Print ISSN : 0972-8813
e-ISSN : 2582-2780

[Vol. 22(2) May-August 2024]

Pantnagar Journal of Research

(Formerly International Journal of Basic and Applied Agricultural Research ISSN : 2349-8765)

G.B. Pant University of Agriculture & Technology, Pantnagar

PANTNAGAR JOURNAL OF RESEARCH

Vol. 22(2) May-August 2023

CONTENTS

Efficiency assessment of classifiers for sugarcane area mapping: A machine learning approach with Google Earth Engine

POOJA YADAV*[1], AJEET SINGH NAIN[1] and SHIVANK DEVLIYAL[2]

[1]*Department of Agrometeorology, College of Agriculture and* [2]*College of Technology, G.B. Pant University of Agriculture and Technology, Pantnagar -263145 (U. S. Nagar, Uttarakhand)*
**Corresponding author's email id : 56875@gbpuat.ac.in*

ABSTRACT : One crucial use of remote sensing in agriculture is crop type mapping and area estimation. Recently, crop categorization has seen a rise in the use of machine-learning classification approaches. Google Earth Engine (GEE), a platform where users can explore a variety of satellite data sets without even downloading the satellite data and provides several powerful classification techniques. The main objective of this study is to explore the ability of different machine learning classification techniques like Random Forest (RF), K-Nearest Neighbor (KNN), Classification And Regression Trees (CART), Support Vector Machine (SVM), and Minimum Distance (MD) for crop classification using Google Earth Engine (GEE). Sentinel-2, MSI (10 m) Level-2A_SR dataset was used for crop type mapping and sugarcane area estimation for Udham Singh Nagar district for the year 2022-2023. Normalized difference vegetation index (NDVI) image composite were created at less than 10% cloud cover and using this we can easily identify the sugarcane area from other areas like other crops, forest, water bodies and buildings. Post classification, accuracy assessment analysis was done through the generation of the confusion matrix (producer and user accuracy), and F1 score. The results indicate that using GEE through a cloud platform, satellite data accessing, filtering, and pre-processing of satellite data can be performed very efficiently. In terms of overall classification accuracy and F1 score, RF (97.88%, 97.31%) classifier performed better than SVM (95.61%, 94.42%), MD (94.96, 93.60%), KNN (94.47%, 92.98%), and CART (90.57%, 88.01%) classifiers in the respective growing season of sugarcane.

Key words: Crop classification, Google Earth Engine, machine learning, remote sensing, Sentinel-2, sugarcane

Agriculture is the foundation for social and economic development (Awokuse, 2009; Virnodkar *et al*., 2020; Verma *et al*., 2023), food security (Lu *et al*., 2016; Gilbertson *et al*., 2017) and land resource management (Huang *et al*., 2016; Lebourgeois *et al*., 2017). Crop production is extremely unpredictable over time and geography, which results in chronic food insecurity (Akponikpe *et al*., 2011). Assessments of food security in nations with limited resources frequently depend on early warning systems (EWS) that use abnormalities in the Normalized Difference Vegetation Index (NDVI) (Baruth *et al*., 2008; Lambert *et al*., 2018). A perennial crop of the grass family, sugarcane (Saccharum spp.) is grown in tropical and subtropical regions. (Moore and Botha, 2013; Sindhu *et al*., 2016; Hu *et al*., 2019). Sugarcane is a key crop because it represents a major world source of sugar (Grof and Campbell, 2001; Li and Yang, 2015; Shield, 2016; Hu *et al*., 2019), ethanol (Cardona *et al*., 2010; Sindhu *et al*., 2016), paper, specialist chemicals such as furfural (Gomes *et al*., 2018), and bio-plastics (e.g. poly-lactic acid, Bonsucro, 2017). Planning and managing the sugarcane industry, requires timely and correct information about the sugarcane planting area, harvested area, green-up date, and harvest date. (Mulianga *et al*., 2015). Crop-type mapping can help us to obtain the spatial distribution patterns and proportions and is the basis for yield estimation (Bolton and Friedl, 2013; Song *et al*., 2017; van der Velde *et al*., 2019) and water resources management (Vogels *et al*., 2019). According to Zhong *et al*. (2016), Wei *et al*. (2018), Zhai *et al*. (2019) and Duan *et al*. (2021) traditional methods of area estimation have some disadvantages such as strong subjectivity, time-consuming, labor-intensive, delayed updating, and the lack of spatial distribution information on the other hand, remote sensing technology, has provided extensive coverage, quick data collecting, fast and dynamic updating, and so on. High-quality, low-cost data is always wanted when a proper blend of temporal, spectral, and spatial resolutions is required (Peña-Barragán *et al*., 2011), and Sentinel-2 satellites achieve this excellently (Zhang *et al*., 2020).

Machine learning, which is a branch of Artificial Intelligence (AI) focusing on learning, can determine patterns and correlations and discover knowledge from datasets. The models need to be trained using datasets, where the outcomes are represented based on experience (Van Klompenburg *et al*., 2020). Supervised classification is commonly applied to remotely sensed imagery and involves the use of a reference set of labeled training sites to train the model to train the algorithm to recognize objects, while testing data is used to determine the performance of the algorithms (Munoz-Mari *et al*., 2007; Carreiras *et al*., 2017; Fritz *et al*.; 2017; Li *et al*., 2017; Macintyre *et al*., 2020). The Google Earth Engine (GEE) platform has a variety of machine-learning algorithms that can be used for the supervised classification of image data e.g. Random Forest (RF), K-nearest neighbor (KNN), Classification And Regression Tree (CART), Support Vector Machine (SVM), and Minimum Distance. The research objectives of this paper are: to evaluate the potential of GEE in crop mapping at a regional scale using composite images; to map the spatial distributions of crops at a regional scale and to Check the potential of different classifiers.

MATERIALS AND METHODS

The overall methodology of this study is briefly presented in Figure 1.

