Satellite-based Crop Discrimination with Machine Learning on GEE Platform: Insights from Udham Singh Nagar

Nidhi Chaudhary[1,2], Koneenika Mallick[1,2], Abhishek Danodia[1], Kamal Pandey[1*], Nagendra S.R.[1]
[1] Indian Institute of Remote Sensing (IIRS)-ISRO, Dehradun, Uttarakhand, India.
[2]2Dhirubhai Ambani University (DAU), Gandhinagar, Gujarat, India.
*Corresponding Author email: kamal@iirs.gov.in

(Received on 02 March 2026; In final form on 20 April 2026)

DOI: *https://doi.org/10.58825/jog.2026.20.1.317*

Abstract: Crop discrimination is crucial for environmental monitoring, agricultural planning, and sustainable development. This study assessed the performance of optical (Sentinel-2, 10 m, atmospherically corrected to surface reflectance) and microwave (Sentinel-1, 10 m, preprocessed with radiometric calibration and speckle filtering) remote sensing data for crop classification in Udham Singh Nagar district, Uttarakhand, India, during the June–October 2023 kharif season. Ground truth data for major crops, namely rice (815 samples) and sugarcane (62 samples), were collected through field surveys and split into 70% training and 30% validation subsets to ensure robust model evaluation. Five machine-learning classifiers Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Classification and Regression Tree (CART), and Gradient Boosted Machine (GBM) were applied to individual and fused datasets. GBM consistently achieved the highest classification accuracy at the monthly scale, likely due to its sequential error-correction mechanism that effectively exploits distinct phenological patterns captured in monthly temporal composites, while RF produced the highest overall accuracy (89.21%) for the season-long fused optical and microwave dataset. SVM and KNN showed comparatively lower performance, especially during transitional crop growth stages. The results highlight the effectiveness of ensemble learning methods and demonstrate the benefit of multi-sensor data fusion for accurate and reliable crop discrimination and land use/land cover mapping.

Keywords: Crop discrimination, Machine learning, Google Earth Engine, Rice, Sugarcane

1. Introduction

Monitoring crop classification and land use/land cover (LULC) dynamics is crucial for agricultural management, urban planning, biodiversity conservation, and disaster mitigation. Remote sensing enables timely and large-scale crop mapping through frequent observations. Optical sensors such as Sentinel-2 provide rich spectral information but are limited by cloud cover. In contrast, microwave sensors like Sentinel-1 penetrate clouds and capture structural information, enabling consistent monitoring (Immitzer et al. 2016; Eisfelder et al. 2024).

Recent studies have improved classification accuracy by fusing optical and microwave data with machine learning algorithms on platforms like Google Earth Engine (GEE). Eisfelder et al. (2024) showed that combining Sentinel-1 SAR with Sentinel-2 time series enhances crop-type discrimination in cloud-prone conditions by integrating structural and spectral features. Multi-temporal analysis further improves results by capturing phenological variations across crop cycles. Ensemble models such as Random Forest (RF) and Gradient Boosted Machines (GBM) are particularly effective for handling high-dimensional, multi-temporal data (Vuolo et al. 2018). Here, "crop classification" refers to assigning crop labels to pixels, whereas "crop discrimination" emphasizes distinguishing between similar crop types under challenging conditions. Despite these advances, few studies have evaluated the combined use of Sentinel-1 and Sentinel-2 across the full growing season in cloud-affected regions like Udham Singh Nagar. Rusňák et al. (2023) highlighted the need for systematic comparison of classifiers across temporal composites. Comparative analyses of multiple classifiers (SVM, KNN, CART, RF, GBM) using both individual and fused datasets remain limited.

Therefore, this study evaluates and compares five machine learning classifiers using optical and microwave data, individually and combined, for crop and LULC classification in Udham Singh Nagar, Uttarakhand, during the Kharif season (June–October 2023). It also assesses the role of ensemble methods and temporal data fusion in improving classification accuracy.

Sentinel-2 data at 10 m spatial resolution were atmospherically corrected to obtain surface reflectance, while Sentinel-1 SAR data (10 m) were preprocessed using standard procedures including radiometric calibration and speckle filtering to improve data quality and consistency. A total of 815 rice and 62 sugarcane ground truth samples were collected; these were systematically divided into training and validation datasets to ensure robust and unbiased model evaluation.

2. Material and Methods

2.1 Site Description

Udham Singh Nagar district, located in the Tarai region of the Shivalik range in Uttarakhand, India, covers about

