

Class Based Sensor Independent Indices and Training Parameter Approach in Fuzzy Machine Learning Model for Psyllium Husk, Medicinal Crop Mapping

Anam Sabir and Anil Kumar*

Indian Institute of Remote Sensing, Dehradun, India

anamsabir1331@gmail.com, anil@iirs.gov.in

*Corresponding Author: A. Kumar <anil@iirs.gov.in>

Received: September 25, 2022; Accepted: November 18, 2022; Published: December 23, 2022

Abstract

The use of spectral indices is prevalent in remote sensing data processing for a variety of applications. However, the spectral bands considered to formulate conventional spectral indices may not be the most appropriate ones for a particular application. With the improvement in technology, the available spectral resolution has increased up to a large extent which can be explored to enhance a class much better way for specific class mapping. This study tests the performance of Modified Soil Adjusted Vegetation Index (MSAVI2) for mapping Psyllium Husk (*Plantago Ovata*), a medicinal crop with a relatively small size which increases the impact of soil brightness. Three variants of MSAVI2 with different bands combination were tested i.e. Conventional (Red-NIR), RedEdge1 (705 nm)-NIR, and CBSI-MSAVI2 (Class Based Sensor Independent). The classification technique used was MPCM (Modified Possibilistic *c*-means), for which two training approaches were made use of i.e. Fuzzy mean-based MPCM and Fuzzy Individual-sample-as-mean (ISM) based MPCM. The classifications were carried out considering a range of sizes of training samples starting from 5 to 50. The accuracy of different combinations of index, bands, and number of samples were assessed using Mean Membership Difference (MMD), Variance, Fuzzy Error Matrix, and Sub-pixel Confusion Uncertainty Matrix by making use of soft classified CubeSat (Dove) 3m data. The overall accuracy for different test cases achieved were between 83%-99% while the Kappa coefficient varies from 0.6 to 0.99.

Keywords: Modified Soil Adjusted Vegetation Index; Modified Possibilistic *c*-Means; Individual-sample-as-mean; Class Based Sensor Independent; Fuzzy Error Matrix

1. Introduction

Remote sensing has become an important tool for the analysis and prediction of various parameters using space-borne and air-borne images [1]. The wide variety of sensors on-board satellites with different operating wavelengths and spatial resolution have made it even more feasible to fill the gaps between datasets in order to carry out analysis [2]. The different types of sensors include visible (visual), multispectral, hyperspectral, thermal, and microwave, and platforms include satellite, manned aircraft (airborne or aerial), unmanned aerial vehicles (UAVs), unmanned aerial systems (UASs), or drones [3]. Particularly in agricultural applications, satellite-based multispectral remote sensing data for specific crop mapping can be used by formulating various spectral indices based on the wavelength of different bands [4]. The most commonly used spectral band indices for vegetation mapping are Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), SAVI (Soil Adjusted Vegetation Index), NDWI (Normalized Difference Water Index), etc. [5]–[7]. The availability of higher temporal and spatial resolution satellite remote sensing data has enabled the mapping of crop phenology more accurately [8]. Specific crop mapping has been explored by studies on yield estimation and soil

productivity [9]. However, specific crop mapping using single-date imagery can be a challenging task and temporal data is required to extract the phenological characteristics of the target crop sun, while handling spectral overlap[8]. A sub-activity carried out in a specific crop mapping domain is the conservation of medicinal crops. In order to conserve a particular species of medicinal crop, it is necessary to map its locations and the extent of its spatial distribution [10]. This paper focuses on Psyllium Husk (*Plantago ovata*) which is a herbal crop grown in various parts of India. It is an important medicinal crop that is used for the treatment of astringent, tonic, biliousness, cough, dysentery, and leprosy [11].

The use of optical data has provided an efficient pathway for crop-related studies with the help of a range of spectral indices (e.g. NDVI, NDWI, SAVI, LAI) [12]–[14]. The choice of the spectral index and bands used for formulating them depends on the specific application [14]–[17]. Therefore, a spectral index formulated by a particular spectral band that performed well for a particular task may not be effective for other tasks. The medicinal crop under study (Psyllium Husk) is small (height up to 15cm) and does not cover the crop field densely [18]. Thus the percentage of soil exposed in the field is considerably

high and this increases the impact of reflectance from the soil on the satellite imagery which may lead to erroneous results. For such cases, SAVI or MSAVI2 is preferred over NDVI as the soil background is a major surface component controlling the spectral behaviour. MSAVI2 incorporates the estimated fraction of soil in an area to suppress the soil signature [19]. The recent advent of high-resolution multispectral and hyperspectral satellites made it possible to use finer bandwidths with a greater spatial resolution for different applications using advanced methods [20]. The Sentinel-2 data with multiple red edge spectral bands gives an approximation of chlorophyll status in vegetation [21]. It highlights the stress in crops and as the chlorophyll content increases, the red edge shifts towards a higher wavelength and vice versa. Thus, it can be used in crop mapping to highlight the sowing and harvesting stages of the crop. The use of appropriate methods and a systematic approach with the red edge spectral bands can prove more effective for the mapping of specific crops.

Conventional algorithms classify each pixel of the image entirely to a single class widely called hard classification[22]–[24]. Although, in the case of most satellite images, the portion covered by one pixel has considerable chances of not belonging

to only one particular class which rules out the chances of hard classification techniques being efficient. Hence, the pixels are mixed in nature and the area under one pixel may be composed of more than one class that requires soft classification techniques for appropriate handling[25]–[28]. To achieve the specific objective of this study (i.e. accurate mapping of a medicinal crop), a soft classification approach based on fuzzy logic is better suitable for the application as the crop is sparsely distributed on the ground [29], [30]. Classifiers based on fuzzy logic assign membership values to the pixels according to the information contained in them with respect to each class. Hence, for each pixel, a membership value is generated for each of the classes present in the image [31]. Therefore, the output of the fuzzy-based soft classification algorithms is in the form of fractional images generated for each class of interest. The conventional fuzzy *c*-means (FCM) algorithm was introduced by [32], but it does not efficiently explain the degree of belonging for data [33]. Thus, a Possibilistic *c*-Means (PCM) approach was proposed to overcome its weakness [34], [35] and deliver more accurate memberships for a particular class as required for the study. This was further improved to propose Modified

Possibilistic *c*-Means (MPCM) algorithm to reduce the overlap between classes.

The assessment of soft classification output depends on the application and number of classes. The most preferred method for accuracy assessment is a fuzzy confusion matrix that indicates recall, precision, accuracy and F-measure [8], [9], [36]. The output images are fractional i.e. they hold membership values of each pixel in different classes. Fuzzy error matrix (FERM), MIN-PROD error matrix, and Entropy are other few methods for evaluating soft classified output. FERM and MIN-PROD error matrix compute accuracy parameters like overall accuracy (OA), coefficient of kappa (K), user's accuracy (UA) and producer's accuracy (PA) using a confusion matrix [37].

