Badan Pusat Statistik

Indonesia Experience on

MIXED METHOD 2024

Earth Observation Data for
Paddy Crop Phenology Identification

Jakarta, February, 3rd 2025

Mixed Method Team
METHODOLOGY TEAM

BACKGROUND

CURRENT

Harvested Area

Area Sampling Frame Survey (ASF)
for monitoring the phenology

×

Yield

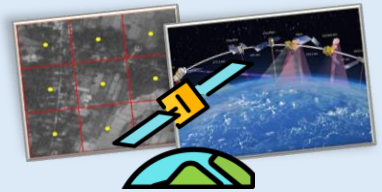
Crop-cutting survey
Estimate the productivity based on ASF samples

=

Rice Production

Dry Harvested Paddy (GKP)
(Ton)

PROPOSE

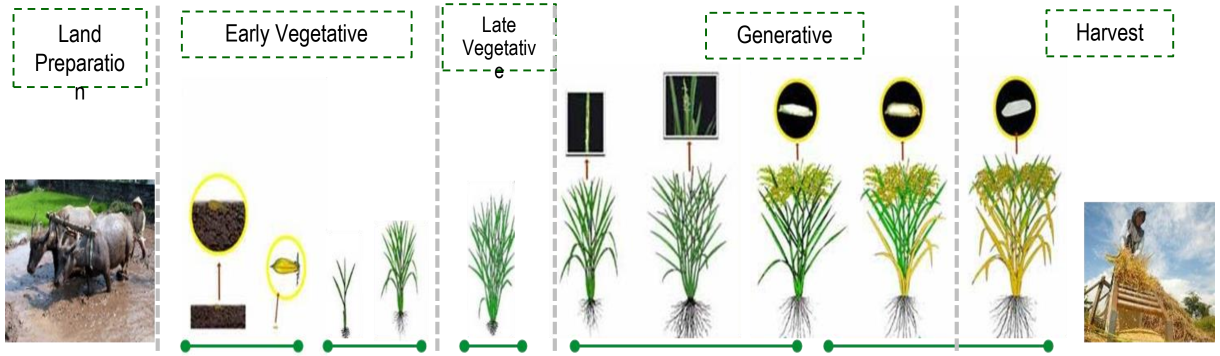


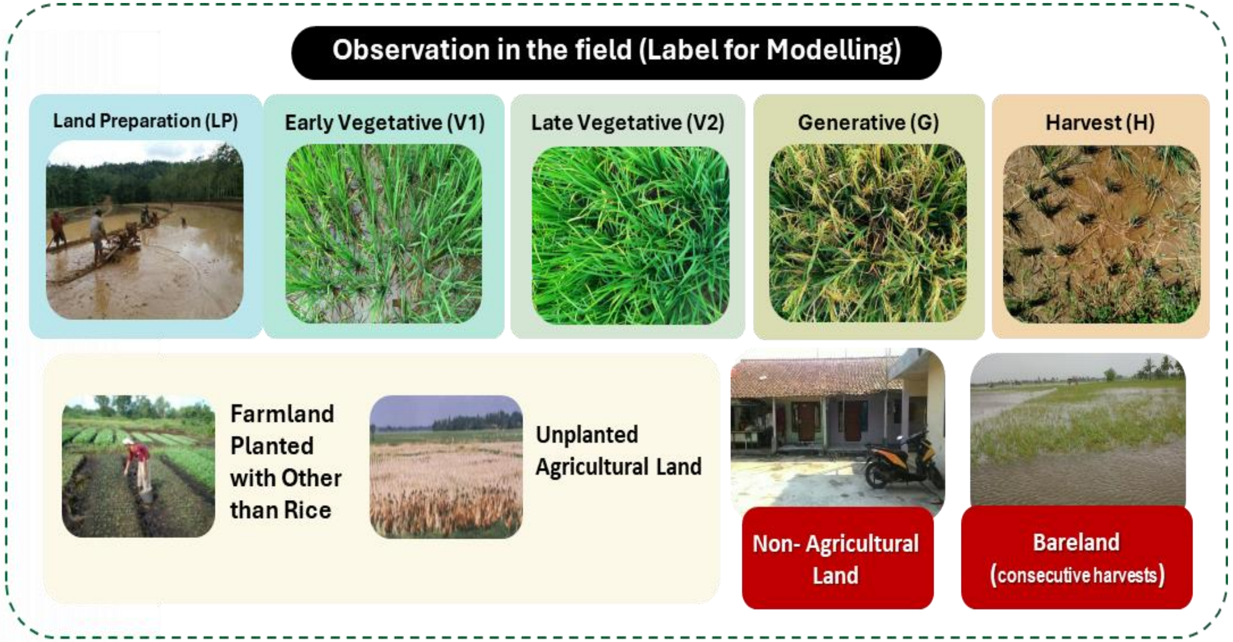
Mixed Method

- Integrating two approaches: **the Area Sample Frame (ASF)** and **Earth Observation (EO) data** to **address the challenges of agricultural modernization by utilizing big data for official statistics**, with purposes such as:
