MAPPING AND YIELD PREDICTION OF CASTOR BEAN (*RICINUS COMMUNIS*) USING SENTINEL-2A SATELLITE IMAGE IN A SEMI-ARID REGION OF INDIA

RITESH KUMAR[1], NARENDRA SINGH BISHNOI[1], NIMISH NARAYAN GAUTAM[2], MUSKAN[3], VARUN NARAYAN MISHRA[4*]

[1]*Haryana Space Applications Centre (HARSAC), Citizen Resources Information Department, Govt. of Haryana, CCS HAU Campus, Hisar-125004, India*
[2]*ICAR-Central Arid Zone Research Institute, Jodhpur-342003, India*
[3]*Department of Environmental Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar-125001, India*
[4]*Amity Institute of Geoinformatics & Remote Sensing (AIGIRS), Amity University, Sector 125, Noida-201313, Gautam Buddha Nagar, India*
**Corresponding author email: varun9686@gmail.com; vnmishra@amity.edu*

Received: 3rd May 2023, **Accepted:** 12th June 2023

ABSTRACT

Castor bean (*Ricinus communis*) indigenous to the southeastern Mediterranean basin, eastern Africa and India is a crop having various industrial and medicinal applications. It is helpful in crop rotation and replenishing the soil nutrients due to less water consumption. The current study explores the utility of Sentinel-2A satellite image for mapping and yield prediction of castor beans. Several classification methods viz. migrating means clustering, maximum likelihood classifier, support vector machine and artificial neural network are used for the classification and mapping of different landscape categories. The overall classification accuracy was achieved to be highest for artificial neural network (85.81 %) subsequently support vector machine (80.12 %), maximum likelihood classifier (74.23 %) and migrating means clustering (73.03 %). The yield prediction is performed using Sentinel-2A-derived indices namely Normalized Difference Vegetation Index and Enhanced Vegetation Index-2. Further, the cumulative values of these two indices are investigated for castor bean yield prediction using linear regression from July 2017 to April 2018 in different seasons (pre-monsoon, post-monsoon, and winter). The regression model provided (adj R2=0.75) value using EVI-2 compared to (adj R2=0.55) using NDVI for yield prediction of *Ricinus communis* crop in the winter season. The methodology adopted in this study can serve as an effective tool to map and predict the productivity of *Ricinus communis*. The adopted methodology may also be extended to a wider spatial level and for other significant crops grown in semi-arid regions of world.

Keywords: Castor bean; Sentinel 2; Classification; Regression; Yield

INTRODUCTION

Castor bean (*Ricinus communis*) is mainly distributed in the arid and semi-arid regions of Eastern India and having medicinal and economic importance. It is commonly known as “Arandi” at local level and a vital biodiesel crop with several industrial applications such as

paints, nylon, and motor oil (Agri Farming, 2023). The castor bean is encouraged to be grown to provide an alternative to farmers when shifting from traditional cropping practices for earning a profit. The importance of cropping practices in terms of productivity, utilization of natural resources, and farmers' income have long been recognized by the international community under the Food and Agriculture Organization (FAO) regulation (Polsot, 2004; Bégué *et al*., 2018). It is increasingly difficult for local governments and consumers to control the food production process at different observational scales. An area's changing land use pattern is a major aspect of the pressure on limited agricultural land resources, driven by various biophysical and anthropogenic factors, mainly due to population growth. Therefore, detecting and validating good cropping practices is vital (Bégué *et al*., 2018). An effective technique is required to provide spatial information on cropping practices and their distribution to ensure sustainable food production and management of water resources.

Space-borne remote sensing technique has been proven as an effective means for observing cropping practices and other land surface features at different spatio-temporal scales (Mishra *et al*., 2017a; Kumar *et al*., 2017). Synoptic and repeated coverage are well-known advantages of remote sensing images that facilitate mapping the multi-temporal pattern and rotation of cropping systems. In the last few decades, images from a large variety of sensors are gaining a lot of interest from the scientific community (Belward, 2015). A cropping system can be stated as the cropping patterns and their management to obtain maximum reimbursement from a given source found under a particular environmental condition. The crop rotation and shifting agriculture with long gaps help in gaining the necessary nutrients for the agricultural land for high productivity. Remotely sensed images are a noticeable and promising source of information for mapping and monitoring crops from the regional to global level (Skakun *et al*., 2017). In several studies, remote sensing-based spectral, phenological, and spatial information is also used for crop classification, crop acreage estimation, crop health, and yield prediction (Kumar *et al*., 2015; Ghazaryan *et al*., 2018).

Then accuracy of classification results is critical based on the remotely sensed images and understanding of the classification algorithms. Several classification methods are available for landscape mapping using images acquired from different satellite sensors (Lu & Weng, 2007). One of the most used conventional methods is the maximum likelihood classifier (MLC) which relies on statistical theory (Richards & Jia, 2006). But advanced non-parametric classifiers are required further and found to be very valuable due to the limitations of normal distribution (Lu *et al*., 2004). Artificial neural network (ANN), a non-parametric classification method, works on back propagation (BP) training algorithm and not influenced by the normal distribution of data (Lu & Weng, 2007). ANN is broadly applied in many studies and found to provide enhanced accuracies for landscape classification compared to the traditional pixel-based method (Pijanowski *et al*., 2002; Mas & Flores, 2008). Cortes & Vapnik (1995) proposed another non-parametric statistical learning method SVM which has been widely used in image classification. Several studies have demonstrated its abilities when only small training samples are available (Chu *et al*., 2012; Mishra *et al.*, 2017b). The SVM's performance has been compared with other classification methods and found to provide improved results (Mishra *et al.*, 2017b). The precise and accurate mapping of the agricultural crop and yield estimation based on remote sensing are of great importance for national food security at different observational scales.

The crop yield prediction or estimation is one of the most significant concerns that appeared, especially in regions with climatic uncertainties. The seasonal rainfall trends, nutrient depletion in soils, cost and availability of fertilizers, pest control, and other aspects bring out the reduction in crop productivity. The crop yield estimation also facilitates governments to put in place strategic emergency plans for the redistribution of food during

times of food crisis. Over the last few decades, remarkable advancements in remote sensing techniques are now supplying valuable information for predicting crop yield at different scales. Several studies were conducted for crop yield prediction, mainly focusing on winter wheat, maize, rice, corn, soybean etc., using remotely sensed data (Mkhabela *et al*., 2011; Esquerdo *et al*., 2011; Huang *et al*., 2013; Shiu & Chuang, 2019). Many vegetation indices, for example, vegetation condition index (VCI), normalized difference vegetation index (NDVI), and enhanced vegetation index (EVI) have been correlated with crop yield estimation. Numerous methods have been developed for crop yield prediction with the help of remotely sensed images. Most frequently used way based on the generation of a regression model, is to create direct empirical relationships between the NDVI derived from the satellite images and the measured crop yield (Mkhabela *et al*., 2011; Boltan & Friedle, 2013). The NDVI, based on red and near-infrared bands, introduces noise because of underlying soil reflectance and atmospheric absorption. The EVI based on the blue band is developed to minimize the soil and atmospheric influences (Huete *et al.,* 2002). Conversely, it is inadequate to produce time series EVI without a blue band in a sensor. Jiang *et al*. (2008) developed and evaluated a 2-band enhanced vegetation index (EVI2) which does not include a blue band. EVI2 has been utilized effectively in classifying croplands and estimating crop yield (Sun *et al*., 2012; Liu *et al*., 2019; Shammi & Meng, 2021). However, very limited studies have been reported till now focusing on multi-cropping practices and its spatial mapping in India.

Crop rotation and shifting agriculture with long gaps help in gaining the necessary nutrients for the agricultural land for high productivity. There are problems with canal-based irrigation and groundwater availability in most areas around the globe. Several studies performed so far are relying on remote sensing dataset for retrieval of information on major crops grown in an area. *Ricinus communis* is a minor crop, helpful in crop rotation and replenishing the soil nutrients and does not need much water for irrigation. It is considered a cash crop for farmers with various applications in industrial and health care sectors. This study provides a knowledge base on simplified technological intervention for better information retrieval about the minor crop having dispersed extent and its quantification through an integrated approach of remote sensing, GIS and yields prediction model. Additionally, the inferences of the present study will be a primary input for research work to be done in future and track the interested groups in this area.

