2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

A Comparative Study and Machine Learning Enabled Efficient Classification for Multispectral Data in Agriculture

Priyanka Gupta *[1], ***Shruti Kanga***[2], ***Varun Narayan Mishra***[3], ***Suraj Kumar Singh***[4], ***Thota Sivasankar***[5],

[1]School of Engineering & Technology, Suresh Gyan Vihar University, Jaipur, Rajasthan, India.
[2]Department of Geography, School of Environment and Earth Sciences, Central University of Punjab, Bathinda, Punjab, India.
[3]Amity Institute of Geo informatics and Remote Sensing (AIGIRS), Amity University, Noida India.
[4]Department of Centre for Sustainable Development, Suresh Gyan Vihar University, Jaipur , India.
[5]Department of GIS, NIIT University, Neemrana, Rajasthan, India.
*Corresponding Author.

Received 17/04/2023, Revised 15/09/2023, Accepted 17/09/2023, Published Online First 25/12/2023, Published 1/7/2024

Abstract

Reliable and accurate crop maps are required for food security from regional to global scale. The increased availability of satellite imagery leads to a “Big Data” problem while producing crop maps. Now, cloud-based platforms have gained a lot of attention for crop classification over large regions. The main goal of the research is to analyze crop classification using various machine learning (ML) such as Support Vector Machine (SVM), Gradient Tree Boosting (GTB), Random Forest (RF), Decision Tree (DT) as well as Classification and Regression Trees (CART) on Google Earth Engine platform. The aim is to explore the Google Earth Engine’s efficiency (GEE) when classification different crops using multi- spectral datasets of Sentinel 2 MSI and Landsat 8 OLI satellites for crop mapping of Mathura district of Uttar Pradesh, India. The best cloud free image (less than 5%) of Landsat 8 OLI and Sentinel 2 MSI datasets ("2020-12-26","2020-12-30") were used for crop classification with the help of automatic filtering i.e. percentage cloud property on the GEE platforms. Moreover that GEE platform perform, acquiring, clarifying as well as preprocessing of satellite dataset could be organized very powerfully. Points as feature spaces were used like training datasets. Furthermore confusion matrixes are used for accuracy assessment (producer and user accuracy) and kappa coefficient. Additionally compare the outcome of the dataset on the basis of overall accuracy (OA), F1 score as well as kappa coefficient. The highest OA is found using GTB (86.7%) followed by RF (82.5%), CART (81.0%), DT (78.1%) and SVM (66.5%) for Landsat 8 OLI image. For the Sentinel 2 image, GTB achieved the highest OA of 84.2% followed by SVM (84%), RF (82.3%), DT (75.2%), and CART (75. 0%) respectively. On the basis of research, found that GTB performed well among all the classifiers to crop mapping using both multi-spectral datasets.

Keywords: Google Earth Engine (GEE), Machine learning (ML), Remote Sensing (RS), Satellite imagery.

Introduction

Information on the spatial disposition of land cover and land use over huge zones are tremendously significant for several ecological observing tasks, including ecosystem dynamics, climate change,

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

management of natural resources, food security, and others[1,2]. Classification of crops is vital to know the climatic requirements and physiological of different crops. The crop maps are very beneficial for the assessment of diversity, food security, sustainable development and management of agricultural fields[3]. Reliable and accurate crop maps with high accuracy can be used for the improved estimation of agriculture statistics for better crop yield[2,4,5] and dryness related danger[6]. Reliable crop mapping is needed for accurate agriculture estimation and better cultivation practices. Several factors including soil condition, contamination of ground, reduction in water resources, and emission of greenhouse gases also affect crop productivity[3]. Traditional ground survey methods and statistical approaches used in crop related studies are tedious, time consuming and expensive. So, it is required to adopt a fast, reliable and automated system for mapping and monitoring croplands at different observational scales[6].

In the last few decades, remotely sensed images became the most capable data sources for the classification of crop types, monitoring the growth of crops as well, and acreage estimation [7-9] .Currently, a wide range of datasets from various satellite sensors are freely available on a regular basis. At the same time, the availability and accessibility of datasets assist a varied group of researchers in expansion insights into complex landscape processes over large areas[10]. In particular, analyses in the arena of agriculture have profited from the increased accessibility of remote sensing imageries. In the view of data science, it is required to use a new approach in extracting relevant information from RS datasets[11]. However, the computational availability of resources required to process and analyze “Big Data” is a considerable hindrance in many studies. As a consequence, several cloud-based computing platforms have been developed[12].

GEE is one of the most commonly used and freely available platforms to process and analyze remotely sensed datasets. It allows users to exploit the increasing library of multi-spectral as well as multi-temporal datasets, by immensely decreasing the time required to download as well as process imagery [10, 13]. GEE with minimal human interaction and interference, performs well in terms of less consuming time and processing complexities[1]. In addition to the fast processing, the accessibility of various package with large numbers of methods simplify access to RS tools making GEE increasingly popular among users. In many studies the applications of GEE are reported for various domains including agriculture, forestry, ecology, LULC studies [13- 15]. The GEE offers a set of advanced supervised machine learning (ML) classifiers as well that can be utilized for crop type mapping. These classifiers are preferred frequently because of the option of setting a predefined number of considered output classes. In many studies, classifiers are well-documented and used for classifying crop types using remotely sensed imageries [7, 16.]

Key challenges for this research:

- Facing a lot of challenges while capturing images when the cloudy climate is above 10% then the images not considered.
- Only below 10% of the cloudy climate is considerable.
- Techniques utilized in crop classification are time-consuming, costly and abstruse.

Main objectives of this research:

- The aim of this study is to assess feasibility of GEE to study crop classification in the Mathura, by determining the extent of crop mapping using various ML methods such as CART, RF, DT and SVM in the GEE platform.
- Additionally is to compare multiple classifiers (RF, SVM, CART, GTB) using GEE platform when classifying with different Sensors (Landsat 8, Sentinel 2) for crop mapping.
- For food security authentic and perfect crop classification is needed from territorial to universal.
- •The production of crop maps is "Big Data" dilemma because of the increased accessibility of satellite imagery.
- Cloud computing platforms have attracted a lot in interest for crop classification over vast areas.

Moreover, to explore crop classification using various ML methods like SVM, RF, DT, CART, as well as GTB on GEE platform. For this purpose Mathura as study area has been selected using RS dataset (Sentinel 2 MSI as well as Landsat 8 OLI satellite) as well as ground truth dataset gathered with the help of a smartphone. Hence, it aims to compare multiple classifiers (RF, SVM, CART,

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

GTB) using GEE platform with multiple Sensors (Landsat 8, Sentinel 2) for crop classification.

The major contributions of this research

- It is to investigate the effectiveness of the GEE for classifying various crops using multi-spectral datasets from the Sentinel 2 MSI as well as Landsat-8 OLI satellite.
- To produce a map of crops for the Mathura district of Uttar Pradesh, India, several supervised ML classifiers, including CART, RF, GTB, DT and SVM are tested.
- For the Landsat 8 OLI image, GTB (86.7%) had the highest overall classification accuracy, followed by RF (82.5%), CART (81.0%), DT (78.1%), and SVM (66.5%).
- GTB achieved the highest overall classification accuracy for the Sentinel 2 image at 84.2%, followed by SVM (84%), RF (82.3%), DT (75.2%), and CART (75.0%), in that order.
- Using both multi-spectral datasets, it was discovered that GTB outperformed all other classifiers in mapping agricultural crops.
- It significantly to the evaluation of the use of a cloud based platform for crop classification.

The rest work is arranged in the following Section. In Section 2, the related literature has been reviewed. In Section 3 the proposed methodology is discussed and implemented in the different crop classifications, Section 4 highlights the result and discussion as well as a comparison that took place to justify results. In the last Section 5 research is concluded with the overall summary of accomplishments provided.

Related Works

Agriculture is the backbone of India around 61.5% of the population depends on agriculture in rural areas. Worldwide for crop productivity, India has second ranked. The classification of crops is very important for agriculture policy making, food security, farmer's income and sustainable agriculture development. Many environmental monitoring activities, including analyzing ecosystem dynamics, food security, climate change, and others, primarily rely on data on the spatial distribution of land cover or land use (LCLU) over large areas. High-accuracy crop maps can be utilized for stratification of agricultural statistics estimation, crop production forecasting, and drought risk assessment[17-19]. For accurate agriculture estimation reliable crop mapping is needed and better cultivation detection. For solving these types of tasks as crop mapping Satellite imagery has become the main data source over the last few years[20]. Many satellites are available which are open access and free with high spatial resolution. Building high quality crop maps on a regular basis is encouraged by new possibilities. Due to the growing usage of remote sensing data from a data science perspective, a big data problem generates new tasks for managing data that require new ways of removing critical information from RS. For high resolution crop maps for a specific area, processing a large number of satellite pictures taken with different sensors is required. Images taken over the same time period allow for comparison of the output from various sensors (Landsat 8 and Sentinel 2) [21, 22].

Crop classification is crucial for the development of sustainable agriculture, farmer income, food security, and agricultural policy. Crop productivity is also influenced by soil quality, ground contamination, shrinking water bodies, and greenhouse gas emissions[23]. Ground survey techniques and statistics were traditionally employed. These were expensive, time-consuming, and labor-intensive. Therefore, a quick, trustworthy, and automated method that uses satellite photos to offer precise crop mapping is needed [24]. Integrating new technology like deep learning, machine learning and Internet of things can be used for crop mapping. It will help to improve accuracy of crop mapping, classification and identification [25]. The creation and choice of classification techniques and characteristics retrieved from satellite data are critical to achieve crop mapping products from remote sensing [26]. The maximum likelihood classifier, classification trees (such as the classification and regression tree), and machine learning classifiers (such as random forest and support vector machines) were the most widely used classification algorithms for single, multi-temporal, and multi-source satellite imagery[27].

GEE has performed quickly and effectively in terms of time and processing complexity with the least interference from and interaction with humans. GEE provide cloud platform to access multiple datasets (Landsat 8, Sentinel 2) free available and it provides many pixel -based classifiers (multiple ML methods) that can be used for crop mapping [28]. For attaining the goals of evaluating the classification of crops using several ML algorithms with numerous datasets (Landsat 8, Sentinel 2) of study region Mathura (Uttar Pradesh), GEE has shown to be reliable. Hence, to compare multiple classifiers (RF, SVM, CART, GTB) using GEE platform when classifying with different Sensors (Landsat 8, Sentinel 2) for crop mapping [29-31]. The author applied a variety of techniques on the GEE platform and compared the results using ML techniques with the aid of OA, the kappa coefficient and the F1 score[32]. The comparative study is described in Table 1.

