CORPUS PUBLISHERS

Archives of Agriculture Research and Technology (AART)

ISSN: 2832-8639

Volume 6 Issue 3, 2025

Article Information

Received date : June 25, 2025

Published date: July 08, 2025

*Corresponding author

Akash Goyal, Regional Remote Sensing Centre-North, New Delhi, India

DOI: 10.54026/AART/1082

Keywords

Machine Learning, Automation; Hyperparameter Tuning; Geoprocessing; SNAP; Sentinel-1A; EOS-4, Digital Elevation Model; Remote Sensing

Distributed under Creative Commons CC-BY 4.0

Research Article

Automated GUI-Based System for In-Season Crop Classification and Acreage Estimation using Multi-Temporal SAR Data and Machine Learning

Akash Goyal[1*], Radhika Aggarwal[2], Shilpa Mahajan[2] and Sameer Saran[1]

[1]*Regional Remote Sensing Centre-North, New Delhi, India*

[2]*Department of CSE, The NorthCap University, Gurugram, Haryana, India*

Abstract

Accurate in-season crop acreage statistics are crucial for agricultural policymakers, stakeholders, and the food security community. However, manual workflows for downloading, preprocessing, and classifying Synthetic Aperture Radar (SAR) data - especially over large areas such as states or country - are time-consuming and prone to inefficiencies. This study presents a Graphical User Interface (GUI)-based automated system designed for multi-class crop classification and acreage estimation using multi-temporal SAR data and machine learning techniques. The system supports data acquisition from two SAR platforms: Sentinel-1A (VV & VH) and EOS-4 (HH & HV), based on user-defined areas of interest (district or state level). The automation pipeline includes comprehensive preprocessing steps such as radiometric calibration, speckle filtering, geometric/terrain correction, mosaicing, crop land masking, and layer stacking. Classification is performed using six different machine learning algorithms, with integrated hyperparameter tuning to enhance model performance. The system outputs class-wise crop area statistics and validates class separability using backscatter temporal profiles. To evaluate its effectiveness, the system was applied to selected districts and achieved an overall classification accuracy of approximately 90% for major crops including paddy, arhar, cotton, and maize. A four-page auto-generated report summarizes the outputs, featuring the classification report, confusion matrix, backscatter curves, variable importance plots, and classified imagery. The system is scalable, efficient, and user-friendly - requiring minimal technical expertise - making it a valuable tool for stakeholders in the agricultural domain. Its methodology can be readily adapted to other datasets and geographic regions, supporting broader applications in operational crop monitoring and decision-making.

Introduction

Accurate and timely estimation of in-season crop acreage is critical for agricultural monitoring, food security planning, and evidence-based policymaking, particularly in large, diverse agrarian economies such as India [1,2]. Traditionally, these estimates have relied on field surveys and manual interpretation of satellite imagery, methods that are not only labour-intensive and time-consuming but also unfeasible for large-scale operational implementation. The emergence of Earth Observation (EO) technologies - especially Synthetic Aperture Radar (SAR) - has transformed the landscape of crop monitoring, offering high-resolution, weather-independent data acquisition capabilities [3,4].

SAR sensors, including Sentinel-1A and EOS-4, provide valuable temporal information on crop phenology and structural changes, even under persistent cloud cover conditions, making them particularly useful during monsoon seasons [5,6]. However, the practical utility of SAR data remains constrained by its complex preprocessing requirements - radiometric calibration, speckle filtering, geometric/terrain correction, mosaicking, and temporal stacking - all of which require substantial domain expertise and computational effort [7].

In recent years, machine learning (ML) techniques have shown considerable promise in automating and enhancing SAR-based crop classification workflows [8,9]. Nonetheless, many existing approaches suffer from several operational shortcomings, including fragmented toolchains, non-intuitive interfaces, and inadequate optimization of ML models. Such limitations restrict the accessibility and scalability of these tools, particularly for end-users in government departments, agribusinesses, or rural advisory services who may lack deep technical expertise.

Moreover, prior systems frequently exhibit poor memory management, inefficient processing routines, and inadequate error handling, especially in terrain correction using DEM data such as SRTM 1 Sec HGT. Challenges related to GUI usability, excessive dependency on external libraries, and limited support for hyperparameter tuning further compromise their effectiveness. In some cases, inaccurate area estimation and non-scalable architectures have diminished the credibility and reliability of SAR-based monitoring tools in real-world applications.

To address these limitations, this study proposes a fully automated, GUI-based system for in-season crop classification and acreage estimation, utilizing multi-temporal SAR datasets from Sentinel-1A (VV & VH) and EOS-4 (HH & HV). The system incorporates a seamless data acquisition module (via ASF API or local input), a robust geoprocessing pipeline implemented through the SNAP Python interface, and ML-driven classification with integrated hyperparameter tuning using the Scikit-learn framework. Importantly, a built-in module automates the download and integration of SRTM elevation data to facilitate accurate terrain correction.

The system is designed to be memory-efficient and operationally robust, featuring a redesigned, intuitive front-end built using PyQt5 that simplifies user interactions and enhances interpretability. Outputs include classified maps, variable importance plots, backscatter temporal curves, and a statistical summary report. With demonstrated classification accuracy approaching 90% for key crops like paddy, arhar, cotton, and maize in selected Indian districts, the proposed system offers a scalable and user-friendly solution for SAR-based crop monitoring.

