Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Integrated Optical and SAR Multi-Index Approach for Crop Discrimination Using Optimized Random Forest

Ajay Prakash M
Tamil Nadu Agricultural University
Ragunath K. P
ragunathkp@tnau.ac.in

Tamil Nadu Agricultural University
Pazhanivelan S
Tamil Nadu Agricultural University
Muthumanickam D
Tamil Nadu Agricultural University
Sivamurugan A. P
Tamil Nadu Agricultural University
Vanitha G
Tamil Nadu Agricultural University

Research Article

Keywords: Optical–SAR data fusion, Multi-crop classification, Sentinel-1 and Sentinel-2, Random Forest optimization, Statistical separability analysis, Precision agriculture

Posted Date: March 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8853952/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An Integrated Optical and SAR Multi-Index Approach for Crop Discrimination Using Optimized Random Forest

Ajay Prakash M[1], Ragunath K.P[2*], Pazhanivelan S[3], Muthumanickam D[4], Sivamurugan A.P [5], Vanitha. G [6],

[1] PhD Scholar, Department of Remote Sensing and GIS, Tamil Nadu Agricultural University, Coimbatore-641003, Tamil Nadu, India. (aj23723@gmail.com)

[2]Associate Professor (SS & AC), Centre for Water and Geo Spatial Studies, Tamil Nadu Agricultural University, Coimbatore-641003, Tamil Nadu, India (ragunathkp@tnau.ac.in)

[3]Director (CWGS), Centre for Water and Geo Spatial Studies, Tamil Nadu Agricultural University, Coimbatore-641003, Tamil Nadu, India. (pazhanivelans@gmail.com)

[4]Professor and Head, (SS & AC) Department of Remote Sensing and GIS, Tamil Nadu Agricultural University, Coimbatore-641003, Tamil Nadu India. (muthutnausac@gmail.com)

[5]Associate Professor (Agronomy), Centre for Water and Geo Spatial Studies, Tamil Agricultural University, Coimbatore-641003, Tamil Nadu, India (apacsivamurugan@gmail.com)

[6]Associate Professor, PG Dean office, Tamil Nadu Agricultural University, Coimbatore-641003, Tamil Nadu, India. (vanithag@tnau.ac.in)

ABSTRACT

Accurately distinguishing multiple crop types in diverse agricultural landscapes is essential for effective crop monitoring, food security assessments, and sustainable land management. However, this task remains challenging in mixed and smallholder farming systems, where crops often share overlapping phenological stages, and optical observations are frequently obstructed by persistent cloud cover. To address these challenges, this study developed a unified optical–SAR multi-index framework that integrates complementary multispectral and radar features with statistical separability analysis and an optimized Random Forest (RF) classifier for robust multi-crop mapping. Multi-temporal Sentinel-2 surface reflectance data were used to derive vegetation greenness, red-edge, moisture, stress, and disturbance indices that capture crop physiological conditions and phenological dynamics. Concurrently, Sentinel-1 SAR VV and VH backscatter coefficients and polarization-based indices were extracted to quantify the canopy structure, surface roughness, and moisture variability under cloud-prone conditions. These optical and SAR predictors were fused into a

high-dimensional feature stack and evaluated using the Bhattacharyya and Jeffries–Matusita distance metrics to quantify pairwise inter-class separability and identify informative, non-redundant features, particularly for spectrally similar crops. Multi-crop classification was performed using an RF classifier with hyperparameter optimization and feature importance–driven top-N predictor selection to reduce dimensionality and improve model generalization. Post-classification refinement using majority filtering and connected pixel cleaning enhanced the spatial coherence of the output maps. The framework was implemented in the coastal agricultural region of Cuddalore District, Tamil Nadu, India, during the Rabi season (2024–2025). The optimized model achieved an overall classification accuracy of approximately 0.81 and produced realistic crop area estimates that were consistent with regional agricultural patterns. The results indicate that Sentinel-2 red-edge and greenness indices were the dominant predictors, whereas Sentinel-1 SAR features provided complementary structural and moisture information that strengthened the discrimination among spectrally ambiguous crops. The framework successfully mapped rice, cotton, maize, groundnut, sugarcane, coconut, cashew, and casuarina, demonstrating a scalable and operational solution for multi-crop mapping in complex and cloud-prone agricultural landscapes.

Keyword: Optical–SAR data fusion; Multi-crop classification; Sentinel-1 and Sentinel-2; Random Forest optimization; Statistical separability analysis; Precision agriculture

Introduction:

Rapid population growth, climate variability, and increasing pressure on agricultural resources have intensified the global demand for precise large-scale crop monitoring to support food security and sustainable land management. Regions experiencing rapid population expansion often encounter significant challenges in ensuring adequate food availability, a situation further exacerbated by climate-induced shifts in temperature, precipitation, and the increasing frequency of extreme events, such as droughts and floods, which disrupt crop yields and traditional cropping systems(Kogo et al., 2021; Molotoks et al., 2021; Thakur et al., 2024).

Agriculture is both impacted by climate change and contributes to it through greenhouse gas emissions resulting from unsustainable practices, thereby complicating efforts

to improve food security(The Bucharest University of Economic Studies et al., 2019). Climate-smart agriculture (CSA) has emerged as a pivotal strategy for enhancing productivity, bolstering resilience, and mitigating emissions through methodologies such as precision agriculture, conservation practices, and regenerative farming. These strategies are heavily dependent on precise, large-scale crop data to optimize inputs and facilitate adaptive decision-making.(Kabato et al., 2025; Mehraj et al., 2024).

Traditional methods for mapping crops, which rely on spectral reflectance and the timing of vegetation growth, are effective in areas with single crop systems. However, they encounter considerable challenges in mixed and fragmented agricultural landscapes. In these settings, crops often display similar spectral characteristics and have overlapping growth periods, which diminishes the accuracy of conventional classification methods.(Gella et al., 2021; Salas et al., 2020). These challenges are particularly exacerbated in regions dominated by smallholder agricultural practices. In these areas, the limited size of landholdings, heterogeneous cropping systems, and diverse management practices result in mixed pixels and spatial variability that are difficult to capture accurately using traditional remote sensing techniques.

Multi-source remote sensing has demonstrated efficacy in overcoming these limitations by integrating complementary data from optical and Synthetic Aperture Radar (SAR) sources. Sentinel-2 offers valuable spectral indices for assessing vegetation conditions, although it is susceptible to cloud cover and atmospheric disturbances. In contrast, Sentinel-1 SAR provides all-weather, day-and-night observations that are sensitive to crop structure and moisture conditions(Chakhar et al., 2021; Holtgrave et al., 2020). The integration of optical and SAR time-series data improves crop discrimination by simultaneously capturing the biochemical, structural, and moisture-related characteristics of crops throughout the growing season(Han et al., 2019).

Machine learning techniques, especially Random Forest (RF), have shown exceptional performance in integrating high-dimensional multisensor datasets for crop classification. RF models are resilient to noise, effective with limited training samples, and capable of ranking feature importance, making them ideally suited for combining spectral indices, SAR features, and phenological metrics(Guo et al. 2022; Hariharan et al. 2018; Orynbaikyzy et al. 2020). In addition, statistical separability metrics, such as the Bhattacharyya and Jeffries–Matusita (JM) distances, serve as essential tools for assessing class differentiation and aiding in feature

selection within multi-temporal optical and SAR datasets(Guo et al., 2022; Wei et al., 2019). Although these measures are effective, they remain underutilized in complex agricultural settings involving high-dimensional and multimodal data(Orynbaikyzy et al., 2020).

To address these challenges, this study proposes a **Unified Optical–SAR multi-index and statistical separability measures framework** for discriminating multi-crops in heterogeneous agricultural landscapes.

- The framework integrates spectral indices derived from Sentinel-2 optical data with structural and moisture-sensitive features from Sentinel-1 SAR, enabling the effective characterization of crop biophysical properties across varying conditions.
- Statistical separability analysis guides feature selection, and an optimized Random Forest classifier is used to bolster classification robustness.

The proposed framework was implemented in a mixed-cropping region in Cuddalore, Tamil Nadu, India, utilizing multi-temporal data from Sentinel-1 and Sentinel-2. The classification performance was thoroughly evaluated to identify the most effective feature combinations and demonstrate the framework's applicability for large-scale agricultural monitoring, crop inventory generation, food security assessment, and sustainable agricultural planning.

2.Materials and methods

The proposed methodology for multi-crop discrimination, which utilizes a unified Optical–SAR multi-index and harmonic phenology framework, is depicted in (Fig.1). The process was initiated by acquiring multi-temporal Sentinel-2 multispectral and Sentinel-1 SAR imagery encompassing the study area in Cuddalore, Tamil Nadu, followed by standard pre-processing to ensure radiometric, atmospheric, and geometric consistency. Sentinel-2 spectral bands were employed to derive indices related to vegetation, moisture, and stress, whereas Sentinel-1 backscatter coefficients and SAR-based indices were extracted to capture crop structural and moisture characteristics. A statistical separability analysis using Bhattacharyya and Jeffries–Matusita (JM) distances was conducted to retain only the most informative and non-redundant features of the data. The selected optical–SAR features were subsequently

integrated and utilized to train a Random Forest classifier, with the dataset divided into training and testing subsets. The model performance was enhanced through systematic hyperparameter tuning, and feature importance analysis was conducted to identify the most discriminative predictors. The optimized model was then applied to generate multi-crop classification maps for the study region. Post-processing techniques, including SNIC-based majority filtering and connected pixel cleaning, were employed to refine the classification outputs. Finally, the classification results were evaluated using standard accuracy metrics, including the confusion matrix, overall accuracy, and Kappa coefficient, to assess the robustness and generalization capability of the proposed framework for large-scale agricultural monitoring.

2.1. Study area

The study area covered a 33 km coastal stretch in the Cuddalore district, situated between the Gaddilam and Vellar Rivers in Tamil Nadu, India (Figure *2*). Geographically, it stretches from 11°28′47″ N to 11°46′36″ N latitude and 79°44′01″ E to 79°49′18″ E longitude, encompassing a predominantly low-lying coastal plain shaped by both fluvial and marine processes and forming part of the Cauvery deltaic system. This region is an intensive agriculture hub, featuring a diverse mosaic of croplands, plantations, water bodies, and settlements. The soils here are varied, encompassing black, red, ferruginous, and arenaceous types, and are further categorized into clayey, loamy, and sandy-textured soils. Black soils are predominant in the taluks of Chidambaram and Virudhachalam, whereas sandy soils are found along the coastal tracts of Cuddalore and Chidambaram. Younger alluvial soils appear as patches along the rivers and streams. Cuddalore experiences a tropical monsoon climate influenced by both the Southwest Monsoon (June–September) and the cyclonic Northeast Monsoon (October–December). The region receives an average annual rainfall of approximately 1,086 mm, ranging from approximately 1,050 mm in Virudhachalam to 1,402 mm in Chidambaram, with Porto Novo receiving approximately 1,347 mm annually. The relative humidity ranges from 60% to 83%, reaching its highest levels during the Northeast Monsoon. Meanwhile, wind velocity remained moderate, with peak speeds occurring in May and the lowest in November. Temperatures generally fluctuate between 24 and 34°C, with recorded extremes reaching 40.3°C in June and dropping to 20.4°C in January. This range supports multiple cropping cycles but also introduces climatic variability into the system. The district is situated at the far end of the Cauvery Delta, receiving irrigation water from canals connected to the Mettur Dam, which is further supplemented by groundwater from the Neyveli lignite mines. Approximately 60% of the farmland relies on rain, and 40% is irrigated. The

irrigation network comprises 592 tanks, 270 canals, and a major reservoir, specifically the Wellington Reservoir in Thittagudi Taluk, along with the Veeranam Tank, which serves the Chidambaram and Kattumannarkoil Taluks. Perumal Lake is a crucial surface water source in Cuddalore Taluk. Agriculture remains the predominant livelihood, supporting nearly 80% of the population, with about 75% of the district's geographical area dedicated to cultivation.The major crops include paddy, cumbu, ragi, cholam, redgram, gingelly, tapioca, greengram, cashew nut, and black gram, alongside cash crops such as sugarcane, coconut, groundnut, and banana.The district is particularly notable for its cashew cultivation. The region's rich crop diversity, mixed cropping systems, predominance of smallholders, and sensitivity to climatic variability create intricate spectral and phenological patterns. This complexity makes it an ideal testbed for assessing multisource remote sensing and machine learning-based crop discrimination frameworks, utilizing multitemporal Sentinel-1 SAR and Sentinel-2 multispectral data.

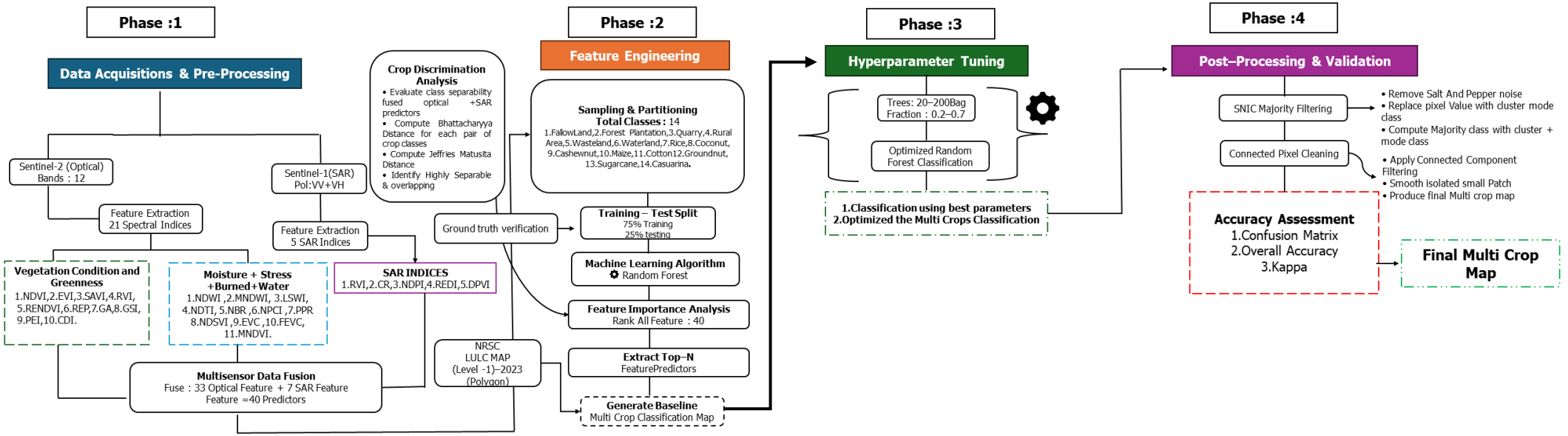


Figure 1. Workflow of the proposed unified Optical–SAR multi-index and Statistical Separability Measures framework for multi-crop discrimination.