Study Site

The present study site is the Udham Singh Nagar district of Uttarakhand state formed in October 1995 which is nestled in the Terai region, which stretches along the foothills of the Himalayan range. It is located between 78° 45' E and 80° 08' E in longitude, and between latitudes 28° 53' N and 29° 23' N (Figure 2) In this region, farmers grow a variety of crops throughout the year. Major field crops in this area are Rice, Wheat, Sugarcane, Mustard, Pea, Maize, and Soybean. The climate of the district is sub-tropical and sub-humid climatic conditions. All year round, the climate remains moderate. In the summertime, temperatures typically range about 30 and 40 degrees Celsius, resulting in scorching. It experiences cool, pleasant winters, with highs of 5 to 20 degrees Celsius. The region has a distinct dip

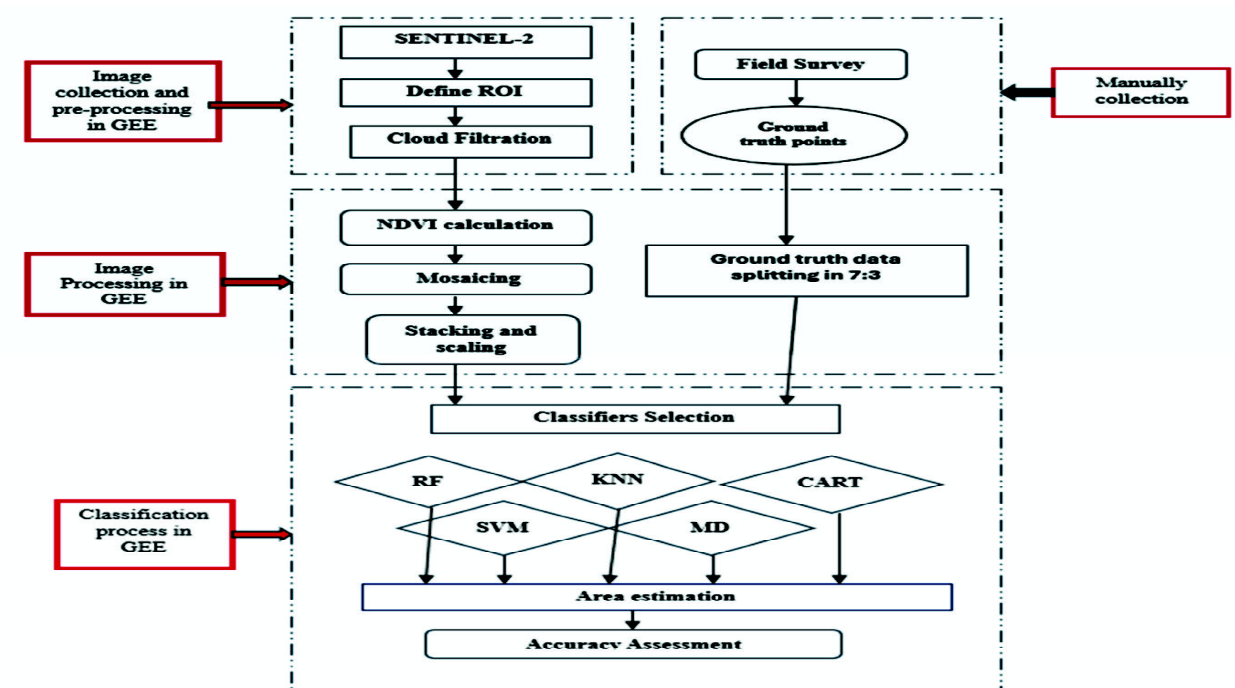

Fig. 1 : Flow chart of the methodology

in temperature during this period of year, which gives the area a cool vibe. Rainfall intensity fluctuates, with minimal levels in November and December and maximum levels in June and July (Bishnoi *et al.,* 2021).

Ground truth points

Ground Truth Data was collected using a smartphone-based Android application i.e. Map Marker. The ground truth was collected from September 2022 to December 2022 and covered the entire study area with major crop information like crop type, crop growth stage, and harvesting information with latitude and longitude information. Major crops were found as rice, sugarcane, soybean, maize, wheat, pea, rapeseed & mustard, and vegetables during the kharif and rabi season and we obtained reference points for Sugarcane(524), Other crops(494), Forest(402), Waterbodies (230) and Urban area (354) with total 2004 in the respective year for the training and validation of different classifiers/machine learning algorithms (Figure 3).

Machine Learning Algorithms

Random Forest (RF) is an algorithm developed from a decision tree based on multiclassification or regression based on many decision trees, where the random forest itself consists of several decision tree models (Jin *et al*., 2019) Based on the recommendations of previous studies and pretests from our data, we selected 100 trees (ntree = 100), while mtry was set to the default value (square root of the total number of features).

K-nearest neighbor (KNN) for classification tasks considered as a “lazy” algorithm because it does not create an explicit model during training. Instead, it stores the entire dataset and makes decisions on new observations instantly (Acito, 2023).

Support Vector Machine (SVM) is an alternative algorithm that can be used to overcome various weaknesses in classification using remote sensing data (Huang, 2002). It is based on an optimal hyperplane to separate classes from training data

