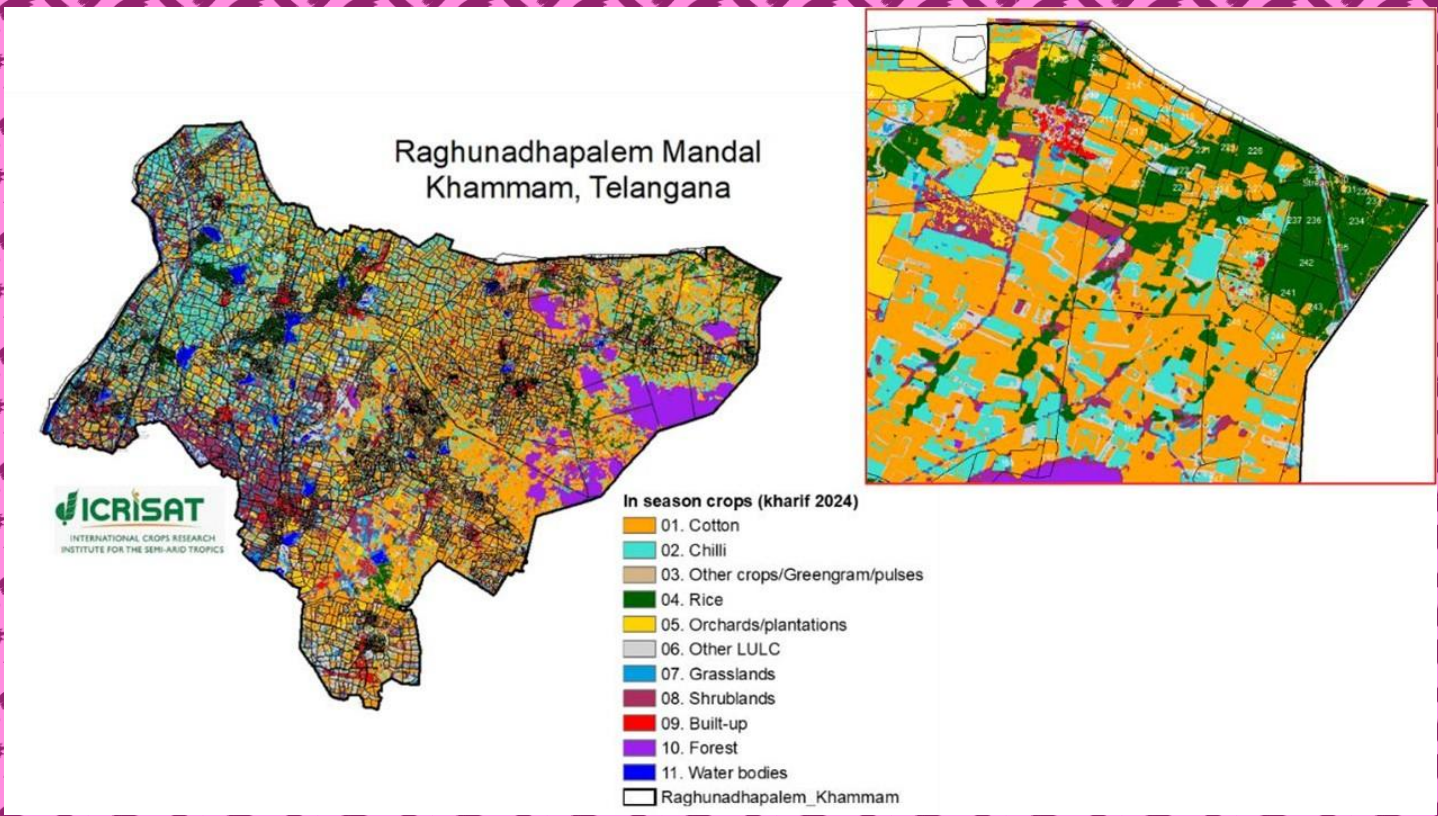
In-season crop monitoring using Machine learning Algorithms

9th September 2025

Conference Paper presented on "On the Road – Bali by Planet Labs"



Contents

In-season crop monitoring using Machine learning Algorithms

1. Introduction

This study was undertaken as a pilot exercise in two mandals: Raghunadhapalem (Khammam district) and Sirikonda (Nizamabad district), to develop and test an analytics workflow for in-season crop monitoring using multi-temporal Earth Observation data and machine-learning methods.