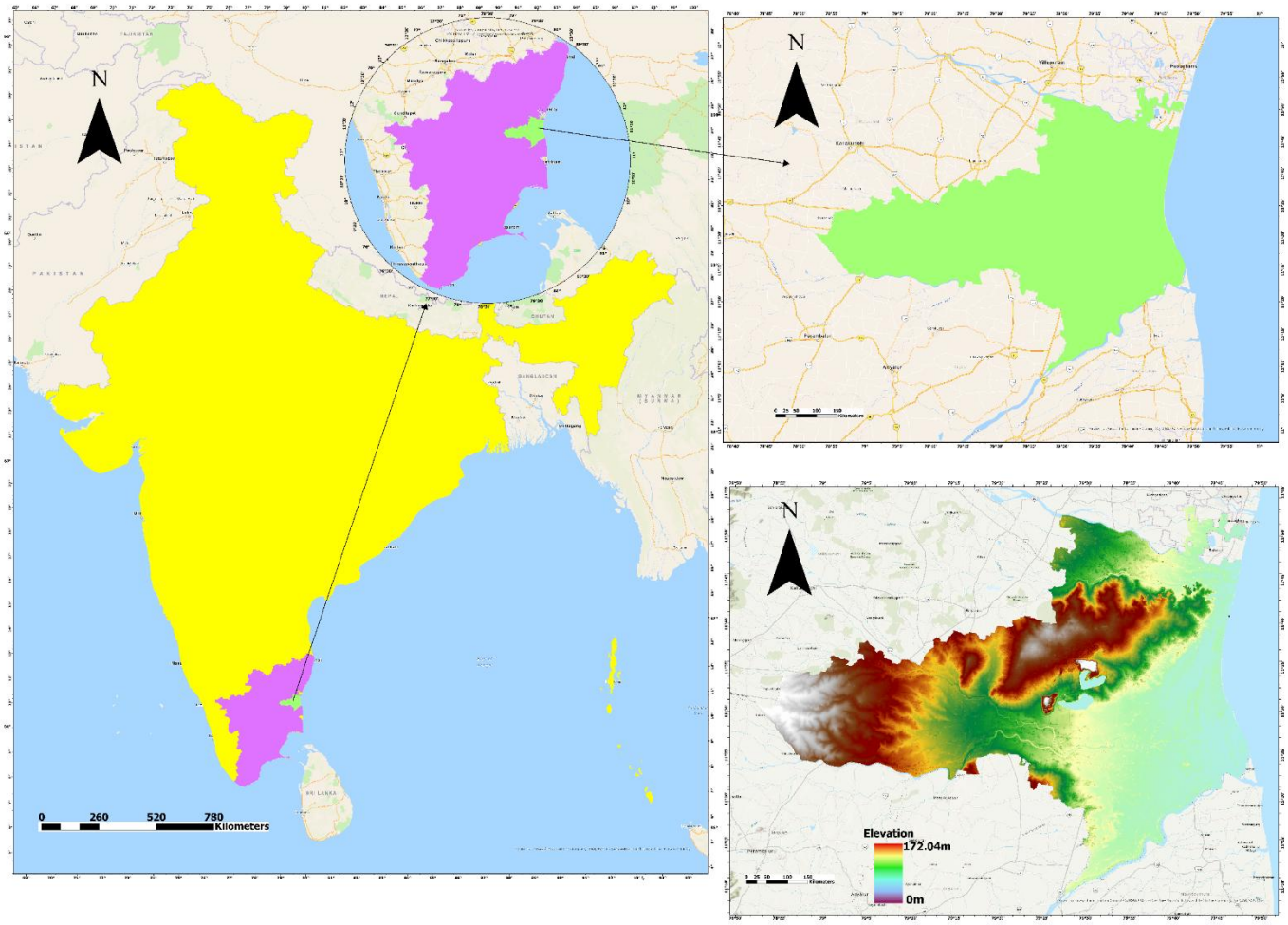


Figure 1. Location and spatial extent of the study area in Cuddalore district, Tamil Nadu, India.

2.2. Data collection

This study employed datasets comprising Sentinel-2 Multispectral Instrument (MSI) images, Sentinel-1 Synthetic Aperture Radar (SAR) images, and ground reference data to identify crop types in the coastal agricultural region of Cuddalore district, Tamil Nadu, India. These datasets were selected to provide complementary spectral, structural, and temporal information essential for accurate multi-crop differentiation in the study area. The specifics of the satellite and ground reference datasets are detailed in the following subsections. Additionally, the 2023 NRSC Land Use/Land Cover (LULC) Level-1 polygon dataset was used as an additional layer to identify agricultural areas. These polygon data were exclusively utilized to create a cropland mask prior to classification, ensuring that the multi-crop analysis was restricted to areas relevant to agriculture.

2.2.1 Sentinel-2 Multispectral and Sentinel-1 SAR Data

Remote sensing datasets comprising Sentinel-1 (S1) SAR and Sentinel-2 (S2) multispectral imagery were accessed and processed using the Google Earth Engine (GEE) platform for crop type identification, as detailed in **Error! Reference source not found.**. Harmonized Sentinel-2 Level-2A surface reflectance (SR) data were employed, ensuring that the observations underwent both radiometric and atmospheric corrections. To reduce cloud contamination while maintaining sufficient temporal coverage, the images were filtered using the metadata-based criterion of a CLOUDY_PIXEL_PERCENTAGE of 60%. Only the spectral bands relevant to vegetation and moisture analysis, including the visible, near-infrared, red-edge, and short-wave infrared regions, were retained for further processing.

Sentinel-1 SAR data were acquired in the Interferometric Wide (IW) mode as Ground Range Detected (GRD) products, featuring dual polarizations (VV and VH) with a spatial resolution of approximately 10 m. These datasets provide reliable insights into the structure and moisture sensitivity, unaffected by cloud cover or lighting conditions. Standard SAR preprocessing procedures, including orbit correction, thermal noise elimination, radiometric calibration, and terrain correction, were automatically executed in the GEE environment. The integration of Sentinel-2 optical data with Sentinel-1 SAR data enabled the combination of spectral, structural, and temporal information, establishing a robust foundation for distinguishing multiple crops in cloud-prone coastal agricultural regions.

Table 1Sentinel-1 SAR and Sentinel-2 multispectral data used for Rabi season crop monitoring in Cuddalore district.

Sensor	Study Area	Acquisition Period	Spatial Resolution (m)	Temporal Resolution
Sentinel-1 (SAR)	Cuddalore district, Tamil Nadu	Rabi season (October 2024 – March 2025)	**10 m**	**6 days**
Sentinel-2 (Optical)	Cuddalore district, Tamil Nadu	Rabi season (October 2024 – March 2025)	**10 m, 20 m, 60 m**	**5 days**

2.2.2 Field Surveying and Ground Reference Data

During the Rabi cropping season of 2024-2025, ground reference data were collected from manually identified ground control points (GCPs). These GCPs were established through visual interpretation, bolstered by field expertise and high-resolution satellite imagery, to accurately reflect the predominant crop and land-cover types in the study area. The study encompassed fourteen land-use and crop categories: fallow land, forest plantation, quarry, rural areas, water bodies, wasteland, rice, coconut, cashew nut, maize, cotton, groundnut, sugarcane, and casuarina. To address the class imbalance and manage the computational demands, the reference samples for each category were limited to 800 points. The entire GCP dataset was randomly divided into training and independent validation subsets for supervised classification and accuracy assessment, respectively. These ground reference samples formed the basis for model training and evaluating the effectiveness of the proposed optical–SAR multi-index and harmonic phenology framework in distinguishing multiple crop types. Figure 2. Location and spatial extent of the study area in Cuddalore district, Tamil Nadu, India. (Figure 2) illustrates the spatial distribution of the resulting multi-crop classification.

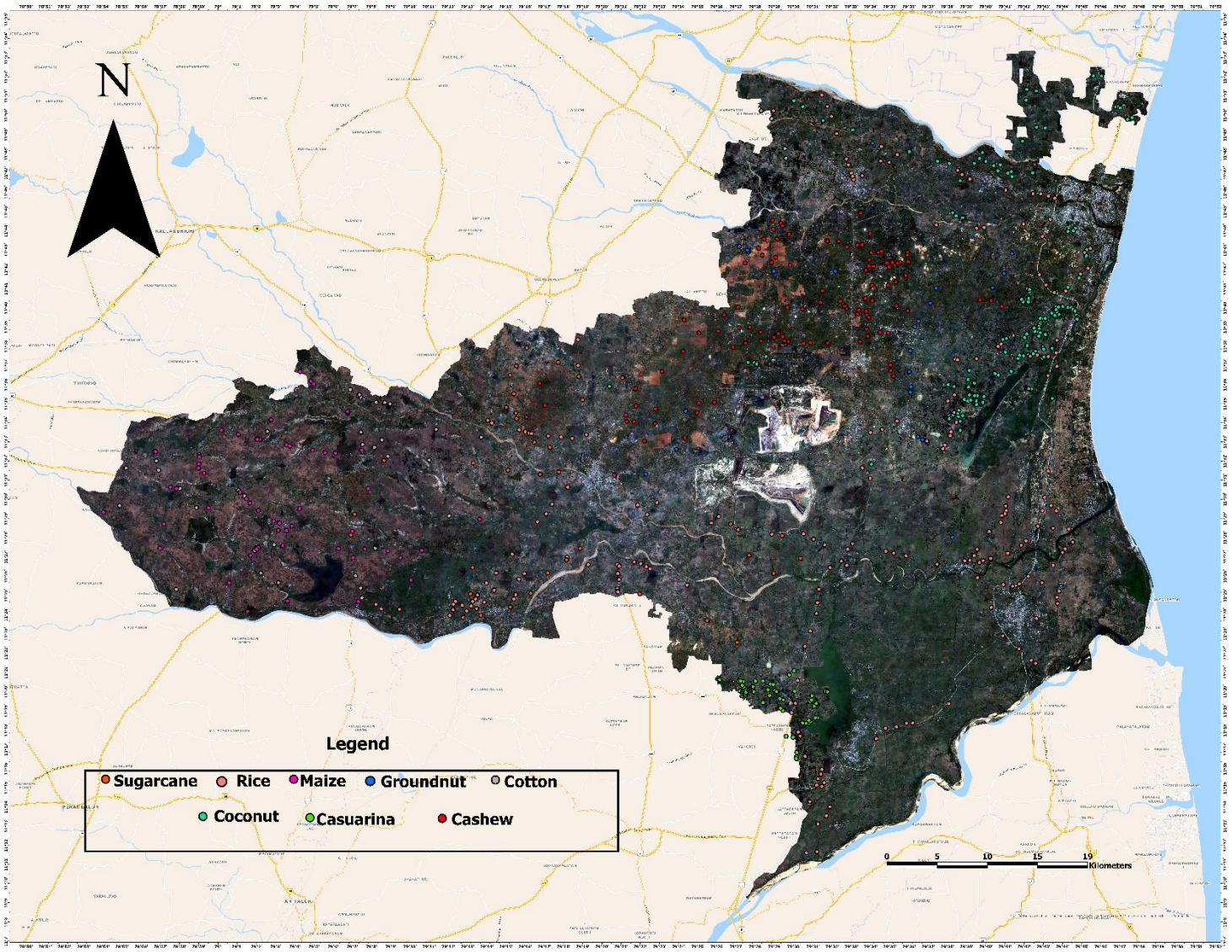


Figure 2 Spatial distribution of ground control points (GCPs) used for training and validation of the multi-crop classification model (Rabi season 2024)

2.2.3 Sentinel-2 Multispectral

To analyze the spectral characteristics and phenological differences of crops, 21 multispectral vegetation indices were extracted from the Sentinel-2 surface reflectance images (Table 1). These indices were organized into two main functional groups: (i) indices related to vegetation condition and greenness and (ii) indices associated with moisture, stress, burned surfaces, and water, to effectively capture the varied biophysical and biochemical attributes of crops in diverse agricultural settings.

(i)Vegetation Condition and Greenness Indices

Indices sensitive to biomass, chlorophyll levels, canopy density, and photosynthetic activity were used to measure the vegetation condition and greenness (Table 2) and (Fig4Table *1*.). This set included the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), Ratio Vegetation Index (RVI), and Modified NDVI (MNDVI), which effectively depict vegetation health while minimizing soil background interference. To improve sensitivity to chlorophyll content and subtle biochemical changes in the canopy, red-edge–based indices, such as the Red-Edge NDVI (RENDVI) and Red-Edge Position (REP), were utilized. Indices related to pigments and nutrients, such as the Normalized Pigment Chlorophyll Index (NPCI), Plant Pigment Ratio (PPR), Green Atmospherically Resistant Index (GA), Green Soil Index (GSI), Plant Enrichment Index (PEI), and Chlorophyll Degradation Index (CDI), were employed to capture variations in plant physiological status and nutrient dynamics. Composite indicators, such as Enhanced Vegetation Condition (EVC) and Further Enhanced Vegetation Condition (FEVC), combine multiple spectral responses to provide a comprehensive view of crop greenness and vitality.

(ii) Moisture, Stress, Burned Surface, and Water Indices

Indices sensitive to leaf water content, soil moisture, residue cover, and disturbance effects were used to characterize moisture availability, crop stress, surface conditions, and water presence (Table 3)and (Fig.5)Table *3*. This group included the Normalized Difference Water Index (NDWI), Modified NDWI (MNDWI), and Land Surface Water Index (LSWI), which are effective in identifying crop water status and irrigation patterns. Conditions related to stress and surface were represented by the Normalized Difference Tillage Index (NDTI), Normalized Burn Ratio (NBR), and Normalized Difference Senescent Vegetation Index