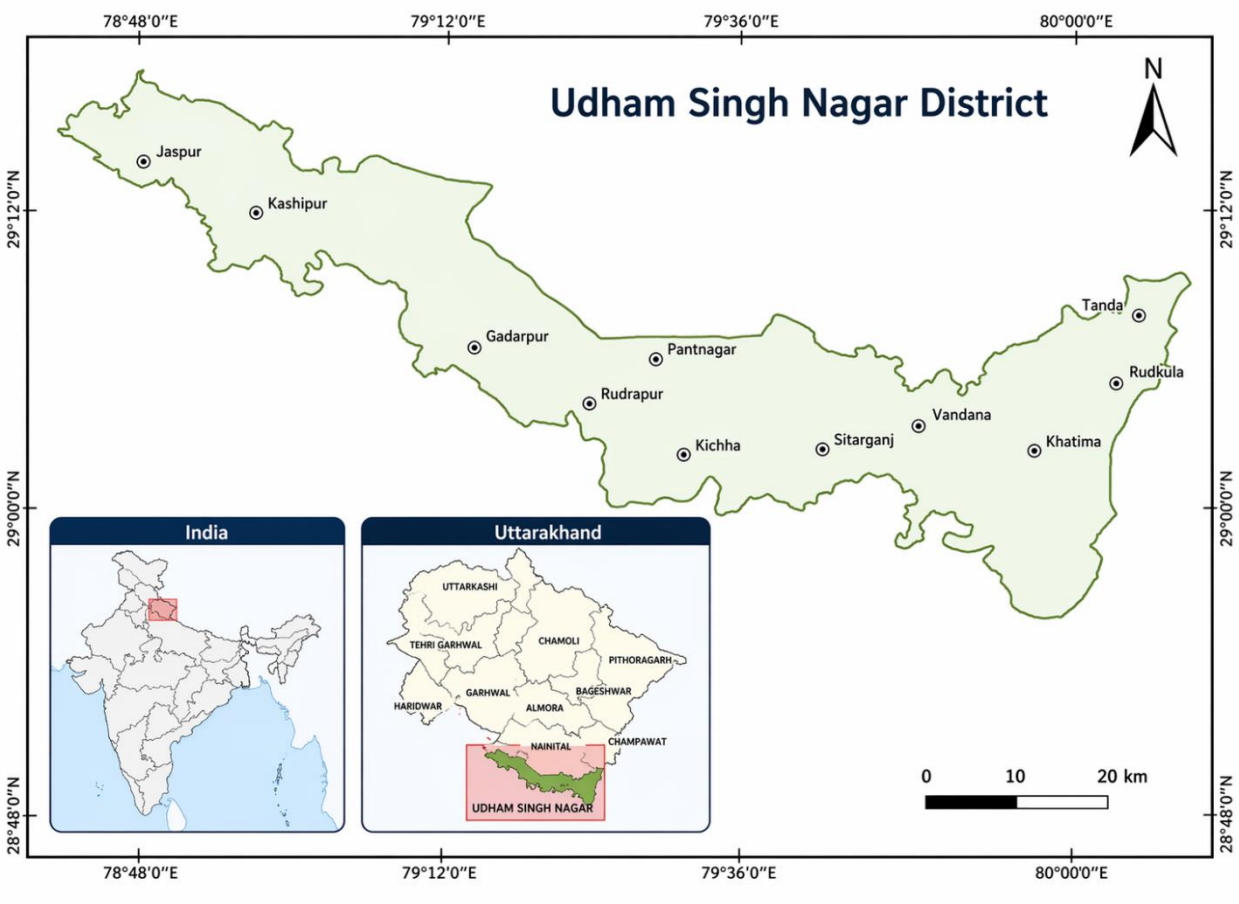

Figure 1: Geographical Extent of the Study Area

2536 km² and lies at an elevation of 210–244 m above mean sea level (Figure 1). It has a humid subtropical climate with distinct summer, monsoon, and winter seasons, receiving most rainfall during June–September. Fertile alluvial soils, seasonal waterlogging, and monsoon rainfall strongly influence LULC dynamics and cropping patterns, with rice and sugarcane as the dominant crops.

2.2 Data Collection

This study utilized both optical and microwave remote sensing data to achieve accurate crop classification. The optical data was sourced from Sentinel-2 imagery, with six spectral bands selected: B2 (Blue), B3 (Green), B4 (Red), B8 (Near Infrared), B11 (Shortwave Infrared 1), and B12 (Shortwave Infrared 2). These bands were chosen for their effectiveness in capturing vegetation and surface characteristics, and the temporal window for data collection spanned from June 1 to October 31, 2023, covering the entire cropping season. In parallel, microwave data was acquired from Sentinel-1 imagery, specifically focusing on the VH (Vertical transmit, Horizontal receive) polarization, which provides structural information regardless of cloud cover or lighting conditions. The same temporal range was maintained for the Sentinel-1 data to ensure consistency in temporal analysis.

Ground truth data were collected through field surveys for two primary crops i.e. rice and sugarcane and other land cover classes such as water bodies, settlements, forests and bare land, which were combined manually delineated using ground survey and high-resolution satellite imagery to strengthen classification accuracy and training data reliability. The entire classification and analysis process was carried out using Google Earth Engine (GEE), a cloud-based geospatial processing platform, in conjunction with GIS for spatial visualization and validation.

3. Methodology

The workflow involved acquiring Sentinel-2 (bands B2, B3, B4, B8, B11, B12) and Sentinel-1 (VH) data for June to October 2023, followed by preprocessing steps that included cloud filtering and generation of monthly median composites for Sentinel-2, and GRD preparation with monthly median composites for Sentinel-1. Feature engineering was performed by appending vegetation and built-up indices such as NDVI, EVI, and NDBI to the spectral and backscatter stacks. Sample preparation utilized field-surveyed rice and sugarcane plots along with manually delineated land use/land cover classes, with 70 and 30 train-test split ratio. Supervised classification was carried out using SVM, KNN, CART, RF and GBM models under both optical-only and Sentinel-1 plus Sentinel-2 fusion setups, across monthly and cumulative time windows. Accuracy assessment included evaluating per-month and cumulative overall accuracy and comparisons across classifiers. To improve prediction stability, ensemble majority voting was applied, followed by post-classification mapping and area estimation. The methodology of the study is depicted in Figure 2.