Under CGIAR's Digital Transformation agenda, Area of Work 1 (AOW1: Data Ecosystem) focuses on building scalable, searchable, and interoperable data systems that enable reuse across research, policy, and operational decision-making. This work aligns with AoW01 HLO 1.2.4 by demonstrating how Earth Observation data, analytics, and visualization can be combined into a single, scalable analytics pipeline for in-season crop monitoring.

This report documents a pilot initiative conducted during the Kharif 2024 season to establish a scalable and searchable geospatial data ecosystem for in-season crop and land use–land cover (LULC) monitoring in Telangana. The objective was not only to generate crop and LULC maps, but to operationalize a reusable data pipeline that integrates Earth observation (EO) time series, field data, and administrative records into a coherent, queryable system.

Furthermore, a dedicated web portal has been developed for parcel-level, evidence-based monitoring of crop sowing and harvest dates. By linking this system with agricultural schemes, it enhances transparency, minimizes the risk of fraud, and enables more informed planning for procurement, input allocation, and timely policy interventions. Collectively, the pilot establishes a technology-driven framework that can be scaled across Telangana, strengthening crop estimation, improving resource efficiency, and ensuring that agricultural incentives reach the intended beneficiaries.

2. Data Ecosystem Design and Data Inputs

This study was designed within the framework of a scalable geospatial data ecosystem, where Earth Observation data, ancillary datasets, and analytics outputs are integrated through a structured processing pipeline. The architecture supports repeatable analytics, temporal querying, and dissemination of results through visualization interfaces, consistent with AoW01 objectives.

The ecosystem follows a layered approach, beginning with standardized data ingestion, followed by analytics and classification, and culminating in user-facing visualization and access services. The datasets and processing steps described below form the foundational inputs to this architecture.

2.1 Overview of Data Inputs

The data inputs were selected to support in-season crop monitoring, land use–land cover classification, and temporal analysis of crop dynamics during the Kharif 2024 season.

All datasets were processed within a cloud-computing environment to ensure consistency, scalability, and reproducibility. Spatial reference systems, temporal intervals, and naming conventions were standardized so that the same workflow can be reused across seasons and geographic regions.

2.2 Data Sources and Data Types

Satellite Data

- Sentinel-2 MSI: Multi-temporal optical imagery used for vegetation analysis, crop type classification, and phenological assessment. Images were composited at monthly or fortnightly intervals to balance temporal detail and noise reduction.

Ancillary and Reference Data

- Mandal-level administrative boundaries for spatial aggregation and reporting
- Crop calendar information relevant to the study region
- Field observations collected using the ICROPS mobile application, used to guide training sample selection and interpretation

Derived Analytics Outputs

- Vegetation indices (e.g., NDVI and related metrics)
- Temporal spectral signatures representing crop phenology
- Classified land use–land cover and crop type layers
- In-season crop extent maps generated at multiple time steps

These datasets collectively support the analytics pipeline described in the subsequent section and enable integration with visualization and decision-support tools.

3. Analytics Pipeline and Step Processing

The in-season crop monitoring workflow was implemented as a stepwise analytics pipeline that integrates data ingestion, pre-processing, feature extraction, machine-learning-based classification, and visualization [1,2,3] (Figure 1). The pipeline is modular in design, allowing individual components to be updated or reused independently, supporting scalability and long-term operational use.

3.1. Satellite Data Processing in Google Earth Engine

Multi-temporal Sentinel-2 imagery was ingested and processed within Google Earth Engine. Pre-processing steps included cloud masking, atmospheric correction, and spatial sub-setting to the study area. Multi-spectral bands were stacked to create temporally consistent datasets.

- Cloud masking and atmospheric correction
- Stacking multi-spectral bands over the crop growing season
- Generating maximum NDVI composites at fortnightly intervals to capture crop phenological patterns.

This processed dataset forms the basis for further analysis and classification. Vegetation indices and spectral features were derived from the pre-processed imagery. Temporal compositing was applied at monthly or fortnightly intervals to capture crop phenological dynamics while reducing noise. These features form the core analytics inputs for crop classification and in-season monitoring.

