# Cotton Crop Classification using Optical and Microwave Remote Sensing Datasets in Google Earth Engine

Benazir Meerasha [1], Martin Sagayam [2]

[1] Research Scholar, Electronics and Communication Engineering Department, Karunya Institute of Technology and Sciences, India – benazirm@karunya.edu.in
[2] Electronics and Communication Engineering Department, Karunya Institute of Technology and Sciences, India – martinsagayam@karunya.edu

**Keywords:** Sentinel 1, Sentinel 2, Random Forest, Google Earth Engine, Crop Mapping

**Abstract**

Cotton maps for Telangana which is one of the cotton producing region in India were produced through supervised classification in Google Earth Engine. To make well-informed decisions, farmers, governments, scientists, and agricultural organizations need accurate information on crop prediction. However, automated crop type mapping remains challenging due to the limited availability of field-level crop labels required to train supervised classification models. Cotton mapping was made more accurate and efficient by using a two-step mapping approach, which consists of mapping the cropland and then extracting the cotton crop for areas with more heterogeneity, this framework increased the accuracy from 83% to 91%. For a more accurate estimate of the cotton crop, this study combined high resolution Sentinel-1 and Sentinel-2 data with several secondary data types in the SMILE Random Forest (RF) model at various stages of the crop growth season. For that First, cropland/non-cropland area were predicted to extract features from time series. Next, cotton crops through RF classifiers were applied on median composites of Sentinel-1 and Sentinel-2 data for each pixel in the region. Furthermore, spectral, structural and phenological feature time-series satellite data were merged and processed into a supervised random forest classifier. The classification of cotton, cropland and noncropland model produced with producers accuracy of 98%,88% and 90%. Through experiments, we also discovered that employing time-series imagery generates substantially higher classification results than single-period images. The inclusion of shortwave infrared bands, followed by the addition of red-edge bands, can increase crop classification accuracy more than using simply traditional bands like the visible and near-infrared bands. Incorporating common vegetation indices and Sentinel-2 data, combining with Sentinel-1 reflectance bands improved the overall crop classification accuracy by 0.2% and 0.6%, respectively. This study demonstrates how combining optical and microwave remote sensing data, the GEE platform, transfer learning, and cotton cropland mapping algorithms can enhance insights into precision agricultural systems.

## 1. Introduction

### 1.1 Introduction

Crop monitoring has emerged as a crucial area in remote sensing-based Earth observation. Remote sensing has become the primary approach for crop mapping at both local and global scales. Unlike labour-intensive and time-consuming field surveys, often complicated by the fragmented and diverse nature of farmland in India, remote sensing offers wide coverage, frequent monitoring, and provides rapid, accurate, and objective crop information. (Dong et al., 2015). For the purposes of crop adaptation assessment and disaster warning, precise and timely identification of crop types and planting locations is crucial. Remote sensing data serves as a resource that can be applied to complement ground statistics (Qiu et al., 2015). Cotton is a significant cash crop in India, thus the variations in its planting area and yield will have an impact on India's cotton-related agricultural development decisions. Accurate and timely mapping of cotton fields is essential to the long-term management and observation of cotton economics. Crop mapping for a million km.sq. expanse frequently requires handling images with tens of thousands of scenes, which is more for the local workstations can handle (Cai et al., 2018). The development of cloud computing and storage platforms, such Google Earth Engine (GEE), has greatly accelerated the growth of crop mapping on a regional and worldwide scale (Xiong et al., 2017). With the help of this platform, more than 70 studies on crop mapping have been carried out throughout the ten years since GEE launched (2010–2019) (Tamiminia et al., 2020). The two primary sources of satellite data are optical and radar remote sensing imagery, which can produce extensive spatial and phenological data with spectral reflectance and backscattering coefficient, respectively (Chen & Zhang, 2023). The combination of Sentinel-1 and Sentinel-2 along with random forest classifier increases the accuracy of the cotton field mapping (Hu et al., 2021a). The different crop type mapping can be analysed effectively by fitting harmonic regression and retrieve the coefficients to extract features from time series (S. Wang et al., 2019). The Sentinel data series have also been used for crop extraction because of their great temporal and spatial resolution. It has established the foundation for efficient and precise crop extraction (Talema & Hailu, 2020). Due to its resistance to sun light and cloud cover, Synthetic Aperture Radar (SAR) circumvents the limits of optical images and can be used for large-scale, high-resolution crop mapping (Fikriyah et al., 2019). The surface information provided by optical and SAR data varies because of their distinct properties. The measured region's spectral properties are provided by the optical data, whereas the SAR data include details regarding soil and vegetation structure (Tian et al., 2019).

Phenological curves and differences are frequently used to identify crops due to phenological variances among different crops (Hu et al., 2021). Textural features like the Gray-Level Co-occurrence Matrix (GLCM) indices have been used to improve crop classification performance (Peña-Barragán et al., 2011).