Table 1. Comparative study

Ref. No	Algorithm	Working ground	Expediency	Limitations	Sensors	Classified Crops	Comparison with proposed work
21	Random Forest	It is a supervised learning that works on bagging concept. A number of module are trained various subset of dataset and final outcome is produced by combining the entire module.	It is accomplished of regression as well as classification. A RF produces accurate predictions that are simple to comprehend. Huge datasets can be handled efficiently. In comparison to the DT method, the RF offers a higher level of accuracy in estimation.	To create maps of specific agricultural fields using the probabilistic result after post-processing.	Digital globe world view2	Off season, growing season	RF may not be as correct as GBT. They can distinguish intricate forms in data because train to accurate each fault. The GBT, however, might overfit if the dataset is noisy.
22	Gaussian mixer model	It is probabilistic method suggests that all data points were formed by combining a limited number of Gaussian scatterings with unknown parameters.	They give evaluations of likelihood that every data point is a member of each cluster. Compared to the single group assignment that the majority of other clustering method offer and lot more related data.	It is necessary to maintain the reliability of the crop yield forecasting model.	MODIS	Winter crops	Compared multiple methods.
25	Method B, C, D, B2, D2	It considers the class of restored pixels and lessens its influence on the outcome. Non-restored pixels received a weight, and each restored pixel received a weight equal to the square of the	Each element that comes close to meeting the requirements of the permitted minimum area dimension, compactness, and class correctness was added to the previously classified parcels.	Does not account for the possibility that a single parcel has many plots. When optical data are present, classification is useful (D1 and D2).	Landsat 8 OLI ad Sentinel 1	Parcel and pixel based	Compared two dataset one is Sentinel 2 MSI other is Landsat 8 OLI.

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

		proportion of clear images to all other images in the time series.					
28	SVM, DT, RF, Naive Bayes and NN	SVM categorizes data points even when they are not otherwise linearly separable by mapping the data to a high-dimensional feature space.	The optimization issue is converted into two dual convex quadratic programs, avoiding the challenges of employing linear functions in the high-dimensional feature space.	When invoked from Python, it typically produced unstable classification results for SVM classifiers and instead returned an Internal Server Error.	Landsat 8 TOA	8	Multiple methods are applied, GTB gives better outcome.
32	SVM, MXL, RF, CART	It is a predictive model that describes how the values of an outcome variable can be anticipated based on other values. A CART output is a DT, where each end node has a prediction for the outcome variable and each fork indicates a split in a predictor variable.	Suitable for both categorical and continuous responses. Manage several missing values and extreme outliers with ease	Single Sensors, In case of vegetables MXL were not performed well	Sentinel 2	Wheat, Mustard	It is easy to comprehend and interpret. little data preparation is necessary. The cost of using the tree increases logarithmically with the volume of data used to train it. It can manage both categorical and numerical data. It can manage issues with several outputs.
33	DCNN (padding technique)	Convolutional neural networks (CNNs) can benefit from padding since it describes the number of pixels that are added to an image during processing by the CNN kernel.	Padding can enhance the performance of the model by minimizing information loss at the input feature map's borders.	Many Arabic have similar sound un efficient to solve them	Voice data of school	Voice data	GEE platform is used which easy to use for remote sensing dataset.
34	MLP, RF,BYO	There are an endless number of hidden layers between the output and input layers that make up the directly connected	It can be applied to resolve challenging nonlinear issues. It effectively manages vast volumes of input data. Following	Worst result using logistic activation function in MLP	LIDAR, IKONOS	Urban area (road, un road)	Comparison with SVM, GBT, DT, RF and CART are conducting on GEE platform.

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

		mechanism of MLP. In MLP, data travels in a forward direction from the input to the output layer, similar to a network operating in feed-forward mode. All of the MLP's nodes are trained with back propagation.	training, quickly makes predictions. Even with less samples, the same accuracy ratio is still possible.				
35	SVM, MNB	The MNB method relies on the Bayes theorem and makes the assumption that, given the class variable, the features are conditionally independent.	MNB takes a features vector this term act as a frequency or it seems number of times.	More feature engineering is need for gained higher accuracy	News sources (twitter)	Textual propeganda	Six classes are used as urban, water, vegetation, mustard, wheat and othercrops
36	KNN, NB, DT, SVM	Regression and classification issues are addressed by it. In order to forecast the class or value of a new data point, it locates the K nearest points in the training dataset and uses their class to do so.	It is simple to comprehend. Regression as well as classification matters can be solved. It is good for nonlinear dataset.	The performance of SVM is reduced due to noise.	NSL_KDD	Attack	Category data are used having six classes

The idea of Table 1. Comparative study is taken from [37].

Materials and Methods

This study has been carried out at Mathura city of Uttar Pradesh (UP) in India and situated in the northern part of India shown in Fig. 1, around 55 km north of Agra and 145 km southeast of Delhi. The latitude and longitude of Mathura are 27.4924° N and 77.6737°E. The administrative center of Mathura district and covers an aerial extent of about 1482 square km and has an average elevation of 174 m above the mean sea level.

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

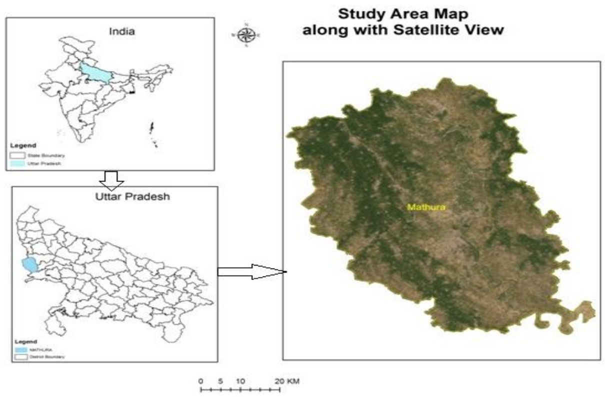

Figure 1. Study area Map along with Satellite View

According to the 2011 census, Mathura has a total population of 349,909 people and a literacy rate of 74.97%. In Mathura, the overall area is 3.32 mha, the total area under cultivation is 3.28 mha, and the total area under irrigation is 3.11 mha. A lot of land is used for crop farming in Mathura. Crops are sown during the Rabi and Kharif seasons. Typically, two important crops—wheat and mustard—are planted during the winter (Rabi) season. For large-scale crop mapping, some research have compared the effectiveness of ML algorithms for multi-sensor categorization optical imaging. The goal of this work is to assess a few supervised machine learning classifiers for crop categorization that are available online through the Java Script API of the GEE cloud computing platform. Its purpose is to investigate how well the GEE medium performs while classifying multi-sensor optical pictures from Sentinel-2 and Landsat-8 for crop mapping with the potential to use it at a larger scale (like the district level). Its work offers substantial information regarding the presence of crop type in agriculturally dominated areas. Such data is dynamic for the successful monitoring and management of crop diversity with its productivity at different scales.

Dataset

The dataset is categorized into two parts. Firstly, multi-spectral remote sensing images acquired from sentinel 2-MSI as well as Landsat 8 OLI sensors are used for comparative analysis of crop classification. Images downloaded from USGS Landsat 8 OLI collection one tier one calibrated top of atmosphere (TOA) reflectance and Sentinel 2MSI from https://scihub.copernicus.eu. A dataset is good because it contains less than 5% cloud cover. The Landsat 8-OLI image has spectral bands (band 1 to 7 as well as band 9) at 30 m spatial resolution as well as panchromatic band 8 at 15 m resolution. In this study, bands are used from band 2 to band 7.

13 spectral bands make up the Sentinel 2 MSI, including 4 at 10 m, 6 at 20 m, and 3 at 60 m spatial resolution. Sentinel scientific data hub provided a multispectral Level-2A (L2A) dataset from Sentinel 2. The reflectance of Sentinel 2's L2A dataset is the bottom of the atmosphere (BOA) with atmospheric correction as well as radiation correction. Images from Sentinel-2 L2A with less than 5% cloud coverage percentage and classified Sentinel-2 imagery using nine spectral bands: green, blue, red, SWIR-2 bands, shortwave infrared band 1 (SWIR-1), near infrared (NIR), red-edge band 1 (RE-1), RE-3 as well as RE-2. In this research B3-Red, B4-Green and B8-NIR bands are used. Sentinel 2 MSI can offer numerous spectral temporal characteristics for crop mapping. Its 5 day average revisits length, high spatial resolutions and 3 red edge bands[37-41]. For crop classification divided total area into six classes such as urban, vegetation, water, wheat, mustard, and other crops. Further multispectral imageries, a wide field review as conducted across the study area to gather ground datasets of various crops using a handheld GPS receiver. Finally, six classes including 2 major crops (wheat and mustard), urban, water bodies, vegetation, and other crops were shown based on field surveys as well as visual image interpretation of area.

Secondly, the ground dataset samples are required to assure crop classification accuracy. In December 2020, conducted a field survey in the Mathura region using a hand-held Global positioning System with a position accuracy of 2 meters where Crop field samples were taken throughout the field survey and determined the boundaries of area of interest using high spatial quality images from Google Earth. At the plot level, crop samples were separated into two sections (80% - 20% for training as well as testing respectively) to ensure that testing and training pixels are in different fields. Fig.2 shows ground dataset of crops (wheat and mustard). Images have taken from GPS devices of mobile phones. The images are taken in winter seasons (Rabi).

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Figure 2. Ground dataset of Wheat and Mustard crops

Google Earth Engine (GEE)

It is a state-of-the-art cloud platform for geospatial and remote sensing data processing. . Petabyte-scale archives of additional data (various composite products) and publicly accessible RS imagery (Sentinel-1 or 2, Landsat-8 or MODIS) are available in GEE. GEE has access to computational substructure of Google for handling geospatial datasets in parallel, python and java script APIs for analysis visualization and an online IDE for rapid spatial analysis visualization and prototyping using the java script API [code editor] [https://developers.google.com/earth-engine/]. The GEE platform offers a variety of cutting edge ML methods for, Naive Bayes, RF, DT, GTB, CART, and SVM[32].