By integrating recent advances in ML and geospatial automation, this research contributes to the growing field of digital agriculture and addresses the longstanding need for accessible, efficient, and accurate tools for operational crop monitoring and acreage forecasting.

Description of Data

Satellite Data

In the proposed automation tool, SAR data has been utilized to bring about information over crop area for an ROI. Table 1 depicts the specifications of the respective satellites and the data used.

Table 1: Tabular description of types of data used: Sentinel-1A, and EOS-4.

S. No.	Satellite	Year Launched	Sensor	Organization	Spatial Resolution	Temporal Resolution	Polarization
1	Sentinel-1A	2014	SAR-C	ESA	10m	12 days	VV, VH
2	EOS-4	2022	SAR-C	ISRO	33 m (MRS)	17 days	HH, HV

Sentinel-1A

Sentinel-1A is a Synthetic Aperture Radar (SAR) satellite mission operated by the European Space Agency (ESA) as a part of the Copernicus program. It provides high-resolution SAR data over a wide range of applications including land, ocean, and cryosphere monitoring. The Sentinel-1A mission carries a C-band SAR sensor, which operates in two polarizations: Vertical-Vertical (VV) and Vertical-Horizontal (VH). The VV polarization is transmitted and received in the vertical direction. This polarization is sensitive to the vertical structure of the target, such as vegetation canopy and roughness of the surface. It is commonly used for land applications such as vegetation mapping, soil moisture estimation, and urban area monitoring. In addition, it is sensitive to oil spills and can be used for marine applications. The VH polarization is transmitted in the vertical direction and received in the horizontal direction. This polarization is sensitive to the orientation of the target, such as man-made objects and surface roughness. It is commonly used for land applications such as detecting man-made objects and monitoring infrastructure. It is also used for sea-ice monitoring and ship detection in marine applications [10].

Sentinel-1A provides SAR data with a spatial resolution of up to 10 meters and a swath width of up to 400 kilometres. It operates in different modes including Stripmap, Interferometric Wide Swath, and Extra-Wide Swath modes, each with different spatial resolutions and swath widths. The data can be acquired in different polarization modes, including single polarization and dual polarization modes [11].

EOS-4

EOS-04 was launched on 14-February 2022 by ISRO"s own PSLV C-52. EOS-04 is a Low Earth Orbit (LEO) satellite to be operated in a Sun Synchronous Polar Orbit (SSPO) with 6 AM-6 PM Equatorial Crossing Time (ECT) at an altitude of 524.87 km carrying a Synthetic Aperture Radar (SAR) payload. EOS-04 Spacecraft is configured using ISRO"s RISAT-1 heritage bus and capabilities are fully exploited with respect to accommodation, power generation, thermal management etc. EOS-04 SAR is capable of providing data in various resolution modes catering to a variety of applications as demonstrated in its precursor mission RISAT-1. The main objective of EOS-04 mission is to provide continuity of data to the users. To cater to the applications, the SAR payload of EOS-04 shall operate in C-Band frequency range (5.4G Hz) and in Side-Looking Radar mode with performance parameters for different modes as specified in the following sections. The EOS-04 SAR will be operating in C-band at a frequency of 5.4 G Hz. The SAR system has been designed to provide constant swath for all elevation pointing for stripmap mode of imaging. Full-polarimetric mode has been introduced newly in EOS04. Quad (Full) polarization will be operational for FRS-1, FRS-2 and for ScanSAR MRS and CRS modes. Quad (Full) polarization is not available for HRS mode. HRS mode configuration in EOS-04 is reconfigured with lower bandwidth (75 MHz) according to available WLAN frequency band.

Experimental Setup

To develop the automation tool, a set of technical requirements were considered. Table 2 describes the technology stack used for implementation.

Table 2: Technology stack for the development of the automation tool.

S. No.	Technology	Use Case
1	Python Programming	v3.6, v3.11
2	GDAL	Data engineering
3	SNAP Python Interface	Data preprocessing
4	ASF-Search	Data acquisition
5	Scikit-learn	Data classification
6	PyQt5	GUI

Python (Python 3.6.0 and Python 3.11.0)

Python is a popular open-source programming language that provides improved process control features. It can create sophisticated multi-protocol network applications while simultaneously preserving a clear and simple syntax. Two versions of Python are required for the proposed work because of the compatibility issues with supporting Python interfaces like "snappy" (SNAP-Py interface) and "scikit-learn" for machine learning classifications. Python 3.6 is used for preprocessing geospatial images. Python 3.11 is used for training the model and predicting the classified images using geospatial data abstraction techniques.

SNAP Python Interface

The primary programming language for the implementation of SNAP (Sentinel Application Platform) is Java, and so is its API a Java API. But somehow, Python has been used here for the SNAP interface because its architecture describes that one can use both SNAP Engine and Desktop with Python too. Using SNAP with Python elevates the overall data preprocessing and accuracy of process outputs, such as terrain correction and speckle filtering. It supports a very limited number of Python versions, i.e., Python 2.7, 3.3, and 3.6. In the proposed automation tool, Python 3.6 has been used to preprocess the satellite data (Standford Network Analysis Project).

For implementation through Python, the steps required for Configurations set up by the user in the SNAP Command Line are discussed and shown in Figure 1.

Install SNAP using the setup file available online.

Keep clicking Next, since no customization is needed. Click on the Install button and close after installation.

Open 'SNAP Command Line' from Start.