3.2. Classification and Analytics

The classification stage integrates clustering, spectral signature development, and spectral matching within a unified analytics workflow. An initial unsupervised classification is applied to group pixels with similar spectral behaviour. This helps identify broad land cover categories and guides the selection of representative training samples. These pre-clusters also help isolate noise and identify spectral confusion zones that may need further ground validation. In cases where crops exhibit similar spectral and phenological profiles, such as rice vs. wetlands or maize vs. sugarcane, classification confusion is more likely. To address this, additional indices like LSWI and EVI, crop calendar information, and region-specific phenology profiles will be used to improve class separability and reduce misclassification.

3.2.1. Development of Spectral Signatures

With support from field data collected using the ICROPS mobile application, spectral signatures are developed for each crop type (e.g., rice, maize) and for other land use/land cover (LULC) classes (e.g., water bodies, fallow, settlements). The process involves:

- Extracting pixel values for ground-verified locations
- Analysing NDVI and other spectral indices over time

- Generating class-wise temporal spectral profiles

This step ensures that each class has a distinct and biologically meaningful spectral identity. To improve adaptability across districts, region-specific classification models and phenology-based crop libraries are being developed. These profiles help capture local crop behaviour and allow the classification approach to be fine-tuned for different agro-climatic zones.

3.2.2. Spectral Matching Techniques

Using the developed signatures, spectral matching techniques are employed to identify and classify each pixel into its most likely crop or LULC class. This step is refined using:

- Field-collected training data
- Secondary datasets such as crop calendars, administrative records, and historical land cover maps
- Expert knowledge to adjust thresholds or resolve class confusion (e.g., rice vs. wetlands)

Classification is performed using algorithms available in Google Earth Engine, such as random forest, support vector machines, or minimum distance classifiers, depending on the data and region. The classification outputs are generated as time-stamped spatial layers, enabling temporal querying and analytics across the cropping season.

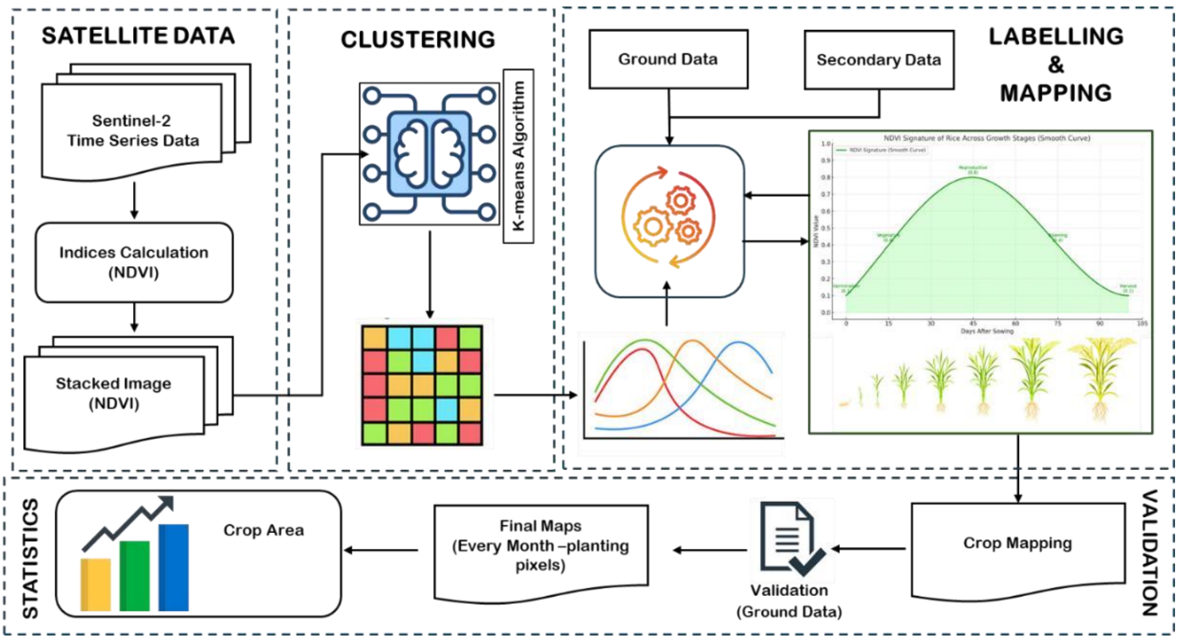


Figure 1: Analytics pipeline for in-season crop monitoring

3.2.3. Output Generation and Visualization

The final outputs include classified crop and LULC maps, in-season crop extent layers, and summary statistics. These outputs are integrated with a web-based visualization platform to enable interactive exploration, supporting search, query, and interpretation by end users. The study mapped crop areas during the growing season to enable timely and informed decision-making (Figure 2). Cotton sowing was recorded at 2,674 ha by June 15. The area expanded to 3,616 ha by June 25 and reached 7,005 ha by August 5. Such in-season monitoring provides early signals on sowing progress, helping planners anticipate input requirements, track adoption trends, and align interventions with the pace of crop expansion.

