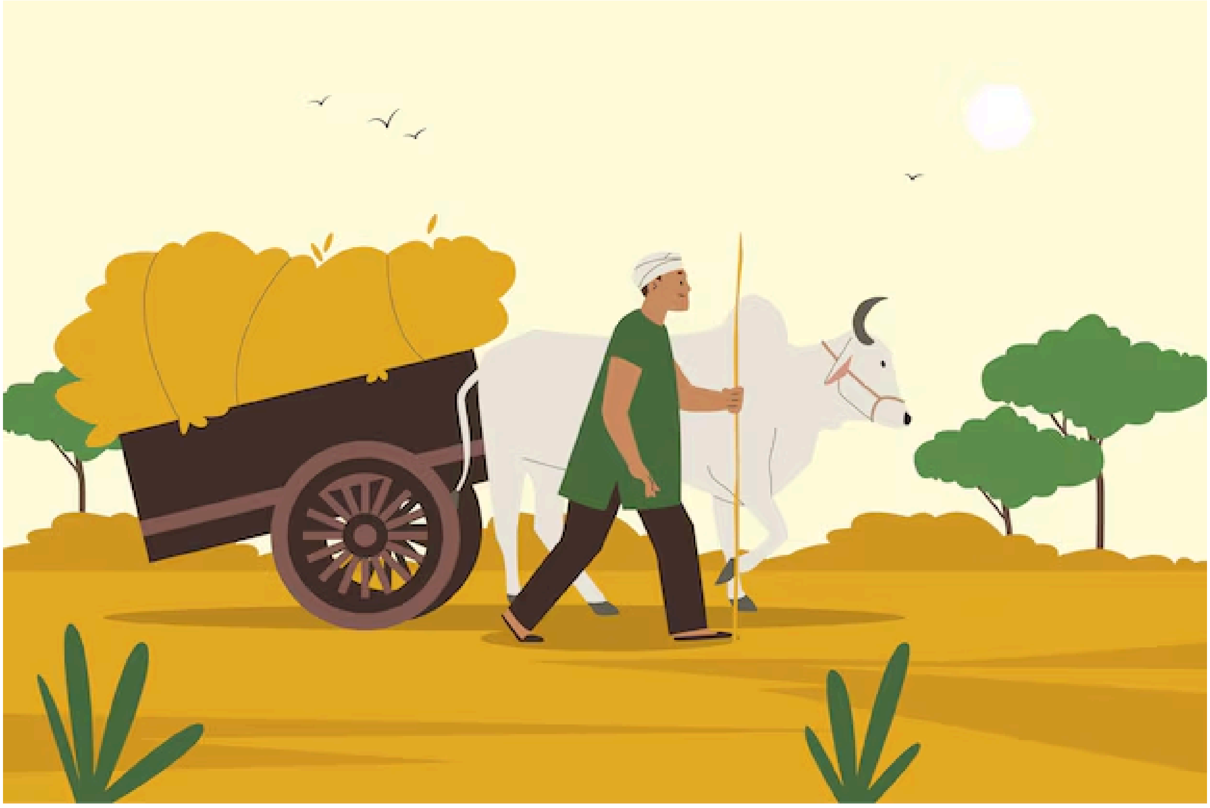
Crop + AI FairForward



Introduction

Crop type mapping using Machine Learning (ML) techniques has become crucial for effective agricultural monitoring and management. Telangana State, situated in southern India, boasts a diverse agricultural landscape characterized by various cropping patterns and smallholder farming systems. The main crops cultivated during the kharif season include paddy, cotton, maize, and red gram. In the rabi season, prominent crops include paddy, maize, Bengal gram, groundnut, and chilli. In recent years, advancements in remote sensing technology, coupled with the proliferation of open-access satellite data sources such as Sentinel-2, have opened up new opportunities for automated crop mapping. ML and AI algorithms offer the capability to process large volumes of satellite imagery and extract meaningful information about crop types with unprecedented speed and accuracy. This study aims to leverage ML and AI techniques to map crop types across Telangana State using Sentinel-2 satellite data. By harnessing the temporal and spectral information captured by Sentinel-2, along with ground-truth data collected through field surveys and agricultural census, we seek to develop robust models for classifying different crop types at a fine spatial resolution.

The main objective of the FAIR-FORWARD Project is to furnish timely and precise data on crop distribution, acreage. This information plays a crucial role in agricultural planning, resource allocation, and informed decision-making processes. Furthermore, by identifying and monitoring changes in cropping patterns over time, the project aims to extract valuable insights into the dynamics of agricultural production in Telangana. These insights encompass the effects of climate variability, shifts in land use, and socio-economic factors, thereby facilitating proactive measures for sustainable agricultural development. The present research focused on the classification of multi-temporal satellite images for cropland monitoring, including Land Use and Land Cover (LULC) mapping. The study utilized Sentinel-1 time series data in conjunction with Sentinel-2 Normalized Difference Vegetation Index (NDVI) data to monitor Rabi and Kharif crops in Telangana State, South India. Given the complex cropping patterns and small average field sizes of 1-2.5 acres in the study area, various classification methods were tested to accurately identify and map crop types. In this study, we demonstrate different ML approaches and SMT techniques to map major crop types using Sentinel-2 time series data and intensive ground data. The main goal of the study is to test different methods to estimate crop areas at the Mandal Level and identify the gaps in heterogeneous smallholder farming systems. We applied supervised classification with intensive training data collected during the Rabi season of 2023–24. Google Earth Engine (GEE) is capable of running various machine learning algorithms such as decision trees (e.g., GBC and RF) and optimization algorithms like Support Vector Machines (SVM), which aid in classification.