Classification Methods

Support Vector Machines (SVM): It is the most standard supervised ML method which is commonly used for both regression as well as classification. The algorithm was developed based on non-parametric statistical learning frameworks by[42]. SVM to understand complex relationships and provide better accuracy with small training data sets[43- 45]. It uses the margins to solve classification problems. Margin is described as the shortest distance between any samples and decision boundary. Margin is maximized hence decision boundary is selected to maximize margin. Margin is perpendicular distance between decision border and data point that is closest to it. When the margin is maximized, a specific decision boundary is chosen. A subgroup of the data sets, called support vectors, defines the placement of this boundary. The employed kernel type is RBF (radial bias function), cost set to 1 and gamma set to 0.0002. Margin expansion results in specific alternatives at the decision boundary. The SVM classifier's output for classification is shown in Fig. 5 for Landsat 8OLI and Fig. 10 for Sentinel 2 MSI.

Classification and Regression Trees (CART)

It is popular non-parametric supervised ML method, proposed by Breiman L[46]. It develops optimal decision trees to provide higher accuracy and precision with smaller trees. The algorithm works based on Gini index metric derived from training data to choose the order of nodes to split into sub-nodes. The Gini index values range from 0 to 1 which determines the degree of unevenness of the sample data. A higher Gini index value indicates the greater unevenness of the sample data. It has been extensively used in RS studies for classification and regression analysis due to its ease to interpret, understand, and visualize[47-49]. It utilized 202 leaves, maximum and feature collection and training, CART classification was performed with 10 training points at depth and 2057 utilizing training points from the data in GEE, a decision tree was generated. As illustrated in Fig. 6 for Landsat 8OLI and Fig. 11 for Sentinel 2 MSI. Classified images are created for six classifications, including urban, water, vegetation, wheat, mustard, and other crops.

Random Forest (RF)

It is an ensemble method, uses a multitude of CART during training for classification and regression analysis[50]. It is developed to minimize the over fitting issues in CART algorithm using bagging technique. The bagging technique draws a subclass of training samples through replacement to generate various DT. In RF algorithm, the robustness, accuracy and precision depend on the user definite factors i.e., number of trees as well as number of samples for training and validation purposes. Its output is defined based on the majority of votes from forest trees[51]. The better accuracy and skill to handle high dataset dimensionality as well as multi-colinearity of RF classifier has made it popular in remote sensing community for various applications including classification [52, 53]. From the training sample, create a 10-tree RF classifier. The

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

output of the RF classifier is displayed in Fig. 7 for Landsat 8 OLI and Fig. 12 for Sentinel 2 MSI.

Gradient Boosting (GB)

GB is a supervised machine learning technique which works on sequentially optimizing to develop a strong model from weak models for better overall performance. The algorithm improves the performance of model by minimizing the cost function i.e., Log loss based on probabilities when it is used as a classifier. The Log loss represents the nearest prediction probability to equivalent actual class. GB has better accuracy and handles missing data making it unique from other machine learning algorithms. Because of these, the utilization of GB for remote sensing applications including classification has been described in detail in the literature [54-56]. However, GB is sensitive to outliers while continuing to minimize errors in the model, resulting in over fitting. Moreover, it is computationally costly in turn takes time and memory for training. The classification output of the GB classifier is displayed in Fig. 8 for Landsat 8OLI and Fig. 13 for Sentinel 2 MSI.

Decision Tree (DT)

DT is a popular non parametric method in RS that is used to solve both (classification as well as regression) problems. A DT is a technique for approximating a discrete value's performance that is resistant to noisy input. A root of tree (which contains all data), internal nodes and multiple leaves make up tree. Till the leaf node reaches, each node makes a binary choice to segregate the distinct groups. Pruning is done with a confidence factor of 0.3 and a minimum of two times per leaf [57]. The classification output of the Decision Tree classifier is displayed in Fig.9 for Landsat 8OLI and Fig. 14 for Sentinel 2 MSI.

Proposed Crop Classified Paradigm

With the increasing demands, Geographical Information Systems (GIS) and Remote sensing platforms have gained a lot of attention to solve real-time hazards for the following areas in recent years, such as Agriculture, land-cover mapping, weather forecasting, natural hazards study, resource exploration as well as environmental study etc. According to survey by the Indian Council of Agriculture and Research have observed that cultivated land is more than 86% in Mathura, UP, India [58]. Due to lack of resources the production ratio is affected as compared to agricultural land. Several machine learning approaches used for crop classification that aims to explore the Google Earth Engine’s efficiency (GEE) when categorizing different crops using multi-spectral datasets of Sentinel-2 MSI as well as Landsat 8 OLI satellites. The proposed methodology trained dataset using different methods such as SVM, RF, DT, GTB and CART. Moreover, better accuracy as compared to previous research using above same techniques. Most importantly, emphasizes the numerous advantages of Landsat and sentinel imagery for various classifiers. GTB classification in both imagery (Landsat and Sentinel) offers the highest classification accuracy than other classifiers. Flow charts are shown in Fig. 4.

The steps in this research's approach were as follows:

- Use the GEE Platform to get the Sentinel-2 MSI as well as Landsat 8 OLI dataset by filtering the area of interest such as a study area.
- Clipping an image to t intended research area (using latitude and longitude) and filtering the image by choosing the more cloud free day obtainable between December 2020.
- The crop characteristics are identified using True Color and False Color computed images.
- Feature collections were created by choosing training points based on ground dataset that was gathered using an android application running on a smartphone.
- All six classes—urban, water, vegetation, wheat, mustard and other crops—were chosen.
- On the basis of the feature collection collected as depicted in Fig. 3, training data sets (80%) were produced.
- Several ML methods, including RF, CART, DT, SVM as well as GTB are employed to categorize the image using GEE after the production of training datasets.

  In addition, Landsat 8 OLI as well as Sentinel-2 MSI dataset (December 2020) are compared, based on OA and kappa coefficient.

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

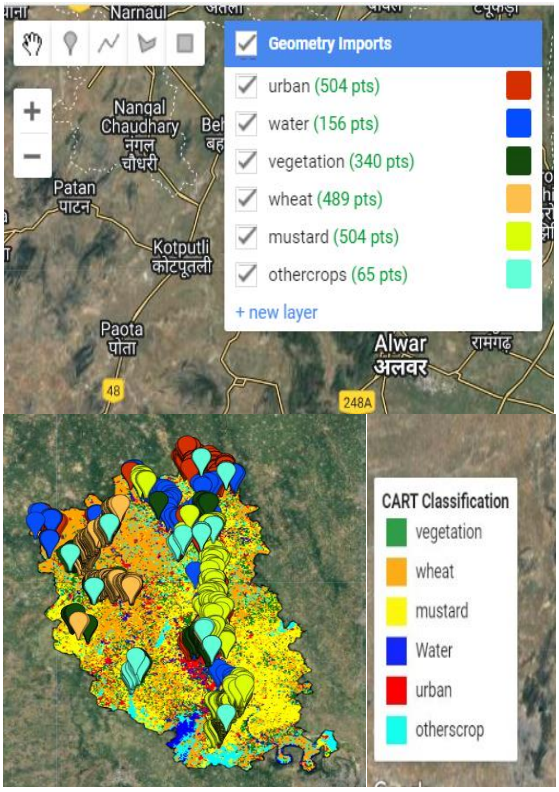

Figure 3. Overlaid feature space on FCC

Table 2. Crop Classes

Data	Description
Urban	Residential area, Industrial, commercial land services, built-up and utilities
Water	Canals, reservoirs, river, and pond
Vegetation	Trees
Wheat	Wheat crops
Mustard	Mustard crops
Other crops	Rice, Barley(jao), Maize(makka), chickpea(chana), Green peas(matar), pink lentils (masoor daal), pigeon pea (Arhar daal)

Table 2 shows types of classes. Urban shows built up areas, residential, industrial as well as commercial area. Water class shows canals, rivers, ponds as well as reservoirs. Vegetation class shows trees. Wheat class shows wheat crops, Mustard class shows mustard crops land. One class shows all types of crops which are not mustard and wheat comes in other classes like barley, maize, chickpea pea etc.

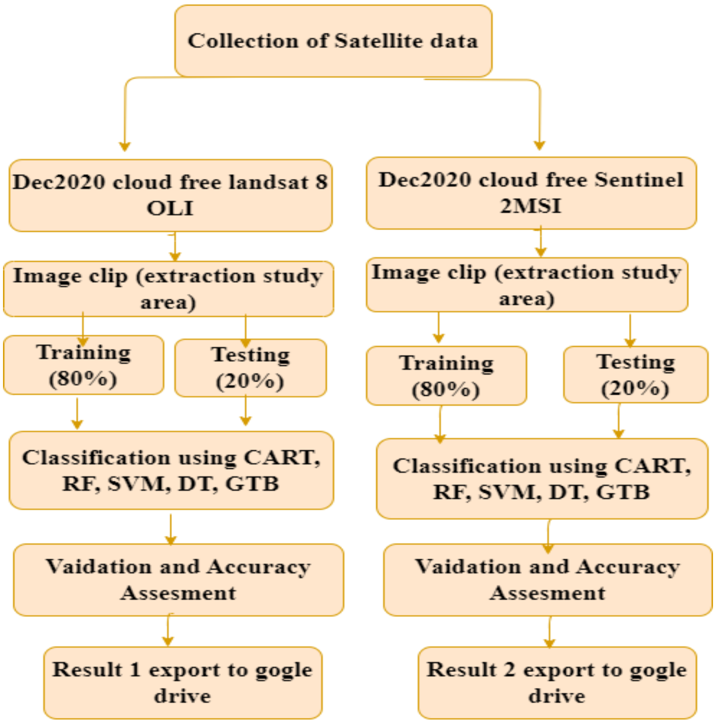

Figure 4. Crop Classification using GEE Method

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Results

Accuracy Assessment:

Analysis of different types of accuracy using different methods such as GTB, RF, CART, DT and SVM on Landsat and Sentinel 2 MSI dataset. For GTB, DT, CART, Random Forest and SVM classification, accuracy has been calculated on a user, producer and overall basis. Along with other courses like Urban and water, vegetation and Other Crops, accuracy assessment has taken major subjects like Wheat and Mustard. Overall accuracy for SVM was found to be 66.58%, while CART, RF, DT and GTB employing machine learning techniques were found to be 81.41%, 82.53%, 78.09% and 86.79% respectively (Table 8). Additionally, the F parameter was calculated using several classifiers. Additionally, GTB outperformed DT, RF, CART and SVM classifiers. Accuracy is calculated using eq. 2. Table 3 to Table 7 shows confusion matrix of all methods. Table 8 shows the overall accuracy[59] and kappa of Landsat 8 OLI.

$$\text{Accuracy} = \frac{(\text{Tpp+ Tnn})}{(\text{Tpp+ Tnn+Fpp+Fnn})} \qquad 1$$

Where Tpp represents True positive, Tnn represents True negative, Fpp represents False positive and Fnn represents False negative.

$$\text{Overall accuracy} = \frac{\text{sum of diagonal element}}{\text{Total number of samples}} * 100 \qquad 2$$

A statistical test to evaluate an accuracy of classification yields the kappa. Kappa essentially evaluates whether the categorization outdone simply randomly assigning values, whether it achieved enhanced than random.

$$\text{Kappa} = \frac{(\text{observed correct} - \text{expected correct})}{(1 - \text{expected correct})} \qquad 3$$

Where observe correct, represents accuracy reported in overall accuracy and expected correct represents correct classification (egyankosh.ac.in/bitstream/123456789/39544/1/Unit-14.pdf).