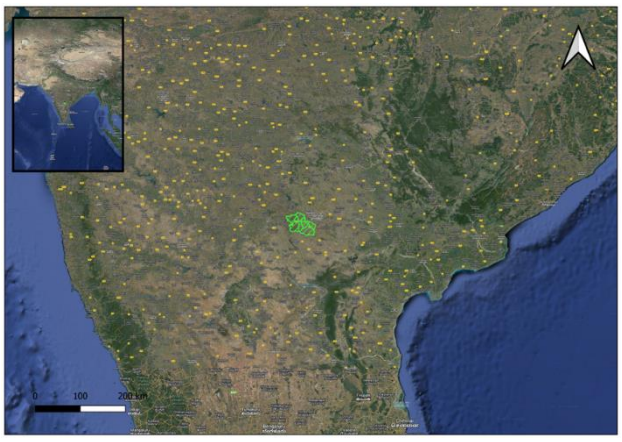
Radar, an active form of remote sensing, can more accurately represent the spatial characteristics of land cover in cloudy and rainy weather than optical remote sensing. Because of their effectiveness when they work together, the combination of radar and optical remote sensing has been an effective tool for classifying crops (Mascolo et al., 2021). Furthermore, there are two categories of typical supplementary data: meteorological and topography variables (Boryan et al., 2011). A digital elevation model (DEM) and slope are two topographic characteristics that may assist to identify terraced fields and irrigated crops, respectively. Temperature and rainfall are two examples of meteorological factors that have a significant impact on crop growth and development and hence have high correlation with the phenology of plants (Tariq et al., 2023). Therefore, the combination of optical and radar remote sensing imageries as well as auxiliary data may offer more information gains for crop classification across the study area Telangana.

In this study 10 m cotton map for 2020 to 2021 were created, utilizing all available Sentinel-1 and Sentinel-2 imageries, together with supplementary data in the GEE platform. Owing to varying image availability between years, we created distinct methods and conducted in-depth analysis and comparison. In the upcoming year, the released cotton maps dataset can serve as a foundation for crop management and policymaking, and the suggested method can be applied to cotton mapping.

## 2. Methods

### 2.1 Study Area

This study area provides the description of the dataset location. The state of Telangana in southern part of India is the home to the research region (Figure.1). The latitude and longitude of the study region are (17.32,77.99). There are mild undulations across the research area due to elevation variations of 520 m to 730 m above Mean Sea Level. The eastern part of the research area gradually falls, whereas the western portion is at a greater level. The average annual rainfall in the study area is 875 mm, with the NW portion of the watershed receiving the most (about 1000 mm) as opposed to the SE, which only receives 700 mm. As the Pediplain Pediment Complex of Denudational Origin covers approximately 85% of the study region, this topography is better suited for agricultural use. The western Deccan traps makes this location perfect for growing cotton crops. Loamy soil is more suited for agricultural use because it is more prevalent in the western region than the eastern one. Consequently, the western part of the study area contains the majority of the double crop zones. The majority of the grounds are utilized for producing vegetables, corn, cotton and receive rainfall.



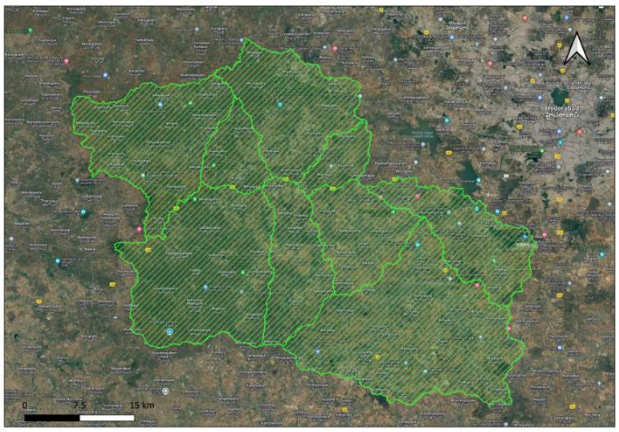


Figure 1. Study area location in Telangana, India Google satellite Image.

### 2.2 Datasets

We carried out field sampling in the main-cotton-planting areas in March 2020. We interviewed the local farmers regarding how many years they have been cultivating cotton over this study area. For the past 10 years they were cultivating cotton over this region. A large number of georeferenced points were collected for both crop types (such as cotton, maize, and corn) and non-crop areas (including barren land, water bodies, forests, and grasslands), as summarized in Table 1. To ensure sufficient, high-quality, and spatially balanced training data across the study area, sample expansion was conducted through visual interpretation of high-resolution imagery available on the Google Earth platform. This process resulted in a total of 2,228 samples for March 2020: 728 samples for non-cotton land cover, 1,000 for cotton cropland, and 500 for another cropland (Table 1).

| Class name (March 2020) | Field Survey | Google Earth | ESA world Cover (2020-2021) | Total |
|---|---|---|---|---|
| cotton | 500 | 500 | | 1000 |
| cropland | 100 | | 400 | 500 |
| Water body, built up, trees and barren land(non-cropland) | 28 | 200 | 500 | 728 |

Table 1. Number of point coordinates for supervised classification.

### 2.3 Sentinel-2 Imagery and Pre-processing

The Sentinel series, comprising optical (Sentinel-2 A/B, S2) and radar (Sentinel-1, S1) imagery, provides multi-band images at a spatial resolution of 10 meters for the study area, covering the period from 2020 to 2021. Harmonized Sentinel-2 Multispectral Instrument (COPERNICUS/S2_SR_HARMONIZED) surface reflectance data from March 2020 to March 2021 is available on the GEE platform for this region. Sentinel-2 surface reflectance imagery is commonly utilized for crop classification (Hu et al., 2021c). To remove clouds and cloud shadows from the Sentinel-2 dataset, we implemented the Clean Pixel Extraction: Utilized

the cloud probability dataset available in GEE to identify and retain cloud-free pixels.