The objective cum innovation of this research work was to study effect of various bands in indices for temporal database generation for Psyllium husk crop mapping. The importance of bands in spectral indices in the temporal domain as the crop stages change has not been explored along with training parameters of classifier must be used to obtain results with better homogeneity in outputs. This study aims at specific crop mapping through Fuzzy MPCM algorithm using

temporal optical data while considering bands that highlight the target crop in the temporal domain. Multiple variants of MSAVI2 index (i.e. conventional (Red-NIR), RedEdge1-NIR and CBSI-MSAVI2) with sample sizes effect were studied for mean-based as well as Individual Sample as Mean (ISM) approach for MPCM algorithm. The ISM method for training the classifier provides a unique way for efficiently considering each sample's temporal phenological signature, so as to obtain a more generalized model which can handle heterogeneity within the input data. The results were evaluated using matrix based image-to-image accuracy assessment methods i.e. FERM and SCM, along with measures like MMD and variance to assess the homogeneity in the output.

2. Mathematical Concepts

2.1 Index: MSAVI2

The choice of index in this study was based on the target crop i.e. Psyllium Husk. Due to considerable coverage of soil on the field, the effect of soil signature on the overall response is higher. Thus MSAVI2 was applied to compensate for the soil's impact.

Within the MSAVI2 index, experiments were carried out to test the performance of different band

combinations. CBSI (Class Based Sensor Independent) provides a method to find out the bands which might perform the best to enhance the class of interest. The band combination for MSAVI2 index was changed keeping same the formula to generate two variants of MSAVI2 used in this study i.e. MSAVI2 Red Edge-1 and CBSI-MSAVI2. As the name suggests, MSAVI2 RedEdge-1 makes use of NIR and Red Edge-1 bands. While in case of CBSI, the bands with minimum and maximum values for the target feature were selected to maximize the value of the index. This helps highlight the target feature more efficiently thus facilitating the further classification.

- MSAVI2 conventionally uses Red and NIR bands as shown in Equation 1.

$$MSAVI2_{Conv} = \frac{(2*NIR+1-\sqrt{(2*NIR+1)^2-8*(NIR-Red)})}{2} \quad (1)$$

- MSAVI2 Red Edge-1 was calculated by replacing the Red band with RedEdge1 in equation (1) as shown in Equation (2).

$$MSAVI2_{RedEdge} = \frac{(2*NIR+1-\sqrt{(2*NIR+1)^2-8*(NIR-RedEdge1)})}{2} \quad (2)$$

- CBSI-MSAVI2 picked up the bands with maximum and minimum value for the target crop and replaced with Red and NIR bands in equation (1) to highlight the target

crop more effectively by maximizing the value calculated for it. The formula for CBSI-MSAVI2 is as given in Equation (3).

$$MSAVI2_{CBSI} = \frac{(2*max+1-\sqrt{(2*max+1)^2-8*(max-min)})}{2} \quad (3)$$

Here max and min are the maximum and minimum valued bands respectively.

2.2 Algorithm: Fuzzy MPCM

As per the requirements of the application, an algorithm with capability to extract single class, handle noise and mixed pixel problem was required. Hence, the Fuzzy MPCM algorithm was used for the mapping of target crop fields.

Fuzzy machine learning algorithms applied for image classification mainly consist of FCM, PCM and MPCM. In the case of standard FCM, the membership values are so assigned that the sum of membership values of a pixel for all classes should come out to be 1[32]. This prevents this algorithm from extracting a single class from the image. This constraint, called a hyper-line constraint, was removed to propose PCM which therefore is capable of single class extraction [34]. But the problem with PCM is that of coincident clusters. MPCM overcomes this drawback and provides an overall fair option for classification on satellite imagery. The

objective function of MPCM is as given in equation 4.

$$J_{MPCM}(U, V) = \sum_{j=1}^c \sum_{i=1}^N \mu_{ij}^m d_{ij}^2 + \eta_j \sum_{i=1}^N (\mu_{ij} \log \mu_{ij} - \mu_{ij})^m \quad (4)$$

Here, U is the matrix containing membership values for each pixel corresponding to each class while V is the matrix containing class centres. Rest of the parameters in equation 4 are explained in equation 5 to equation 7 separately along with respective formulae.

μ_{ij} is the typicality value of pixel i in class j.

$$\mu_{ij} = \frac{\frac{1}{d_{ij}^{\frac{1}{m-1}}}}{\sum_{k=1}^c \frac{1}{d_{ik}^{\frac{1}{m-1}}}}, \text{ for all } i, j \quad (5)$$

d_{ij}^2 is the square of the distance between the measured value of a pixel and that of cluster centre, eq (6).

$$d_{ij}^2 = \|x_i - v_j\|^T A^{-1} \quad (6)$$

In equation 7 and 8, x_i refers to the measured value whereas v_j is the cluster centre as fuzzy mean:

$$\mu_{ij} = \exp(d_{ij}^2/\eta_j), \text{ for all } i, j \quad (7)$$

$$v_j = \frac{\sum_{i=1}^N \mu_{ij} x_i}{\sum_{i=1}^N \mu_{ij}} \quad (8)$$

2.3 Accuracy Assessment Metrics

FERM (Fuzzy Error Matrix) is a modified form of traditional error matrix

for soft classification. It is a square matrix with values varying between 0 and 1. The column R_n usually represents the sample elements assigned to the reference class n whereas the rows indicate the sample elements assigned to the classified class m [38] (Binaghi et al. 1999). The element in the fuzzy error matrix (M) at row m and column n for a feature vector x is computed as mentioned in equation (9).

$$M(m, n) = \sum \min(\mu_{C_m}(x), \mu_{R_n}(x)) \quad (9)$$

Here, x is the overall sampled data set. μ_{C_m} and μ_{R_n} are the membership values for the referenced and the classified data. The "min" operator is a traditional fuzzy set operator that returns the minimum membership value among the classified and referenced data set for a specific class. The various indices for accuracy assessment like overall accuracy, user's accuracy, and producer's accuracy can be calculated from FERM.

SCM (Sub-pixel Confusion Uncertainty Matrix) is also a modification of the traditional error matrix unlike which, here, the entries are based on the agreement and disagreement measure for a class, calculated using the area overlap, between the classified output and the referenced data at pixel level [39]. Different operators used for SCM in this study are MIN-MIN, MIN-LEAST and MIN-MAX.

3. Proposed Approach

3.1 Study Area and Data Description

The area under study is that of a portion of Jalore district, Rajasthan, India. This area has an abundance of medicinal crop fields, out of which Psyllium Husk is the target crop. A field visit was done for collection of ground truth from this location

on 8th January 2021 where various field coordinates were recorded throughout Jalore and parts of Badmer district of Rajasthan. Several other varieties of medicinal crops were also reported in this area as well. The study area is shown in figure 1 along with the field photographs in figure 2.

