Original Research Article

Remote Sensing-Based Crop Identification and Acreage Estimation of Rabi Wheat in Anand, Gujarat

ABSTRACT

In the comprehensive evaluation of crop identification methods, classified methods/algorithms were thoroughly evaluated. The supervised classification approach encompassed four classifier methods, namely Random Forest (RF), Minimum distance (MD), Support Vector Machine (SVM), and Smile Cart (sCART), all of which demonstrated significant results. Notably, the RF classifier exhibited the highest overall average accuracy, reaching 91%, with an average kappa value of 87% across two seasons of experimentation. Conversely, the unsupervised classification method employed the Isoclustering algorithm, achieving an average accuracy and kappa value of 84% for the first season and 80% for the second season. Comparative analysis revealed that the random forest method consistently outperformed other classifiers, with MD, SVM, and sCART following in accuracy for Sentinel-2 imagery. Acreage estimation, employing both supervised and unsupervised classification techniques, involved the multiplication of spatial resolution with the number of pixels. The RF classifier estimated 69 thousand hectares (2019-20) and 64 thousand hectares (2020-21), aligning closely with the district statistical yield data of 61 thousand hectares (2019-20) and 58 thousand hectares (2020-21). In contrast, the accuracy of the sCART and SVM classifiers was notably lower at 46% and 36%, respectively, when compared with the statistical yield data.

Key words: Crop classification, Sentinel-2 imagery, Random Forest, Support vector machine, Smile cart, Wheat

1. INTRODUCTION

Numerous factors, such as population growth, increased biofuel consumption, and rising demands for dairy and meat products, will significantly shape the future of global agriculture(Godfray *et al.*, 2010). (Huang *et al.*, 2004; Tilman *et al.*, 2011)These challenges loom ahead due to widespread reports of yield stagnation in various cereal crops, including rice, wheat, and maize. Projections indicate that global food demand will double by 2050. Recent global crop production has fallen short of meeting these anticipated demands, prompting us to question which geographic regions are best suited to generate bountiful harvests that can satisfy the needs of our growing population(Finger, 2010; Jiao *et al.*, 2014; Peltonen-Sainio *et al.*, 2009). Researchers from around the world have consistently noted the issue of yield stagnation in various cereal crops.

Monitoring crops plays a crucial role in numerous agricultural and ecological applications(Harfenmeister *et al.*, 2021; Prudente *et al.*, 2019). These applications encompass estimating crop yields, warding off disease and insect infestations, applying fertilizers, and effectively managing water resources. The identification and prediction of phenological stages supply indispensable insights for precision agriculture(Dineshkumar *et al.*, 2019).Delving into specific phenological stages can optimize schedules for

irrigation and fertilizer application. Certain phenological stages exhibit heightened susceptibility to pests and diseases; thus, foreseeing and pinpointing these stages can forestall pest outbreaks, curbing the need for excessive pesticide usage(Lopez-Sanchez *et al.*, 2012). Furthermore, phenological stages can serve as indicators of the impact of global warming on terrestrial ecosystems.(Franko *et al.*, 2007; Jones *et al.*, 2003; Keating *et al.*, 2003)In pursuit of maximizing crop yields, numerous researchers have delved into exploring the intricate connection between crop development and its environmental conditions. They have also pioneered the creation of crop models aimed at replicating the intricate dynamics of crop growth(Nendel *et al.*, 2011). Over the course of nearly four decades, (Steduto *et al.*, 2009; Stöckle *et al.*, 2003)these crop models have evolved significantly, progressing from their early qualitative representation of crop development to their present capability of quantitatively mirroring crop growth. Furthermore, they have transitioned from solely simulating individual physiological and ecological growth aspects to encompassing the entire continuum of the growth process.

The recent availability of Synthetic Aperture Radar (SAR) Sentinel-1 (S-1) and optical Sentinel-2 (S-2) sensors has opened up a distinctive opportunity for regular, high-resolution crop monitoring. These sensors capture image time-series at a high temporal frequency, with intervals ranging from 5 to 12 days, contingent on the acquisition mode and geographic location. Additionally, they provide high spatial resolution, featuring 2.3 meters and 13.9 meters in the range and azimuth directions for S-1 bands, and spatial resolutions of 10, 20, and 60 meters for S-2 bands. The added advantage is that S-1 & 2 data are freely accessible under an open license(Felegari *et al.*, 2021).