This study aims to map and predict the yield of *Ricinus communis* using remote sensing-based information and ground data. More explicitly, the present study aims to: (1) identify and map the industrial crop (*Ricinus communis*) using Sentinel 2A satellite data; (2) estimate the acreage of the crop; and (3) use a linear regression algorithm for yield prediction of this industrial crop. Therefore, analyzing and modeling the cropping pattern and productivity of *Ricinus communis* will provide vital information to the research community, agriculture planners, decision-makers, and other stakeholders to understand its importance and benefits to farmers in India. The outcomes of this study could also be applied to other blocks or even at the state level in India and other geographical sites around the world.

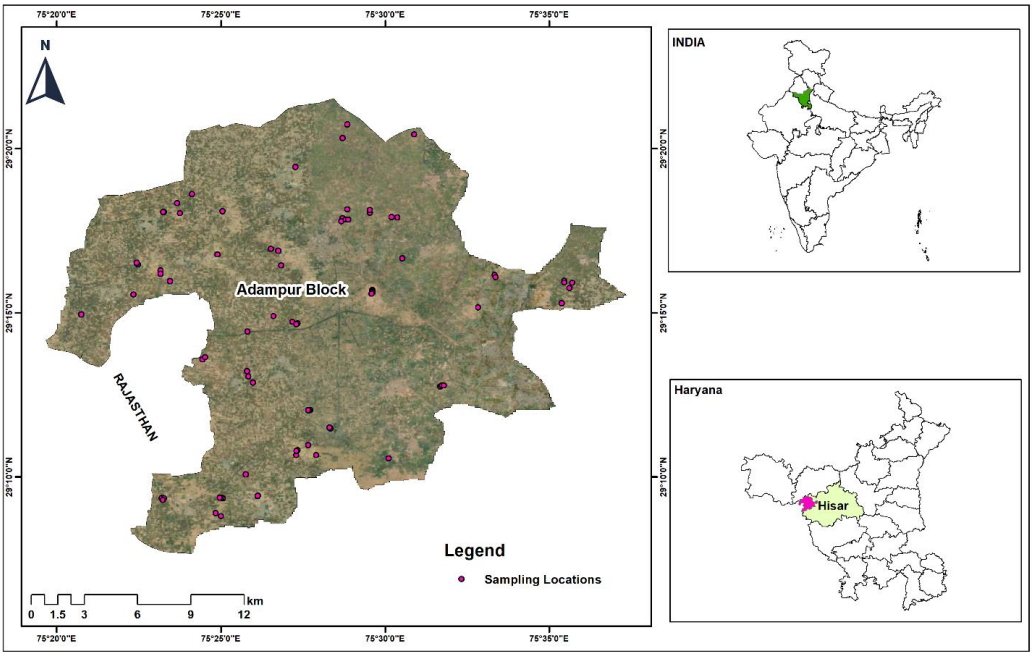
MATERIALS AND METHODOLOGY

Study site

The geographical location of Adampur block is between 29° 06′ to 29° 23′ latitudes and 75° 22′ to 75° 37′ longitudes, covering an area of about 37596 ha (Figure 1). This block shares its administrative boundary with Balsamand, Hisar-II and Barwala blocks in Hisar district and with Fatehabad district of Haryana state, India. It is well known for the

agricultural production of Cotton, Cluster bean (Gawar), Wheat, Gram, etc. Climatically, it experiences a sub-tropical climate with a considerable temperature difference between the summer and winter seasons. The normal annual rainfall in Adampur block is around 330 mm. The rainfall in the block usually increases from southwest to northeast. The soil type of study area is grouped as Sierozem and Desert soils.

Fig. 1: Location map of the study area



Castor bean has a wide range of adaptability as it grows throughout warm, temperate, and tropical regions, the limiting factor being frost. The seedlings of castor plants take 10-20 days to emerge and flowering after 40-70 days. Every single plant bear flower at various stages of its development. It takes 140 – 170 days to attain full maturity. The ripening of fruits is very uneven, the lower fruits maturing earlier than the upper one and in the region of water scarcity or at elevation and sandy soil, the period of first and last mature fruits may be of many weeks. The fruits occur thrice (post-monsoon, winter, and pre-monsoon) on the same plant every year.

Data description

In this study, a level-1C product of Sentinel 2A-Multi-Spectral Instrument (MSI) images for the years 2017 (July to December) and 2018 (January to April) were downloaded from ESA website (http://scihub.copernicus.eu/) and subjected to further processing. During the project work only, Level-1C product was available for India and it was required to convert it into Bottom of Atmosphere (BOA) reflectance data before utilizing. A systematic generation of Level-2A products are carried out at the ground segment over Europe since March 2018, and was extended to global scale in December 2018 (https://sentinels.copernicus.eu /web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a). Sentnel-2 carries the

Multispectral Imager (MSI) covering 13 spectral bands (443 to 2190 nm) with 12-bit quantization, a revisit time of 10 days and a swath width of 290 km. The MSI image is having a spatial resolution of 10 m (four; 490nm, 560nm, 665nm (visible) and 842nm (near-infrared) bands), 20 m (three; 705nm, 740nm and 783 nm (red edge) and one; 865nm (near-infrared), two; 1610nm, 2190nm (shortwave infrared bands) and 60 m (three atmospheric correction bands).

Image pre-processing

The Sentinel 2A-MSI images were downloaded as top-of-atmosphere (TOA) reflectance products. It is processed further to convert it into bottom-of-atmosphere (BOA) or surface reflectance values, including atmospheric correction. The Sentinel toolbox provides the Sen2Cor processor for converting TOA into BOA reflectance to derive L2A MSI land products. An operational atmospheric correction based on the Sen2Cor algorithm is applied to the MSI bands for retrieving the atmospheric variables from the image itself, with correction at 1375 nm; water vapour retrieval at 865, 945 nm and AOD retrieval (Martins *et al.*, 2017). Therefore, the Sen2Cor algorithm implements a semi-empirical method by associating the atmospheric characteristics derived from the image with the help of a pre-computed look-up table (LUT) using the libRadtran radiative transfer model.

Extraction of NDVI

The NDVI is a dimensionless index that describes the difference between visible and near-infrared reflectance of vegetation cover and can be used to estimate the density of green on an area of land (Weier & Herring, 2000). The broadly used NDVI was employed to distinguish vegetation and non-vegetation area from the satellite images. The NDVI was calculated using the surface reflectance of red (R) and near-infrared (NIR) bands of Sentinel 2A-MSI image using the formula given as

$$\mathrm{NDV} = \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \mathrm{R}} \quad (1)$$

Extraction of EVI-2

It was originally built to optimize the signals from vegetation in high biomass areas for enhancing the sensitivity to better monitoring of vegetation by separating the signals coming from the background of canopies and reduction in atmosphere effects. It is particularly developed to use MODIS data products (Huete *et al.*, 1999). The EVI uses atmospherically corrected surface reflectance of blue (B), red (R) and near-infrared (NIR) bands. The coefficients used in EVI are, L=1, C1=6, C2=7.5, and G (gain factor) = 2.5. In EVI, the blue band does not supply any added biophysical information on the properties of vegetation. It is somewhat focused at reducing the noise and uncertainties related to the enormously variable atmospheric aerosols. Therefore, it is likely to develop an alternative method by decomposing equation (2) into a 2-band EVI i.e., EVI2 without a blue band. Since the visible bands are extremely correlated to each other over many earth's surface features. Therefore, the reflectance from blue band can be given as a function of the reflectance of red band (Jiang *et al.*, 2008). Equation (3) can be diminished to a 2-band EVI by assuming the relationship, Red = c × Blue, with the values of L, C1, and C2 as given above.

$$\mathrm{EVI2} = \mathrm{G} \times \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + (6 - 7.5/\mathrm{c})\mathrm{R} + 1} \quad (2)$$

where G is to be determined in accordance with the c value.

The vegetation indices thus generated were subjected to regression against the crop productivity data collected from the field.

Satellite image classification

In this study, four different classification methods, including an unsupervised classifier (MMC), a traditional supervised classifier (MLC), and two machine learning classifiers (SVM and ANN) are applied to investigate the best-performing classifiers for mapping *Ricinus communis* in the study region. A classification schema considering Land Use in the study area was defined to classify the images into classes: *Ricinus communis*, Cotton, Cluster bean (Gawar) Waterbody, Sand Dunes, Built-up, Open Space and Others. The intention behind this was to find out the different land use area and finally to delineate the minor crop castor beans dispersed percentage coverage. A set of training pixels were collected on satellite images from different training sites to classify the study area into classes as per defined classification schema. A total of 160 training sites were identified including 20 sites in each class category. The same training samples are utilized for each classifier used in this study.