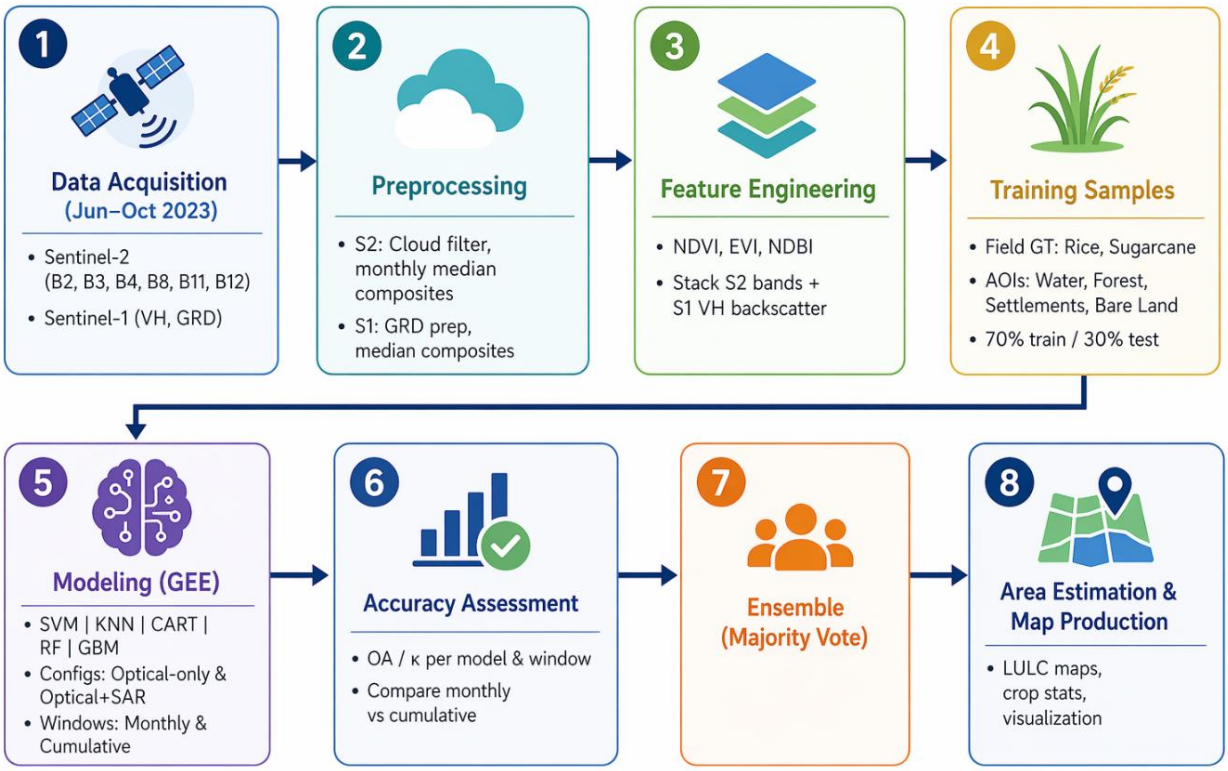

Figure 2: Methodology Flowchart

All data processing, classification, and accuracy assessments in this study were conducted using the cloud-based Google Earth Engine (GEE) platform. GEE provides powerful tools for handling large-scale Earth observation datasets and allows the seamless integration of optical and microwave imagery for supervised classification tasks. This filtering minimized atmospheric interference while ensuring sufficient temporal resolution. Monthly median composites were generated to further reduce residual cloud effects and anomalous reflectance values. Like the optical data, Sentinel-1 GRD products were mosaicked into monthly composites using a median reducer, which helped suppress speckle noise and preserved the overall surface texture and pattern useful for classification. To enrich the feature space for classification, several widely used remote sensing indices were computed directly within GEE. These included the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and Normalized Difference Built-up Index (NDBI).

3.1 Classification Workflow

Five machine learning classifiers were implemented using GEE's built-in ML libraries, allowing all computations to be conducted in the cloud without requiring local downloads or heavy computation. These classifiers were as followed:

Random Forest (RF): A widely used ensemble learning method that builds multiple decision trees from randomly sampled data and features. In GEE, the <ee.Classifier.smileRandomForest> function was used. RF is known for its robustness against overfitting and its ability to handle high-dimensional data effectively. Its internal feature importance ranking also makes it suitable for interpreting the contribution of different bands and indices. RF consistently delivered high accuracy in this study, especially when combining temporal and multi-sensor inputs.

Support Vector Machine (SVM): It was implemented using <ee.Classifier.libsvm>, this non-parametric classifier creates optimal hyperplanes to separate different land cover classes in a high-dimensional feature space. SVM is effective when the training data is limited but can be sensitive to overlapping class boundaries. In our results, SVM performed reasonably well but struggled in the post-harvest months when class separability was reduced due to spectral similarities.

K-Nearest Neighbors (KNN): A simple, instance-based classifier implemented using <ee.Classifier.smileKNN>. KNN classifies each pixel based on the majority vote of its k-nearest neighbors in the feature space. While computationally lightweight, its performance is influenced by class density and spatial heterogeneity. KNN was less effective in mixed-class regions but still provided useful complementary predictions for ensemble modeling.

Classification and Regression Trees (CART): This decision tree-based model, implemented via <ee.Classifier.smileCart>, constructs a hierarchy of decisions by splitting the dataset based on feature thresholds. CART models are interpretable and fast but tend to overfit if not carefully tuned. In this study, CART performed moderately well and added diversity to the ensemble classifier.

Gradient Boosted Machine (GBM): A sequential ensemble method where each new model aims to correct the errors of the previous one. In GEE, GBM was implemented using <ee.Classifier.gbdt>. This classifier is

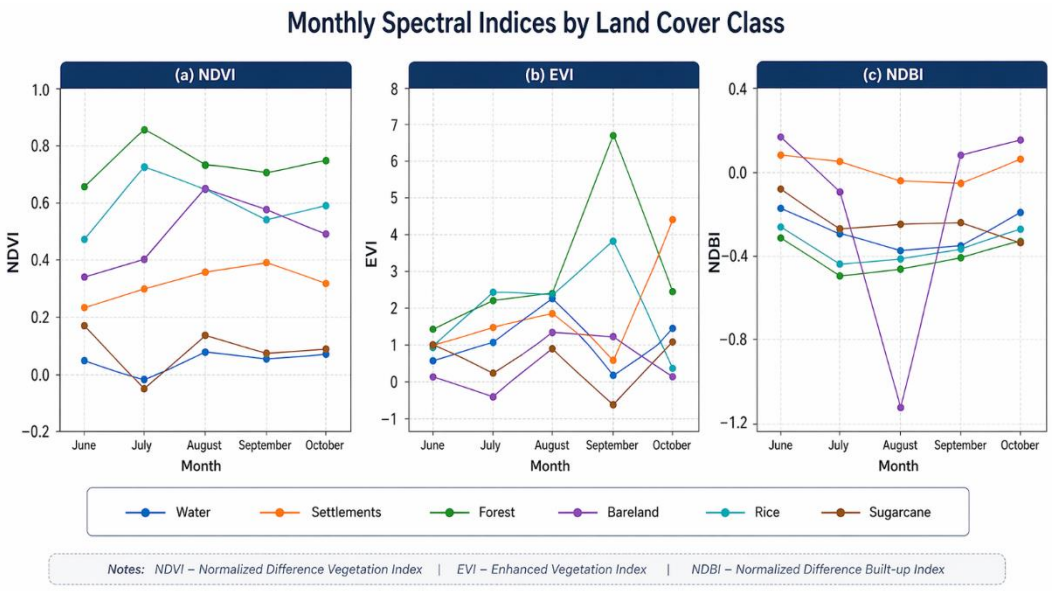

Figure 3: Monthly Variation of NDVI, EVI, and NDBI in the Study Area

Table 1: Monthly and Multi-Month Classification Accuracy (%) with only Optical Data

Month	RF Accuracy (%)	SVM Accuracy (%)	KNN Accuracy (%)	CART Accuracy (%)	GBM Accuracy (%)
June	75.91	69.68	75.69	69.68	76.77
July	86.86	81.51	85.08	83.07	88.42
August	85.53	73.24	82.01	79.61	84.87
September	82.99	66.21	77.55	75.28	82.08
October	83.04	72.39	80.65	78.04	85.21
June+July	76.11	68.51	76.95	69.75	78.19
June+July+August	80.65	73.89	77.62	75.52	82.98
June+July+August+September	84.54	72.18	81.23	75.28	85.43
June to October	85.99	82.54	85.13	78.66	85.78

Table 2: Performance (%) of Machine Learning Models Using Integrated Optical and Microwave Data

Month	RF Accuracy (%)	SVM Accuracy (%)	KNN Accuracy (%)	CART Accuracy (%)	GBM Accuracy (%)
June	77.97	73.65	71.27	70.41	77.54
July	88.27	78.89	85.71	85.5	88.49
August	87.1	56.77	80.65	81.94	88.25
September	82.98	74.58	78.36	79.2	84.03
October	80.96	74.9	78.45	77.4	83.47
June+July	82.7	79.11	76.37	80.8	83.12
June+July+August	87.85	85.93	77.18	84.43	88.27
June+July+August+September	89.18	82.78	81.01	83.22	88.14
June to October	89.21	80.95	83.45	86.17	88.66