  - Optimization of **Cost-Effective Data Collection**
  - **Addressing non-response issues** in remote areas
  - Reducing the **surveyor burden** (for instance: permission to access the rice field.)
- The Mixed Method **does not entirely replace ASF** activities but can reduce the number of samples of ASF in areas where the Mixed Method can be applied.
- The use of **standardized methodologies and business processes**, along with a **comprehensive evaluation**, is necessary to produce high-quality data that aligns with the **10 Principles of Official Statistics**.

SATELLITE DATA AND CROP PADDY PHASE

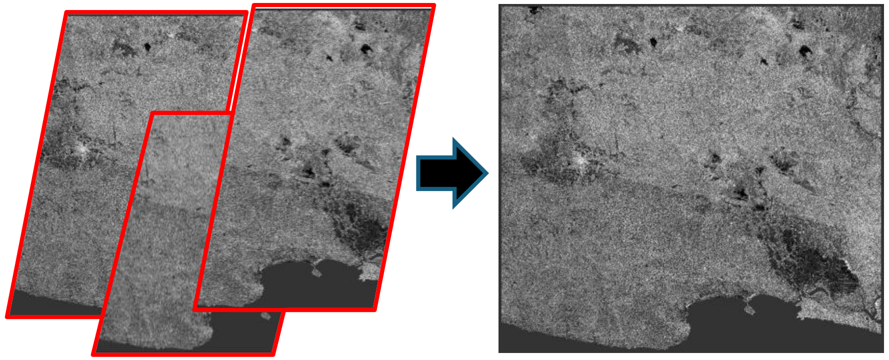
Paddy Phenology based on ASF data

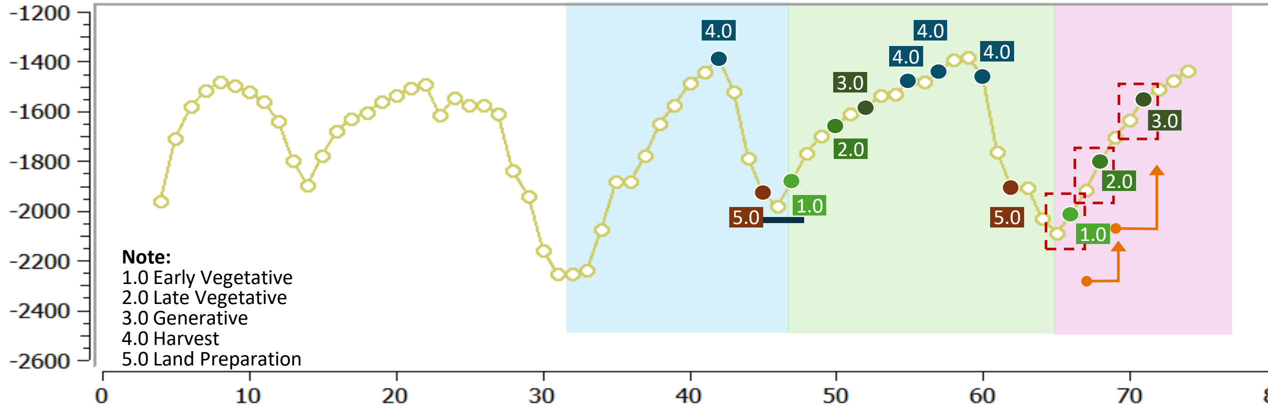




Satellite Imagery Data

The satellite data used corresponds to the field data collection for ASF.