MMC-based classification

The MMC is the most frequently used unsupervised classification method. The user does not require the fore knowledge of the landscape types. It uses some clustering algorithm around assumed means to classify image data (Richards, 1993). This method can be used to find the number and location of unimodal spectral categories. The MMC classifier marks each pixel to unknown cluster centers after that, moves to another cluster center in a way to reduce the Error Sum of Squares (SSE) measure of the former segment (Richards, 1993). The clustering algorithm was optimized with derivation of 60 clusters, 12 iterations and 0.95 clustering threshold.

MLC-based classification

MLC is the most widely used supervised classification approach. It assumes that every spectral category can be explained by a multivariate normal distribution. It utilizes the directly captured training signatures from the satellite data to be classified. Conversely, the efficacy of MLC relies convincingly on the exact assessment of the mean vector in addition to the covariance matrix for every spectral category data (Richards, 1993). The likelihood is computed for a pixel belonging to each category under consideration. Then a category with maximum likelihood is assigned to the pixels in an image. Thus, it is essential to choose a suitable classification scheme so as to individual category follows a Gaussian distribution (Lillesand *et al*., 2008). Nearly 20 training sites for each desired class were considered to train the classifier. The same polygons of training sites were considered for the different supervised classifiers.

ANN-based classification

ANN is a non-linear mapping structure synonymic to the function of the human brain (Mas & Flores, 2008). It is capable of simulating underlying non-linear and complex data relationships with appropriate topological configuration (Atkinson & Tatnall, 1997). The classification of satellite images by ANN is generally trained by the back-propagation algorithm. The topological structure of ANN includes: (1) a neuron in an input layer including spectral bands, (2) each neuron in the output layer delineating the landscape categories to be mapped and (3) one or more hidden layers based on weighted channel linking components of input and output layers (Srivastava *et al*., 2012; Mishra *et al*., 2017a).

ANN is the best-suited method for the modeling of agriculture-related data, which are recognized to be complex and often non-linear in nature. Here to optimize the process the ANN classifier was configured with two neurons and 1000 iterations.

SVM-based classification

SVM is statistical learning-based non-parametric classification method (Vapnik, 1998). It offers some system-implicit benefits in comparison to other methods. In case of complex and noisy data, SVM produces good classification outcomes. SVM based on the structural risk minimization principle, minimizes the possibility of misclassifying unknown data sets (Vapnik, 1998). The primary objective of SVM is the construction of an optimal separating hyperplane between categories which is linearly distinguishable in a multi-dimensional feature space (Pal & Foody, 2010). The closest training samples to this optimal separating hyperplane are known as support vectors (Vapnik, 1998). SVM uses optimization methods to detect the best possible decision boundaries between the categories. SVM uses linear decision boundaries to distinguish linearly separable categories. The concept of kernel is introduced to handle the problems due to nonlinearly separable categories during the classification process (Cortes & Vapnik, 1995; Mishra *et al*., 2017a). In this work, radial basis function (RBF) kernel is applied for classifying the image. The RBF kernel needs fewer computational efforts and can take control over the non-linear relationship between the training and the total data sets.

Assessment of classification accuracy

The classification accuracies of different methods were assessed using the error matrix approach. It compares the classification outcome with ground truth data to identify pixel misclassification. The performance of classification methods was evaluated in terms of producer's accuracy (PA), user's accuracy (UA), overall accuracy (OA), and kappa statistics (K_c). For each classified result, an error matrix was generated to estimate the mapping accuracies of individual categories. In addition, the F-score was measured to avoid the independent class imbalance for reliable evaluation of classified results (Puissant *et al*., 2014; Mishra *et al*., 2017). A total of 256 random points were used as sampling sites for accuracy assessment of all the classified images derived from different classification methods. All these points were verified with the help of ground truth data collected during field visits and other ancillary data with FCC image.

Data was also collected for the yield of *Ricinus communis* from 72 plots of different farmers in the Adampur block. The yield data was collected per acre and was done on the dates of crop cutting. The plot size from where yield data were collected are of 1 acre each. Usually, the agriculture plot size in the study area is of 1 acre each, but farmers cultivate this industrial crop on more than one plot lying nearby, usually on dunes.

Regression analysis for yield estimation

A linear regression analysis method establishes a linear relationship between a response variable Y which varies according to the value of an intervention variable X. It is used due to the consideration that the vegetation indices are linearly related to crop photosynthetic capacity (Roujean *et al*., 1995), which can be applied to model crop biomass and measured yield (Liu *et al*., 2010; Bolton & Friedl, 2013). The model based on cumulative seasonal NDVI and EVI-2 can efficiently predict the crop yield (Jaafar & Ahmad, 2015). In this work, linear regression analysis was conducted between vegetation indices (NDVI and EVI-2) and crop yields for three different seasons separately. The NDVI and EVI-2 were calculated for the months of July 2017 to April 2018. Further the values of NDVI and EVI-2 were extracted

by spatial analyst tool available with ArcGIS for the polygons of *Ricinus communis* obtained by ANN classification. The cumulative values of crop yield data were calculated according to the seasons as pre-monsoon (March, April, May), post-monsoon (October and November) and winter (December, January, February) crops.

The dependent variable (estimated yield) on y-axis is modelled against the independent variables (cumulative NDVI and cumulative EVI2 values) on the x-axis. Thus, a linear regression model was developed between the explanatory variable (X_i) and dependent variable (Y_i) that can be represented as equation (4)

$$Y_i = a + bX_i \quad (3)$$

where *i = 1,…,n. a* and *b* are the intercept and slope, respectively.

Performance indicators of model

The performance of developed regression models was evaluated using different indices such as the squared sum of analytical error (SSPE) and Theil's fractional inequality coefficient (U). Different sources of analytical errors were differentiated by utilizing several inequality coefficients such as a fraction linked with mean differences of observed and predicted values (U_{bias}), the fraction linked with the slope (β) of the fitted model & the base line ($U_{\beta-1}$), and the fraction linked with the unexplained variance (U_e). The formulae for different indices are given below:

$$\mathrm{SSPE} = \sum_n (\mathrm{obs_i} - \mathrm{pre_i})^2 \quad (4)$$

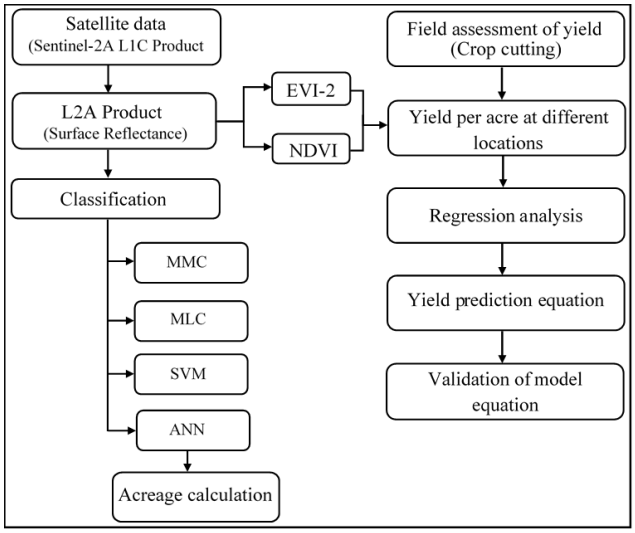
$$\mathrm{U_{bias}} = \frac{n(\mathrm{OBS} - \mathrm{PRE})^2}{\mathrm{SSPE}} \quad (5)$$

$$\mathrm{U_{\beta-1}} = \frac{(\beta - 1)^2 \sum_n (\mathrm{pre_i} - \mathrm{PRE})^2}{\mathrm{SSPE}} \quad (6)$$

$$\mathrm{U_e} = \frac{\sum_n (\mathrm{est_i} - \mathrm{obs_i})^2}{\mathrm{SSPE}} \quad (7)$$

where n is the number of sites; *pre* and *obs* are predicted and observed values correspondingly; *PRE* and *OBS* are the mean values of the predicted and observed values and *est* are estimated values of the model of fitted regression (Paruelo *et al.*, 1998). The methodology flow chart adopted in this study is given in Figure 2.

