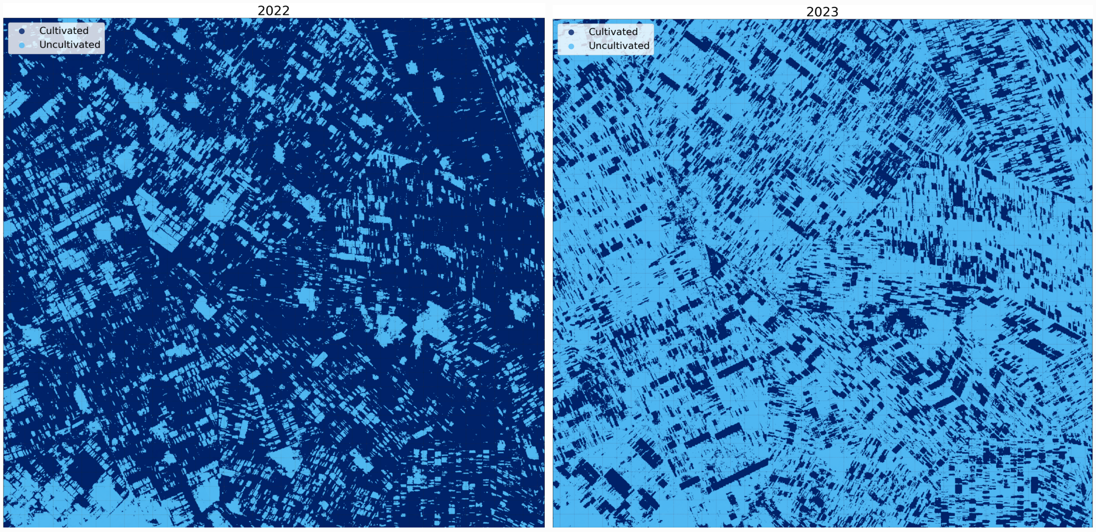

Figure 6: 2022 vs 2023 Sample Inference (Gazira Scheme)

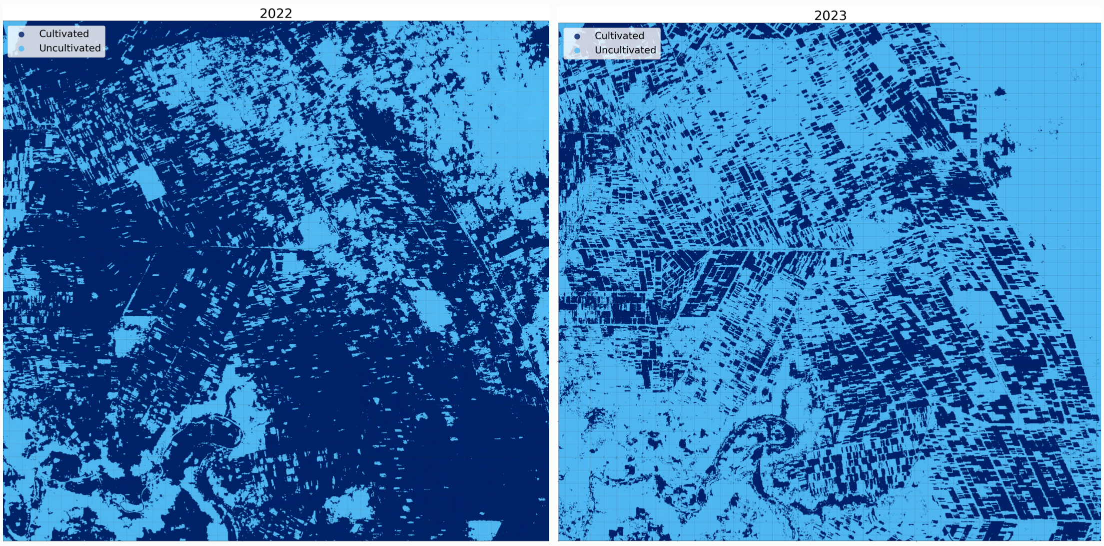

Figure 7: 2022 vs 2023 Sample Inference (Rahad Scheme)

Comparison to secondary resources.

Use Cases & Future Directions:

Improving data quality:

- The labeled data generated was generated visually, by looking at true color images and visually identifying the cultivated areas, vs the uncultivated areas.

- Sentinel hub free data, 10m pixel resolution.
- Gezira only

Develop a model for Rainfed areas:

- The labeled data collected is limited to Irrigated fields only.
- Less organized and harder to tell between Natural bushes vs cultivated crops.
- Difficulty generating labeled data; rainfed fields are more tricky to distinguish using satellite images.

Develop a model to identify which crops have been grown.

Resources:

