International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Jun-24)

http://dx.doi.org/10.12785/ijcds/1501123

Mapping Crop Types At a 10 m Scale Using Sentinel-2 Data and Machine Learning Methods

Atiya Khan[1], Chandrashekhar H. Patil[1,*], Amol D. Vibhute[2] and Shankar Mali[1]

[*]*Corresponding author: Chandrashekhar H. Patil, chpatil.mca@gmail.com*
[1]*School of Computer Science, Dr. Vishwanath Karad MIT World Peace University, Pune, MH, India.*
[2]*Symbiosis Institute of Computer Studies and Research (SICSR), Symbiosis International (Deemed University), Pune-411016, MH, India.*

Received 8 Nov. 2023, Revised 4 Apr. 2024, Accepted 6 Apr. 2024, Published 1 Jun. 2024

Abstract: Crop classification plays a vital role in crop status monitoring, crop area estimation, and food production. Remote sensing data is widely accepted for crop classification at remote locations. However, crop classification is challenging due to spectral and spatial similarities, complex land structures, temporal inconsistencies, and environmental parameters. Machine learning models must be robust, particularly when dealing with a variety of crop types and changing environmental factors. This study examines the extent to which various algorithms generalize, emphasizing the importance of adaptability to various farming systems and geographical conditions. It explores transfer learning and ensemble approaches as possible ways to improve the resilience and flexibility of the model. In the present study, an effort has been made to identify, classify, and map multiple crops from the complex environment using the Sentinel-2 dataset and advanced machine learning methods such as random forest, Spectral Angle Mapper (SAM), Maximum Likelihood Classifier (MLC), K-means clustering and Iterative Self-Organizing Data Analysis Technique (ISODATA). The crop spectral features were identified using the Normalized Difference Vegetation Index (NDVI). The NDVI outcomes ranged between -0.91 and 0.54, which were then used to identify crop areas. Ground reference data, Google map, and Google Earth data were used to determine the crop classes, train the data, and validate the results. The five major crops viz. Cotton, Paddy, Orchard, Yellow split Pigeon peas, Chickpeas, and Other crops were identified and classified efficiently. According to the experimental results, the random forest approach had the best overall accuracy, 87.71%, and a kappa value of 0.86 than other methods. Alternatively, the ISODATA method provided an overall accuracy of 85.01% with a kappa value of 0.82. The agricultural decision-makers can use the results of this study for decision-making and management.

Keywords: Crop types mapping, Sentinel-2 data, Machine learning, Random Forest,NDVI.

1. Introduction

Agricultural productivity must be improved with the latest techniques to fulfill future food demand and sustainability goals. In this case, precision farming is increasing rapidly due to its high performance in growing crops and enhancing food production. Precision farming comprises a wide variety of technical capabilities, such as satellite data, unmanned aerial systems applications, machine learning, route planning, and conversation, with a focus on assisting farmers and maintaining a healthy atmosphere to attain sustainable development, climate-related goals, and financial gains [1]. Farmers are more conscious about crop growth, water or fertilizer use, hazardous infections, and food production. Thus identification of crops and their diseases is essential in crop monitoring, crop area estimation, and crop classification [2] for better food production. However, crop identification and its classification is a challenging task by traditional methods. Recently, remotely sensed satellite images [3] play an essential role in agricultural practices to attain precision agriculture goals with satisfactory accuracy. Remote sensing technology offers multi-temporal datasets that can be used in crop monitoring, allowing farmers to react quickly before possible problems proliferate and negatively affect crop production. The crop identification and mapping are enhanced using multi-temporal remote sensing imagery, which changes reflectance as a result of plant phenology, i.e., stages of plant growth to facilitate classification. Therefore, multi-temporal remote sensing sensors are required to monitor the crops throughout the growing season. Furthermore, crop classification is a useful tool for crop insurance and agricultural yield modeling. Given the growing emphasis on crop insurance in India, field-level stakeholders must have access to geographic data about crop stress, crop health, and crop type [4]. Moreover, vegetation fingerprints are measured by remote sensing to assess the impact of regional and global droughts on agricultural productivity. In some places, the NDVI method can also be used to differentiate between cultivated and

E-mail address: atiyakhan15@gmail.com, chpatil.mca@gmail.com, amolvibhute2011@gmail.com

wild plants. The vegetation indices help distinguish between areas with and without vegetation as well as between living and dead vegetation. To generate a vegetation index, spectral values from multiple bands are often combined, multiplied, or split [5]

However, crop identification and classification remain challenging due to the complexity of agricultural landscapes and the comparable spectral signatures of different crops [6]. Therefore, the use of high spatial and spectral resolution satellite data for crop classification is essential. The Sentinel-2 satellite has been launched recently and has 13 multispectral bands. Its short revisit time makes it a valuable tool for mapping vegetation [7]. With some restrictions, research on the classification of crops using satellite images has been carried out in this field. For instance, in the research paper [2], vegetation and land cover in a specific region of Italy were detected using Sentinel-2 multispectral data. Significant vegetative growth information and geographical NDVI patterns were utilized to evaluate crop phenological cycles using the most accurate multi-temporal imagery that was selected using the separability measure [2]. The crop areas were determined by using the NDVI data, according to the research findings [8]. The extraordinary rainfall increased crop area and enhanced agricultural conditions [8]. The NDVI crop forecast model was also used to determine and monitor crop status [9]. Additionally, the vegetation cover was classified using K-means clustering algorithms after the NDVI values were computed using Sentinel-2 data [10]. A distinct study [11] claims that the red-edge band 1 and near-infrared band 1 of Sentinel-2 are better at classifying crops than the other bands. The paper's authors [12] used multispectral images with Sentinel-2 imagery to improve the spatial resolution for precisely identifying crop field borders in small, fragmented agricultural areas.

Absence plots, such as those including crops, water bodies, or impermeable areas, can be generated using comparatively large geographical remote sensing data to evaluate the accuracy of natural vegetation maps [13]. The implementation of spectral indices improved classification accuracy, according to the research result [14]; nevertheless, the combined use of spectral indices with reflectance increased the negative effects of large sets of related elements and decreased accuracy. [15] state that each variable's relevance indicates how much it contributes to the categorization process.

According to authors [16], integrating red-edge bands and combining SWIR can greatly increase crop identification accuracy when compared with using traditional bands for visible and NIR bands [16]. Nonetheless, few studies have used machine learning methods to categorize crops. For instance, the RF classification techniques help manage land, categorizing crops, identifying regions with erroneous labels, and categorizing a variety of crop species [17].

Higher accuracy in crop classification has been shown using the SVM on GLSM-based spatial features and high-resolution IRS-LISS-IV images [18]. The excellent precision of the results, particularly for times of vegetative development, highlights the significance of merging Sentinel-2 data with ML algorithms to identify rice crops [19]. The unsupervised ISODATA and NDVI methods were used for harvested area estimate using Sentinel-2 data [10]. The work [20] focused on crop area classification using IRS-LISS-III pictures, maximum likelihood classification, and advanced fuzzy convolution. The study found that crop types may be effectively classified using machine learning approaches including SVM, ANN, and RF [4].

On the other hand, NDVI-based vegetation cover classification was the main focus of earlier research. These investigations have also employed low-resolution datasets, which yield little information. Low-resolution records make it difficult to identify several crops in complicated locations, and the topic is not well-researched. These works concentrated on low-resolution datasets for crop recognition with a moderate level of accuracy, together with supervised or unsupervised approaches.

Models developed for one place cannot be transferred to the other region and vice versa due to spatiotemporal changes. It is challenging to distinguish and classify the different crops from the complex locales. Therefore, the current study detects and classifies many crops from complex locales using NDVI, supervised and unsupervised algorithms, high-resolution Sentinel-2 data, and various other methodologies. This work mapped crop kinds using a range of advanced machine learning algorithms, such as MLC, Random Forest, K-means clustering, ISODATA, and SAM, using the Sentinel-2 dataset.

The current study used machine learning (ML) approaches to identify crop types using Sentinel-2 data, choosing them according to their distinct qualities and capacities. As Random Forest can handle complicated, high-dimensional data with exceptional precision and robustness, it is a good choice for utilizing the extensive spectrum data that the Sentinel-2 bands provide. Because the SAM can assess spectral similarity, it is used. This skill comes in particularly handy when working with crops that have different spectral fingerprints. The Maximum Likelihood Classifier (MLC) is a straightforward method that works well if classes have a normal distribution within the feature space because of its simplicity and interpretability.

A technique for handling big datasets and identifying natural geographical patterns in Sentinel-2 images is presented using the K-means clustering technique. The ISODATA approach is used in scenarios where the number of clusters cannot be fixed and may vary over the terrain due to its adaptability in dynamically altering cluster shape and size during clustering. Together, these algorithms aim to offer a complete crop-type mapping solution that addresses the different challenges posed by the dynamic and intricate nature of agricultural landscapes.

Our goal is to retain fast and accurate crop classification while minimizing dependency on large sample datasets. The primary goals are:
(1) to use Sentinel-2 satellite data,
(2) to use the NDVI approach for determining the crop features
(3) to apply supervised and unsupervised machine learning methods such as random forest, SAM, MLC, K-means clustering, and ISODATA for multiple crop classification,
(4) To evaluate crop mapping's potential at a regional level, and
(5) To evaluate and compare the accuracy.

The current paper is organized into five sections. Background information, prior study limitations, and the significance of the current investigation are covered in the first section. In part two, the examined area and datasets are provided. Section three is devoted to the adopted methodology. In section four, the findings are discussed in previous research. In the final section, the investigation is concluded and future scope is discussed.

2. Study area and data

A. Study Area

The Nagpur district of Maharashtra, India, was chosen as a study area between 21°9' 9.432" N latitudes and 79° 5' 17.52" E longitudes as shown in Figure 1. The Nagpur District covers approximately 9897 km2 area. The forest region covers 28 percent of the area, with an average altitude of 274.5 meters to 652.70 meters above sea level. The spatial reference (projection) was set to be Universal Transverse Mercator (UTM) zone 43 North with WGS-84 datum. The cultivated land of Nagpur district falls into three categories such as watered or garden land, rice lands, and dry cropland. However, due to irregular monsoon, the dry agricultural lands are classified as Rabi (late monsoon) and Kharif (early monsoon).

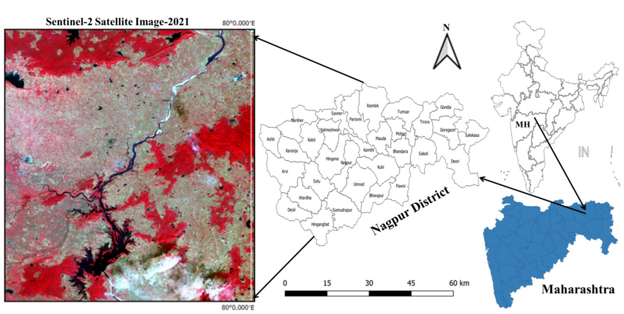


Figure 1. The satellite image of study area

The major Kharif crops of the studied region are Gossypium, Sorghum bicolor, Oryza sativa, Arachis hypogaea, Vigna mungo, Vigna radiata, Sesamum indicum, Ricinus communis, Macrotyloma uniflorum, Capsicum annuum Cowpea, Cucurbitaceae, Solanum melongena, Abelmoschus esculentus, and green vegetables. The Kharif season begins in the middle of June and ends in the middle of July. These crops are typically harvested during October and December. The Rabi season begins in the middle of October and concludes in the middle of February. However, the Rabi crops are more essential in the middle region of the district, which includes the southern half of Ramtek Tehsil and the eastern parts of Nagpur and Umrer Tehsils. The southwest monsoons are few and unpredictable [16] [21].

B. Sentinel-2 data

The European Space Agency (ESA) launched the Sentinel-2 satellite on June 23, 2015, from the Copernicus programme to deliver a plethora of data and imagery. The Sentinel-2 is an optoelectronic multispectral sensor with 13-spectral bands with 10m, 20m, and 60 m resolutions in the visible, NIR, and SWIR spectral regions. The details of Sentinel-2 data are listed in Table 1, along with their names, spatial resolution, and wavelengths [10].

This study obtained a Level 1C Sentinel-2 image on 19 November 2021 from ESA scientific hub. Sentinel-2's four bands (band 8 (NIR), band 4(Red), band 3 (Green), and band 2 (Blue)) at 10m resolutions were used for crop identification. These four high spatial resolution bands have provided better features and a wide range of spectral data to describe crop topology. Every pixel in the Sentinel-2 image has provided the best features which were used in crop classification.

C. Ground truth dataset

The ground dataset for 300 sample points was collected from road travel from October 10 to December 4, 2021, covering about 500 kilometers of area. A specific quantity of ground truth samples is required to ensure crop classification accuracy. During the field study for the Nagpur district area was carried out using a portable GPS Map Camera App with a positioning precision of ± 2 m. The 300 fields' boundaries were subsequently identified by utilizing Google Earth's high spatial resolution images.

The image training was done based on comprehensive information about specific spots, such as identifying crop classes [22] and their labeling. The information, such as GPS points, terrestrial objects, and crop types, was obtained from various field sites. Simultaneously, discussions were done with local farmers to confirm the cropping patterns. The latitude/longitude values and crop type information were collected at the point location to validate the crop classes. The ground samples were collected within the large continuous regions of specific terrestrial objects. As shown in Figure 2, the Latitude and longitude information were gathered along with crop information in the field. During the season, the main crops identified were wheat, cotton, chicken peas, oranges, and vegetables.

The figure 2 displays the training reference samples that were gathered during field survey. In this study, every effort

Table I: List of multispectral bands of Sentinel-2

Bands Name	Central Wavelengths (µm)	Spatial Resolution (m)
Coastal aerosol- Band1	0.443	60
Blue - Band 2	0.49	10
Green - Band 3	0.56	10
Red- Band 4	0.665	10
Vegetation Red Edge - Band 5	0.705	20
Vegetation Red Edge - Band 6	0.74	20
Vegetation Red Edge - Band 7	0.783	20
NIR- Band 8	0.842	10
Vegetation Red Edge- Band 8 A	0.865	20
Water vapour- Band 9	0.945	60
SWIR – Cirrus- Band 10	1.375	60
SWIR- Band 11	1.61	20
SWIR- Band 12	2.19	20

(a) Vegetable plantations in the study area

(b) Corn plantations in the study area.

(c) Cotton plantations in the study area.

(d) Chicken peas plantations in the study area.

Figure 2. Ground truth images gather using smart phone and GPS MAP CAMERA APP during the field survey of Nagpur region.

is taken to gather data covering all class variations in order to improve the machine learning algorithms' accuracy.

In addition, the standard False Color Composite (FCC) of the Sentinel-2 image was generated and used for crop identification and classification. The ground sampling was done based on Google Maps, which was used for detecting the road network and regular footpaths. Every five or ten kilometers, either by car or on foot, the sample positions were selected along the road networks. Figure 3 illustrates a standard FCC of the studied area generated from Sentinel-2 images where NIR (8th band), Red (4th band), and Green (3rd band) bands are displayed in red, green, and blue colors accordingly.

A single-season cropping system is applied between October to December. The studied areas' crop classes were divided into several classes such as Cotton, Chickpeas, Paddy, Yellow split Pigeon peas, Orchard, and other crops. The term "orchard" refers to a group of trees planted in the form of a fruit garden, such as guava, orange, and mango trees, in the field and along the fields' margins. Other crops include mixed crops such as trees and cotton, chickpeas and yellow split pigeon peas, and vegetable crops (Cauliflower, Radish, and Cabbage). The main crops in the study area are Cotton, Paddy, Yellow Split Pigeon Peas, and Chickpeas. Table 2 shows the available crop calendar specifying the sowing and harvest period followed in the studied region.

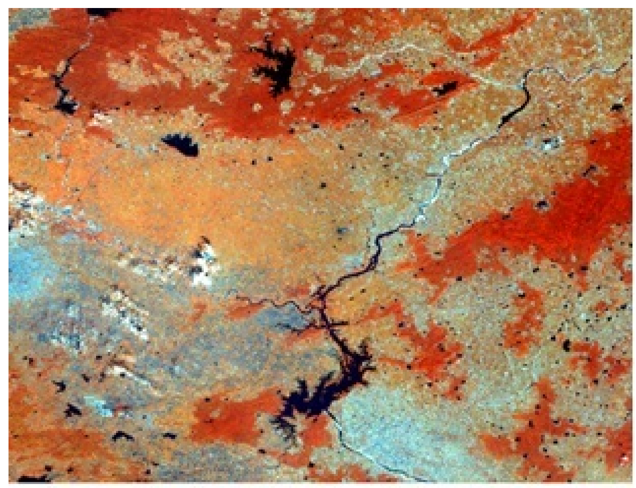


Figure 3. Standard false color composite of the study region derived from the Sentinel-2 image (RGB (8-4-3))

3. The methodology

The implemented methodology and workflow are provided in four sections: data pre-processing, crop feature extraction using the NDVI method, data classification in different crops with supervised and unsupervised algorithms, and accuracy assessment of classification. Figure 4 shows the process of the implemented methodology. In this study, we have used QGIS 3.10 software for the processing of

Table II: Phenological calendar of major crop types followed in the Nagpur District

Crop Type	Code	Sowing Period	Harvesting Period
Cotton	CO	June - July	December-February
Chick pea	CH	June - July	December-January
Paddy	PA	June - July	Mid - September
Pigeon pea	PP	June - July	January-February
Orchard	OR	Early June	October-December

Sentinel-2 data and results evaluation.

A. Data pre-processing

To correct the crucial information, the data that was downloaded was first preprocessed. Level-1C formatted raw data was accessible; level-2A formatted data must be transformed. Top of Atmosphere (TOA) reflectance was converted to Bottom of Atmosphere (BOA) reflectance, which produced the original Sentinel-2 data [23]. ROIs were used to crop the study area's geographic subset. The Sentinel-2 image was corrected for atmospheric and radiometric factors using QGIS software. After the data products' spectral bands were recovered, atmospheric alterations were made to each band to improve it. Furthermore, the information was converted into a Level 2A version for additional processing [10]. The raw image was transformed into an appropriate format through preprocessing. At a spatial resolution of 10m, band stacking (NIR, Red, Blue, and Green) was performed on the preprocessed image. Ultimately, NDVI-based crop feature extraction, feature selection, and feature integration were carried out using these preprocessed and stacked image data (level-2A).

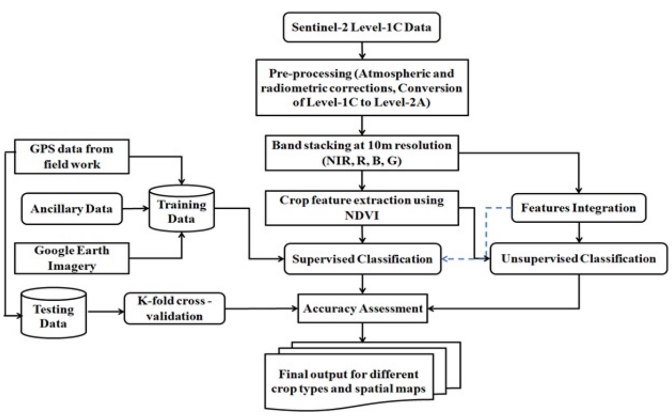


Figure 4. The implemented methodology's workflow

B. NDVI-based crop feature extraction:

A useful tool for distinguishing between areas that are not vegetated and those that are, as well as among living and dead vegetation, is a vegetation index. Among the most popular indices used in several types of research is the NDVI [24]. Consequently, the NDVI approach was used in the current study to extract the crop features for every category of interest from Sentinel-2 data. Equation 1 was used to take into consideration for red bands and the NIR ratio when applying the NDVI to Sentinel-2 data [10].

$$NDVI = \frac{\rho(B8) - \rho(B4)}{\rho(B8) + \rho(B4)} \quad (1)$$

Where ρ is the reflectance value for the corresponding band, B8- NIR band, and B4-Red band. Equation 1 is based on the fact that vegetation leaves absorb the red band and strongly reflect infrared light due to their chlorophyll concentration. Healthy and well-nourished vegetation will receive the majority of visible wavelengths and reflect a substantial proportion of NIR light. In contrast, poor or weak vegetation will reflect less NIR and more visible wavelength light. As a result, a significant variation between the red and NIR values of the same pixel is obtained [5]. These NDVI-based spectral features were used in differentiating various crops. Vegetated areas in the study region are stated as pixels with an NDVI value ranging between -0.76 and 0.80. Figure 5 depicts the thematic map derived using the NDVI method. The crop features have provided positive values near the positive one (0.034 to 0.545) and other class values were negative (Figure 5). The crop features were successfully identified and provided to the classification methods. The NDVI-based spectral features have given relevant information on various crops that were used in classification.

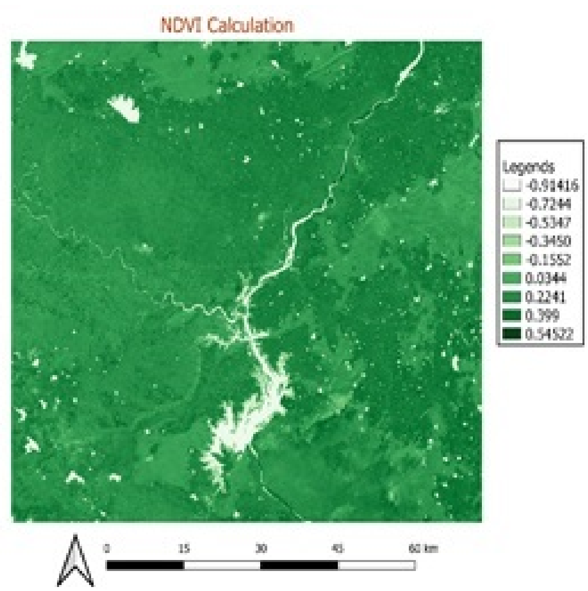


Figure 5. The thematic map derived from NDVI method

C. Data classification

In the present study, three supervised methods, such as random forest, SAM, and MLC, and two unsupervised methods, such as ISODATA and K-means clustering, have been implemented on pre-processed Sentinel-2 data for the classification of different crops. The goal of both approaches is to reveal the spatial distribution of various crops. The supervised and unsupervised classification approaches required some understanding of the subject matter. The essential factors are (1) the spectral data quality to be used in the classification process and (2) the depth of class detail necessary for classification.

1) Random forest (RF) method

The RF is one of the popular algorithms based on Breiman's ensemble concept, which uses many decision trees or classification regression trees to classify the features [25]. The Gini index (Eq. 2) has been used in the random forest method to determine the connections between the nodes on a decision tree branch.

$$Gini = 1 - \sum_{i=1}^{c} (pi)2 \tag{2}$$

Where c is the no. of classes and pi is the relative frequency of the class in the dataset. Eq. 2 determines the Gini of each branch on a node depending on the class and probability and then decides which branch is better for that class. Additionally, entropy is used to determine how the nodes in a decision tree branch are based on the probability of a particular outcome. The entropy was calculated using Eq. 3.

$$Entropy = \sum_{i=1}^{c} - pi * \log 2(pi) \tag{3}$$

The random forest approach has effectively reduced the model fitting by providing randomization to training samples and classification variables [11]. Two elements were modified for RF: (i) the number of trees formed by simply picking samples from the training data is given by the n-tree parameter, and (ii) the number of parameters required for tree node splitting is given by the m-tree (m-try) parameter. The Gaussian process was used to find the best combinations of these hyperparameters. The Bayesian optimization techniques are widely used to tune hyperparameters in machine learning techniques [14]. The n-tree parameter was set to 1000 for the data. Since earlier research has shown trees based on randomly selected samples, there is no rise in the number of faults above 1000. We utilized a Sentinel-2 image using ten bands in this investigation to identify crops, hence the input data consisted of ten variables. Several values (m-try=1:10; n-tree=100, 200) were verified for all datasets to identify the best RF classifier parameters.

2) Spectral Angle Mapper

Another supervised approach used in this study is the SAM method, also known as a spectral similarity metric. It determines the degree of spectral similarity between reference or end members and the spectra of each image pixel. In SAM, the image's pixels are matched to the reference spectra using the n-D angle. Discrimination is determined by the angle that is the smallest between two spectra; that is, a closer match is seen between reference and image pixels that correlate with a smaller angle. For classification, other pixels were disregarded. The method calculates the "spectral angle" between both spectra and reads them as vectors inside a space whose dimension equals the number of bands to find their similarity (nb). The SAM classification technique received the trained ROIs as input. To calculate the degree of similarity between an unknown spectra t and a reference spectra r, the SAM approach has been applied using Equation 4 [26].

$$\alpha = \cos^{-1}\left[\frac{\sum_{i=1}^{nb} t_i r_i}{\left[\sum_{i=1}^{nb} t^2 i\right]^{1/2}\left[\sum_{i=1}^{nb} r^2 i\right]^{1/2}}\right] \tag{4}$$

where, α- spectral angle, t- the test spectrum, nb- band numbers, and r- the reference spectrum. In QGIS, the SAM approach is applied by considering the number of train classes or reference spectrum derived from the selected ROI area. As a result, a classified image is created for every pixel with the most vital SAM matches. The rule images show greater striking similarity with the reference spectrum and lighter pixels at smaller spectral angles.

3) Maximum Likelihood Classification

Bayes' decision-making theorem and the normal distribution of cells in every category sample in multidimensional space serve as the foundation for the MLC technique. An algorithm for multivariate statistical classification is called the MLC. Conversely, the MLC technique analyzes the variance as well as the covariance of the test dataset, assigning each pixel to a certain class that is defined in the training set. The discriminant functions that need to be applied to each pixel of a picture to attain maximum likelihood classification are found using Eq. 5 [24], [27].

$$gi(x) = lnp(\omega i) – 1/2ln|\sum i| – 1/2(x - mi)T \sum_{i}^{-1} (x - mi) \tag{5}$$

Here, class is denoted by i, x is n-dimensional data, n is the number of bands, p(ωi) is the chance that class, ωi, occurs in the image and is taken to be the same for all classes,$\sum$i is the determinant of the covariance matrix of the data in class ωi, $\sum$i-1 is its inverse matrix, and mi is the mean vector.

4) K-means clustering

By randomly allocating samples to the cluster centroid that is closest to each sample, as measured by Euclidean distance, the K-means clustering technique divides n samples into k groups. The cluster centroids are adjusted using the cluster average of observations [28]. The strategy makes use of a simple technique to categorize a given data set into a predefined number of clusters (assuming k clusters). The idea is to establish k centers, one for each cluster. Reducing the squared error function, or total intra-cluster variance, to the lowest possible level is the aim of k-means clustering [10]. To calculate the k-means clustering technique, use equation 6.

$$\mathbf{J} = \sum_{\mathbf{j=1}} \mathbf{k} \sum_{\mathbf{i=1}} \mathbf{n} \parallel \mathbf{x}i \ \ (j) - \mathbf{c}j \| 2 \tag{6}$$

Where: J=objective function, k = no. of clusters, n = no. of classes, cj = centroid for the cluster j.

5) ISODATA

Except for allowing for a variable number of clusters, the ISODATA algorithm and the k-means clustering algorithm are quite similar. However, the k-means algorithm requires that the number of clusters be known prior.

4. Analyzing the accuracy of the classification

To evaluate the acquired categorization accuracy, QGIS software was employed. The confusion matrix has been used to assess how accurate satellite image classification is. Moreover, the mapping of producer and user accuracy (PA) has been done through the computation of overall accuracy (OA). The testing samples were used to perform the accuracy assessment. The confusion matrix has been used in conjunction with the kappa coefficient to produce PA, OA, and user's accuracy (UA) [24]. The efficacy of each class was evaluated using the producers' and users' accuracy methodologies. The kappa coefficient is calculated within the range of -1 to 1 to express the precision. [29] states that an accuracy of more than 0.70 is considered suitable. Equations (7) are used to calculate the OA, PA, UA, and kappa coefficient, respectively [29].
i) OA % = (Properly classified pixels / Total number of pixels)
ii) UA = (Properly classified pixels / Classified total pixels)
iii) PA = (Appropriately classified pixels / Reference total pixels)
iv) Kappa coefficient (K)

$$K = \frac{N \ \sum_{i=1}^{n} m \ \ i,i \quad - \sum_{i=1}^{n} (Gi \ \ Ci)}{N2 \ - \ \sum_{i=1}^{n} (Gi \ \ Ci)} \tag{7}$$

Where: J=objective function, k = no. of clusters, n = no. of classes, cj = centroid for the cluster j.
Where: class number is indicated by i; N is the number of values that have been classified in contrast to truth values; mi, i is the the amount of truth class i points that have also been classified as class i; Ci is the total number of values that have been predicted to belong to class i; and Gi is the overall quantity of truth values in class i.

5. Results and discussions

The study under consideration constitutes a significant advancement in the domain of agricultural mapping, particularly in the integration of state-of-the-art ML techniques with the Sentinel-2 dataset. The machine learning algorithms are pioneers in this study. The innovative aspect is in the extensive study and use of these cutting-edge techniques to effectively map many crops in sophisticated environments. This research aims to push beyond the limits of current knowledge by combining cutting-edge remote sensing data with advanced ML algorithms. By doing so, it will provide new insights and methodologies for precise and detailed crop mapping, thereby making a significant contribution to the field of agricultural research and spatial analysis.

A. Training and Testing data sample

High-resolution Google Earth images and the original Sentinel-2 data were manually inspected in the field to provide the training data. The ROI tool in QGIS 3.10 was used to build 100 polygons for every vegetative category to collect training data samples. The pixel value for every crop category is differed due to differing polygon sizes, as shown in Table 3.

The samples were split into three groups in a stratified random sampling technique. Classification models are created using the training set, and classification precision is assessed using the testing set [30]. Table 3 indicates the number of field for every kind of crop.