Type cd C:\Program Files\snap\bin

Type snappy-conf C:\Python36\python.exe

Copy 'C:\Users*username\.snap\snap-python\snappy' to 'C:\Python36\Lib\site-packages\\'

Users can set up a memory limit for SNAP to consume since its major drawback is memory and space management.

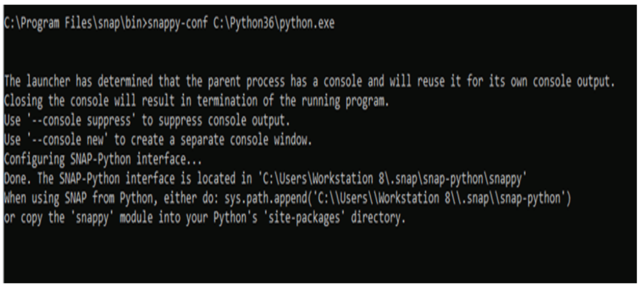


Figure 1: SNAP Command-Line.

GDAL (Geospatial Data Abstraction Library)

It is a Python library, released in 2000, used to manipulate and study geospatial data in raster and vector formats, e.g., '.tif' for raster, '.shp' for vector. It is a free, open-source library widely provided in the form of wheel files for almost all Python versions. A good number of geospatial/remote-sensing-based software uses this library in their backend programs, such as ArcGIS, QGIS, ERDAS, Google Earth, etc. GDAL wheel files are usually available at the GDAL Documentation.

ASF-ALASKA Python Interface

ASF-Alaska is a website managed by NASA to download SAR spatial imagery. It has provided a Python interface namely "asf_search" which allows users to download required SAR satellite images.

It is simple to use and provides a bridge between NASA's Alaska Satellite Facility Distribution Active Archive Center (ASF DAAC), which allows the user to download satellite data by specifying parameters such as start and end dates, flight direction, platform, well-known text, processing level, and a relative orbit. To access it through Python, one needs to have a login username and password for authentication purposes [12].

SCIKIT-LEARN Machine Learning Library

It is a Python library used for the implementation of machine learning algorithms and techniques on a set of training and testing data. It consists of various regression, classification, and clustering algorithms such as linear regression, logistic regression, SVMs, decision trees, random forest, gradient boosting, k-means, etc. to manipulate and apply the property of prediction and estimation on a given set of data.

The proposed automation software is used for working with random forests, decision trees, k-nearest neighbors, support vector machines, naïve Bayes, and multi-layer perceptron.

Py-QT5 Technology

It is one of the options available in Python used to develop GUI-based applications to provide an easy-to-interact, user-friendly frontend so that the communication between processes and the end goal receiver becomes faster and clearer [13]. It comes with a set of tools, according to the user requirements; one of them being QtDesigner which has an interface very easy to use and design '.ui' files.

System Configuration

Table 3 describes the ideal system configuration for the efficient and smooth functioning of the automation tool.

Table 3: System configuration for smooth functioning of the automation tool.

S. No.	Requirement Type	Configuration
1	GPU	NVidia
2	Workstation	High-end
3	Installed RAM	16.0 GB or more
4	System Type	64-bit operating system, x64-based processor

Methodology

The existing workflow by Filipponi describes the optimal approach for preprocessing SAR data to correct distortions and granular noises in the imagery [14]. After multiple hits, trials, and tests, a flow of classification and acreage forecasting was devised to bring about the desired outcomes of the data. Figure 2 shows the flow of steps followed to estimate the spread of each crop in the satellite data.

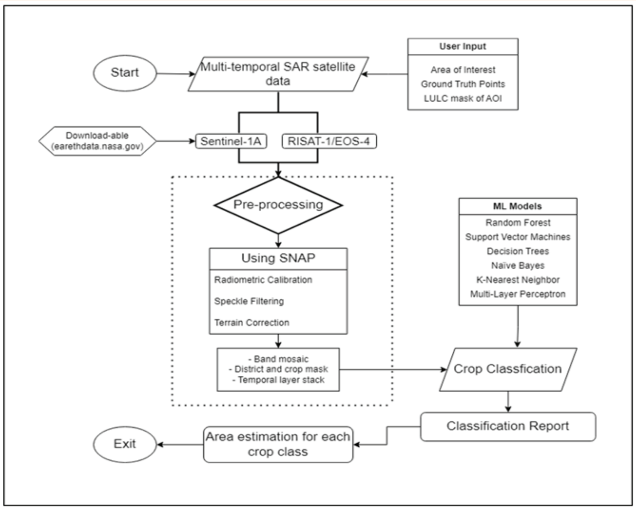


Figure 2: Methodology.

The data processing workflow consists of preprocessing and classification. Once the data has been preprocessed, calculations can be performed on each classified crop class. The automated workflow is described in subsequent sections.

Data Download using ASF-Search

SENTINEL-1A

Due to high temporal resolution, the phenological information of the target crops can be easily captured by the satellite through the multi-date analysis. The satellite captures imagery at a high spatial resolution of 10m and a temporal resolution of 12 days. The high temporal and spatial resolution of the Sentinel-1 data makes it perfect for the study of the monitoring of crops. The satellite operates at a central wavelength of 5.04 GHz, at a spatial resolution of 10m, and a swath of 250 Km.