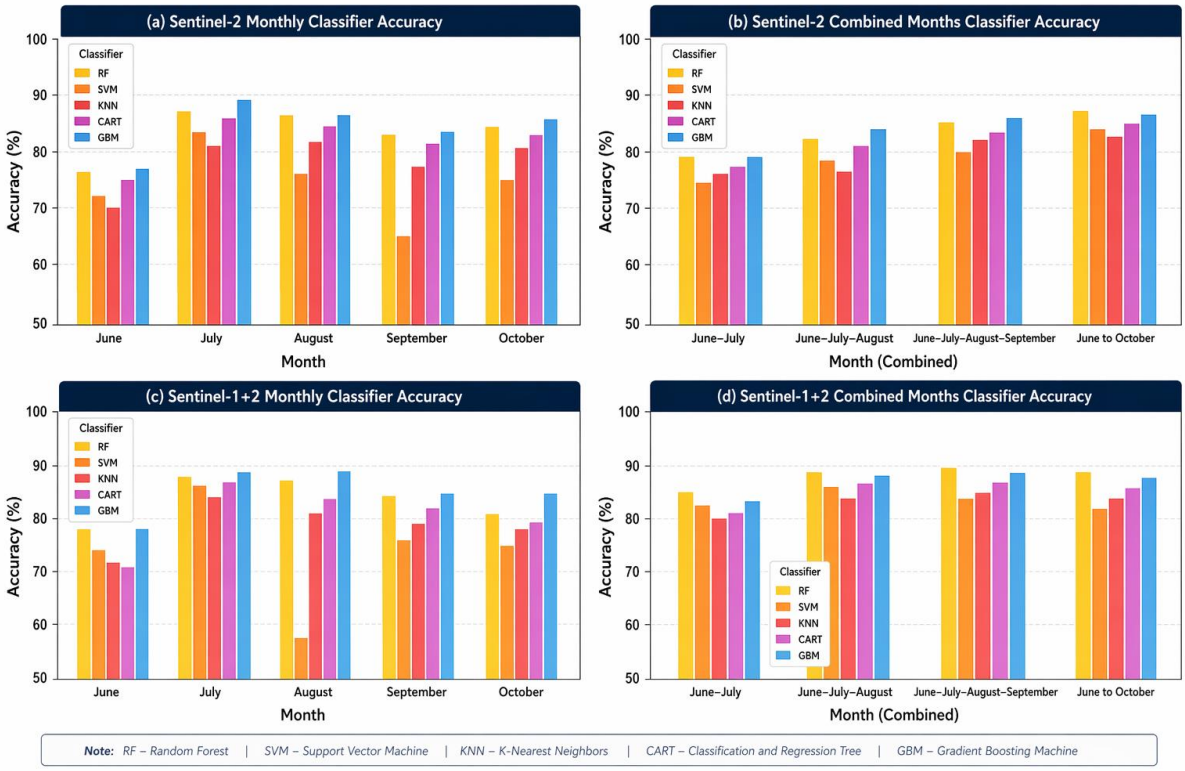

Figure 4: Classifier Accuracy (%) for Sentinel-1 and Sentinel-2 (Monthly and Combined)

known for its high predictive power, especially in cases where subtle patterns must be learned. GBM delivered the highest individual monthly accuracy in most scenarios due to its strength in capturing complex interactions between features over time.

Each classifier was trained and evaluated under two primary configurations: one using only the Sentinel-2 optical dataset and the other using a fusion of Sentinel-2 and Sentinel-1 data. To assess temporal variation, classifications were performed for each month (June, July, August, September, and October) as well as for combined periods (June–July, June–August, June–October). This approach allowed for both time-specific analysis and assessment of the cumulative impact of data integration on classification accuracy.

To further enhance the robustness of the final land cover maps, an ensemble classification approach was adopted. Majority voting was performed across the classification outputs from RF, SVM, CART, and GBM. Each pixel was assigned the most frequently predicted class, leveraging the individual strengths of each model. This ensemble approach reduced individual model biases and consistently yielded the highest overall accuracy, making it highly effective for crop discrimination over dynamic agricultural landscapes like Udham Singh Nagar.

4. Results and Discussion

4.1 Monthly Analysis of Indices

Vegetation indices (NDVI, EVI) and urbanization index (NDBI) were analyzed for each month from June to October 2023 (Figure 3) to assess land surface changes with LULC (Tassi and Vafeidis, 2022; Vuolo et al. 2023). NDVI values across this period consistently reflected the health and density of vegetation. Forest and rice exhibited the highest NDVI values, with peaks occurring in August and September, which aligns with the monsoon season and the primary growth phase of these crops (Shelestov et al. 2017; Abubakar et al. 2023). Sugarcane showed a steady and moderate rise in NDVI throughout the months, indicative of its long-duration, irrigated growth cycle (Achahboun et al. 2022). Water bodies, as expected, recorded the lowest or even negative NDVI values throughout, affirming the index's sensitivity in distinguishing non-vegetative surfaces (Guo et al. 2021). On the other hand, bareland and settlements maintained lower NDVI values, with minor fluctuations possibly due to temporary surface vegetation or environmental factors (Yao et al. 2022).

EVI, which is designed to improve sensitivity in areas of high biomass, especially in densely vegetated zones, amplified the vegetative signal even further (Vuolo et al. 2023). Forest regions displayed remarkably high EVI values, peaking sharply in September, suggesting lush, healthy canopy conditions during this time (Tassi and Vafeidis, 2022). Sugarcane also exhibited elevated EVI values, reflecting its dense and robust foliage. Interestingly, water bodies registered unusually high EVI in October, a deviation that may indicate the presence of floating vegetation, algal blooms, or even potential misclassification errors that warrant further ground verification (Guo et al. 2021). The EVI values for settlements and bareland remained comparatively low, reinforcing their non-vegetated nature.

NDBI values effectively captured the extent of built-up and impervious surfaces (Guo et al. 2021). Settlements and bareland consistently showed positive NDBI values, confirming their association with human-modified

surfaces and lack of vegetation (Yao et al. 2022). Meanwhile, forest, rice, and sugarcane recorded negative NDBI values across all months, consistent with their vegetative land cover (Tassi and Vafeidis, 2022). A significant dip in NDBI values for sugarcane and rice was observed in August and September, coinciding with their periods of maximum canopy cover (Abdali et al. 2023). Anomalies such as the sharp drop in bareland NDBI in August suggest temporary cover or water presence, further demonstrating the dynamic nature of these indices (Guo et al. 2021).