The satellite's image, which was utilized to determine the crop class and macro class and calculate the spectral signatures, received the created training pixels as an input. Each ROI is identified by a class ID, and each ROI is assigned to a land cover category using a macro class ID [31]. Eq. 8 was used to calculate the number of samples [32].

$$N = (\sum i = 1 \ (Wi * S\,i / S\,0) 2) \tag{8}$$

In equation 8 ,Where, Wi = Portion of class i's allocated area; Si = stratum i standard deviation; So = overall accuracy's predicted standard deviation; c = overall number of classes. The Table 4 shows the generated training pixels using the Eq.8.

Table 4 shows the sum of pixels and percentage of each class found in the classification report, divided by 100 to get the needed portion of each class i's allocated area denoted by Wi. Let us assume overall accuracy's predicted standard deviation as S0=0.01 and calculate the stratum i standard deviation as Si. Therefore, N = (0.2591/0.01)2=671 is the number of samples we should distribute among classes.

Considering the complexity of interpreting more spectral bands, a great volume of information in Sentinel-2 images presents both possibilities and operational challenges.

Table III: Crop classes along with training and testing samples used in this experiment

Class ID	Macro class name	Training Data (polygon/pixels)	Validation Data (Pixels)	Test Data (pixels)
1	Cotton	100/1265	600	667
2	Chickpea	100/1427	586	672
3	Paddy	100/900	630	682
4	Pigeon pea	100/743	625	645
5	Orchard	100/1305	592	620

Table IV: Classification report

Class	**Pixel Sum**	**Percentage%**	**Area** [metre^2]	**Wi**	**Si**	**Wi*Si**
1	4694009	3.969405599	469400900	0.039694056	0.5	0.019847028
2	5625042	4.75671717	562504200	0.047567172	0.4	0.019026869
3	32713231	27.66336457	3271323100	0.276633646	0.3	0.082990094
4	72251997	61.09862197	7225199700	0.61098622	0.2	0.122197244
5	2970429	2.511890689	297042900	0.025118907	0.6	0.015071344
			Total			0.259132578

Subsequently, the training pixels were given to supervised algorithms such as random forest, SAM, and MLC and classified the images accordingly. The random forest identified the crop features based on accurate training pixels, NIR bands, and NDVI-based [33] input values. The MLC method predicted results based on training ROI resulting in optimum statistical likelihood values. The spectral angle between two pixels and spectral signals that are integrated over the ROIs have been used in SAM based method for classification. Finally, the ISODATA method combined identical clusters and partitioning clusters with large standard deviations and the k-means clustering method arranged features into N-clusters from a vector layer. The classification results of random forest, SAM, MLC, k-means clustering, and ISODATA methods are given in Figure 6. It was observed from Figure 6 that the cotton crop was accurately classified by random forest, SAM, k-means, and ISODATA methods and occupied most of the region. The chickpea crop was accurately identified by unsupervised methods (Figure 6) with the highest accuracy. The pigeon pea, paddy, and orchard crops were identified with MLC and unsupervised methods (Figure 6). The random forest and SAM method has resulted very well for cotton and pigeon pea crops. However, there was misclassification between pigeon pea and other classes using the supervised methods. The random forest method classified more accurate classes than the SAM and MLC methods (Figure 6). Alternatively, the ISODATA method has also been performed accurately for all classes. There were only a few pixels that were unclassified by the MLC method. It was observed that from Figure 5 and Table 11, the results were satisfactory. The random forest and ISODATA methods produced the highest results (87.71% and 85.01%), followed by MLC (82.70%), k-means clustering (81.39%), and SAM (80.25%) method. The reason behind low accuracy with the SAM method is spectral similarity and the most significant used angle. Hence, the angle of the SAM method has produced spectral confusion between classes. However, the random forest method has also produced confusion between water and other classes. Besides, the spectral properties and spatial information of crop classes were similar cause of misclassification in some cases (Figure 6).

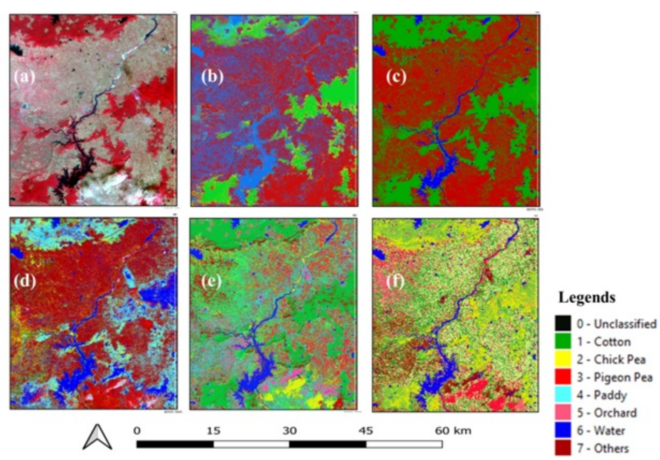


Figure 6. Shows (a) Pre-processed Sentinel-2 images and the spatial distribution maps derived from (b) Random forest, (c) SAM method, (d) MLC method, (e) k-means clustering, and (f) ISODATA method, respectively.

Table V: Confusion Matrix for Random Forest method results (in %)

	Classification Data								
	Class	**Cotton**	**Chick pea**	**Pigeon pea**	**Paddy**	**Orchard**	**Water**	**Others**	**Total**
Ground Truth	**Cotton**	320	6	7	3	8	0	20	364
	Chick pea	9	330	12	6	9	0	30	396
	Pigeon pea	4	9	290	8	10	0	29	350
	Paddy	10	3	8	270	8	0	26	325
	Orchard	8	4	3	0	190	0	22	227
	Water	0	0	0	0	0	300	0	300
	Others	8	4	5	9	3	0	376	405
	Total	359	356	325	296	228	300	503	2367
	OA=87.71%, K=0.86								

Table VI: Confusion Matrix for SAM method results (in %)

	Classification Data								
	Class	**Cotton**	**Chick pea**	**Pigeon pea**	**Paddy**	**Orchard**	**Water**	**Others**	**Total**
Ground Truth	**Cotton**	300	8	10	6	3	0	40	367
	Chick pea	20	270	11	9	3	0	50	363
	Pigeon pea	10	20	260	5	3	0	30	328
	Paddy	11	6	8	200	34	0	20	279
	Orchard	16	5	4	0	190	0	32	247
	Water	0	0	0	0	0	240	0	240
	Others	30	12	5	20	7	0	320	394
	Total	387	321	298	240	240	240	492	2218
	OA=80.25%, K=0.76								

TABLE II

Table VII: Confusion Matrix for MLC method results (in%)

	Classification Data								
	Class	**Cotton**	**Chick pea**	**Pigeon pea**	**Paddy**	**Orchard**	**Water**	**Others**	**Total**
Ground Truth	**Cotton**	290	7	10	3	8	0	34	352
	Chick pea	10	310	15	10	3	0	30	378
	Pigeon pea	14	9	280	8	10	0	43	364
	Paddy	19	3	8	210	8	0	40	288
	Orchard	12	8	4	0	260	0	32	316
	Water	0	0	0	0	0	250	0	250
	Others	19	20	5	9	7	0	350	410
	Total	364	357	322	240	296	250	529	2358
	OA=82.70%, K=0.81								

Table VIII: Confusion Matrix for k-means clustering method results (in %)