[1] Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y. and Ranagalage, M., 2020. Sentinel-2 data for land cover/use mapping: A review. Remote Sensing, 12(14), p.2291.
[2] Dave, C.P., Joshi, R. and Srivastava, S.S., 2015. A survey on geometric correction of satellite imagery. International Journal of Computer Applications, 116(12).
[3] Sowmya, D.R., Deepa Shenoy, P. and Venugopal, K.R., 2017. Remote sensing satellite image processing techniques for image classification: a comprehensive survey. International Journal of Computer Applications, 161(11), pp.24-37.
[4] Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience & Remote Sensing, 57(1), pp.1-20.
[5] Rußwurm, M. and Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. ISPRS International Journal of Geo-Information, 7(4), p.129.
[6] Mandanici, E. and Bitelli, G., 2016. Preliminary comparison of sentinel-2 and Landsat 8 imagery for combined use. Remote Sensing, 8(12), p.1014.
[7] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I. and Zhou, T., 2015. Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), pp.1-4.
[8] Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.A. and Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. Remote Sensing, 12(7), p.1135.
[9] Stehman, S., 1996. Estimating the kappa coefficient and its variance under stratified random sampling. Photogrammetric engineering and remote sensing, 62(4), pp.401-407.
[10] Ahmed, A., 2009, January. Gezira Scheme irrigation system performance after 80 years of operation. In Proceedings of the International Conference on Water, Environment, Energy and Society, New Delhi, India (pp. 12-16).
[11] Osman, I.S., Schultz, B., Suryadi, F.X., Mohamed, Y. and Osman, A.K., 2011. Improving the operation and maintenance for better sediment and water management in Gezira Scheme, Sudan. In Proceedings of the ICID 21st International Congress on Irrigation and Drainage, Tehran, Iran, 15-23 October, 2011 (pp. 51-62).
[12] JOUR, Goelnitz, Anna, Al-Saidi, Mohammad, 2020/11/01, 106396, Too big to handle, too important to abandon: Reforming Sudan's Gezira scheme, 241, 10.1016/j.agwat.2020.106396, Agricultural Water Management.

[13] Hamid, S.H., Mohamed, A.A. and Mohamed, Y.A., 2011. Towards a performance-oriented management for large-scale irrigation systems: case study, Rahad scheme, Sudan. Irrigation and drainage, 60(1), pp.20-34.
[14] XGBoost. (n.d.). XGBoost Documentation. Retrieved [Oct. 14, 2023], from https://xgboost.readthedocs.io/en/latest/index.html
[15] J. Bergstra, D. Yamins, and D. D. Cox, "Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms," in Proceedings of the 12th Python in Science Conference, 2013, pp. 13-20
[16] Scikit-learn. (n.d.). RandomizedSearchCV. Retrieved [accessed Oct. 14, 2023] from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html
[17] Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. and Meygret, A., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote sensing of Environment, 120, pp.25-36.
[18] Documentation of sentinelhub Python package — Sentinel Hub 3.9.1 documentation (sentinelhub-py.readthedocs.io)
[19] Waśniewski, A., Hościło, A. and Aune-Lundberg, L., 2023. The impact of selection of reference samples and DEM on the accuracy of land cover classification based on Sentinel-2 data. Remote Sensing Applications: Society and Environment, 32, p.101035.
[20] Sentinel Hub. (n.d.). Sentinel Hub EO Browser. Retrieved from https://apps.sentinel-hub.com/
[21] Karasiak, N., Dejoux, J.F., Monteil, C. and Sheeren, D., 2022. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. Machine Learning, 111(7), pp.2715-2740.

Appendix 1: Calculation of Confidence Intervals

To calculate the confidence intervals for a sample size *N* and a known true positive rate *tpr* and true negative rate *TNR*, we start by working out confidence bounds for the number of false positives and false negatives separately.

Part 1: False Negatives Bounds:

For each true pixel, there are two possible outcomes, either it's classified as true or as false.
This can be thought of as a Bernoulli trial[2], with the probability of success being the probability of a false negative, i.e. *(1-tpr)*.

Given this setup and using the binomial distribution formula we can work out the expected number of false negatives, let's call it *E(FN)*:

$$E_{max}(FN) = \max C \cdot (1 - tpr)$$

Where we define maxC to be a conservative estimate of the largest possible number of true pixels in the sample. This estimate will be based on secondary knowledge of the target region.

The confidence interval for *FN* for a binomial distribution with a large number of true pixels maxC can be calculated as follows *CI(FN)* is:

$$CI_{max}(FN) = \max C \cdot z \cdot \sqrt{\frac{tpr(1 - tpr)}{\max C}}$$

Hence for a given confidence level we can be sure that the number of false positives can not be larger than $E_{\max}(FN) + CI_{\max}(FN)$.

Similarly, we can also calculate the expected value and the confidence levels for false positives as follows:

$$E_{min}(FN) = minC \cdot (1 - tpr)$$

$$CI_{min}(FN) = minC \cdot z \cdot \sqrt{\frac{tpr(1 - tpr)}{minC}}$$

Where minC is defined as a conservative estimate of the lowest possible number of true pixels in the sample. This estimate will be based on secondary knowledge of the targeted region.

Similarly, for a given confidence level we can be sure that the number of false positives can not be less than $E_{min}(FN) + CI_{min}(FN)$.

---

[2] This is a safe assumption as the prediction of each pixel is independent of the previous predictions.

Part 2: False Positive Bounds:

By making the same Arguments for false pixels we get the following formulas for the false positives;

$$E_{max}(FP) = (N - \min C) \cdot (1 - tnr)$$

$$CI_{max}(FP) = (N - minC) \cdot z \cdot \sqrt{\frac{tnr(1 - tnr)}{(N - minC)}}$$

$$E_{min}(FP) = (N - maxC) \cdot (1 - tnr)$$

$$CI_{min}(FP) = (N - maxC) \cdot z \cdot \sqrt{\frac{tnr(1 - tnr)}{(N - maxC)}}$$

Part 3: Putting it all together:

We now show how the confidence for the number of cultivated pixels *X* in the sample *N* are defined by the equations provided above.

For a given confidence level, we first note that our estimated value for the variable X can only decrease from the real value by a value that is smaller than the number of false positives. Similarly, the estimated value for the variable X can only decrease from the real value by a value that is smaller than the number of false positives.

Hence, for a given confidence level. the lower confidence interval for the variable X, is defined as the upper confidence bound for the false positives minus the lower confidence bounds for the false negatives which is equal to: