Journal of Systems Engineering and Electronics (ISSN NO: 1671-1793) Volume 34 ISSUE 12 2024

A Machine Learning Approach for Crop Classification Using Sentinel -1 Data

Mr. V. Praveen Kumar
Associate Professor
Department of ECE
Sree Venkateswara College of Engineering ,Northrajupalem, Nellore, A.P, India.

Dr. P.Giri Prasad
Professor &HOD
Department of ECE
Sree Venkateswara College of Engineering ,Northrajupalem, Nellore, A.P, India.

Abstract— The national economy of India is greatly influenced by agriculture, and the majority of important choices are based on agricultural statistics. The latest generation of satellite and aerial imaging sensors captures massive amounts of Earth's images with excellent spatial, spectral, and temporal resolution. These sensors are widely used for precision agriculture, urban planning, and natural disaster monitoring. In this work, crop classification of Sentinel-1 satellite temporal remote sensing image data has been done using machine learning models in satellite remote sensing image data processing. A village called Pendlimarri and T. Velamvaripalli is located in the Andhra Pradesh state of India's Kadapa district. The availability of ground truth and the variety of agricultural crops in this area are the primary factors in the selection.

Keywords-Satellite based imaging sensors, Spatial ,Spectral, Temporal resolution, Precision farming, Machine learning models, Sentinal-1 Satellite

I. INTRODUCTION

1.1 Introduction

Agriculture holds a critical role in Indian society due to its impact on the economy, employment, food security, self-reliance, and overall well-being. Key stakeholders, such as producers, resource managers, marketers, financiers, and government officials, recognize the importance of fast and reliable information on crop area and production for making informed decisions. With the global shift toward market economies, the demand for dependable agricultural data has risen significantly. Remote sensing offers a practical solution by providing decision-makers with vital crop-related information through crop information systems.

Currently, remote sensing is the only technology capable of delivering precise and timely data on crop inventories. Among the various remote sensing systems, orbiting satellites stand out by offering repeat coverage. Space-based sensors operating in the visible and near-infrared spectrum have achieved reasonable success in crop identification and area estimation in recent decades (Brisco et al., 1998; Dadhwal et al., 2000; Fisette et al., 2005). However, optical sensors, which rely on sunlight, face challenges in cloudy or rainy conditions, limiting their ability to capture cloud-free images for in-season crop monitoring (Schuster et al., 2011). This highlights the need for a remote sensing system that can operate reliably in all weather conditions, making radar a promising alternative. Studies (Moran et al., 2011; Chang et al., 2009; McNairn et al., 2013) have demonstrated the strong potential of Synthetic Aperture Radar (SAR) imagery for estimating agricultural acreage and monitoring crops.

Despite these advancements, crop classification accuracy with current SAR interpretation techniques often falls short of decision-making requirements. To improve accuracy in crop identification and area estimation, we need to: (a) deepen our understanding of crop and soil properties influencing radar backscatter during the growing season, (b) select effective techniques for extracting crop data from SAR imagery, and (c) analyze multitemporal SAR data for crop identification. Building on this framework, the present study focuses on assessing multitemporal SAR imagery to estimate the crop area and identify crops like cotton and bananas in the villages of Pendilimari and Velamvaripalli in the Kadapa district. The purpose of this work is to evaluate Sentinel-1 backscatter's capabilities for monitoring vegetation changes in more detail. In order to achieve success, we considered utilizing appropriate classification algorithms that can take advantage of the information present in both multi-polarization and multi-temporal SAR observations. schemes for classifying different kinds of schemes according to the level of each pixel or item. A group of pixels with similar characteristics is processed using an object-based approach, and this method shows great potential when working with high-resolution images. For automatic supervised classification, many options can be selected such as: image type, sensor, time scene, training data sample, classification algorithm.

1.2 Problem statement

The project's premise is that farmers are producing more than needed and requesting the government to provide better prices for their produce. However, it is unfeasible for the government to do so as it would lead to wastage of the surplus, which is not beneficial. Therefore, utilizing the freely available sentinel data helps the government in estimating the yield and categorizing the crops.

MATERIALS AND METHODS

2.1 Study Area

Fig 2.1 Study Area

Fig 2.2: Andhra Pradesh

Fig 2.3: Pendlimarri village

Fig 2.4: Velamvaripalli village

Pendlimarri, a village in the Kadapa district of Andhra Pradesh, falls under the Pendlimarri mandal within the Kadapa revenue division. Its geographical coordinates are 14.4491° N and 78.6314° E.

T.velamvaripalli, a village panchayat in Cuddapah district of Andhra Pradesh, India, can be located at the geocoordinates 14.48 latitude and 78.81 longitude. Preserving the accuracy of the specifications.

2.2 Data Used

Sentinel1-A was a C-band radar system that was used systematically for crop inventory. It had VV (Vertical–Vertical) and VH (Vertical–Horizontal) polarization, which were obtained at twelve-day intervals. Sentinel-1A features four common operating modes that are intended to facilitate system interoperability. For this study, the interferometric wide (IW) swath mode (1) of High Resolution (HR) yielded the Level-1 ground range (GRD) result. To provide complete coverage during the groundnut and cotton crop growth season, satellite data covering the dates of June 6, 2021, through December 28, 2021, was retrieved. We have gathered ground truth data on cotton and groundnuts from the villages of Velamvaripalli and Pendilimari. Vectors are trained and tested using this data. The information is displayed below..

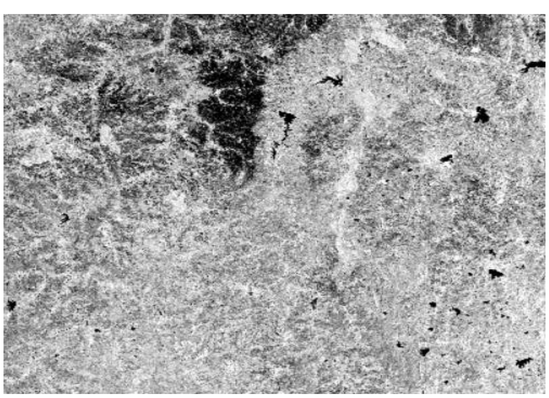


Fig 2.5**:** Sentinel-1 image

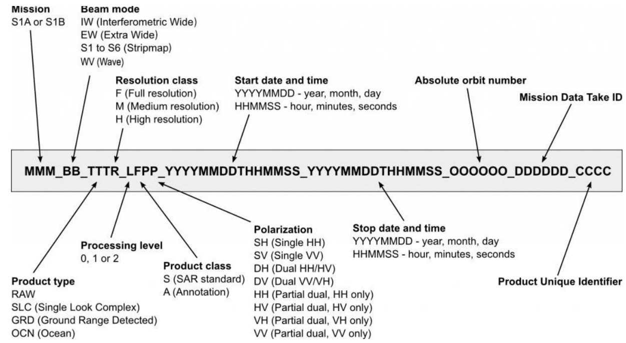


Fig 2.6: Sentinel1 naming convention

Example:

S1A_IW_GRDH_1SDV_20160731T001533_20160731 T001558_012385_0134FE_1 CFC

2.3 Tools Used

2.3.1 SNAP Tool Box

The Sentinel-1 Toolbox (S1TBX) is a suite of processing tools, data product readers and writers, and a display and analysis application designed to manage the extensive data archives from ESA SAR missions, including Sentinel-1, ERS-1 & 2, and ENVISAT. It also supports third-party SAR data from platforms like ALOS PALSAR, TerraSAR-X, COSMO-SkyMed, and RADARSAT-2. The toolbox offers features such as calibration, speckle filtering, coregistration, orthorectification, mosaicking, data conversion, polarimetry, and interferometry.

Array, in collaboration with DLR, Brockmann Consult, and OceanDataLab, is developing the Sentinel-1 Toolbox for ESA.

The SNAP platform looks like this where we can do all the processing steps in easy way which is free of cost.

Features

- Standard design shared by all Toolboxes.
- Detailed region-of-interest definitions for generating statistics and different types of plots.
- Effortless bitmask definition and overlay.
- Flexible band arithmetic through the use of arbitrary mathematical expressions.
- Very fast image display and navigation, capable of handling giga-pixel images.
- Graph Processing Framework (GPF): enables the creation of custom processing chains by users.
- Advanced layer management that enables the addition and manipulation of new overlays.
- Employing ground control points for geo-coding and rectification, ensuring accurate reprojection and ortho-rectification to standard map projections.
- A product library for efficiently scanning and classifying large files, with support for multithreading and multi-core processors. It also includes automatic SRTM DEM download and tile selection, along with

integrated World Wind visualization.

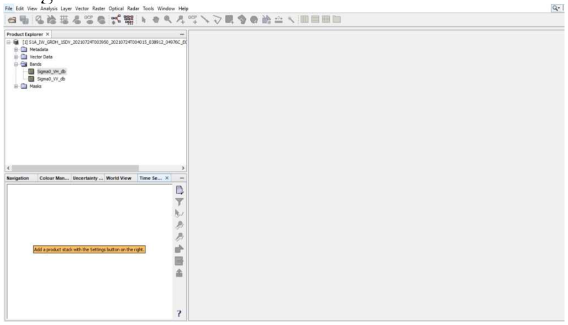


Fig 2.7: SNAP platform

2.3.2 QGIS

QGIS is a free, open-source, cross-platform desktop geographic information system (GIS) application that enables users to view, edit, print, and analyze geospatial data. It serves as GIS software, allowing spatial data to be analyzed and edited, while also enabling the creation and export of graphical maps. QGIS supports raster, vector, and mesh layers, with vector data stored as points, lines, or polygons. It supports various raster image formats and can georeference images.

QGIS supports a variety of industry-standard formats, including shapefiles, personal geodatabases, DXF, MapInfo, PostGIS, and more. It also allows the use of external data through web services, such as Web Map Service (WMS) and Web Feature Service (WFS).

QGIS integrates seamlessly with other open-source GIS tools, such as PostGIS, GRASS GIS, and MapServer. Its functionality can be extended through plugins written in Python or C++, which can perform tasks like geocoding using the Google Geocoding API, executing geoprocessing functions similar to those in ArcGIS, and connecting with databases like PostgreSQL/PostGIS, SpatiaLite, and MySQL. QGIS also works with SAGA GIS and Kosmo. Additionally, QGIS supports displaying multiple layers from different sources or representations of those sources.

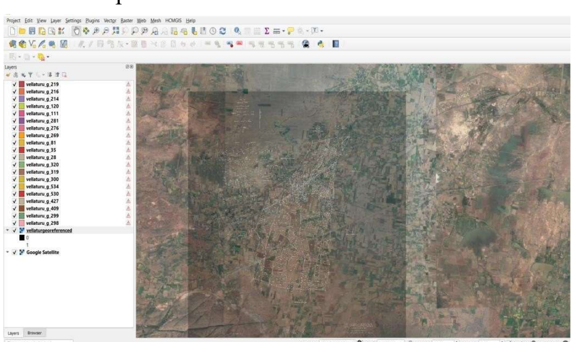


Fig 2.8**:** QGIS Software

2.3.3 Coper Nicus Open Access Hub

Through self-registration, anyone can register online. Self-registration is a quick and simple process. Access privileges to search and download Sentinels products are granted upon registration. Anyone can obtain Sentinel's products for free. By using the Sentinel data, the User is considered to have accepted the Legal Notice on the use of Copernicus Sentinel Data and Service Information, which governs the data made accessible through the Data Hub. Access to the data sources is open and free.

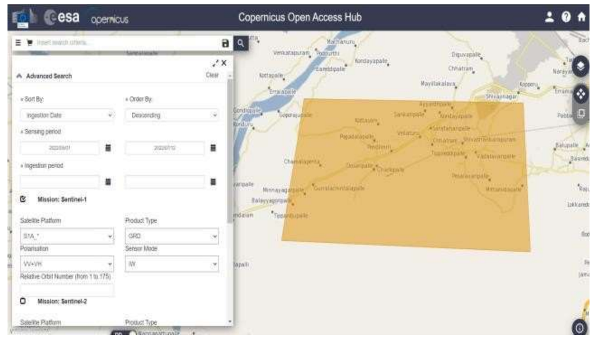


Fig 2.9: Coper nicus open access

II. PROPOSED SYSTEM

Crop classification utilizing multi-temporal sentinel data is the proposed system. This aids in crop productivity estimation and crop classification according to temporal and backscatter characteristics.

3.1 Methodology

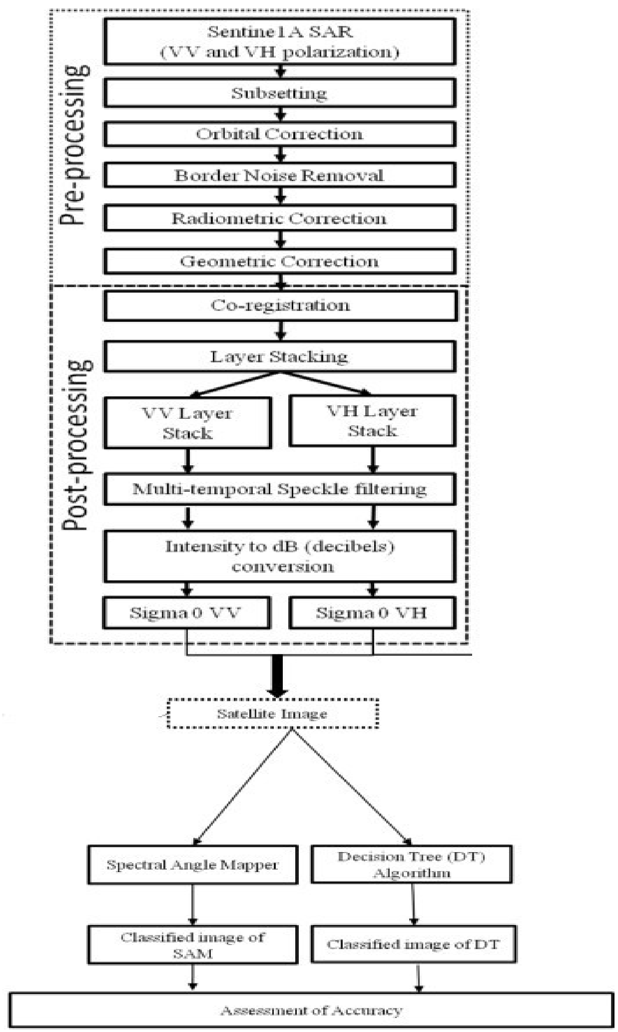


Fig 3.1: Methodology for classification for multi-temporal SAR data

3.2 Preprocessing Techniques

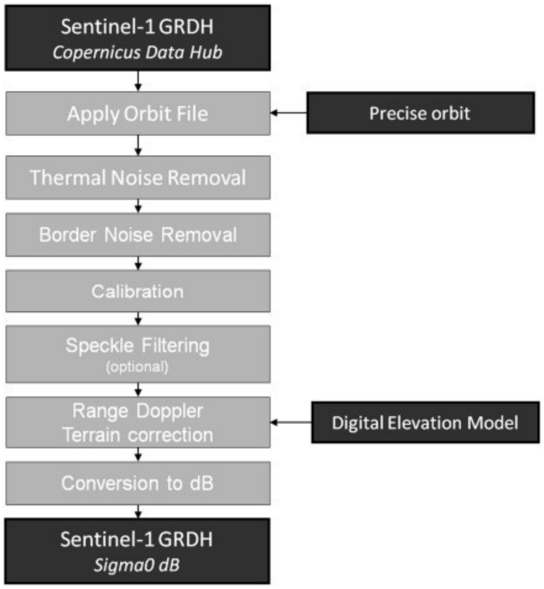


Fig 3.2: Sentinel-1 Ground Range Detected (GRD) preprocessing workflow

3.3 Work Flow

A standard workflow for preprocessing Copernicus Sentinel-1 GRD data is outlined here. This workflow was developed for use within the Sentinel Application Platform (SNAP), a common framework for all Sentinel satellite toolboxes. The processing graph, available in 'xml' format, enables the processing of Sentinel-1 GRD data via the command line graph processing framework, which supports batch processing of large datasets. The preprocessing procedure includes seven steps, each designed to minimize error propagation in subsequent processes, which are detailed in the following subsections. The code required to implement this preprocessing workflow is available on the GitHub repository and in the Supplementary Materials under Computer Code 1.

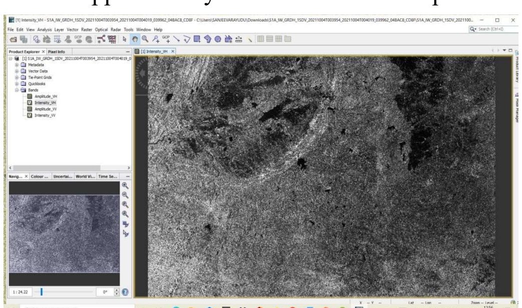


Fig 3.3 Sentinel-1 Image

Launch and Present the Sentinel-1 Picture

1. Launch the SNAP application
2. Select File > Open Product from the File menu in the SNAP interface.
3. Choose the folder containing the data from Sentinel-1.
4. Click on the picture.
5. To see the directories inside the file, double-click the file name.
6. The coverage of the opened image is displayed in the Worldview window, which is located on the lower left corner.Click Intensity_VH twice.
7. Press Intensity_VH twice.

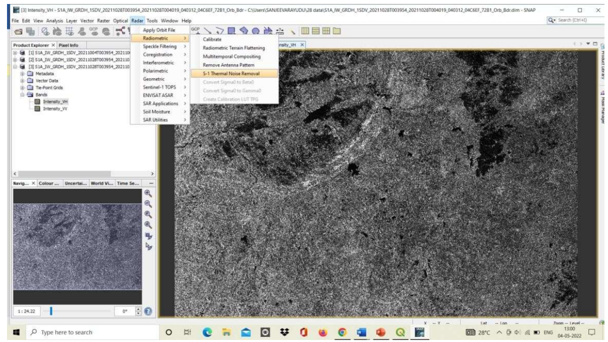
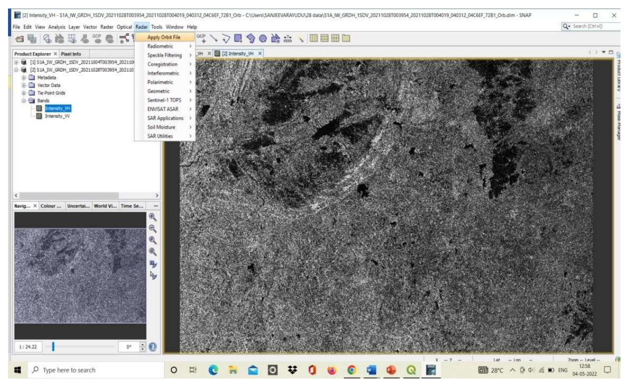


Fig 3.4

3.3.1 Apply Orbit File

Typically, the orbit state vectors provided in the metadata of SAR products lack precision. It takes a few days for the exact orbits of the satellites to be determined, with the precise data being made available days or even weeks after the product is generated. By applying an accurate orbit from SNAP, users can automatically download and update the orbit state vectors for each SAR scene in the product metadata, offering precise details on the satellite's position and velocity.(Fig. 3.5)



3.3.2 Thermal Noise Removal

Additive thermal noise is commonly observed in Sentinel-1 images, particularly in the cross-polarization channel. The treatment of thermal noise helps normalize the backscatter signal across the entire Sentinel-1 image, reducing discontinuities between sub-swaths, especially in multi-swath acquisition modes. This process minimizes the noise impact in the inter-sub-swath texture. Additionally, the thermal noise removal operator in SNAP for Sentinel-1 data not only updates product annotations to allow for the reapplication of the correction but also has the capability to restore the noise signal that may have been removed during the level-1 product development. Each Sentinel-1 level-1 product includes a noise look-up table (LUT) in linear power, which is used to generate calibrated noise profiles that align with the calibrated GRD data.

3.3.3 Border Noise Removal

To compensate for variations in Earth's curvature, the sample start time must be adjusted during the production of level-1 products. Azimuth and range compression can also introduce radiometric artifacts at the image edges. The border noise removal algorithm, which is an operator in SNAP, aims to eliminate erroneous data and low-intensity noise from the edges of scenes.

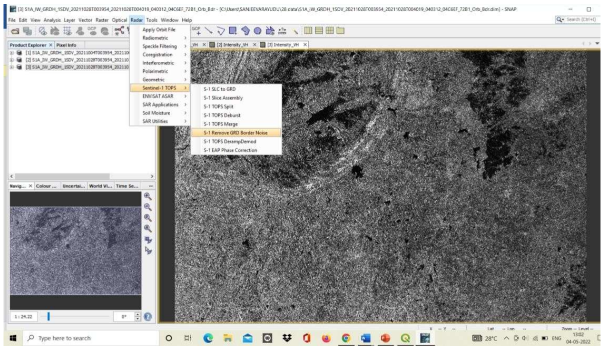


Fig 3.6

3.3.4 Calibration

Calibration is the process of converting digital pixel data into radiometrically calibrated SAR backscatter. The Sentinel-1 GRD product includes the necessary data to apply the calibration equation, with a calibration vector provided as an annotation in the product, making it straightforward to convert image intensity values into sigma nought values. The calibration process incorporates a range-dependent gain, a constant offset, and the absolute calibration constant, in addition to reversing the scaling factor applied during the level-1 product generation. A look-up table (LUT) is used in the preprocessing workflow to generate sigma nought values, providing radiometrically calibrated SAR backscatter relative to the nominally horizontal plane. Sigma represents the radar cross section of a distributed target compared to that predicted for an area of one square meter, reflecting the strength of the reflection in terms of the geometric cross section of a conducting sphere. Sigma nought is heavily influenced by the scattering surface's characteristics, wavelength, polarization, and incidence angle.

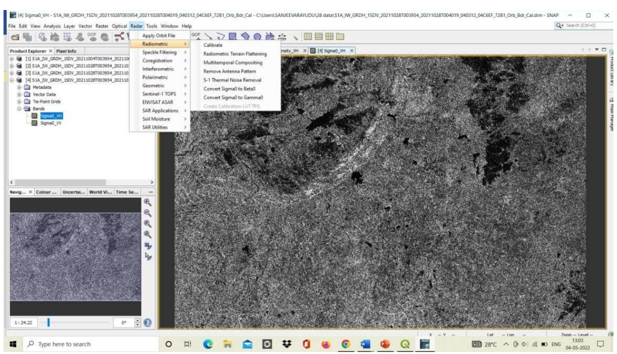


Fig 3.7 Calibrating pixel values

3.3.5 Speckle Filtering

Speckle, which appears as granular noise in SAR images, is caused by the interference of waves reflected from numerous small scatterers. The process of "speckle filtering" reduces this noise to enhance image quality. Applying speckle filtering early in the processing stages prevents the spread of speckle during subsequent procedures, such as terrain correction or conversion to dB. However, when attempting to identify small-scale spatial features or texture within an image, speckle filtering should be avoided, as it may remove important details. The refined Lee filter has demonstrated superior performance in preserving edges, linear features, point targets, and texture information, making it more effective than other single-product speckle filters for visual interpretation. In recent times, multitemporal speckle filters have been created to minimize speckle by utilizing numerous SAR measurements at different times. The 'None' filter type allows you to bypass the speckle filtering phase that is part of the suggested preprocessing strategy. The SNAP single product speckle filter operator now offers one of the following filters: "Boxcar," "Median," "Frost," "Gamma Map," "Lee," "Refined Lee," "Lee Sigma," and "IDAN."

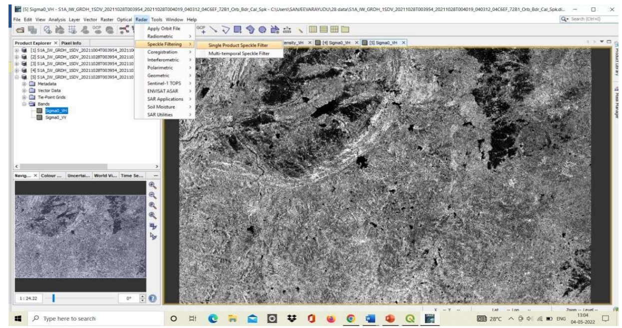


Fig 3.8 Filtering the speckle

3.3.6 Range Doppler Terrain Correction

SAR data often produces images with distortion due to the side-looking geometry, typically captured with a variable viewing angle greater than 0 degrees. Terrain corrections aim to compensate for these distortions and bring the image's geometric representation as close to reality as possible. Range Doppler terrain correction uses a digital elevation model (DEM) to adjust the position of each pixel, correcting for geometric distortions caused by topography, such as shadows and foreshortening. The Range Doppler orthorectification method for geocoding SAR scenes from radar geometry is implemented by the Range Doppler terrain correction operator in SNAP. This operator uses reference DEM data, slant-to-ground range conversion parameters, radar time annotations, and available orbit state vector information from the metadata to determine precise geolocation. Additionally, the target Coordinate Reference System (CRS) can be selected, typically aligning with the UTM zone of overlapping Sentinel-2 granules. The operator allows for choosing the target pixel spacing in the CRS and selecting the image resampling technique. This processing step also enables the spatial snapping of Sentinel-1 GRD data to Sentinel-2 MSI grids, promoting the use of satellite virtual constellations.

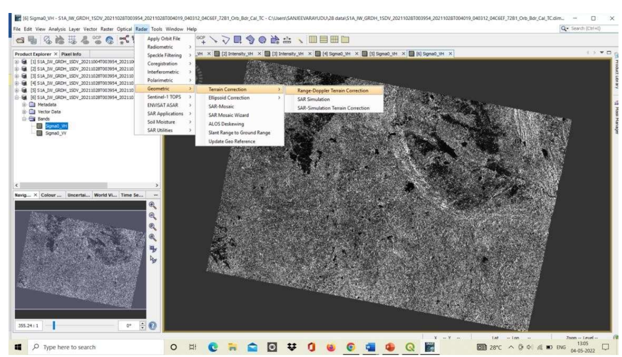
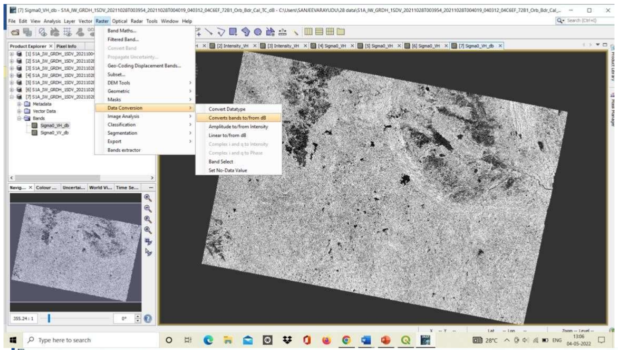


Fig 3.9 Correcting the Terrain

3.3.7 Conversion to dB

The unitless backscatter coefficient is converted to dB through a logarithmic transformation as the final step in the preprocessing process.

Fig 3.10 Converting coefficients to dB



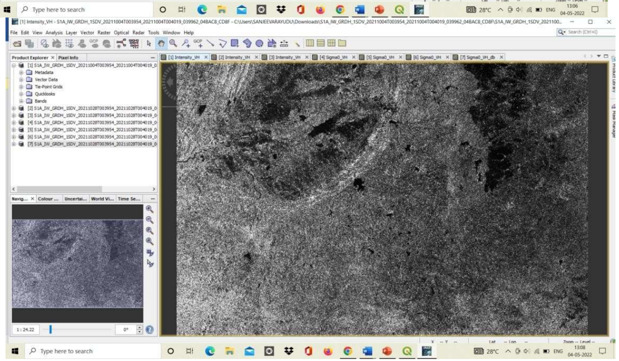


Fig 3.11: Before Preprocessing

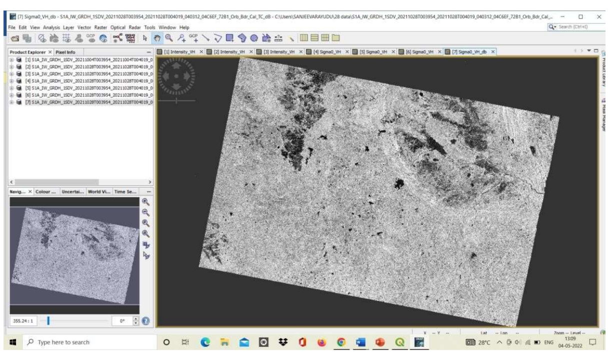


Fig 3.12: After Preprocessing

3.4 Training vectors in QGIS

Fig 3.13: Training vectors

We used QGIS to train the crop vectors from various villages. Snap imports a few of the vectors onto the sentinel image.

The time series data of crop vectors from various villages as shown in the figure below.

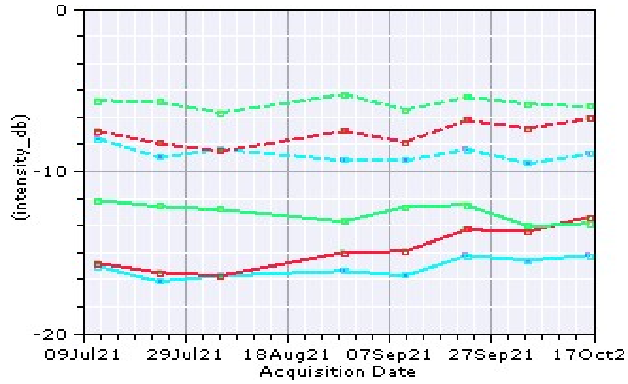


Fig 3.14: Time series data of crops

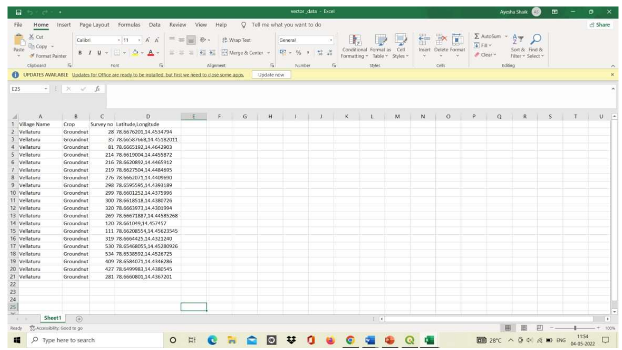


Fig 3.15: Data Collection

The vectors' latitudes and longitudes were entered into an Excel sheet.

In order to verify the outcome, we tested a few of the vectors and ran the machine learning algorithm.

We launch QGIS and open the Sentinel image in order to run the algorithm.

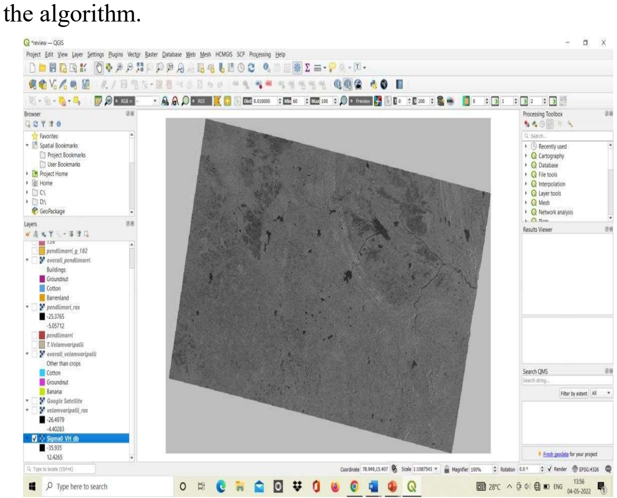


Fig 3.16 Sentinel image

We import the shapefile of the villages on QGIS.

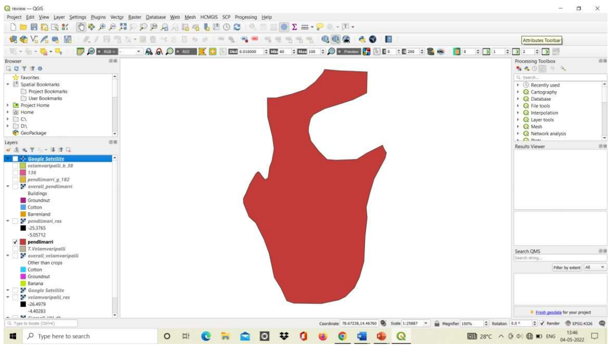


Fig 3.17 Shapefile of Pendilimari village

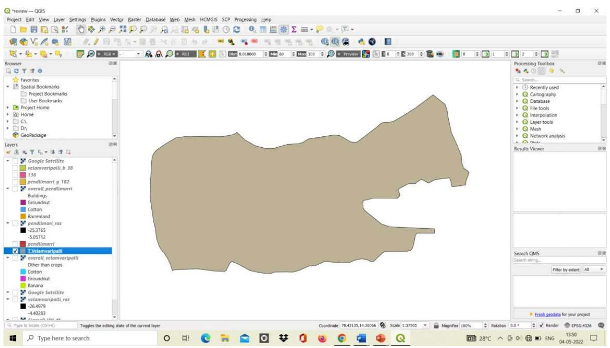


Fig 3.18 Shapefile of Velamvaripalli Village

We extracted the raster of the villages by clipping the raster by mask layer.

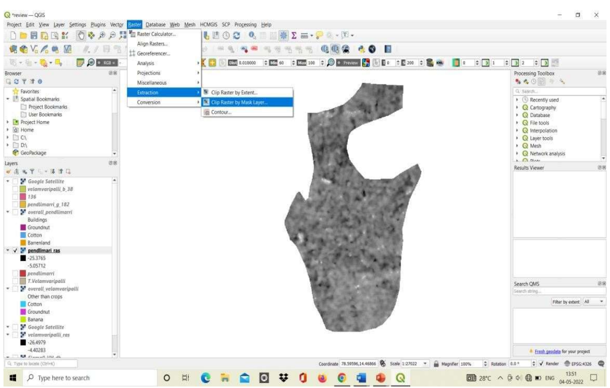


Fig 3.19

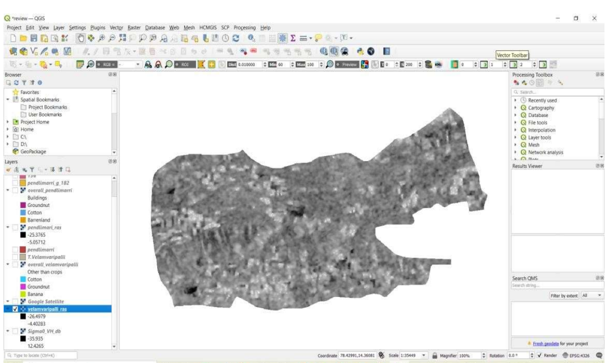


Fig 3.20

Installing the Semi Automatic Classification Plugin is necessary in order to run the Random Forest Classifier Algorithm.

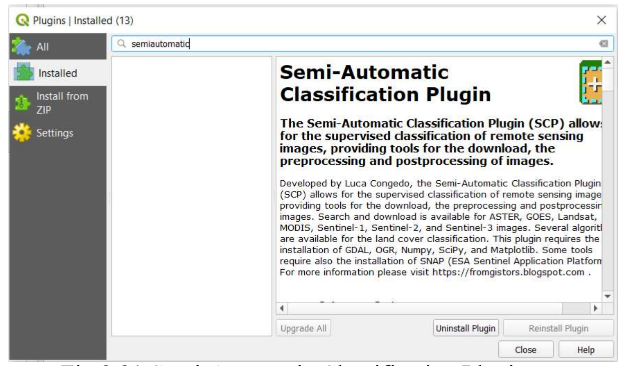


Fig 3.21 Semi-Automatic Classification Plugin

Go to SCP -> Band Processing -> Random Forest.

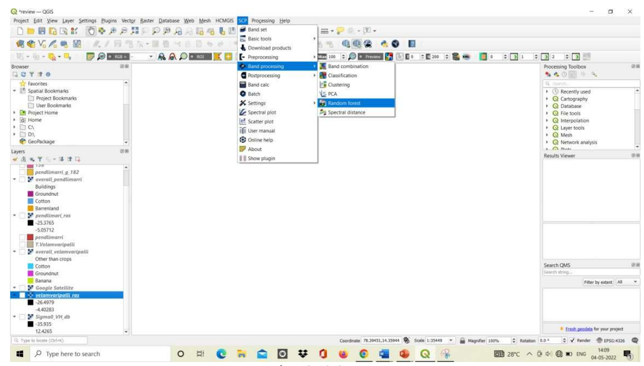


Fig 3.22

Give all the inputs in random forest.

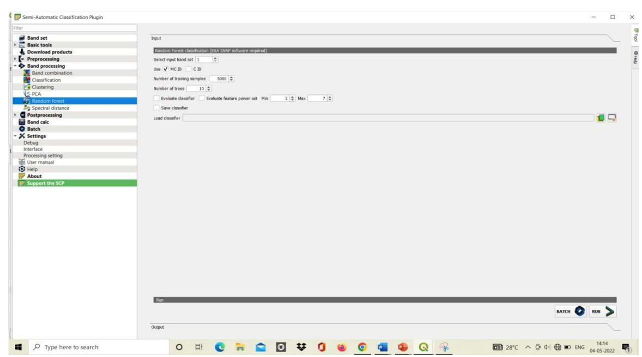


Fig 3.23

After running the algorithm, we get the output as

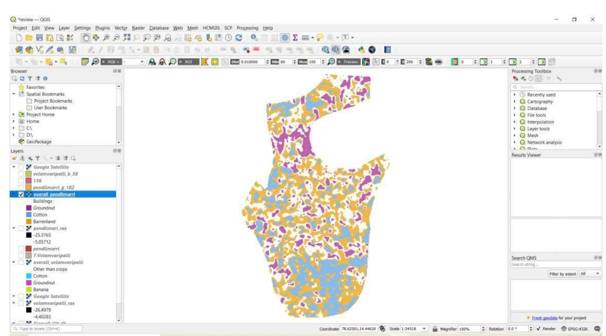


Fig 3.24

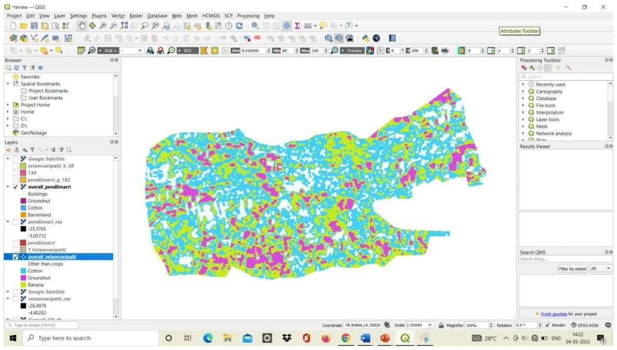


Fig 3.25

We label the classes by checking the vectors in properties.

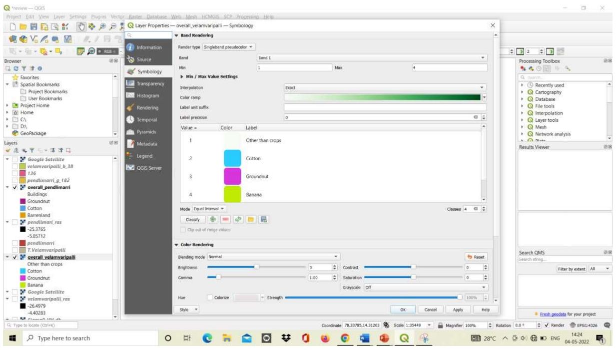


Fig 3.26

We can calculate the area of our classes by using raster layer unique value report.

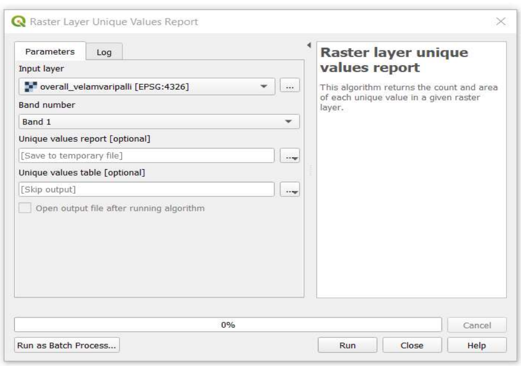


Fig 3.27

We obtain the pixel count of the classes after execution, translate it into square meters, and then add the results.

Analyzed file: C:/Users/SANJEEVARAYUDU/Documents/overall_velamvaripalli.sdat (band 1)

Extent: 78.3437291107500045,14.3038450558499992 : 78.4303266767499991,14.3522642343499989

Projection: EPSG:4326 - WGS 84

Width in pixels: 964 (units per pixel 8.98315e-05)

Height in pixels: 539 (units per pixel 8.98315e-05)

Total pixel count: 519596

NODATA pixel count: 197400

Value	Pixel count	Area (deg²)
1	46153	0.0003724407898974888
2	136206	0.001099141339214728
3	40897	0.0003300264551478257
4	98940	0.0007984159589291605

Fig 3.28

III. MACHINE LEARNING ALGORITHM

Random Forest Algorithm

Random Forest is a machine learning algorithm within the supervised learning framework, used for both classification and regression tasks. It works by building multiple decision trees on different subsets of the dataset and then averaging the results to enhance predictive accuracy. This technique of combining several classifiers to address a complex problem and improve model performance is known as ensemble learning. Instead of relying on a single decision tree, the random forest makes predictions based on the majority vote of all the trees in the model. Increasing the number of trees in the forest helps reduce overfitting and improves the overall accuracy of the predictions.

What purpose does Random Forest serve?

- It provides highly accurate predictions, even with large datasets, and operates efficiently.
- It demands less training time compared to other algorithms.
- It can maintain accuracy even when a substantial amount of data is missing.

The two phases of Random Forest's operation involve first building the random forest by combining N decision trees, followed by making predictions based on the outputs of each tree generated in the initial phase.

Algorithm for Random Forest:

The stages and graphic below can be used to demonstrate the working process:

Step 1: Randomly select K data points from the training set.

Step 2: Build decision trees based on the selected data points (subsets) from step one.

Step 3: Choose the number N, which determines how many decision trees you want to create.

Step 4: Repeat Steps 1 and 2.

Step 5: Gather the predictions from each decision tree for the new data points and assign the final classification to the group that receives the majority of votes.

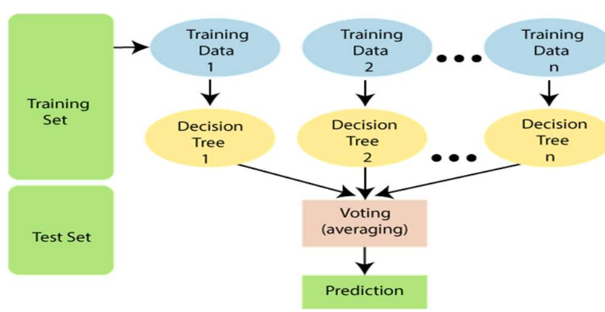


Fig 4.1: Working of the Random Forest algorithm

Benefits of Random Forest:

- It can handle jobs related to both classification and regression.
- It has the ability to handle big, highly dimensional datasets.

- It avoids the overfitting problem and improves the model's accuracy.

IV. ADVANTAGES

- The suggested method for predicting crop yield is more practically implementable and yields results more accurately than other conventional ways because it makes use of both ground and remote sensing data.
- Policy makers find it simple to implement agro-economic initiatives such as price prediction.
- The scheme was expanded to include crop insurance and local department of agriculture officials estimating crop yields allowing the bank to issue more loans.
- This initiative aids seed generation organizations in comprehending how diverse environmental circumstances and regional heterogeneity affect seed genotype performance in real-time scenarios.

V. RESULTS

Pendilimari village

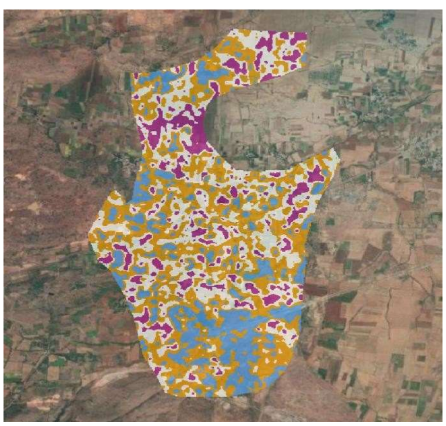


Fig 6.1 Pendilimari Village

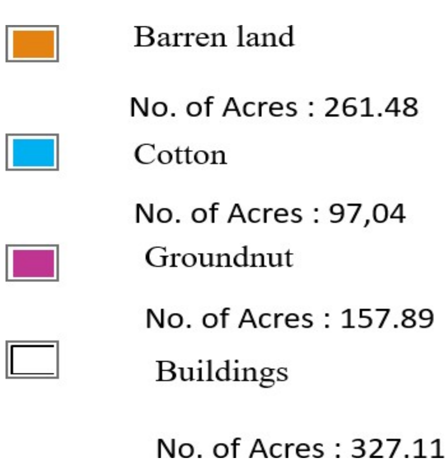


Fig 6.2 output after classification using Random Forest Classifier

Velamvaripalli village

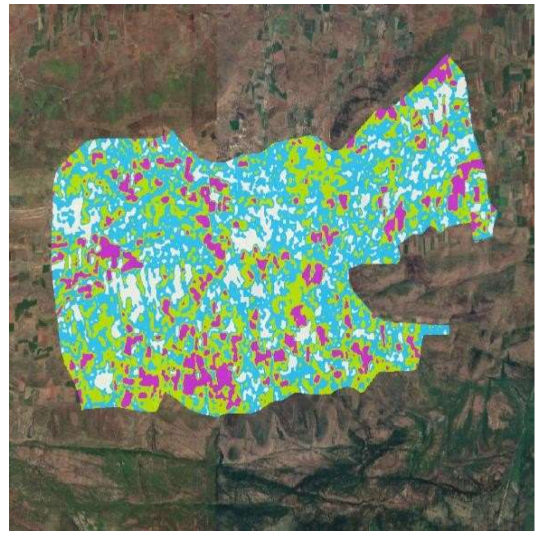


Fig 6.3 Velamvaripalli Village

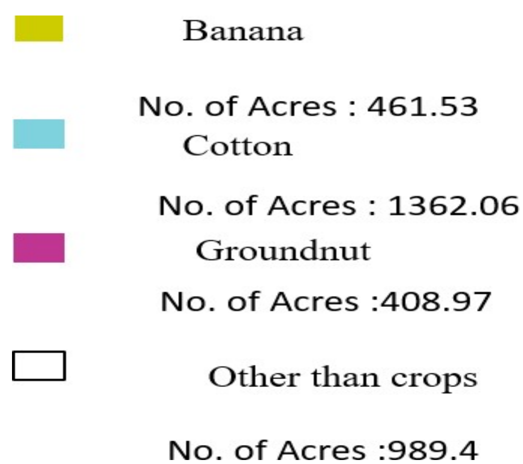


Fig 6.4: output after classification using Random Forest Classifier

VI. CONCLUSION AND SCOPE FOR FUTURE WORK

7.1 Conclusion

This study utilized VV and VH polarization to explore the potential of multi-temporal Sentinel 1-A SAR data for mapping groundnut and cotton areas. The mean backscattering values for groundnut and cotton increased by 1.0 dB to 2 dB at the peak crop growth stage (from σ0 D3 to σ0 D4-D9), compared to the initial stage (σ0 D3). The findings demonstrate that the high temporal resolution of the multi-temporal Sentinel 1-A SAR sensor, which captures the entire crop phenology throughout the cropping season, proves to be effective for distinguishing between cotton and maize crops.

7.2 Scope for Future Work

Future research will focus on investigating temporal models incorporating polarimetric features and evaluating a combined model using dual-polarimetric SAR intensities. Additionally, the generalization of these temporal models will be explored. The complexities of crop growth under different environmental conditions could be better understood by linking the temporal models to climate factors, such as seed type, latitude, altitude, irrigation, and temperature. In the future, a temporal model library could be developed to aid in crop classification, yield estimation, and market price prediction. This project could also be expanded to include yield estimation in subsequent studies.

REFERENCES

[1] Aubert, M., N. Baghdadi, M. Zribi, A. Douaoui, C. Loumagne, F. Baup, M. El Hajj and S. Garrigues. 2011. Analysis of TerraSAR-X data sensitivity to bare soil moisture, roughness, composition and soil crust. Remote Sensing Environment, 115, pp. 1801–1810.

[2] Boerner, W.M., B.Y. Foo and H.J. Eom. 1987. Interpretation of the polarimetric co-polarization phase term in radar images obtained with JPL airborne L-band SAR system. IEEE Transactions on Geoscience & Remote Sensing, 25(1), pp.77- 82.

[3] Bouman, B. and H. Van Kasteren. 1990. Ground-based x-band (3cm wave) radar backscattering of agricultural crops. ii: Wheat, barley and oats; the impact of canopy structure. Remote Sensing Environment, 34(2), pp. 107–118.

[4] github.com, Ffilipponi Repository—Sentinel-1_GRD_Preprocessing: StandardWorkflow for the Preprocessing of Sentinel-1 GRD Satellite Data. Available online: https://github.com/ffilipponi/Sentinel-1_GRD_preprocessing (accessed on 21 May 2019).

[5] Park, J.W.; Korosov, A.; Babiker, M. Efficient thermal noise removal of Sentinel-1 image and its impacts on sea ice applications. In the Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 23–28 April 2017; Volume 19, p. 12613.

[6] SNAP Software, Help Document 2019. Available online: https://step.esa.int/main/toolboxes/snap (accessed on 1 May 2018). 4. Guillaume, H. Masking "No-value" Pixels on GRD Products generated by the Sentinel-1 ESA IPF. ESA Tech. Rep, 2015. Reference MPC-0243, Issue 1.0.

[7] https://en.wikipedia.org/wiki/QGIS

[8] https://sentinels.copernicus.eu/web/sentinel/toolboxes/sentinel-1

[9] https://colhub.copernicus.eu/userguide/

[10] Skriver, H.; Mattia, F.; Satalino, G.; Balenzano, A.; Pauwels, V.R.N.; Verhoest, N.E.C.; Davidson, M. Crop classification using short-revisit multitemporal sar data. IEEE J-STARS 2011, 4, 423–431.

[11] Skakun, S.; Kussul, N.; Shelestov, A.Y.; Lavreniuk, M.; Kussul, O. Efficiency assessment of multitemporal c-band Radarsat-2 intensity and Landsat-8 surface reflectance satellite imagery for crop classification in Ukraine. IEEE J-STARS 2016, 9, 3712–3719.

[12] Kenduiywo, B.K.; Bargiel, D.; Soergel, U. Higher order dynamic conditional random fields ensemble for crop type classification in radar images.IEEE Trans. Geosci. Remote Sens. 2017, 55, 4638–4654.