The information on crop area statistics is backbone of agricultural statistical system of which reliable and timely information on crop area is of great importance to planners and policy makers. This information is useful for efficient and timely agricultural development and making important decisions with respect to procurement, storage, public distribution, export, import and other related issues. Crop yield estimates are generally portrayed as the product of two components: area to be harvested and expected yield per unit area (You *et al.*, 2014). Making timely and accurate regional predictions of crop yield is of great importance for agricultural management and food security warning purposes (Piao *et al.*, 2010; Fritz *et al.*, 2018). Applications of remote sensing technique in crop acreage estimation has becoming increasingly demanding and dominating in India (Mosleh *et al.*, 2015) due to low cost and this approach with combination of ground truth data will retrieve the best area estimate. Several techniques have been adapted for crop estimation using aerial photographs and satellite images, including: pixel count (Gallego *et al.*, 2014), supervised classification (Kussul *et al.*, 2017), Bayesian/fuzzy classification and spectral un-mixing (Mann and Joshi 2017) and area frame sampling (Pradhan 2001, Boryan *et al.*, 2017). Wu and Li (2012) studied crop planting and type proportion method for crop acreage estimation of complex diverse agricultural landscapes.

Thus, obtaining reliable information on crop classification and acreage in the mixed cropping situation is of paramount importance for farming policies regarding import/ export, procurement, storage etc. However, the use of satellite data for

identification of various crop under multiple cropping system is limited, against this backdrop, the present study was undertaken to evaluate the crop classification and acreage of wheat crop in Anand district.

MATERIAL AND METHODS

2.1 Study area

The study area was in Anand(latitude 22° 35' N, longitude 72° 58' E), Gujarat, India, as shown in Fig. 1. The area is under Middle Gujarat Agro Ecological Situation Zone, Zone-III**.** During the winter/rabi (November to March), mean monthly maximum and minimum temperature varies from 20° to 36°C and 7° to 20°C, respectively. Normal annual rainfall of the location is 882 mm, of which maximum amount of rainfall is received during June to September (south west monsoon) and meagre rain is receives during *rabi* season.

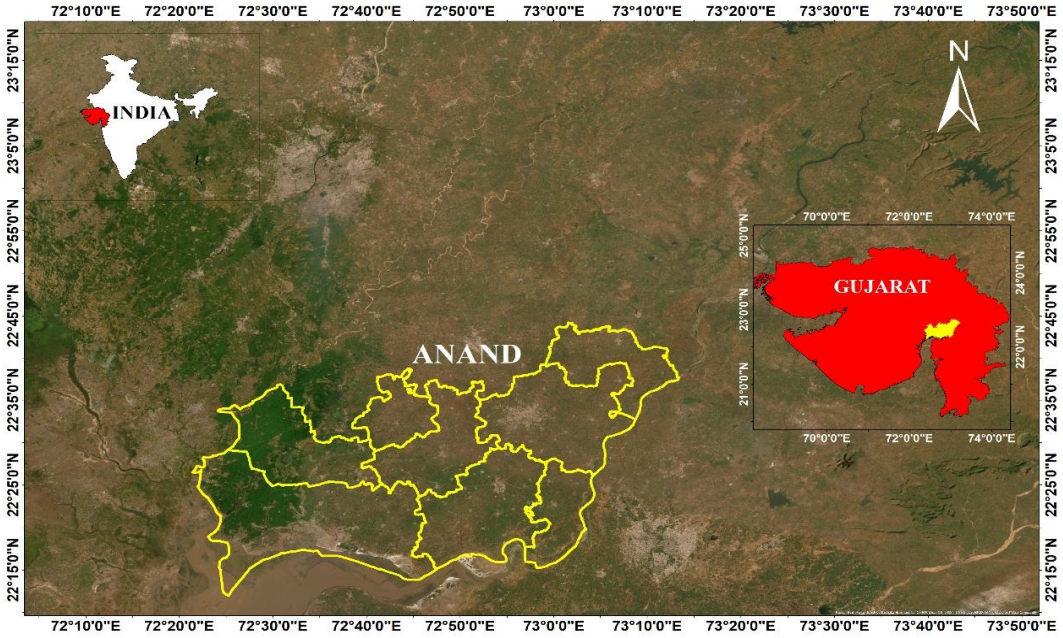

Fig. 1. Test site location, Anand District of Gujarat (India)

2.2 Remote Sensing Classification Methods

Crop classificationwas conducted using the different approaches(Table 1)to analyse land use and land cover patterns. Both supervised and unsupervised classification techniques were employed to achieve accurate and meaningful results.

Table 1. Methods used for Cropland classification.

Classification Type	Method	Validation
Supervised Classification	Random Forest (RF)	Validated using ground truth data, demonstrating high accuracy and robustness.
	Support Vector Machine (SVM)	Validated with ground truth data, performing well on nonlinear and

		complex datasets.
	Minimum Distance (MD)	Validated with ground truth data,demonstrating higher accuracy.
	SmileCart (sCART)	Validated using ground truth data, providing reliable results with optimized parameters.
Unsupervised Classification	K-Means Clustering	Validated with ground truth data through visual inspection and comparisons with existing maps.