	Classification Data								
	Class	**Cotton**	**Chick pea**	**Pigeon pea**	**Paddy**	**Orchard**	**Water**	**Others**	**Total**
Ground Truth	**Cotton**	265	7	11	8	3	0	9	303
	Chick pea	23	290	19	9	3	0	67	411
	Pigeon pea	10	21	300	7	9	0	23	370
	Paddy	9	13	8	190	34	0	20	274
	Orchard	16	5	4	0	188	0	32	245
	Water	0	0	0	0	0	230	0	230
	Others	13	8	9	3	7	0	330	370
	Total	336	344	351	217	244	230	481	2203
	OA=81.39%, K=0.78								

Table IX: Confusion Matrix for ISODATA method results (in %)

	Classification Data								
	Class	**Cotton**	**Chick pea**	**Pigeon pea**	**Paddy**	**Orchard**	**Water**	**Others**	**Total**
Ground Truth	**Cotton**	330	4	22	3	5	0	9	373
	Chick pea	8	320	13	4	22	0	31	398
	Pigeon pea	6	2	310	18	22	0	17	375
	Paddy	3	5	1	90	0	0	34	133
	Orchard	6	13	4	0	188	0	27	238
	Water	0	0	0	0	0	270	0	270
	Others	4	8	18	1	7	0	290	328
	Total	357	352	368	116	244	270	408	2115
	OA=85.01%, K=0.82								

Table X. The class producer's accuracy and specific user's of different classification methods' (in %)

Class	Random Forest		SAM Classifier		MLC		K-Mean		ISODATA	
	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)
Cotton	89.14	87.91	77.52	81.74	79.67	82.39	78.87	87.46	92.44	88.47
Chickpea	92.70	83.33	84.11	74.38	86.83	82.01	84.30	70.56	90.91	80.40
Paddy	89.23	82.86	87.25	79.27	86.96	76.92	85.47	81.08	84.24	82.67
Pigeon Pea	91.22	83.08	83.33	71.68	87.50	72.92	87.56	69.34	77.59	67.67
Orchard	83.33	83.70	79.17	76.92	87.84	82.28	77.05	76.73	77.05	78.99
Water	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Others	74.75	92.84	65.04	81.22	66.16	85.37	68.61	89.19	71.08	88.41

Table XI. Overall Accuracy of Different Classification Methods and Kappa Statistics

Classification Methods	Overall Accuracy	Kappa Coeeficients
Random Forest	87.71	0.86
ISODATA	85.01	0.82
MLC	82.70	0.81
K-Means	81.39	0.78
SAM	80.25	0.76

Table XII: The results of the current study's compared with standard literature

Satellite Datasets	Crop pattern/target classes	Classification methods	Accuracy (in %)	References
Sentinel-2	Multiple crops	Random Forest	91.2	[2]
Sentinel-2	Forest , grassland, water bodies, alpine area	Random Forest	75.62	[5]
Sentinel-2	Built up ,fodder, orchard, sugarcane, wheat, forest, crops, water	Random Forest	84.22	[7]
Multi temporal,Gaofen 1,satellite (GF-1)	Multiple crops	Random forest, K-means and ISODATA	RF-79,K-means-74, ISODATA-75	[28]
Sentinel-2	Multiple crops	Random Forest	92	[1]
Sentinel-2	Multiple crops	Random Forest	90	[11]
Sentinel-2	Multiple crops	Random Forest	PB-RF 86.67, OB-RF 88.57	[12]
Sentinel-2	Bare land, forest, agriculture, residential, water, Impervious surface	Random Forest	94.59	[34]
Sentinel-2	Multiple crops	Random Forest	93	[34]
Sentinel-2	Cropland mapping	K-means clustering	81	[30]
SPOT 5	Urban, grassland, water, vegetation, barren land	MLC and ISODATA technique	MLC- 90.28, ISODATA-80.56	[15]
Sentinel-2	Multiple crops	MLC	MLC-76	[26]
Sentinel-2	Multiple crops	RF, SAM, MLC, K-means clustering, and ISODATA,	Random Forest-87.71, SAM-80.25, MLC-82.70, K-means-81.39, ISODATA-85.01	**This study**

The confusion matrix was created using the classified images, and the kappa coefficient, producers' accuracy, overall accuracy, and users' accuracy were all acquired. The confusion matrix produced by the RF, SAM, MLC, k-means clustering, and ISODATA techniques is displayed in Tables 5, 6, 7, 8, and 9. The kappa coefficient values, producers', users', and overall accuracy were calculated using the confusion matrix. The reference values are displayed in the confusion matrix's columns, and the classified values are highlighted in the rows. The correctly classified values are present in the diagonal position of the confusion matrix [27]. Table 5 shows that the random forest approach has the highest overall accuracy (87.71 percent) with a kappa value of 0.86 in terms of classification accuracy. With kappa values of 0.76 and 0.81 for the SAM and MLC procedures, respectively (Tables 6 and 7), the total accuracy was 80.25 and 82.70 percent. However, with 81.39 and 85.01 percent kappa values of 0.78 and 0.82, the unsupervised approaches have likewise yielded acceptable accuracy levels. (Tables 8 and 9).

Table 10 shows a comparative examination of all the algorithms used for classification as well as producer and user accuracy for each class in the analyzed region. With over 2000 validation sample pixels, each method was run individually. For every classification method, the accuracy of the water class as reported by the producer and user was 100%. Nonetheless, identifying and categorizing the crop characteristics that were adequately identified was our goal. The accuracy of the crop-specific user (table 10) for the cotton crop was 89.14 percent for the random forest, 77.52 percent for the SAM, 79.67 percent for the MLC, 78.87 percent for the k-means, and 92.44 percent for the ISODATA techniques. With results of 92.70, 89.23, 91.22, and 83.33 percent, respectively, the random forest yielded high user accuracy for chickpea, paddy, pigeon pea, and orchard. The user has received accuracy ranging from 77 to 92.44% using the SAM, MLC, k-means, and ISODATA approaches. The producer's accuracy was found to vary from 71.68 to 88.47% across all approaches. When applying the random forest approach, the maximum user's accuracy for chickpeas was 92.44%, while the producer's accuracy utilizing the ISODATA method for cotton was 88.47% (table 10). With the SAM approach, the user's accuracy for the following crops: cotton, chickpea, paddy, pigeon pea, and orchard crops were 77.52, 84.11, 87.25, 83.33, and 79.17 percent. Similarly, the MLC method has also given satisfactory user accuracy for all crops, with 87.74 percent for orchard crops.

Although resampling techniques particularly for k-fold cross-validation (CV) have been used extensively in machine learning algorithms like SVM and RF, there hasn't been much research done on how they affect accuracy relations. Based on these resampling techniques, we recognized images for fulfilling our objectives to: (i) compare the reflectance values of the satellite's bands; and (ii) identify which resampling technique (CV) and classifier (RF or SVM) get higher results in terms of accuracy metrics at the class and overall levels. In the classification training and testing process, the balance of the dataset was assessed using tenfold cross-validation.

On the other hand, ISODATA has also performed well for cotton, chickpea, and paddy crops with 92.44, 90.91, and 84.24 percent user accuracy. The performance of the k-means clustering method was a little bit low for the cotton and chickpea crops in case of user's accuracy. In addition, the producer's accuracy was also good for a cotton crop with the k-means clustering method. The pigeon pea and orchard crops producer's accuracy was very low with the k-means and ISODATA method as compared to other methods.

Table 11 and Figure 7 present a comparison and performance analysis of all the classification methods utilized in this study, along with the corresponding kappa coefficient values and overall accuracy. It is noted that all of the methods yield satisfactory categorization results. Random forest and ISODATA improved the kappa statics and overall accuracy in all approaches (Table 11 and Figure 7). The ISODATA, MLC and RF approaches likewise yielded good results when the kappa values were used. Nonetheless, the outcomes obtained from the SAM and k-means clustering techniques are likewise satisfactory.