Study Area

The study area in Telangana State, situated in southern India, the region lies between 15°50'10" N and 19° 55'4" N latitudes and 77°14'8" E and 81°19'16" E longitudes. encompasses a diverse agricultural landscape characterized by varying topography, soil types, and climatic conditions. Telangana State is known for its rich agricultural heritage and is a significant contributor to the country's food production. Geographically, Telangana State is bordered by the states of Maharashtra to the north, Chhattisgarh to the northeast, Karnataka to the west, and Andhra Pradesh to the south and east. The state is predominantly located on the Deccan Plateau, with the Godavari and Krishna rivers flowing through its northern and southern regions, respectively. These rivers, along with their tributaries, play a crucial role in irrigation and agriculture in the state. The climate of Telangana varies from semi-arid to tropical, with hot summers and moderate winters. Rainfall is predominantly received during the monsoon season, which occurs from June to September. The state experiences diverse cropping patterns, with both Kharif (summer) and Rabi (winter) crops being cultivated. Major crops grown in Telangana include paddy, cotton, maize, red gram, Bengal gram, groundnut, chili, and various horticultural crops. The study area within Telangana State is characterized by complex cropping patterns, with a mix of rainfed and irrigated agriculture. The average field size in the study area is relatively small, ranging from 1 to 2.5 acres, reflecting the prevalence of smallholder farming systems. This presents challenges for accurate crop mapping and monitoring, necessitating the use of advanced remote sensing techniques and machine learning algorithms.

Districts (33)

Source: Telangana State Remote Sensing Applications Centre (TRAC)

S.No	District	No. of Revenue Divisions	No. of Revenue Mandals	No. of Revenue Villages	No. of Gram Panchayats
1	Adilabad	2	18	509	468
2	Kumuram Bheem	2	15	434	335
3	Mancherial	2	18	385	310
4	Nirmal	2	19	430	396
5	Nizamabad	3	29	450	530
6	Jagtial	3	18	290	380
7	Peddapalli	2	14	215	266
8	Jayashankar	1	11	223	241
9	Bhadradri Kothagudem	2	23	377	481
10	Mahabubabad	2	16	288	461
11	Warangal	2	13	191	323
12	Hanumakonda	2	14	163	208
13	Karimnagar	2	16	210	313
14	Rajanna Sircilla	2	13	171	255
15	Kamareddy	3	22	473	526
16	Sangareddy	4	27	600	647
17	Medak	3	21	381	469
18	Siddipet	3	24	383	499
19	Jangaon	2	12	176	281
20	Yadadri Bhuvanagiri	2	17	321	421
21	Medchal-Malkajgiri	2	15	162	61
22	Hyderabad	2	16	66	0
23	Rangareddy	5	27	599	558
24	Vikarabad	2	19	508	566
25	Mahabubnagar	1	16	307	441
26	Jogulamba Gadwal	1	12	203	255
27	Wanaparthy	1	14	223	255
28	Nagarkurnool	4	20	358	461
29	Nalgonda	3	31	566	844
30	Suryapet	3	23	279	475
31	Khammam	2	21	380	589
32	Mulugu	1	9	336	174
33	Narayanpet	1	11	252	280
	Total	**74**	**594**	**10,909**	**12,769**

Data Use

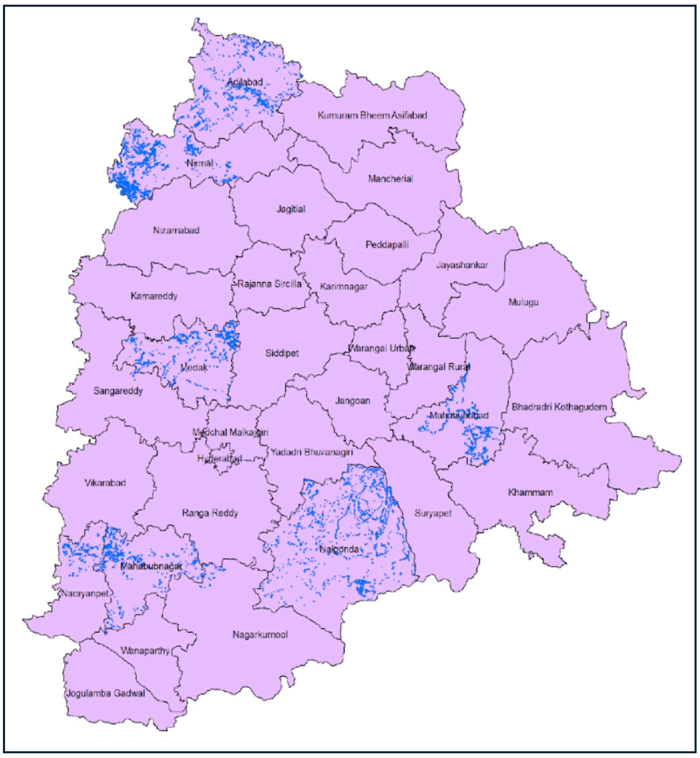
Satellite imagery/indices	Band	Significance
Sentinel 1	VH	Helps in identification of temporal changes
Sentinel 2	B2,B3, B4 and B8	
NDVI	B8 – B4)/(B8 +B4)	Vegetation identification

Image Acquisition

The present study uses Sentinel-2 satellite imagery for classification sentinel 2 images were collected for November 2023 to April 2024.