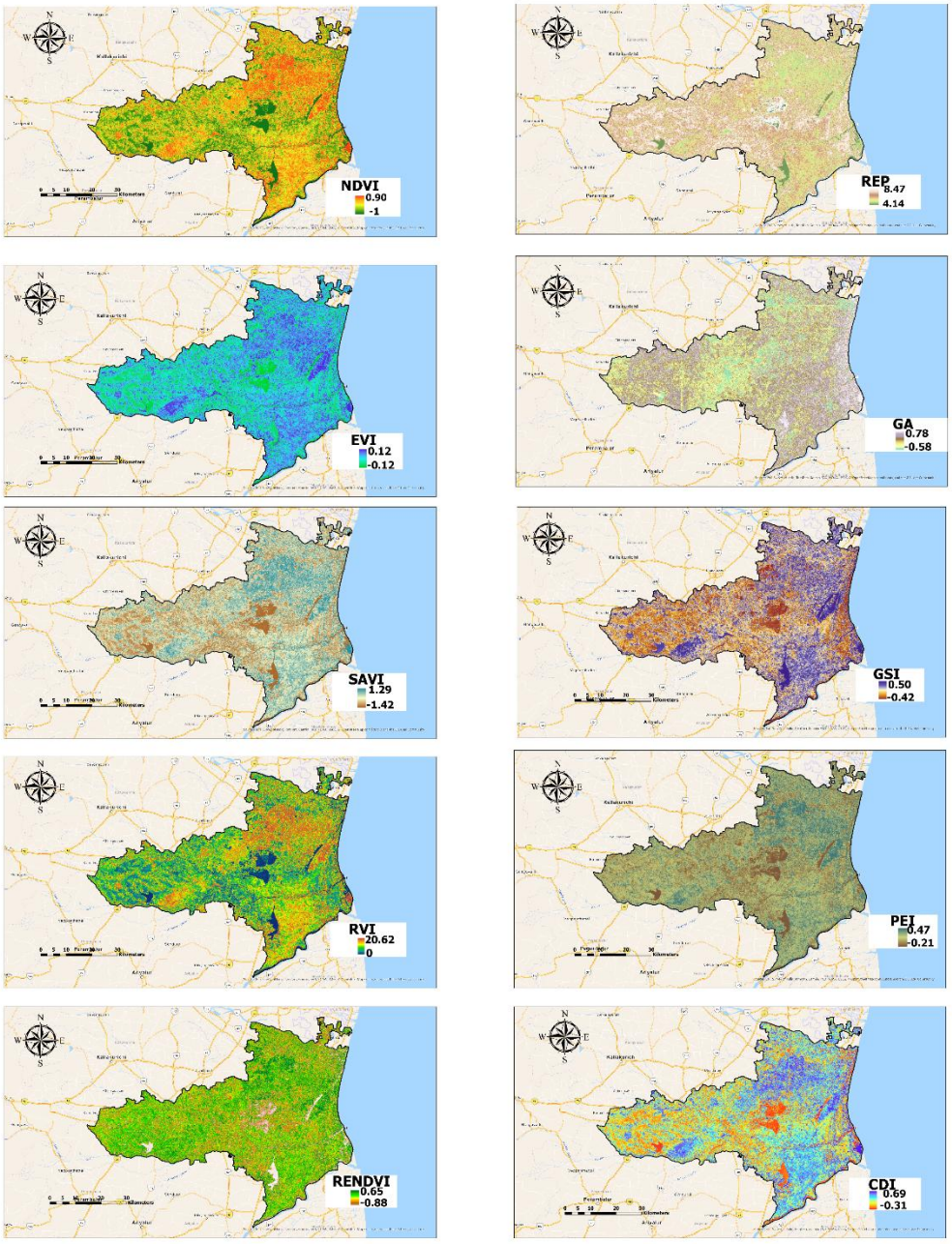
(NDSVI), allowing for the differentiation of stressed vegetation, senescent crops, and surfaces that were disturbed or covered with residue. Together, these two categories of Sentinel-2 multispectral indices offer a complementary and comprehensive description of crop biophysical conditions, physiological status, and moisture dynamics, forming an essential part of the feature space utilized in the proposed optical–SAR multi-index and harmonic phenology framework for distinguishing multiple crops.

Table 2 Vegetation conditions and greenness indices derived from Sentinel-2 multispectral data.

S. No.	Index	Equation	Purpose	Sentinel-2 Bands Used
1	NDVI	(B08 − B04) / (B08 + B04)	Vegetation greenness and biomass	B04, B08
2	EVI	2.5 × (B08 − B04) / (B08 + 6×B04 − 7.5×B02 + 1)	Reduces atmospheric & soil effects	B02, B04, B08
3	SAVI	(B08 − B04) × 1.428 / (B08 + B04 + 0.428)	Minimizes soil background	B04, B08
4	RVI	B08 / B04	Canopy density and vigor	B04, B08
5	MNDVI	(B08 − B04) / (B08 + B04 − 2×B01)	High-biomass sensitivity	B01, B04, B08
6	RENDVI	(B08 − B06) / (B08 + B06)	Chlorophyll sensitivity	B06, B08
7	REP	700 + 40 × ((B04 + B07)/2 − B05)/(B06 − B05)	Red-edge inflection	B04, B05, B06, B07
8	GA	(B03 − B01) / (B03 + B01)	Atmospherically resistant greenness	B01, B03
9	GSI	(B03 − B04) / (B03 + B04)	Vegetation–soil contrast	B03, B04
10	PEI	(B07 − B06) / (B07 + B06)	Nutrient enrichment	B06, B07
11	NPCI	(B04 − B03) / (B04 + B03)	Pigment degradation	B03, B04
12	PPR	(B03 − B02) / (B03 + B02)	Pigment composition	B02, B03
13	CDI	(B05 − B04) / (B05 + B04)	Chlorophyll degradation	B04, B05
14	EVC	NDVI − LSWI	Vegetation condition	NDVI, LSWI
15	FEVC	EVC × EVI	Enhanced vegetation condition	EVC, EVI

Table 3 . Moisture, stress, burned surface, and water indices were derived from Sentinel-2 multispectral imagery.

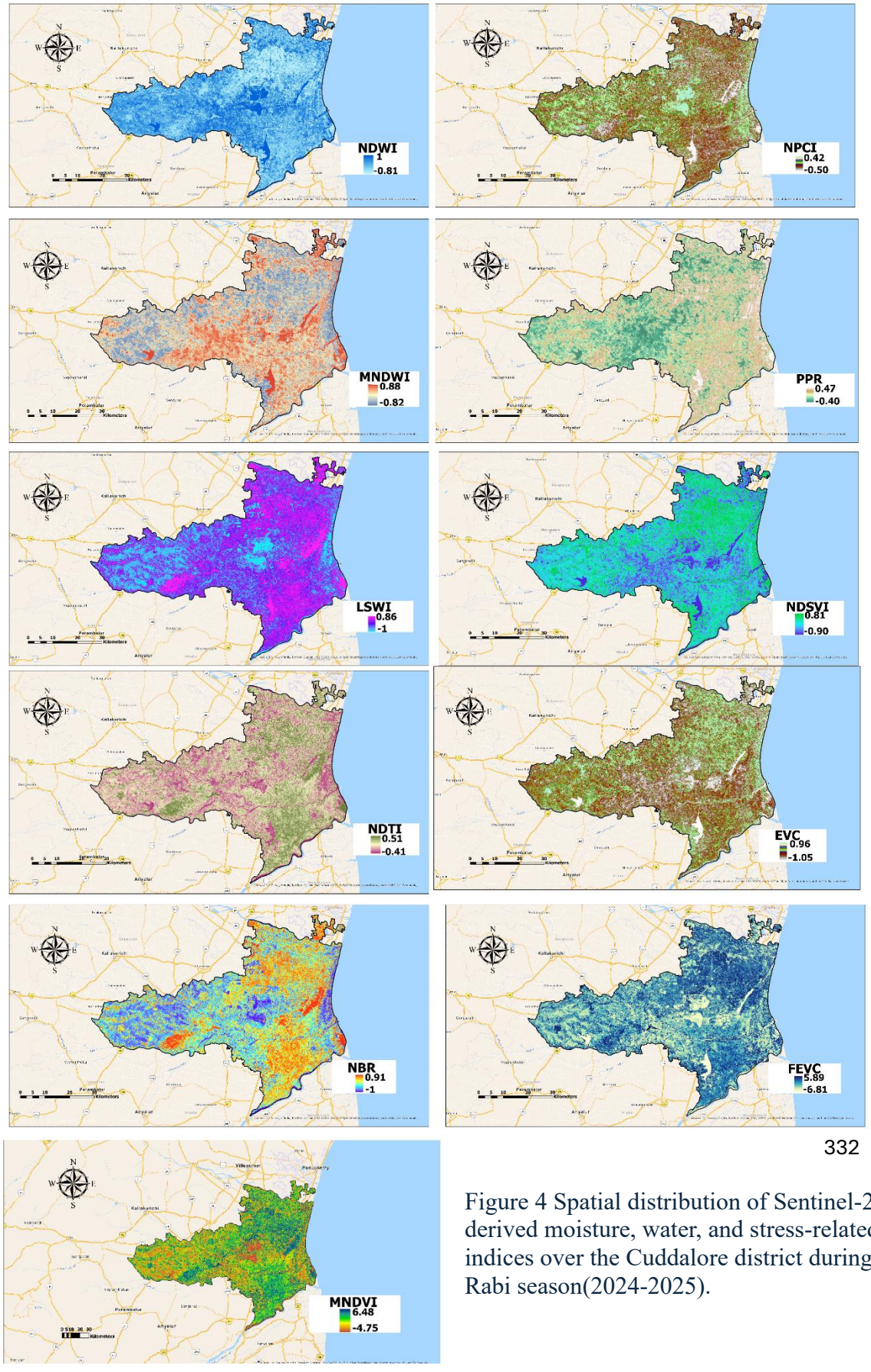
S. No.	Index	Equation	Purpose	Sentinel-2 Bands Used
16	NDWI	(B03 − B08) / (B03 + B08)	Surface water & moisture	B03, B08
17	MNDWI	(B03 − B11) / (B03 + B11)	Open water detection	B03, B11
18	LSWI	(B08 − B11) / (B08 + B11)	Vegetation water content	B08, B11
19	NDTI	(B11 − B12) / (B11 + B12)	Residue & soil disturbance	B11, B12
20	NDSVI	(B11 − B04) / (B11 + B04)	Senescent vegetation	B04, B11
21	NBR	(B08 − B12) / (B08 + B12)	Burned/stressed vegetation	B08, B12



309

Figure 3 Spatial distribution of Sentinel-2–derived vegetation indices over the Cuddalore district during the Rabi season(2024-2025).

311



332

Figure 4 Spatial distribution of Sentinel-2–derived moisture, water, and stress-related indices over the Cuddalore district during the Rabi season(2024-2025).

2.2.4 Sentinel-1 SAR Indices

To obtain information on crop structure and moisture that is unaffected by weather conditions, a series of indices based on Synthetic Aperture Radar (SAR) were developed using Sentinel-1 Ground Range Detected (GRD) data collected in Interferometric Wide (IW) mode with VV and VH polarizations (Table *4* & Figure 5) Table *4*. These indices take advantage of the sensitivity of SAR backscatter to factors such as canopy structure, biomass distribution, surface roughness, and moisture content, which are especially useful in cloudy and varied agricultural settings. Indices such as the Radar Vegetation Index (RVI) and Dual Polarization Vegetation Index (DPVI) were utilized to assess vegetation density and canopy structure complexity by combining data from both VV and VH backscatter components. The Cross Ratio (CR) and Normalized Difference Polarization Index (NDPI) were applied to improve the differentiation between crop types by highlighting variations in scattering mechanisms and polarization responses. The Radar Forest Degradation Index (RFDI) was used to detect changes related to canopy disturbance, structural degradation, and biomass variability. Collectively, these Sentinel-1 SAR indices offer complementary structural and moisture-sensitive data that improve crop differentiation when combined with multispectral Sentinel-2 indices in the proposed optical–SAR multi-index framework.

Table 4. Sentinel-1 SAR indices were derived from the VV and VH polarizations.

S. No.	Index	Equation	Purpose	Sentinel-1 Bands Used
1	RVI (Radar Vegetation Index)	$4 \times VH / (VV + VH)$	Quantifies vegetation structure and canopy density	VV, VH
2	CR (Cross Ratio)	VV / VH	Highlights polarization contrast related to crop structure	VV, VH
3	NDPI (Normalized Difference Polarization Index)	$(VV - VH) / (VV + VH)$	Enhances discrimination of scattering mechanisms	VV, VH
4	RFDI (Radar Forest Degradation Index)	$(VV - VH) / (VV + VH)$	Indicates canopy disturbance and structural degradation	VV, VH

5	DPVI (Dual Polarization Vegetation Index)	VH / VV	Represents vegetation density and moisture sensitivity	VV, VH

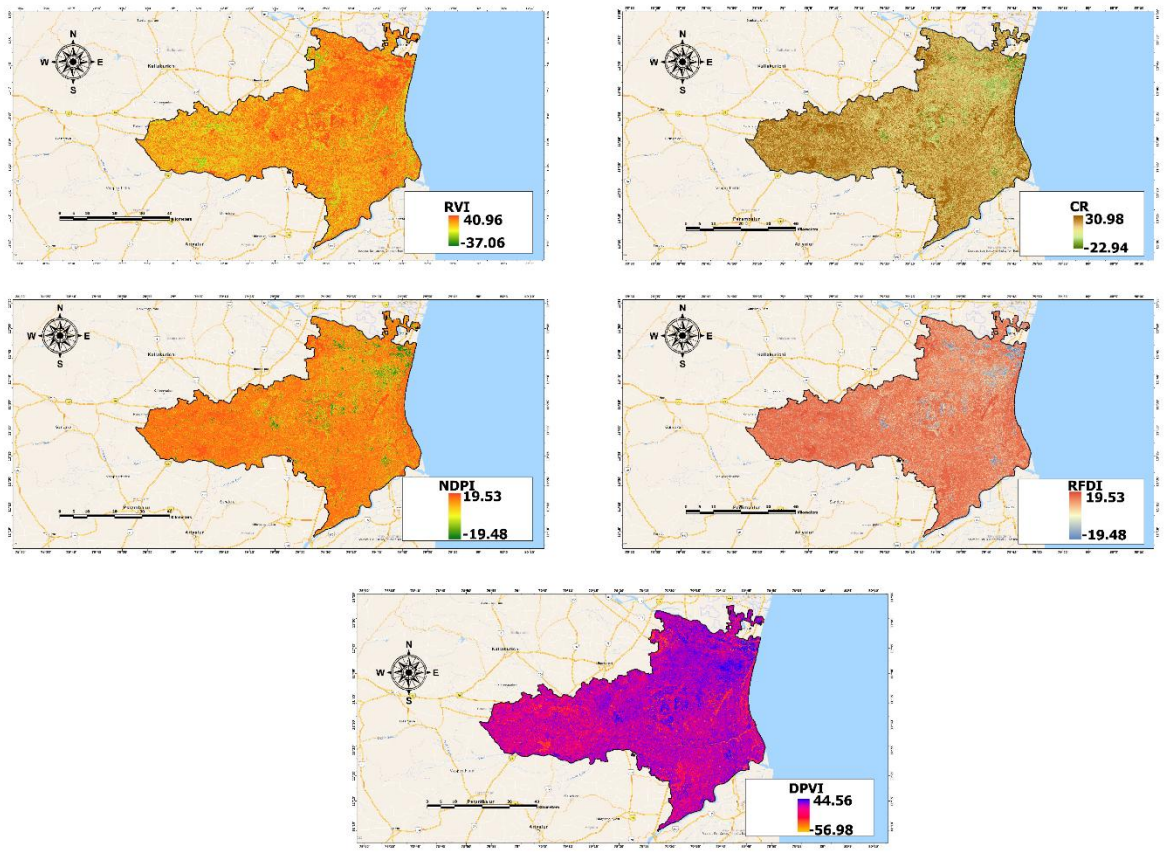


Figure 5 Spatial distribution of ratio- and difference-based vegetation indices derived from Sentinel-2 imagery over the Cuddalore district during the Rabi season(2024-2025).