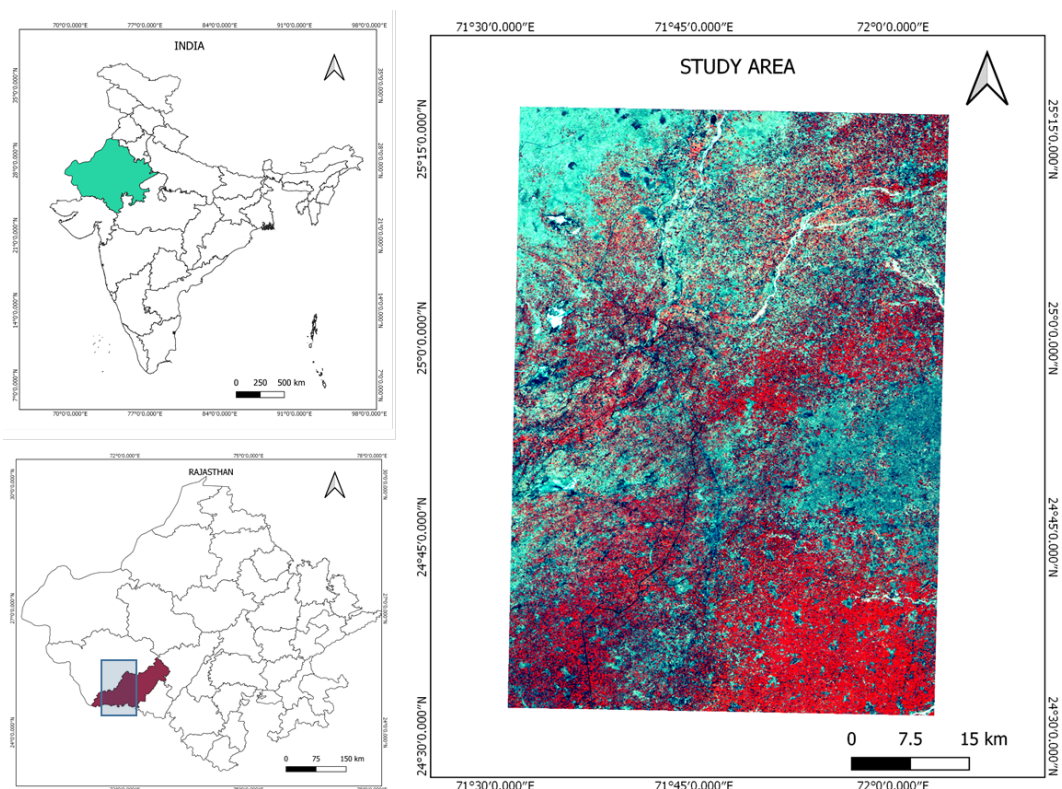


Figure 1: Study Area: Jalore district, Rajasthan, India



Psyllium Husk



Mustard



Cumin



Wheat



Fenugreek



Castor

Figure 2: Field Photographs taken on 8th January 2021 showing target crop Psyllium Husk along with different crops present in the study area

Psyllium Husk (scientific name: *Plantago Ovata*) is a medicinal crop used in various medications. It is a short-stemmed crop with a number of shoots flowering from the base. It grows up to a height of 30-40cm. It is cultivated in Mediterranean regions and West Asia as it requires a dry climate since rain can spoil the crop

completely. It is a Rabi crop with a period of 110-120 days. Its seed is covered in white husk. At maturity, the leaves become yellowish, spikes turn brownish in colour and husk opens exposing the dark brown seeds. The crop stages with respect to different months are as shown in Table 1.

Table 1: Psyllium Husk Crop Stages

Month	Stage	Specific Stage
November	Sowing	Sowing
December	Growth	Seedling
January		Budding
February	Harvesting	Flowering, Pollination
March		Ripening, Maturing, Harvesting

Optical Level-2 data from Sentinel 2A and 2B was utilized for the study. The data consists of 13 bands in VNIR and SWIR bandwidth range. Out of these 13 bands, 60m resolution bands were omitted and only those with resolution 10m and 20m were utilized for the study. Also CubeSat data with 3m spatial resolution was downloaded from PlanetScope and classified for the same study area to be used as soft reference data for accuracy assessment using FERM and SCM. The data was available with no cloud coverage

due to the geographic location of the area and good contrast was observed in the images which facilitated the further processing as well as accuracy assessment. It was made sure that the temporal data under use covered the whole phenology of target crop. Since Psyllium Husk is a Rabi crop, dates were covered starting from December to March. The specific dates for which Sentinel and CubeSat data (for accuracy assessment) was downloaded is as mentioned in table 2.

Table 2: Temporal Data used for the study: Sentinel-2 from Copernicus and Dove from PlanetScope

Sentinel Data	Dove Data
31 st Dec 2020	30 th Dec 2020
10 th Jan 2021	11 th Jan 2021
4 th Feb 2021	4 th Feb 2021
14 th Feb 2021	14 th Feb 2021
1 st Mar 2021	1 st Mar 2021
6 th Mar 2021	15 th Mar 2021
21 st Mar 2021	20 th Mar 2021
26 th Mar 2021	26 th Mar 2021

3.2 Methodology

The data for Sentinel 2 was downloaded from Copernicus' website. Out of the 13 bands available, only 10m and 20m bands were used for the study i.e. ten

bands in total. These bands were resampled to maintain 10m spatial resolution throughout, followed by indices calculation and separability analysis. Here the optimum dates were selected according to the phenology of the target crop to maximize

the spectral distance between target and non-target classes. After the generation of the dataset, training samples were picked up and classification was done using the Fuzzy MPCM classifier. The process was then repeated for CubeSat data, to have soft classified reference data for accuracy

assessment using FERM and SCM. The outputs were also evaluated on the basis of MMD and variance within the field. The flowchart for the methodology proposed for the study has been as shown in the figure (3).

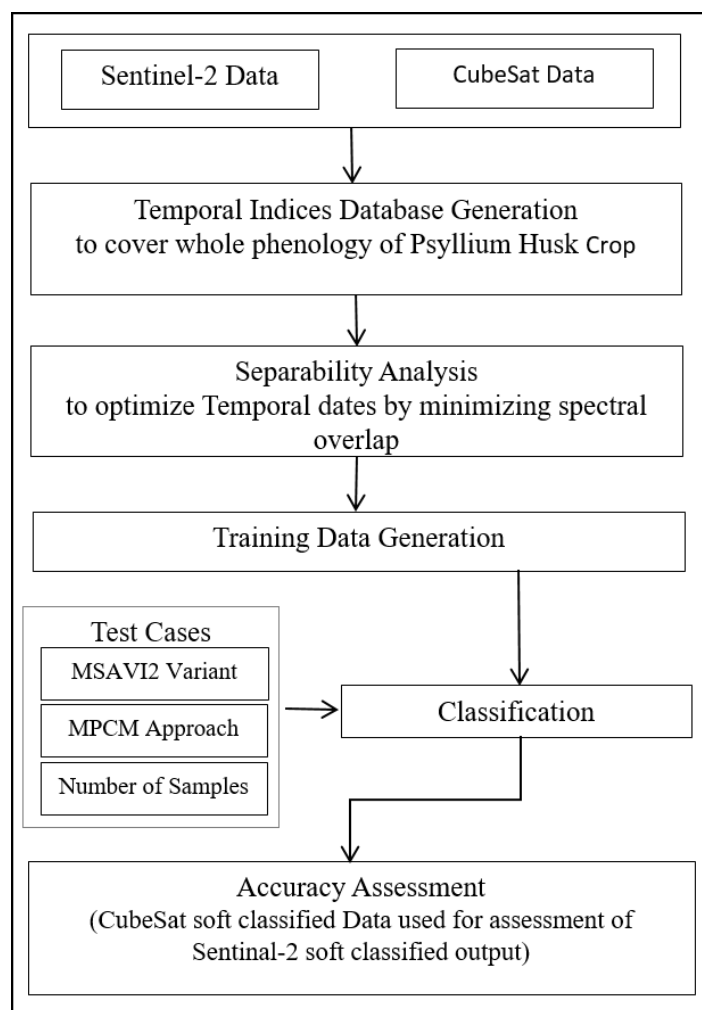


Figure 3: Proposed Methodology for this research work

The bands to be used for the study i.e. 10m and 20m bands of Sentinel 2 data were resampled to 10m resolution and stacked. This was done for all the dates followed by calculation of MSAVI2 for each date's stack utilizing three different

indices combinations. First approach is the conventional way, using Red and NIR bands. Second approach makes use of Red Edge-1(RE-1) and NIR bands. The third approach used is Class Based Sensor Independent (CBSI) which works on the

idea of selecting bands with minimum and maximum values for the target class and hence highlights the target class irrespective of its phenological stage.