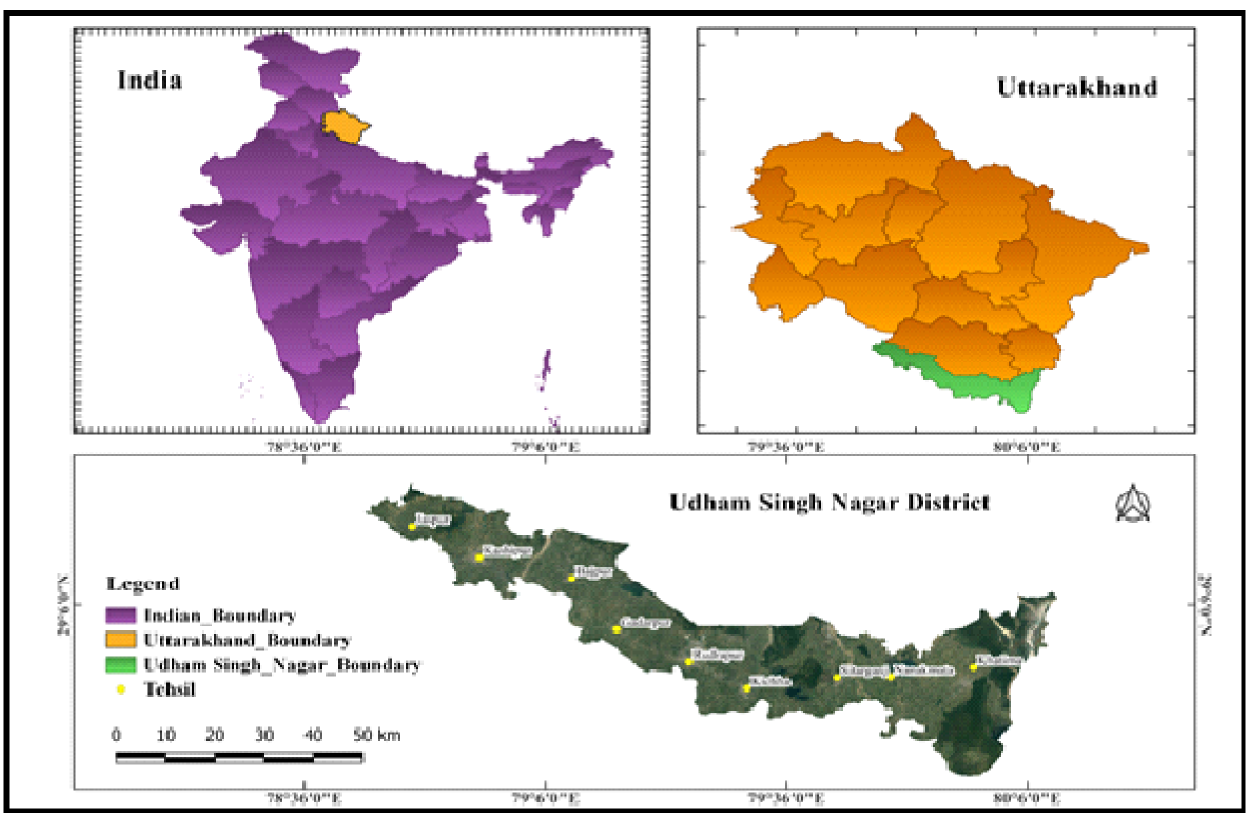

Fig. 2 : Location of the study area

based on the process that is performed (Mathur and Foody, 2008; Srivastava *et al*., 2012; Awad and Khanna, 2015). To map non-linear decision boundaries into linear ones in a higher dimension, the four most frequently used types of kernel functions in SVM algorithms are linear, polynomial, radial basis function (RBF) and sigmoid kernels (Kavzoglu and Colkesen, 2009).

Classification and Regression Tree (CART) is a classification algorithm. In this a tree can be developed in a binary recursive partitioning procedure by splitting the training sample set into subsets based on an attribute value test and then repeating this process on each derived subset. The tree-growing process stops when no further splits are possible for subsets. The maximum depth of the tree is the key tuning parameter in CART, which determines the complexity of the model (Qian *et al*., 2014).

The Minimum Distance (MD) classification algorithm first calculates the mean vectors and draws the decision boundary for each class (Torabi *et al*., 2019). It is a linear classification rule which assigns each pixel to the class whose mean is closest to Euclidean distances. It is a processor in LARSYS (Fuhs and Akiyama, 1980).

Cloud computing on Google Earth Engine (GEE) and image data

GEE has the entire Landsat and Sentinel archive along with many openly available raster datasets from NASA, the European Space Agency (ESA), and other imagery. Due to large area classification, the size of the satellite data is very high (running into petabytes when time series is involved). It is very hard to process the data in regular systems because they require high storage and take time to process, so we performed using the Google Earth Engine cloud computing platform for image processing. It can write scripts in Python and Java language to process petabyte-scale data processing over very large areas within minutes (Gorelick *et al*., 2017); In this study, we used the Sentinel 2 image collection from March 2022 to February 2023 for the entire Udham Singh Nagar District (Figure 2). The entire classification process was performed in GEE using the following steps:

- We used Sentinel 2 SR HARMONIZED image collection (i.e., surface reflectance atmospherically corrected, with masks for clouds, cloud shadows, snow, and water) (Hagolle *et al*., 2015). Sentinel-2 L2 assets have the following format: eg. COPERNICUS/S2_SR/20151128T002653_20151128T102149_T56MNN. Here the first numeric part represents the sensing date and time, the second numeric part represents the product generation date and time, and the final 6-character string is a unique granule identifier indicating its UTM grid reference.
- There is a significant cloud issue associated with optical sensor images to resolve this we applied a filter code at less than 10%. After this, we get images on different dates which are listed in Table 1.
- In the case of sentinel-2 entire study area is covered with 5 tiles (i.e. RKT, RLS, RMS, RMT, RLT) for this we need mosaicking to cover the entire study area in a single image and displayed in the output window console using the browser data catalog from March 2022 to February 2023 for the Udham Singh Nagar district in GEE JavaScript. From here we can export the image to the drive.
- Using the NIR (Band 8) and RED (Band 4), we applied the NDVI function to each image and produced NDVI Pixel composites at 10m spatial scale.
- Next, we used the BLUE, GREEN, RED, NIR, SWIR1, and SWIR2 bands to build high-quality visual mosaicking for every image. After that, we'll utilize this to generate the greenest NDVI image composites.
- Adding all of the NDVI bands together, we finally produced a staked image for the crop type mapping. For improved comprehension, apply the scale function if we wish to raise the image's scale.
- For training and validation, we split the ground truth data in a 7:3 ratio. After that, divide all of the 5 classes (i.e. Sugarcane, Other Crops, Forest, Waterbodies, Urban area) into two

sets—one for training and the other for validation—by using these ground truth points.

- In GEE we used Random Forest, KNN, CART, SVM, and Minimum Distance classifiers (MD) for the cropland classification. Generally, the image classification script sequence procedure in all the above algorithms is similar. The only difference is the machine learning algorithm itself (Kamal *et al.*, 2019).

Table 1 : Sentinel -2 Acquisition Dates for the area estimation of sugarcane in 2022-2023

Months	Sentinel -2 Acquisition Dates (2022-2023)	
	Available Images Days (without cloud filter)	Used Images Days (<10% cloud filter)
March	2,7,12,17,22,27	12,22
April	1,6,11,16,21,26	1
May	1,6,11,16,21,26,31	21,26
June	5,10,15,20,25,30	15
July	5,10,15,20,25,30	N/A
August	4,9,14,19,24,29	N/A
September	3,8,13,18,23,28	N/A
October	3,8,13,18,23,28	23,28
November	2,7,12,17,22,27	2,17,22,27
December	2,7,12,17,22,27	7,12,17
January	1,6,11,16,21,26,31	21
February	5,10,15,20,25	10,25

Accuracy assessment

The GEE has an inbuilt function 'errorMatrix' which calculates the 'ConfusionMatrix' to assess the accuracy of the classifier. A confusion matrix provides a more detailed breakdown of the model's performance by showing the counts of true positive, true negative, false positive, and false negative predictions for each class. From the confusion matrix, various metrics can be derived, including:

Overall Accuracy (OA): Represents the total classification accuracy. It is obtained by dividing the total numbers of correctly classified pixels by the total number of reference pixels. The drawback of this measure is that it does not tell us about how

Table 2 : Results of trained models in terms of Resubstitution error matrix (training dataset) and training accuracy for sugarcane area estimation.