In this study, the data download is automated using a Python module called asf-search which can be used to fetch the orbit number from the date of pass input by the user and calculate the relative orbit number using the formula:

$$\text{Relative Orbit Number} = \text{mod}\,(\text{Absolute Orbit Number orbit} - 73,\ 175) + 1$$

It is observed that Relative orbit numbers for the images with the same paths (12 days' time gap) are the same. Using the calculated relative orbit number, all the tiles from the input range of dates are downloaded to the setup path. The user is provided with the choice to either download the data with the input of the area of interest shapefile or browse the data files from the system. Data availability for EOS-4 satellite imagery is yet to be made public by its organization.

Preprocessing using SNAP-Py

Preprocessing for both SENTINEL-1A and EOS-4 has been approached in different ways. The Sentinel-1A data is available as Single Look Complex (SLC), and Ground Range Detected (GRD). For the current study, the GRD product of Sentinel-1A was employed and further pre-processed. The GRD data products have already been focused, multi-looked, and projected in the ground range. Further preprocessing of the data was carried out to obtain the final backscatter image and values. The data were pre-processed using the following steps:

Copyright © : Goyal A

Speckle Filter using Lee Filter

The speckle in the SAR imagery is the granular noise, which is mainly because of the interference of waves reflected from many elementary scatters. Speckle filtering is the process of reducing the granular noise present in the imagery to increase image quality. This step is performed before radiometric calibration and terrain correction, so the speckle present in the image does not get propagated further. In the current study, the Lee filter was used to perform speckle filtering using a 3X3 window size [15]. The large window size of 5X5 and 7X7 was avoided as at these window sizes small information is lost which can further affect the accuracy of classification.

Radiometric Calibration

The radiometric calibration process is done to convert the digital pixel values to radiometrically calibrated backscatter values [16]. To calibrate EOS-4 data, instead of using SNAP features, a formula to transform the pixel values is used

$$DN' = DN^2 \frac{\sin(i)}{10^{\frac{k}{10}}}$$

Where
DN = backscatter value in the raw satellite image
i = incidence angle
k = calibration constant (HH or HV)
DN' = modified value of the pixel

Unit conversion of pixel values to decibels is also done here in both types of data using the formula:

$$DN'' = 10 * \log 10(DN')$$

Where
DN'' = pixel value in decibel

Range Doppler Terrain Correction using SRTM 1 Sec (30m) HGT DEM Data

The SAR data is majorly captured at a slanting angle, i.e., greater than 0°, therefore, there can be few distortions corresponding to a side-looking angle. Terrain correction is therefore carried out to eliminate these distortions and match the imagery to the actual effective world. These distortions are majorly caused by foreshortening and shadow using Digital Elevation Models (DEM). The publicly available SRTM DEM data of 30m resolution was used to carry out Range Doppler terrain correction [17].

Feature Engineering

After the preprocessing, images from common date of the pass are seamlessly aligned to form a single mosaic. The mosaic then undergoes an extract by mask process using the Crop Mask of the area of interest which filters out the actual overall agricultural land in that area. It helps in cleaning the data, and makes it less computationally complex for the machine learning model. After all the preprocessed images are mosaiced, and crop masked, a single multi-layered stack is created to make it easier for the classification, and extraction of backscatter coefficients from the images for calculating the zonal statistics of each crop class. In this tool, only mean statistic is used to plot the backscatter coefficient against all dates of passes for each crop class using the ground truth data.

Training data

Ground truth and land use categories were collected so that ground truth information is widely spread and represents all land use categories within the region of interest or district. The ground truth information for both target and non-target crops was collected using MapPad software. The Polygon shapefiles for each of the fields for target and non-target crops were collected. To perform supervised classification, good quality Ground Truth (GT) points or training data are required. GT points are the polygon or point features using which supervised classification classifies the images into different categories. The spread and number of training points play a significant role in supervised classification. As a thumb rule, the number of points for the optimum number of training points should preferably be 10 times the number of variables used in the classification. The larger the size of the training sample, the more the spread of training points, and the higher the accuracy. Therefore, theoretically, the total training points to be collected should equal 10 or 100 times the number of variables. For the current study, the number of variables differs with the region of interest. A common understanding is minimum of 10 variables will be present in each district therefore, the total points that can be collected are equal to approximately 100 points for each crop type. The higher number and spread over the district will capture the geographical variability in the crop. All the non-target classes in training samples should be present otherwise there is a chance of misclassification. The Ground truth points collected are split into training and validation points. Out of the total points, 75% of the samples were used for training the model and 25% of the samples were taken to validate the final classification result [18,19].

The spatial distribution of training points also matters to achieve a higher classification accuracy. In different parts of the AOI, the classes or the crop in this case may have some different variations and complications. Therefore, to better capture the dynamics of crops with varying sowing and harvesting times across the district, the spatial distribution of the training points should be well distributed across the region of interest.

Machine Learning-based Classification

The pre-processed image obtained from the above steps was further used for classification. In the proposed study, ML algorithms have been used for image classification. Many of them have been developed over the past decade to carry out the land use land cover classification [20,21]. Out of all the available machine learning, Random Forest, Support Vector Machine, and K-Nearest Neighbor have gained much popularity as these algorithms are insensitive to noise data which makes them convenient to use in unbalanced data [22]. Here, the following algorithms have been used:

a) Random Forests
b) K-Nearest Neighbors
c) Support Vector Machines
d) Decision Trees
e) Naïve Bayes
f) Multi-Layer Perceptron

These classifiers are the most utilized in Machine learning-based classification studies. Table 4 Defines the hyperparameter tuning that has been done to enhance the performance of the mentioned algorithms.

