ASSESSMENT OF SENTINEL-1 AND SENTINEL-2 SATELLITE IMAGERY FOR CROP CLASSIFICATION IN INDIAN REGION DURING KHARIF AND RABI CROP CYCLES

Jitendra Singh, Aniruddha Mahapatra, Saurav Basu, Biplab Banerjee

IBM Research-India, IIT Roorkee, IIT Bombay

ABSTRACT

Real-time monitoring of agricultural crops is an important exercise because of it's huge impact on agri-business and agricultural policy management. Identification of crops during multiple crop growth stages can help formulate better agricultural policies and management strategies. In this context, the objective of this article is to evaluate the potential of Sentinel-1 Synthetic Aperture Radar (SAR) and Sentinel-2 optical imagery in crop identification for an Indian region. A multi-class classification algorithm based on the support vector machine (SVM) is applied to the temporal features extracted from the above mentioned satellite data sets. The experiments are conducted for Kharif and Rabi crop cycles with major crops in the region. The experiments suggest that the joint use of optical and radar imagery results in better classification accuracy compared to using them individually. An overall accuracy of 89% and 96% is obtained for Kharif and Rabi crops, respectively.

Index Terms— Crop Identification, Synthetic Aperture Radar, Normalized Difference Vegetation Index, SVM

1. INTRODUCTION

Crop classification approaches using remote sensing data sets are increasingly becoming more reliable (besides being scalable, cost and time effective) in acreage and yield estimation, thanks to advancing technologies in remote sensing and easier access to high resolution temporal/spatial satellite data.

Early crop classification approaches were mostly based on using different vegetation indices derived from optical remote sensing data, but their accuracy for a given region depends upon availability of sufficient number of cloud and noise free pixel tiles. This issue can be prominent for India and other equatorial regions. For instance, during Kharif crop season in India (May-November), due to an intersecting monsoon period (July-Sept), a critical crop growth period suffers cloud occlusion in optical satellite data. The Synthetic Aperture Radar (SAR) data is a promising option to address this challenge as cloud occlusion has an insignificant effect on radar back scattering. As a result, many crop-identification approaches have been proposed which benefit from the joint use of optical and SAR data to achieve high accuracy [1–3].

The main challenges in crop-identification for Indian regions are as follows: firstly, the size of the fields are typically much smaller in size with average field size less than quarter of an hectare. Secondly, croping patterns are very hetrogenous. Thirdly, as mentioned before, optical imagery data often have significant cloudy pixels during the crop growth cycle. Fourthly, the availability of ground reference data is limited and collection of such data is a difficult task. Despite these challenges, several interesting papers have been published on this topic for Indian regions using different machine learning methods and remote sensing satellite data sets. For instance, Kumar et al. performed multi-crop classification using different methods using sensor data from Resourcesat-2 [4]. The authors in in [5]presented initial results for Kharif crop cycles with few crops using a random forest algorithm and Sentinel-1 and Sentinel-2 derived features.

The objective of the present study is two fold: 1. Better understanding of the Sentinel-1 radar and Sentinel-2 optical imagery datasets for crop classification in the Indian region. 2. Analyse the relative importance of different features derived from these satellites using standard classification methods. To gain confidence, the analysis is performed over Kharif and Rabi crop cycles since the ground sample data is limited.

The paper is organised as follows: Section 2 briefly describes the study area and data sets used. Section 3 is on methods and analysis of the results. Key take-aways are summarised in section 4.

2. STUDY AREA AND DATA

2.1. Study Area

The study area lies in district Bulandshahr located in the state of Uttar Pradesh, India. It lies between 28.14-28.15 latitude and 77.76-77.78 longitude. Agriculture is the major land use type in the study area. The ground reference data was collected for Kharif (May-Nov 2017) and Rabi (Nov 2017-April 2018) crop cycles. For the Kharif 2017 crop cycle 293 fields samples were collected (April-Oct) for sugarcane (172), paddy (61), millet (33), sorghum (27), pigeon pea (10). Additionally, two more classes of built-up (34)and non-cropland (209) were added to this data. For the Rabi 2017-18 crop cycle, 301 samples were collected during the

end of the crop cycle for sugarcane (142), wheat (120) and mustard (39). Sugarcane is a perennial crop in the region and samples collected during Rabi crop cycle correspond to late sowed sugarcane fields.

2.2. Satellite Imagery

2.2.1. Sentinel-1 Synthetic Aperture Radar

The Sentinel-1 is a C-band SAR instrument operating in single (HH or VV) and dual polarisation mode (HH+HV or VV+VH). We have selected the dual polarised VV+VH SAR images (Interferometric Wide Swath, Ground Range Detected) from European Space Agency for this study. A set of 13 images for Kharif and 12 images for Rabi crop cycles were acquired between May 2017 and April 2018 from the same orbit, as listed in Tables 1 & 2. The Sentinel -1 images are available every 12-14 days with exceptions in June and October 2017.