Classifier	Classes	Resubstitution error matrix (Training)					Training accuracy
		2022-2023					
		Other crops	Sugarcane	Forest	Water bodies	Urban	
RF	Other crops	366	1	0	0	0	99.78%
	Sugarcane	2	344	0	0	0	
	Forest	1	0	278	0	0	
	Water bodies	0	0	0	161	0	
	Urban	0	0	0	0	236	
KNN	Other crops	367	0	0	0	0	100.00%
	Sugarcane	0	346	0	0	0	
	Forest	0	0	279	0	0	
	Water bodies	0	0	0	161	0	
	Urban	0	0	0	0	236	
CART	Other crops	367	0	0	0	0	100.00%
	Sugarcane	0	346	0	0	0	
	Forest	0	0	279	0	0	
	Water bodies	0	0	0	161	0	
	Urban	0	0	0	0	236	
SVM	Other crops	285	51	26	0	5	85.24%
	Sugarcane	28	297	21	0	0	
	Forest	21	31	226	0	1	
	Water bodies	8	0	0	152	1	
	Urban	12	0	0	0	224	
MD	Other crops	239	88	34	0	6	75.88%
	Sugarcane	35	294	17	0	0	
	Forest	22	59	197	0	1	
	Water bodies	9	0	0	113	39	
	Urban	12	0	0	13	211	

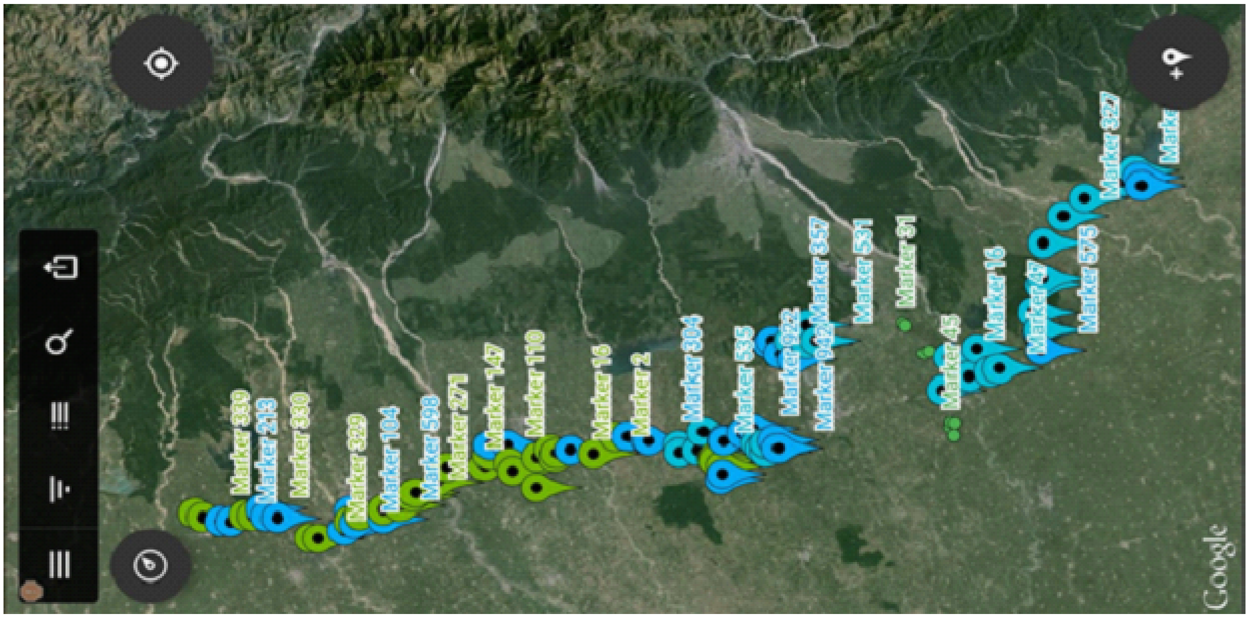

Fig. 3 : Ground truth points of different classes (Sugarcane, Other crops, Forest, Waterbodies, Urban) using Map marker app.

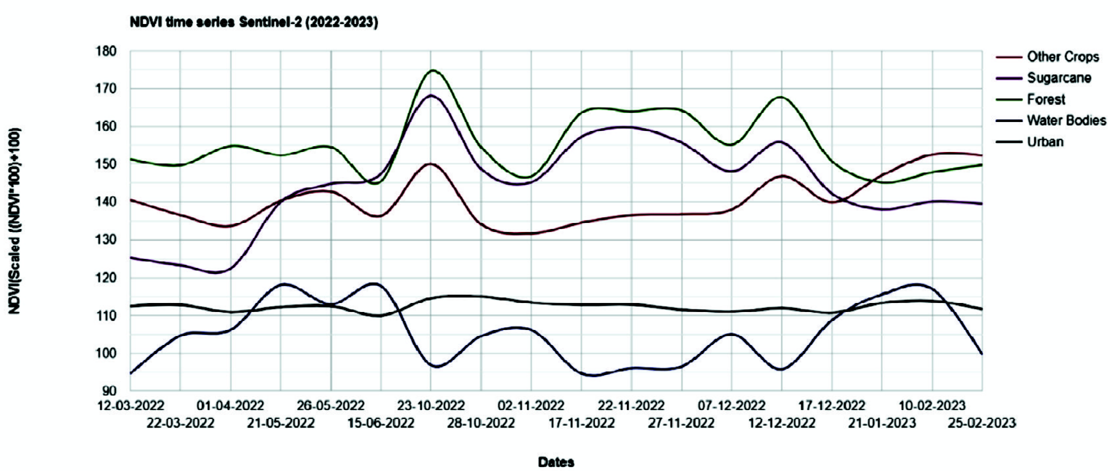

Fig. 4 : NDVI time series curve of different classes using sentinel- 2 (A &B) data.

well individual classes are classified. The producer and user accuracy are two widely used measures of class accuracy depend21 on the omission and commission accuracy respectively.

$$\text{Overall accuracy} = \frac{\text{Number of correctly classified}}{\text{Total Number of reference sites}}$$

Producer's Accuracy (PA): It is defined as the ratio of the number of validation sites of a class classified accurately to the total number of validation sites for the class.

$$\text{Producer's Accuracy (PA)} = \frac{\text{Numbers of pixels correctly classified in each category}}{\text{Numbers of sample pixels taken for each category (column total)}}$$

User's Accuracy (UA): It refers to the probability that a pixel labeled as a certain class in the map is this class. It is obtained by dividing the accurately classified pixels by the total number of pixels classified in this category. It is defined as the ratio of the number of validation sites of a class correctly classified to the number of sites classified as that class.

$$\text{User's Accuracy (UA)} = \frac{\text{Numbers of pixels correctly classified in each category}}{\text{Numbers of sample pixels taken for each category (row total)}}$$

Kappa Coefficient/F1-score: It is the harmonic mean of UA and PA. It is used to test the consistency of ground data and classified data, where K = 1 means that all pixels are correctly identified (Zhang *et al.*, 2019). This statistic is particularly useful for assessing classification accuracy when class distributions are imbalanced. Accuracy can be used when the class distribution is similar while F1-score is a better metric when there are imbalanced classes as in the above case.