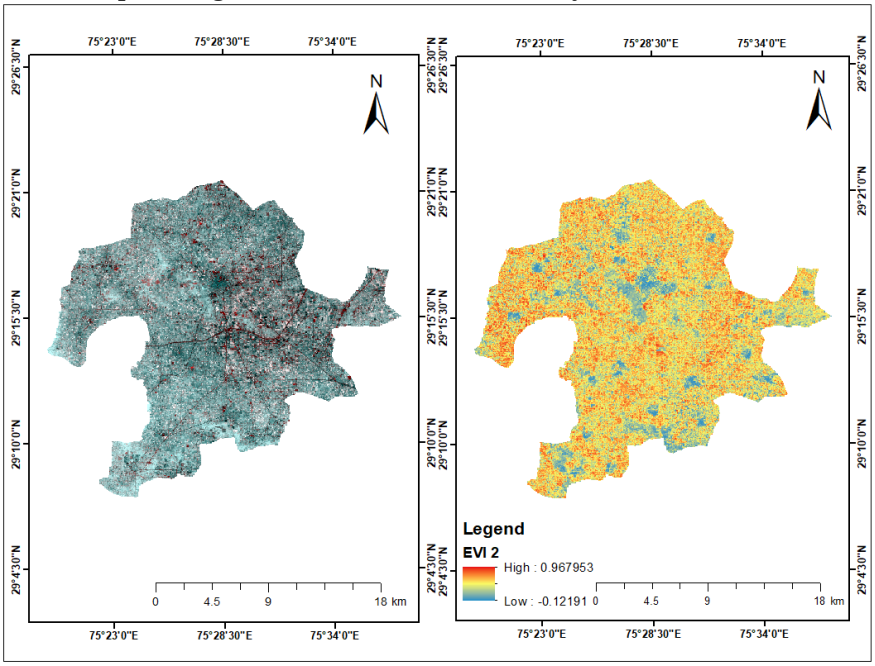
Fig. 2: Flow chart of the methodology



RESULTS

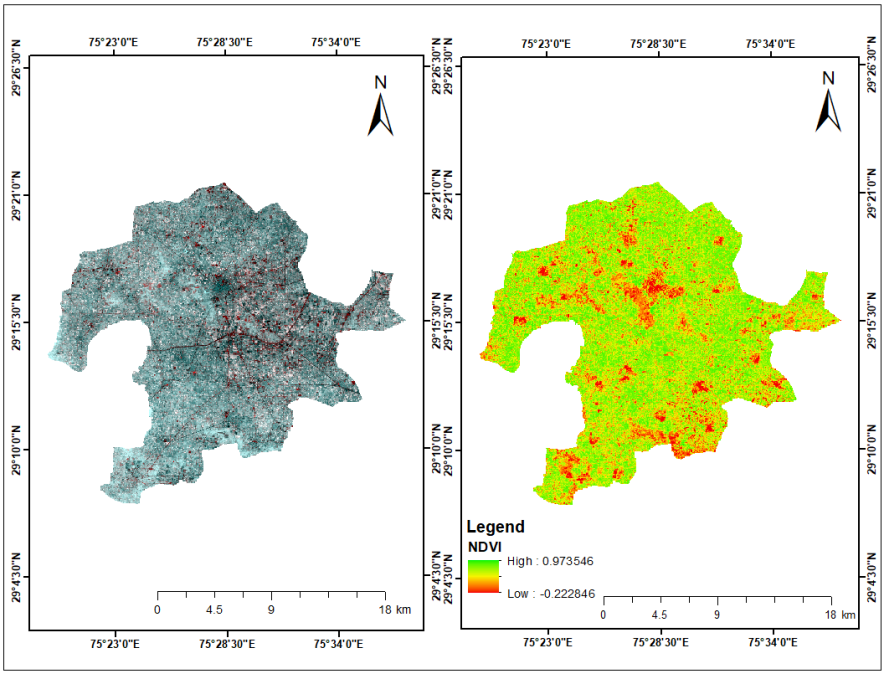
Although time-series EVI-2 image is generated using Sentinel-2A satellite data, a single timestamp for April month (pre-monsoon) is shown in Figure 3.

Fig. 3: Standard FCC and EVI-2 generated using sentinel-2 data shows the dry soil condition and sparse vegetation distribution in the study area



The values of EVI-2 range from 0.967 (highest) to -0.121 (lowest) in the study area during pre-monsoon season. Similarly, time-series NDVI images were derived from Sentinel-2A satellite data, although it is shown for April month (pre-monsoon) season in Figure 4. The values of NDVI range from 0.973 (highest) to -0.222 (lowest) in the study area. The derivation of multi-date NDVI and EVI-2 suggests that the NDVI values saturates early in high biomass condition as compared to EVI-2.

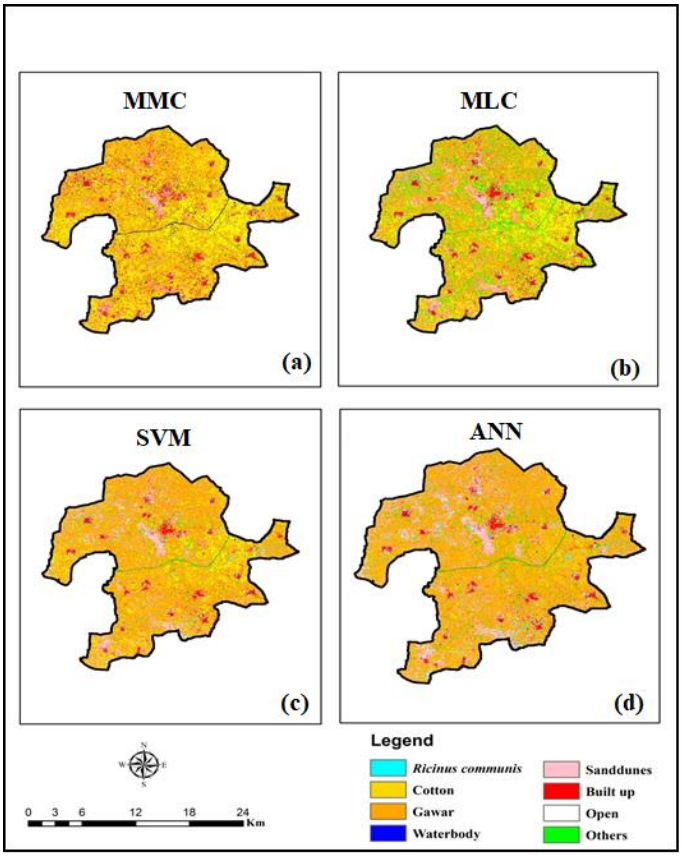
Fig. 4: Standard FCC and NDVI generated using sentinel-2 data shows the dry soil condition and sparse vegetation distribution in the study area



Assessment of classification accuracies using different methods

The image was classified into eight categories: agricultural crops (*Ricinus communis*, cotton and cluster bean), water body, sand dunes, built-up, open/barren land, and others (tree cover, grassland and pastureland). The maps generated using all four classification methods are shown in Figure 5

Fig. 5: Classified maps of the study area based on (a) MMC; (b) MLC; (c) SVM; and (d) ANN methods



ANN achieved the highest OA 85.81 % in comparison to those of SVM, MLC and MMC algorithms with 80.12, 74.23 and 73.03 %, respectively. The K_c values of 0.85, 0.77, 0.71 and 0.69 were also obtained for four different classifiers, namely ANN, SVM, MLC and MMC, respectively. The statistics of PA, UA, OA, K_c and F-score attained using four classifiers used in this study are presented in Table 1. The highest PA of *Ricinus communis* (88.32%) was achieved among all categories by using ANN in comparison to those of SVM, MLC and MCC with 79.41, 74.24 and 71.76 % respectively. The ANN method provided the highest UA (92.37 %) for *Ricinus communis* among all categories in comparison to those of SVM, MLC and MMC with 85.71, 69.01 and 67.63 %, respectively. In terms of F-score, ANN provided the highest value (90.30 %) for Ricinus *communis* compared to SVM, MLC and MMC with 82.44, 71.53, and 69.33 % respectively. A detailed description of PA, UA, OA, K_c and F-score values for all other categories is given in Table 1.

The aim was to identify and delineate the minor crop *Ricinus communis* in the study area using the best classification method. The ANN method was found to provide the highest OA in comparison to other methods. So, it was decided to consider the ANN classified data to be used further in the modeling and analysis. The ANN based classification was carried forward to extract the dispersed polygons of *Ricinus communis* for calculating its total area.