Spectral Values

We learn **the temporal pattern of spectral values from satellite imagery** to identify each phenology stage based on field conditions (from ASF data). Unique temporal patterns are observed for each phenology stage during each planting period.

METHODOLOGY (1)

I. Satellite Imagery Acquisition and Preprocessing

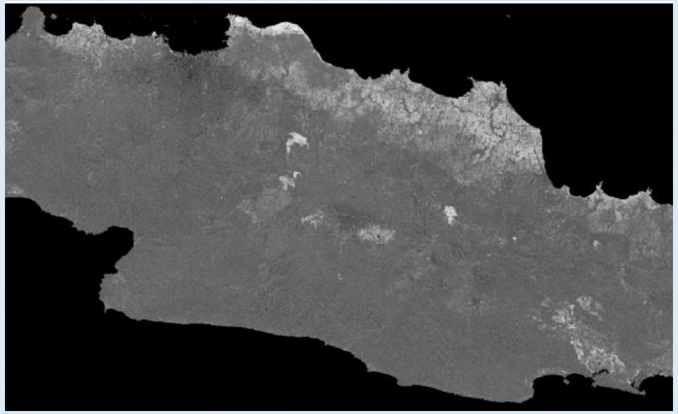
1. Intersecting the 10 selected provinces with the **Military Grid Reference System (MGRS).**

2. Download Sentinel-1 data using the **ASF Database API** for the selected boundary and time period.

3. **Preprocessing using SNAP,** including *Apply Orbit File, Thermal Noise Removal, Border Noise Removal, Calibration, Speckle Filtering, Range Doppler Terrain correction, and Conversion to dB.*

4. **For each grid, mosaicking is performed** with the following steps: *Subseting, Band selection (VH), and Mosaicking.*

Sample Results



- Sentinel-1 focuses on the **VH band**
- **Temporal resolution:** 12-day intervals
- **Period:** January 2021 – December 2023
- **Spatial resolution**: 20 meters

II. Spatial Clustering

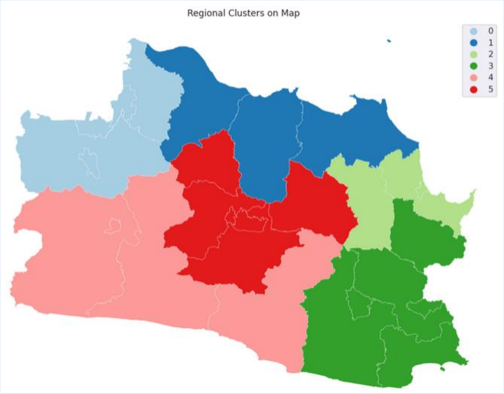
1. Retrieval of **Regional Characteristic Data** for Clustering Analysis as stated on the table

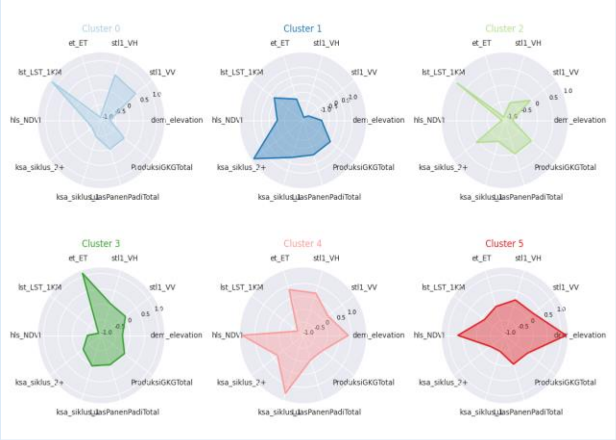
2. Clustering data using **SKATER** *(Spatial 'K'luster Analysis by Tree Edge Removal).*

3. **Profiling** the clustered data results.