Ground Truth Distribution

In the ground truth collection for the Rabi season, a total of 9102 samples were collected across various crop types. The breakdown of samples collected for each crop is as follows: Bengal Gram: 990 samples, Chilli: 1340 samples, Groundnut: 1070 samples, Maize: 1725 samples, Paddy: 2255 samples, Red Gram: 60 samples



Sl No.	Name of Dist.	Bengal Gram	Chilli	Groundnut	Maize	Paddy	Red Gram	Total No of GT
1	Adilabad	497	3	0	172	12	0	684
2	Nirmal	483	3	2	1312	85	0	1885
3	Medak	9	26	1	213	942	11	1202
4	Mahabubabad	0	1281	2	25	4	1	1313
5	Nalgonda	0	0	200	0	1004	0	1204
6	Mahabubnagar	0	1	609	1	153	1	765
7	Kamareddy	0	0	1	1	1	0	3
8	Khammam	0	5	0	0	0	0	5
9	Nagarkurnool	0	0	28	0	5	0	33
10	Narayanpet	0	1	210		52	46	309
11	Vikarabad	0	0	11	0	0	1	12
12	Wanaparthy	0	0	6	0	0	0	6
13	Warangal	0	2	0	0	0	0	2
14	Siddipet	1	0	0	1	6	0	8

Methodology

1. Machine Learning Approach(mL)
2. Spectral Matching Approach(SMT)

Machine Learning approach crop type mapping

Leveraging the Google Earth Engine (GEE) API within the Python Colab environment, this methodology utilizes various algorithms such as Decision Tree, Random Forest, and Gradient Boost to classify crops based on multispectral data. The dataset encompasses all optical bands along with derived indices like Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI), collected throughout the Rabi Season.

Google Earth Engine API for Python Colab

The Google Earth Engine API facilitates the seamless integration of Earth observation data for analysis and visualization within the Python Colab environment. By leveraging GEE's extensive data repository and powerful computational capabilities, researchers gain access to a wealth of satellite imagery and geospatial datasets for crop classification tasks.

Utilized Algorithms:

1. Decision Tree: A versatile and interpretable algorithm, Decision Trees partition the feature space into regions, making it an ideal choice for classification tasks. By recursively splitting the data based on feature thresholds, Decision Trees construct a tree-like structure where each leaf node represents a class label.

2. Random Forest: Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. It mitigates overfitting and enhances prediction accuracy by aggregating the outputs of multiple decision trees.

3. Gradient Boost: Gradient Boosting is a machine learning technique that builds an ensemble of weak learners, typically decision trees, in a sequential manner. Unlike Random Forest, Gradient Boosting optimizes the model by minimizing a loss function through iterative refinement, sequentially adding predictors that minimize the loss.

Data Description

The dataset used for crop classification comprises all available optical bands, including visible, near-infrared, and short-wave infrared, along with derived spectral indices such as NDVI and NDWI. These spectral signatures provide valuable information about vegetation health, water content, and land surface characteristics, essential for distinguishing different crop types. Moreover, data collection spans the entire year, capturing seasonal variations and enabling robust classification across different growth stages of crops.

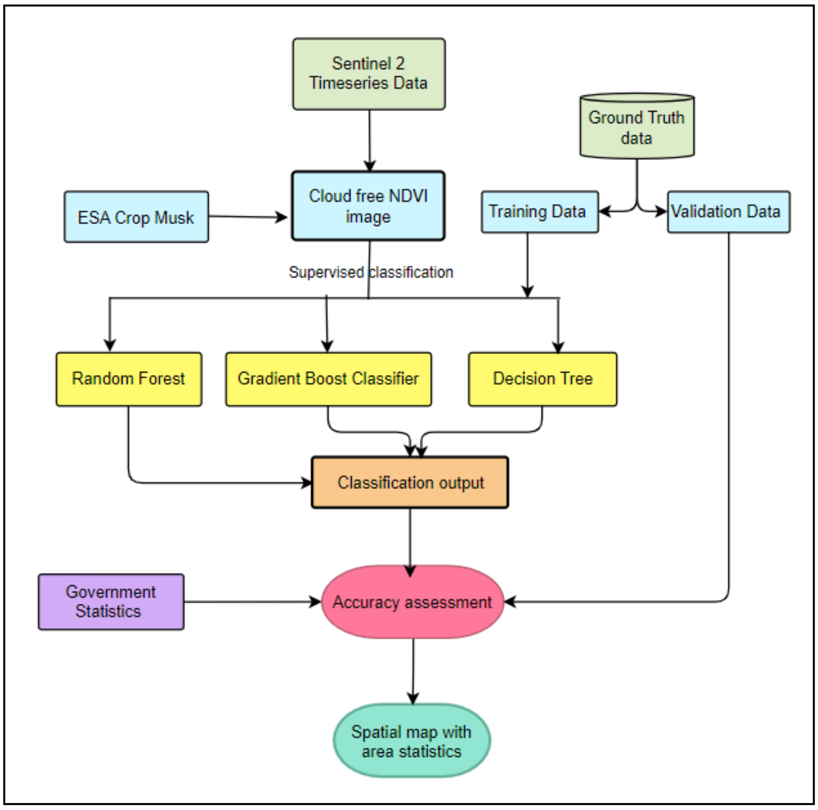


Fig. 1 Flow chart showing methodology used in GEE API for crop classification

Spectral Matching Techniques (SMT)

Spectral Mixture Technique (SMT) is a well-known technique in crop classification using NDVI time series data. This technique primarily involves generating ideal spectral signatures using ground data and correlating them with other spectral signatures to label the classes. The process begins with Google Earth Engine (GEE) downloading Sentinel-2 satellite data for the Rabi season (November 2023 to April 2024). Mean NDVI images are generated for each month and stacked in chronological order. Unsupervised K-means classification is then applied to the stacked NDVI image with 120 classes, which are grouped based on spectral similarity. Ideal spectral signatures are derived using ground data by extracting spectral values, including NDVI values. The 120 classes are compared with the ideal spectral signatures and ground data, and classes with similar time series are merged into a single class and labeled accordingly. Any discrepancies in classes are addressed by masking out the problematic class and repeating the process. Finally, accuracy assessment is conducted using field ground truth data within Google Earth Engine to validate the classification results.