Table 1: Comparison of classification accuracy results

	MMC			**MLC**			**SVM**			**ANN**		
Landscape classes	PA (%)	UA (%)	F-score (%)	PA (%)	UA (%)	F-score (%)	PA (%)	UA (%)	F-score (%)	PA (%)	UA (%)	F-score (%)
Ricinus communis	71.76	67.63	69.33	74.24	69.01	71.53	79.41	85.71	82.44	88.32	92.37	90.30
Cotton	70.31	73.17	71.71	71.43	73.17	72.29	76.98	83.62	80.17	84.30	84.30	84.30
Cluster bean (Gawar)	72.09	69.92	70.99	73.77	69.77	71.71	79.20	80.49	79.84	87.60	80.92	84.13
Waterbody	79.31	77.97	78.63	78.45	79.13	78.79	83.47	79.53	81.45	85.12	87.29	86.19
Sand dunes	73.02	77.97	75.41	75.41	76.67	76.03	83.05	80.33	81.67	87.07	84.17	85.59
Built up	73.55	75.42	74.48	72.66	77.50	75.00	80.31	79.69	80.00	85.71	87.80	86.75
Open	71.09	70.00	70.54	71.97	76.00	73.93	78.05	76.80	77.42	81.68	83.59	82.63
Others	73.77	73.77	73.77	76.42	74.02	75.20	80.80	75.37	77.99	86.72	86.05	86.38
OA (%)	73.03			74.23			80.12			85.81		
Kappa statistics	0.69			0.71			0.77			0.85		

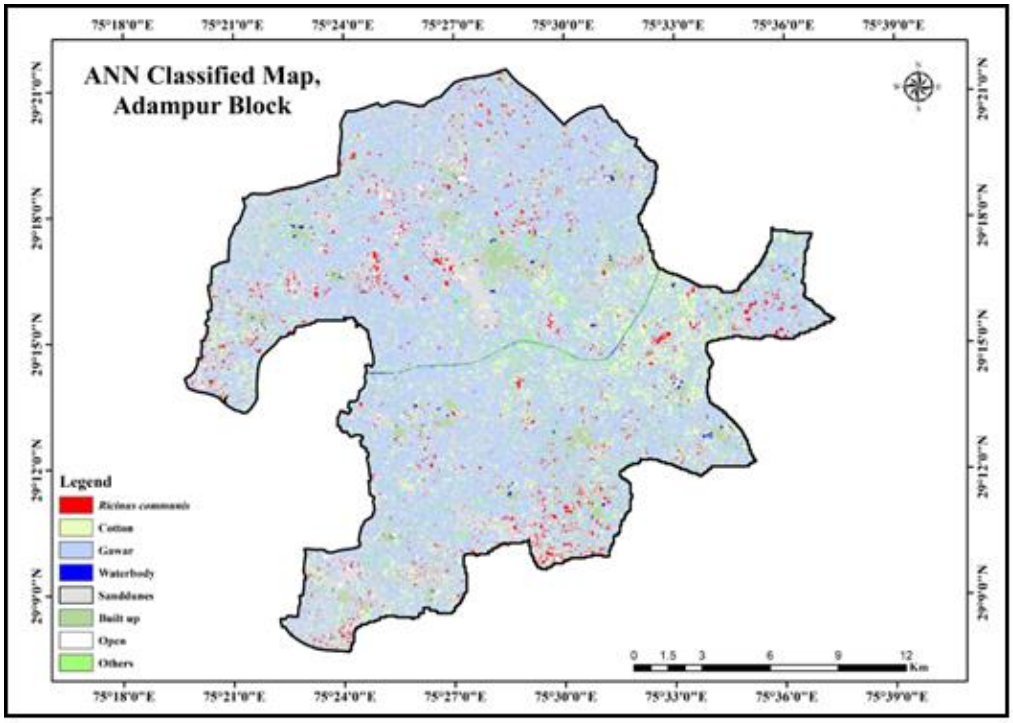
The area statistics of different landscape classes based on ANN method are shown in Table 2. Other classes in the study area were also considered while landscape classification to show the percentage distribution of different classes and highlight the usefulness of the classification technique in extracting a class with the least coverage with good enough accuracy for further processing and inclusion in the model with confidence. The total area of *Ricinus communis* was calculated as 2372.81 acres in the Adampur block, which has only 2.55 % of the coverage area. Despite the low coverage area, the ANN classifier was able to extract the castor bean area with good enough accuracy to carry the work forward. Other crops grown in the area were also considered while classification to show the cropping pattern in the study area. The cropping pattern in turn, signifies the arid conditions prevailing in the region. The classified output based on ANN is shown in Figure 6.

Table 2: Total area (acre) of different categories based on ANN classifier

Class Name	**Area (acre)**	**Area (in %)**
Ricinus communis	2372.81	2.55
Cotton	13025.34	14.02
Cluster bean (Gawar)	53517.30	57.61
Waterbody	305.49	0.33
Sand dunes	15281.44	16.45
Built up	3738.73	4.02
Open	1409.44	1.52
Others	3251.19	3.50
Total Area	92901.74	100

Ricinus communis is having very sparse and scattered distribution in the study area because farmers cultivate it only on plots where soil moisture is less, or it is difficult to irrigate due to sand dune elevations. The crop was identified, and its distribution was mapped using the ANN classification method is shown in a part of the study area overlaid on the standard FCC of Sentinel-2 data (Figure 7). The tonal variations on the standard FCC generated using a multi-band image shows that the area is dry and soil moisture is less, with a light cyan colour corresponding to the area utilized by farmers for growing castor bean.

Fig. 6: Classified map of Adampur block based on ANN method.



Enhanced Vegetation Indexe-2, originally developed for extracting vegetation information from MODIS data efficiently, without saturation as NDVI in high biomass regions, was utilized to develop the crop productivity model for *Ricinus Communis*. The time series EVI-2 was able to capture the phenological changes of the crop. The areas of *Ricinus communis* extracted using ANN classification algorithm are shown over the FCC image of time series EVI-2 (Figure 8). It shows the tonal key for identifying the crop on multi-layer vegetation index and the potential of EVI-2 in detecting *Ricinus communis* in Adampur block, Hisar. Time series Normalized Difference Vegetation Index, the commonly used index in researcher community was also calculated with Sentinel-2 data to capture phenology of the crop. The intention behind was to compare the applicability of phenology capturing using vegetation indices and its utility in crop yield modeling. In Figure 9 the extracted areas of *Ricinus communis* based on ANN algorithm are shown over the FCC image of time-series NDVI. The tonal variations shown on the FCC generated with time series NDVI is useful as key for identifying *Ricinus communis* and to demonstrate the utility of NDVI in detecting *Ricinus communis* and modeling the yield.

Fig. 7: Ricinus communis areas based on ANN classification algorithm overlaid on standard FCC of Sentinel-2 data.

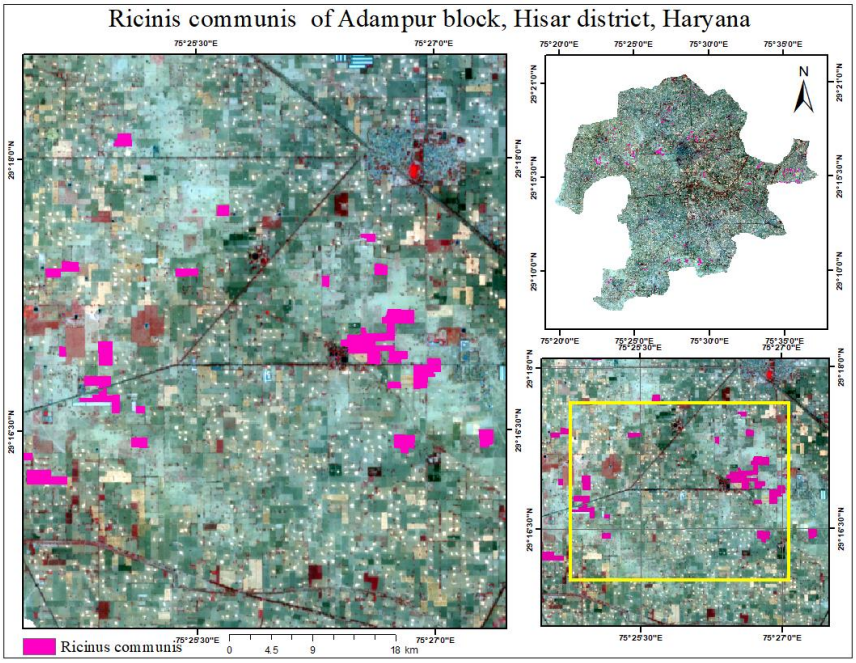


Fig. 8: Ricinus communis area based on ANN classification algorithm overlaid on FCC of time series EVI-2.