Table 3. Confusion matrix of GTB method for Landsat 8 OLI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy(%)
Urban	91	3	4	0	1	0	99	91.91
Water	2	25	3	0	0	0	30	83.33
Vegetation	1	0	51	4	2	1	59	86.44
Wheat	3	0	9	84	5	0	101	83.16
Mustard	2	0	2	3	99	0	106	93.39
Other crops	0	0	3	4	2	5	14	35.71
Column Total	99	28	72	95	109	6	409	-
Producer 'accuracy (%)	91.91	89.28	70.83	88.42	90.82	83.33	OA	86.79

Table 4. Confusion matrix of RF method for Landsat 8 OLI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	99	0	1	0	0	0	100	99
Water	2	32	4	1	0	0	39	82.05
Vegetation	4	2	60	17	4	2	89	67.41
Wheat	5	0	12	78	2	1	98	79.59
Mustard	1	0	1	6	93	0	101	92.07
Other crops	0	0	6	3	3	2	14	14.285
Column Total	111	34	84	105	102	5	441	-
Producer 'accuracy (%)	89.18	94.11	71.42	74.28	91.17	40	OA	82.53

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 5. Confusion matrix of CART method for Landsat 8 OLI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	83	2	6	7	1	0	99	83.83
Water	1	28	1	0	0	0	30	93.33
Vegetation	1	4	44	6	2	2	59	74.576
Wheat	5	1	13	78	2	2	101	77.22
Mustard	3	1	1	5	94	2	106	88.67
Other crops	1	1	1	3	2	6	14	42.85
Column Total	94	37	66	99	101	12	409	-
Producer 'accuracy (%)	88.29	75.67	66.66	78.78	93.06	50	OA	81.41

Table 6. Confusion matrix of DT method for Landsat 8 OLI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	83	0	6	5	3	0	97	85.56
Water	4	25	6	0	0	2	37	67.56
Vegetation	2	0	45	14	2	0	63	71.42
Wheat	5	0	8	74	7	10	104	71.15
Mustard	2	0	7	1	97	2	109	88.99
Other crops	1	0	2	2	1	4	10	40
Column Total	97	25	74	96	110	18	420	-
Producer 'accuracy (%)	85.56	100	60.81	77.08	88.18	22.22	OA	78.09

Table 7. Confusion matrix of SVM method for Landsat 8 OLI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	92	0	0	6	1	0	99	92.92
Water	4	22	0	4	0	0	30	73.33
Vegetation	3	0	20	35	1	0	59	33.89
Wheat	13	0	0	81	7	0	101	80.19
Mustard	4	0	4	18	80	0	106	75.47
Other crops	2	0	0	11	1	4	18	22.22
Column Total	118	22	24	155	90	10	413	-
Producer 'accuracy (%)	77.96	100	83.33	52.25	88.88	40	OA	66.58

All above table from Table 3 to Table 7 shows confusion matrix of all methods (GTB, RF, CART, DT, SVM). Diagonal values shows correct prediction (correct classified) while other than diagonal specific to class shows misclassified (wrong prediction).

Table 8. Overall Accuracy of Landsat 8 OLI

S.NO	Classifier	OA %	Kappa
1	GTB	86.79	0.83
2	RF	82.53	0.77
3	CART	81.41	0.76
4	DT	78.09	0.72
5	SVM	66.58.	0.61

In this section, shows an analysis of multispectral datasets like Sentinel 2 MSI as well as Landsat 8 OLI. For experimentation purposes, GEE is used to implement. It runs on an Intel Core i3 7th generation processor held 8 GB of RAM. Now first describe the analysis of Landsat 8 OLI using RF, DT, SVM, GBT and CART. This study used Google Earth Engine platform for this purpose. Dataset is considered in the image form. Dataset was split into train and test with 80%-20% ratio respectively. Predicting a variable that is categorical by many independent attributes, uses classification models. It has six classes such as wheat, mustard,

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

vegetation, urban area, water body and other crops. Trained CART, DT, RF, GTB and SVM models on train datasets that consist of all the features. PA as well as UA also calculated using eqs. 4 and 5.Table 9 shows other accuracy such as PA and UA of Landsat 8 OLI.

$$\text{Producer Accuracy} = \frac{\text{Total number of correct pixel in a category}}{\text{Total number of pixel of that category derived from the reference data (row data)}} \quad 4$$

$$\text{User Accuracy} = \frac{\text{Total number of correct pixel in a category}}{\text{Total number of pixel of that category derived from the reference data (column data)}} \quad 5$$

Table 9. PA and UA Accuracy of Landsat 8 OLI.

	CART		RF		GTB		SVM		DT	
Classes	**PA %**	**UA%**	**PA %**	**UA**	**PA %**	**UA%**	**PA%**	**UA%**	**PA%**	**UA%**
Urban	88.29	83.83	89.18	99	91.91	91.91	77.96	92.92	85.56	85.56
Water	75.67	93.33	94.11	82.05	89.28	83.33	100	73.33	100	67.56
Vegetation	66.66	74.57	71.42	67.41	70.8	86.44	83.33	33.89	60.81	71.42
Wheat	78.78	77.77	74.28	79.59	88.42	83.16	52.25	80.19	77.08	71.15
Mustard	93.06	88.67	91.17	92.07	90.82	93.39	88.88	75.47	88.18	88.99
Other crops	50	42.85	40	14.28	83.33	35.71	40	22.22	22.22	40

Classified Map: Fig. 5 to Fig. 9 shows classified map of Landsat 8 OLI and also Landsat classified map for SVM, DT, CART, RF and GTB all methods. In classified map red color is used for urban area, blue color is used for water bodies, green color is used for vegetation, orange is used for wheat, yellow is used for mustard and cyan color shows other crops area. Images that have been classified are displayed using the various classifiers listed above while using GEE in the code editor (developers.google.com/earth-engine/tutorials).