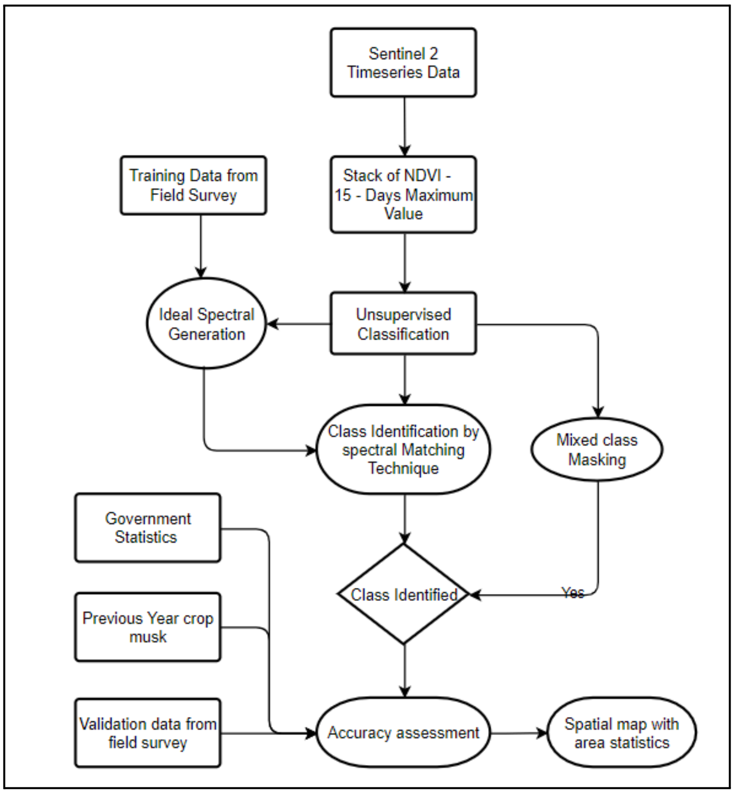


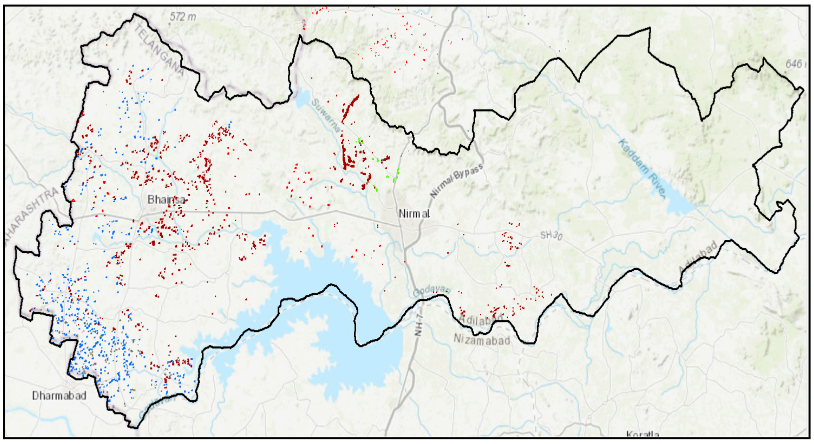
Fig. 2. Flow chart showing methodology of SMT for crop classification

Ideal NDVI Spectral signatures

Before employing Spectral Mixture Analysis (SMA), it's crucial to define ideal NDVI temporal signatures for each class of interest. In this study, Sentinel-2 time-series and ground survey data collected from homogeneous patches, coupled with comprehensive information about the cropping system and irrigation methods, were utilized to generate these ideal temporal NDVI signatures. The ideal spectra for six major crops in Telangana State—Paddy, Bengal gram, Groundnut, Cotton, Chilli, and Maize—were developed. Each major crop has distinct sowing and harvesting times: for instance, Cotton is sown from June/July to January/February, Paddy has Kharif season from June/July to October/November and Rabi season from December/January to April/May, Bengal gram is sown from October/November to January, and Groundnut from October/November to April. These differing time characteristics were exploited to effectively distinguish between various crops.

Image Classification

Crop type mapping in Telangana State involves two approaches to image classification: ML-based supervised classification and spectral matching technique. Below, we will discuss how these models are executed and the accuracy achieved from both approaches.



Class	No of GT from field visit	Training Point	Validation Point
Paddy	85	1994	522
Maize	1166	70883	17578
Groundnut	2	114	24
Fallow		6776	1685
Chilli	3	160	40
Bengal Gram	483	34566	8775

Fig.3 Ground Truth distribution of Nirmal dist. in Telangana state.

Machine Learning Classification for Multi Crop Mapping of Nirmal District

For multi crop mapping, three machine learning models were employed: Random Forest (RF), Gradient Boosting Classifier (GBC), and Decision Tree Classifier (DC). Among these, Gradient Classifier (RBC) yielded particularly impressive results. Below are the performance scores for each model, including Accuracy, Precision and Recall:

Classification Report:

	precision	recall	f1-score	support
0.0	0.39	0.79	0.52	1685
1.0	0.16	0.69	0.26	522
3.0	0.92	0.50	0.65	17578
5.0	0.01	0.83	0.01	24
6.0	0.03	0.78	0.06	40
7.0	0.69	0.74	0.72	8775
accuracy			0.59	28624
macro avg	0.37	0.72	0.37	28624
weighted avg	0.80	0.59	0.65	28624