### 2.4 Sentinel-1 imagery and pre-processing

In this research, all images for VV (single polarization, vertical emission/vertical reception) and VH (double polarization, vertical emission/horizontal reception) polarization from March 2020 to March 2021, were selected. In total, 42 Sentinel-1 dual polarized C-band SAR instrument images were analysed for cotton crop classification, and these data were archived in GEE in the form of Sentinel-1 SAR Ground Range Detected (GRD) datasets. All images were pre-processed by the Sentinel-1 toolbox using thermal noise removal, radiometric calibration, terrain correction using the Shuttle Radar Terrain Mission (SRTM) or Advanced Spaceborne Thermal Emissions and Reflection Radiometer (ASTER) digital elevation model (DEM) and conversion to a backscattering coefficient (σ0) in decibels (dB). Sentinel-1 has an average a maximum of 42 observations. Finally, the Refined Lee Filter has been used to filter the backscattering time series to remove speckle noise in the Sentinel-1 dataset (Lee et al. 2008; Yommy et al. 2015). For this classification task, this paper selected the VV polarization and VH polarization of Sentinel-1, which are commonly used in land use classification tasks.
Based on previous studies in crop classification and mapping, a two-step strategy has been employed to first separate the cropland from non-croplands and then cotton, cropland and non-cropland has been classified within the study area. A two-step process enables the elimination of non-cropland types and has a significant application in cotton mapping.

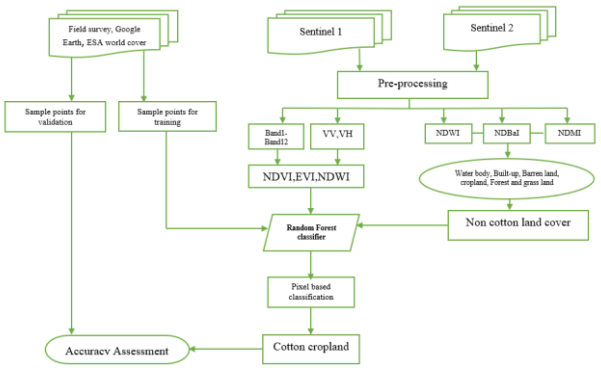



Figure 2. Methodology presented in this study.

### 2.5 Feature Selection

The Google Earth Engine (GEE) platform offers various machine learning techniques for imagery classification, including random forest (RF), decision tree, support vector machine (SVM), and naive Bayes classifiers. Among these, RF is widely used for crop classification due to its ability to leverage the bagging technique, which optimizes feature selection through an automated process to enhance classification accuracy. In our study, we trained the models using a consistent set of features for both cropland and cotton mapping tasks. As an illustration, the technical framework for the second phase of cotton mapping is presented in Figure 2. Implementing the vegetation index has made it possible to assess vegetation growth and coverage qualitatively, which is useful for crop monitoring. This study computed two vegetation indices: the normalized difference vegetation index (NDVI) and the EVI stands for enhanced vegetation index (Gong et al., 2024). These indices offer essential data, regarding the growth of vegetation and biomass, respectively. Additionally, several indices were computed to capture essential features related to vegetation, water, and land cover. These include the Normalized Difference Water Index (NDWI), Normalized Difference Moisture Index (NDMI) and Normalized Difference Bareness Index (NDBaI), which are widely applied in studies focusing on water bodies, built-up areas, vegetation canopies, and related phenomena (Habibie et al., 2024). NDWI emphasizes water body information, making it effective for detecting surface water features. VV and VH polarizations capture vegetation-related information by accounting for factors such as canopy structure and observation angle, offering advantages for cropland monitoring. Red Edge Band is particularly sensitive to chlorophyll content, making it a valuable indicator for assessing crop growth and health. These indices, combined with SAR data, provide complementary insights into vegetation dynamics, enhancing the accuracy of crop monitoring and classification.

### 2.6 Pixel-Based Classifier: Random Forest

Classification tree or K tree, P random choices is nothing but the classification trees which belongs to family with name Random Forest classifiers. Breiman was the first to propose this idea, in 2001(Breiman, 2001). Classes that submit to K-tree P with variables and prediction over each k when combining via majority category voted in the full process. Each tree is trained on random points for each node, and a binary problem to split the learning set at this level of each tree chosen from p input variables (picked randomly). The random forest is a bagging machine learning technique that fits many decision trees on subsamples of the data. Random forest classifiers are also good for interpreting important of indicators and feature selection (Maxwell et al., 2018), which making it high necessity but data rich module to be able to employed better than other models Lawrence and Moran (2015) has compared the Random Forest with other machine learning models analysis. The classification accuracy based on attributes obtained higher for pre-processing data among different ways used, where the average of prediction by random forest was greater than others methods: Support Vector Machines (SVM). Therefore, the random forest model has higher computational cost than other decision tree of ensemble methods for better performance cotton crop extraction. They take care of the variables in a more optimal way and require very less to zero work on hyperparameter tuning.

### 2.7 Accuracy Assessment

High-quality data can be used to assess the accuracy of outputs at a relevant spatial and temporal scale. Both of these verification datasets are therefore essential components to the mapping project. Ground-truth data of 2020 were collected with a stratified random sampling of the accuracy assessment, based on different sources. The cotton, cropland and non-cropland field samples were interpreted on the High Spatial Resolution images in Google Earth or ESA Worldcover with an S2 image 2020 as a reference dataset to produce the final results of Cotton Croplands Map. During the collection of the ground sample points, the third Derivative Land Survey data was applied as an auxiliary reference source in visual process handling. In this study collected a total of 2228 validate samples which is combination with 1000 cotton field, 500 cropland and 728 non-cropland field. The confusion matrix was calculated using the validation samples for accuracy calculation. Accuracy including overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA) and kappa coefficient (k) were computed by following equations:

$$OA = TP + TN+TS / TP + FPS + FPN +FSP+FSN+FNS+FNP+ TN+TS \quad (1)$$

$$PA=TP/(TP+FSP+FNP) \quad (2)$$

$$UA=TP/(TP+FPS+FPN) \quad (3)$$

$$k=po-pe/1–pe \quad (4)$$

where TP represents the actual number of cotton fields predicted as cotton fields, TN is the actual number of non-cropland fields predicted as non-cropland fields, TS actual number of cropland predicted as cropland ,FPS is the actual number of cotton fields predicted as cropland fields, FPN is the actual number of cotton fields predicted as non-cropland, FSP actual number of cropland predicted as cotton, FSN actual number of cropland predicted as non-cropland, FNS actual number of non-cropland predicted as cropland, FNP actual number of non-cropland predicted as cotton , po represents the overall classification accuracy, and pe represents the ratio of the sum of the product of the number of real samples and the number of predicted samples to the square of the total number of samples.

## 3. Results and Discussions

### 3.1 Mapping of cropland and non-cropland

For the cropland classification, we trained random forest models using the sentinel 2 dataset. Breiman (2001) describes random forests as an ensemble machine learning technique that combines many decision trees to produce exceptional accessibility and performance. For crop type mapping, they have been shown to yield higher accuracies than maximum likelihood classifiers, support vector machines, and other techniques (Pelletier et al., 2016). They are extensively utilized in the field of remote sensing for mapping crop types and classifying land cover (Ghazaryan et al., 2018). The detailed landscape mosaic that is observed in the study area comprises five classes – (1) water bodies – rivers, lakes; (2) cropland- permanent crops; (3) built up- urban areas; (4) treecover- forests, like those growing along rivers or lake shore with the fifth category-called others for non-irrigated agricultural land and miscellaneous other lands. A total of 1228 training points were generated randomly and labelled manually through a visual interpretation of the 'ESA/WorldCover/v100'. The literature study makes extensive use of this strategy. Transfer learning approach has been adopted in this research that is training the model with pretrained datasets like ESA Worldcover increases the overall accuracy from 83% to 91%. Earth Engine has a classifier called random forests that is computationally efficient and capable of achieving high accuracies. A minimum leaf population of 10 and 100 trees for an ee.Classifier object has been assigned. Thus, the landcover of the study area has been mapped.

### 3.2 Mapping of cotton crop

In this study, ground samples were taken from cotton fields and several Index values were calculated for selected point coordinates through March 2020-21 using Sentinel-1 and Sentinel-2 satellite images. For this, the study centered on the phenology of cotton crops (specifically at peak growing months). The findings were expressed with the help of graphs, which showed that the cotton plants experienced observable growth between June 2020 and January 2021, which also represented the peak season for growing. This data enables tracking the health and growth of cotton crops over time and aid in precision farming practices leading to improved yield.

In Google Earth Engine (GEE), a composite image was created incorporating spectral bands from Sentinel-1 and Sentinel-2, along with various spectral indices such as EVI, NBR, NDMI, NDWI, NDBI, and NDBaI. Additionally, elevation data from NASA’s Shuttle Radar Topography Mission (SRTM) at 30 m resolution (“USGS/SRTMGL1_003”) was included. We generated variables like classvalue, classnames, columns, and features. The classvalue had values from 0 to 2, representing cotton, cropland, and non-cropland as classnames. We created samples based on these classnames and split the data into training (80%) and testing (20%) sets. Samples were extracted using the sampleRegions function, where each pixel was treated as a feature with an associated classvalue. We then trained the model using the smile Random Forest (RF) algorithm, which is based on multiple decision trees to classify crop types. The model involved determining the number of trees and estimating the out-of-bag error. The results were assessed using a confusion matrix, overall accuracy, and kappa coefficient. The confusion matrix compared the model’s predicted classifications with the actual ones, yielding an overall accuracy of 0.91 and a kappa coefficient of 0.88.

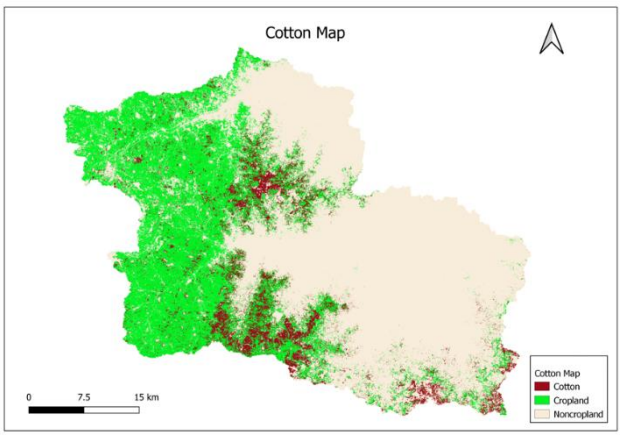



Figure 3. The classified image cotton map of the study area.