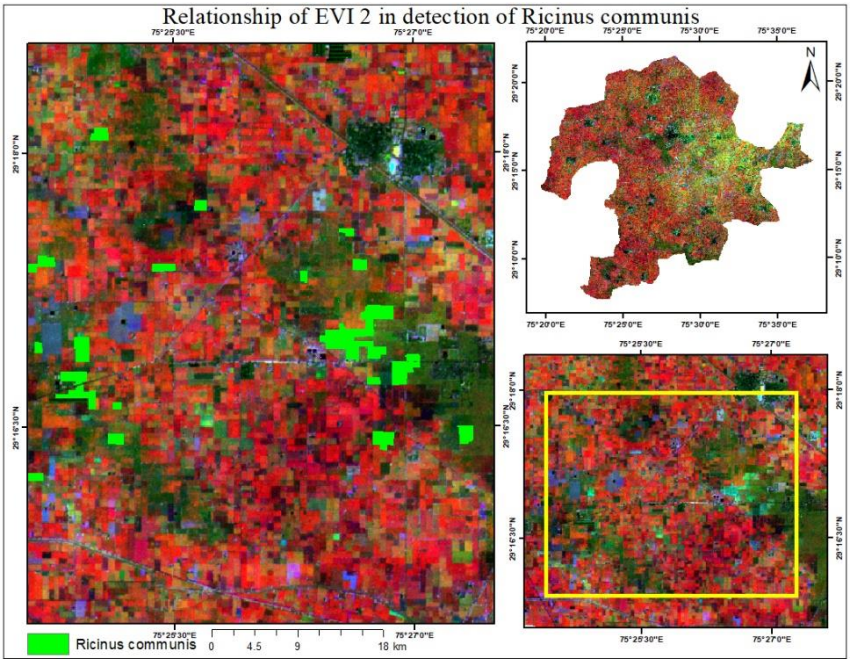
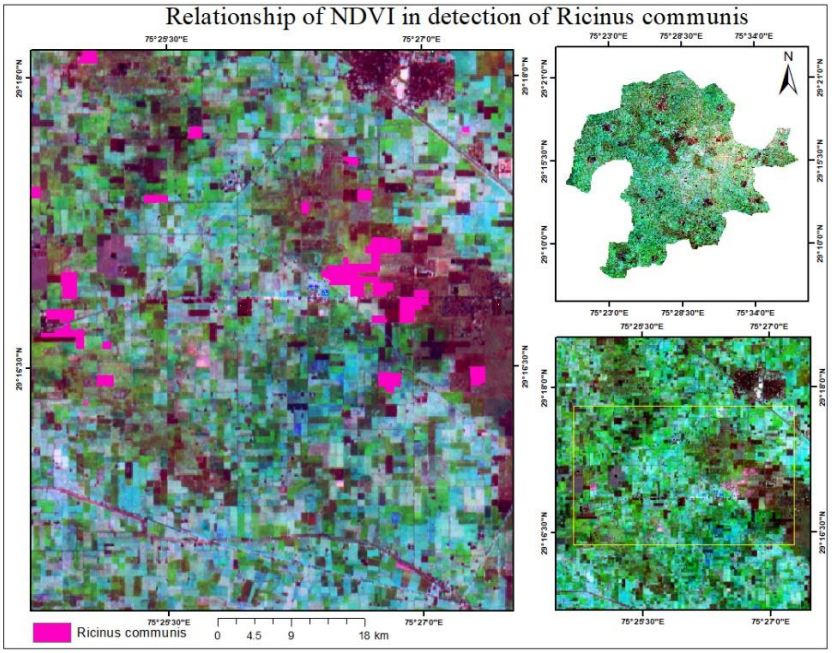


Fig. 9: Ricinus communis area based on ANN classification algorithm overlaid on FCC of time series NDVI.



Comparison of yield prediction models

The models for yield prediction are derived using cumulative NDVI and EVI-2 of different seasons and compared. The regression analysis was done between collected yield data (ground truth) on the y-axis and the cumulative values calculated seasonally on the x-axis. For regression analysis, 50 % of ground collected yield data was used and the rest 50 % of data, was used for testing the model generality. Only the positive values of NDVI collected randomly from *Ricinus communis* fields were considered for analysis and modeling the crop yield. The season wise scatter plot diagram for crop yield regressed against NDVI and EVI-2 are analyzed. The results show that pre-monsoon crop yields are poorly correlated with the cumulative NDVI and cumulative EVI-2 having coefficient of determination (R^2=0.386) and (R^2=0.094) respectively (Figure 10). Due to severity of higher range temperature prevailing in the area during pre-monsoon season the crop manifests less spread of leaf and hence reduced canopy cover. Although the NDVI and EVI-2 shows lesser values during the season, the crop yield is not affected or higher. For the post-monsoon season the crop yields also show poor correlation with the cumulative NDVI and cumulative EVI-2 having coefficient of determination (R^2=0.080) and (R^2=0.227) respectively (Figure 11). The canopy cover of the crop shows mild vigour after meagre rainfall in the area during monsoon season leading to very little increase in NDVI and EVI-2 values. However, even though there is little increase in NDVI and EVI-2 values there is no increase in crop yield, clearly showing the adaptability of the *Ricinus communis* for arid environment demanding less soil moisture. This also justifies the practice of growing the crop in this region as the castor bean crop demands less water and is an industrial utility crop, fetching good income for the farmers. The results show that winter crop yields are highly correlated with both cumulative vegetation indices given as NDVI (R^2=0.555) and EVI-2 (R^2=0.755) in the study area (Figure 12). The winter crop has

higher vigour in terms of canopy cover and shows higher values for NDVI and EVI-2 which corresponds well with the productivity in the season, showing large correlation between both vegetation indices and castor bean productivity. However, the correlation of crop productivity with winter season cumulative EVI-2 is highest and the regression model thus developed can be considered as the most successful predictor for castor bean productivity in any semi-arid region.

Fig. 10: Scatter plots of observed and predicted values for (a) NDVI and (b) EVI-2 with crop yield of *Ricinus communis* in pre-monsoon season

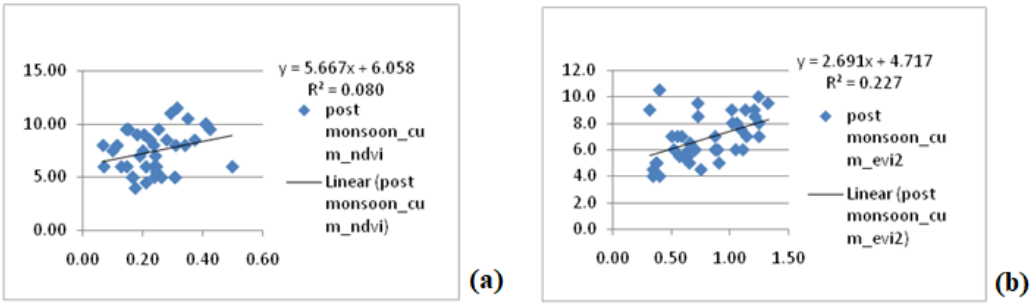


Fig. 11: Scatter plots of observed and predicted values for (a) NDVI and (b) EVI-2 with crop yield of *Ricinus communis* in post-monsoon season

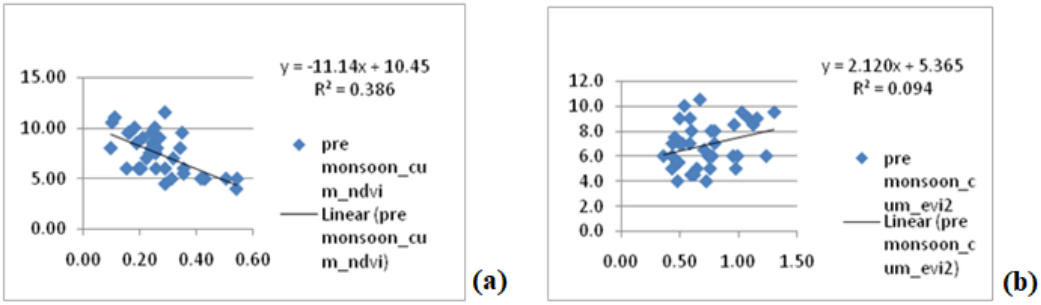
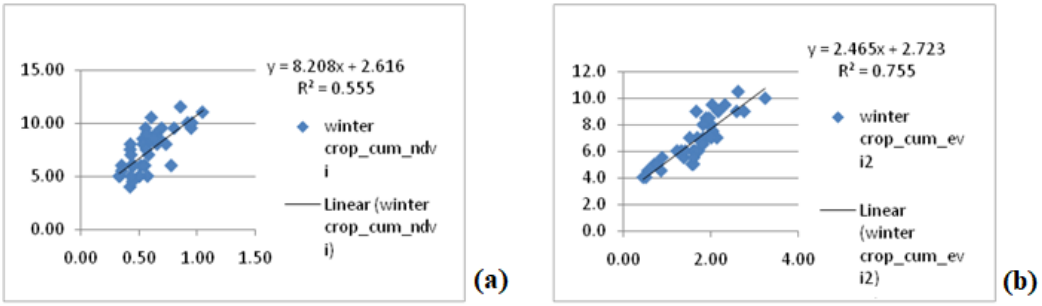


Fig. 12: Scatter plots of observed and predicted values for (a) NDVI and (b) EVI-2 with crop yield of *Ricinus communis* in winter season



The slope and y-intercept for the regression models were calculated by using the equation of cumulative winter crop yields for both the vegetation indices and the ground truth data. The regression analysis Y on X (observed over predicted) is given by Y = a + b*X, where a and b are y-intercept and slope respectively. The normal equations formed for testing generality of the derived prediction model are given below:

$$\sum Y = N * a + b * \sum X \quad (8)$$

$$\sum XY = a * \sum X + b * \sum X^2 \quad (9)$$

The observed and predicted values of NDVI for *Ricinus communis* showed a weaker agreement (R^2= 0.55, n=40) as shown in Table 3.