Using both trained and unsupervised techniques, a multilayer satellite input image is transformed into a single-layer thematic map for classification. While unsupervised classification groups pixels according to spectral value similarity, as illustrated in technique flowcharts (Figures 2 and 3), supervised classification depends on user-selected sample pixels (training sites) that direct the software in classifying all image pixels. Google Earth Engine (GEE) was used to pre-process the sentinel data for this study in order to exclude cloud cover and compute indices like NDVI and NDWI. NDVI, SAVI, NDBI, and NDWI indices were added to 30-meter-resolution seasonal composite images for the rabi wheat seasons (2019–20 and 2020–21) in order to increase mapping accuracy. Random Forest (RF), Support Vector Machine (SVM), Minimum Distance (MD), and Smile CART (sCART) were used for classification.

High-resolution Google Earth imagery and ground truth data from the Anand region were utilized to create training sites, of which 70% were used for training and 30% for validation. Both years' classifications included six land-use classes: urban, water, forest, wheat, and fallow land. SVM employed a radial basis function (RBF) kernel with a gamma value of 0.075, and RF had 100 decision trees. For wheat, the training site points were 70 for the 2019–20 season and 60 for the 2020–21 season. The sample proportions were based on the land-use area. By guaranteeing strong classification results, these strategies demonstrated the relative advantages and disadvantages of each technique.

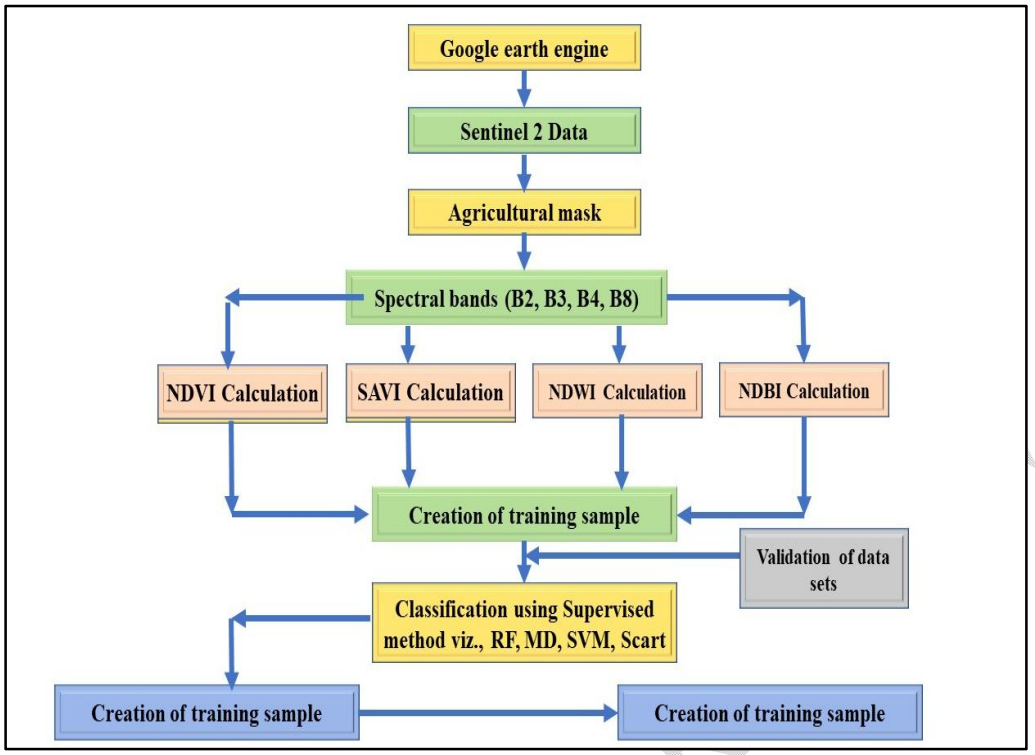

Figure 2: Methodology flowchart used for supervised imageclassification using GEE