Table 4: Hyperparameter tuning variables and their values of the used ML classifiers.

S. No.	Algorithm/ Classifier	Tuned Hyperparameters	Used Values
1	Random Forests	Number of features to consider at every split	['sqrt'] [1000,2000] [3, 5] [1, 2] [True, False]
		Maximum number of levels in the tree	
		Minimum number of samples required to split a node	
		Minimum number of samples required at each leaf node	
		Method of selecting samples for training each tree	
2	K-Nearest Neighbors	Leaf size	[1,50] [1,30] [1,2]
		Number of neighbors	
		Number of candidates	
3	Support Vector Machines	Cost	[0.1, 1, 5, 10] [1,0.1,0.01,0.001] [1,0.1,0.01,0.001] ['rbf', 'poly', 'sigmoid']
		Gamma	
		Kernel	
4	Decision Trees	Splitter	["best", "random"] [1,3,5,7,9,11,12] [1,2,3,4,5,6,7,8,9,10] [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9] ["log2", "sqrt", None] [None,10, 20, 30, 40, 50, 60, 70, 80, 90]
		Maximum depth	
		Minimum sample leaf	
		Minimum weight fraction leaf	
		Maximum features	
		Maximum leaf nodes	

5	Naïve Bayes	Priors	[None, [0.1,] * len(number_of_class)] [1e-9, 1e-6, 1e-12]
		Var smoothing	
6	Multi-Layer Perceptron	Hidden Layer Size	[(10,30,10), (20,)] ['tanh', 'relu'] ['sgd', 'adam'] [0.0001, 0.05] ['constant', 'adaptive']
		Activation function	
		Solver	
		Alpha	
		Learning Rate	

GUI and Testing

Graphical User Interface

A user-friendly graphical user interface was designed to ensure effective communication between the user and the system for clearer inputs to the software. To set up the inputs, the main window i.e. Figure 3. is displayed. The GUI flow consists of a set of inputs, which includes choosing data type, operation type, and browsing AOI, ground truth data, and crop mask data. Figure 4 Shows a detailed flow of user inputs for the GUI [23-26].

Figure 3: GUI window to enter user inputs.

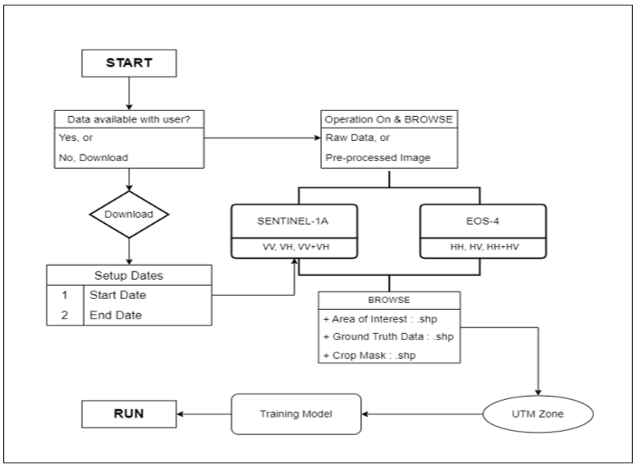


Figure 4: GUI flow for user inputs.

On clicking the DOWNLOAD button, the user can select the start and end date of Sentinel-1A data to be downloaded.

Experimentation and Testing on Sample Data

The current test has been carried out with data from the Aligarh district of Uttar Pradesh, located between latitudes 27°57'N and 28°18'N and longitudes 77°48'E and 78°61'E, in the southeastern part of the state. E. The district is 193 meters above the mean sea level. The great rivers Ganga and Yamuna, which come from the northeast and northwest sides, respectively, border the district. From the northwest, the Palwal district of Haryana, from the northeast, Badaun, from the north, Bulandshahar, from the north, Mathura, from the west to the southeast, Hathars, and from the south and east, Etah. The district's median annual rainfall is 662.8 mm (Department of Agriculture Cooperation & Farmers Welfare, 2019). The district's typical temperature is 25.2°C. The 2011 census revealed that Aligarh has a population of roughly 3.67 million (Census of India, 2011). The district's reported total area is 371300 ha, of which around 304000 ha are net sown (Agriculture, Cooperation and Farmers Welfare, 2012) as shown in Figure 5 [27-31].