RESULTS AND DISCUSSION

NDVI multi-temporal profile

The spatial variations in the NDVI of different classes, such as Sugarcane, Other Crops, Forest, Water bodies, and Urban areas, are indicated by the NDVI time series. These were extracted using the average NDVI of all the training ground points. NDVI time series charts were exported using the chart function in GEE. These variations are primarily caused by the different spectral reflectance properties of the classes, which depend on their growth stages and types. In other crops, we included the Kharif (i.e. Rice, Soybean, Maize) and Rabi season crops (i.e., wheat, pea, mustard, vegetables). The multi-sensor and multi-temporal variance in the NDVI of several classes for 2022-2023 is displayed in Figure 4, NDVI serves as an input band for creating the classification map. Comparing sugarcane (pink line) with other crops and forests, it has a lower NDVI value at the beginning of the

Table 3 : Government-reported area and the area estimated by the different classifiers

Satellite	Year	Reported area (ha) (2020-2021)	CLASSIFIERS									
			RF		KNN		CART		SVM		MD	
			Estimated area (ha)	Devi-ation	Estimated area (ha)	Devi-ation	Estimated area (ha)	Devi-ation	Estimated area (ha)	Devi-ation	Estimated area (ha)	Devi-ation
Sentinel 2	2022-2023	12575	14012	1437	23755	11180	17530	4955	14393	1818	16735	4160

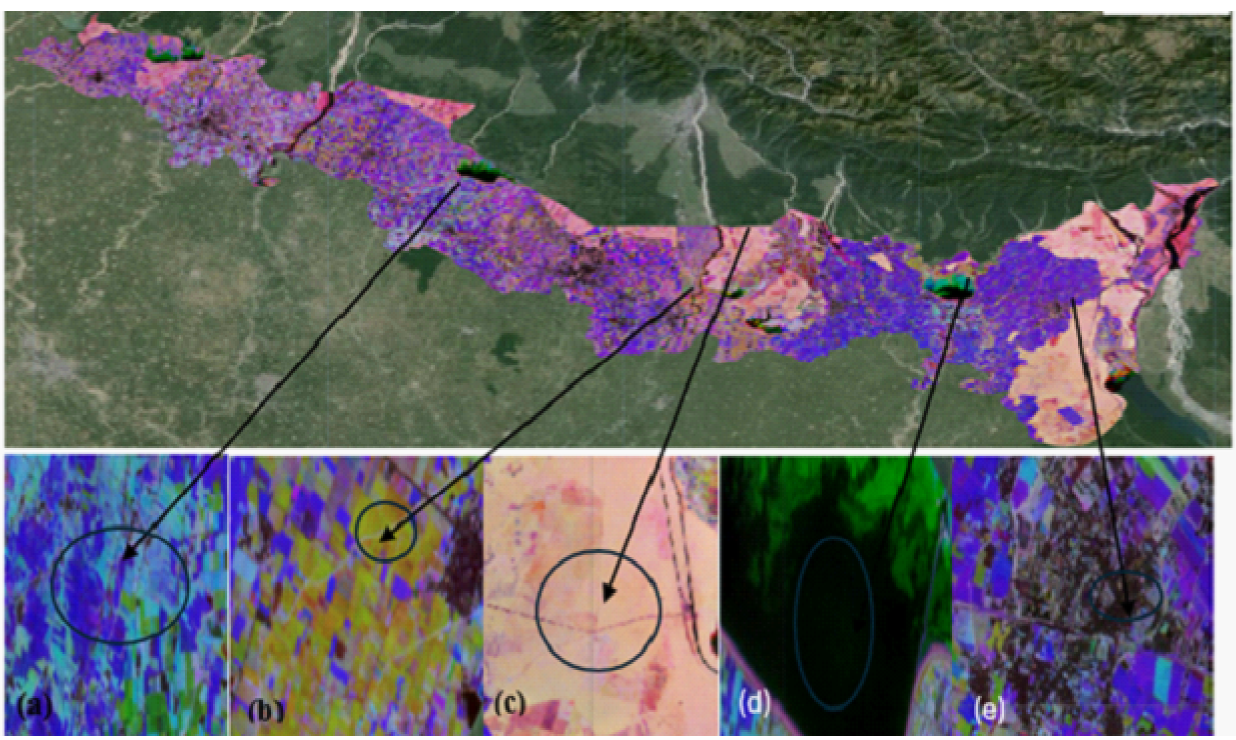

Fig. 5 : Stacked NDVI scaled image composite map of 2022-2023. Visualization of map (a) Other crops (b) Sugarcane (c) Forest (d) Waterbodies (e) Urban

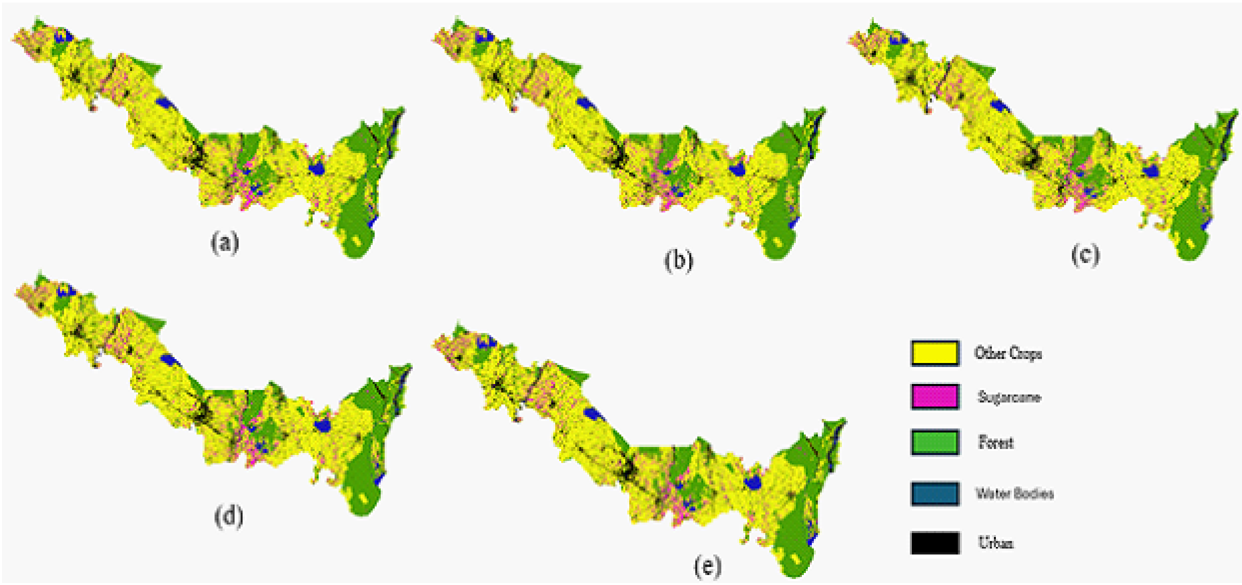

Fig. 6 : Crop classification map produced using (a) Random Forest (RF), (b) K-Nearest Neighbors (KNN), (c) Classification and Regression Tree (CART), (d) Support Vector Machine (SVM) and (e) Minimum Distance (MD) from Sentinel-2 satellite imagery for the Udham Singh Nagar

Table 4 : Results of validated models in terms of Confusion matrix (validation dataset).