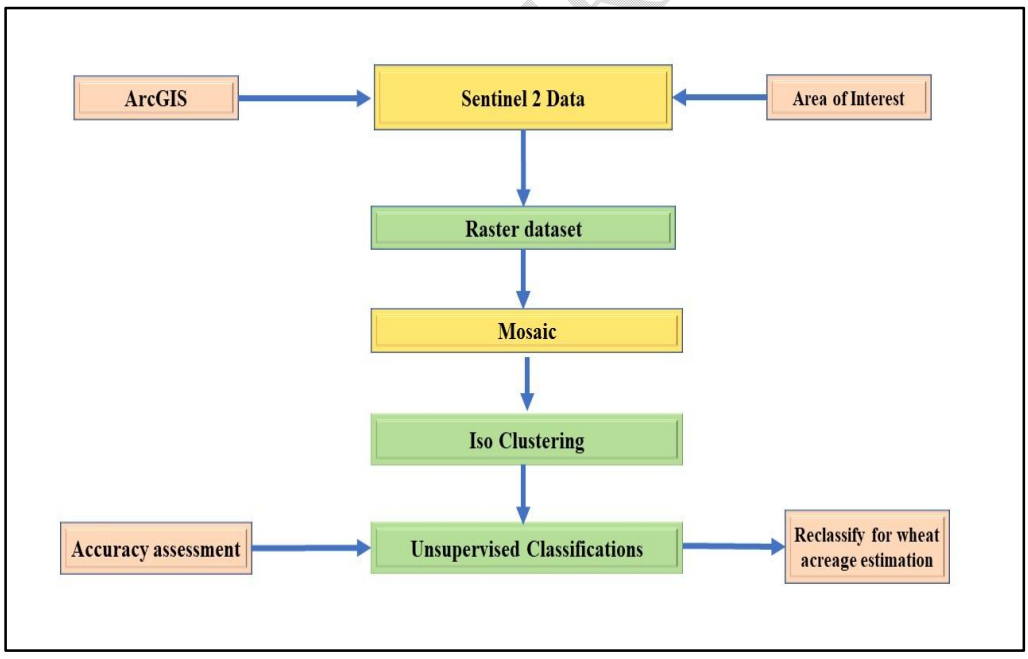

Figure 3: Methodology flowchart used for unsupervised image classification using ArcGIS

2.3 Training Data Collection

Training datasets were derived from high-resolution Sentinel-2B imagery. Six land-use classes were defined: Urban, Water, Forest, Wheat, Fellow Land, and Tobacco/Other Crops.

Table 2. Number of training and validation samples, for each class investigated.

Class	**2019-20**	**2020-21**
Urban	161	170
Water	117	117
Forest	100	100
Wheat	100	100
Fellow Land	50	50
Tobacco and Other Agricultural Crops	160	160

2.4 Classification accuracy assessment and Statistical Validation

The accuracies of the pixel- and object-based classifications obtained were evaluated in terms of overall accuracy, producer's accuracy, user's accuracy metrics(Congalton, 1991), and kappa coefficient(Cohen, 1960). The validation samples available for study area are shown in Table 2. The classification results were validated through ground truth data, visual interpretation, and comparisons with existing authoritative data.

3. RESULTS AND DISCUSSION:

3.1. Supervised Classification of Cropland using Sentinel-2 data

Supervised classification methods were applied for the years 2019-20 (first year of experimentation) and 2020-21 (second year of experimentation). Classifier performances were assessed based on accuracy percentages (Table 4).Figures 4 illustrate the classification results for Random Forest (RF) for wheat, which achieved approximately 90% accuracy in most parts of the Anand region. However, areas such as Khambhat and Bhal showed some misclassification, particularly between aestivum and durum wheat, likely due to variations in spectral properties. Additionally, in 2019-20, some wheat spectral properties were misclassified as barren land or, to a lesser extent, water bodies.

Across both years, RF, MD, and SVM performed well in classifying wheat due to the dense plantations exhibiting similar reflectance patterns. While insCARTmisclassifications occurred near barren land, other agricultural crops, and forests, as evident in classifier results. Water bodies, forests, barren land, and built-up areas were correctly classified by RF and other algorithms.The sCART algorithm misclassified different classes for wheat, tobacco, water bodies, built-up areas, barren land, and forests for both years, outperforming other methods. For SVM, wheat was misclassified as other vegetation, barren land, or built-up areas. Similarly, MD effectively

classified most regions but misclassified forests and certain built-up areas, as reflected in its confusion matrices.

The classification changes across the years 2019-20 and 2020-21, as derived from Sentinel-2 images, are presented in Figures 4 to 7, demonstrating the performance of different algorithms.For all methods from the two years, the RF classifier performed well in comparison to the other classifiers *i.e.*, SVM, MD and sCART.

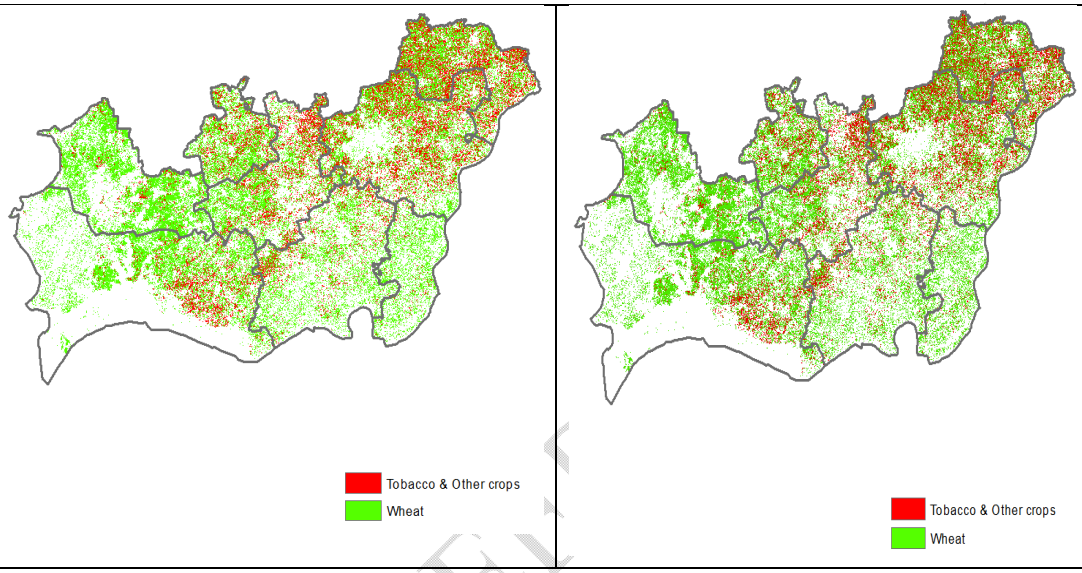

Figure 4: Supervised classification map of wheat using random forest classifiers for year 2019-20and 2020-21