The bands used in conventional MSAVI2 i.e. Red and NIR enhance the vegetation while reducing the impact of soil brightness. The Red band helps detect chlorophyll while NIR band is used for monitoring stress in the vegetation. This may/may not be stress depending upon the

expected stage on the basis of the phenology of specific crop. The second variant utilizes the Red-Edge (705 nm) band which is used for detecting early stress as well as changes in chlorophyll content. As the chlorophyll content increases, Red-Edge shifts towards a longer wavelength region. The different bands picked up by CBSI for mapping of Psyllium Husk using MSAVI2 in the study area are as listed in Table 3.

Table 3: Conventional MSAVI2, RedEdge MSAVI2 and CBSI-MSAVI2 values for the target crop along with the bands selected for CBSI-MSAVI2 for each date

Date	Conventional MSAVI2 Value (Red-NIR)	RedEdge MSAVI2 Value (RedEdge1-NIR)	CBSI-MSAVI2		
			Maximum Valued Band	Minimum Valued Band	CBSI-MSAVI2 Value
31 st Dec 2020	0.5529	0.4901	SWIR (1,610 nm)	Blue (490 nm)	0.647059
10 th Jan 2021	0.4156	0.4509	SWIR (1,610 nm)	Blue (490 nm)	0.588235
4 th Feb 2021	0.6980	0.6235	NIR (842 nm)	Blue (490 nm)	0.784314
14 th Feb 2021	0.7568	0.7450	Red Edge (865 nm)	Blue (490 nm)	0.92549
1 st Mar 2021	0.7882	0.7176	NIR (842 nm)	Blue (490 nm)	0.843137
16 th Mar 2021	0.6	0.5843	NIR (842 nm)	Blue (490 nm)	0.678431
21 st Mar 2021	0.5529	0.5176	SWIR (1,610 nm)	Blue (490 nm)	0.709804
26 th Mar 2021	0.5529	0.5254	SWIR (1,610 nm)	Blue (490 nm)	0.705882

The CBSI approach picked up the Blue band in place of the Red band for all the dates under study as shown in table 3. This is because the minimum value it detected for the Psyllium Husk field was

that of Blue band. This indicates that the lower value of vegetation in the Blue band in comparison with that in the Red band can prove useful for further increasing the index value for highlighting the vegetation. On

the other hand, in place of NIR, the CBSI approach picked up NIR (842 nm), Red-Edge (862 nm) and SWIR (1610 nm). The MSAVI temporal dates indices thus generated were stacked together which gave the temporal indices database. The

temporal behaviour of different crops within the eight dates considered for the study area was analysed in terms of the MSAVI2 values obtained. The curves for MSAVI2 values for six crops were plotted against eight dates as shown in figure 4.

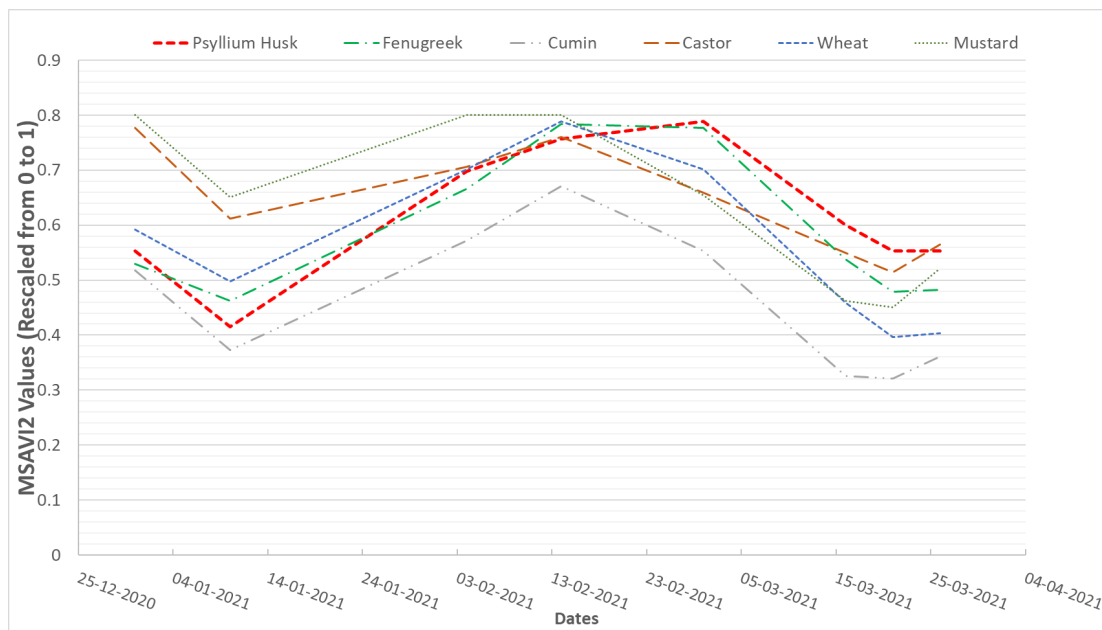


Figure 4: MSAVI2 curves for different crops in the study area

Figure 4 shows the curves followed by MSAVI2 values during different phenological stages. The lower values indicate watering of the crop which leads to a decrease in MSAVI2 values whereas at the peak of vegetation (around mid-February), highest values are observed which correspond to increased chlorophyll content in the crops. Now as the crop starts ripening, it turns yellowish from green and due to this decrease in chlorophyll, the MSAVI2 values drop and are minimum in the last of March which marks the harvest of the crop.

From figure 4, it can be observed that the fenugreek exhibits the closest profile to that of Psyllium Husk, followed by wheat. These three curves have similar shapes and differ just by a small offset. It indicates the need for separability analysis in order to bring out the unique behaviour of the target crop while suppressing non-target crops. For the study of target crop separately, the values for all three variants of MSAVI2 i.e. conventional, Red Edge-1 and CBSI were plotted against the eight dates as shown in figure 5.