Variable	Source
Elevation	DEM
Number of Planting in a year	ASF data
VV+VH	Sentinel-1
NDVI	Sentinel-2 HLS
ET & LST	MODIS
Harvest Area & Paddy Productivity	Statistics Report

Sample Results





METHODOLOGY (2)

III. Data Preprocessing

1 Prepare ASF data by **cloning each KSA observation point** within subsegments to reach 25 points. Filter data for labels included in the Mixed Method model.

2 **Extract band values** for each pixel within each KSA subsegment.

3 Impute data using the **Whittaker-Eilers** approach for handling missing data.

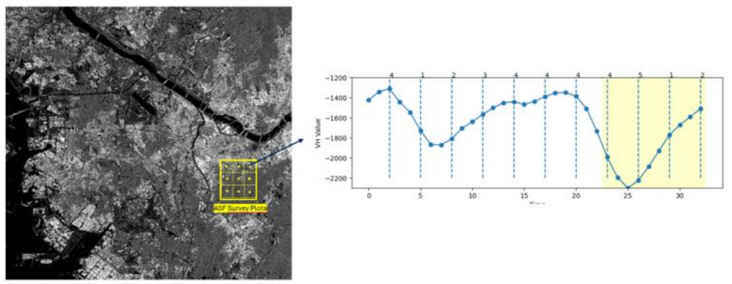
4 Filter the data by examining the **pixel variance** within each subsegment,

5 Filter the data using the **Self-Organizing Maps (SOM)** approach for each cluster, resulting in a more representative training data subset.

Label For Mixed Method

Label	Description
0	Bareland; if Harvest is consecutive
1	Vegetative 1
2	Vegetative 2
3	Generative
4	Harvest
5	Land Preparation
6	Non Paddy and Non Vegetation

Sample Results



id	nth	VH_t0	...	VH_t31	cluster	class

IV. Modelling

1 **Split the data** with a 70% training and 30% testing data.

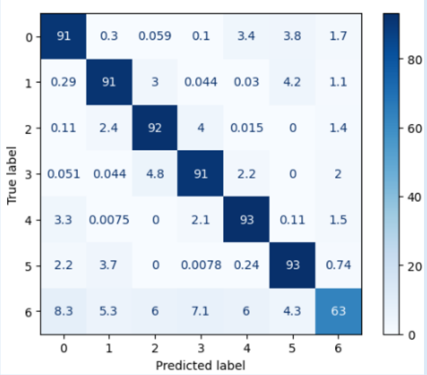
2 **Tune the XGBoost model** for each cluster. **Perform XGBoost modeling** with the parameters obtained from tuning, using 3-fold cross-validation to ensure valid results.

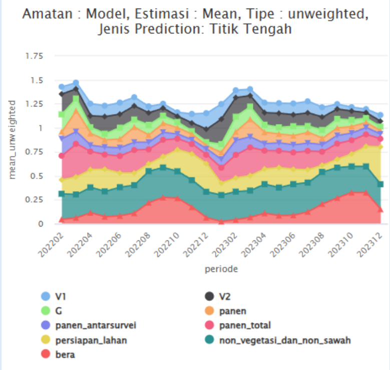
3 **Evaluate the modeling results** by examining several metrics, including: Accuracy, F1-macro, F1-micro, ROC-AUC, Relative accuracy

4 **Perform paddy phase predictions** for the period from Jan 2022 to Dec 2023.

5 **Estimating harvested area and** compared with ASF estimates .