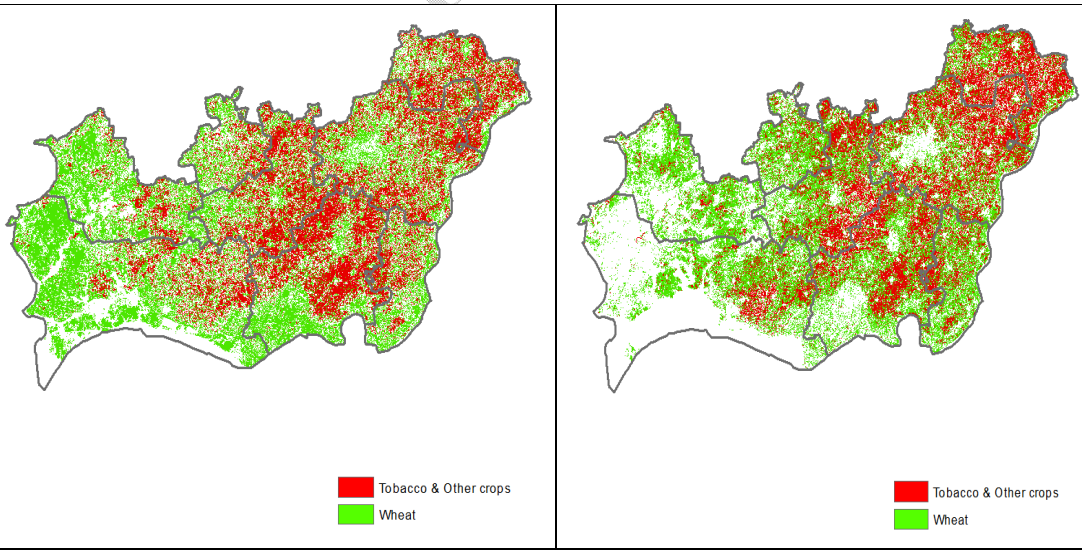

Figure 5: Supervised classification map of wheat using support vector machine (SVM) classifiers for year 2020-21and 2020-21

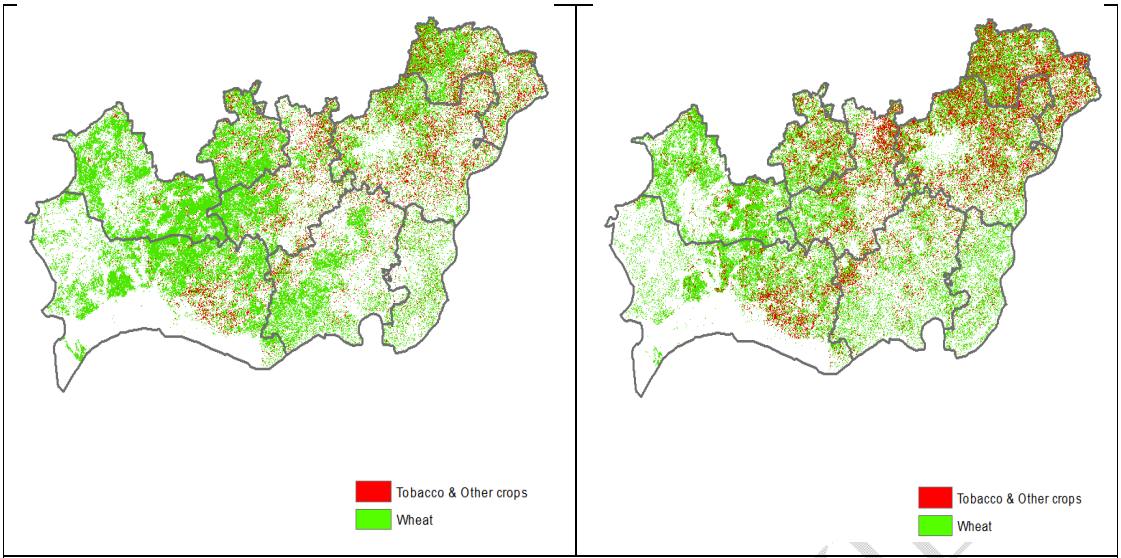

Figure 6: Supervised classification map of wheat using minimum distance classifiers for year 2019-20 and 2020-21