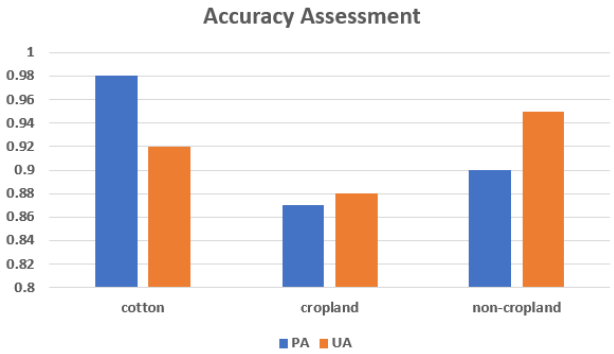



Figure 4. Accuracy Assessment

Graph comparing Producer Accuracy (PA) and User Accuracy (UA) for cotton, cropland & non-cropland. PA is the performance of the model and refers how well it detects pixels belonging to class x, and UA measures how often the predictions made by a model is correct. Cotton has approximately 0.98 for PA and 0.92 for UA which is high accuracy. On the other hand, while cropland also shares PA and UA having lower values than all other land classes (approx. 0.84), it shows more classification errors, non-cropland had an intermediate value of balanced PA around 0.95, which provides some confidence that the drop in performance for this class was not due to severe misclassification errors.

### 3.3 Time series profile of cotton map

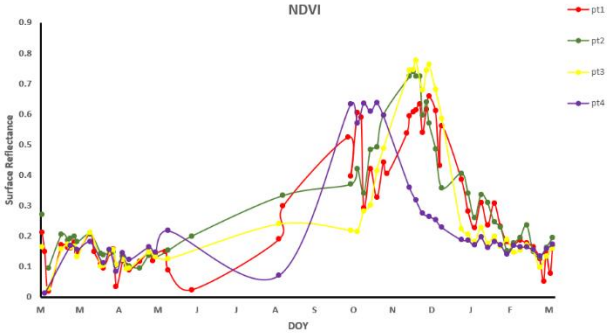



Figure 5. NDVI of 4 cotton point coordinates selected from the classified image.

NDVI (Normalized Difference Vegetation Index) is a remote sensing tool GIS analysts use for cotton crop analysis to examine plant health and development throughout the growing season. NDVI values are between -1 and +1 which allows us to learn about the photosynthetic energy, hence, vegetative greening changes in the quantum of light absorbed by green vegetation cover initiated by the surface.

The chart shows the seasonality of NDVI values for four different data points (pt1, pt2, pt3, and pt4) in the one-year whereby each line represents relevant to a specific location/pixel within a cotton field. The NDVI curves show some of the following trends.

At the early season (March-May), all points exhibit low NDVI values, ranging between 0.1 and 0.3 This indicates a cotton field in pre-plant or early growth stage. Given that cotton is a warm season crop, this information means that at the time these NDVI images were captured in spring, this has essentially zero or some vegetation cover, as it may be either bare or extremely early stages of its growth. At the growing season (June - August), NDVI values increase notably especially from June. Pt1 and pt2 both have a rapid rise in NDVI values that peaks around August, with values in the range of 0.6–0.7 This corresponds to the early-mid-season active growth period under dense canopy cover and thus maximal photosynthetic activity in cotton. The curve for pt3, appears to be shallower and does not show the NDVI peak value, which could again indicate a difference in health of the crop or possibly planting density or some external variable like soil quality, watering levels pesticide/ pest pressure. The pt4 shows a relatively lower peak compared to pt1 and pt2, suggesting a variation in crop conditions or environmental factors in that area.

At the peak season (August- September) the observed NDVI values are high which indicates the maximum (greenness) levels of the cotton crop, reflecting a healthy canopy with maximum leaf area. The cotton crops are well developed and active photosynthesis. The productivity and the yield of the cotton crop are largely decided in these months. Following the peak, during November NDVI values steeply drop. This period reflects the natural senescence of the cotton crop to maturity and harvesting. The drop in NDVI is correlated to the decline in green biomass as cotton plants shed their leaves and mature bolls, the result of both stages having less photosynthetic resources. During this period, NDVI show a significant fall for all points which is common in cotton harvest cycles. From December, the NDVI values once again returns to post-harvest low near 0. The cotton fields are likely to be recently ploughed with little or no residue left behind and thus all points have low NDVI. This time is also crucial for assessing the condition of field and soil to plan next seasons planting cycles.

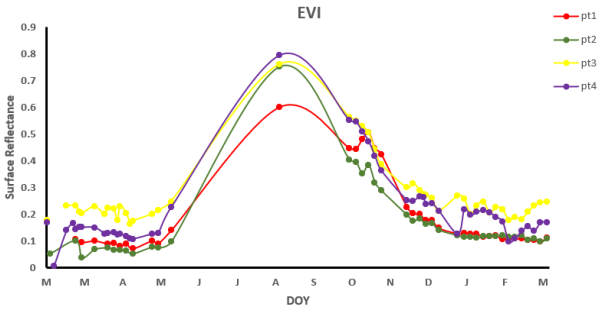



Figure 6. EVI of 4 cotton point coordinates selected from the classified image.