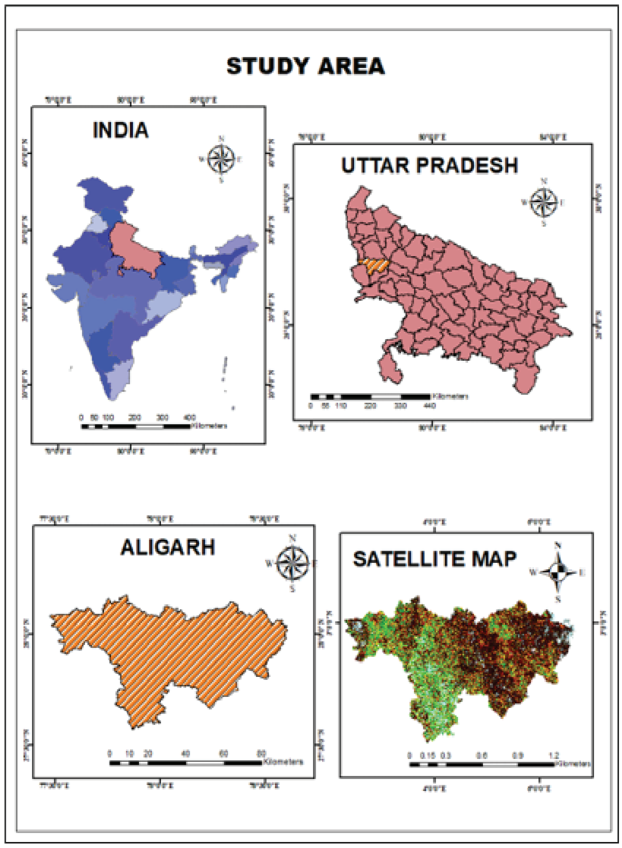
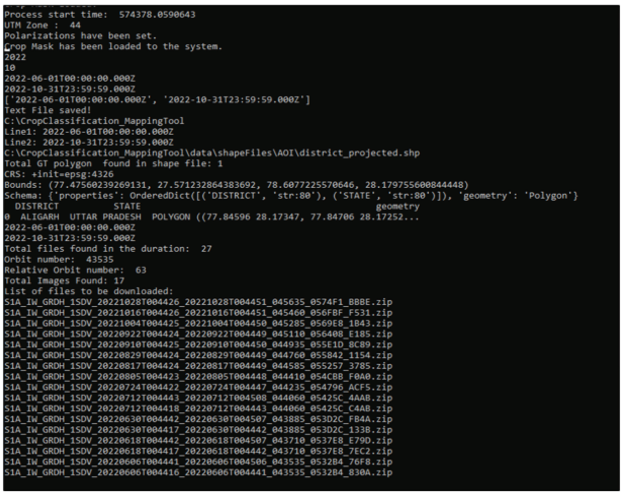


Figure 5: The study area for testing.

Data Download

While downloading, a complete list of the data to be collected is displayed in the command line, along with the relative orbit of all the data files as shown in Figure 6 [32,33].



```
Process start time:  574378.0590643
UTM Zone :  44
Polarizations have been set.
Crop Mask has been loaded to the system.
2022
10
2022-06-01T00:00:00.000Z
2022-10-31T23:59:59.000Z
['2022-06-01T00:00:00.000Z', '2022-10-31T23:59:59.000Z']
Text File saved!
C:\CropClassification_MappingTool
Line1: 2022-06-01T00:00:00.000Z
Line2: 2022-10-31T23:59:59.000Z
C:\CropClassification_MappingTool\data\shapeFiles\AOI\district_projected.shp
Total GT polygon  found in shape file: 1
CRS: +init=epsg:4326
Bounds: (77.47560239269131, 27.571232864383692, 78.6077225570646, 28.179755600844448)
Schema: {'properties': OrderedDict([('DISTRICT', 'str:80'), ('STATE', 'str:80')]), 'geometry': 'Polygon'}
  DISTRICT         STATE                                       geometry
0  ALIGARH  UTTAR PRADESH  POLYGON ((77.84596 28.17347, 77.84706 28.17252...
2022-06-01T00:00:00.000Z
2022-10-31T23:59:59.000Z
Total files found in the duration:  27
Orbit number:  43535
Relative Orbit number:  63
Total Images Found: 17
List of files to be downloaded:
S1A_IW_GRDH_1SDV_20221028T004426_20221028T004451_045635_0574F1_BBBE.zip
S1A_IW_GRDH_1SDV_20221016T004426_20221016T004451_045460_056FBF_F531.zip
S1A_IW_GRDH_1SDV_20221004T004425_20221004T004450_045285_0569E8_1B43.zip
S1A_IW_GRDH_1SDV_20220922T004424_20220922T004449_045110_056408_E185.zip
S1A_IW_GRDH_1SDV_20220910T004425_20220910T004450_044935_055E1D_8C89.zip
S1A_IW_GRDH_1SDV_20220829T004424_20220829T004449_044760_055842_1154.zip
S1A_IW_GRDH_1SDV_20220817T004424_20220817T004449_044585_055257_3785.zip
S1A_IW_GRDH_1SDV_20220805T004423_20220805T004448_044410_054CBB_F0A0.zip
S1A_IW_GRDH_1SDV_20220724T004422_20220724T004447_044235_054796_ACF5.zip
S1A_IW_GRDH_1SDV_20220712T004443_20220712T004508_044060_05425C_4AAB.zip
S1A_IW_GRDH_1SDV_20220712T004418_20220712T004443_044060_05425C_C4AB.zip
S1A_IW_GRDH_1SDV_20220630T004442_20220630T004507_043885_053D2C_FB4A.zip
S1A_IW_GRDH_1SDV_20220630T004417_20220630T004442_043885_053D2C_133B.zip
S1A_IW_GRDH_1SDV_20220618T004442_20220618T004507_043710_0537E8_E79D.zip
S1A_IW_GRDH_1SDV_20220618T004417_20220618T004442_043710_0537E8_7EC2.zip
S1A_IW_GRDH_1SDV_20220606T004441_20220606T004506_043535_0532B4_76F8.zip
S1A_IW_GRDH_1SDV_20220606T004416_20220606T004441_043535_0532B4_830A.zip
```

Figure 6: Command line output for data download.

Backscatter Curve

After preprocessing, a backscatter curve is generated using the layer-stacked preprocessed image and the ground truth points for each crop class over the temporal data collected. Figure 7 shows the curve generated for the data from June 2022 to October 2022, and crop classes Arhar, Bajra, Cotton, Fallow Land, Jowar, Maize, Paddy, and Sugarcane. The more the backscatter coefficient value the greater the probability of the crop class being present at the field during that time [34-37].