Overall, this spectral analysis across time and land cover types reinforces the value of NDVI, EVI, and NDBI as complementary tools for land use classification, crop monitoring, and environmental assessment (Vuolo et al. 2023; Tassi and Vafeidis, 2022). The trends observed can help inform agricultural planning, urban sprawl detection, and climate-resilient land management practices (Shelestov et al. 2017). For instance, understanding when crops reach peak greenness or when built-up expansion intensifies allows for more targeted policy and conservation decisions (Abubakar et al. 2023). The synthesis of these indices offers a robust approach to monitoring and managing rapidly changing landscapes in a data-driven, seasonally aware manner (Abdali et al. 2023).

These observed peaks in NDVI for rice and forest during August and September align with the findings of Shelestov et al. (2017) and Abubakar et al. (2023), who identified this window as the primary growth phase driven by monsoon rains. The gradual rise in sugarcane NDVI is consistent with the long-duration growth cycle described by Achahboun et al. (2022). The index's sensitivity in distinguishing non-vegetative surfaces, such as water and bareland, mirrors the results reported by Guo et al. (2021) and Yao et al. (2022). Furthermore, the enhanced sensitivity of EVI in high-biomass zones matches the observations of Vuolo et al. (2023) and Tassi and Vafeidis (2022). The NDBI trends, specifically the positive values for human-modified surfaces and the significant dips for crops during peak canopy cover, are supported by the work of Abdali et al. (2023) and Guo et al. (2021).

While analyzing the accuracy of various model (Table 1) it is observed that across all months, the GBM classifier consistently achieved the highest accuracy, peaking in July at 88.42% (Abdali et al. 2023). RF closely followed particularly strong during the peak growing season (June–August), highlighting its robustness for agricultural applications (Vuolo et al. 2023). SVM underperformed relative to other models, particularly during transitional months like September where spectral variability likely increased due to harvesting activities (Yao et al. 2022). Aggregating multiple months improved classification performance significantly, as longer temporal composites captured both phenological development and harvesting, offering richer spectral variation for the classifiers (Tassi and Vafeidis, 2022). GBM remained the top performer overall, while RF achieved excellent performance for the full season (June–October) (Abdali et al. 2023).

The superior performance of the GBM classifier in agricultural mapping is corroborated by Abdali et al. (2023), while the robustness of RF during peak growing seasons aligns with Vuolo et al. (2023). The decline in SVM accuracy during transitional months like September is likely due to increased spectral variability from harvesting activities, a phenomenon noted by Yao et al. (2022). Additionally, the significant improvement observed when using multi-temporal composites reinforces the conclusions of Tassi and Vafeidis (2022), who noted that longer temporal windows capture essential phenological and harvesting variations.

4.2 Integration of Optical and Microwave Data

The accuracy results of the integrated optical and microwave data set for various models is presented in Table 2. It is observer from the Table that the fusion of optical and microwave data improved the monthly accuracies by 2–5% compared to optical-only datasets, particularly under challenging monsoon conditions in agreement with (Guo et al. 2021; Dube et al. 2024). SAR information effectively complemented spectral data, enabling better discrimination of mature crops, harvested fields, and settlements (Tassi and Vafeidis, 2022). GBM again outperformed all other classifiers, peaking at 88.49% in July, followed closely by RF (Abdali et al. 2023). The highest overall accuracy (89.21%) was achieved by Random Forest (RF) when using the full June–October optical + microwave information, demonstrating that multi-sensor integration significantly enhances model robustness over longer periods encompassing complex land cover transitions (Dube et al. 2024; Guo et al. 2021).

The observed accuracy gain through sensor fusion is consistent with the findings of Guo et al. (2021) and Dube et al. (2024), who emphasized that SAR data provides structural information that complements spectral signals. The effectiveness of SAR (specifically VH polarization) in overcoming cloud-induced gaps during the monsoon is supported by Tassi and Vafeidis (2022). Furthermore, the peak performance of the RF model on multi-sensor datasets demonstrates the model's robustness in managing complex land cover transitions over long periods, as highlighted by Dube et al. (2024) and Abdali et al. (2023) (Figure 4).

4.3 Area Estimation

Rice dominated the agricultural landscape, covering approximately 1,77,749 hectares, accounting for nearly 98% of the classified crop area, while sugarcane occupied around 2838 hectares, representing about 2% of the total area (Abubakar et al. 2023; Shelestov et al. 2017).

Ensemble classification:

The ensemble classification, created through majority voting among RF, SVM, KNN, CART, and GBM outputs achieved an overall accuracy of 90.25% demonstrating the strength of integration multiple machine learning models for LULC mapping.

The dominance of rice in the agricultural landscape is in agreement with the regional crop distributions documented by Abubakar et al. (2023) and Shelestov et al. (2017). The success of the ensemble method (90.25%) demonstrates that integrating multiple machine learning models minimizes individual model errors, particularly during