EVI is an indicator of vegetation health and is commonly used to monitor crop phenology. During (March- May), all cotton field presented an EVI that has a slight increase, showing the crop growth began. EVI reaches its maximum in the growing season (June to August) when photosynthesis is at its maximum. Among them, Pt 4 is the highest in EVI value (about 0.8), indicating the cotton crop with the greatest growth vigor; Such could mean a likely optimal growing conditions and very high productivity. On the other hand, Pt 1 has a slower growth trajectory and later peak, which means different cotton field with delayed phenological events.

From the peak in summer, all cotton fields show a decreasing trend in EVI values from August to October. This decline in integration suggests the crops are headed toward senescence, a time during which growth slows and leaves begin to yellow or drop from plants leading to reduced photosynthetic activity. The EVI values for all-time series were low during the winter (November–February), indicating that there was almost no vegetation activity over this period. This is the dormancy of many agricultural systems, in which crops either have been harvested or are not actively growing. The cotton crops time series from the temporal domain has a general seasonal growth pattern, starting in (March-May) and reaching its maximum vegetative productivity within (Jun-Aug), then undergoing a slow senescence fall dormancy winter cycle. Pt 4 is the rapid growth stage and may be a fast growing or prolific seeding crop, while Pt 1 could be an earlier maturing type of crop.

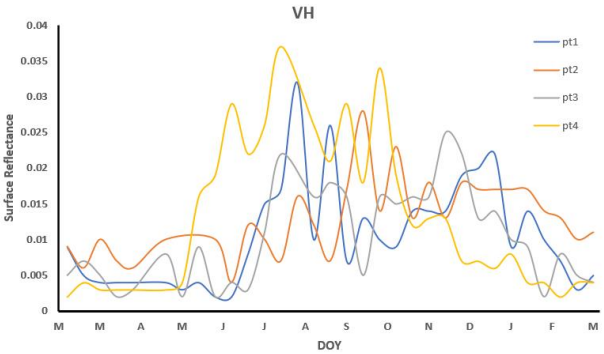



Figure 7. VH of 4 cotton point coordinates selected from the classified image.

The graph shows the phenology of that crop based on VH (Vertical-Horizontal) polarization from SAR (Synthetic Aperture Radar) data. Maps the surface reflectance over time, against Day of Year (DOY) The Y-axis shows the backscatter intensity (VH polarization) representing how much of the radar signal is being reflected back from the surface of cotton crop at different growth stages. That is simply the day of year (DOY) of that year on days

and time scale, which goes through many months through March from next year to around March again.

Each line corresponds to one of the different crop fields (pt1 to pt4) where the reflectance data were collected. There are many seasonal patterns observations which represent the four growth phases of the cotton crop: planting, growing, peak and harvest periods. Across all points, there is an obvious spike in surface reflectance from July to about September, which we assume reflects the seasonal peak vegetation or floral stage of the cotton crop. This is the period in which the whole of the crop is good established and will communicate more offering with the radar signal.

Once this peak is reached, reflectance drops more abruptly for all but the first point, typically indicating the end of the growing or harvesting season. Up to the early months (Mar-May), the reflectance was quite low, probably due to either bare soil or initial vegetative stages of cotton crop. However, the signal from each line (pt1, pt2, pt3, pt4) can be mainly attributed to variability of soil moisture and crop density or very small phenological variation considering both locations. As the VH polarization in SAR data is sensitive to vegetation structure and moisture, the peaks and troughs signify large changes in these parameters throughout the season.

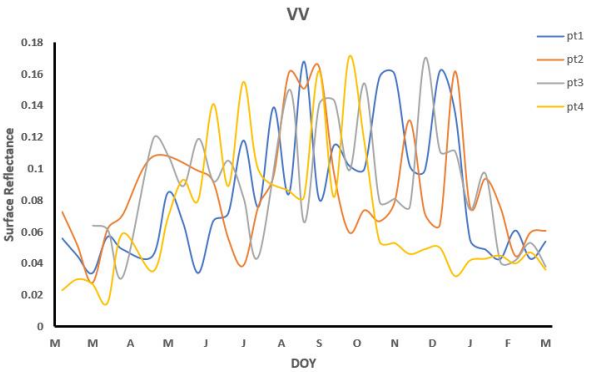



Figure 8. VV of 4 cotton point coordinates selected from the classified image.

The following figure 8 is the cotton phenology which was derived from VV polarized data of Synthetic Aperture Radar (SAR), within this the plant surface reflectance is plotted against time with DOY represented on X-axis and reflectance in Y-axis. The graph shows very high reflectance immediately after emergence in April-May, with a definite but uncertain spike between 7 and 8 weeks (more visible for pt3); this is likely soil moisture as well as early vegetative stages or rapid growth in that region. The VV signal which is widely interfered by the structural properties of crops and surface roughness, especially for early stages of growth.

All the points show a rise in peak reflectance during (June-July), which would be expected as it is around the mid to late growth stage of the cotton crop and is when flowering occurs, and continues until about peak vegetative phase of canopy cover. We can see that certain regions have clearer peaks or more stable reflectance values (for example, pt2 has sharper high peaks) that may suggest differences in crop condition, moisture content or regional growth dynamics. Following September, there is a downward slope in the surface reflectance corresponding to the senescing phase of crop when the canopy density is decreasing down due to harvesting.