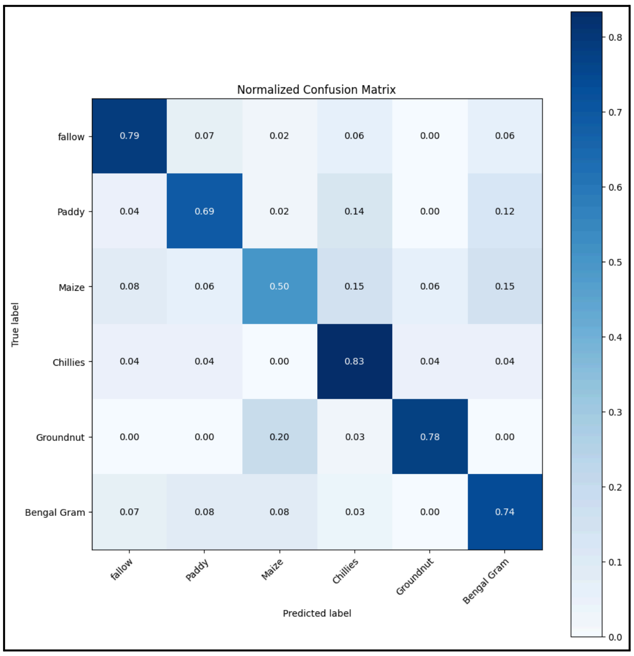


Fig. 4. Decision Tree Classification Report (left) and Confusion Matrix (right)

Classification Report:

	precision	recall	f1-score	support
0.0	0.44	0.83	0.57	1685
1.0	0.23	0.72	0.35	522
3.0	0.94	0.48	0.64	17578
5.0	0.01	0.83	0.02	24
6.0	0.02	0.93	0.03	40
7.0	0.68	0.83	0.75	8775
accuracy			0.62	28624
macro avg	0.39	0.77	0.39	28624
weighted avg	0.82	0.62	0.66	28624

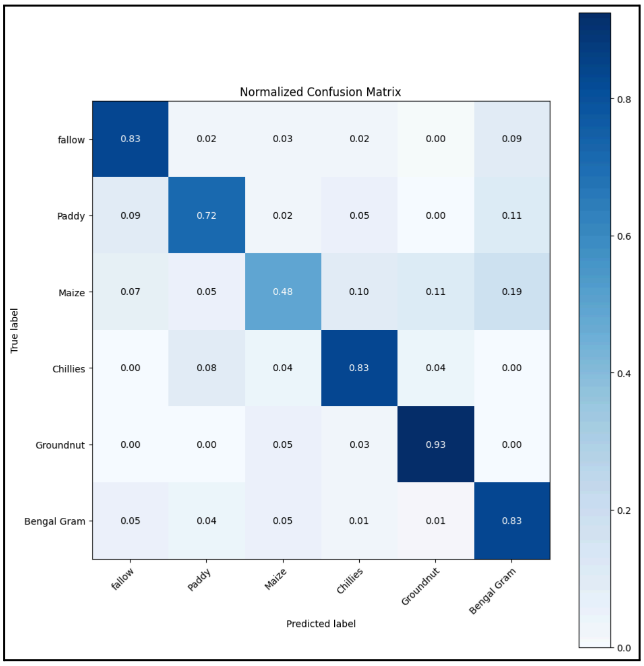


Fig. 5. Random Forest Classification Report (left) and Confusion Matrix (right)

Classification Report:

	precision	recall	f1-score	support
0.0	0.84	0.49	0.62	1685
1.0	0.85	0.47	0.61	522
3.0	0.85	0.89	0.87	17578
5.0	0.21	0.17	0.19	24
6.0	0.50	0.25	0.33	40
7.0	0.77	0.78	0.77	8775
accuracy			0.82	28624
macro avg	0.67	0.51	0.56	28624
weighted avg	0.82	0.82	0.82	28624

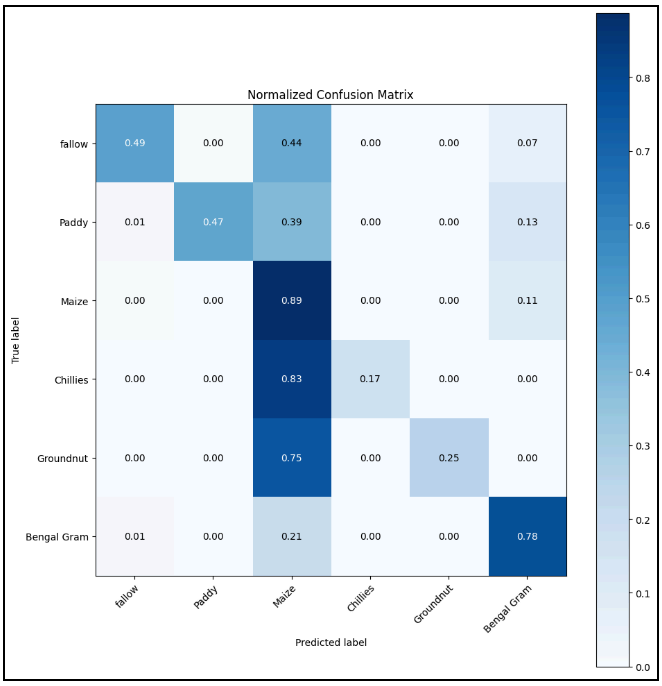


Fig. 6. Gradient Boost Classification Report (left) and Confusion Matrix (right)