2.2.5 Optical–SAR Feature Fusion and Predictor Stack Construction

After gathering and preparing the data, a combined optical–SAR feature fusion approach was employed to develop a comprehensive predictor stack for distinguishing multiple crops. This involved merging spectral bands and vegetation indices from Sentinel-2 with backscatter coefficients and radar-based indices from Sentinel-1 to collectively capture the spectral responses of crops, as well as their canopy structure, moisture dynamics, and phenological variations(Saad El Imanni et al., 2022). For Sentinel-2, the multispectral bands in the visible, near-infrared, red-edge, and shortwave infrared regions were enhanced with various vegetation indices that reflect crop greenness, pigment concentration, nutrient status, moisture conditions, and stress responses (Bhattarai et al. 2020). These indices improved the detection

of subtle biochemical and structural differences among crops that are often not discernible with raw spectral bands. Sentinel-1 SAR features, such as VV and VH backscatter and derived polarization-based indices, were incorporated to provide additional insights into crop geometry, surface roughness, and canopy water content. The inclusion of SAR features ensured a robust characterization of crop conditions, even under persistent cloud cover and varying illumination, which are typical in coastal agricultural areas. To minimize noise and maintain temporal consistency, all optical and SAR features were aggregated using median compositing over the study period and spatially clipped to the areas of interest. The final predictor stack comprised Sentinel-2 spectral bands, optical vegetation indices, Sentinel-1 backscatter coefficients, and SAR-derived indices, creating a unified, multidimensional feature space. This integrated optical–SAR predictor stack was used as the input dataset for subsequent feature selection, Random Forest classification, and statistical separability analysis, facilitating the effective discrimination of various crop types in diverse and mixed-cropping landscapes.

2.2.6 Statistical Separability Analysis for Feature Evaluation

To quantitatively assess the ability of the extracted optical–SAR features to distinguish between classes, a class separability analysis was performed using the Bhattacharyya Distance (BD) and the Jeffries–Matusita (JM) distance(Sen et al., 2019). These statistical tools evaluate the extent of overlap between class probability distributions and are commonly used to determine the effectiveness of spectral, structural, and phenological features in crop discrimination studies. The Bhattacharyya distance measures class separability by considering both the mean vectors and covariance matrices of the feature distributions, thus accounting for the variability within classes and differences between classes. However, because the Bhattacharyya distance is unbounded and can be challenging to interpret directly, the Jeffries–Matusita (JM) distance, a normalized version of the Bhattacharyya distance, was also utilized. The JM distance is defined as follows:

$$JM=2(1-e-B)$$

The Bhattacharyya distance is denoted by B. The JM distance ranges from 0 to 2, which enhances its interpretability when evaluating class separation.

For interpretation, the JM distance values were categorized as follows:

- **1.9–2.0**: Excellent separability
- **1.0–1.9**: Moderate separability
- **< 1.0**: Poor separability

Separability metrics were calculated for each pair of crop classes using all predictors, which included Sentinel-2 spectral bands, vegetation indices, Sentinel-1 SAR backscatter coefficients, radar-derived indices, and combined optical–SAR features. This pairwise analysis facilitated the identification of features that offered strong differentiation between crops that were spectrally similar and those with overlapping phenological stages.

The class separability analysis results were employed to (i) measure the overlap of distributions based on the mean and covariance, and (ii) pinpoint weak or redundant predictors. The integration of the Bhattacharyya and Jeffries–Matusita (JM) distance metrics offers a quantitative foundation for assessing inter-class separability among crop types. By incorporating a statistical separability analysis into the feature evaluation phase, the proposed framework enhances feature interpretability, minimizes redundancy in the predictor set, and strengthens classification robustness in diverse and mixed-cropping settings.

3. Methodology

3.1 Overview of the Crop Discrimination Framework

This study introduces a comprehensive framework that combines optical–SAR multi-index and harmonic phenology approaches to accurately distinguish multiple crops in diverse and mixed-cropping agricultural settings. The framework merges multi-temporal data from Sentinel-2 multispectral imagery and Sentinel-1 Synthetic Aperture Radar (SAR) with statistical feature analysis and an optimized Random Forest (RF) classifier to leverage complementary spectral, structural, and temporal data for crop mapping. The process began with gathering multi-temporal Sentinel-2 surface reflectance images and Sentinel-1 SAR data for the Rabi cropping season (2024–2025) in the Cuddalore district, Tamil Nadu. Sentinel-2

images were used to extract spectral bands and a wide range of vegetation, moisture, stress, and disturbance indices that describe crop greenness, biochemical traits, and water dynamics. Concurrently, Sentinel-1 SAR data collected in Interferometric Wide (IW) mode were processed to obtain backscatter coefficients and polarization-based indices that are sensitive to canopy structure, surface roughness, and moisture conditions, ensuring reliable crop monitoring even in cloudy conditions. Next, optical and SAR features were combined to create a unified optical–SAR predictor stack by integrating multispectral indices, radar-derived metrics, and multi-temporal data into a shared feature space. Before classification, a statistical separability analysis using Bhattacharyya and Jeffries–Matusita (JM) distance measures was performed to quantitatively assess inter-class discrimination and evaluate the effectiveness of individual predictors. The Random Forest classifier was then trained using ground reference data collected during the Rabi season of 2024. Feature importance was assessed solely using the RF model, and all predictors were ranked based on their relative impact on the classification performance. From this ranking, a Top-N subset of the most significant predictors was chosen to minimize redundancy, enhance computational efficiency, and improve model generalization. Hyperparameter tuning was then applied to optimize the RF classifier using the selected predictors. After classification, post-processing steps were implemented to refine the classification results. These steps include SNIC-based majority filtering to reduce salt-and-pepper noise and connected pixel cleaning to eliminate isolated misclassified patches, thereby improving the spatial coherence of the crop classes. Finally, the classification accuracy was evaluated using independent validation samples through standard accuracy metrics, such as the confusion matrix, overall accuracy, and Kappa coefficient. The proposed optical–SAR multi-index and harmonic frame discrimination workflow is depicted in **Error! Reference source not found.**showcasing the sequential integration of data acquisition, feature extraction, fusion, statistical evaluation, classification, post-processing, and accuracy assessment.

3.2 Feature Importance Analysis Using Random Forest

To evaluate the contribution of each predictor to distinguishing between multiple crops, a feature importance analysis was conducted using the inherent variable importance measure of the Random Forest (RF) classifier. The RF model was trained with the full optical–SAR predictor set, and importance scores were calculated based on the average decrease in node impurity across all decision trees. This RF-based method offers a robust, data-driven assessment of predictor significance, considering nonlinear relationships and complex interactions among spectral, structural, and moisture-sensitive features. The importance

ranking revealed that the Sentinel-2 optical predictors were the primary contributors to the overall classification performance. Indices based on vegetation greenness and the red-edge, such as NDVI, EVI, RENDVI, REP, and EVC, consistently showed high importance scores, indicating their strong sensitivity to crop biochemical properties, canopy vigor, and phenological changes. Indices related to moisture and stress, such as LSWI, NDWI, NBR, and NDTI, further improved class discrimination by capturing crop water status, residue cover, and surface disturbance conditions, which are particularly significant during the Rabi season. Although optical predictors had a higher overall importance, Sentinel-1 SAR features provided complementary and unique information. Backscatter coefficients (VV and VH) and polarization-based indices, such as the Radar Vegetation Index (RVI), Cross Ratio (CR), and Normalized Difference Polarization Index (NDPI), significantly contributed to characterizing canopy structure, surface roughness, and moisture conditions. These SAR-derived features are particularly useful for distinguishing crops with similar spectral responses but different structural attributes and for maintaining classification robustness in cloud-prone coastal areas. Overall, the RF-based feature importance analysis underscores the dominant role of optical multispectral indices while highlighting the added value of SAR features within a combined optical–SAR framework, thus offering a quantitative basis for subsequent feature ranking and dimensionality reduction.

3.3 Top-N Predictor Selection for Dimensionality Reduction

Utilizing the feature importance ranking derived from RF, a Top-N predictor selection method was employed to decrease the feature dimensionality while maintaining the classification accuracy. This approach focuses on dimensionality reduction by retaining only the most significant predictors, as determined by their importance scores. This method reduces redundancy among correlated variables, lowers computational demands, and improves model generalization in high-dimensional multi-sensor datasets. After training the RF with the complete optical–SAR predictor stack, all predictors were ranked by importance in descending order. From this list, the top 15 predictors (Top-N = 15) were automatically chosen to form a reduced-feature subset. This subset included a balanced mix of Sentinel-2 multispectral indices, mainly related to vegetation greenness, red-edge sensitivity, and moisture dynamics, and Sentinel-1 SAR backscatter and polarization-based indices, ensuring the retention of both spectral and structural information. The reduced Top-N predictor stack was then extracted from the full feature space and used for further model training and evaluation. By focusing on the most discriminative predictors, the Top-N feature selection method effectively reduced the

dataset dimensionality while preserving the essential information required for accurate multi-crop discrimination. This RF-guided dimensionality reduction enhanced computational efficiency, improved model stability, and facilitated more interpretable classification outcomes in diverse and mixed-cropping agricultural landscapes than previous methods.

3.4 Random Forest Classification Framework

To conduct multi-crop mapping, the Random Forest (RF) classification method was utilized with the optimized Top-N Optical–SAR predictor set. RF is an ensemble learning technique that generates numerous decision trees during the training phase and determines the class labels through majority voting. This approach enhances the classification stability and minimizes overfitting. Its proficiency in managing high-dimensional data, nonlinear feature interactions, and intricate class boundaries makes it especially suitable for applications involving multi-sensor remote sensing systems. One of the primary benefits of RF is its ability to seamlessly merge complementary data from optical and SAR data sources. Optical datasets offer detailed spectral and textural insights that are sensitive to vegetation health, biochemical attributes and surface features. In contrast, SAR datasets provide structural and moisture-related information that is largely unaffected by atmospheric conditions or lighting. This improvement is evident in research where the fusion of optical and SAR data using RF significantly increased the overall accuracy compared to using a single sensor (Hasituya et al., 2020). For instance, in the mapping of urban impervious surfaces and vegetation, integrating Sentinel-1 SAR and Landsat 8 optical features with RF resulted in better classification outcomes because of the exploitation of complementary features(Shao et al., 2024).

To develop and validate the Random Forest (RF) classifier, the reference dataset was randomly divided into training and testing subsets using a stratified sampling strategy. A 75–25% split was adopted to ensure an optimal balance between model learning and independent validation, as Random Forest classifiers require a sufficiently large training set to reliably capture complex, nonlinear relationships in high-dimensional feature spaces, while retaining an adequate proportion of unseen samples for unbiased performance evaluation. Accordingly, 75% of the ground reference samples were used for model training, providing a robust representation of crop variability, and the remaining 25% were reserved for independent testing and accuracy assessment. This partitioning strategy is widely recommended in remote sensing–based classification studies and has been shown to yield stable and reliable accuracy estimates

for ensemble machine-learning models. Random partitioning was performed using fixed seeds to ensure the reproducibility of the results. In addition to its ability to classify, RF naturally facilitates feature assessment and selection, which is especially beneficial in multi-sensor settings with correlated and diverse predictors. By concentrating on the Top-N most informative features identified through importance ranking, the RF model highlights the most distinguishing spectral indices, red-edge metrics, SAR backscatter components, and polarization-based indices, thus enhancing the classification efficiency and robustness(Zhang et al., 2021).Additionally, RF does not rely on assumptions about data distribution and is robust against noise and missing values, which are prevalent in optical–SAR datasets owing to cloud interference or sensor-specific artifacts. The Random Forest classifier, when utilizing an optimized Top-N Optical–SAR feature set, offers a strong and adaptable approach for classifying multiple crops. Its capacity to incorporate spectral, structural, and moisture-related data makes it especially suitable for practical crop mapping in diverse and cloud-prone agricultural regions.

3.5 Hyperparameter Optimization of the Random Forest Model

To improve model generalization and ensure effective multi-crop discrimination, a systematic approach to hyperparameter optimization was applied to the Random Forest (RF) classifier. RF was chosen for its established ability to manage high-dimensional, diverse datasets from multispectral remote sensing imagery and ground reference samples while maintaining computational efficiency and interpretability. The performance of Random Forest models can be enhanced by fine-tuning hyperparameters, such as the number of trees, tree depth, and feature selection, using advanced methods such as Harris Hawks Optimization or other metaheuristic approaches. This optimization process effectively minimizes overfitting and boosts generalization capabilities(Sabbaghi & Fathianpour, 2025).The refined model demonstrated exceptional classification accuracy, consistently differentiating between various crop types using spectral and vegetation index features derived from high-resolution remote sensing imagery. The enhanced Random Forest model operates with high efficiency, facilitating precise multi-class classification, which is crucial for agricultural surveillance, yield prediction, and crop management strategies. By incorporating feature fusion and optimization, the final model achieved classification accuracies that surpassed those of standard Random Forest techniques and other conventional machine learning methods, often reaching over 80-90% accuracy depending on the specific data and regional conditions (Lu et al., 2022).The optimized model was then applied to the complete optical–SAR predictor stack

for multi-crop classification. The specific hyperparameter values selected and their corresponding classification performances are detailed and discussed in the Results section.

3.6 Multi-Crop Classification Using the Optimized Random Forest Model

After ranking the feature importance, selecting the top-N predictors, and optimizing the hyperparameters, the final multi-crop classification was executed using the fine-tuned Random Forest (RF) classifier. This model was trained using the chosen Top-N Optical–SAR predictors, which combined multispectral indices, SAR backscatter elements, and polarization-based metrics into a single feature space. The optimized RF classifier was then applied to the complete optical–SAR predictor stack to produce a spatially continuous multi-crop classification map for the study area. By integrating spectral, structural, and moisture-sensitive data, the RF model effectively managed class complexity owing to diverse land-use patterns, mixed cropping systems, and overlapping crop phenology. To create and test the Random Forest (RF) classifier, the reference dataset was split into training and testing groups using a stratified sampling method. Specifically, 75% of the ground reference samples were allocated for training the model, whereas the remaining 25% were set aside for independent testing and evaluating accuracy. This division ensured that there were sufficient samples for the model to learn effectively while also allowing for an unbiased assessment of its classification performance. Random division was conducted using a fixed seed to ensure reproducibility of the results. The ensemble-based approach of RF facilitates robust decision-making through majority voting across trees, minimizing sensitivity to noise, outliers, and correlated predictors. Post-classification processing was performed to enhance the thematic consistency of the output map, including majority filtering to reduce salt-and-pepper noise and connected pixel cleaning to remove isolated misclassified pixels and improve the spatial coherence of crop patches. The resulting classification map represents the final multi-crop distribution used for further validation and accuracy assessments.