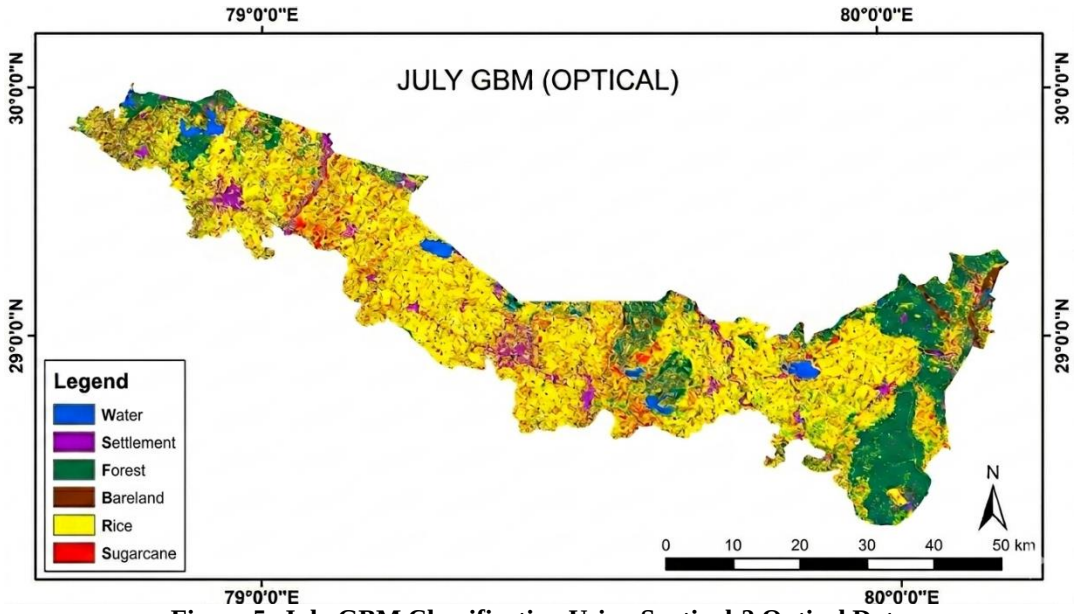

Figure 5: July GBM Classification Using Sentinel-2 Optical Data

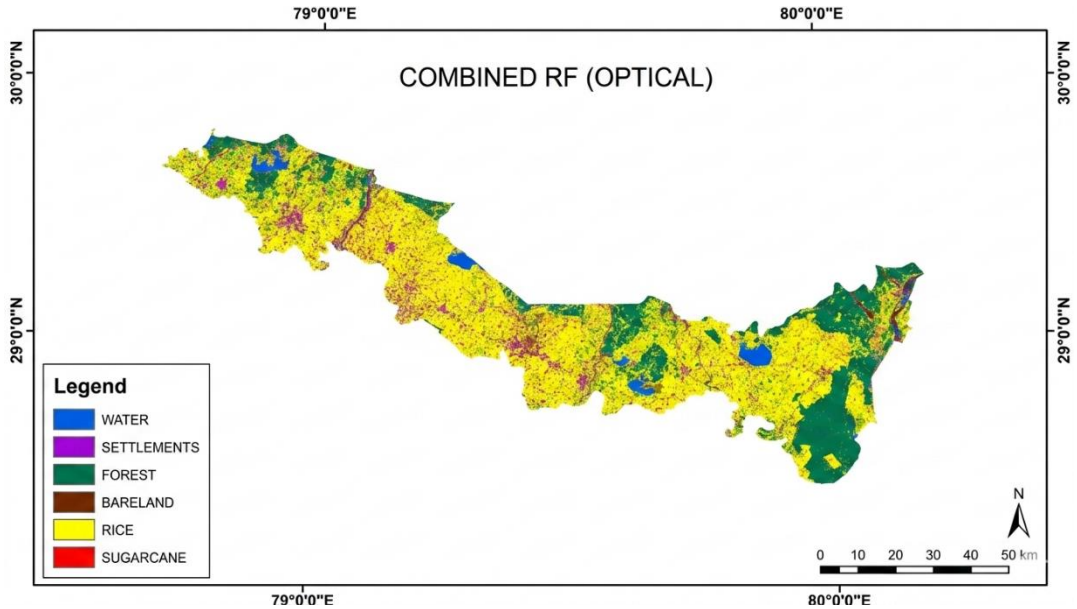

Figure 6: LULC Classification using Random Forest on Optical Data alone

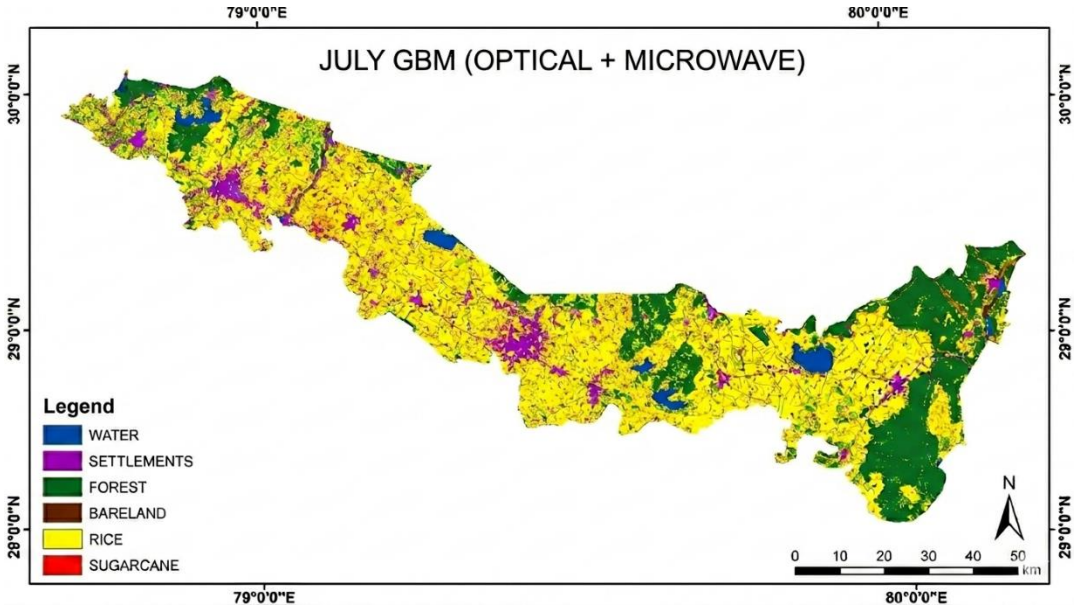

Figure 7: GBM Classification Using Fused Sentinel-1 and Sentinel-2 Data (July)

transitional crop stages, as suggested by Yao et al. (2022) and Abdali et al. (2023). Even though models like KNN and CART showed lower individual accuracies, their contribution to the ensemble provided useful spatial features, a benefit noted by Achahboun et al. (2022).

4.4 Discussion

This study demonstrates the effectiveness of integrating multi-temporal, multi-sensor remote sensing data with machine learning techniques for crop and land cover classification (Tassi and Vafeidis, 2022; Guo et al. 2021). The fusion of Sentinel-1 (SAR) and Sentinel-2 (optical) data significantly enhanced classification accuracy by leveraging both spectral and structural information, especially during cloud-covered monsoon periods (Dube et al. 2024). Major temporal patterns of vegetation indices (NDVI, EVI) and urban indices (NDBI) aligned with crop phenology, highlighting the importance of seasonal dynamics in classification (Vuolo et al. 2023).

Among classifiers, ensemble models like Random Forest and Gradient Boosted Machine outperformed simpler algorithms, achieving a highest overall accuracy of 90.25% (Abdali et al. 2023). Ensemble learning proved effective in minimizing individual model errors, especially during transitional crop stages (Yao et al. 2022). While KNN and CART had lower individual accuracy, they contributed useful spatial features to the ensemble model (Achahboun et al. 2022). The fusion of SAR and optical data improved accuracy by 2–5%, a valuable gain in large-scale applications such as crop monitoring, insurance estimation, and early warning systems (Dube et al. 2024).

Sentinel-1's VH polarization data was particularly useful during monsoon-induced optical data gaps (Tassi and Vafeidis, 2022). Implementation on Google Earth Engine enabled efficient, scalable processing without the need for high-end infrastructure, making the workflow replicable for similar agricultural settings (Shelestov et al. 2017; Abdali et al. 2023). Practically, the results confirmed rice as the dominant crop, with limited sugarcane coverage (Abubakar et al. 2023). Future studies could expand to include additional crop types, incorporate various SAR polarizations, and explore SAR-based indices for richer feature representation (Guo et al. 2021; Vuolo et al. 2023).