Classifier	Classes	Confusion Matrix (Validation) 2022-2023: Other crops	Sugarcane	Forest	Water bodies	Urban
RF	Other crops	156	1	0	0	0
	Sugarcane	9	138	1	0	0
	Forest	1	0	122	0	0
	Water bodies	0	0	0	69	0
	Urban	1	0	0	0	117
KNN	Other crops	147	6	4	0	0
	Sugarcane	12	131	5	0	0
	Forest	0	1	122	0	0
	Water bodies	0	0	0	64	5
	Urban	1	0	0	0	117
CART	Other crops	142	7	6	1	1
	Sugarcane	20	124	3	1	0
	Forest	10	0	113	0	0
	Water bodies	0	0	0	62	7
	Urban	2	0	1	0	116
SVM	Other crops	153	1	2	0	1
	Sugarcane	8	135	5	0	0
	Forest	10	0	123	0	0
	Water bodies	0	0	0	63	6
	Urban	4	0	0	0	114
MD	Other crops	153	1	2	0	1
	Sugarcane	15	127	6	0	0
	Forest	1	0	122	0	0
	Water bodies	0	0	0	69	0
	Urban	4	0	0	1	113

planting season (i.e. March and April) and at the time of harvesting (i.e. December to February) of sugarcane. Compared to other classes, the forest (green line) has the greatest NDVI, with a modest fall in value noted during the winter. No variations in NDVI were seen over the whole period in urban areas (black line). With this spatio-temporal variation in NDVI, the sugarcane crop can be distinguished from the other classes (other crops, forest, water bodies, and urban) and used it as a unique feature for classification in further steps. Vuolo *et al.* (2018) results show how the addition of multi-temporal information increases the crop type classification accuracy.

Sugarcane crop area classification

After combining the NDVI bands to improve our ability to distinguish between different areas based on their NDVI values, we selected NDVI bands from three specific months—March 12th, June 15th, and October 28th, 2022—for visualization using the "add to the map layer" function in GEE. With a primary focus on sugarcane, we assigned these NDVI bands to the channels BGR (Blue, Green, Red). The map displays areas according to their NDVI values during different months. Regions with high NDVI values in January, February, and March are represented in

Table 5: Accuracy assessment of different classifiers with the help of Commission error, User accuracy, Omission error, Producer accuracy, Overall accuracy, and F1 score

Classifier	Class	Commission error	User's Accuracy (UA)	Omission error	Producer's Accuracy (PA)	Overall Accuracy (OA)	F1 Score
RF	Other crops	8.13%	91.88%	0.64%	99.36%	**97.88%**	**97.31%**
	Sugarcane	5.07%	94.93%	6.76%	93.24%		
	Forest	6.87%	93.13%	0.81%	99.19%		
	Water bodies	0.00%	100.00%	0.00%	100.00%		
	Urban	0.00%	100.00%	0.85%	99.15%		
KNN	Other crops	8.13%	91.88%	6.37%	93.63%	94.47%	92.98%
	Sugarcane	5.04%	94.96%	11.49%	88.51%		
	Forest	6.15%	93.85%	0.81%	99.19%		
	Water bodies	0.00%	100.00%	7.25%	92.75%		
	Urban	4.10%	95.90%	0.85%	99.15%		
CART	Other crops	18.39%	81.61%	9.55%	90.45%	90.57%	88.01%
	Sugarcane	5.34%	94.66%	16.22%	83.78%		
	Forest	7.38%	92.62%	8.13%	91.87%		
	Water bodies	3.13%	96.88%	10.14%	89.86%		
	Urban	6.45%	93.55%	1.69%	98.31%		
SVM	Other crops	7.27%	92.73%	2.55%	97.45%	95.61%	94.42%
	Sugarcane	0.74%	99.26%	8.78%	91.22%		
	Forest	5.39%	94.62%	0.00%	100.00%		
	Water bodies	0.00%	100.00%	8.70%	91.30%		
	Urban	5.79%	94.21%	3.39%	96.61%		
MD	Other crops	11.56%	88.44%	2.55%	97.45%	94.96%	93.60%
	Sugarcane	0.78%	99.22%	14.19%	85.81%		
	Forest	6.15%	93.85%	0.81%	99.19%		
	Water bodies	1.43%	98.57%	0.00%	100.00%		
	Urban	0.88%	99.12%	4.24%	95.76%		

blue, while those with high NDVI values in May, June, and July are shown in green. Water bodies appear in dark blue, and urban areas in black. Since sugarcane typically exhibits high NDVI values from May to January, it appears in shades of yellow and yellow-orange on the map (Figure 5).

Crop classification map produced using Machine Learning Algorithms

Using the methodology discussed above, the crop classification maps were generated for the five different algorithms i.e., RF, KNN CART, SVM, and Minimum Distance are shown in Figure 6. For the training of the model; 367 points of sugarcane, 346 points of other crops, 279 points of forest, 161 points of water bodies, and 236 points of urban were used for crop classification and area estimation process. All the different algorithm maps were first exported to the drive from there maps were downloaded and processed in QGIS mainly for clipping and better visualization with more clarity of images. To check the performance of the trained model we used Resubstitution error matrics for training accuracy (Table 2). KNN and CART trained models show 100% training accuracy followed by RF, SVM, and MD. However, only training the model does not provide accurate information that which model is suitable for sugarcane area estimation. Different classifier area estimation varies with each other (Table 3). In 2022-2023 area estimated by RF (14012 ha) was closer to the reported area in 2020-2021 (12575 ha) followed by SVM (14393 ha) MD (16735 ha), KNN (23755 ha), and CART (17530 ha). There were lots of variations observed in area estimation to check the accuracy of different classifiers we need validation of the models.

Accuracy Assessment and Validation

To check the authenticity of the gap between the government-reported area and the area estimated by the different classifiers we went for the validation of the model in the classification process. For this 157 points of sugarcane, 148 points of other crops, 123 points of forest, 69 points of water bodies, and 118 points of urban were used to complete the process. Accuracy assessment was done with the help of the Confusion matrix (Table 4). It is used for the calculation of Commission error, User accuracy, Omission error, Producer accuracy, Overall accuracy, and F1 score. The results of the accuracy assessment of the validation model for the sugarcane class versus other classes (other crops, forest, water bodies, and urban) are provided in Table 5. Among all the classifiers RF classifier performed outstanding than other classifiers with 97.88% overall accuracy and 99.31% F1 score followed by SVM (95.61%, 94.42%), MD (94.96%, 93.60%), KNN (94.47%, 92.98%), and CART (90.57%, 88.01%). These results were favored by the results of Kussul *et al*. (2016), Neetu and Ray (2019), Ge *et al*. (2020), and Panjala *et al*. (2022). They also determined that the Random Forest classifier excels over other classifiers when it comes to accuracy assessment. Mahdianpari *et al*. (2017) and Xia *et al*. (2017) explain the reasons why Random Forest has received considerable interest over the last two decades (1) Good handling of the outliers and noisier datasets; (2) Good performance with high dimensional and multi-source datasets; (3) Higher accuracy than other popular classifiers, such as SVM, KNN or MLC in many applications (Rodriguez-Galiano and Chica-Rivas, 2014; Abdel-Rahman *et al*., 2014) and (4) Increasing the processing speed by selecting important variables (Van Beijma *et al*., 2014).