The VV polarization is generally more sensitive to surface roughness and the vertical crop structure, including stalks and leaves. Consequently, the high reflectance in the middle of the year indicate maximum growth and variability through points could be attributed to differences in crop health or soil moisture across cotton fields.

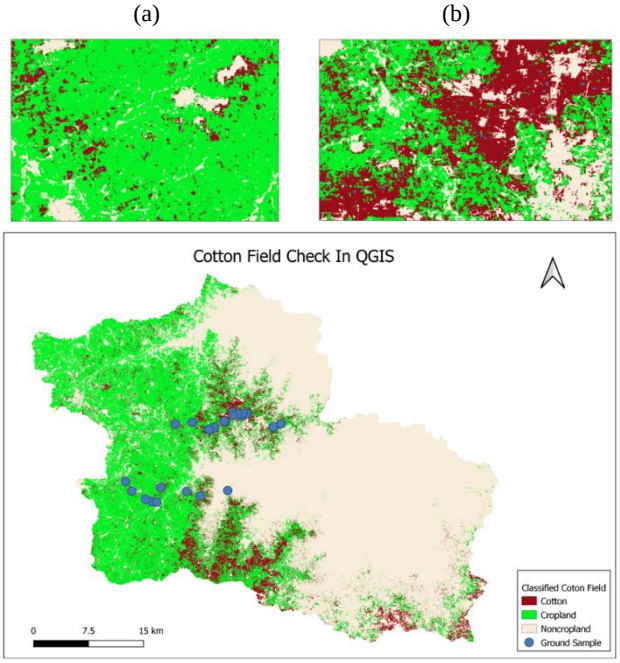



Figure 9. Cotton Field check in QGIS

Figure 9 shows the classified image is analysed in QGIS, to verify if the point coordinates collected from the cotton fields align with the predicted cotton crop areas. In Figure a and b ensures that the ground coordinates collected from the field survey falls under the classified cotton. This assessment ensures that the classification accurately represents the areas designated as cotton crops based on the collected coordinates.

In addition, some general characteristics of its flowering stage and boll-opening remain phenologically indicative for cotton-mapping. Optical imagery from Sentinel 2 can distinguish the cotton flower, but it is difficult, because of the cotton consist tens of flowers have different color, such as white and red color that has apparent distinct colors (Liang et al., 2023). In particular, the color of petals might change as a flower develops. Though different cotton varieties show variations in the flowers colors, even on same plant flowers of different colors are present. Another important phenological stage is the boll-opening period (Jiang et al., 2020). White fibre contains most of the cotton color with a small amount of fibre. Some recent studies found that boll-open time can be used as a phenological feature for cotton mapping and yield estimation (Xu et al., 2021). Wang et al. (2021) established the white bolls index (WBI) from Sentinel 2 imagery in the boll-opening period and conducted a county-level cotton mapping. The mapping performance was similar to our study (OA=95%, Kappa =0.88) to the best of our knowledge, in which WBI (N. Wang et al., 2021) was able to achieve this accuracy. However, there are some drawbacks: It needs the temporal data at a higher resolution, temporally (every 5 days) for the boll-opening timing determination or in other words it still does not decrease up the same data as our monthly vegetation index. A second point is that it heavily utilizes the optical imagery in boll-opening. However, it is very hard to ensure that all the cotton fields can have one clear observation during the boll-opening stage because of bad weather (clouds, mist, rain), etc. This makes it difficult to transfer the WBI approach to cotton where early detection is possible. Specifically, the WBI is employed based on determining cotton and non-cotton, therefore another technical means should be used to identify cropland. To

address this issue, the study first differentiated between cropland and non-cropland areas using remote sensing techniques. This separation was achieved by analyzing the land cover data. Once cropland was identified, cotton crops were predicted based on their unique temporal and spatial features in Google Earth Engine using SMILE Random Forest classifier. By using these methods, the study effectively identified cotton fields, distinguishing them from other types of crops or land cover, ensuring more accurate detection and classification of cotton crops.

## 4.Conclusion

The cotton-growing area was estimated using both SAR and optical data, applying a supervised classification technique in Google Earth Engine to distinguish cotton pixels. By fusing SAR and optical data, the classification achieved an overall accuracy of 91% and a kappa coefficient of 88%, with the cotton area calculated to be 14,666.35 hectares. Temporal analysis of backscattering coefficients and spectral reflectance for cotton, cropland, and non-cropland was conducted throughout their vegetative and reproductive stages. This research could be further extended to predict cotton yield and biomass estimation.

## References

Boryan, C., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: The US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, *26*(5), 341–358. https://doi.org/10.1080/10106049.2011.562309

Breiman, L. (2001). *Random Forests* (Vol. 45).