Table 3: The values of Y-intercept, Slope and coefficient of determination attributed to both vegetation indices (NDVI and EVI-2) of regression fitted model to the observed and predicted values

Attribute	y-intercept	Slope	R^2
NDVI	2.96	0.70	0.55
EVI-2	1.61	0.65	0.75

The fraction linked with the unexplained variance had a great manipulative power on the sum of squared values of the analytical error (SSPE), a measure of goodness of fit (Table 4). The slope from base line of fitted regression was = 0.70, telling an over-estimation of the comparative fraction of NDVI values.

The observed and predicted values of regression fitted model of the EVI-2, the coefficient of determination and total number of observations was (R^2=0.75, n=40). The slope (0.65) and y-intercept (1.61) calculated do not differ considerably from 1 and 0 respectively. The lack of fit or the contribution to SSPE was largely related to the mean differences of observed and predicted values or termed as bias as shown in Table 4. Out of two cumulative indices analyzed for its accuracy as predictor, the EVI-2 outperformed the NDVI at the time of model generation as well as testing of the model generality for yield estimation.

Table 4: SSPE, U_{bias}, $U_{\beta-1}$, and U_e values for NDVI and EVI-2

Attribute	*SSPE*	U_{bias}	$U_{\beta-1}$	U_e
NDVI	89.39	0.36	0.10	0.54
EVI-2	92.07	0.59	0.11	0.30

Total production of *Ricinus communis* using the yield prediction model

The total production of the *Ricinus communis* crop in Adampur block was calculated using a developed yield prediction model. For the calculation of yield based on the developed model the EVI-2 values were extracted with the help of a zonal statistic value extraction tool. Moreover, the values of EVI-2 were extracted from all the images for December to April using the crop acreage polygons of *Ricinus communis* obtained from ANN-based classification results. In addition, the EVI-2 values were added to find the cumulative EVI-2

values for winter crop season. Finally, the yield per acre of *Ricinus communis* crop was calculated utilizing the developed model.

The minimum yield per acre was found to be 3.45 quintals, the maximum yield per acre was 12.11 quintals and the average or mean yield per acre was found to be 7.17 quintals with a standard deviation of 1.76 as given by the developed yield prediction model. The area of polygons having *Ricinus communis* crop was calculated in acres with the help of tools available in ArcGIS. Finally, the total production was calculated, and it was around 13475.10 quintals in Adampur block.

The crop yield depends on various parameters like rainfall, temperature, humidity, soil moisture availability, etc., during different phases of the crop cycle. The data published by Central Ground Water Board (CGWB), Chandigarh for Ground Water Potential in the Adampur block suggests the status to be critical. The normal annual rainfall in Hisar district is around 330 mm which is unevenly distributed over the area. The rainfall received in the region is scanty and limited mostly within the Monsoon season (June to September). Considering the effect of all the parameters on biomass accumulation and it is quite imperative that very good assessment of biomass can be done indirectly by the EVI-2 derived from satellite imageries. Only EVI-2 was considered for building the yield prediction model for *Ricinus communis*.

There are several challenges in the identifying and mapping of cropland in view of the diversity of cropping systems, management practices and field extents. It could further be explored to improve the accuracy of classification results by integrating multi-source satellite datasets.

DISCUSSION

Different types of satellite based remote sensing images were widely used to provide information in recognizing plant stress conditions, quantifying crop needs, predicting crop yield or the optimum harvesting time. Such images would also be of interest for sparsely grown castor bean crops. The result of this study is promising for utilizing remote sensing technology for identifying and delineating any minor crop like castor bean within a mixed land use pattern. The earlier study conducted for identifying agriculture crop is mostly dedicated to identifying and delineating major crops grown in an area. Although the land use pattern and agriculture crops show variability in the study area, the applied methodology for acreage estimation of minor crop having sparse and dispersed distribution is established with good confidence, not shown in any earlier study. In a recent study conducted by Cavalaris *et al* (2022), nineteen vegetation indexes (VIs) were derived from a multispectral camera mounted on UAS and NDVI presented a higher accuracy in predicting the harvesting time of castor bean. This study used multispectral camera instead of satellite image for the study of castor bean in the farm of University of Thessaly, central Greece.

The time-series of different vegetation indices derived from satellite remote sensing images was used earlier also to delineate crops but never for a minor crop like castor bean that grown sparsely. In the present study, the stacking of time-series EVI-2 images together with ANN classifier outperformed the time-series NDVI images in delineating different crops including castor bean and land use precisely. The time series EVI-2 values have been used to predict the yield of castor bean.

Water Potential in the Adampur block suggests the status to be critical. Although the castor bean is relatively easy to grow. But the crop yield depends on various parameters like rainfall, temperature, humidity, soil moisture availability, etc., during different phases of the crop cycle. The data published by Central Ground Water Board (CGWB), Chandigarh for

Ground The normal annual rainfall in Hisar district is around 330 mm which is unevenly distributed over the area. The rainfall received in the region is scanty and limited mostly within the Monsoon season (June to September). Considering the effect of all the parameters on biomass accumulation and it is quite imperative that very good assessment of biomass can be done indirectly by the EVI-2 derived from satellite imageries. Only EVI-2 was considered for building the yield prediction model for *Ricinus communis*.

There are several challenges in the identifying and mapping of cropland in view of the diversity of cropping systems, management practices and field extents. It could further be explored to improve the accuracy of classification results by integrating multi-source satellite datasets.

CONCLUSION

Mapping of sparse minor crops using remote sensing images is a challenge. To some extent these challenges are overcome with the advent and availability of high resolution multispectral remote sensing dataset. Sentinel- 2A multispectral dataset is found to be very useful in deciphering information on the extent and coverage of a sparsely grown minor crop. The classification and mapping of the minor crop *Ricinus communis* was carried out through various classifiers for the Adampur block of Haryana state, India can establish a method for identifying such sparsely sown crops in mixed land use and cropping pattern. ANN method applied on time-series stacked EVI-2 was found to be the best among all classifiers used in this study for the mapping of minor crop having dispersed distribution.

For the winter season, the regression prediction model based on cumulative values of EVI-2 with a coefficient of determination (r^2=0.75) was observed to be better than that of NDVI with a coefficient of determination (r^2=0.55). The established model showed a robust potential of cumulative vegetation indices for crop yield prediction. The contribution to SSPE for EVI-2 based model was related mainly due to bias or mean difference between observed and predicted values in contrast to NDVI based model. The SSPE was more associated with unexplained variance, resulting in overestimation of associated fraction.

Additional empirical studies should be executed by exploiting Sentinel 2-derived NDVI and EVI-2 for predicting the yield of other minor crops to further establish and improve its prediction ability. This work may be improved if conducted by considering the socio-economic factors, weather conditions, soil properties etc. that affect the crop yield in addition to satellite images. Various data mining and statistical methods can also be tested to analyze several influencing factors.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge HARSAC, Hisar for providing lab facilities to carry out this study. Authors are also thankful to the European Space Agency (ESA) for the free access to Sentinel-2 MSI data.

REFERENCES

Agri Farming, (2023). *Castor cultivation information guide*. Retrieved June 4, 2023, from https://www.agrifarming.in/castor-cultivation-information-guide.