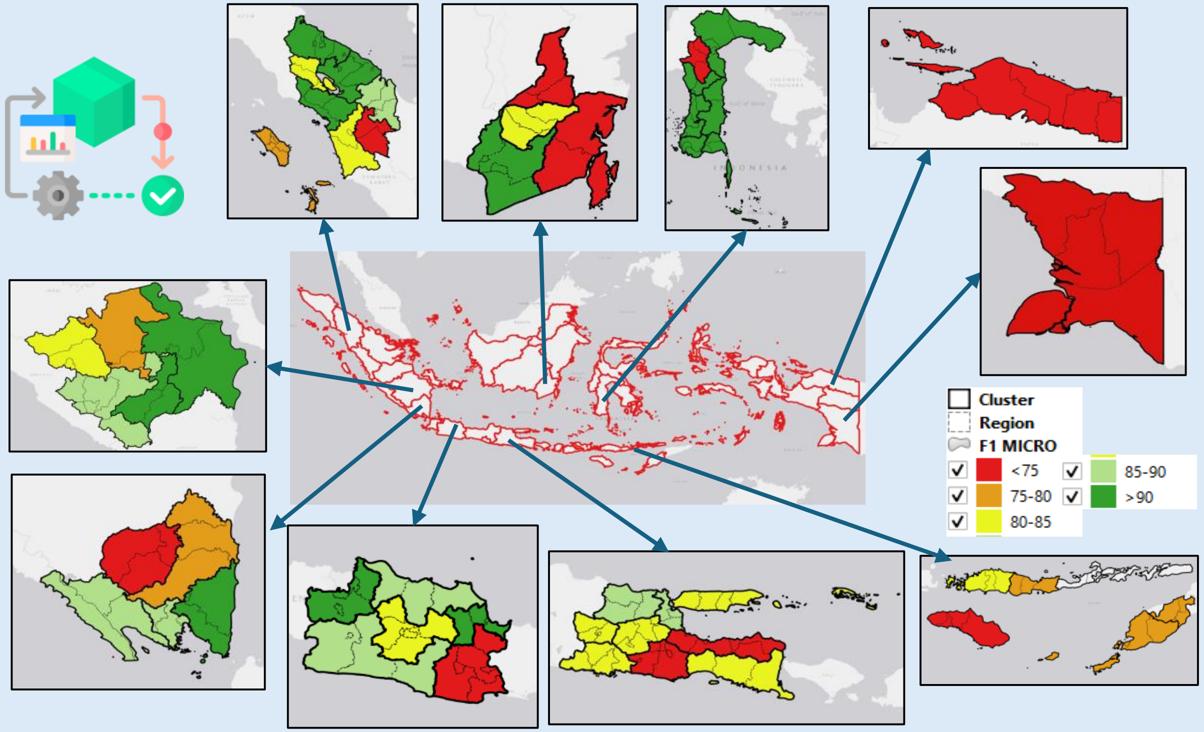
Sample Results





RESULTS AND FINDINGS

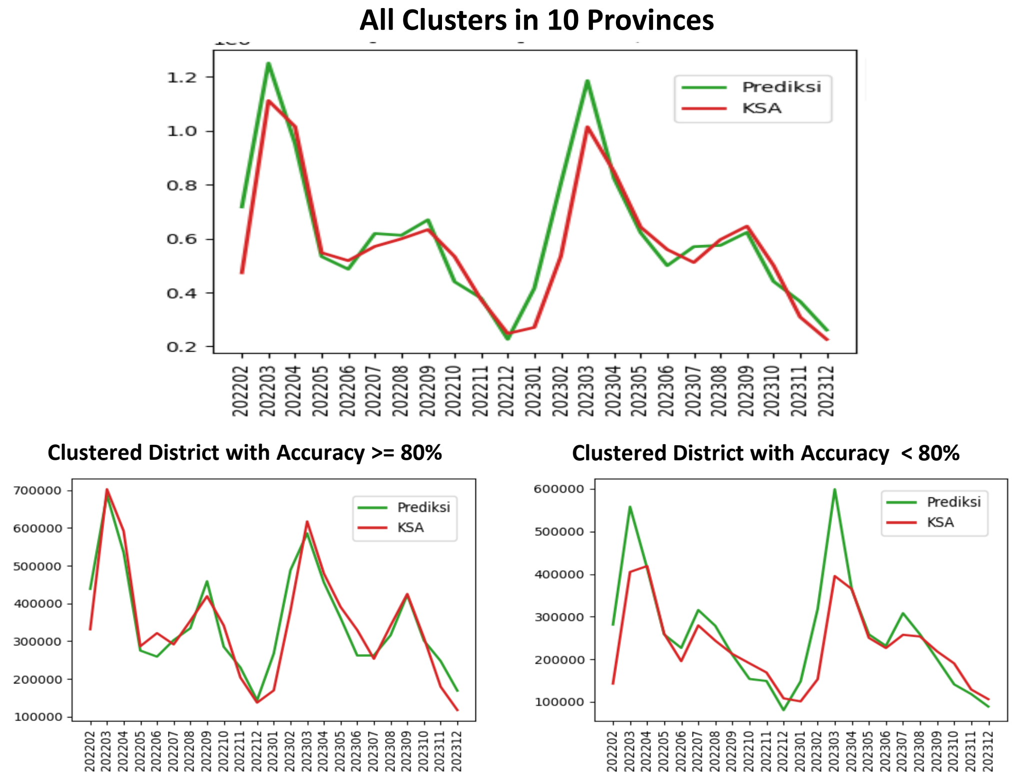
Accuracy for Each Cluster



Initial identification shows that the **model performs poorly** in regions with the following characteristics:

- **Mountains** (Terracing)
- **Small rice cultivation areas**
- **Dry rice fields** (NTT)
- **Swamp rice fields** (South Kalimantan)
- **Irregular planting patterns** (Papua)