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., & Li, Z. (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, *210*, 35–47. https://doi.org/10.1016/J.RSE.2018.02.045

Chen, J., & Zhang, Z. (2023). An improved fusion of Landsat-7/8, Sentinel-2, and Sentinel-1 data for monitoring alfalfa: Implications for crop remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *124*, 103533. https://doi.org/10.1016/J.JAG.2023.103533

Dong, J., Xiao, X., Kou, W., Qin, Y., Zhang, G., Li, L., Jin, C., Zhou, Y., Wang, J., Biradar, C., Liu, J., & Moore, B. (2015). Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. *Remote Sensing of Environment*, *160*, 99–113. https://doi.org/10.1016/J.RSE.2015.01.004

Fikriyah, V. N., Darvishzadeh, R., Laborte, A., Khan, N. I., & Nelson, A. (2019). Discriminating transplanted and direct seeded rice using Sentinel-1 intensity data. *International Journal of Applied Earth Observation and Geoinformation*, *76*, 143–153. https://doi.org/10.1016/J.JAG.2018.11.007

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

Gong, Z., Ge, W., Guo, J., & Liu, J. (2024). Satellite remote sensing of vegetation phenology: Progress, challenges, and opportunities. In ISPRS Journal of Photogrammetry and Remote Sensing (Vol. 217, pp. 149–164). Elsevier B.V. https://doi.org/10.1016/j.isprsjprs.2024.08.011

Habibie, M. I., Ramadhan, Nurda, N., Sencaki, D. B., Putra, P. K., Prayogi, H., Agustan, Sutrisno, D., & Bintoro, O. B. (2024). The development land utilization and cover of the Jambi district are examined and forecasted using Google Earth Engine and CNN1D. *Remote Sensing Applications: Society and Environment*, *34*. https://doi.org/10.1016/j.rsase.2024.101175

Hu, T., Hu, Y., Dong, J., Qiu, S., & Peng, J. (2021a). Integrating sentinel-1/2 data and machine learning to map cotton fields in northern xinjiang, china. *Remote Sensing*, *13*(23). https://doi.org/10.3390/rs13234819

Jiang, Y., Li, C., Xu, R., Sun, S., Robertson, J. S., & Paterson, A. H. (2020). DeepFlower: a deep learning-based approach to characterize flowering patterns of cotton plants in the field. *Plant Methods*, *16*(1). https://doi.org/10.1186/s13007-020-00698

Lawrence, R. L., & Moran, C. J. (2015). The AmericaView classification methods accuracy comparison project: A rigorous approach for model selection. *Remote Sensing of Environment*, *170*, 115–120. https://doi.org/10.1016/J.RSE.2015.09.008

Liang, Q., Jin, Y., Zhu, Q. H., Shao, D., Wang, X., Ma, X., Liu, F., Zhang, X., Li, Y., Sun, J., & Xue, F. (2023). A MYB transcription factor containing fragment introgressed from Gossypium bickii confers pink flower on Gossypium hirsutum L. *Industrial Crops and Products*, *192*, 116121. https://doi.org/10.1016/J.INDCROP.2022.116121

Mascolo, L., Martinez-Marin, T., & Lopez-Sanchez, J. M. (2021). Optimal grid-based filtering for crop phenology estimation with sentinel-1 sar data. *Remote Sensing*, *13*(21). https://doi.org/10.3390/rs13214332

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. In *International Journal of Remote Sensing* (Vol. 39, Issue 9, pp. 2784–2817). Taylor and Francis Ltd. https://doi.org/10.1080/01431161.2018.1433343

Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, *187*, 156–168. https://doi.org/10.1016/J.RSE.2016.10.010

Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E., & Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, *115*(6), 1301–1316. https://doi.org/10.1016/J.RSE.2011.01.009

Qiu, B., Li, W., Tang, Z., Chen, C., & Qi, W. (2015). Mapping paddy rice areas based on vegetation phenology and surface moisture conditions. *Ecological Indicators*, *56*, 79–86. https://doi.org/10.1016/J.ECOLIND.2015.03.039

Talema, T., & Hailu, B. T. (2020). Mapping rice crop using sentinels (1 SAR and 2 MSI) images in tropical area: A case study in Fogera wereda, Ethiopia. *Remote Sensing Applications:*

*Society and Environment*, *18*, 100290. https://doi.org/10.1016/J.RSASE.2020.100290

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *164*, 152–170. https://doi.org/10.1016/J.ISPRSJPRS.2020.04.001

Tariq, A., Yan, J., Gagnon, A. S., Riaz Khan, M., & Mumtaz, F. (2023). Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest. *Geo-Spatial Information Science*, *26*(3), 302–320. https://doi.org/10.1080/10095020.2022.2100287

Tian, F., Wu, B., Zeng, H., Zhang, X., & Xu, J. (2019). Efficient identification of corn cultivation area with multitemporal synthetic aperture radar and optical images in the google earth engine cloud platform. *Remote Sensing*, *11*(6). https://doi.org/10.3390/RS11060629

Wang, N., Zhai, Y., & Zhang, L. (2021). Automatic cotton mapping using time series of sentinel-2 images. *Remote Sensing*, *13*(7). https://doi.org/10.3390/rs13071355

Wang, S., Azzari, G., & Lobell, D. B. (2019). Crop type mapping without field-level labels: Random Forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, *222*, 303–317. https://doi.org/10.1016/j.rse.2018.12.026

Xiong, J., Thenkabail, P. S., Gumma, M. K., Teluguntla, P., Poehnelt, J., Congalton, R. G., Yadav, K., & Thau, D. (2017). Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *126*, 225–244. https://doi.org/10.1016/J.ISPRSJPRS.2017.01.019

Xu, W., Chen, P., Zhan, Y., Chen, S., Zhang, L., & Lan, Y. (2021). Cotton yield estimation model based on machine learning using time series UAV remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, *104*, 102511. https://doi.org/10.1016/J.JAG.2021.102511