Temporal spectral signatures derived from satellite data clearly depict crop growth stages, including establishment, peak vegetative phase, senescence, and harvest. By analysing these phenology curves, variations in sowing dates and harvesting windows across fields can be identified (Figure 3). This technical insight directly translates into decision-usefulness: seed procurement can be aligned with actual sowing patterns, while production procurement and supply chain planning can be synchronized with expected harvest timelines.

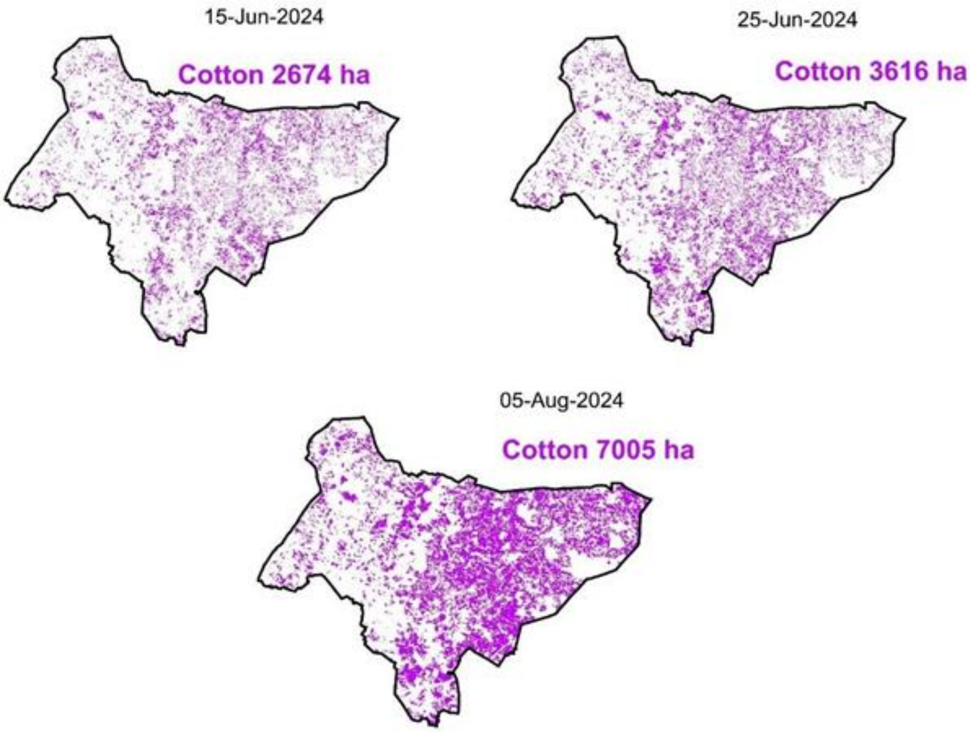


Figure 2:Temporal crop classification

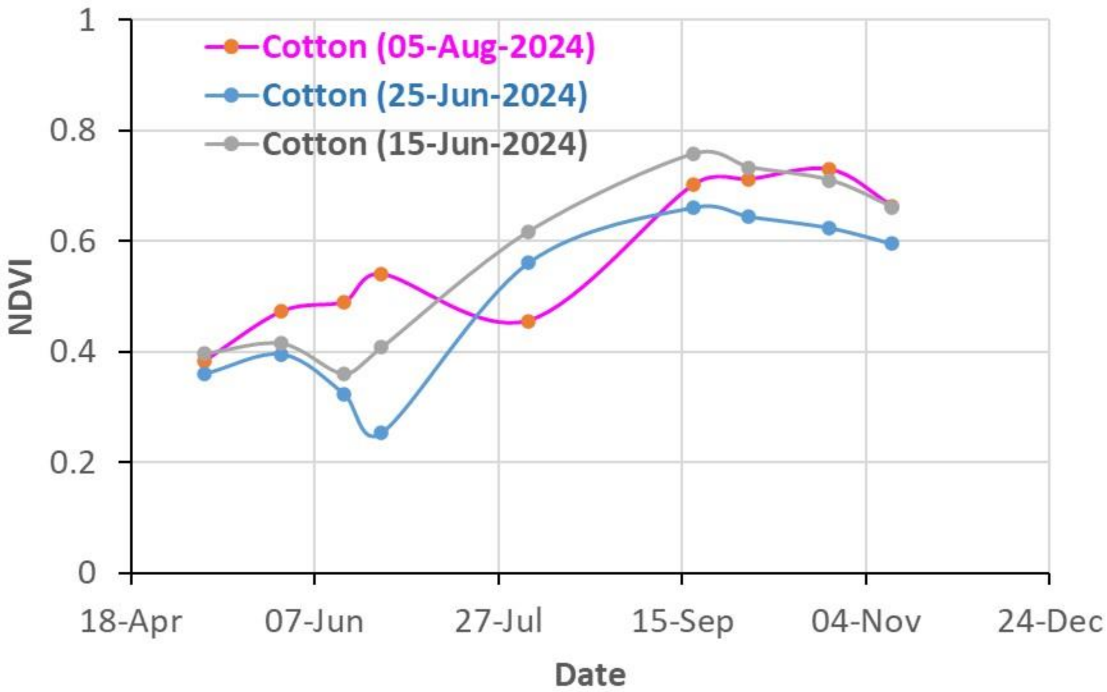


Figure 3:Temporal spectral signatures

4. LULC and Crop type Mapping

4.1 Raghunadhapalem Mandal

The LULC analysis covered a total area of 47,292 acres in Raghunadhapalem (Figure 4, Table 1). Cotton emerged as the dominant crop, occupying 17,309 acres, followed by chilli with 6,106 acres. Other crops such as greengram and pulses accounted for 625 acres, while rice covered 1,270 acres. Orchards and plantations extended over 3,254 acres. Non-cropped areas were also significant, with 5,053 acres under fallows, 4,223 acres classified as grasslands, and 5,693 acres under shrublands. Built-up land comprised 1,001 acres, forest areas covered 2,058 acres, and water bodies or wetlands occupied 699 acres. This distribution highlights the prominence of cotton cultivation, while also reflecting the considerable share of non-crop categories that influence land use planning and resource management in the region.