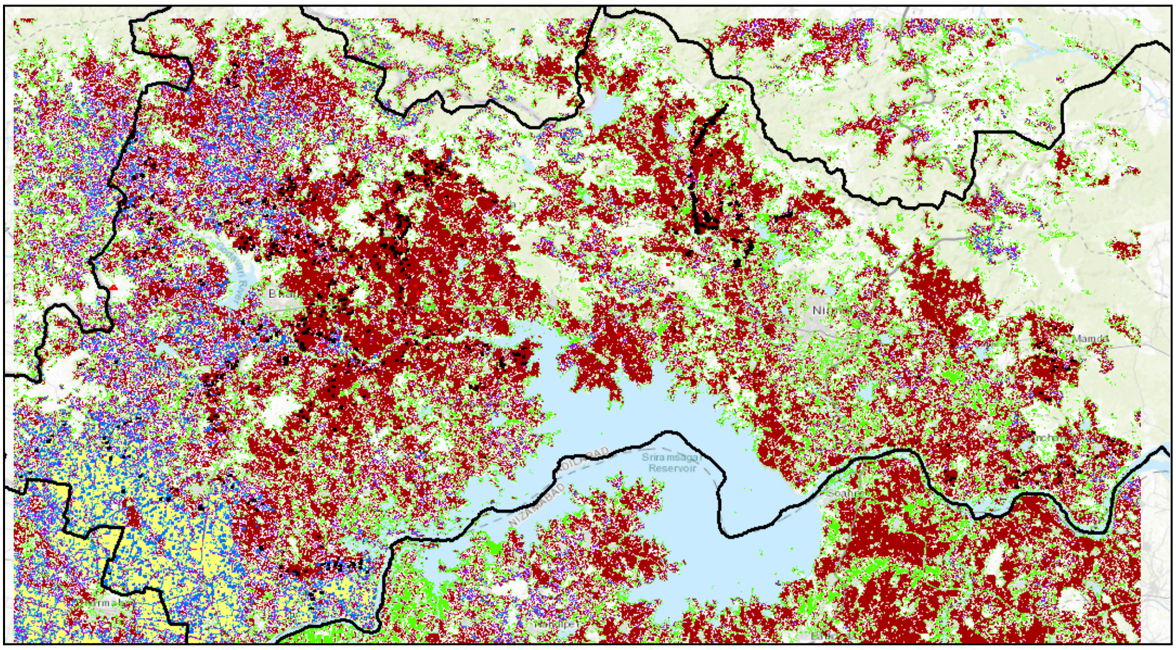


Fig. 7. ML Based Gradient Boost Classification for multi-crop type mapping

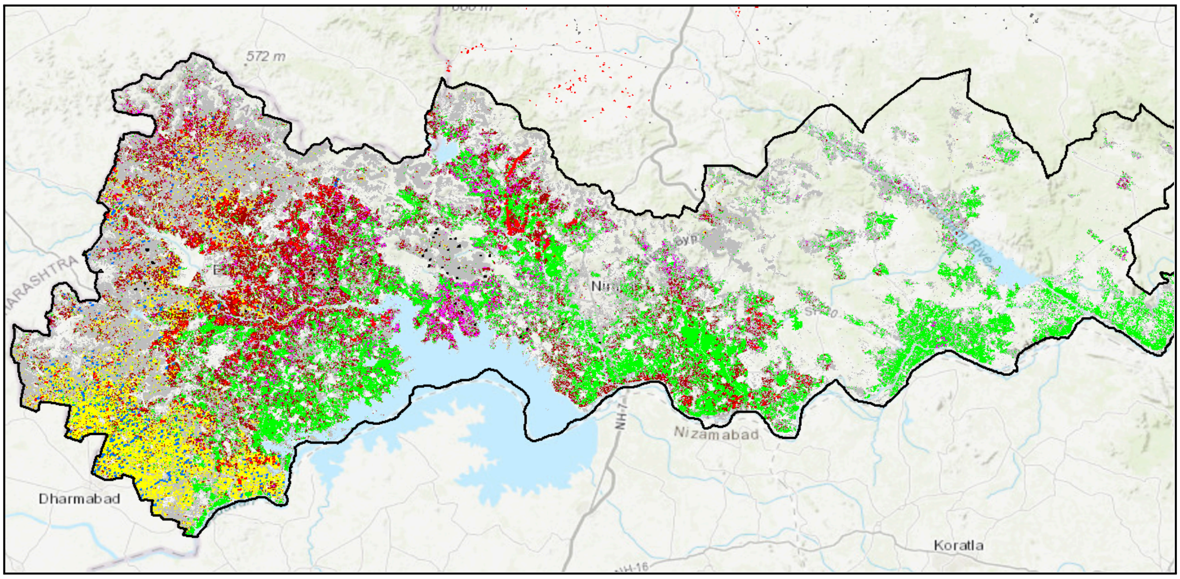


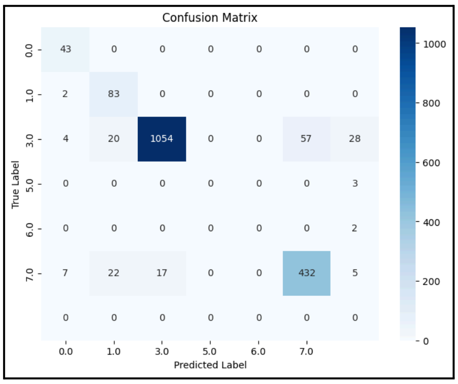
Fig. 8. Spectral Matching Techniques (SMT) Classification for multi-crop type mapping

```
Classification Report:
              precision    recall  f1-score   support

         0.0       0.77      1.00      0.87        43
         1.0       0.66      0.98      0.79        85
         3.0       0.98      0.91      0.94      1163
         5.0       0.00      0.00      0.00         3
         6.0       0.00      0.00      0.00         2
         7.0       0.88      0.89      0.89       483
         8.0       0.00      0.00      0.00         0

    accuracy                           0.91      1779
   macro avg       0.47      0.54      0.50      1779
weighted avg       0.93      0.91      0.92      1779
```