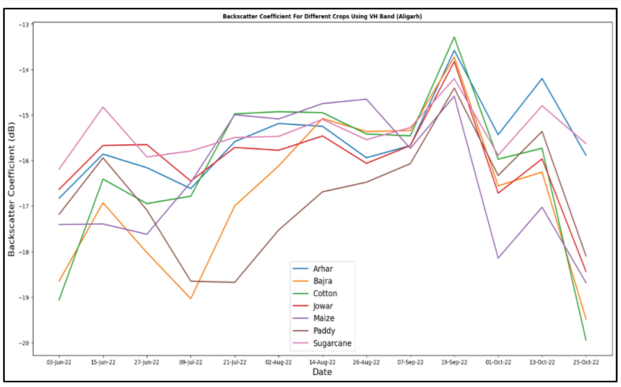
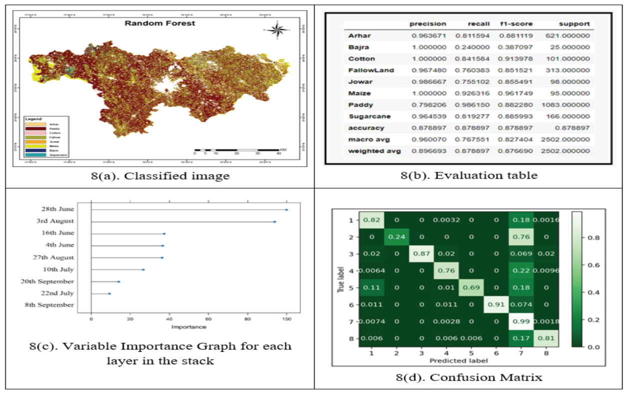


Figure 7: The backscatter curve is generated by the automation tool.

Classification Report

After preprocessing, the image is classified using the user-selected machine learning algorithm. Figure 8 shows the classification results for a set of crop classes over the area of the Aligarh district generated as a part of an additional automated 4-page report at the end to summarize a set of outputs. Figure 8(a) shows the resultant classified image, the spatial representation of classified values over the area, 8(b) displays an evaluation table of performance metrics for the machine learning model, 8(c) shows a variable importance graph to provide a plot of most significantly contributing image to the ML model in the layer stack, and 8(d) plots a confusion matrix to describe the difference between predicted and actual values of the classifier for each crop class.



	precision	recall	f1-score	support
Arhar	0.963671	0.811594	0.881119	621.000000
Bajra	1.000000	0.240000	0.387097	25.000000
Cotton	1.000000	0.841584	0.913978	101.000000
FallowLand	0.967480	0.760383	0.851521	313.000000
Jowar	0.986667	0.755102	0.855491	98.000000
Maize	1.000000	0.926316	0.961749	95.000000
Paddy	0.798206	0.986150	0.882280	1083.000000
Sugarcane	0.964539	0.819277	0.885993	166.000000
accuracy	0.878897	0.878897	0.878897	0.878897
macro avg	0.960070	0.767551	0.827404	2502.000000
weighted avg	0.896693	0.878897	0.876690	2502.000000

Figure 8: Classification results generated by the automation tool.

Conclusion

This study presents a comprehensive and fully automated system for multi-crop classification and acreage estimation using multi-temporal SAR datasets and machine learning algorithms. The GUI-based tool effectively streamlines complex geospatial workflows - ranging from data acquisition and preprocessing to classification and reporting - thereby reducing manual intervention and operational overhead. With classification accuracies exceeding 90% for key crops, the system demonstrates notable reliability across diverse agro-ecological regions. Compared to traditional methods, the proposed system improves predictive model performance, significantly reduces processing time from several weeks to a few days, and optimizes memory usage based on user-defined parameters. Its intuitive design and robust backend make it scalable, reproducible, and user-friendly, even for non-expert stakeholders. Looking ahead, the integration of advanced deep learning techniques—such as Convolutional Neural Networks (CNNs) -alongside a richer set of annotated ground truth data holds promise for further enhancing classification accuracy and spatial detail extraction. Such extensions can enable more nuanced modeling of crop phenology and support a wider range of applications in agricultural monitoring, yield forecasting, and climate resilience planning. Overall, the study underscores the transformative potential of automated SAR-based analytics in operational crop monitoring, offering a scalable solution that aligns with the goals of digital agriculture and evidence-based policymaking.

Acknowledgement

The authors are grateful to Regional Remote Sensing Centre-North, National Remote Sensing Centre, Indian Space Research Organization, Government of India, for support and sincerely acknowledge the technical and logistic provided by RRSC-North. This paper includes information generated in a project on "Automation of In-Season Crop Acreage Forecasting".

References

1. FAO (2017) The future of food and agriculture – Trends and challenges. Food and Agriculture Organization of the United Nations.
2. Lobell DB, Thau D, Seifert C, Engle E, Little B (2015) A scalable satellite-based crop yield mapper. Remote Sensing of Environment 164: 324-333.
3. Inglada J, Vincent A, Arias M, Marais SC (2015) Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. Remote Sensing 7(12): 15612-15638.
4. Veloso A, Ceschia M, Inglada J, Dedieu G, Sepulcre G, et al. (2017) Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2 data: A case study in France. Remote Sensing of Environment 199: 415-426.