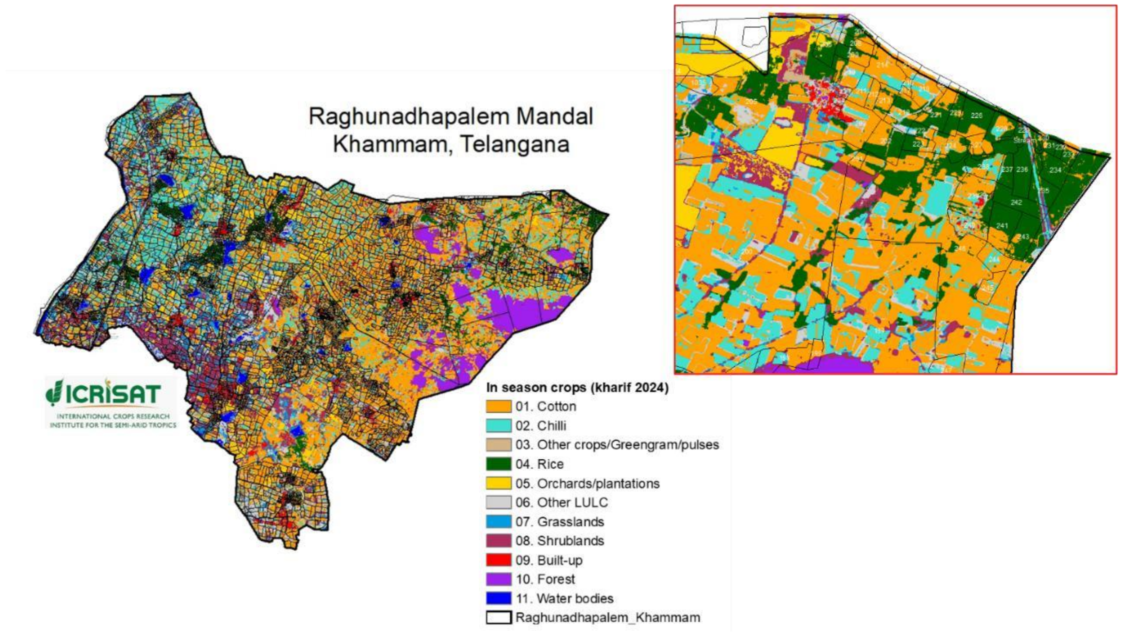


Figure 4:LULC and crop type map of Raghunadhapalem

Table 1:Area Statistics of Raghunadhapalem mandal

LULC#	Area (Acre)
01. Cotton	17309
02. Chilli	6106
03. Other crops/Greengram/pulses	625
04. Rice	1270
05. Orchards/plantations	3254
06. Fallows	5053
07. Grasslands	4223
08. Shrublands	5693
09. Built-up	1001
10. Forest	2058
11. Water bodies /wetlands	699
Total area	47292

4.2 Sirikonda Mandal

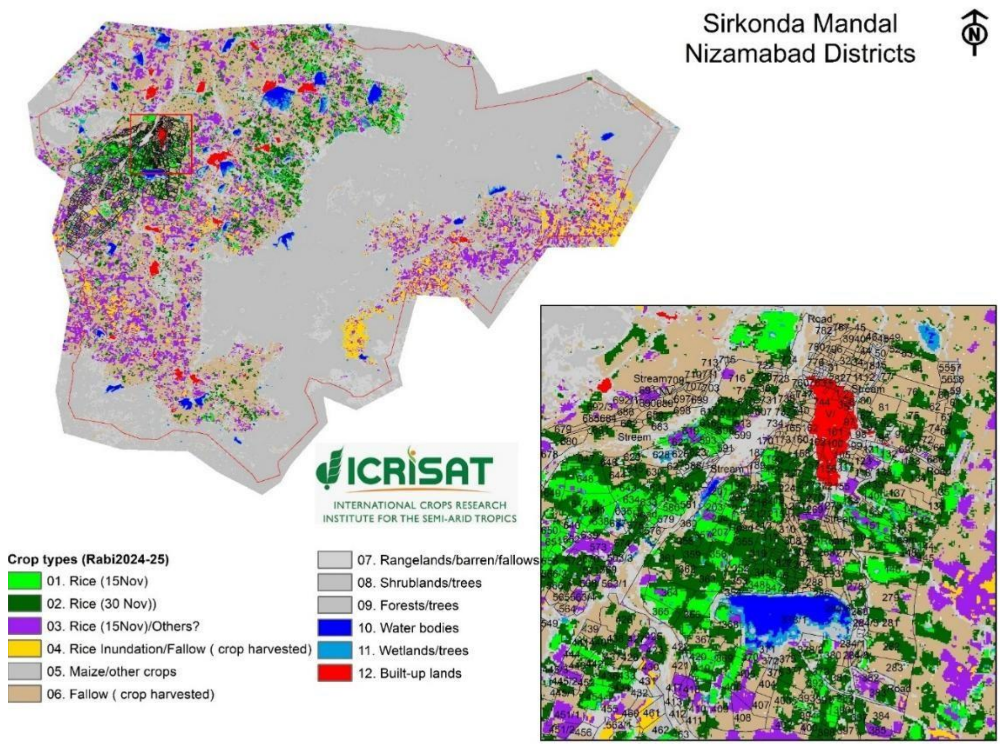


Figure 5:LULC and crop type map of Sirikonda mandal

The LULC assessment covered a total of 67,214 acres in Sirikonda (Figure 5, Table 2). Rice cultivation was mapped in different stages, with 980 acres sown by November 15, expanding to 3,566 acres by November 30, and an additional 6,298 acres where rice was followed by other crops. Areas categorized as rice inundation or harvested fallows accounted for 2,382 acres. Maize and other crops together covered 90 acres. A large portion of land, 11,824 acres, was under harvested fallows, while rangelands and fallows added another 8,704 acres. Shrublands and tree cover occupied 9,432 acres, and forests contributed the largest share with 22,002 acres. Water bodies and wetlands accounted for 718 acres and 576 acres respectively, while built-up lands extended over 641 acres. This distribution underscores the dominance of forest and tree cover, along with substantial areas of fallows and rangelands, while rice continues to be the primary cultivated crop with varying temporal patterns during the season.