3.7 Accuracy Assessment and Validation

To quantitatively assess the classification performance, an accuracy evaluation was performed using independent validation samples that were not part of the model training process. These validation points were sourced from ground reference data gathered during the Rabi season and were randomly divided from the complete dataset to ensure an unbiased evaluation(Congalton & Green, 2019).A confusion matrix was created to compare the reference class labels with the corresponding classified output. From this matrix, standard

accuracy metrics were derived, including Overall Accuracy (OA) (Equation(1)), the kappa coefficient, producer's accuracy (PA)(Equation(2)), and user's accuracy (UA)(Equation(3)) for each specific crop. Overall Accuracy offers a comprehensive measure of classification accuracy, whereas the Kappa (Equation(4)) coefficient considers the agreement that occurs by chance. The producer and user accuracies for each crop were calculated to assess the omission and commission errors for each crop type, respectively, allowing for a detailed evaluation of the class-specific discrimination performance. This crop-wise accuracy assessment helps identify well-classified crops and classes that show spectral or structural confusion, providing a solid foundation for interpreting the classification reliability.

$$OA = Number\ of\ correct\ predictions \ / \ Total\ number\ of\ Predictions \quad (1)$$

$$PA = Number\ correctly\ idenitied\ in\ a\ given\ class \ / \ Number\ actually\ in\ the\ class \quad (2)$$

$$UA = Number\ correctly\ idenitied\ in\ given\ class \ / \ Number\ claimed\ to\ be\ in\ the\ class \quad (3)$$

$$\textbf{F-Score} = 2\times \frac{PA\times UA}{PA+UA} \quad (4)$$

4.Result:

4.1 Feature Importance and Top-N Predictor Analysis

The random forest-based feature importance analysis (Fig.7) revealed clear differences in the relative contributions of Optical and SAR predictors for multi-crop discrimination. Overall, the results demonstrate strong complementarity between Sentinel-2 multispectral data and Sentinel-1 SAR observations. Among all the predictors, the Sentinel-1 backscatter coefficient VH emerged as the most influential variable, followed closely by VV, highlighting the critical role of C-band SAR data in capturing crop structural characteristics, canopy roughness, and moisture variability. The dominance of SAR backscatter features indicates their effectiveness in distinguishing crops with similar spectral reflectance but differing canopy architectures, particularly under cloud-prone coastal conditions, where optical data availability can be limited.

Several Sentinel-2 optical bands also ranked prominently in the importance hierarchy, notably B9 (water vapor), B6 and B5 (red-edge bands), B8A and B8 (NIR), B3 (green), B1 (coastal aerosol), and B11 (SWIR-1). The high importance of the red-edge and NIR bands reflects their sensitivity to chlorophyll content, biomass accumulation, and phenological variation across

crop types, whereas SWIR-based bands contribute to detecting plant moisture status, soil background effects, and crop residue conditions during the Rabi season. The prominence of B9 further suggests an indirect role for atmospheric and moisture-related information in crop separability.

In addition to the raw spectral bands, the optical vegetation and moisture indices demonstrated strong discriminative power. Indices such as PPR, PEI, NDTI, LSWI, MNDWI, NBR, GA, NDVI, EVI, and RENDVI consistently ranked among the highest-ranked predictors. These indices effectively capture vegetation vigor, pigment concentration, surface moisture, irrigation dynamics, and disturbance signals, thereby enhancing crop class separability beyond what is achievable using spectral bands alone. The strong performance of pigment- and nutrient-sensitive indices (e.g., PPR and PEI) indicates their ability to highlight subtle biochemical differences among the crop types.

SAR-derived indices, including $DPVI_{s1}$, CR_{s1}, RVI_{s1}, $NDPI_{s1}$, and $RFDI_{s1}$, contributed moderately but consistently to the overall feature- These indices describe the polarization behavior and scattering mechanisms and are particularly valuable during periods of optical signal ambiguity caused by cloud cover or temporal inconsistencies. Their inclusion improved the classification robustness by reinforcing the structural and moisture-related information.

Based on the ranked importance scores, a Top-N predictor subset (N = 15) was selected to reduce the feature dimensionality while retaining the most informative variables. This subset comprised a balanced combination of Sentinel-1 SAR backscatter (VH and VV), key Sentinel-2 optical bands, and high-performing optical indices related to vegetation conditions and moisture dynamics. RF-guided feature selection significantly reduced redundancy among correlated predictors, improved model interpretability, and preserved classification performance. This optimized feature configuration provides a robust foundation for subsequent crop classification experiments and confirms that integrating spectral–biochemical information from optical data with structural–moisture sensitivity from SAR observations is essential for reliable multi-crop classification.

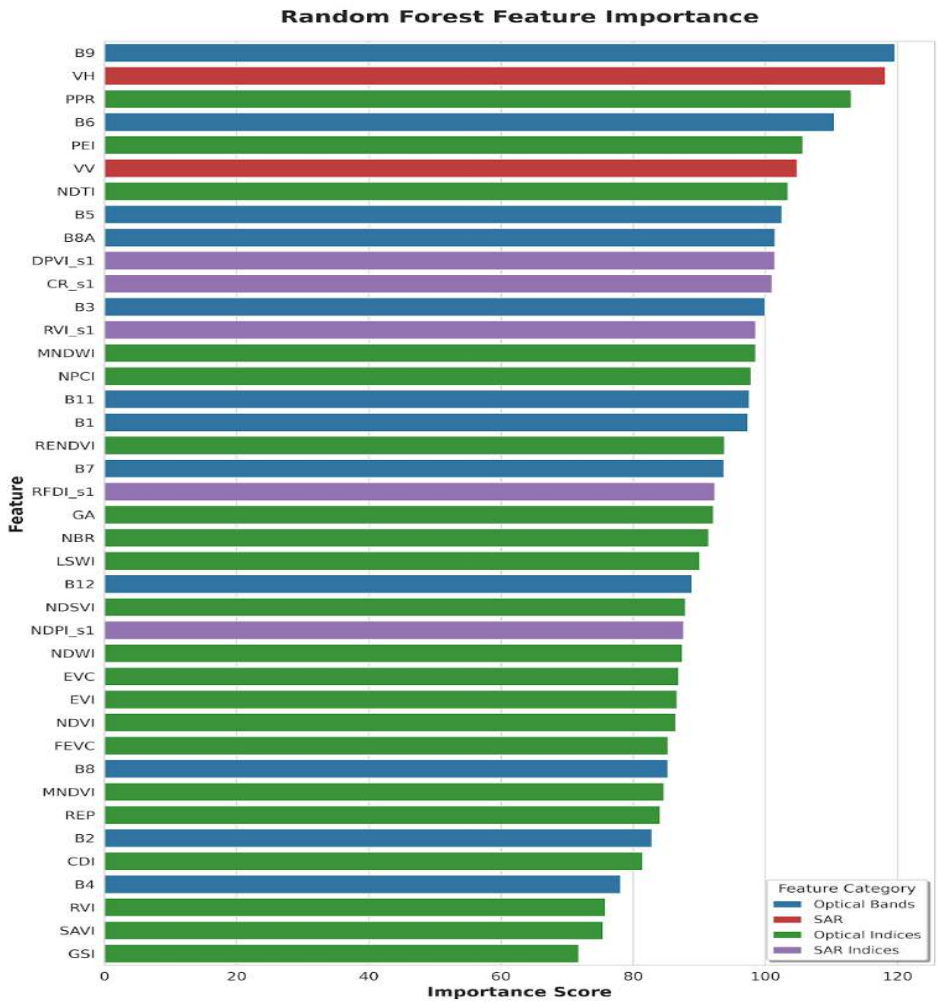


Figure 6 Random Forest Feature Importance Analysis For Rabi season (2024-2025)

675

676 **4.2 Class Separability Analysis for Multi-Crops.**

677 Using the Jeffries–Matusita (JM) distance, class separability analysis quantitatively evaluated
678 the ability of individual Optical and SAR predictors to distinguish between various crop pairs.
679 The ranking of the combined optical–SAR features revealed that indices derived from SAR—
680 specifically, the Cross Ratio (CR), Dual Polarization Vegetation Index (DPVI), Radar Forest
681 Degradation Index (RFDI), Normalized Difference Polarization Index (NDPI), and Radar
682 Vegetation Index (RVI)—achieved the highest average JM distances (≈ 1.0), indicating
683 excellent separability for crops with differing canopy structures and biomass characteristics.
684 Among the optical predictors, indices sensitive to the red-edge, moisture, and stress, such as
685 REP, LSWI, NBR, RENDVI, and PEI, also showed strong discriminative capabilities,
686 particularly for crops with distinct phenological behaviors and water-use patterns. The Average
687 crop separability analysis (Figure 8a) demonstrated significant discrimination between
688 perennial plantation crops (Casuarina, Coconut, and Cashew) and seasonal field crops (Cotton,
689 Maize, and Rice), with JM distance values generally above 1.0. Conversely, lower separability
690 was noted among spectrally and structurally similar crop pairs, such as cotton–Groundnut,
691 Groundnut–Maize, and Rice–maize, suggesting an inherent overlap in canopy architecture,
692 growth stages, and surface conditions during the study period. The heatmap-based separability
693 results further highlighted the spatial and index-wise discrimination patterns across the crop
694 pairs (Fig. 8 b). Optical indices related to moisture dynamics and red-edge response (LSWI,
695 NBR, REP, and RENDVI) demonstrated enhanced separability for water-intensive crops such
696 as Rice and Sugarcane. Similarly, SAR-based separability heatmaps (Fig. 8 c) consistently
697 showed strong discrimination across most crop combinations, especially under dense canopy
698 and high-biomass conditions, emphasizing the robustness of SAR features in mitigating
699 spectral confusion caused by cloud cover and atmospheric variability. Overall, the JM
700 distance–based separability analysis confirmed the complementary strengths of Optical and
701 SAR predictors, supported the selection of the Top-N feature subset, and demonstrated their
702 effectiveness for robust and reliable multi-crop discrimination in diverse and mixed-cropping
703 agricultural landscapes.

704

705

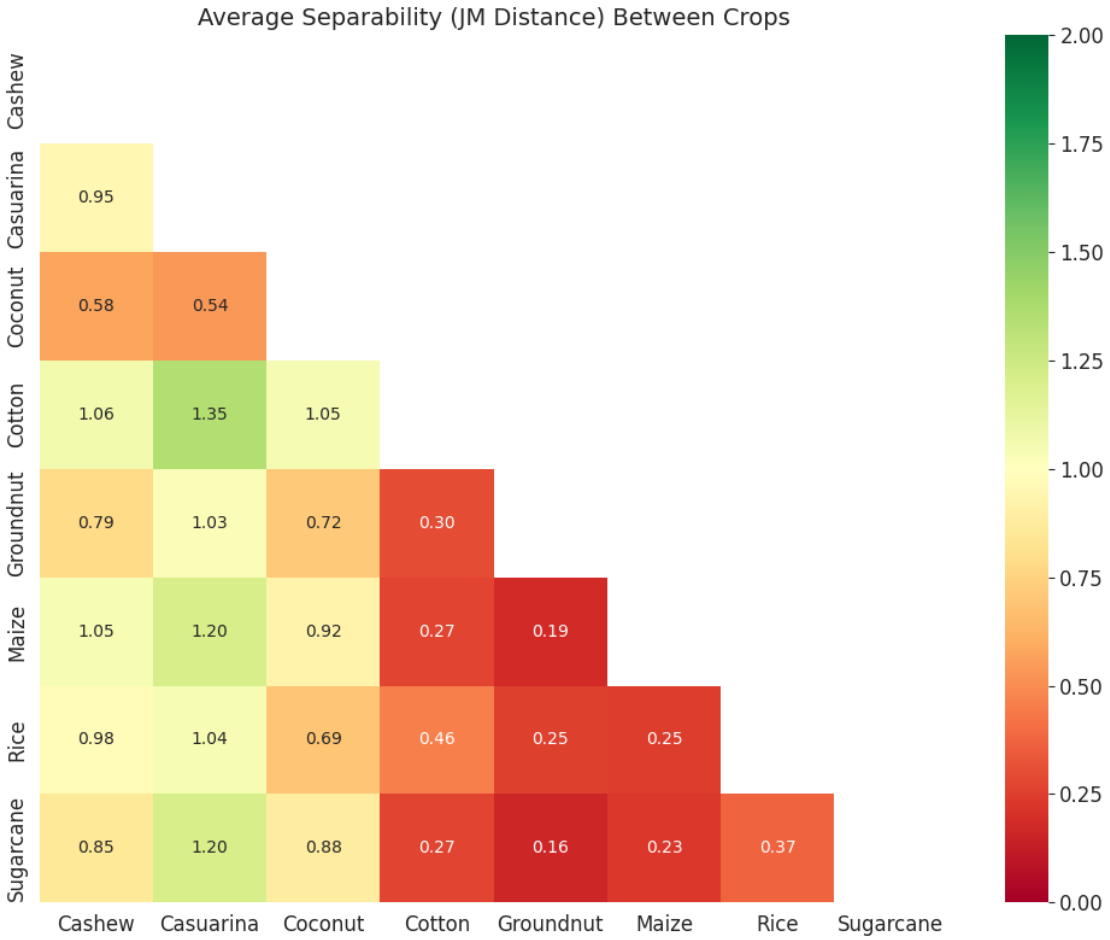


Figure 7 a. Average JM distance between crops pairs for Rabi Sesaon (2024-2025)

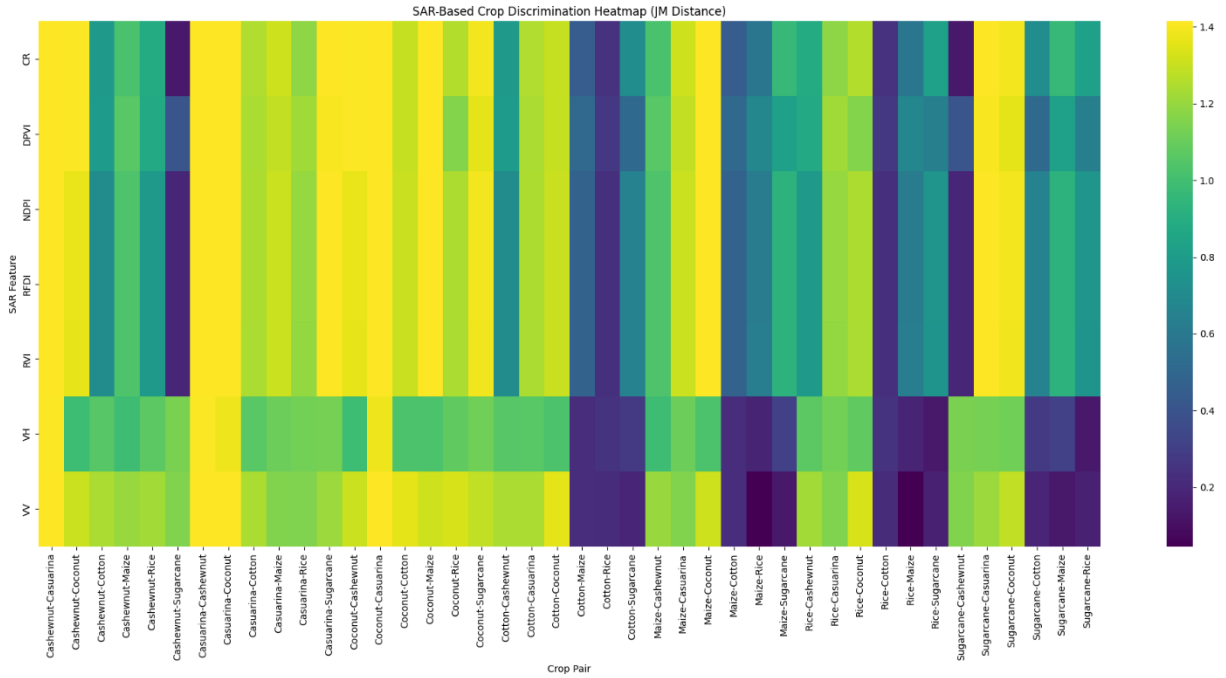


Figure 8 b. SAR-based crop discrimination heatmap (JM distance)