SMT Classification Report



SMT Confusion Matrix

Machine Learning Classification for Single crop Mapping

For single crop mapping, three machine learning models were employed: Random Forest (RF), Gradient Boosting Classifier (GBC), and Decision Tree Classifier (DC). Among these, Gradient Classifier (RBC) yielded particularly impressive results. Below are the performance scores for each model, including Accuracy, Precision and Recall:

```
Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.81      0.81      4849
           1       0.84      0.85      0.84      5786

    accuracy                           0.83     10635
   macro avg       0.83      0.83      0.83     10635
weighted avg       0.83      0.83      0.83     10635
```

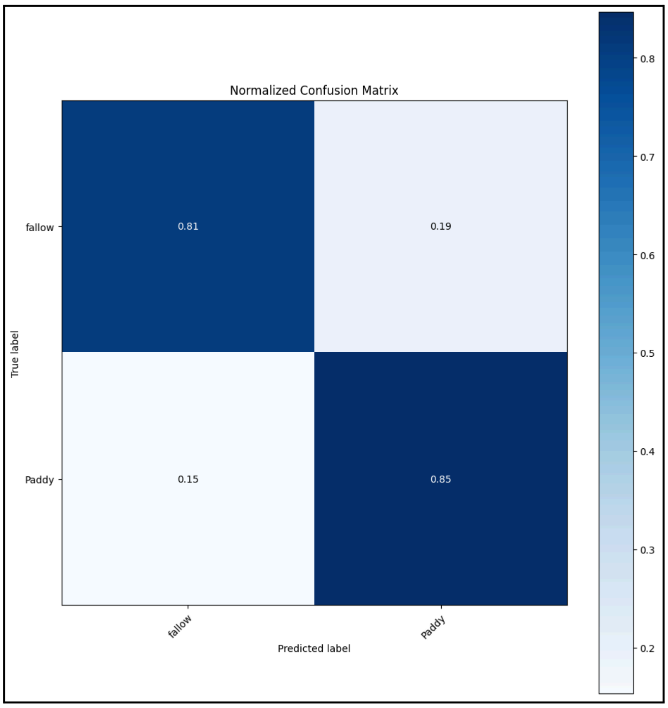


Fig. 9 Decision Tree Classification Report (left) and Confusion Matrix (right)

```
Classification Report:
              precision    recall  f1-score   support

           0       0.78      0.88      0.83      4849
           1       0.88      0.80      0.84      5786

    accuracy                           0.83     10635
   macro avg       0.83      0.84      0.83     10635
weighted avg       0.84      0.83      0.83     10635
```

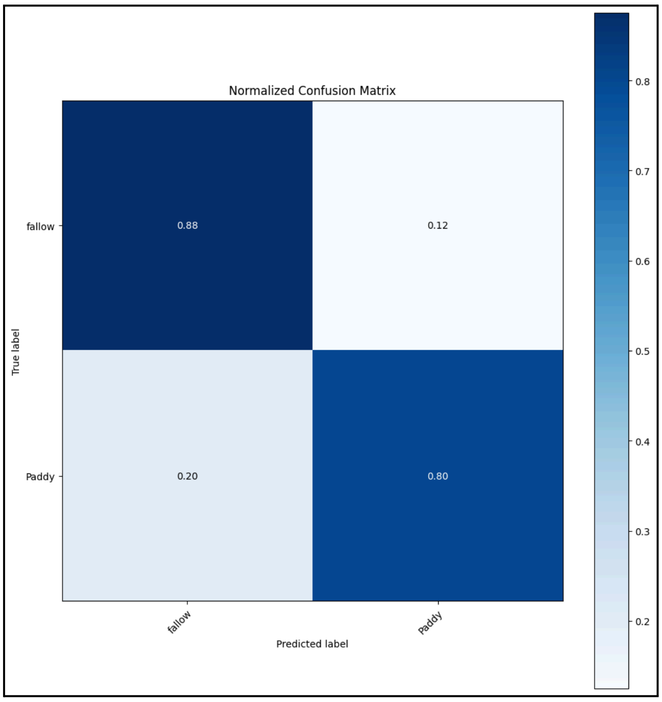


Fig. 10. Random Forest Classification Report (left) and Confusion Matrix (right)

```
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.85      0.87      4849
           1       0.88      0.90      0.89      5786

    accuracy                           0.88     10635
   macro avg       0.88      0.88      0.88     10635
weighted avg       0.88      0.88      0.88     10635
```

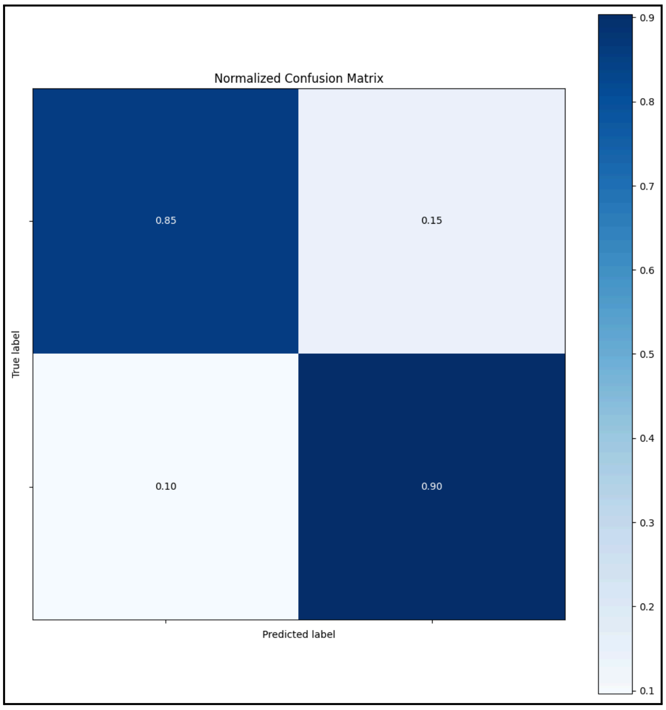


Fig. 11. Gradient Boost Classification Report (left) and Confusion Matrix (right)