5. Mandal D, Pradhan R, Patel NR (2021) In-season rice crop mapping using dual-polarized Sentinel-1 SAR data and machine learning approach. Remote Sensing Applications: Society and Environment 21, 100454.

6. Dutta S, Patel NR, Saha SK, Saha S (2020) Assessment of monsoon rice acreage using multi-temporal Sentinel-1 SAR data and cloud computing. GIScience & Remote Sensing 57(5): 652-669.

7. Mekonnen AD, Vrieling A, Herold M, Tadesse T (2021) Pre-processing of Sentinel-1 SAR data for crop monitoring in Africa. Remote Sensing 13(2): 272.

8. Belgiu M, Drăguţ L (2016) Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114: 24-31.

9. Pelletier C, Webb GI, Petitjean F (2019) Temporal convolutional neural network for the classification of satellite image time series. Remote Sensing 11(5): 523.

10. Torres R, Snoeij P, Geudtner D, Bibby D, Davidson M, et al. (2012) GMES Sentinel-1 mission. Remote sensing of environment 120: 9-24.

11. Khabbazan S, Vermunt P, Dunne SS, Arntz LR, Marinetti C, et al. (2019) Crop monitoring using Sentinel-1 data: A case study from The Netherlands. Remote Sensing 11(16): 1887.

12. Arko SA, Allen AR, Dixon IR (2014) Near-Real-Time, Global Radar Data at the Alaska Satellite Facility DAAC from NASA's SMAP Satellite. AGU Fall Meeting Abstracts.

13. Willman J, Joshua W (2021) Overview of pyqt5. Modern PyQt: Create GUI Applications for Project Management, Computer Vision, and Data Analysis p. 1-42.

14. Filipponi F (2019) Sentinel-1 GRD preprocessing workflow. International Electronic Conference on Remote Sensing.

15. Sen LJ (1980) Digital image enhancement and noise filtering by use of local statistics." IEEE transactions on pattern analysis and machine intelligence 2(2): 165-168.

16. Mohit K, Akash G, Neeti N, Maity S, Mukesh K, et al. (2022) Machine learning-based meta-classifier for Kharif Bajra (pearl millet) discrimination in the mixed cropping environment using multi-temporal SAR data. Geocarto International 37(27): 16671-16686.

17. Abdikan S, Sanli FB, Ustuner M, Calò F (2016) Land cover mapping using sentinel-1 SAR data." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 41: 757-761.

18. Patel P, Srivastava HS (2013) Ground truth planning for synthetic aperture radar (SAR): addressing various challenges using statistical approach. International Journal of Advancement in Remote Sensing, GIS and Geography 1(2): 1-17.

19. Shu ZZ, Boerner WM, Sato M (2004) Development of a ground-based polarimetric broadband SAR system for noninvasive ground-truth validation in vegetation monitoring. IEEE Transactions on Geoscience and Remote Sensing 42(9): 1803-1810.

20. Lu X, Zhang H, Wang C, Zhang B, Liu M (2018) Crop classification based on temporal information using sentinel-1 SAR time-series data. Remote Sensing 11(1): 53.

21. Talukdar S, Talukdar S, Singha P, Mahato S, Shahfahad, et al. (2001) Land-use land-cover classification by machine learning classifiers for satellite observations - A review. Remote Sensing 12(7): 1135.

22. Leo B (2001) Random forests. Machine learning 45: 5-32.

23. Orynbaikyzy A, Ursula G, Christopher C (2019) Crop type classification using a combination of optical and radar remote sensing data: A review. International Journal of Remote Sensing 40(17): 6553-6595.

24. Viskovic L, Ivana NK, Toni M (2019) Crop classification using multi-spectral and multitemporal satellite imagery with machine learning. 2019 International conference on software, telecommunications and computer networks (SoftCOM). IEEE.

25. Arias M, Campo BMA, Jesús ÁM (2018) Crop type mapping based on Sentinel-1 backscatter time series. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE.

26. Arias M, Campo BMÁ, Mozos JÁ (2020) Crop classification based on temporal signatures of Sentinel-1 observations over Navarre province, Spain. Remote Sensing 12(2): 278.

27. Subrahmanyam KV, Bothale RV, Swapna M, Chauhan P (2022) Deciphering the signatures of convective rain cells using simultaneous observations from C-band Synthetic Aperture Radar onboard EOS-4 satellite and GPM measurements. Authorea Preprints.

28. https://snap.stanford.edu/snappy

29. https://towardsdatascience.com/getting-started-with-snap-toolbox-in-python-89e33594fa04

30. https://senbox.atlassian.net/wiki/spaces/SNAP/pages/50855941/Configure+Python+to+use+the+SNAP-Python+snappy+interface

31. https://gdal.org/gdal.pdf

32. https://pypi.org/project/asf-search/

33. https://scikit-learn.org/stable/supervised_learning.html#supervised-learning

34. Yang L, Xingmin M, Xiaoqiang Z (2011) SRTM DEM and its application advances." International Journal of Remote Sensing 32(14): 3875-3896.

35. https://medium.com/geekculture/bulk-download-sentinel-1-sar-data-d180ec0bfac1

36. https://dges.carleton.ca/courses/IntroSAR/LAB%203A%20%20Carleton%20SAR%20Training%20-%20SAR_processing_FINAL.pdf

37. https://www.nrsc.gov.in/sites/default/files/pdf/ebooks/UIM-2022/uim_1.pdf