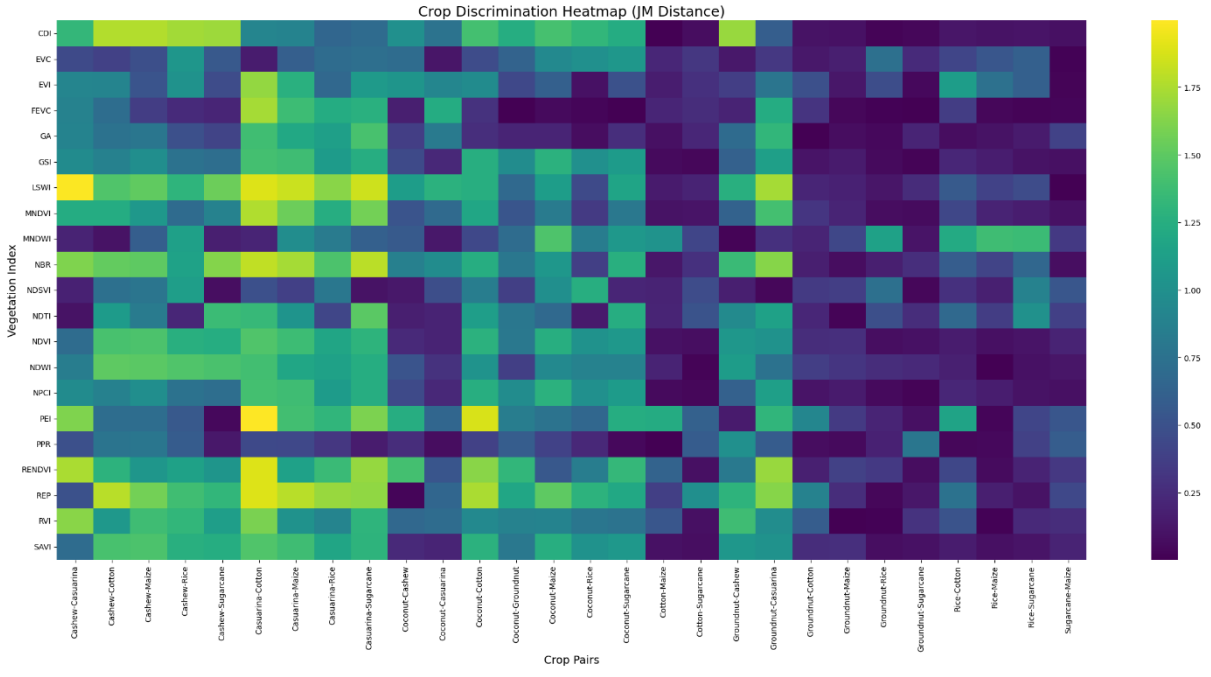


Figure 8 c. Optical-based crop discrimination heatmap (JM distance)

4.5Accuracy Assessment and Validation

The performance of the optimized Random Forest model in classification was quantitatively assessed using independent validation samples, focusing on class-specific producer's accuracy (PA), user's accuracy (UA), and F1-score (Figure 9). The findings revealed significant variability in classification reliability among different crop types, which can be attributed to variations in crop structure, phenology, and spatial distribution. Perennial plantation crops showed the highest classification accuracy. Cashew stood out with a PA of 0.97, UA of 0.90, and an F1-score of 0.94, indicating excellent differentiation and minimal errors of omission and commission. Coconut also achieved high accuracy (PA = 0.83, UA = 0.60, F1 = 0.69), benefiting from its unique canopy structure and consistent spatial patterns. Among seasonal field crops, maize showed strong performance (PA = 0.95, UA = 0.68, F1 = 0.79), whereas rice had moderate accuracy (PA = 0.57, UA = 0.52, F1 = 0.54), likely due to phenological overlap and variable field conditions during the study period. Lower classification accuracies were noted for crops with fragmented spatial distributions and spectral similarities. Cotton (F1 = 0.56) and groundnut (F1 = 0.57) exhibited moderate performance, reflecting confusion with other seasonal crops at similar growth stages. Sugarcane and Casuarina recorded relatively low F1-scores (0.20 and 0.14, respectively), mainly due to limited sample representation and spectral–structural overlap with surrounding vegetation. Overall, the

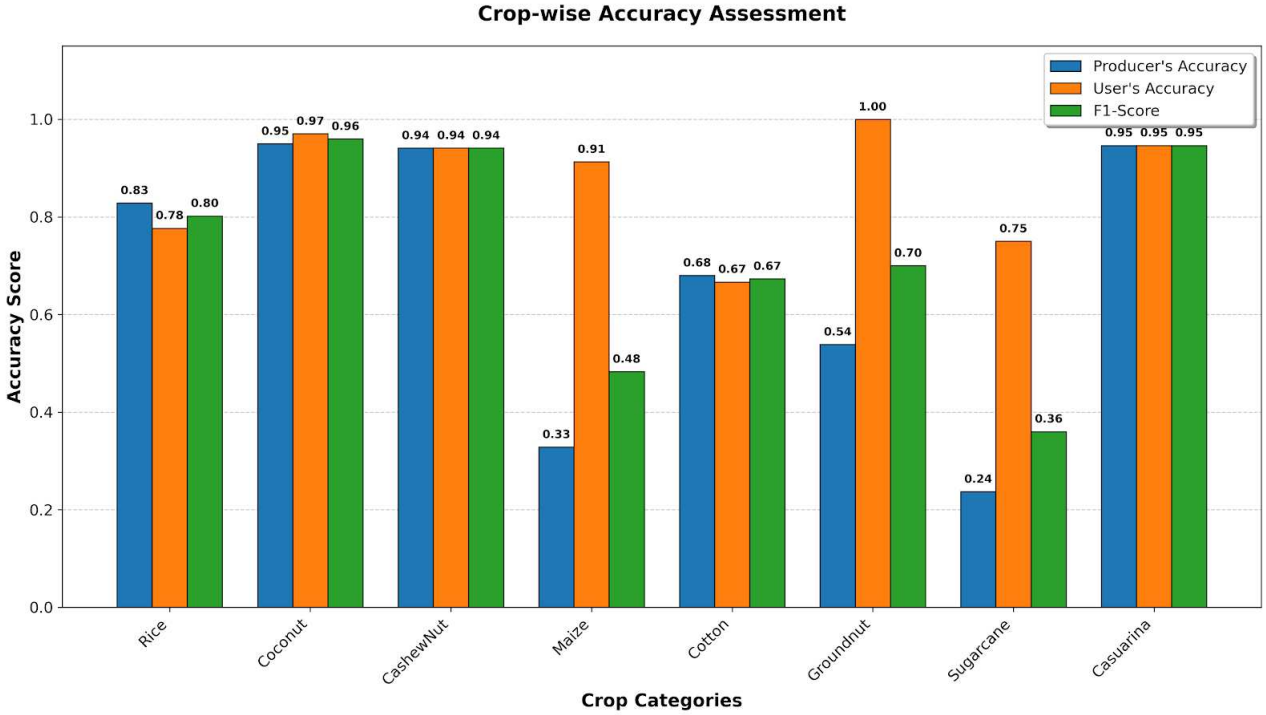


Figure 8 Crop-wise classification performance of the optimized Random Forest model showing Producer's Accuracy (PA), User's Accuracy (UA), and F1-score for major crop types for Rabi Sesaon (2024-2025).

accuracy assessment confirmed that the optimized optical–SAR Random Forest framework performed reliably for dominant and structurally distinct crops, while highlighting challenges in distinguishing minor or spectrally similar crop classes. The combined use of PA, UA, and F1-score offers a comprehensive evaluation of classification reliability and supports the robustness of the proposed methodology for operational multi-crop mapping in the future.

4.2 Effect of Hyperparameter Optimization on Model Performance

Optimizing the hyperparameters had a significant impact on the performance of the Random Forest (RF) classifier. (Figure 10)Figure 10 shows the two-dimensional accuracy surface derived from a grid search over the number of trees and bagging fraction, showing distinct performance variations across different parameter settings. The classification accuracy consistently increased with a larger ensemble size up to a certain point, after which the improvements levelled off or slightly decreased. The peak validation accuracy was reached with a bagging fraction of 0.6 and 140 trees, indicating an ideal balance between ensemble diversity and model generalization. (

Figure 10*) further illustrates the link between the classification accuracy and the number of trees for various bagging fractions. The accuracy increased sharply from 20 to 60 trees, beyond which additional trees yielded diminishing returns, indicating ensemble saturation. Lower bagging fractions (≤ 0.3) led to relatively lower accuracies, whereas moderate to higher bagging fractions (0.5–0.7) consistently delivered better and more stable performance. The optimized setup (60 trees, bag fraction = 0.7) achieved the highest overall accuracy (~0.816) (Fig.11)*

Figure *10*, surpassing non-optimized configurations across the tested parameter space. Overall, the hyperparameter optimization process significantly enhanced the classification robustness and stability by minimizing overfitting and boosting generalization. These findings confirm that the meticulous tuning of RF hyperparameters is crucial for multi-crop

classification using high-dimensional optical–SAR feature spaces. Thus, the chosen optimized configuration was adopted for final crop mapping and subsequent analysis.

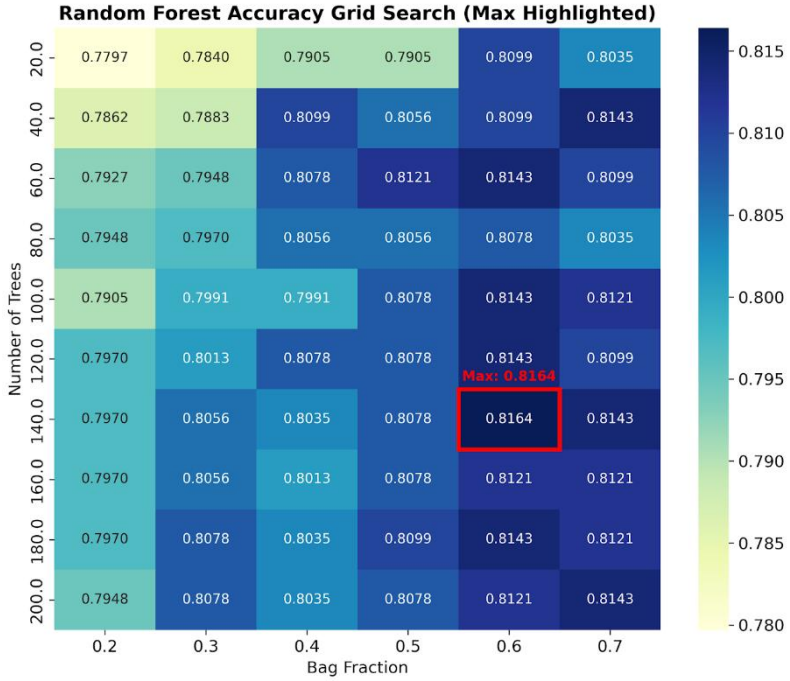


Figure 9 Validation accuracy heatmap showing the combined effect of number of trees and bagging fraction on Random Forest performance Rabi Season (2024-2025).

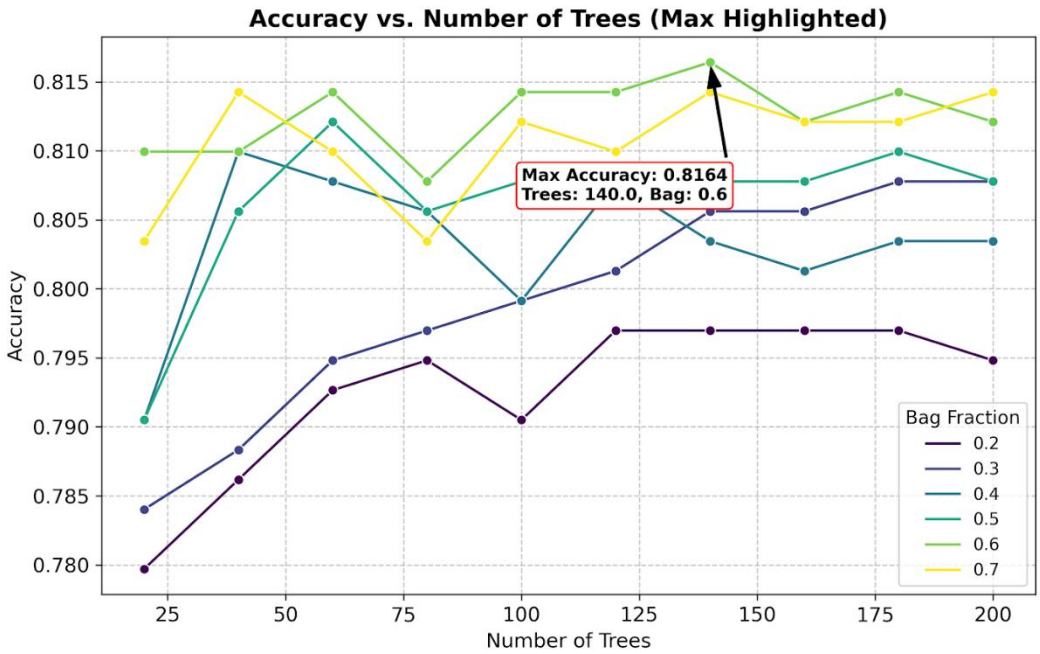


Figure 10 Variation of Random Forest classification accuracy with the number of trees for different bagging fractions, highlighting the optimal model configuration.

4.3 Multi-Crop Classification Results Using Optimized Random Forest and Post-Classification Processing

The improved Random Forest (RF) classifier was first applied to the combined Top-N Optical–SAR predictor stack to produce an initial multi-crop classification map for the study region (Fig.12 and Table 5). This preliminary classification effectively identified major crop types, including rice, cotton, maize, groundnut, sugarcane, coconut, casuarina, and cashew, along with other land-cover classes, such as non-agricultural (forest plantations, fallow land, wasteland, rural/built-up areas, waterbodies, and quarry sites). However, it also revealed localized salt-and-pepper noise and fragmented pixel clusters, particularly along field edges and in areas with diverse cropping patterns. To correct these spatial irregularities, the RF output underwent further refinement through hyperparameter-optimized classification (Fig.13) and

spatial post-processing methods, including cluster filtering and connected pixel analysis(Fig.14). These enhancements led to a spatially coherent and thematically consistent final classification map with clearly defined field shapes and significantly reduced noise**Error! Reference source not found.**.Crop-specific area statistics were calculated from the refined classification by determining the total number of pixels for each crop class and converting them into area estimates in hectares and as percentages of the total cultivated area. The findings show that seasonal field crops, particularly Rice, Cotton, and Maize, were predominant in the agricultural landscape during the study period. In contrast, perennial plantation crops, such as Coconut, Casuarina, and Cashew, were mainly found in contiguous and spatially stable blocks, contributing a smaller but steady portion of the total agricultural area. Sugarcane and Groundnut showed moderate spatial coverage, often linked with irrigated areas and mixed-cropping zones. Overall, the integration of Optical and SAR features within the optimized RF framework, along with post-classification spatial refinement, improved thematic clarity, reduced misclassification among spectrally and structurally similar crops, and provided realistic crop area estimates. The final refined classification and derived area statistics (Fig.14

and Table.6) offer reliable quantitative data for agricultural inventory development, seasonal crop assessment, and regional land-use planning applications.