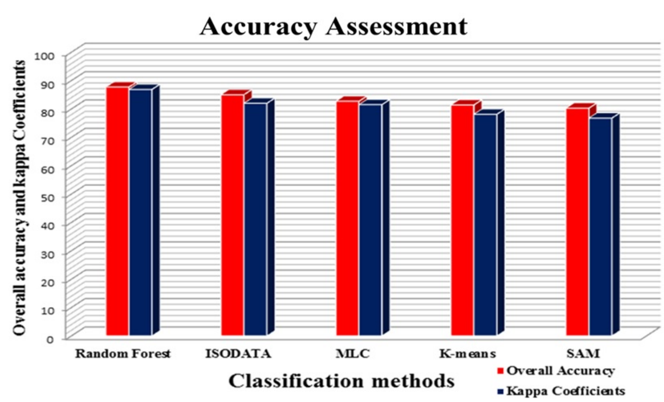


Figure 7. Accuracy comparisons for random forest, ISODATA, MLC, k-means clustering, and SAM methods

B. Discussions and findings

In the present study, Sentinel-2 satellite images were collected at 10m spatial resolutions throughout the growing season and successfully identified and classified multiple crops. The supervised (random forest, SAM, and MLC) [35] and unsupervised (NDVI, k-means, and ISODATA) approaches were successfully implemented on preprocessed data. The results were tested using a confusion matrix to determine the methods' performance.

Since the present study has offered essential information on crop type identification and classification, this study's findings can help solve the problem of crop classification. The study revealed that remotely sensed Sentinel-2 data are the most efficient for crop classification. It also revealed that the most efficient and dependable methods for differentiating between various crop kinds in order to achieve high precision are machine learning algorithms. This lays the groundwork for selecting the optimal blend of machine learning algorithms and remotely sensed data for crop type mapping with a desirable result. Prior research has concentrated on using Sentinel-2 photos to classify crop types because of their significant temporal resolutions in all seasons.

Additionally, earlier studies have shown the importance of either supervised methods or unsupervised methods, specifically SVM [18], ANN, k-means clustering, ISODATA [34] , decision tree [14], RF [36], the fuzzy classifier [18] for crop type classification. However, these studies do not combine a focus on both the supervised/unsupervised approaches. In addition, these methods have not successfully identified crop types in lesser areas or mixed cropping areas. The present study focused on multiple crop type identification, classification, and mapping in miniature and mixed cropping areas. The obtained results were validated with ground truth points and field visits to the studied regions. Furthermore, we compared our results with other standard literature (Table 12) and found the superior performance of our implemented methods. The comparisons with other studies are highlighted in Table 12 with the importance of the present study.

The results achieved in the present study are satisfactory with adequate accuracy for the complex region (Table 12). However, there are significant differences in accuracy and geographical inconsistencies between the data sources due to variances in remotely sensed images, classification schemes, and classification algorithms, resulting in significant differences in the classification findings for multiple crops.

6. Conclusions and Future Work

In the current findings, the optimization parameters of the classifiers are adjusted for desired values to generate a better accurate outcome. The bands are examined, and feature importance for the classifiers is determined. According to this study, the Sentinel-2 data has more potential and generates more high accurate outcomes for crop classification. The NIR is the most suitable band for the random forest method according to feature computation. The results demonstrate that the RF classifier surpasses the other classification methods. The crop varieties and agricultural fields identified can be used to predict yields shortly. Predicting future outcomes is difficult, but it is required to implement real-world applications. In future, the present studies' accuracy can also be enhanced with a hybrid classification approach or deep learning-based models with hyperspectral or RADAR datasets. Furthermore, the integration of Sentinel-2 data and hyperspectral images can enhance accuracy by merging high spatial and spectral characteristics for sub-pixel classification.

Authors' contributions: Atiya Khan, Chandrashekhar H. Patil, Amol D. Vibhute and Shankar Mali have equally contributed.

References

[1] M. Marszalek, M. Lösch, M. Körner, and U. Schmidhalter, "Multitemporal crop type and field boundary classification with google earth engine," Apr. 2020.

[2] A. Khaliq, L. Peroni, and M. Chiaberge, "Land cover and crop classification using multitemporal sentinel-2 images based on crops phenological cycle," in *2018 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS)*. IEEE, Jun. 2018.

[3] S. Khanal, K. Kc, J. P. Fulton, S. Shearer, and E. Ozkan, "Remote sensing in agriculture—accomplishments, limitations, and opportunities," *Remote Sens. (Basel)*, vol. 12, no. 22, p. 3783, Nov. 2020.

[4] M. K. Gumma, K. Tummala, S. Dixit, F. Collivignarelli, F. Holecz, R. N. Kolli, and A. M. Whitbread, "Crop type identification and spatial mapping using sentinel-2 satellite data with focus on field-level information," *Geocarto Int.*, vol. 37, no. 7, pp. 1833–1849, Apr. 2022.

[5] A. Droin, C. Sobe, L. Stanzel, and W. Sulzer, "Vegetation mapping of mountainous areas using sentinel-2 satellite imagery, schneealpe, steiermark," *Grazer Schriften der Geographie und Raumforschung*, no. 48, pp. 103–110, 2018.

[6] A. Khan, C. H. Patil, A. D. Vibhute, and S. Mali, "A u-net based approach for high-accuracy land use land cover classification in hyperspectral remote sensing," in *Soft Computing and Its Engineering Applications*, ser. Communications in computer and information science. Cham: Springer Nature Switzerland, 2024, pp. 94–106.

[7] R. Saini and S. K. Ghosh, "Crop classification on single date sentinel-2 imagery using random forest and suppor vector machine," *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-5, pp. 683–688, Nov. 2018.

[8] S. Saxena, A. Rabha, P. Tahlani, and S. S. Ray, "Crop situation in india, before, during and after COVID-19 lockdown, as seen from the satellite data of resourcesat-2 AWiFS," *J. Ind. Soc. Remote Sens.*, vol. 49, no. 2, pp. 365–376, Feb. 2021.

[9] S. V. Gaikwad, A. D. Vibhute, and K. V. Kale, "Development of NDVI prediction model using artificial neural networks," in *Communications in Computer and Information Science*, ser. Communications in computer and information science. Cham: Springer International Publishing, 2022, pp. 376–384.

[10] S. V. Gaikwad, A. D. Vibhute, K. V. Kale, and A. V. Mane, "Vegetation cover classification using sentinal-2 time-series images and K-Means clustering," in *2021 IEEE Bombay Section Signature Conference (IBSSC)*. IEEE, Nov. 2021.

[11] Z. Yi, L. Jia, and Q. Chen, "Crop classification using multi-temporal sentinel-2 data in the shiyang river basin of china," *Remote Sens. (Basel)*, vol. 12, no. 24, p. 4052, Dec. 2020.

[12] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote Sensing of Environment*, vol. 204, pp. 509–523, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425717304686

[13] N. Matton, G. Canto, F. Waldner, S. Valero, D. Morin, J. Inglada, M. Arias, S. Bontemps, B. Koetz, and P. Defourny, "An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series," *Remote Sens. (Basel)*, vol. 7, no. 10, pp. 13 208–13 232, Oct. 2015.

[14] N. Kobayashi, H. Tani, X. Wang, and R. Sonobe, "Crop classification using spectral indices derived from Sentinel-2A imagery," *J. Inf. Telecommun.*, vol. 4, no. 1, pp. 67–90, Jan. 2020.