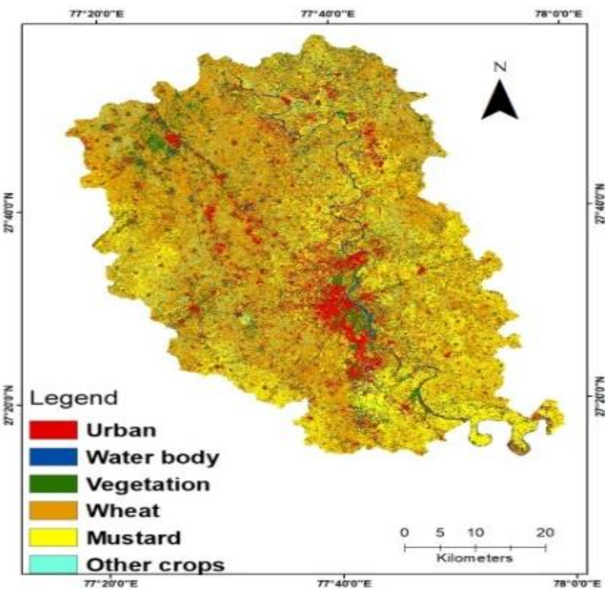

Figure 5. Crop Map using SVM

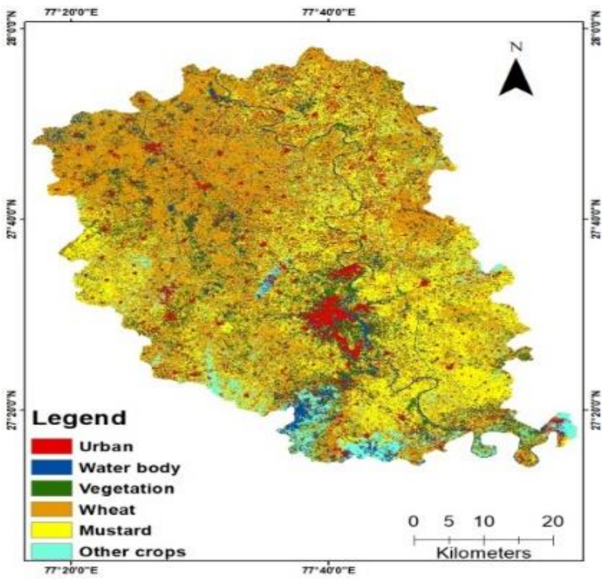

Figure 6. Crop Map using CART

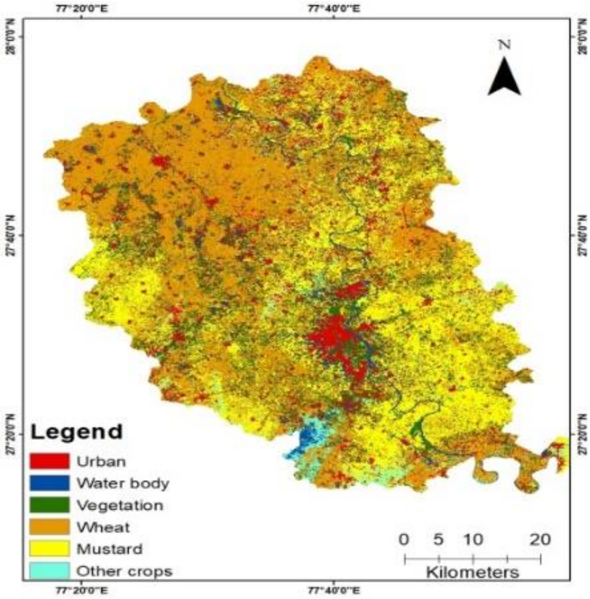

Figure 7. Crop Map using RF

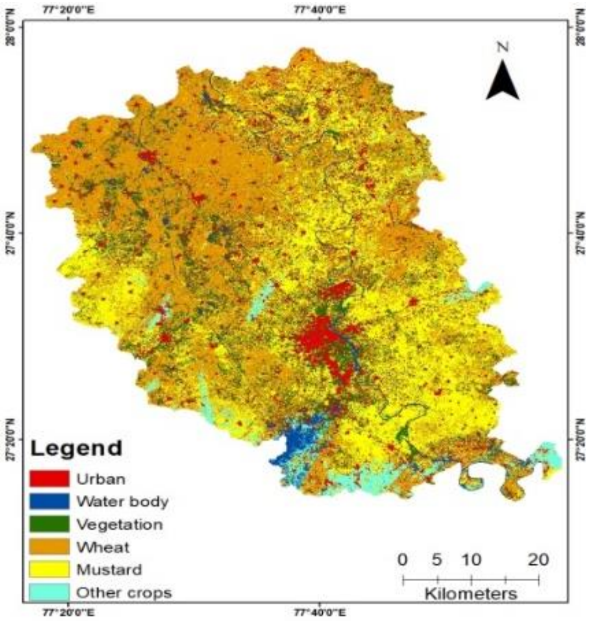

Figure 8. Crop map using GTB

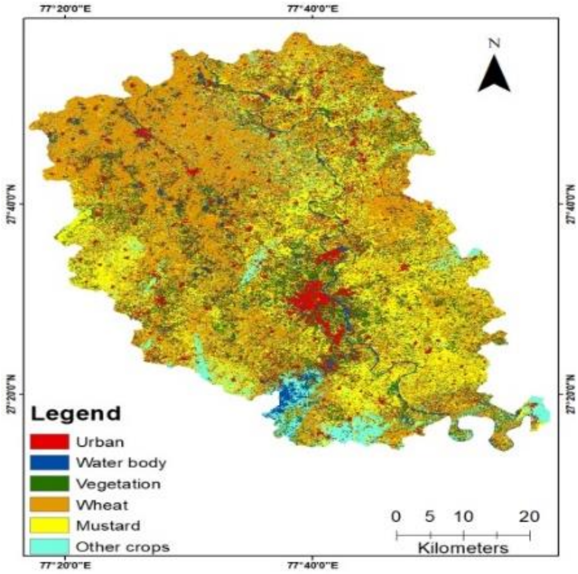

Figure 9. Crop Map using DT of Landsat 8 OLI

Classified Map and Various Accuracy of Sentinel 2 MSI

Table 10 to Table 14 shows confusion matrix of various methods (GTB, RF, CART, DT and SVM). Table 15 shows kappa and overall accuracy for classification of Sentinel dataset for Dec 2020 season. GTB achieved 84.23%, SVM 84.06, RF 82.35, DT 75.29%, and CART 75.29%. Other accuracies such as PA and UA are shown in Table 16. Fig. 10 to Fig. 14 shows a sentinel classified map for all methods. In classified map red color is used for urban areas, blue color is uses for water bodies, green color is uses for vegetation, orange is uses for wheat, yellow is uses for mustard and cyan color is uses for other crops areas. Overall Accuracy of Sentinel 2 is also calculated using equation 2 as well as kappa coefficient is calculated using eq. 3. Table 15 shows overall accuracy of Sentinel 2 MSI.

Table 10. Confusion matrix of GTB method for Sentinel 2 MSI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	54	0	0	3	0	0	57	94.73
Water	1	59	5	2	0	0	67	88.05
Vegetation	1	9	28	10	0	0	48	58.33
Wheat	6	1	3	122	11	1	144	84.72
Mustard	1	0	0	7	92	0	100	92
Other crops	0	0	0	5	1	3	9	33.33
ColumnTotal	63	69	36	149	104	4	425	
Producer 'accuracy (%)	85.71	85.50	77.77	81.87	88.46	75	OA	84.23

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 11. Confusion matrix of RF method for Sentinel 2 MSI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	53	1	1	2	0	0	57	92.98
Water	2	57	5	2	1	0	67	85.07
Vegetation	0	9	33	6	0	0	48	68.75
Wheat	9	2	9	114	10	0	144	79.16
Mustard	1	0	0	8	91	0	100	91
Other crops	0	0	0	5	2	2	9	22.22
ColumnTotal	65	69	48	137	104	2	425	
Producer 'accuracy (%)	81.53	82.60	68.75	83.21	87.5	100	OA	82.53

Table 12. Confusion matrix of CART method for Sentinel 2 MSI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	47	3	3	2	2	0	57	82.45
Water	1	53	7	5	1	0	67	79.10
Vegetation	1	14	22	9	0	2	48	45.83
Wheat	4	3	4	116	10	7	144	80.55
Mustard	1	0	1	15	81	2	100	81
Other crops	0	0	0	7	1	1	9	11.11
ColumnTotal	54	73	37	154	95	12	425	
Producer 'accuracy (%)	87.03	72.60	59.45	75.32	85.26	8.33	OA	75.29

Table 13. Confusion matrix of DT method for Sentinel 2 MSI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	47	3	3	2	2	0	57	82.45
Water	1	53	7	5	1	0	67	79.10
Vegetation	1	14	22	9	0	2	48	45.83
Wheat	4	3	4	116	10	7	144	80.55
Mustard	1	0	1	15	81	2	100	81
Other crops	0	0	0	7	1	1	9	11.11
Column Total	54	73	37	154	95	12	425	
Producer 'accuracy (%)	87.03	72.60	59.45	75.32	85.26	8.33	OA	75.29

Table 14. Confusion matrix of SVM method for Sentinel 2 MSI

Classes	Urban	Water	Vegetation	Wheat	Mustard	Other crops	Row Total	User's Accuracy (%)
Urban	53	1	1	2	0	0	57	92.98
Water	2	57	6	2	0	3	70	81.428
Vegetation	1	7	35	5	0	0	48	72.91
Wheat	5	2	2	124	11	0	144	86.11
Mustard	1	0	0	4	95	0	100	95
Other crops	0	1	1	6	1	5	14	35.71
Column Total	62	68	45	143	107	8	433	
Producer 'accuracy (%)	85.48	83.82	77.77	86.71	88.78	62.5	OA	84.06

The above Tables from 10 to Table 14 show the confusion matrix of all methods (GTB, RF, CART, DT, SVM). Diagonal values show correct prediction (correct classified) while other than diagonal specific to class shows misclassified (wrong

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

prediction). The above Tables 10 to Table 14 confusion matrix are conducted for Sentinel 2 MSI.

Table 15. Overall Accuracy and Kappa of Sentinel 2 MSI

S.NO	Classifier	OA %	Kappa
1	GTB	84.23	0.79
2	RF	82.35	0.77
3	CART	75.29	0.67
4	DT	75.29	0.67
5	SVM	84.06	0.81

Producer and User accuracy calculated using eqs. 4 and 5. Different types (PA, UA) of accuracy are shows in Table 16 for Sentinel 2MSI dataset.

Table 16. PA and UA accuracy of Sentinel 2 MSI dataset

	CART		RF		GTB		SVM		DT	
Classes	PA %	UA%	PA %	UA%	PA %	UA%	PA%	UA%	PA%	UA%
Urban	87.03	82.45	81.53	92.98	85.71	94.73	85.48	92.98	87.03	82.45
Water	72.60	79.10	82.60	85.07	85.50	88.05	83.82	81.42	72.60	79.10
Vegetation	59.45	45.83	68.75	68.75	77.77	58.33	77.77	72.91	59.45	45.83
Wheat	75.32	80.55	83.21	79.16	81.8 7	84.77	86.71	86.11	75.32	80.55
Mustard	85.26	81	87.51	91	88.46	92	88.78	95	85.26	81
Other crops	8.33	11.11	100	22.22	75	33.33	62.51	35.71	8.33	11.11

Classified Map for Sentinel 2 MSI: Fig. 10 to Fig. 14 shows classified map for Sentinel 2 data. Fig. 10 to Fig. 14 shows Sentinel 2 classified map for all methods. In classified map red color is used for urban area, blue color is used for water bodies, green color is used for vegetation, orange is used for wheat, yellow is used for mustard and cyan color shows other crops area.