CONCLUSION

With the wide use of RS technologies, the accuracy of classifiers in classification has become increasingly critical. In this paper, five classifiers RF, KNN, CART, SVM, and Minimum Distance were used to estimate the area of sugarcane for 2022-2023. Land-use classification is created by sensing images taken from the sentinel-2 satellite for the area Udham Singh Nagar which is located in Uttarakhand State. The whole classification and area estimation process was performed in the Google Earth Engine. It is found that with the help of this cloud platform, satellite data accessing, filtering, pre-processing, and export of satellite data can be done very efficiently and quickly as compared to other software used for similar work. GEE has great potential in crop mapping at a regional scale using composite images. We can perform crop-type mapping for any region at any scale. With the help of stacking different vegetation indices time series images, we can easily check the spatial distribution of different crops in the targeted region with the help of ground truth points or knowledge of that particular region. Among the various classifiers, RF outperforms others such as SVM, MD, KNN, and CART, achieving higher overall accuracy and F1 scores. This superior performance allows RF to estimate the area more accurately compared to the reported area. In the regions of the world where sugarcane is cultivated, continuous research and innovation in the area can solve issues related to food security, environmental sustainability, water scarcity, and climate change adaptation. Governments, planners, and decision-makers can use to gather vital information for policy-making.

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to the Department of Agrometeorology, G. B. Pant University of Agriculture & Technology, Pantnagar for extending their guidance and technical assistance in conducting this research work.

REFERENCES

Abdel-Rahman, E. M., Mutanga, O., Adam, E. and Ismail, R. (2014). Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88: 48-59.

Acito, F. (2023). k Nearest Neighbors. In Predictive Analytics with KNIME: Analytics for Citizen Data Scientists, Cham: *Springer Nature Switzerland*, Pp 209-227.

Akponikpe, P., Minet, J., Gerard, B., Defourny, P. and Bielders, C. (2011). Spatial fields dispersion as a farmer strategy to reduce agro-climatic risk at the household level in pearl millet-based systems in the Sahel: a modeling perspective. *Agricultural and Forest Meteorology*, 151 (2): 215–227.

Awad, M. and Khanna, R. (2015). Support vector machines for classification. In Efficient Learning Machines; Apress: Berkeley, CA, USA, 39–66p.

Awokuse, T. O. (2009). Does agriculture really matter for economic growth in developing countries? Technical Report.

Baruth, B., Royer, A., Klisch, A. and Genovese, G. (2008). The use of remote sensing within the MARS crop yield monitoring system of the European Commission. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37: 935-940.

Bishnoi, A., Kumar, D., Nain, A. S., Singh, A., Mor, A. and Bhardwaj, S. (2021). Geospatial technology application for prioritization of land resources in Udham Singh Nagar District of Uttarakhand, India. *Indian Journal of Traditional Knowledge (IJTK)*, 20(2): 595-603.

Bolton, D. K. and Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and* Forest *Meteorology*, 173: 74–84.

Bonsucro,(2017). Corbion Launches its New PLA Bioplastics Produced from Bonsucro Certified Sugarcane [WWW Document]. URL https://www.bonsucro.com/corbionlaunches-new-pla-bioplastics-produced-bonsucro-certified-sugarcane/ (accessed 1.22.20).

Cardona, C. A., Quintero, J. A. and Paz, I. C. (2010). Production of bioethanol from sugarcane bagasse: status and perspectives. *Bioresource Technology*, 101: 4754–4766.

Carreiras, J. M., Jones, J., Lucas, R. M. and Shimabukuro, Y. E. (2017). Mapping major land cover types and retrieving the age of secondary forests in the Brazilian amazon by combining single-date optical and radar remote sensing data. *Remote Sensing of Environment*, 194: 16–32.

Duan, Y., Zhang, W., Huang, P., He, G. and Guo, H. (2021). A New Lightweight Convolutional Neural Network for Multi-Scale Land Surface Water Extraction from GaoFen-1D Satellite Images. *Remote Sensing*, 13(22): 4576.

Fritz, S., See, L., Perger, C., Mccallum, I., Schill, C., Schepaschenko, D., Duerauer, M., Karner, M., Dresel, C., Laso–Bayas, J.C. and Lesiv, M. (2017). A global dataset of crowdsourced land cover and land use reference data. *Scientific Data*, 4: 170075.

Fuhs, N. and Akiyama, T. (1980). Evaluation of several schemes for classification of remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 46(12):1547-1553.

Ge, G., Shi, Z., Zhu, Y., Yang, X. and Hao, Y. (2020). Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms. *Global Ecology and Conservation*, 22: e00971.

Gilbertson, J. K., Kemp, J. and Van Niekerk, A. (2017). Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. *Computers and Electronics in Agriculture*, 134: 151–159.

Gomes, G. R., Rampon, D. S. and Ramos, L. P. (2018). Production of furan compounds from sugarcane bagasse using a catalytic system containing ZnCl2/HCl or AlCl3/HCl in a biphasic system. *The Journal of the Brazilian Chemical Society.*

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R. (2017). Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202: 18–27.

Grof, C. P. L. and Campbell, J. A. (2001). Sugarcane sucrose metabolism: scope for molecular manipulation. *Australian Journal of Plant Physiology*, 28: 1–12.

Hagolle, O., Sylvander, S., Huc, M., Claverie, M., Clesse, D., Dechoz, C., Lonjou, V. and Poulain, V. (2015). SPOT-4 (Take 5): simulation of Sentinel-2 time series on 45 large sites. *Remote sensing*, 7(9):12242-12264.

Hu, S., Shi, L. S., Huang, K., Zha, Y. Y., Hu, X. L., Ye, H. and Yang, Q. (2019). Improvement of sugarcane crop simulation by SWAP-WOFOST model via data assimilation. *Field Crop Research*, 232: 49–61.

Huang, J., Blanz, V. and Heisele, B. (2002). Face recognition using component-based SVM classification and morphable models. In International Workshop on Support Vector Machines (Pp. 334-341). Berlin, Heidelberg: Springer Berlin Heidelberg.

Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X., Fan, J. and Wu, W. (2016). Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to estimate regional winter wheat yield. *Journal of Agricultural Meteorology*, 216:188–202.

Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M. and Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, 228: 115-128.

Kamal, M., Jamaluddin, I., Parela, A. and Farda, N. M. (2019) . Comparison of Google Earth Engine (GEE)-based machine learning classifiers for mangrove mapping. In Proceedings of the 40th Asian Conference Remote Sensing*, ACRS*, Pp. 1-8.