[15] M. Hasmadi, H. Z. Pakhriazad, and M. F. Shahrin, "Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data," *Geografia: Malaysian Journal of Society and Space*, vol. 5, no. 1, pp. 1–10, 2009.

[16] C. LUO, H. jun LIU, L. ping LU, Z. rong LIU, F. chang KONG, and X. le ZHANG, "Monthly composites from sentinel-1 and sentinel-2 images for regional major crop mapping with google earth engine," *Journal of Integrative Agriculture*, vol. 20, no. 7, pp. 1944–1957, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2095311920633299

[17] D. T. Nirbhay Bhuyar, Samadrita Acharya, "Crop classification with multi-temporal satellite image data," *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH TECHNOLOGY (IJERT)*, vol. 09, 2020.

[18] R. K. Dhumal, A. D. Vibhute, A. D. Nagne, M. M. Solankar, S. V. Gaikwad, K. V. Kale, and S. C. Mehrotra, "A spatial and spectral feature based approach for classification of crops using techniques based on glcm and svm," *Lecture Notes in Electrical Engineering*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:125985424

[19] F. J. López-Andreu, M. Erena, J. A. Dominguez-Gómez, and J. A. López-Morales, "Sentinel-2 images and machine learning as tool for monitoring of the common agricultural policy: Calasparra rice as a case study," *Agronomy (Basel)*, vol. 11, no. 4, p. 621, Mar. 2021.

[20] R. K. Dhumal, A. D. Vibhute, A. D. Nagne, Y. D. Rajendra, K. V. Kale, and S. C. Mehrotra, "Fuzzy convolution tactic for classification of spatial pattern and crop area," in *2017 11th International Conference on Intelligent Systems and Control (ISCO)*. IEEE, Jan. 2017.

[21] A. Khan, A. D. Vibhute, C. H. Patil, and S. Mali, "Spectral unmixing for end member extraction and abundance estimation," in *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*. IEEE, Jan. 2023.

[22] S. Rapinel and L. Hubert-Moy, "One-class classification of natural vegetation using remote sensing: A review," *Remote Sens. (Basel)*, vol. 13, no. 10, p. 1892, May 2021.

[23] S. V. Gaikwad, A. D. Vibhute, and K. V. Kale, "Estimation of area sown and sowing dates of in-season rabi crops using sentinel-2 time series data," *The Journal of Research ANGRAU*, vol. 49, no. 1, pp. 69–81.

[24] A. D. Vibhute, K. V. Kale, S. V. Gaikwad, R. K. Dhumal, A. D. Nagne, A. B. Varpe, D. B. Nalawade, and S. C. Mehrotra, "Classification of complex environments using pixel level fusion of satellite data," *Multimed. Tools Appl.*, vol. 79, no. 47-48, pp. 34 737–34 769, Dec. 2020.

[25] A. Khan, A. D. Vibhute, S. Mali, and C. Patil, "A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications," *Ecological Informatics*, vol. 69, p. 101678, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954122001285

[26] S. Fal, M. Maanan, L. Baidder, and H. Rhinane, "The contribution of sentinel-2 satellite images for geological mapping in the south of tafilalet basin (eastern anti-atlas, morocco)," *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-4/W12, pp. 75–82, Feb. 2019.

[27] A. D. Vibhute, R. K. Dhumal, A. D. Nagne, Y. D. Rajendra, K. V. Kale, and S. C. Mehrotra, "Analysis, classification, and estimation of pattern for land of aurangabad region using high-resolution satellite image," in *Advances in Intelligent Systems and Computing*, ser. Advances in intelligent systems and computing. New Delhi: Springer India, 2016, pp. 413–427.

[28] Z. Ma, Z. Liu, Y. Zhao, L. Zhang, D. Liu, T. Ren, X. Zhang, and S. Li, "An unsupervised crop classification method based on principal components isometric binning," *ISPRS Int. J. Geoinf.*, vol. 9, no. 11, p. 648, Oct. 2020.

[29] J. Gao, *Digital analysis of remotely sensed imagery*. McGraw-Hill Education, May 2009.

[30] R. Sonobe, Y. Yamaya, H. Tani, X. Wang, N. Kobayashi, and K.-I. Mochizuki, "Crop classification from sentinel-2-derived vegetation indices using ensemble learning," *J. Appl. Remote Sens.*, vol. 12, no. 02, p. 1, May 2018.

[31] L. Congedo, "From GIS to remote sensing," accessed: 2024-3-12. [Online]. Available: https://fromgistors.blogspot.com/2019/09/Accuracy-Assessment-of-Land-Cover-Classification.html

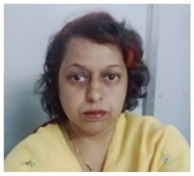
[32] "Good practices for estimating area and assessing accuracy of land change," *Remote Sensing of Environment*, vol. 148, pp. 42–57, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425714000704

[33] A. Mullapudi, A. D. Vibhute, S. Mali, and C. H. Patil, "Spatial and seasonal change detection in vegetation cover using time-series landsat satellite images and machine learning methods," *SN Comput. Sci.*, vol. 4, no. 3, Mar. 2023.

[34] T. Noi and P. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors*, vol. 18, no. 1, 2017.

[35] A. Mullapudi, A. D. Vibhute, S. Mali, and C. H. Patil, "A review of agricultural drought assessment with remote sensing data: methods, issues, challenges and opportunities," *Appl. Geomat.*, vol. 15, no. 1, pp. 1–13, Mar. 2023.

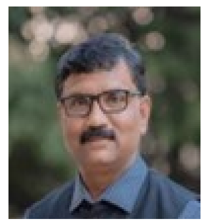
[36] K. Tatsumi, Y. Yamashiki, M. A. Canales Torres, and C. L. R. Taipe, "Crop classification of upland fields using random forest of time-series landsat 7 ETM+ data," *Comput. Electron. Agric.*, vol. 115, pp. 171–179, Jul. 2015.



Atiya Khan is a Ph.D. Scholar at the School of Computer Science, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India.She received her MCA (Computer Science) postgraduate degree from RTMNU, Nagpur Maharashtra,India. Her area of interest is remote sensing, image processing, machine learning and pattern recognition. Her main topic of research is crop classification using deep learning approach. She has published 15+ research papers in international journals like Elsevier, Springer, and IEEE explore conference proceedings.

Amol D. Vibhute has received his PhD in Computer Science under the domain of Geospatial Technology, M. Phil and M. Sc. in Computer Science from the Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MH-India. He is an Assistant Professor at the Symbiosis Institute of Computer Studies and Research (SICSR). He authored/co-authored over 65+ referred journals, book chapters, and conference papers in reputed international journals and conferences indexed by Scopus/SCIE/UGC.

Chandrashekhar H. Patil is an Associate Professor at the Department of Computer Science and applications, Faculty of Science, MITWPU, Pune, India. His fields of research interest include Digital Document Processing, Optical Character Recognition, Biometrics, Word spotting, Video Document Analysis, Signal processing etc. He has published 65 research papers in international journals like Elsevier, springer, and calendar conference proceedings. He is a recognized PhD guide in computer science at MITWPU, Pune. Currently 4 PhD. Scholars are working under him.



Shankar Mali is a Professor at the School of Computer science, MIT World Peace University, Pune, India. His research area includes Digital Image processing, Pattern Recognition, Natural language processing and Networking. He has published 30 research papers in National/International journals. He received a Ph. D. degree from Solapur University in 2011 and qualified SET in Computer Science in 2004. He guides 6 Ph. D. scholars at Savitribai Phule Pune University and MIT World Peace University.