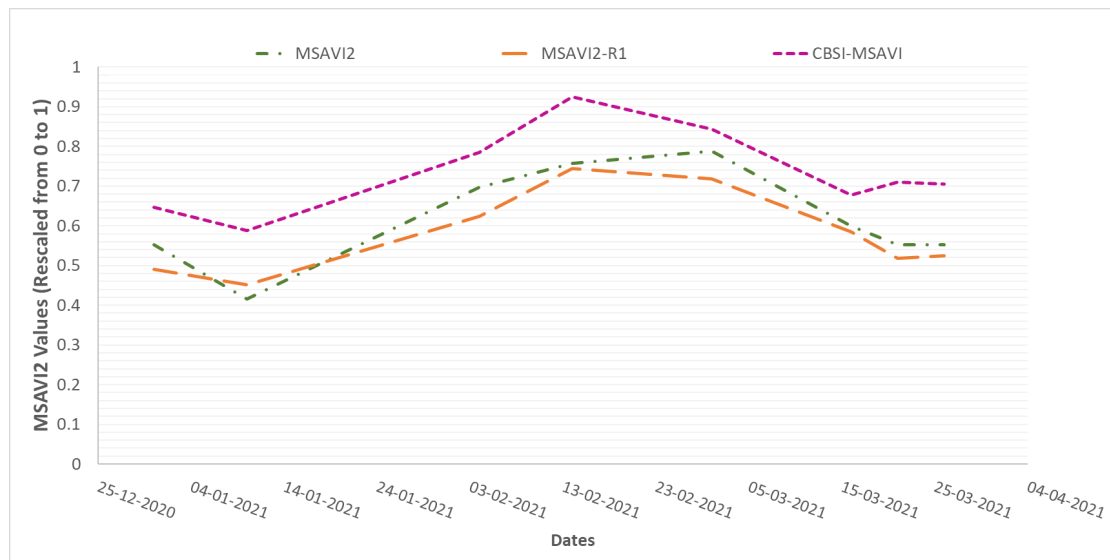


Figure 5: Conventional MSAVI2, RedEdge-1 MSAVI2 and CBSI-MSAVI2 curves for Psyllium Husk

Figure 5 shows the phenological profile of Psyllium Husk using different versions of the MSAVI2 index. It was observed that RE-1 variant gave higher value when the crop has less chlorophyll content i.e. at the time of sowing and harvest, except that it maintained comparatively lower value than conventional MSAVI2. Apart from this, the whole curve for CBSI-MSAVI2 is elevated in comparison with the other two. This is because it works on the principle of selecting bands which highlight the target crop the most, irrespective of the wavelength. It picks up the bands with minimum and maximum value for the target class thus resulting in higher value of the index.

This optimized temporal indices database was further used for picking samples for different crops present in the study area with the help of ground truth data collected from field visit. Separability analysis was carried out using signatures thus recorded where the separability between target and non-target crop were analysed using Euclidean distance to select the temporal dates which maximized the separation between the target crop and the non-target crop, which is closest to Psyllium Husk. These layers correspond to dates representing specific crop stages of target crop, which are essential for mapping the crop according to its phenology, while maximizing the spectral gap. In the case of Psyllium Husk, the closest crop in terms of temporal behaviour/phenology came out to

be Fenugreek. Thus, attempts were made to retain those bands which differentiated best between Psyllium Husk and Fenugreek. Therefore, three sets of bands were

finalized to be used for each of the three indices as shown in table 4. These bands were stacked in ascending chronological order for further processing.

Table 4: Optimum dates for Psyllium Husk Mapping according to Separability Analysis Results

MSAVI2 Variant	Optimum Dates
Conventional (Red, NIR)	31 st December 2020, 10 th January 2021, 1 st March 2021, 16 th March 2021, 21 st March 2021
Red Edge 1, NIR	31 st December 2020, 10 th January 2021, 4 th February 2021, 1 st March 2021, 16 th March 2021
CBSI MSAVI2	31 st December 2020, 10 th January 2021, 14 th February 2021, 16 th March 2021, 26 th March 2021

4. Results and Discussion

4.1 Soft Classification Output

Two different approaches with the MPCM classifier were tested for mapping Psyllium Husk crops. The basic idea was to reduce heterogeneity within class while mapping the target crop which may be a result of slightly different spectral response from two different parts of the same field. This variation may arise due to non-uniform water/fertilizer application. Hence to compensate for this, ISM approach was applied for the classification along with the standard MPCM classification which makes use of the mean of the training data. While in case of ISM approach, instead of utilizing the mean of the training dataset, each training sample is used to its full extent equally in the training process by running the algorithm n number of times,

where n denotes the number of training samples. Thus n membership values of each pixel are obtained corresponding to each of the training sample. Among these, the highest membership value is retained and rest are discarded. This gives rise to a more inclusive method in terms of unique behaviour exhibited by each training sample which results in better mapping ultimately.


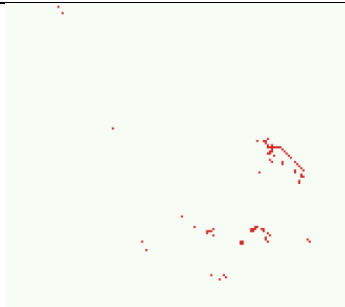
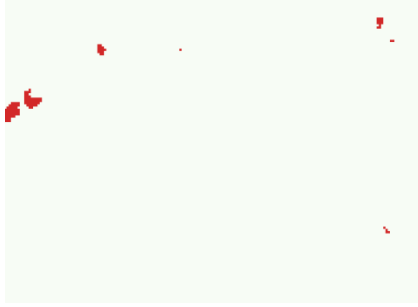
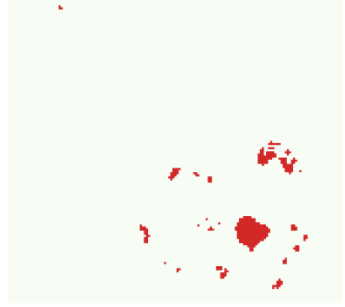
The three databases generated corresponding to the three indices, were used for picking training samples varying from 5 to 50 in number. It is a general practice in classification using remote sensing to take training samples at least ten times the number of wavelength bands present in the imagery [36]. Therefore, different sample sizes were considered for


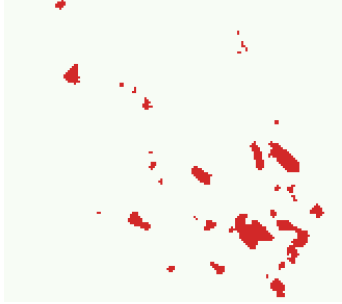


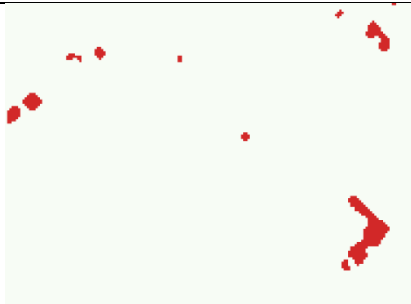

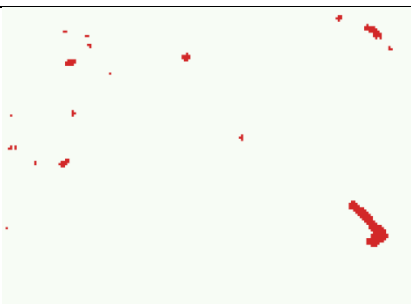

this study i.e. 5, 10, 15, 20, 25 and 50 to check the variation in performance. With the help of SMIC (Sub-pixel Multispectral Image Classifier) Tool [40], the outputs were generated for all the cases.

For each of the outputs, taking testing points into consideration, the corresponding membership values were used to calculate Mean Membership Difference (MMD). MMD is the difference between the mean of the membership values of different points, It can be calculated as inter-crop as well as intra-crop. Within the target crop i.e. in case of intra-crop analysis, the value of MMD should be as less as possible since the

membership values within the target class should not vary significantly. While In case of inter-crop analysis, the value of MMD should be as high (close to 1) as possible since the membership value of target crop should be ideally 1 and that of non-target crop should ideally be 0 which gives a high difference. It was observed from the MMD and variance values that the outputs in most of the cases saturated at 10 or 15 training samples i.e. by further increasing the number of samples, significant change was not observed. Thus the outputs for different variants of MSAVI2 using both the MPCM approaches are shown in the table 5.