In summary, this study highlights a robust, scalable, and adaptable approach for crop classification using modern remote sensing and ensemble machine learning, with potential for real-time applications and integration into broader agricultural decision-making frameworks.

Highest-performing classifier maps visualization:

The final maps generated using the highest- performing classifiers and the ensemble model clearly depict the distribution of major land covers including rice fields, sugarcane plantations, water bodies, forests, settlements, and bare lands across the Udham Singh Nagar district. Each map represents land cover over Udham Singh Nagar during the 2023 cropping season. The maps differ in terms of time window (July vs. June–October composite), machine learning model (GBM vs. RF), and sensor data used (optical only vs. optical + microwave).

4.5 Classifiers

4.5.1 July GBM (Optical)

The GBM classification for July using Sentinel-2 optical data highlights strong spectral separability during peak vegetative growth, particularly for rice, while showing limited discrimination between settlement and bare land classes (Figure 5). Rice fields (yellow) dominate the map, with sugarcane (red) appearing in scattered patches. Forest areas are well delineated due to their distinct spectral response. However, spectral overlap during this period reduces the clarity between urban and bare land classes, resulting in noticeable mixing, especially in the central regions of the study area.

4.5.2 Combined RF (Optical)

The Random Forest (RF) classification using multi-temporal Sentinel-2 optical data (June–October 2023) effectively captures seasonal phenological variability, resulting in improved class separability compared to the single-month GBM output (Figure 6). While rice (yellow) remains the dominant class, there is a clearer delineation of settlements (purple) and bare land (brown). The use of multi-month inputs allows RF to incorporate crop growth and post-harvest stages, enhancing discrimination among land cover types. The resulting map appears smoother, with better representation of smaller patches, and clearly highlights the rice-dominated agricultural landscape of Udham Singh Nagar.

4.5.3 July GBM (Optical and Microwave)

Figure 7 presents the GBM classification for July using fused Sentinel-1 (VH polarization) and Sentinel-2 data, demonstrating the advantages of SAR–optical integration during peak monsoon conditions. The inclusion of SAR data reduces spectral confusion and significantly improves the delineation of settlements, forests, and bare land. Crop classes such as rice and sugarcane are more clearly defined, with sharper boundaries and reduced noise, particularly in urban and mixed land cover areas. This highlights the effectiveness of data fusion in enhancing classification accuracy under dense vegetation and cloud-affected conditions.

4.5.4 Combined RF (Optical and Microwave)

Figure 8 presents the most comprehensive LULC classification, generated using the Random Forest (RF) model on fused Sentinel-1 and Sentinel-2 data from June to October 2023. By integrating multi-temporal and multi-sensor inputs, this map achieves the highest overall accuracy (89.21%) and provides the clearest delineation of all land cover classes. Settlements and bare land are distinctly separated due to the structural sensitivity of SAR data combined with seasonal variability captured in optical imagery. Sugarcane fields are more accurately mapped and spatially coherent compared to single-month outputs. Overall, the classification offers a reliable and detailed representation of Udham Singh Nagar's agricultural landscape, making it well suited for operational planning and large-scale monitoring.