Atkinson, P.M., Tatnall, A.R.L. (1997). Introduction neural networks in remote sensing. *Int. J. Remote Sens*. 18(4):699-709. https://doi.org/10.1080/014311697218700

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., Abelleyra, D., Ferraz, R. P. D., Lebourgeois, V., et al., (2018). Remote Sensing and Cropping Practices: *A Review. Remote Sens.*, 10, 99. https://doi.org/10.3390/rs10010099

Belward, A.S., Skøien, J.O., (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian Earth observation satellites. *ISPRS J. Photogramm. Remote Sens*., 103, 115-128. https://doi.org/10.1016/j.isprsjprs.2014.03.009

Bolton, D.K., Friedl, M.A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* 173, 74-84. https://doi.org/10.1016/j.agrformet.2013.01.007

Cavalaris, C., Latterini, F., Stefanoni W., Karamoutis, C., Luigi Pari, L., Alexopoulou, E., (2022) Monitoring Chemical-Induced Ripening of Castor (*Ricinus communis* L.) by UAS-Based *Remote Sensing. Agriculture*, 12, 159. https://doi.org/10.3390/agriculture 12020159

Chu, C., Hsu, A.L., Chou, K.H., Bandettini, P., Lin, C., ADNI Initiative (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *NeuroImag*e, 60(1):59-70. https://doi.org/10.1016/j.neuroimage.2011.11.066

Cortes, C., Vapnik, V.N. (1995). Support-vector networks. *Mach. Learn*. 20:273-297. https://doi.org/10.1007/BF00994018

Esquerdo, J., Zullo, J., Antunes, J.F.G. (2011). Use of NDVI/AVHRR time-series profiles for soybean crop monitoring in Brazil. Int. *J. Remote Sens*. 32: 3711-3727. https://doi.org/10.1080/01431161003764112

Ghazaryan, G., Dubovyk, O., Löw, F., Lavreniuk, M., Kolotii, A., Schellberg, J., Kussul, N., (2018). A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. *Eur. J. Remote Sens*., 51:1, 511-524. https://doi.org/10.1080 /22797254.2018.1455540.

Huang. J., Wang, X., Li, X., Tian, H., Pan, Z. (2013). Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's-AVHRR. *PLoS ONE* 8(8): e70816. https://doi.org/10.1371/journal.pone.0070816.

Huete, A. R., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.,* 83, 195-213. https://doi.org/10.1016/S0034-4257(02)00096-2

Huete, A., Justice, C., Van Leeuwen, W. (1999). MODIS vegetation index (MOD13), *Algorithm theoretical basis document*, 3, 213.

Jaafar, H. H., Ahmed, F. A. (2015). Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands. *Int. J. Remote Sens*., 36(18), 4570-4589. https://doi.org/10.1080/01431161.2015.1084434.

Jiang, Z., Huete, A.R, Didan, K., Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ*. 112(10), 3833-3845. https://doi.org/10.1016/j.rse.2008.06.006.

Kumar, P, Gupta, D. K., Mishra, V.N., Prasad, R. (2015). Comparison of support vector machine, artificial neural network and spectral angle mapper algorithms for crop classification using LISS IV data. *Int. J. Remote Sens*. 36(6):1604-1617. https://doi.org/10.1080/2150704X.2015.1019015.

Kumar, P., Prasad, R., Choudhary, A., Mishra, V. N., Gupta, D.K. and Srivastava, P.K. (2017). A statistical significance of differences in classification accuracy of crop types using different classification algorithms. *Geocarto Int*., 32(2), 206-224. https://doi.org/10.1080/10106049.2015.1132483.

Lillesand, T.M., Kiefer, R.W., Chipman, J.W. (2008). *Remote sensing and image interpretation*, 6th Edn. Wiley, New York.

Liu, J., Shang, J., Qian, B., Huffman, T., Zhang, Y., Dong, T. Jing, Q., Martin, T. (2019). Crop Yield Estimation Using Time-Series MODIS Data and the Effects of Cropland Masks in Ontario, Canada. *Remote Sens*. 11(20), 2419. https://doi.org/10.3390/rs11202419.

Lu, D., Mausel, P., Batistella, M., Moran, E. (2004). Comparison of landcover classification methods in the Brazilian Amazon basin. *Photogramm Eng Remote Sens* 70(6):723-731. https://doi.org/10.14358/PERS.70.6.723.

Lu, D., Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens*., 28(5):823-870. https://doi.org/10.1080/01431160600746456.

Martins, V. S., Barbosa, C. C. F., De Carvalho, L. A. S., Jorge, D. S. F., Lobo, F. D. L., and Novo, E. M. L. de Mores. (2017). Assessment of atmospheric correction methods for Sentinel-2A MSI images applied to Amazon floodplain lakes. *Remote Sens*., 9 (4), 322. https://doi.org/10.3390/rs9040322.

Mas J.F., Flores, J.J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *Int. J. Remote Sens*. 29:617-663. https://doi.org/10.1080/01431160701352154.

Mishra, V.N., Prasad, R., Kumar, P., Gupta, D.K., Srivastava, P.K (2017b). Dual polarimetric C-band SAR data for land use/land cover classification by incorporating textural information. Environ. *Earth Sci*. 76(1):26. https://doi.org/10.1007/s12665-016-6341-7.

Mishra, V.N., Prasad, R., Kumar, P., Srivastava, P.K., Rai, P. K., (2017a). Knowledge-based decision tree approach for mapping spatial distribution of rice crop using C-band synthetic aperture radar-derived information. *J. Appl. Remote Sens*. 11(4), 046003. https://doi.org/10.1117/1.JRS.11.046003.

Mkhabela, M.S., Bullock, P., Raj, S., Wang, S., Yang, Y. (2011). Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. For. Meteorol.* 151, 385-393. https://doi.org/10.1016/j.agrformet.2010.11.012.

Pal, M., Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. IEEE *Trans. Geosci. Remote Sens.* 48, 2297-2307. https://doi.org/10.1109/TGRS.2009.2039484.

Pareulo, J. M., Jobbagy, E. G., Sala, O. E., Lauenroth, W. K. and Burke, I. C. (1998). Functional and Structural convergence of temperate grassland and shrubland ecosystems. *Ecol Appl*, 8 (1), 194-206. https://doi.org/10.1890/1051-0761(1998)008[0194:FASCOT]2.0.CO;2.

Pijanowski, B.C., Brown, D.G., Shellito, B.A., Manik, G.A. (2002). Using neural networks and GIS to forecast land use changes: a land transformation model. *Comput Environ Urban*

Syst 26:553-575. https://doi.org/10.1016/S0198-9715(01)00015-1.

Polsot, A.-S., Speedy, A., Kueneman, E., (2004). *Good Agricultural Practices-A Working Concept*; Background Paper for the FAO Internal Workshop on Good Agricultural Practices; FAO: Rome, Italy, p. 41.

Puissant, A., Rougiera, S., Andre´, S. (2014). Object-oriented mapping of urban trees using random forest classifiers. *Int J Appl Earth Obs Geoinf* 26:235–245. http://dx.doi.org/10.1016/j.jag.2013.07.002.

Richards, J.A, Jia, X. (2006). *Remote sensing digital image analysis*, 4th edn. Springer, Heidelberg.

Richards, J.A. (1993). *Remote Sensing Digital Image Analysis*, Springer.

Roujean, J. L., Breon, F.M. (1995). Estimating par absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ*., 51, 375-384. https://doi.org/10.1016/0034-4257(94)00114-3.

Shammi, S. A., Meng, Q. (2021). Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. *Ecol Indic*, 121, 107124, https://doi.org/10.1016/j.ecolind.2020.107124.

Shiu, Y.-S., Chuang, Y.-C. (2019). Yield Estimation of paddy rice based on satellite imagery: comparison of global and local regression models. *Remote Sens*. 11, 111. https://doi.org/10.3390/rs11020111.

Skakun, S., Franch, B., Vermote, E., Roger, J-C., Becker-Reshef, I., Justice, C., Kussul, N., (2017). Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. *Remote Sens. Environ*., 195, 244-258. https://doi.org/10.1016/j.rse.2017.04.026.

Srivastava, P.K., Han, D., Rico-Ramirez, M.A., Bray, M., Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Adv. Space Res*. 50:1250-1265. https://doi.org/10.1016/j.asr.2012.06.032.

Sun, H., Xu, A., Lin, H., Zhang, L., Mei, Y. (2012). Winter wheat mapping using temporal signatures of MODIS vegetation index data. *Int. J. Remote Sens*., 33(16), 5026-5042. https://doi.org/10.1080/01431161.2012.657366.

Vapnik, V.N. (1998). *Statistical learning theory*. Wiley, New York.

Weier, J. and Herring, D. (2000). *Measuring Vegetation (NDVI & EVI)*. NASA Earth Observatory, Washington DC.