Table 5: Outputs corresponding to different test cases for two sites covering target fields

Index Variant	Mapped fields on Site 1	Mapped fields on Site 2
MPCM with Mean Approach		
Conventional MSAVI2		
MSAVI2 Red-Edge1		

CBSI MSAVI2		
ISM- Individual Sample as Mean Approach		
Conventional MSAVI2		
MSAVI2 Red-Edge1		
CBSI MSAVI2		

From the results presented in table 5, it can be observed that within the Mean based approach applied for MPCM, the outputs for CBSI MSAVI2 have efficiently mapped the target fields in the area while conventional MSAVI2 and MSAIV2 Red-Edge failed to do so. Whereas in case of

ISM approach, all the three index variants have mapped all the target fields, though with varying accuracy. It can be observed that the boundaries and shape of all the mapped fields are retained and homogeneity within the mapped fields is maintained. Also, the overall noise in the

output is quite less compared to the Mean based approach.

4.2 Accuracy Assessment

The results were assessed using FERM, SCM, MMD (Mean Membership Difference) and variance. MMD was calculated within target crop as well as between target crop and non-target crop. In addition to this, for the accuracy in terms of potential to handle heterogeneity within class, outputs were assessed using variance inside the testing field. Matrix based accuracy measures (FERM and SCM) were calculated using classified CubeSat data for the cases which gave best MMD and Variance results to determine the User's Accuracy, Producer's Accuracy, Overall Accuracy and Kappa Coefficient. For calculating FERM and SCM, two different datasets are required which serve as classified and reference datasets. Using these two sets of fractional images, image to image accuracy assessment is carried out in terms of agreement within the two since they are at different spatial resolution. The reference dataset was prepared using the

CubeSat (Dove) optical temporal data available at 3m spatial resolution. The classified outputs were generated for conventional MSAVI2 and CBSI-MSAVI2 for Dove data, leaving behind the RedEdge-1 MSAVI2 variant which couldn't be tested due to availability of only four bands in the 3m dataset. Thus for MSAVI2, Band4 (NIR) and Band3 (Red) were used. Whereas for CBSI-MSAVI2, Band4 (NIR) and Band1 (Blue) were used. The classification was carried out for two classes, one being the target class, Psyllium Husk and one additional class Fenugreek since the method requires at least two classes for the generation of matrices. Fenugreek was chosen due to the spectral and phenological similarity so that it can be detected with the same combination of dates as that taken for Psyllium Husk. The Sentinel data was resampled to 9m spatial resolution for maintaining a ratio of 1:3 between classified (Sentinel) and reference (Dove) data. The results for accuracy assessment using FERM and SCM taking 600 samples are as shown in table 6.

Table 6: Accuracy Assessment Measures using Matrix-based Image-to-Image Accuracy Assessment Techniques

Measures	FERM (%)			
	MSAVI2		CBSI-MSAVI2	
	Psyllium Husk	Fenugreek	Psyllium Husk	Fenugreek
User's Accuracy	91.55	92.09	82.00	81.44
Producer's Accuracy	92.49	92.98	82.34	84.56

Overall Accuracy	92.73	83.44		
Min Least				
Measures	MSAVI2		CBSI-MSAVI2	
	Psyllium Husk	Fenugreek	Psyllium Husk	Fenugreek
User's Accuracy	86.53	85.74	82.06	77.33
Producer's Accuracy	86.21	86.07	77.22	82.16
Overall Accuracy	86.14		79.623	
Kappa Coefficient	0.72		0.593	
Min Max				
Measures	MSAVI2		CBSI-MSAVI2	
	Psyllium Husk	Fenugreek	Psyllium Husk	Fenugreek
User's Accuracy	99.84	99.78	99.03	98.38
Producer's Accuracy	99.73	99.83	98.39	99.02
Overall Accuracy	99.78		98.70	
Kappa Coefficient	0.99		0.97	

The accuracy assessment measures mentioned in table 6 show that the classification carried out using MSAVI2 as well as CBSI-MSAVI2 with MPCM following mean approach lead to results exhibiting good accuracy with MSAVI2 concluded as the best method for the crop under study. Both the crops i.e. Psyllium Husk as well as Fenugreek show appreciable results. Small difference in

values of accuracy and kappa coefficient for different variants of MSAVI2 was observed. The MMD values were calculated using samples from ground truth for target crop (Psyllium Husk) and other crops present in the study area. The inter-crop and intra-crop MMD values as well as variance values are shown in table 7 and table 8 respectively.

Table 7: MMD between Psyllium Husk and non-target crops for MSAVI2 variants and MPCM approaches

Number of Samples	MSAVI2		MSAVI2 Red Edge 1		CBSI MSAVI2	
	Mean	ISM	Mean	ISM	Mean	ISM
MMD between Psyllium Husk and Cumin using different approaches						
5	0.601494	0.578525	0.466853	0.449486	0.683287	0.672829
10	0.567881	0.578525	0.56620	0.449486	0.665546	0.682166
15	0.446872	0.464239	0.554435	0.368814	0.65901	0.640336
20	0.585247	0.464239	0.555182	0.392344	0.661999	0.637722
25	0.585434	0.464239	0.563399	0.392344	0.661251	0.637722

50	0.546592	0.435107	0.51746	0.44127	0.650233	0.691317
MMD between Psyllium Husk and Mustard using different approaches						
5	0.729085	0.722222	0.796405	0.801307	0.445751	0.425491
10	0.583987	0.448693	0.713399	0.569935	0.433333	0.43987
15	0.26732	0.448693	0.753595	0.569935	0.441503	0.430392
20	0.674183	0.448693	0.758823	0.593464	0.471242	0.431699
25	0.678432	0.448693	0.750654	0.593464	0.450326	0.257189
50	0.713399	0.283006	0.783006	0.485948	0.459477	0.396079
MMD between Psyllium Husk and Fenugreek using different approaches						
5	0.026144	0.022222	0.046013	0.036602	0.16732	0.156863
10	0.05699	0.022222	0.001045	0.000523	0.174902	0.171242
15	0.046536	0.022222	0.015163	0.000523	0.171503	0.170981
20	0.02092	0.022222	0.017255	0.020915	0.182222	0.165229
25	0.01804	0.018300	0.013595	0.016994	0.173856	0.150327
50	0.02092	0.014379	0.018300	0.025359	0.179869	0.153726
MMD between Psyllium Husk and Wheat using different approaches						
5	0.350849	0.328889	0.171503	0.159738	0.30536	0.288628
10	0.233987	0.263791	0.212026	0.159738	0.218039	0.27634
15	0.254379	0.258301	0.208104	0.149543	0.213856	0.231373
20	0.291242	0.258301	0.208627	0.173072	0.219085	0.221699
25	0.297255	0.258301	0.212026	0.173072	0.213856	0.221699
50	0.288889	0.361830	0.175163	0.135948	0.306928	0.486275
MMD between Psyllium Husk and Castor using different approaches						
5	0.487908	0.480065	0.452288	0.442484	0.238889	0.201961
10	0.320262	0.183987	0.333987	0.254248	0.217647	0.21634
15	0.083006	0.183987	0.375164	0.248366	0.20915	0.209804
20	0.428105	0.183987	0.372549	0.271896	0.21634	0.211111
25	0.432353	0.183987	0.357516	0.259151	0.202287	0.193464
50	0.48987	0.245752	0.389869	0.269281	0.189869	0.366667

In case of inter-crop analysis, as it can be seen from table 7, the MMD values obtained were higher which shows that there is separation between target and non-target crop. This separation should be as high as possible to be able to highlight the class of interest. The MMD values for fenugreek and wheat were less. This is because of similarity in the phenology and spectral profile of Psyllium Husk, Fenugreek and Wheat. Fenugreek is quite similar to Psyllium Husk visually too, i.e. both have a small structure, which further

increases the spectral similarity between the two. While in case of wheat, the crop structure is different, wheat is denser and more continuous with respect to Psyllium Husk are distant from each other. Still there is spectral similarity between the two which accounts for the fact that MSAVI2 has worked for the benefit and has reduced the soil impact up to a large extent which has ultimately brought these two crops to be comparable. CBSI-MSAVI2 based both the approaches i.e. mean as well as ISM can be observed to have differentiated between

Psyllium Husk and Fenugreek the best among the three.