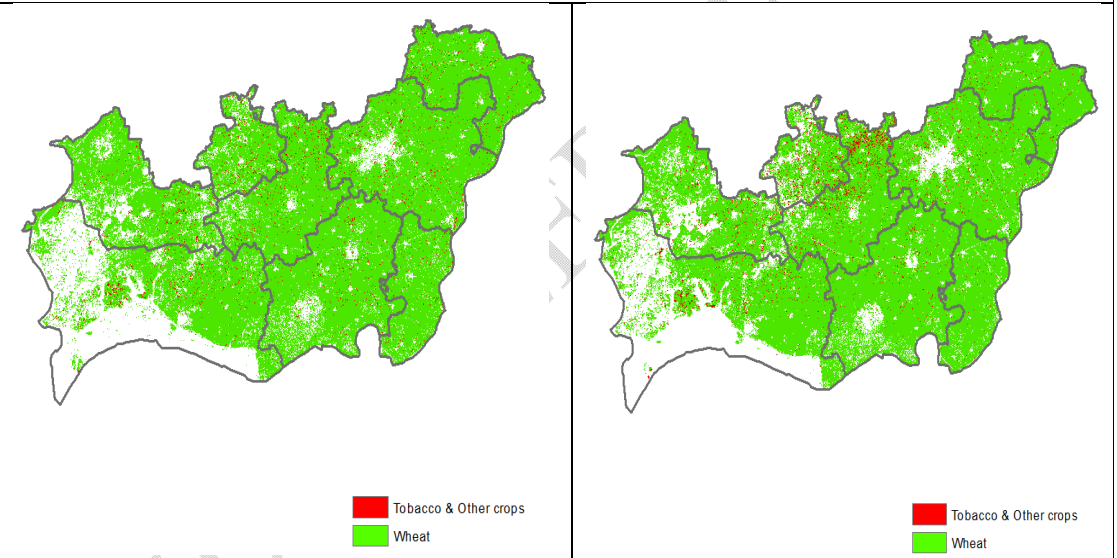

Figure 7: Supervised classification map of wheat using smile cart (sCART) classifiers for year 2020-21 and 2020-21

3.2. Unsupervised Classification of Cropland using Sentinel-2 data

The sentinel satellite images pertaining to Anand district were also classified based on unsupervised classification method using Iso clustering approach (Venables and Ripley, 2002). In the unsupervised classification method, pixels were classified based on spectral values only without any ground truth information. The area which was classified as water body and other class by the supervised classification method has now been classified as wheat class by unsupervised classification technique.

In the unsupervised classification output (Figure 8) of Anand district, result shows. the mask of the area of study which shows wheat vegetation of district in rabi season shows the spatial distribution of the crop. Wheat observed to be the major crops in Anand district, as it classified has the major coverage in the district during *rabi* season 2019-20 and 2020-21.

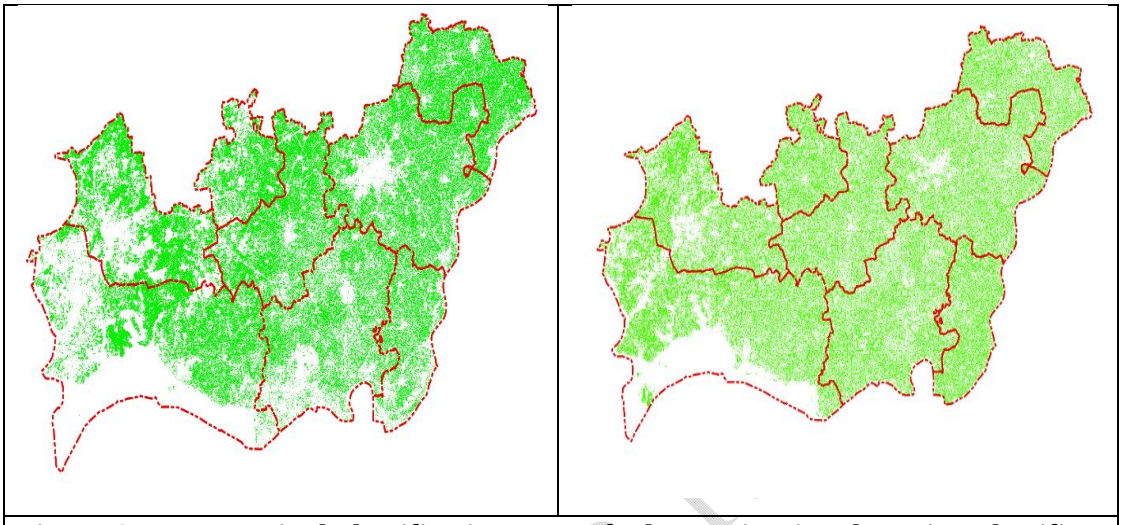

Figure 8: Unsupervised classification map of wheat using iso-clustering classifiers for year 2019-20and 2020-21

In the Anand district, the results of the comparison between supervised and unsupervised classification methods revealed that the supervised approach resulted in greater accuracy. The difference in accuracy between the two methods were found to be the least in the first and second seasons, respectively.The supervised classification had an overall accuracy of above 90 % while the unsupervised classification had an overall accuracy of 85 %. Thus, random forest method is much more capable in crop classification and identification for supervised classification while iso clustering classifier performed reasonable in crop classification for both years of experimentation.