- **Abandoned rice fields after harvest** (South Papua)

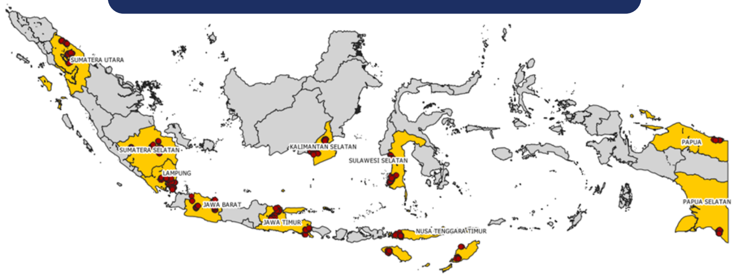
Accuracy for Each Cluster



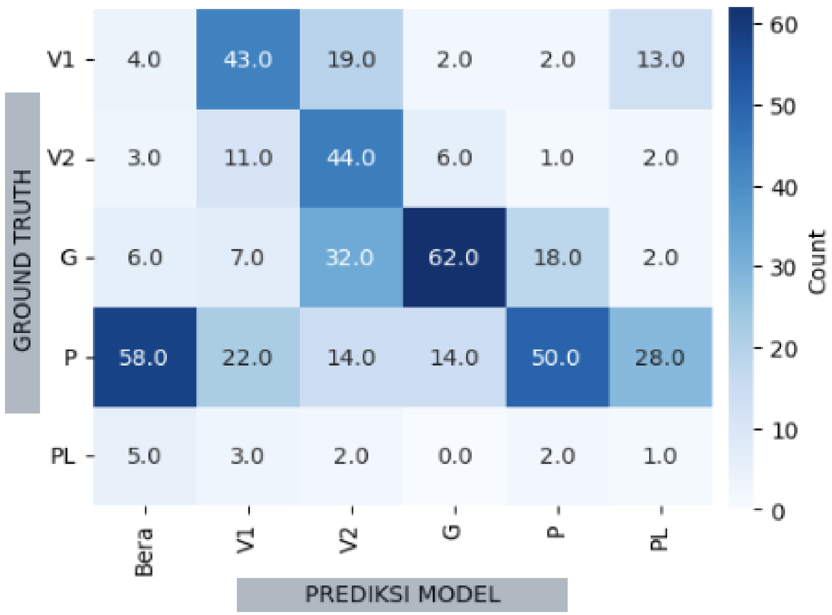
- In general, the estimated total harvest area using the KSA method and RS-Modeling method show **similar patterns and figures.**
- However, for regions within clusters with **accuracy <80%,** the RS-Modeling estimates tend to **overestimate**, especially during peak harvest periods.