As it can be observed from table 8, MMD values within the Psyllium Husk crop field are very low which shows that there is less difference between the membership values of training and testing class and mapping has been carried out efficiently. The results for accuracy show consistent and close values irrespective of the technique and number of samples used.

On the basis of overall performance, the optimum approach and number of samples that balance the variance within the field, minimizes MMD values within target class and maximizes inter-class MMD values were concluded to be ISM approach using CBSI-MSAVI2 index. Also the MMD as well as variance values can be observed as saturating around 10 and 15 training samples thus showing that the algorithm doesn't require all of 50 samples to provide fruitful results.

Table 8: MMD and Variance within Psyllium Husk crop for MSAVI2 variants and MPCM approaches

Number of Samples	Mean		ISM	
	MMD	Variance	MMD	Variance
Conventional MSAVI2				
5	0.025882	0.002446	0.013594	0.137638
10	0.075294	0.001487	0.052287	0.000752
15	0.056732	0.004603	0.044967	0.000770
20	0.014379	0.002269	0.044967	0.000770
25	0.006536	0.002172	0.048104	0.000649
50	0.043398	0.003901	0.222745	0.002981
MSAVI2 Red Edge-1				
5	0.018823	0.016880	0.074771	0.094211
10	0.117124	0.004745	0.036601	0.001362
15	0.087843	0.007962	0.027973	0.000995
20	0.087320	0.008192	0.037124	0.000448
25	0.092810	0.007404	0.043921	0.000361
50	0.056731	0.014915	0.059346	0.000286
CBSI MSAVI2				
5	0.024575	0.003994	0.022745	0.003769
10	0.032418	0.003765	0.023791	0.004315
15	0.033725	0.004041	0.031111	0.004187
20	0.049411	0.004636	0.042614	0.004187
25	0.041568	0.004563	0.071111	0.001600
50	0.053333	0.005233	0.090196	0.002981

This work can further be continued to evaluate the performance of CBSI approach for other spectral indices across different targets. The noise in the output can be dealt with by application of standard or dedicated filters.

5. Conclusion

In order to carry out efficient mapping for specific crop, it is important to process temporal data with acquisition dates that cover unique crop stages. This further helps in differentiating the target crop and highlighting it among the presence of other similar crops. The bands selected for vegetation indices calculation for crop mapping influence the classified outputs and hence should be selected carefully so as to match the target under study. The classification technique used in this study

i.e. Fuzzy MPCM suits the applications since the pixels are not pure and may contain some other class. The ISM approach for MPCM classification shows less variance within the target fields detected which points that the output is unaffected by heterogeneity within the class. The results obtained after varying the number of training samples show that at a certain point that was 10 to 15 training samples, the results get saturated and are unaffected or negligibly affected by an increase in the number of training samples. The accuracy assessment results show that both the variants of MSAVI2 i.e. conventional and CBSI-MSAVI2 give fair results with good overall accuracy as well as kappa coefficient.

Conflict of interest: There is no conflict of interest in this research paper

References

- [1] T. Zhou, Z. Li, and J. Pan, "Multi-feature classification of multi-sensor satellite imagery based on dual-polarimetric sentinel-1A, landsat-8 OLI, and hyperion images for urban land-cover classification," *Sensors (Switzerland)*, vol. 18, no. 2, pp. 1–20, 2018, doi: 10.3390/s18020373.
- [2] Y. Xue, Y. Li, J. Guang, X. Zhang, and J. Guo, "Small satellite remote sensing and applications - History, current and future," *Int. J. Remote Sens.*, vol. 29, no. 15, pp. 4339–4372, 2008, doi: 10.1080/01431160801914945.
- [3] S. Khanal, K. C. Kushal, J. P. Fulton, S. Shearer, and E. Ozkan, "Remote sensing in agriculture—accomplishments, limitations, and opportunities," *Remote Sens.*, vol. 12, no. 22, pp. 1–29, 2020, doi: 10.3390/rs12223783.
- [4] 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, 2020, doi: 10.1080/24751839.2019.1694765.
- [5] P. Naik, M. Dalponte, and L. Bruzzone, “A comparison on the use of different satellite multispectral data for the prediction of aboveground biomass,” no. July, p. 40, 2020, doi: 10.1117/12.2572807.
- [6] I. Pôças, A. Calera, I. Campos, and M. Cunha, “Remote sensing for estimating and mapping single and basal crop coefficients: A review on spectral vegetation indices approaches,” *Agric. Water Manag.*, vol. 233, no. December 2019, p. 106081, 2020, doi: 10.1016/j.agwat.2020.106081.
- [7] L. Korhonen, Hadi, P. Packalen, and M. Rautiainen, “Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index,” *Remote Sens. Environ.*, vol. 195, pp. 259–274, 2017, doi: 10.1016/j.rse.2017.03.021.
- [8] C. Sun, Y. Bian, T. Zhou, and J. Pan, “Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region,” *Sensors (Switzerland)*, vol. 19, no. 10, pp. 1–23, 2019, doi: 10.3390/s19102401.
- [9] P. Villa, D. Stroppiana, G. Fontanelli, R. Azar, and P. A. Brivio, “In-season mapping of crop type with optical and X-band SAR data: A classification tree approach using synoptic seasonal features,” *Remote Sens.*, vol. 7, no. 10, pp. 12859–12886, 2015, doi: 10.3390/rs71012859.
- [10] B. Biswas, S. Walker, and M. Varun, “Web GIS based identification and mapping of medicinal plants : A case study of Agra (U.P.), India,” *Plant Arch.*, vol. 17, no. 1, pp. 8–20, 2017.
- [11] Y. C. Tripathi, V. V Prabhu, R. S. Pal, and R. N. Mishra, “Medicinal plants of rajasthan in Indian system of medicine,” *Anc. Sci. Life*, vol. 15, no. 3, pp. 190–212, 1996, [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22556743> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3331209>.
- [12] A. Bégué et al., “Agricultural Systems Studies using Remote Sensing To cite this version : HAL Id : hal-02098284,” Hal, 2019.
- [13] M. Weiss, F. Jacob, and G. Duveiller, “Remote sensing for agricultural applications: A meta-review,” *Remote Sens. Environ.*, vol. 236, no. August 2019, p. 111402, 2020, doi: 10.1016/j.rse.2019.111402.
- [14] A. Karakoç and M. Karabulut, “Ratio-based vegetation indices for biomass estimation depending on grassland characteristics,” *Turk. J. Botany*, vol. 43, no. 5, pp. 619–633, 2019, doi: 10.3906/bot-1902-50.
- [15] Z. Zhang, M. Liu, X. Liu, and G. Zhou, “A new vegetation index based on multitemporal sentinel-2 images for discriminating heavy metal stress levels in rice,” *Sensors (Switzerland)*,