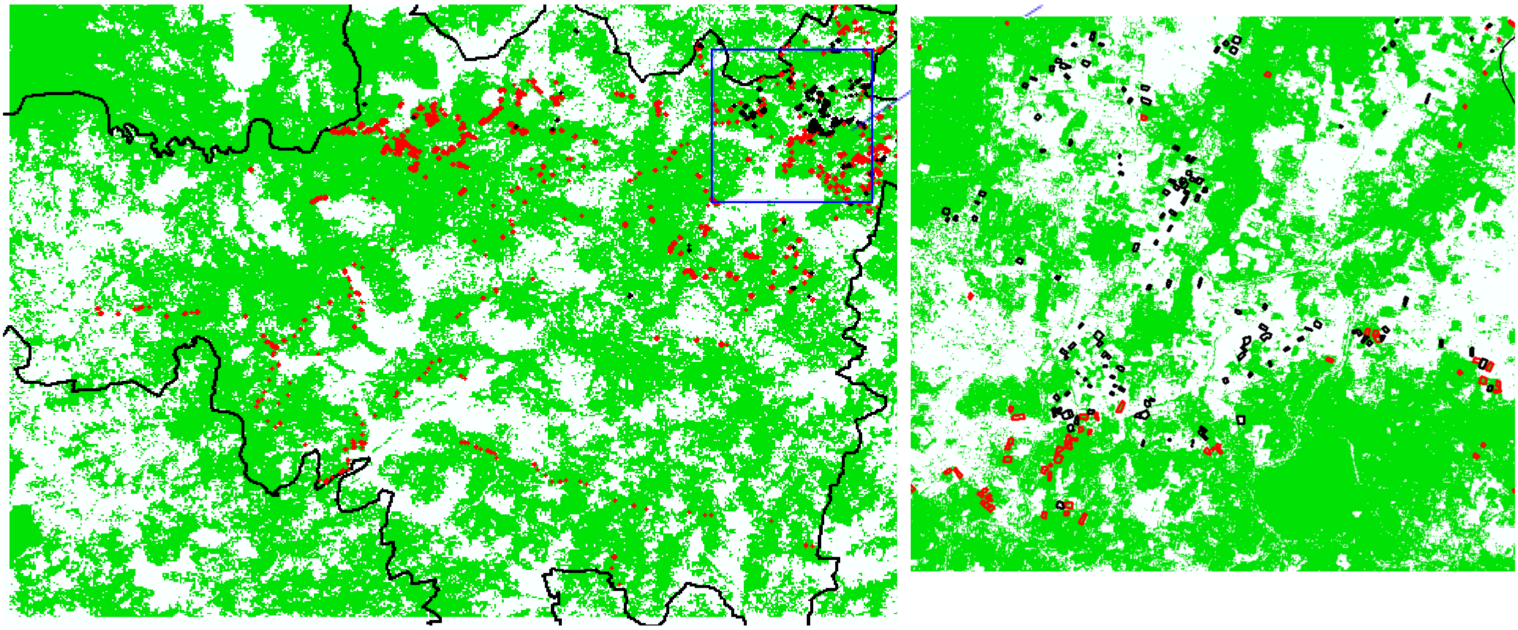


Fig. 12. Paddy crop classification map by Gradient Boosting Classifier model for Medak district

Red polygon is Paddy and Black polygon is Maize.

The results demonstrate the effectiveness of the machine learning approach in accurately classifying Paddy crops within the study area. Out of a total of 1202 ground truth samples, 942 were correctly classified as Paddy, indicating a high level of accuracy. However, challenges arose when attempting to map multicrop areas, particularly in distinguishing between mono and mixed crops due to their similar spectral signatures. To address this issue, the focus shifted to mapping areas dominated by a single crop. By refining the model to concentrate solely on the predominant crop type, significant improvements were observed in classification accuracy. The Accuracy, a measure of model performance, was reported at 88%, indicating strong agreement between the predicted and observed values.

In the classified image provided above, the red polygons represent areas classified as Paddy, while the black polygons represent areas classified as Maize. The machine learning model performed exceptionally well for single crop classification, as it accurately differentiated between Paddy and Maize. This accuracy demonstrates the effectiveness of the model in accurately mapping individual crop types in medak district.

Conclusion

The current Telangana Rabi crop study focuses on classifying crop types in smallholding farms using various ML and SMT techniques. Each classifier has its advantages in distinguishing different crops. The Gradient Boosted Trees (GBT) algorithm excelled in classifying mono-crops accurately, however, it has its limits in classifying mixed crops.Spectral Matching Technique (SMT) successfully classified every class and achieved good accuracy by reclassifying disparities. Gradient Boosted Trees (GBT) performed similarly to SMT. However, other classifiers like Decision Trees (DT) and RF achieved satisfactory results but occasionally misclassified other crops.

The study highlighted the significance of ground data in crop classification. RF and classification requires a larger amount of evenly distributed ground data. In terms of accuracy assessment, SMT outperformed RF classification. SMT initially classified satellite images into numerous classes using unsupervised methods, which were later simplified into major classes using ground data spectral values. In contrast, ML approaches solely classified images based on supervised classification using ground data training points. Therefore, ML classification accuracy heavily relies on the equal distribution of training points.