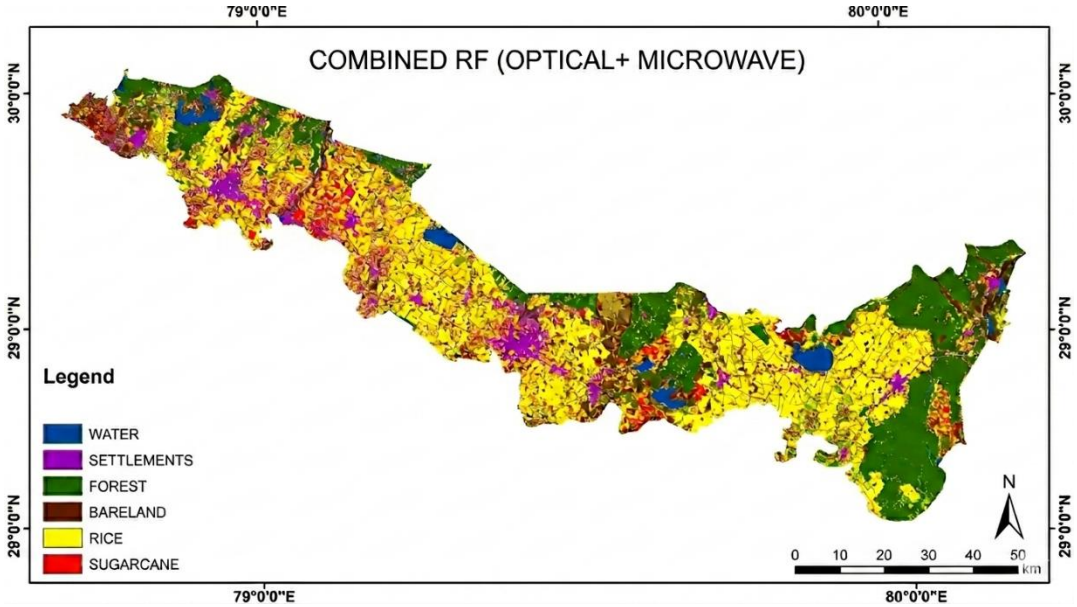

Figure 8: RF-Based LULC Classification using Optical and Microwave Data

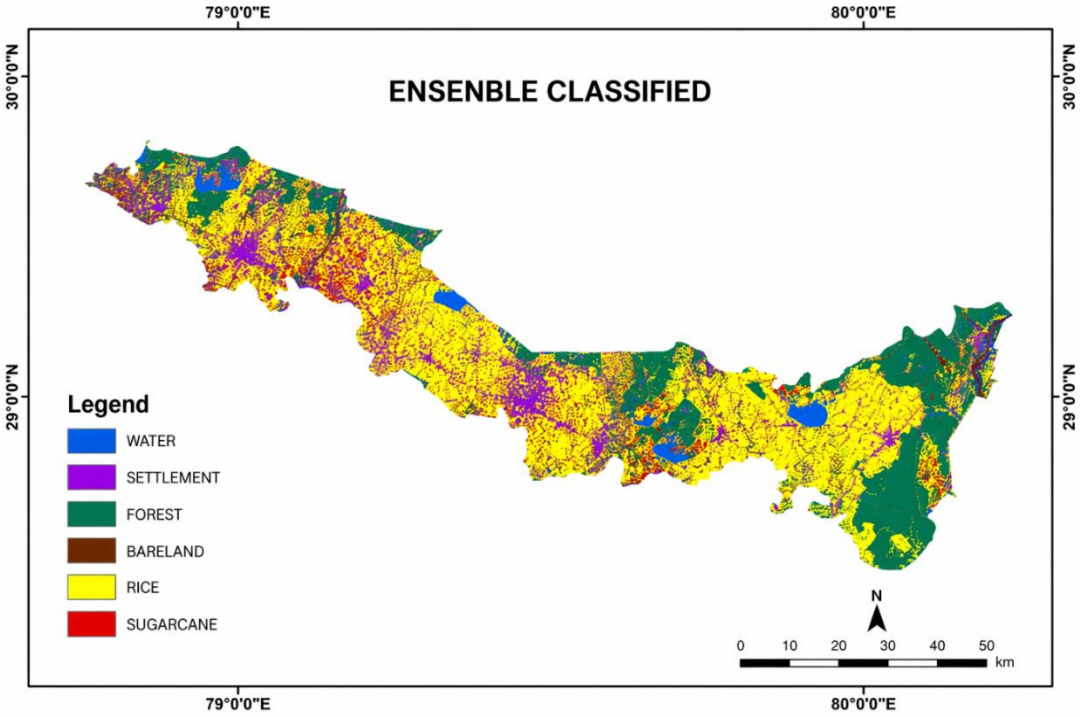

Figure 9: Ensemble classification results showing crop area distribution

The improved clarity achieved through SAR–optical fusion reinforces the value of multi-sensor approaches for applications such as crop monitoring and insurance estimation, as highlighted in recent studies. Additionally, the efficient generation of high-accuracy maps using platforms like Google Earth Engine supports scalable workflows for regional analysis. These results emphasize the importance of seasonally informed, data-driven frameworks for agricultural planning and climate-resilient land management.

4.5.5 Ensemble Classification

The ensemble classification, implemented through majority voting across RF, SVM, KNN, CART, and GBM outputs, achieved the highest overall accuracy of 90.25%, demonstrating greater robustness and consistency compared to individual classifiers (Figure 9). The resulting spatial distribution clearly highlights the dominance of agricultural land, with rice covering 177,749 ha and sugarcane occupying 2,838 ha, reflecting the prevailing cropping pattern in the study area.

5. Conclusions

This study comprehensively evaluated the effectiveness of integrating Optical and microwave remote sensing datasets for multi-temporal crop and land use/land cover (LULC) classification in Udham Singh Nagar district, Uttarakhand, India, during the 2023 agricultural season. By leveraging a combination of Sentinel-2 optical data and Sentinel-1 microwave backscatter information, classification accuracies improved by approximately 2–5%, with the

highest overall accuracy reaching 89.21% using Random Forest over the full June–October period, compared to 85.99% for the optical-only equivalent. Among the five machine learning classifiers tested- Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Classification and Regression Tree (CART), and Gradient boosted machine (GBM)- ensemble methods, particularly RF and GBM, Consistently and cumulative datasets. The highest overall accuracy achieved was 89.21% using the Random Forest classifier over the June-October period when combining optical and microwave datasets.

Seasonal vegetation dynamics, as captured through NDVI, EVI and NDBI closely followed known agricultural cycles, underscoring the value of time specific indices for improving classification separability. The integration of microwave data, less affected by atmospheric conditions was particularly valuable during cloud- prone monsoon months and in discriminating post-harvest bare fields and built-up areas.

Overall, the findings emphasize the critical role of multi-sensor data fusion and ensemble machine learning models in enhancing the reliability. Robustness and applicability of LULC mapping for operational and scientific purposes. Nevertheless, this study is limited to two primary crop types (rice and sugarcane), and spatial generalization of the proposed framework to other agro-climatic zones and cropping systems warrants further validation.

Acknowledgements

This research work was carried out as part of a final project under the academic program at the Indian Institute of Remote Sensing (IIRS), Dehradun, funded and supported by ISRO. Field survey were conducted under the TDP research project of IIRS-ISRO. The ground truth and field survey information were documented and shared from Agriculture & Soil Department, IIRS-ISRO, Dehradun.

Declaration of Interests

The authors declare that there is no conflict of interest in authorship and research work.

Data Availability

Data are available on request basis only.

Authors Contribution

Nidhi Chaudhary & Koneenika Mallick: Data Analysis, Visualization, Writing-original draft; Abhishek Danodia: Conceptualization, Resources, Data collection, Supervision, Methodology, Writing-review and editing; Kamal Pandey: Conceptualization, Supervision, Methodology; Nagendra S.R.: Data collection, Methodology

References

Abdali, E., M. J. Valadan Zoej, A. Taheri Dehkordi, and E. Ghaderpour. 2023. "A parallel-cascaded ensemble of machine learning models for crop type classification in Google Earth Engine using multi-temporal Sentinel-1/2 and Landsat-8/9 remote sensing data." *Remote Sensing* 15(4):1063.

Abubakar, G. A., K. Wang, A. F. Koko, M. I. Husseini, K. A. M. Shuka, J. Deng, and M. Gan. 2023. "Mapping maize cropland and land cover in semi-arid region in northern Nigeria using machine learning and Google Earth Engine." *Remote Sensing* 15(11):2835.

Achahboun, C., M. Chikhaoui, M. Naimi, and M. Bellafkih. 2022. "Crop classification using machine learning and Google Earth Engine." Pp. 1–6 in *Proceedings of the 2022 International Conference on Intelligent Systems and Computer Vision (ISCV).*

Dube, T., M. Sibanda, and C. Shoko. 2024. "Mapping smallholder crops using integrated Sentinel data and random forest in Google Earth Engine." *GIScience & Remote Sensing* 61(1):77–92.

Guo, H., L. Zhang, Z. Li, and Y. Zhao. 2021. "Land cover classification using optical and SAR data fusion with random forest and support vector machines on Google Earth Engine." *Remote Sensing Letters* 12(1):43–52.

Shelestov, A., M. Lavreniuk, N. Kussul, A. Novikov, and S. Skakun. 2017. "Large-scale crop classification using Google Earth Engine platform." Pp. 777–780 in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS).*

Tassi, A., and A. T. Vafeidis. 2022. "Multi-temporal crop classification using Sentinel-1 and Sentinel-2 data with machine learning." *ISPRS Journal of Photogrammetry and Remote Sensing* 186:98–114.

Vuolo, F., M. Neuwirth, M. Immitzer, C. Atzberger, and W. T. Ng. 2023. "Improved crop type mapping using Sentinel-2 time series and ensemble learning methods in Google Earth Engine." *International Journal of Applied Earth Observation and Geoinformation* 115:103107.

Yao, J., J. Wu, C. Xiao, Z. Zhang, and J. Li. 2022. "Classification methods for crop remote sensing using deep learning, machine learning, and Google Earth Engine." *Remote Sensing* 14(12):2758.