3.3 Acreage Estimation

Remote sensing has significant potential for estimating acreage and identifying different classes. After classifying the image using supervised and unsupervised techniques, aggregation was performed to estimate the wheat area for both study seasons. The wheat area was calculated by multiplying the spatial resolution by the number of pixels, as detailed in Table 3.

Table 3: Area of Wheat in Anand district

Sr. No.	Classifiers	2019-20		2020-21	
		Estimated Area (ha^{-1})	Reported Area (ha^{-1})	Estimated Area (ha^{-1})	Reported Area (ha^{-1})
01	RF	68996	61000	64255	58000
02	MD	77263	61000	73568	58000
03	SVM	84364	61000	81626	58000
04	sCART	115549	61000	98236	58000
05	Isoclustering	74253	61000	69258	58000

3.4 Classification Accuracy Assessment and Statistical Comparison

Classification methods yield different results when applied to the same data. In this study, we calculated accuracy matrices—Kappa coefficient, producer's accuracy, user's accuracy, and overall accuracy—using confusion matrices derived from training and validation sample points. Pixel-based algorithms used confusion matrices based on pixel counts.

3.5 Comparison of Classifiers

Table 4 summarizes the classifier performance for two experimental years. The Random Forest (RF) classifier outperformed others, with the highest overall accuracy (92% and 90%) and Kappa values (88% and 86%) for 2019-20 and 2020-21, respectively. Iso-clustering achieved 86% and 85% overall accuracy, with Kappa values of 82% and 76%.

Table 4: Accuracy and the kappa of the classifiers for two seasons of the study

Year	Classifier	Overall Accuracy	Kappa Coefficient
2019-20	RF	92%	0.88
	SVM	77%	0.77
	MD	86%	0.76
	sCART	83%	0.70
	Iso-cluster	86%	0.82
2020-21	RF	90%	0.86
	SVM	74%	0.65
	MD	84%	0.79

	sCART	86%	0.82
	Iso-cluster	85%	0.76

3.6 User's Accuracy

User's accuracy for urban and water classes was high across classifiers due to distinct spectral properties. However, misclassification issues arose with land cover types such as wheat and forest, where spectral overlaps led to errors. For instance, wheat pixels were misclassified as forest or fellow land due to similar spectral signatures. Table 5 shows the user's accuracy for different classes.

Table 5: Users' accuracy of based on the stacked images for supervised classification

Year	Classifier	Urban	Water	Forest	Wheat	Tobacco & Other Crops	Fellow Land
2019-20	RF	1.00	1.00	0.95	0.94	0.90	1.00
	SVM	1.00	0.97	0.85	0.61	0.87	0.46
	MD	1.00	0.98	0.82	0.88	0.89	1.00
	sCART	0.98	0.98	1.00	0.61	0.81	0.40
	Iso-clustering	1.00	0.98	0.90	0.84	0.84	0.88
2020-21	RF	0.99	1.00	0.99	0.93	0.88	0.89
	SVM	0.97	0.97	1.00	0.98	0.83	0.25
	MD	0.98	0.98	0.81	0.98	0.86	0.89
	sCART	0.98	0.97	0.99	0.55	0.81	0.67
	Iso-clustering	1.00	0.94	0.75	0.82	0.90	0.90

3.7 Confusion Matrices

The confusion matrices for 2019-20 and 2020-21show classification errors for each method. The RF classifier, despite its high overall accuracy, misclassified a small number of tobacco and other crops points. The SVM and sCART methods highlighted misclassification of fellow land, while Iso-clustering showed errors in classifying wheat.

4. Conclusion and Future Works:

This study assessed various supervised and unsupervised classification methods for land cover and wheat acreage estimation in Anand district, Gujarat, using Sentinel-2 remote sensing data. Key findings include:

- Supervised vs. Unsupervised Classification: The Random Forest (RF) method outperformed others, including Iso-clustering, with the highest accuracy in land cover classification and wheat acreage estimation.
- Accuracy Assessment: Urban and water classes had high accuracy due to distinct spectral properties. Misclassifications occurred between wheat and classes like forest due to spectral overlaps.
- Acreage Estimation: RF and MD classifiers closely matched statistical yield data, with RF providing the most accurate acreage estimation.
- Implications: Accurate wheat acreage estimation is crucial for crop yield prediction and food security planning. Remote sensing data combined with advanced classification methods offers a promising approach for agricultural monitoring.
- Future Research: Future studies could explore multi-sensor data and advanced machine learning techniques to enhance classification accuracy and monitoring.

In conclusion, RF demonstrated its effectiveness in land cover classification and wheat acreage estimation, highlighting the potential of remote sensing and machine learning for agricultural monitoring and informed decision-making.