GROUND TRUTH

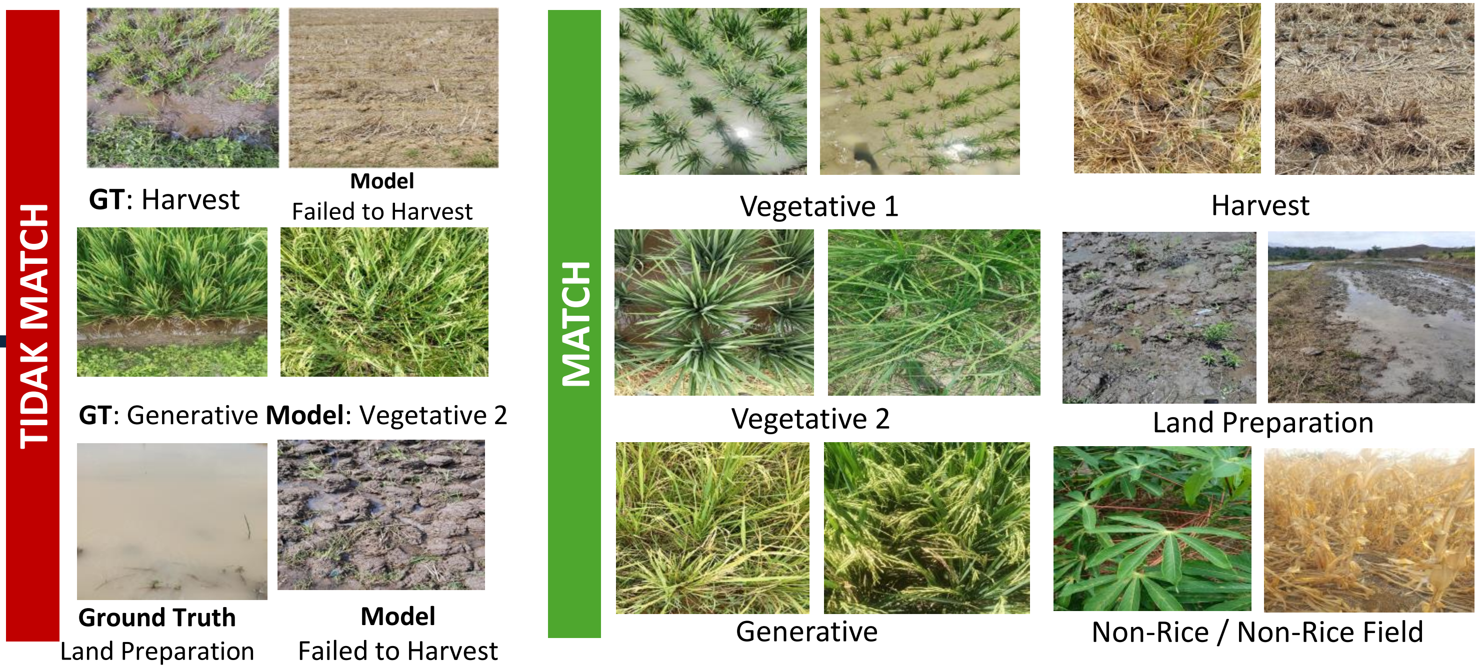
10 Provinces Ground Truth Mixed Method and Geotagging Results



Comparison Results



Comparison of Ground Truth Results and Model Prediction Results



GENERAL FINDINGS

- For the **Sumatra** region, the lowest matching evaluation results are found in rice fields of the **Swampy Lowland (Lebak)** type
- For the **Java** region, the lowest matching evaluation results are found in rice fields on slopes with the **Terraced** type.
- For **other regions**, the lowest matching evaluation results are found in **Rainfed Rice Field**s, **Terraced Rice Fields**, and **Heterogeneous** areas.

ADIOS