- vol. 18, no. 7, 2018, doi: 10.3390/s18072172.
- [16] L. Cabrera-Bosquet, G. Molero, A. Stellacci, J. Bort, S. Nogués, and J. Araus, “NDVI as a potential tool for predicting biomass, plant nitrogen content and growth in wheat genotypes subjected to different water and nitrogen conditions,” *Cereal Res. Commun.*, vol. 39, no. 1, pp. 147–159, 2011, doi: 10.1556/CRC.39.2011.1.15.
- [17] C. Kalaitzidis, V. Heinzl, and D. Zianis, “A Review of Multispectral Vegetation Indices for Biomass Estimation,” *Imagin {E,G} Eur.*, no. January 2016, pp. 201–208, 2010, doi: 10.3233/978-1-60750-494-8-201.
- [18] R. Masood and M. MirafTAB, “Psyllium: Current and Future Applications,” *Med. Healthc. Text.*, pp. 244–253, 2010, doi: 10.1533/9780857090348.244.
- [19] A. R. Huete, “A Soil-Adjusted Vegetation Index (SAVI),” *Bangladesh Med. Res. Counc. Bull.*, vol. 22, no. 1, pp. 27–32, 1996.
- [20] P. Naik, M. Dalponte, and L. Bruzzone, “A DISENTANGLED VARIATIONAL AUTOENCODER FOR PREDICTION OF ABOVE Department of Information Engineering and Computer Science , University of Trento , Italy . Sustainable Agro-ecosystems and Bioresources Department , Fondazione Edmund Mach , Italy .” *Int. Geosci. Remote Sens. Symp.*, pp. 2991–2994, 2021.
- [21] D. N. H. Horler, M. Dockray, and J. Barber, “The red edge of plant leaf reflectance,” *Int. J. Remote Sens.*, vol. 4, no. 2, pp. 273–288, 1983, doi: 10.1080/01431168308948546.
- [22] A. Moumni and A. Lahrouni, “Machine Learning-Based Classification for Crop-Type Mapping Using the Fusion of High-Resolution Satellite Imagery in a Semiarid Area,” *Scientifica (Cairo)*, vol. 2021, 2021, doi: 10.1155/2021/8810279.
- [23] L. Viskovic, I. N. Kosovic, and T. Mastelic, “Crop classification using multi-spectral and multitemporal satellite imagery with machine learning,” 2019 27th Int. Conf. Software, Telecommun. Comput. Networks, SoftCOM 2019, 2019, doi: 10.23919/SOFTCOM.2019.8903738 .
- [24] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” *Sensors (Switzerland)*, vol. 18, no. 8, pp. 1–29, 2018, doi: 10.3390/s18082674.
- [25] V. K. Dadhwal, R. P. Singh, S. Dutta, and J. S. Parihar, “Remote sensing based crop inventory: A review of Indian experience,” *Trop. Ecol.*, vol. 43, no. 1, pp. 107–122, 2002.
- [26] N. Kakhani and M. Mokhtarzade, “A new neuro-fuzzy-based classification approach for hyperspectral remote sensing images,” *J. Earth Syst. Sci.*, vol. 128, no. 2, pp. 1–21, 2019, doi: 10.1007/s12040-018-1054-9.

- [27] O. Mohd, N. Suryanna, S. Sahib Sahibuddin, M. Faizal Abdollah, and S. Rahayu Selamat, "Thresholding and Fuzzy Rule-Based Classification Approaches in Handling Mangrove Forest Mixed Pixel Problems Associated with in QuickBird Remote Sensing Image Analysis," *Int. J. Agric. For.*, vol. 2, no. 6, pp. 300–306, 2012, doi: 10.5923/j.ijaf.20120206.06.
- [28] M. Singha, A. Kumar, A. Stein, P. N. L. Raju, and Y. V. N. K. Murthy, "Importance of DA-MRF Models in Fuzzy Based Classifier," *J. Indian Soc. Remote Sens.*, vol. 43, no. 1, pp. 27–35, 2015, doi: 10.1007/s12524-014-0382-8.
- [29] G. Misra, A. Kumar, N. R. Patel, R. Zurita-Milla, and A. Singh, "Mapping specific crop- A multi sensor temporal approach," in 2012 IEEE International Geoscience and Remote Sensing Symposium, 2012, pp. 3034–3037, doi: 10.1109/IGARSS.2012.6350786.
- [30] S. Kaur and R. Bansal, "Mixed Pixel Classification by Using Hybridization of Evolutionary Method With Neural Networks," *Int. J. Appl. Eng. Res.*, vol. 13, no. 8, pp. 5736–5744, 2018, [Online]. Available: <http://www.ripublication.com5736>.
- [31] G. M. Foody, N. A. Campbell, N. M. Trodd, and T. F. Wood, "Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 9, pp. 1335–1341, 1992.
- [32] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2, pp. 191–203, 1984, doi: [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7).
- [33] O. OZDEMİR and A. KAYA, "Ecoli Veri Protein Lokalizasyonunda Bulanık ve Olabilirlikli Kümeleme Algoritmalarının Analizi," *Afyon Kocatepe Univ. J. Sci. Eng.*, vol. 19, no. 1, pp. 92–102, 2019, doi: 10.35414/akufemubid.429540.
- [34] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, 1993, doi: 10.1109/91.227387.
- [35] A. Singh, A. Kumar, and P. Upadhyay, "Modified possibilistic c-means with constraints (MPCM-S) approach for incorporating the local information in a remote sensing image classification," *Remote Sens. Appl. Soc. Environ.*, vol. 18, no. March, p. 100319, 2020, doi: 10.1016/j.rsase.2020.100319.
- [36] T. G. Van Niel, T. R. McVicar, and B. Datt, "On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification," *Remote Sens. Environ.*, vol. 98, no. 4, pp. 468–480, 2005, doi: 10.1016/j.rse.2005.08.011.
- [37] M. Suresh and K. Jain, "Subpixel level mapping of remotely sensed

- image using colorimetry,” *Egypt. J. Remote Sens. Sp. Sci.*, vol. 21, no. 1, pp. 65–72, 2018, doi: 10.1016/j.ejrs.2017.02.004.
- [38] E. Binaghi, P. A. Brivio, P. Ghezzi, and A. Rampini, “A fuzzy set-based accuracy assessment of soft classification,” *Pattern Recognit. Lett.*, vol. 20, no. 9, pp. 935–948, 1999, doi: 10.1016/S0167-8655(99)00061-6.
- [39] J. L. Silván-Cárdenas and L. Wang, “Sub-pixel confusion-uncertainty matrix for assessing soft classifications,” *Remote Sens. Environ.*, vol. 112, no. 3, pp. 1081–1095, 2008, doi: 10.1016/j.rse.2007.07.017.
- [40] A. Kumar, S. K. Ghosh, and V. K. Dadhwal, “ALCM: Automatic land cover mapping,” *J. Indian Soc. Remote Sens.*, vol. 38, no. 2, pp. 239–245, 2010, doi: 10.1007/s12524-010-0030-x.