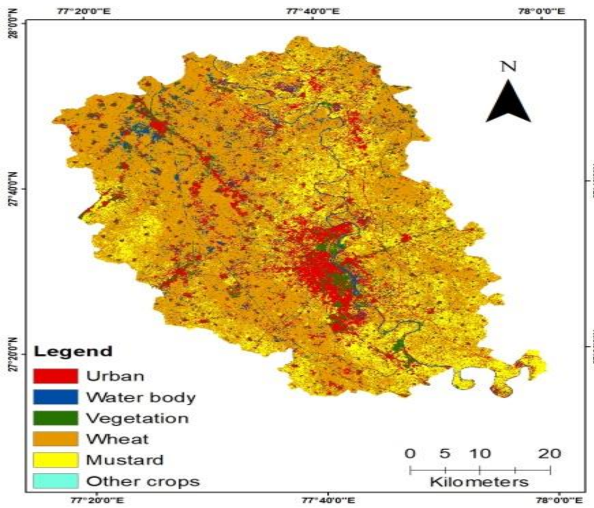

Figure 10. Crop Map using SVM

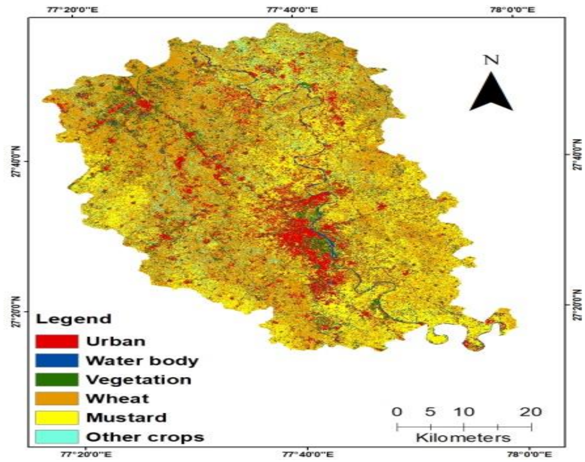

Figure 11. Crop map using CART

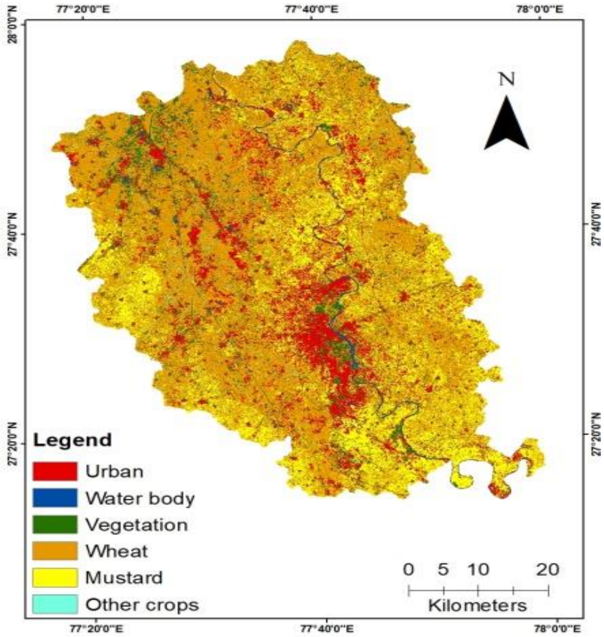

Figure 12. Crop map using RF

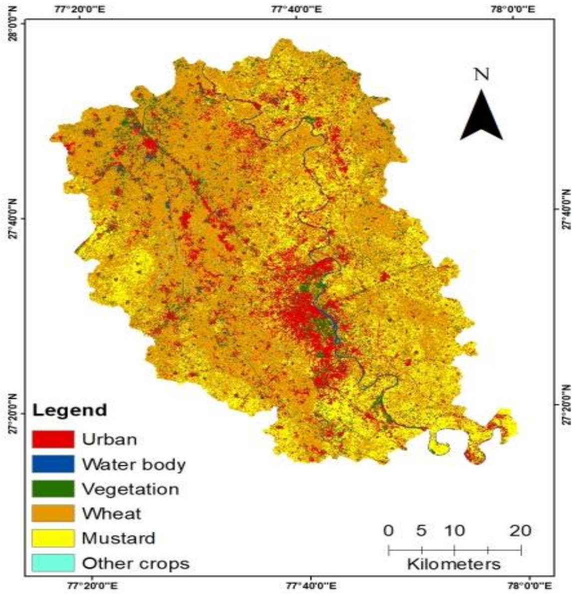

Figure 13. Crop map using GTB

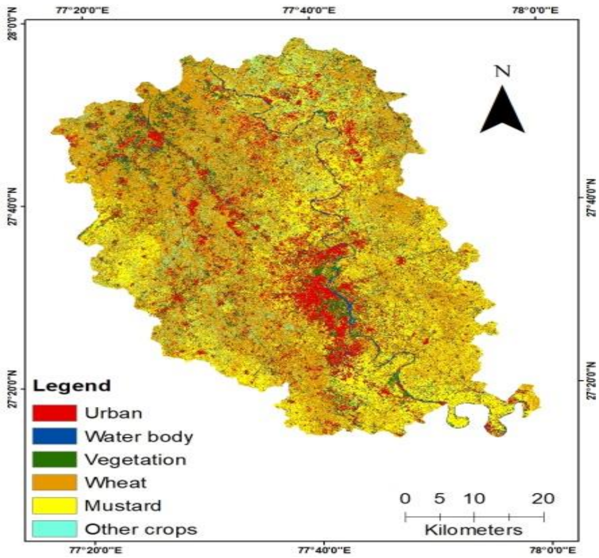

Figure 14.Crop map using DT of Sentinel 2 MSI

To train datasets using different methods such as SVM, RF, DT, GTB and CART with the same training and testing data. During classification process this can help selecting the best dates for the remote sensing image the above Fig. 5 to Fig. 9 show the result of Landsat 8 OLI image classification by GTB, RF, CART, DT and SVM methods. The classes in the crop maps are six as urban, vegetation, water, wheat, mustard, and other crops. The OA is computed. Tables 8 and 9 shows the accuracies of Landsat 8 OLI classification using different methods. GTB gives the highest accuracy in both datasets.

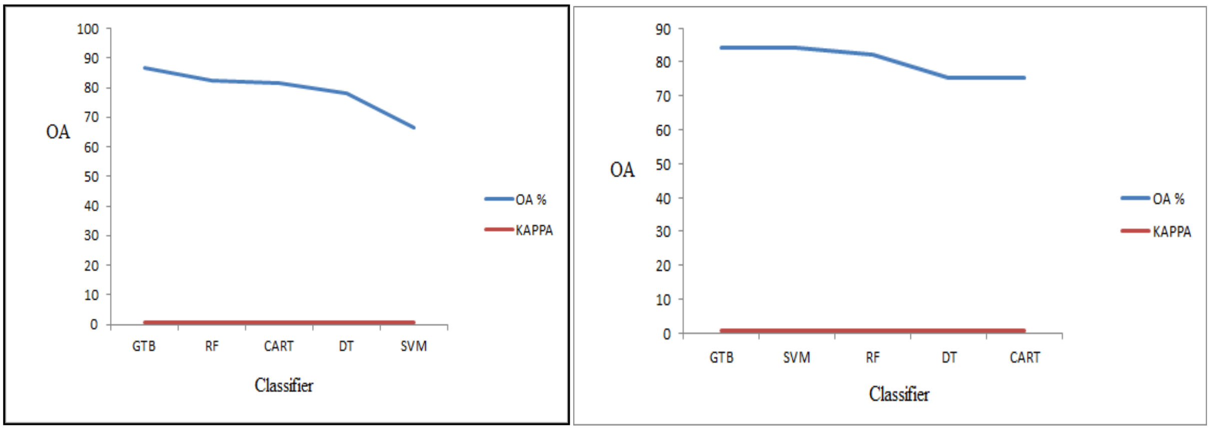

Figure 15. Overall accuracy and kappa coefficient of Landsat 8 OLI and Sentinel 2 MSI

Table 17. Comparative table of Landsat 8 OLI and Sentinel 2 MSI dataset on the bases of F1 Score

S.NO	Methods	Classes	F1 Score using Landsat 8 OLI Dataset (%)	F1 Score using Sentinel 2 MSI Dataset (%)
1	Gradient Tree Boosting	Urban	0.96	0.96
		Water	0.93	0.93
		Vegetation	0.94	0.94
		Wheat	0.95	0.95
		Mustard	0.96	0.96
		Other crops	0.76	0.76
2	Random Forest	Urban	0.97	0.97
		Water	0.96	0.96
		Vegetation	0.96	0.94
		Wheat	0.97	0.96
		Mustard	0.98	0.98
		Other crops	0.90	0.81
3	Classification and Regression Tree	Urban	1	1
		Water	1	0.99
		Vegetation	1	0.99
		Wheat	1	1
		Mustard	1	1
		Other crops	1	1
4	Decision Tree	Urban	1	0.99
		Water	1	1
		Vegetation	1	1
		Wheat	1	1
		Mustard	1	1
		Other crops	1	1
5	Support Vector Machine	Urban	0.85	0.99
		Water	0.75	0.98
		Vegetation	0.80	0.99
		Wheat	0.58	1
		Mustard	0.83	1
		Other crops	0.52	0.99

In both qualitative and quantitative ways, a detailed analysis and discussion were carried out. This looked at the performance and possibility of utilized GEE to speed up remote sensing and GIS procedures. For crop classification divided total area in to six classes as urban, vegetation, water, wheat, mustard, and other crops. GTB has repeatedly demonstrated good accuracy results based on these data. Table 8 shows OA and kappa for classification of Landsat 8 OLI data for

2024, 21(7): 2462-2484
https://doi.org/10.21123/bsj.2023.8952
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

December 2020 season. To applies different methods for obtained consistent result. The finding shows in Table 4 the overall accuracy are GTB achieved 86.79%, RF 82.53%, CART 81.41%, DT 78.09% and SVM 66.58% for Landsat 8 OLI. Fig. 15 represents graphical view of overall accuracy for all methods, which can easily compared to both sensors.

Discussion

In order to produce a crop map easily compare the outcome with both sensors Sentinel 2 MSI as well as Landsat 8 OLI. Earlier studies were more focused on other regions and only one sensor (Sentinel 2 or Landsat 8) are used[28, 32]. In previous research, only one sensor[26]. Sentinel-2 MSI is distinctive multi-spectral band that has received a lot of interest since its 3 red-edge bands can provide a wealth of spectrum data for monitoring vegetation. Integrating various bands improves crop classification overall accuracy. The most crucial elements in the multi-spectral crop classification were the RE-1 as well as SWIR-1 of Sentinel-2. Sentinel 2 MSI B3-Red, B4-Green and B8-NIR are utilized and Landsat 8 OLI bands 2 through band 7 bands are utilized. The research is achieving more accuracy and classifying crop of Mathura region due to highly cultivated area and also comparing the outcome with multiple sensors. It achieved a high F1 score which is shown in Table 17 in all methods as compared to previous study [32]. Five ML methods (SVM, DT, GTB, CART as well as RF) have been directly compared to see how well they perform using the specified performance assessment measures. Different data sets are examined for each algorithm's training and testing phases as model input.

The error variance between the actual and projected values is typically used to gauge the model accuracy, but other research employed confusion metrics to analyze performance which is shown in Table 3 to Table 7 and Table 10 to Table 14. It focused the work to estimate the effect of spectral-temporal variables on crop mapping in Mathura using GEE platform. GEE provide easy and powerful platform for handling huge amount of RS imagery used for crop mapping. In both Sensors, GTB achieved 86.79% overall accuracy in Landsat and 84.23% in Sentinel. While another method such as RF, CART, DT and SVM also achieve good accuracy. Compare the outcome of various classifiers such as CART, RF, DT, SVM and GTB by calculating different factors such as OA, PA, UA, and Kappa Coefficient and also classified map for all methods. Different machine learning methods have been tried in various regions in order to find adequate models for an accurate prediction due to the increased demand for crop mapping. As a result, research and development of an expert model for applications including crop mapping are ongoing challenges.