Kavzoglu, T. and Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(5): 352-359.

Kussul, N., Lemoine, G., Gallego, F. J., Skakun, S. V., Lavreniuk, M. and Shelestov, A. Y. (2016). Parcel-based crop classification in Ukraine using Landsat-8 data and Sentinel-1A data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6), 2500-2508.

Lambert, M. J., Traoré, P. C. S., Blaes, X., Baret, P. and Defourny, P. (2018). Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sensing of Environment*, 216: 647-657.

Lebourgeois, V., Dupuy, S., Vintrou, E., Ameline, M., Butler, S. and Begue, A. (2017). A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). *Remote Sensing*, 9: 259.

Li, X., Ling, F., Foody, G. M., Ge, Y., Zhang, Y. and Du, Y. (2017). Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps. *Remote Sensing of Environment*, 196: 293–311.

Li, Y.R. and Yang, L.T. (2015). Sugarcane agriculture and sugar industry in China. *Sugar Technology*, 17:1–8.

Lu, M., Wu, W., Zhang, L., Liao, A., Peng, S. and Tang, H. (2016). A comparative analysis of five global cropland datasets in china. *Science China Earth Sciences*, 59: 2307–2317.

Macintyre, P., Van Niekerk, A. and Mucina, L. (2020). Efficacy of multi-season Sentinel-2 imagery for compositional vegetation classification. *International Journal of Applied Earth Observation and Geoinformation*, 85: 101980.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F. and Motagh, M. (2017). Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and

TerraSAR-X imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130: 13-31.

Mathur, A. and Foody, G. M. (2008). Multiclass and binary SVM classification: Implications for training and classification users. *Geoscience and Remote Sensing Letters*, 5: 241–245.

Moore, P.H. and Botha, F.C. (2013). Sugarcane: Physiology, Biochemistry and Functional Biology. John Wiley & Sons.

Mulianga, B., Begue, A., Clouvel, P. and Todoroff, P. (2015). Mapping cropping practices of a sugarcane-based cropping system in Kenya using remote sensing. *Remote Sensing*, 7: 14428–14444.

Munoz-Mari, J., Bruzzone, L. and Camps–Valls, G. (2007). A support vector domain description approach to supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 45:2683–2692.

Neetu and Ray, S. S. (2019). Exploring machine learning classification algorithms for crop classification using Sentinel 2 data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42: 573-578.

Panjala, P., Gumma, M. K. and Teluguntla, P. (2022). Machine learning approaches and sentinel-2 data in crop type mapping. *Data science in agriculture and natural resource management*, Pp161-180.

Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E. and Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment,* 115: 1301–1316.

Qian, Y., Zhou, W., Yan, J., Li, W. and Han, L. (2014). Comparing machine learning classifiers for object-based land cover classification using very high-resolution imagery. *Remote Sensing*, 7(1): 153-168.

Rodriguez-Galiano, V. F. and Chica-Rivas, M. (2014). Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. *International Journal of Digital Earth*, 7(6): 492-509.

Shield, I. (2016). 11 - sugar and starch crop supply chains. In: Holm-Nielsen, J.B., Ehimen, E.A. (Eds.), Biomass Supply Chains for Bioenergy and Biorefining. Woodhead Publishing, Pp 249–269.

Sindhu, R., Gnansounou, E., Binod, P. and Pandey, A. (2016). Bioconversion of sugarcane crop residue for value added products - an overview. *Renewable Energy,* 98: 203–215.

Song, X. P., Potapov, P. V., Krylov, A., King, L., Di Bella, C. M., Hudson, A., Khan, A., Adusei, B., Stehman, S. V. and Hansen, M. C. (2017). National scale soybean mapping and area estimation in the united states using medium resolution satellite imagery and field survey. *Remote Sensing of Environment,* 190: 383–395.

Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M. and Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research*, 50: 1250–1265.

Torabi, M., Hashemi, S., Saybani, M. R., Shamshirband, S. and Mosavi, A. (2019). A Hybrid clustering and classification technique for forecasting short-term energy consumption. *Environmental Progress & Sustainable Energy*, 38: 66–76.

Van Beijma, S., Comber, A. and Lamb, A. (2014). Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote Sensing of Environment*, 149:118-129.

Van der Velde, M., Biavetti, I., El-Aydam, M., Niemeyer, S., Santini, F. and van den Berg, M. (2019). Use and relevance of European union crop monitoring and yield forecasts. *Agricultural Systems*, 168, 224–230.

Van Klompenburg, T., Kassahun, A. and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177: 105709A.

Verma, A. K., Garg, P. K., Prasad, K. H. and Dadhwal, V. K. (2023). Variety-specific sugarcane yield simulations and climate change impacts on sugarcane yield using DSSAT-CSM-CANEGRO model. *Agricultural Water Management*, 275:108034.

Virnodkar, S. S., Pachghare, V. K., Patil, V. C. and Jha, S. K. (2020). Remote sensing and machine learning for crop water stress determination in various crops: a critical review. *Precision Agriculture,* 21: 1121–1155.

Vogels, M. F., De Jong, S. M., Sterk, G. and Addink, E. A. (2019). Mapping irrigated agriculture in complex landscapes using SPOT6 imagery and object based image analysis–a case study in the Central Rift Valley, Ethiopia. *International Journal of Applied Earth Observation and Geoinformation,* 75: 118–129.

Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C. and Ng, W. T. (2018). How much does multi-temporal Sentinel-2 data improve crop type classification?. *International journal of applied earth observation and geoinformation*, 72: 122-130.

Wei, M., Qiao, B., Zhao, J. and Zuo, X. (2018). Application of remote sensing technology in crop estimation, in: 2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing,(HPSC) and IEEE International Conference on Intelligent Data and Security (IDS), IEEE., Pp. 252–257.

Xia, J., Falco, N., Benediktsson, J. A., Du, P. and Chanussot, J. (2017). Hyperspectral image classification with rotation random forest via KPCA. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(4): 1601-1609.

Zhai, H., Zhang, H., Zhang, L. and Li, P. (2019). Total variation regularized collaborative representation clustering with a locally adaptive dictionary for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57: 166–180.

Zhang, H., Kang, J., Xu, X. and Zhang, L. (2020). Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China. *Computers and Electronics in Agriculture*, 176: 105618.

Zhang, X. W., Liu, J. F., Qin, Z. Y. and Qin F. (2019). Winter wheat identification by integrating spectral and temporal information derived from multi-resolution remote sensing data. *Journal of Integrative Agriculture*, 18: 2628–2643.

Zhong, L., Hu, L., Yu, L., Gong, P. and Biging, G. S. (2016). Automated mapping of soybean and corn using phenology. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119: 151–164.

Received: June 20, 2024
Accepted: July 26, 2024