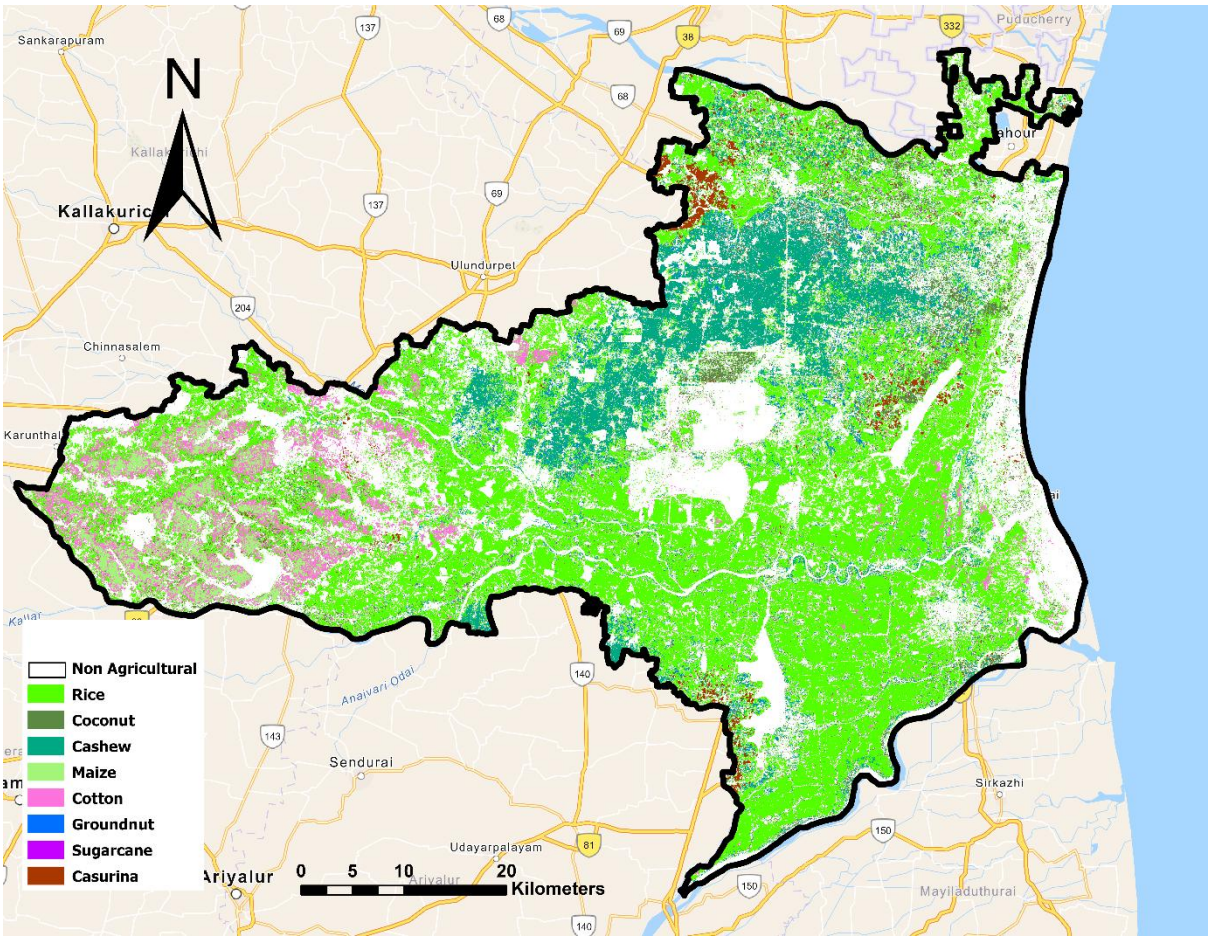


Figure 11 Initial Random Forest–based land use and multi-crop classification in the Cuddalore district derived from fused optical–SAR data for Rabi Season (2024-2025).

Table 5 Crop-wise and land-cover area distribution (ha and %) during the Rabi season derived from Random Forest classification.

S.No	Land Cover Class	Area (ha)	Percentage (%)
1	Rice	144,785.22	39.21%
2	Non-Crop	140,251.68	37.98%
3	Cashew Nut	36,596.00	9.91%
4	Cotton	15,036.01	4.07%

5	Coconut	13,433.53	3.64%
6	Maize	9,619.29	2.61%
7	Casuarina	5,096.04	1.38%
8	Groundnut	3,046.14	0.83%
9	Sugarcane	1,406.99	0.38%
Total	All Classes	369,270.90	100.00%

*Non-Crop: Includes Fallow, Forest, Quarry, Rural areas, Waterbodies, and Wasteland

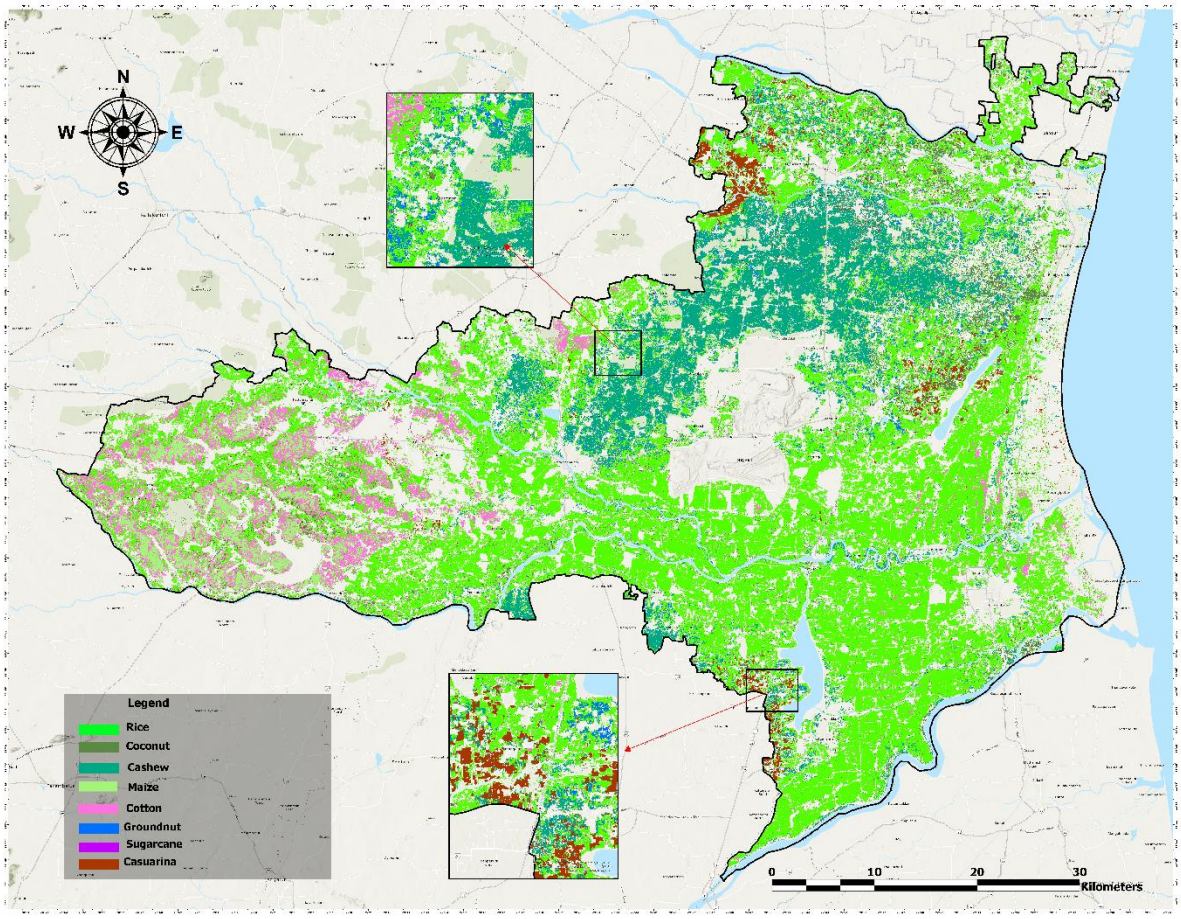


Figure 12 Initial Random Forest–based multi-crop classification using fused Optical–SAR predictors for Rabi Season (2024-2025).

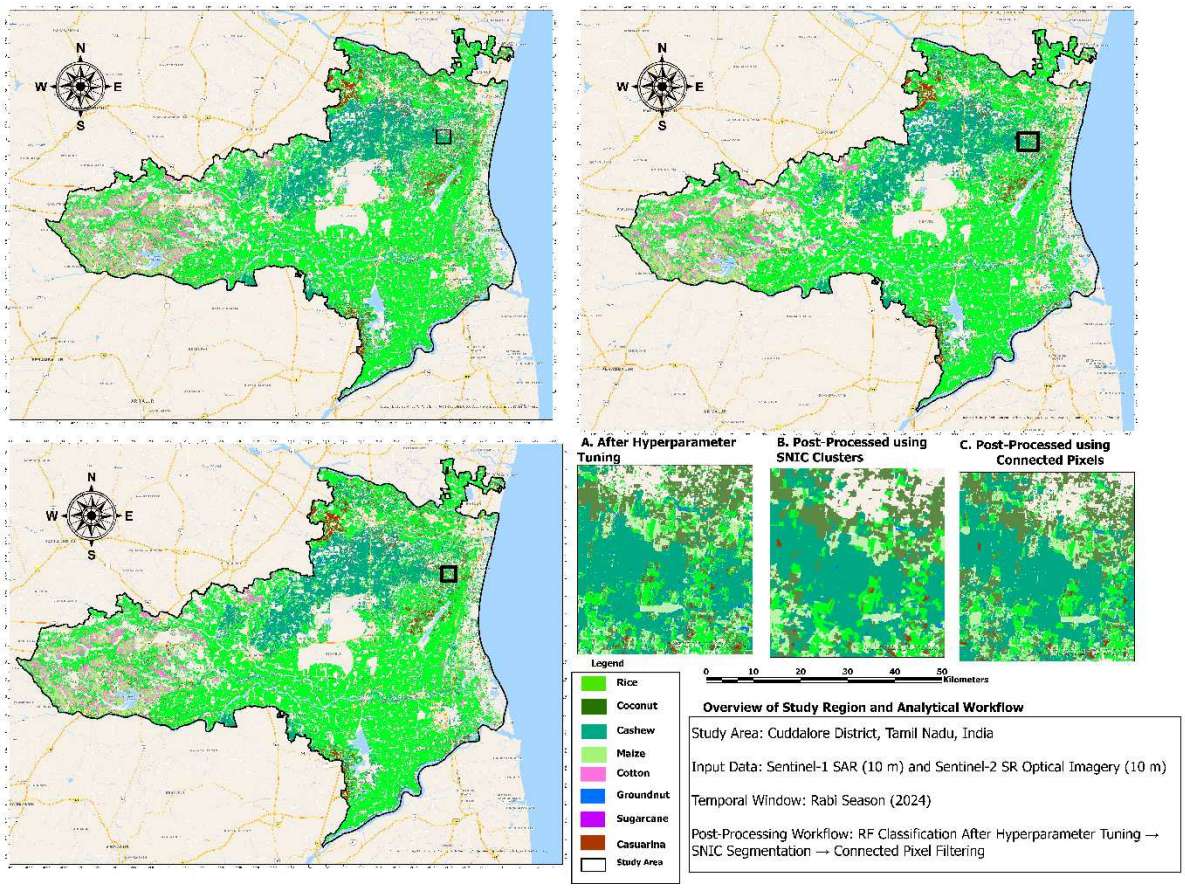


Figure 13 Optimized and post-processed Random Forest multi-crop classification for Rabi Season (2024-2025).

Table 6. Crop-wise area distribution (ha and %) for the Rabi season 2024 -2025 derived from the optimized Random Forest classification.

Crop	Area (ha)	Area (%)
Rice	133,547.42	64.09
Cashew Nut	34,022.23	16.33
Cotton	13,833.16	6.64
Coconut	9,068.76	4.35
Maize	8,955.22	4.3
Casuarina	4,807.84	2.31
Groundnut	2,875.65	1.38
Sugarcane	1,168.53	0.56
Total	208,278.82	100

4. Discussion and limitations

4.1 Discussion

This study illustrates that combining Sentinel-2 multispectral data with Sentinel-1 SAR observations in a refined Random Forest (RF) model significantly improves the ability to distinguish between multiple crops in diverse and mixed-cropping agricultural areas. The enhanced classification results from effectively utilizing spectral, structural, and moisture-sensitive data, which are crucial for accurate crop mapping in cloud-covered and intricate agro-ecosystems(Gašparović & Dobrinić, 2020; Meroni et al., 2021; Sun et al., 2019).The analysis of feature importance highlighted that vegetation indices derived from Sentinel-2, especially those based on greenness and red-edge metrics, such as NDVI, EVI, RENDVI, REP, and EVC, were the main factors contributing to classification precision. These indices are highly responsive to crop biochemical characteristics, canopy health, and growth stage dynamics, facilitating the effective differentiation of crops with overlapping growth phases during the Rabi season(Hatfield & Prueger, 2010; Nagai et al., 2012).

Additionally, indices related to moisture and stress, such as the LSWI, NDWI, NBR, and NDTI, improved class separability by capturing differences in irrigation methods, aging, and surface disruptions, which are vital in intensively managed farming systems. While optical features showed greater relative importance, Sentinel-1 SAR predictors provided complementary and unique information, especially for crops with similar spectral characteristics. SAR backscatter coefficients (VV and VH) and polarization-based indices, such as RVI, CR, and NDPI, enhanced differentiation by detailing canopy structure, biomass distribution, and surface roughness, while also providing consistent observations despite persistent cloud cover(Ha Tran & Ha Tuan, 2025; Rasul, 2025).This finding aligns with previous research emphasizing the value of combining SAR and optical data for crop classification in monsoon-affected and coastal areas. The RF-based Top-N feature selection method effectively reduced the number of predictors while maintaining the classification accuracy, showing that a small set of highly informative features can capture most of the class-discriminative information. This RF-guided dimensionality reduction has been proven to enhance computational efficiency, minimize redundancy, and improve model generalization in high-dimensional remote sensing applications(Orynbaikyzy et al., 2020; Sukawattanavijit et al., 2017).Moreover, optimizing the hyperparameters of the RF classifier increased the model stability and resistance to overfitting, allowing for the robust differentiation of crops with

complex growth patterns and mixed cropping practices. Overall, the proposed optical–SAR fusion framework offers a scalable and dependable solution for operational multi-crop mapping and agricultural monitoring.