Conclusion

The aim of the research is to assess crop classification in Mathura using GEE platform. GEE provide an easy and powerful platform for handling huge amount of remote sensing imagery used for crop mapping. With the help of GEE platform, one can easily access remote sensing data and can apply different types of classification methods with minimum interaction and effort. The GEE platform offers many classification methods. GEE provides a good performance in allowing access to RS products based on cloud platform for satellite data flow. In both Sensors, GTB achieved 86.79% overall accuracy in Landsat and 84.23% in Sentinel. While another method such as RF, CART, DT and SVM also achieve good accuracy. Compare the performance of various classifiers such RF, DT, SVM, CART, GTB by calculating different factors such as OA, PA, UA and Kappa coefficient and also classified map for all methods. In view of F1 score this study achieved good F1 score as previous study[32]. It is conducted not just to investigate ML methods but also to compare the outcome with multiple sensors. Furthermore, compared pixel based crop mapping methods in Mathura and investigated the effectiveness of the GEE cloud platform for extensive crop mapping. Furthermore, GEE delivered excellent performance in facilitating access to RS images through the cloud platform as well as great computational abilities that may assist in dealing with large-scale crop mapping. It will analyze different Geo-location with various deep learning methods to train our method for crop classification. It is limited to the Mathura city and its agricultural regions only. If similar agricultural regions with the similar hydro-meteorological conditions exists then this study might apply. And some limitations, like dataset resolution (temporal and spatial), no field expeditions in especially the past time. It is limited to noisy or cloudy data. Future research could use microwaves and several dates of satellite data to extend the current investigation. There is also scope of cooperation in the outcome with more sensors and can utilize multiple dates with multiple regions.

List of abbreviations:

GTB Gradient Tree Boosting
DT Decision Tree
ML Machine Learning
SVM Support Vector Machine
CART Classification and Regression Trees
RF Random Forest
PA Producer Accuracy
TOA Top of atmosphere
UA User Accuracy
GB Gradient Boosting
GEE Google Earth Engine
MSI Multi Spectral Images
OLI Operational Land Imager
GIS Geographical Information System
IDE Integrated Development Environment

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at Suresh Gyan Vihar University.

Authors' Contribution Statement

P. G. conceptualization, data set curation, methodology, formal analysis. V. N. M. conceptualization. S. K. formal analysis. T. S. S. formal analysis, investigation. S. K.S. formal analysis. All authors have read as well as agreed to the published version of the manuscript.

References

1. Shelestov A, Lavreniuk M, Kussul N, Novikov A, Skakun S. Exploring Google Earth Engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. Front Earth Sci. 2017: 17. https://doi.org/10.3389/feart.2017.00017
2. Weiss M, Jacob F, Duveiller G. Remote sensing for agricultural applications: A meta-review. Remote Sens. Environ. 2020; 236: 111402. https://doi.org/10.1016/j.rse.2019.111402
3. Basso B, Liu L. Seasonal crop yield forecast: Methods, applications, and accuracies. Adv Agron. 2019 Jan 1; 154: 201-55. https://doi.org/10.1016/bs.agron.2018.11.002
4. Gallego FJ, Kussul N, Skakun S, Kravchenko O, Shelestov A, Kussul O. Efficiency assessment of using satellite data for crop area estimation in Ukraine. Int J Appl Earth Obs. Geoinf. 2014; 29: 22-30. https://doi.org/10.1016/j.jag.2013.12.013
5. Lobell DB, Thau D, Seifert C, Engle E, Little B. A scalable satellite-based crop yield mapper. Remote Sens Environ. 2015; 164: 324-33. https://doi.org/10.1016/j.rse.2015.04.021
6. Skakun S, Vermote E, Franch B, Roger JC, Kussul N, Ju J, et al. Winter wheat yield assessment from Landsat 8 and Sentinel-2 data: Incorporating surface reflectance, through phenological fitting, into regression yield models. Remote Sens. 2019; 11(15):1768. https://doi.org/10.3390/rs11151768
7. Mishra VN, Prasad R, Kumar P, Srivastava PK, Rai PK. Knowledge-based decision tree approach for mapping spatial distribution of rice crop using C-band synthetic aperture radar-derived information. J Appl Remote Sens. 2017; 11(4): 046003-. https://doi.org/10.1117/1.JRS.11.046003
8. Ji S, Zhang C, Xu A, Shi Y, Duan Y. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote

Sens. 2018; 10(1): 75. https://doi.org/10.3390/rs10010075
9. Sicre CM, Fieuzal R, Baup F. Contribution of multispectral (optical and radar) satellite images to the classification of agricultural surfaces. Int J Appl Earth Obs Geoinf. 2020 ;84: 101972. https://doi.org/10.1016/j.jag.2019.101972
10. Orieschnig CA, Belaud G, Venot JP, Massuel S, Ogilvie A. Input imagery, classifiers, and cloud computing: Insights from multi-temporal LULC mapping in the Cambodian Mekong Delta. Eur J Remote Sens. 2021; 54(1): 398-416. https://doi.org/10.1080/22797254.2021.1948356
11. Cavallaro G, Riedel M, Richerzhagen M, Benediktsson JA, Plaza A. On understanding big data impacts in remotely sensed image classification using support vector machine methods. IEEE J Sel Top Appl Earth Obs Remote Sens. 2015; 8(10): 4634-46. https://doi.org/10.1109/JSTARS.2015.2458855
12. Carretero J, Blas JG. Introduction to cloud computing: platforms and solutions. Cluster Comput. 2014; 17: 1225-9: https://doi.org/10.1007/s10586-014-0352-5
13. Kumar L, Mutanga O. Google Earth Engine applications since inception: Usage, trends, and potential. Remote Sens. 2018; 10(10): 1509.https://doi.org/10.3390/rs10101509
14. Kumar L, Mutanga O. Remote sensing of above-ground biomass. Remote Sens. 2017; 9(9): 935. https://doi.org/10.3390/rs9090935
15. Tamiminia H, Salehi B, Mahdianpari M, Quackenbush L, Adeli S, Brisco B. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. J Photogramm. Remote Sens. 2020; 164: 152-70. https://doi.org/10.1016/j.isprsjprs.2020.04.001
16. Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. Int J Remote Sens. 2007; 28(5): 823-70. https://doi.org/10.1080/01431160600746456
17. Sishodia RP, Ray RL, Singh SK. Applications of remote sensing in precision agriculture: A review. Remote Sens. 2020; 12(19): 3136.https://doi.org/10.3390/rs12193136
18. Amani M, Ghorbanian A, Ahmadi SA, Kakooei M, Moghimi A, Mirmazloumi SM, et al. Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review. IEEE J Sel Top Appl Earth Obs Remote Sens. 2020; 13: 5326-50. https://doi.org/10.1109/JSTARS.2020.3021052
19. Pelletier C, Webb GI, Petitjean F. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. 2019; 11(5): 523.https://doi.org/10.3390/rs11050523
20. Leslie CR, Serbina LO, Miller HM. Landsat and agriculture—Case studies on the uses and benefits of Landsat imagery in agricultural monitoring and production. U S Geol Surv. 2017. https://doi.org/10.3133/ofr20171034
21. Debats SR, Luo D, Estes LD, Fuchs TJ, Caylor KK. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. Remote Sens Environ. 2016; 179: 210-21. https://doi.org/10.1016/j.rse.2016.03.010
22. Skakun S, Franch B, Vermote E, Roger JC, Becker-Reshef I, Justice C, et al. Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. Remote Sens. Environ. 2017; 195: 244-58. https://doi.org/10.1016/j.rse.2017.04.026
23. Benami E, Jin Z, Carter MR, Ghosh A, Hijmans RJ, Hobbs A, et al. Uniting remote sensing, crop modelling and economics for agricultural risk management. Nat Rev Earth Environ. 2021 ; 2(2): 140-59. https://doi.org/10.1038/s43017-020-00122-y
24. Skakun S, Franch B, Vermote E, Roger JC, Justice C, Masek J, et al. Winter wheat yield assessment using Landsat 8 and Sentinel-2 data. InIGARSS 2018-2018. Int Geosci Remote Sens Symp. 2018 (pp. 5964-5967). IEEE. https://doi.org/10.1109/IGARSS.2018.8519134
25. Kussul N, Lemoine G, Gallego FJ, Skakun SV, Lavreniuk M, Shelestov AY. Parcel-based crop classification in Ukraine using Landsat-8 data and Sentinel-1A data. IEEE J Sel Top Appl Earth Obs Remote Sens. 2016; 9(6): 2500-8. https://doi.org/10.1109/JSTARS.2016.2560141
26. Lavreniuk M, Kussul N, Meretsky M, Lukin V, Abramov S, Rubel O. Impact of SAR data filtering on crop classification accuracy. IEEE First Ukraine. Conf Electr Comput Eng. UKRCON 2017 (pp. 912-917). IEEE.https://doi.org/10.1109/UKRCON.2017.8100381
27. Claverie M, Ju J, Masek JG, Dungan JL, Vermote EF, Roger JC, et al. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens Environ. 2018; 219: 145-61. https://doi.org/10.1016/j.rse.2018.09.002
28. Shelestov A, Lavreniuk M, Kussul N, Novikov A, Skakun S. Large scale crop classification using Google earth engine platform. IEEE Int Geosci Remote Sens Symp .2017 (pp. 3696-3699).. https://doi.org/10.1109/IGARSS.2017.8127801
29. Kussul N, Mykola L, Shelestov A, Skakun S. Crop inventory at regional scale in Ukraine: developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery. Eur. J. Remote Sens. 2018;51(1):627-36. https://doi.org/10.1080/22797254.2018.1454265
30. Chen J, Cao X, Peng S, Ren H. Analysis and applications of GlobeLand30: a review. Int J Geo-Inf Remot sens. 2017; 6(8): 230. https://doi.org/10.3390/ijgi6080230