Table 2: Area Statistics of Sirikonda mandal

LULC	Area (Acre)
01. Rice (15Nov)	980
02. Rice (30 Nov))	3566

03. Rice (15Nov)/Others Crops	6298
04. Rice Inundation/Fallow (crop harvested)	2382
05. Maize/other crops	90
06. Fallow (crop harvested)	11824
07. Rangelands/fallows	8704
08. Shrublands/trees	9432
09. Forests/trees	22002
10. Water bodies	718
11. Wetlands/trees	576
12. Built-up lands	641
Total area	67214

4.3 Comparison of LULC Statistics with Rythu Bharosa Records

In Raghunadhapalem mandal, the total geographical area is 47,292 acres, of which 33,618 acres are cultivable, and 13,675 acres are non-cultivable. However, the extent of land for which payments were made under the Rythu Bharosa (RB) scheme stands at 38,988 acres, an overestimation of 5,370 acres or about 16 percent above the mapped cultivable area. In Sirikonda mandal, the total area is 67,213 acres. Cultivable land accounts for 25,140 acres, while non-cultivable land extends over 33,369 acres. RB records show payments made for 28,800 acres, exceeding the cultivable extent by 3,660 acres, which translates to an overestimation of about 15 percent.

Table 3:Area Statistics comparison with RB

Class	Raghunadhapalem (Acres)	Sirikonda (Acres)
Total Area	47,292	67,213
Cultivable Land	33,618	25,140
Non-Cultivable Land	13,675	33,369
RB Paid Land Extent	38,988	28,800
Difference (RB vs Cultivable)	5,370	3,660
Overestimation (%)	**16**	**15**

This comparison clearly shows a consistent pattern where the RB-recorded extent surpasses the actual cultivable land mapped through LULC. The overestimation, ranging from 15 to 16 percent, indicates the need for geospatial validation of scheme data to ensure more accurate targeting of agricultural incentives.

4.4 Repository Links and Data Access

The analytics workflows, derived datasets, and visualization outputs generated under this study are intended for dissemination through institutional repositories and web-based platforms.

The following web applications provide access to crop and LULC analytics outputs:

Land Use Land Cover Viewer:

https://ee-my-pytest.projects.earthengine.app/view/telanganacrops

Cropland Monitoring Viewer:

https://ee-my-pytest.projects.earthengine.app/view/tgscrops

Users can interactively explore spatial layers, zoom to parcel or mandal levels, and examine crop status across different time periods.

4.5 Data Download and Reuse

Derived spatial datasets can be exported from the analytics platform in standard geospatial formats (GeoTIFF, vector layers) using predefined export configurations. These datasets are suitable for integration into GIS software and downstream analysis by partner institutions. These access mechanisms demonstrate how analytics outputs are made discoverable and reusable within a scalable data ecosystem, consistent with AoW01 HLO 1.2.4.

References

1. Gumma, M. K., Thenkabail, P. S., Panjala, P., Teluguntla, P., Yamano, T., & Mohammed, I. (2022). Multiple agricultural cropland products of South Asia developed using Landsat-8 30 m and MODIS 250 m data using machine learning on the Google Earth Engine (GEE) cloud and spectral matching techniques (SMTs) in support of food and water security. GIScience & Remote Sensing, 59(1), 1048–1077. https://doi.org/10.1080/15481603.2022.2088651
2. Gumma, M. K., Tummala, K., Dixit, S., Collivignarelli, F., Holecz, F., Kolli, R. N., & Whitbread, A. M. (2022). Crop type identification and spatial mapping using Sentinel-2 satellite data with focus on field-level information. Geocarto International, 37(7), 1833–1849. https://doi.org/10.1080/10106049.2020.1805029
3. Gumma, M.K.; Panjala, P.; Dubey, S.K.; Ray, D.K.; Murthy, C.S.; Kadiyala, D.M.; Mohammed, I.; Takashi, Y. Spatial Distribution of Cropping Systems in South Asia Using Time-Series Satellite Data Enriched with Ground Data. Remote Sens. 2024, 16, 2733. https://doi.org/10.3390/rs16152733

A pilot analytics pipeline for scalable, in-season crop monitoring and data-driven decision support

This report presents a pilot implementation of a scalable analytics pipeline for in-season crop monitoring using Earth Observation data and machine-learning methods. Developed under CGIAR's Digital Transformation agenda, the workflow integrates data ingestion, analytics, and visualization to support transparent, timely, and evidence-based agricultural decision-making. The approach is designed for reuse and scaling across regions and seasons.

CGIAR DIGITAL TRANSFORMATION