4.2 Limitations

Although the results are promising, it is important to recognize several limitations. First, the analysis depended on seasonal median composites, which might obscure short-term phenological changes and lessen sensitivity to quick crop transitions or varied sowing dates. Utilizing dense time-series analysis and specific phenological metrics can enhance crop differentiation, particularly for crops with similar seasonal patterns. Second, the study employed Level-1 NRSC LULC polygons mainly as a crop-masking tool rather than as direct training inputs. Although this method helped limit the classification to agricultural zones, inaccuracies or timing discrepancies in the LULC data could lead to boundary uncertainties. Future research should incorporate higher-resolution or field-validated crop masks to enhance spatial accuracy. Third, although Random Forest provides strong performance and interpretability, it does not explicitly account for temporal dependencies. More advanced methods, such as temporal ensemble learning, harmonic regression-based classifiers, or deep learning models (such as LSTM or Transformer-based architectures), could be investigated to better capture crop phenology, particularly in areas with multiple cropping cycles. Finally, the transferability of the proposed framework across different agro-climatic zones was not specifically tested. Although the methodology is sensor-agnostic and scalable, its effectiveness across diverse soil types, crop calendars, and management practices requires further exploration using multiregional datasets. Overall, addressing these limitations through improved temporal modelling, better reference data, and cross-regional validation would enhance the applicability of the proposed optical–SAR framework for large-scale and operational crop monitoring.

Conclusions

This study established and executed a comprehensive framework combining optical–SAR multi-index and machine learning techniques to accurately distinguish multiple crops in diverse and mixed-cropping agricultural settings. By incorporating multi-temporal Sentinel-2 multispectral data, Sentinel-1 SAR observations, Random Forest (RF) classification, feature

importance analysis, and top-N predictor selection within the Google Earth Engine (GEE) platform, the proposed method effectively addressed issues related to cloud interference, mixed cropping patterns, and overlapping crop growth stages.

The following key conclusions can be drawn from this study:

1.Effectiveness of optical -SAR data fusion

The combination of Sentinel-2 multispectral indices with Sentinel-1 SAR backscatter and polarization-based indices greatly enhanced the ability to distinguish between different crops compared with using a single sensor. Optical indices, especially those based on greenness and red-edge metrics, such as NDVI, EVI, RENDVI, REP, and EVC, were particularly successful in capturing the biochemical properties and phenological changes of crops during the Rabi season. Additionally, SAR features offer strong structural and moisture-related insights, facilitating accurate crop mapping even in cloud-prone coastal areas.

2.Robustness of the Optimized Random Forest Framework

The Random Forest classifier exhibited excellent capability in managing high-dimensional, multi-sensor feature spaces and intricate class boundaries that emerge from mixed cropping systems. Optimizing the hyperparameters further improved the model's stability and ability to generalize, whereas the ensemble-based decision-making process effectively minimized the sensitivity to noise, outliers, and correlated predictors.

3. Importance-Guided Feature Selection and Dimensionality Reduction

RF-based feature importance analysis revealed the relative contributions of Optical and SAR predictors, highlighting the dominant role of multispectral vegetation indices while confirming the complementary value of SAR features. The Top-N predictor selection strategy successfully reduced feature dimensionality without compromising classification accuracy, improving computational efficiency and interpretability, while retaining the most discriminative information for crop mapping.

4.Spatial Consistency through Post-Classification Processing

By employing post-processing methods, such as majority filtering and connected-pixel cleaning, the spatial consistency of the classified crop areas was significantly enhanced. These techniques effectively reduced salt-and-pepper noise and eliminated isolated errors, thereby increasing the thematic accuracy of the final multi-crop classification map, which is now more suitable for practical agricultural use.

5.Operational Applicability for Coastal and Mixed-Cropping Regions

The suggested framework is both computationally efficient and scalable, relying entirely on freely accessible satellite data and cloud-based processing. Its ability to incorporate spectral, structural, and moisture-sensitive data makes it particularly suitable for crop mapping in coastal and intensively managed agricultural areas, which are often characterized by frequent cloud cover, smallholder farming, and varied cropping practices.

Although the proposed method performs well, it depends on representative ground reference samples for supervised classification. Future studies should investigate semi-supervised or transfer-learning methods to reduce reliance on field data and enhance the adaptability of the model across different regions and seasons. Moreover, integrating more detailed temporal phenological metrics and expanding the framework to monitor crops over multiple years could further boost its usefulness in long-term agricultural evaluation and policymaking. Overall, this study illustrates that the combined use of optical–SAR data, RF-based feature optimization, and spatial post-processing offers a strong and dependable solution for distinguishing multiple crops, providing valuable insights for precision agriculture, crop monitoring, and sustainable land management.

Acknowledgements

The authors express their sincere gratitude to the Tamil Nadu Agricultural University (TNAU), Coimbatore, for providing academic support, access to laboratory facilities, and institutional resources that enabled the successful completion of this study. The authors also thank the Center for Water and Geo Spatial Studies (CWGS) and the Department of Remote Sensing and GIS for their technical guidance, access to geospatial datasets, and continuous encouragement during the preparation of this manuscript. The constructive feedback and suggestions provided by colleagues and domain experts are gratefully acknowledged.

Declaration of Interest statement

Funding: This research did not receive funding from any source.

Ethics approval: The manuscript was conducted in an ethical manner, as advised by the

targeted journal.

Consent to participate: Not applicable.

Clinical trial number: Not applicable.

Consent for publication: The research was scientifically consented to be published.

Conflict of interest: The authors declare no conflicts of interest to disclose.

Code availability: Not applicable.

Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials: Data used in this study were downloaded from vari ous data repositories and platforms through their websites. These data sets are freely available for use.

Human face Consent to publish a declaration: We have carefully reviewed all the images.
in our manuscript and confirmed that no human faces were present.

Authors Contributions

Ajay prakash M: Conceptualization, Supervision, Data curation, Methodology, ReviewEditing **Ragunath K.P:** Methodology, Formal analysis, Investigation, Writing - Original Draft **Pazhanivelan S:** Supervision, Review, and Editing **Muthumanickam D:** Review and Editing **Sivamurugan A.P:** Review and Editing **Vanitha. G:** Review & Editing

References

Chakhar, A., Hernández-López, D., Ballesteros, R., & Moreno, M. A. (2021). Improving the Accuracy of Multiple Algorithms for Crop Classification by Integrating Sentinel-1 Observations with Sentinel-2 Data. *Remote Sensing*, *13*(2), 243. https://doi.org/10.3390/rs13020243

Congalton, R. G., & Green, K. (2019). *Assessing the accuracy of remotely sensed data: Principles and practices* (Third editon). CRC Press. https://doi.org/10.1201/9780429052729

Gašparović, M., & Dobrinić, D. (2020). Comparative Assessment of Machine Learning Methods for Urban Vegetation Mapping Using Multitemporal Sentinel-1 Imagery. *Remote Sensing*, *12*(12), 1952. https://doi.org/10.3390/rs12121952

Gella, G. W., Bijker, W., & Belgiu, M. (2021). Mapping crop types in complex farming areas using SAR imagery with dynamic time warping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *175*, 171–183. https://doi.org/10.1016/j.isprsjprs.2021.03.004

Guo, Z., Qi, W., Huang, Y., Zhao, J., Yang, H., Koo, V.-C., & Li, N. (2022). Identification of Crop Type Based on C-AENN Using Time Series Sentinel-1A SAR Data. *Remote Sensing*, *14*(6), 1379. https://doi.org/10.3390/rs14061379

Ha Tran, P., & Ha Tuan, C. (2025). Multi-Source Fusion Algorithms for Satellite Image Resolution Enhancement in Urban Land Cover Mapping of Ho Chi Minh City, Vietnam. *E3S Web of Conferences*, *626*, 01005. https://doi.org/10.1051/e3sconf/202562601005

Han, D., Liu, S., Du, Y., Xie, X., Fan, L., Lei, L., Li, Z., Yang, H., & Yang, G. (2019). Crop Water Content of Winter Wheat Revealed with Sentinel-1 and Sentinel-2 Imagery. *Sensors*, *19*(18), 4013. https://doi.org/10.3390/s19184013

Hariharan, S., Mandal, D., Tirodkar, S., Kumar, V., Bhattacharya, A., & Lopez-Sanchez, J. M. (2018). A Novel Phenology Based Feature Subset Selection Technique Using Random Forest for Multitemporal PolSAR Crop Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(11), 4244–4258. https://doi.org/10.1109/JSTARS.2018.2866407

Hasituya, Zhongxin, C., Fei, L., & Yuncai, H. (2020). Mapping plastic-mulched farmland by coupling optical and synthetic aperture radar remote sensing. *International Journal of Remote Sensing*, *41*(20), 7757–7778. https://doi.org/10.1080/01431161.2020.1763510

Hatfield, J. L., & Prueger, J. H. (2010). Value of Using Different Vegetative Indices to Quantify Agricultural Crop Characteristics at Different Growth Stages under Varying Management Practices. *Remote Sensing*, *2*(2), 562–578. https://doi.org/10.3390/rs2020562

Holtgrave, A.-K., Röder, N., Ackermann, A., Erasmi, S., & Kleinschmit, B. (2020). Comparing Sentinel-1 and -2 Data and Indices for Agricultural Land Use Monitoring. *Remote Sensing*, *12*(18), 2919. https://doi.org/10.3390/rs12182919

Kabato, W., Getnet, G. T., Sinore, T., Nemeth, A., & Molnár, Z. (2025). Towards Climate-Smart Agriculture: Strategies for Sustainable Agricultural Production, Food Security, and Greenhouse Gas Reduction. *Agronomy*, *15*(3), 565. https://doi.org/10.3390/agronomy15030565

Kogo, B. K., Kumar, L., & Koech, R. (2021). Climate change and variability in Kenya: A review of impacts on agriculture and food security. *Environment, Development and Sustainability*, *23*(1), 23–43. https://doi.org/10.1007/s10668-020-00589-1

Lu, T., Wan, L., & Wang, L. (2022). Fine crop classification in high resolution remote sensing based on deep learning. *Frontiers in Environmental Science*, *10*, 991173. https://doi.org/10.3389/fenvs.2022.991173

Mehraj, J., Ahmed, L., Qayoom, K., Rasool, F., Bhat, R. A., Fayaz, U., Abdullah, U. R., Dar, S. Y., & Mir, A. H. (2024). Climate Smart Agriculture: A Roadmap to Sustainable Food Security. *International Journal of Environment and Climate Change*, *14*(7), 234–253. https://doi.org/10.9734/ijecc/2024/v14i74266

Meroni, M., d'Andrimont, R., Vrieling, A., Fasbender, D., Lemoine, G., Rembold, F., Seguini, L., & Verhegghen, A. (2021). Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and -2. *Remote Sensing of Environment*, *253*, 112232. https://doi.org/10.1016/j.rse.2020.112232

Molotoks, A., Smith, P., & Dawson, T. P. (2021). Impacts of land use, population, and climate change on global food security. *Food and Energy Security*, *10*(1), e261. https://doi.org/10.1002/fes3.261

Nagai, S., Saitoh, T. M., Kobayashi, H., Ishihara, M., Suzuki, R., Motohka, T., Nasahara, K. N., & Muraoka, H. (2012). *In situ* examination of the relationship between various vegetation indices and canopy phenology in an evergreen coniferous forest, Japan. *International Journal of Remote Sensing*, *33*(19), 6202–6214. https://doi.org/10.1080/01431161.2012.682660

Orynbaikyzy, A., Gessner, U., Mack, B., & Conrad, C. (2020). Crop Type Classification Using Fusion of Sentinel-1 and Sentinel-2 Data: Assessing the Impact of Feature Selection, Optical Data Availability, and Parcel Sizes on the Accuracies. *Remote Sensing*, *12*(17), 2779. https://doi.org/10.3390/rs12172779

Rasul, A. (2025). *Crop Classification in Uzbekistan Using Random Forest: Integrating Sentinel-1 SAR and Sentinel-2 Optical Data with Ground-Truth Validation*. In Review. https://doi.org/10.21203/rs.3.rs-6822402/v1

Saad El Imanni, H., El Harti, A., Hssaisoune, M., Velastegui-Montoya, A., Elbouzidi, A., Addi, M., El Iysaouy, L., & El Hachimi, J. (2022). Rapid and Automated Approach for Early Crop Mapping Using Sentinel-1 and Sentinel-2 on Google Earth Engine; A Case of a Highly Heterogeneous and Fragmented Agricultural Region. *Journal of Imaging*, *8*(12), 316. https://doi.org/10.3390/jimaging8120316

Sabbaghi, H., & Fathianpour, N. (2025). Hybrid Harris hawks-optimized random forest model for detecting multi-element geochemical anomalies related to mineralization. *Scientific Reports*, *15*(1), 23662. https://doi.org/10.1038/s41598-025-07534-0

Salas, E. A. L., Subburayalu, S. K., Slater, B., Zhao, K., Bhattacharya, B., Tripathy, R., Das, A., Nigam, R., Dave, R., & Parekh, P. (2020). Mapping crop types in fragmented arable landscapes using AVIRIS-NG imagery and limited field data. *International Journal of Image and Data Fusion*, *11*(1), 33–56. https://doi.org/10.1080/19479832.2019.1706646

Sen, R., Goswami, S., & Chakraborty, B. (2019). Jeffries-Matusita distance as a tool for feature selection. *2019 International Conference on Data Science and Engineering (ICDSE)*, 15–20. https://doi.org/10.1109/ICDSE47409.2019.8971800

Shao, Z., Ahmad, M. N., & Javed, A. (2024). Comparison of Random Forest and XGBoost Classifiers Using Integrated Optical and SAR Features for Mapping Urban Impervious Surface. *Remote Sensing*, *16*(4), 665. https://doi.org/10.3390/rs16040665

Sukawattanavijit, C., Chen, J., & Zhang, H. (2017). GA-SVM Algorithm for Improving Land-Cover Classification Using SAR and Optical Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, *14*(3), 284–288. https://doi.org/10.1109/LGRS.2016.2628406

Sun, C., Bian, Y., Zhou, T., & Pan, J. (2019). Using of Multi-Source and Multi-Temporal Remote Sensing Data Improves Crop-Type Mapping in the Subtropical Agriculture Region. *Sensors*, *19*(10), 2401. https://doi.org/10.3390/s19102401

Thakur, Y., Sharma, A., & Sharma, V. K. (2024). A Review on the Relationship of Climate Variability and Extremes with Crop Production. *International Journal of Environment and Climate Change*, *14*(4), 513–529. https://doi.org/10.9734/ijecc/2024/v14i44136

The Bucharest University of Economic Studies, Istudor, N., Ion, R. A., The Bucharest University of Economic Studies, Petrescu, I. E., The Bucharest University of Economic Studies, Hrebenciuc, A., & The Bucharest University of Economic Studies. (2019). Agriculture and the Twofold Relationship between Food Security and Climate Change. Evidence from Romania. *Www.Amfiteatrueconomic.Ro*, *21*(51), 285. https://doi.org/10.24818/EA/2019/51/285

Wei, S., Zhang, H., Wang, C., Wang, Y., & Xu, L. (2019). Multi-Temporal SAR Data Large-Scale Crop Mapping Based on U-Net Model. *Remote Sensing*, *11*(1), 68. https://doi.org/10.3390/rs11010068

Zhang, X., Xu, J., Chen, Y., Xu, K., & Wang, D. (2021). Coastal Wetland Classification with GF-3 Polarimetric SAR Imagery by Using Object-Oriented Random Forest Algorithm. *Sensors*, *21*(10), 3395. https://doi.org/10.3390/s21103395

1147

1148

1149

1150

1151

1152