31. Yang N, Liu D, Feng Q, Xiong Q, Zhang L, Ren T, et al. Large-scale crop mapping based on machine learning and parallel computation with grids. Remote Sens. 2019; 11(12): 1500. https://doi.org/10.3390/rs11121500
32. Neetu, Ray SS. Exploring machine learning classification algorithms for crop classification using Sentinel 2 data. Int Arch Photogramm Remote Sens Spat Inf Sci. 2019; 42: 573-8. https://doi.org/10.5194/isprs-archives-XLII-3-W6-573-2019, 2019
33. Asroni A, Ku-Mahamud KR, Damarjati C, Slamat HB. Arabic speech classification method based on padding and deep learning neural network. Baghdad Sci. J. 2021;18(2 (Suppl.)):0925-. http://dx.doi.org/10.21123/bsj.2021.18.2(Suppl.).0925
34. Jasim OZ. Using of machines learning in extraction of urban roads from DEM of LIDAR data: Case study at Baghdad expressways, Iraq. Period. Eng. Nat. Sci. 2019; 7(4):1710-21. http://dx.doi.org/10.21533/pen.v7i4.914
35. Khanday AM, Khan QR, Rabani ST. Detecting textual propaganda using machine learning techniques. Baghdad Sci. J. 2021; 18(1):0199-. https://doi.org/10.21123/bsj.2021.18.1.0199
36. Mahmood RA, Abdi A, Hussin M. Performance evaluation of intrusion detection system using selected features and machine learning classifiers. Baghdad Sci. J. 2021;18(2 (Suppl.)):0884-. http://dx.doi.org/10.21123/bsj.2021.18.2(Suppl.).0884
37. Ashraf S, Alfandi O, Ahmad A, Khattak AM, Hayat B, Kim KH, et al . Bodacious-instance coverage mechanism for wireless sensor network. Wirel Commun Mob Comput. 2020; 2020:1-1. https://doi.org/10.1155/2020/8833767
38. Campos-Taberner M, García-Haro FJ, Martínez B, Sánchez-Ruíz S, Gilabert MA. A copernicus sentinel-1 and sentinel-2 classification framework for the 2020+ European common agricultural policy: A case study in València (Spain). Agronomy. 2019; 9(9):556. https://doi.org/10.3390/agronomy9090556
39. Immitzer M, Vuolo F, Atzberger C. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Remote Sens. 2016; 8(3): 166. https://doi.org/10.3390/rs8030166
40. Nasrallah A, Baghdadi N, Mhawej M, Faour G, Darwish T, Belhouchette H, et al. A novel approach for mapping wheat areas using high resolution Sentinel-2 images. Sensors. 2018; 18(7): 2089. https://doi.org/10.3390/s18072089
41. Piedelobo L, Hernández-López D, Ballesteros R, Chakhar A, Del Pozo S, González-Aguilera D, Moreno MA. Scalable pixel-based crop classification combining Sentinel-2 and Landsat-8 data time series: Case study of the Duero river basin. Agric Syst. 2019; 171: 36-50. https://doi.org/10.1016/j.agsy.2019.01.005
42. Vapnik V, Vapnik V. Realism and Instrumentalism: Classical Statistics and VC Theory (1960–1980). Estimation of Dependences Based on Empirical Data. 2006:411-24. https://doi.org/10.1007/0-387-34239-7_11
43. Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: A review. J Photogramm. Remote Sens. 2011; 66(3): 247-59. https://doi.org/10.1016/j.isprsjprs.2010.11.001
44. Maulik U, Chakraborty D. Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques. IEEE Trans Geosci Remote Sens. 2017; 5(1): 33-52. https://doi.org/10.1109/MGRS.2016.2641240
45. Sivasankar T, Kumar D, ShankerSrivastava H, Patel P. Wheat leaf area index retrieval using RISAT-1 hybrid polarized SAR data. Geocarto Int. 2020; 35(8):9 05-15. https://doi.org/10.1080/10106049.2019.1566404
46. Breiman L. Classification and regression trees. Routledge; 2017 Oct 19. https://doi.org/10.1201/9781315139470
47. Bittencourt HR, Clarke RT. Use of classification and regression trees (CART) to classify remotely-sensed digital images. Int Geosci Remote Sens Symp. 2003; 7: 3751-3753 (IEEE Cat. No. 03CH37477). https://doi.org/10.1109/IGARSS.2003.1295258
48. Zhang N, Wu Y, Zhang Q. Detection of sea ice in sediment laden water using MODIS in the Bohai Sea: A CART decision tree method. Int J Remote Sens. 2015; 36(6):1661-74. https://doi.org/10.1080/01431161.2015.1015658
49. Han J, Mao K, Xu T, Guo J, Zuo Z, Gao C. A soil moisture estimation framework based on the CART algorithm and its application in China. J Hydrol. 2018; 563: 65-75. https://doi.org/10.1016/j.jhydrol.2018.05.051
50. Tamy S, Belhadaoui H, Rabbah MA, Rabbah N, Rifi M. An evaluation of machine learning algorithms to detect attacks in SCADA network. In2019 7th Mediterranean CMT 2019 (pp. 1-5). IEEE. https://doi.org/10.1109/CMT.2019.8931327
51. Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 2016; 114: 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011
52. Guan H, Li J, Chapman M, Deng F, Ji Z, Yang X. Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. Int J Remote Sens. 2013; 34(14): 5166-86. https://doi.org/10.1080/01431161.2013.788261
53. Zhang H, Li Q, Liu J, Du X, Dong T, McNairn H. Object-based crop classification using multi-temporal SPOT-5 imagery and textural features with a Random Forest classifier. Geocarto Int. 2018;

33(10): 1017-35. https://doi.org/10.1080/10106049.2017.1333533
54. Shukla G, Garg RD, Srivastava HS, Garg PK. Performance analysis of different predictive models for crop classification across an aridic to ustic area of Indian states. Geocarto Int. 2018; 33(3): 240-59. https://doi.org/10.1080/10106049.2016.1240721
55. Man CD, Nguyen TT, Bui HQ, Lasko K, Nguyen TN. Improvement of land-cover classification over frequently cloud-covered areas using Landsat 8 time-series composites and an ensemble of supervised classifiers. Int J Remote Sens. 2018; 39(4): 1243-55. https://doi.org/10.1080/01431161.2017.1399477
56. Bui QT, Chou TY, Hoang TV, Fang YM, Mu CY, Huang PH, et al. Gradient boosting machine and object-based CNN for land cover classification. Remote Sens. 2021; 13(14): 2709. https://doi.org/10.3390/rs13142709
57. Sharifi A. Yield prediction with machine learning algorithms and satellite images. J Sci Food Agric. 2021; 101(3): 891-6. https://doi.org/10.1002/jsfa.10696
58. Grandini M, Bagli E, Visani G. Metrics for multi-class classification: an overview. arXiv preprint arXiv: 2008.05756. 2020. https://doi.org/10.48550/arXiv. 2008.05756

دراسة مقارنة التعلم الآلي في التصنيف الفعال للبيانات متعددة الأطياف في الزراعة

بريانكا غوبتا[1]، شروتي كانجا[2]، فارون نارايان ميشرا[3]، سوراج كومار سينغ[4]، ثوتا سيفاسانكار[5]

[1]قسم علوم وهندسة الحاسوب، جامعة سوريش جيان فيهار، جايبور، الهند.
[2]قسم الجغرافيا، كلية البيئة وعلوم الأرض، جامعة البنجاب المركزية، باتيندا، البنجاب، الهند.
[3]معهد أميتي للمعلوماتية الجغرافية والاستشعار عن بعد(AIGIRS) ، جامعة أميتي، نويدا الهند.
[4]قسم مركز التنمية المستدامة، جامعة سوريش جيان فيهار، جايبور، الهند.
[5] قسم نظم المعلومات الجغرافية، جامعة NIIT، نيمرانا، راجاستان، الهند.

الخلاصة

يتطلب أن تكون خرائط محاصيل موثوقة ودقيقة لتحقيق الأمن الغذائي من المستوى الإقليمي إلى المستوى العالمي. يؤدي التوفر المتزايد لصور الأقمار الصناعية إلى مشكلة "البيانات الضخمة" أثناء إنتاج خرائط المحاصيل. الآن، اكتسبت المنصات السحابية الكثير من الاهتمام لتصنيف المحاصيل في مناطق واسعة. الهدف الرئيسي من البحث هو تحليل تصنيف المحاصيل باستخدام مختلف التعلم الآلي (ML) مثل آلة دعم المتجهات (SVM)، وتعزيز شجرة التدرج (GTB)، والغابات العشوائية (RF)، وشجرة القرار (DT) بالإضافة إلى التصنيف والتصنيف. أشجار الانحدار (CART) على منصة محرك Goggle Earth. الهدف هو استكشاف كفاءة محرك Google Earth (GEE) عند تصنيف المحاصيل المختلفة باستخدام مجموعات البيانات متعددة الأطياف من Sentinel 2 MSI بالإضافة إلى الأقمار الصناعية Landsat 8 OLI لرسم خرائط المحاصيل في منطقة ماثورا في ولاية أوتار براديش بالهند. تم استخدام أفضل صورة خالية من السحابة (أقل من 5%) لمجموعات بيانات Landsat 8 OLI و"2020-12-Sentinel 2 MSI (" 26"، و"30-12-2020") لتصنيف المحاصيل بمساعدة التصفية التلقائية، أي النسبة المئوية. الملكية السحابية على منصات GEE. علاوة على ذلك، يمكن تنظيم أداء منصة GEE والحصول عليها وتوضيحها وكذلك معالجتها المسبقة لمجموعة بيانات الأقمار الصناعية بقوة كبيرة. تم استخدام النقاط كمساحات مميزة مثل مجموعات بيانات التدريب. علاوة على ذلك، يتم استخدام مصفوفات الارتباك لتقييم الدقة (دقة المنتج والمستخدم) ومعامل كابا. بالإضافة إلى ذلك، قم بمقارنة نتائج مجموعة البيانات Sentinel 2 MSI) وLandsat 8 OLI) على أساس الدقة الإجمالية (OA) ودرجة F1 بالإضافة إلى معامل كابا. تم العثور على أعلى وصول حر باستخدام GTB (86.7% يليه RF (82.5%)، CART (81.0%)، DT (78.1% وSVM (66.5%) لصورة Landsat 8 OLI. بالنسبة لصورة Sentinel 2، حققت GTB أعلى وصول وصول بنسبة 84.2% يليها SVM (84%)، RF (82.3%)، DT (75.2%))، و75.0% CART) على التوالي. على أساس البحث، وجد أن أداء GTB كان جيدًا بين جميع المصنفات في رسم خرائط المحاصيل باستخدام مجموعتي البيانات متعددة الأطياف.

الكلمات المفتاحية: محرك جوجل إيرث (GEE)، التعلم الآلي (ML)، الاستشعار عن بعد (RS)، صور الأقمار الصناعية.