Disclaimer (Artificial intelligence)

Option 1:

NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2: ----NO

5. REFERENCES

1. Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1), 37-46.
2. Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment, 37(1), 35-46.
3. Dineshkumar, C., Kumar, J. S., & Nitheshnirmal, S. (2019). Rice Monitoring Using Sentinel-1 Data in the Google Earth Engine Platform. June 2019, 4. https://doi.org/10.3390/iecg2019-06206
4. Felegari, S., Sharifi, A., Moravej, K., Amin, M., Golchin, A., Muzirafuti, A., Tariq, A., & Zhao, N. (2021). Integration of sentinel 1 and sentinel 2 satellite images for crop mapping. Applied Sciences (Switzerland), 11(21). https://doi.org/10.3390/app112110104
5. Gan, L., Cao, X., Chen, X., Dong, Q., Cui, X., & Chen, J. (2020). Comparison of MODIS-based vegetation indices and methods for winter wheat green-up date detection in Huanghuai region of China. Agricultural and Forest Meteorology, 288–289, 108019. https://doi.org/https://doi.org/10.1016/j.agrformet.2020.108019
6. Harfenmeister, K., Itzerott, S., Weltzien, C., & Spengler, D. (2021). Agricultural monitoring using polarimetric decomposition parameters of sentinel-1 data. Remote Sensing, 13(4), 1–28. https://doi.org/10.3390/rs13040575
7. Ibrahim, E. S., Rufin, P., Nill, L., Kamali, B., Nendel, C., & Hostert, P. (2021). Mapping crop types and cropping systems in nigeria with sentinel-2 imagery. Remote Sensing, 13(17), 1–24. https://doi.org/10.3390/rs13173523
8. Nasrallah, A., Baghdadi, N., El Hajj, M., Darwish, T., Belhouchette, H., Faour, G., Darwich, S., & Mhawej, M. (2019). Sentinel-1 data for winter wheat phenology monitoring and mapping. Remote Sensing, 11(19). https://doi.org/10.3390/rs11192228
9. Ndikumana, E., Ho, D., Minh, T., Nguyen, D., Thu, H., Baghdadi, N., Courault, D., Hossard, L., & Moussawi, I. El. (2018). Rice height and biomass estimations using multitemporal SAR Sentinel-1 : Estimation of Rice Height and Biomass Using Multitemporal SAR Sentinel-1 for Camargue ,. October. https://doi.org/10.1117/12.2325174
10. Nendel, C., Berg, M., Kersebaum, K. C., Mirschel, W., Specka, X., Wegehenkel, M., Wenkel, K. O., & Wieland, R. (2011). The MONICA model: Testing predictability for crop growth, soil moisture and nitrogen dynamics. Ecological Modelling, 222(9), 1614–1625. https://doi.org/10.1016/j.ecolmodel.2011.02.018
11. Ok, A. O., Akar, O., & Gungor, O. (2012). Evaluation of random forest method for agricultural crop classification. European Journal of Remote Sensing, 45(1), 421–432. https://doi.org/10.5721/EuJRS20124535
12. Pan, Z., Huang, J., Zhou, Q., Wang, L., & Cheng, Y. (2015). Mapping crop

phenology using NDVI time-series derived from HJ-1 A/B data - Research Portal | Lancaster University. International Journal Of, 1–19. http://www.sciencedirect.com/science/article/pii/S0303243414001755%5Cnhttp://www.research.lancs.ac.uk/portal/en/publications/mapping-crop-phenology-using-ndvi-timeseries-derived-from-hj1-ab-data(166c7adf-2ea1-4451-ad30-d0cae4a333f5).html

13. Peltonen-Sainio, P., Jauhiainen, L., & Laurila, I. P. (2009). Cereal yield trends in northern European conditions: Changes in yield potential and its realisation. Field Crops Research, 110(1), 85–90. https://doi.org/10.1016/j.fcr.2008.07.007
14. Prudente, V. H. R., Oldoni, L. V., Vieira, D. C., Cattani, C. E. V., & Del'Arco Sanches, I. (2019). Relationship between SAR/Sentinel-1 polarimetric and interferometric data with biophysical parameters of agricultural crops. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(3/W6), 599–607. https://doi.org/10.5194/isprs-archives-XLII-3-W6-599-2019
15. Saini, R., & Ghosh, S. K. (2018). Crop Classification on Single Date Sentinel-2 Imagery Using Random Forest and Suppor Vector Machine. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII–5(November), 683–688. https://doi.org/10.5194/isprs-archives-xlii-5-683-2018
16. Tatsumi, K., Yamashiki, Y., Canales Torres, M. A., & Taipe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. Computers and Electronics in Agriculture, 115, 171–179. https://doi.org/10.1016/j.compag.2015.05.001
17. Zhou, T., Pan, J., Zhang, P., Wei, S., & Han, T. (2017). Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region. Sensors (Switzerland), 17(6), 1–16. https://doi.org/10.3390/s17061210