The SAR data is processed using the SNAP toolbox in the following steps: 1. Calibration to obtain the backscatter coefficient 2. Multi-looking the data with a window size of 2 $\times$ 2 pixels to reduce the speckle noise effect at a resolution of 20m. 3. Terrain correction using the digital elevation model from the Shuttle Radar Topography Mission 4. A simple Lee filter with a 3 $\times$ 3 pixel window is applied to further reduce the speckle effect while preserving the 20m resolution.

2.2.2. Sentinel-2 Normalized Difference Vegetation Index

A set of 11 and 20 Sentinel-2 optical images were acquired for Kharif and Rabi crop cycles. From July 2017 the data from Sentinel-2B was also available, providing more comprehensive temporal coverage every 5 days. However, additional data from Sentinel-2B was not helpful for Kharif crop cycle as the images had a high percentage of cloud cover, as reflected in 1. Sentinel-2 tiles with more than 80% cloud cover were completely discarded. The images were processed to level 2A (i.e., surface reflectance values with masks for clouds, cloud shadows, snow and water) at 20m spatial resolution using the sen2cor toolbox. Table 1 & 2 lists the Sentinel-2 images used in this study. The NDVI values were computed from band 4 (red) and band 8 (near infrared, NIR) reflectance values as follows: NDVI = $\frac{NIR-red}{NIR+red}$. 23% and 26% NDVI time stamps were affected by cloud cover during Kharif and Rabi crop cycles, respectively. These were interpolated using neighbouring time stamps.

Table 1. Acquisition dates and sensor type for the data used in this study for Kharif crop cycle

Months	Sentinel -2	Sentinel-1 SAR
May	20	15,27
June	09	08
July	14	02,14,26
August	08,23	07,19,31
September	12,17,	12,24
October	02,07,22,27	06,18

Table 2. Acquisition dates and sensor type for the data used in this study for Rabi crop cycle

Months	Sentinel -2	Sentinel-1 SAR
October	22, 27	NA
November	NA	11,23
December	01,06	05,17
January	10,15,20,30	10,22
February	09,14	03,27
March	01,06,11,26,31	11,23
April	05,10,15,20,25	04,16

3. METHODOLOGY AND RESULTS

We built five multi-class classifiers using support vector machines (SVM) with following the features

- Time series of Sentinel-2 NDVI
- Time series of Sentinel-1 SAR consisting VV and VH seperately
- Time series of Sentinel-1 SAR jointly consisting VV and VH
- Time series of NDVI, VV and VH together as feature vectors

In all the experiments an 80-20 split was performed for training and testing data respectively. We tried the ratio of VV and VH as a feature but the classification accuracy worsens, hence the results are not included for the same.

3.1. Kharif crop cycle

For Kharif 2017 the crop classes considered are sugarcane, paddy, pearl millet, sorghum and pigeon pea. Sugarcane and paddy are the major Kharif crops in the region. Due to unavailability of good quality land use maps we have included non-cropland and built-up classes in our analysis. Classification results are summarised in Tables 3,4,5,6 and 7. The key observations are as follows:

- Since VH polarization contains volume scattering information it appears to be more useful than VV polarization.
- Though SAR based features provide similar classification accuracy, the best results are obtained by the joint use of NDVI, VV,VH features.

Overall, paddy is not classified as accurately as sugarcane. In an earlier work, we have showed that with dense VV and VH features derived from SAR images taken from different orbits the accuracy can be significantly improved. However, considering different orbits require significant amounts of pre-processing from an operational perspective.

3.2. Rabi crop cycle

The major Rabi crops are wheat and mustard. Harvesting of late sown sugarcane is also done during the same period as of Rabi crops, thus sugarcane is also included for classification. Classification results are summarized in Tables 8,9,10,11 and 12. The key observations are as follows:

- The cloud occlusion issue is less prominent for Rabi crop cycle. Due to dense NDVI time series and good cloud free data, using NDVI as the only feature results in a good classification accuracy of 91.8 %.
- SAR features perform equally well wrt classification accuracy as NDVI even if the number of Sentinel-2 tiles used are more than twice of that of Sentinel-1.
- Similiar to Kharif crop cycle observations, the VH feature is better than VV for classification. The best results are obtained by the joint use of NDVI, VV,VH features.

4. CONCLUSION

In this study we have analysed the joint use of optical and radar data from Sentinel-2 and Sentinel-1 for multi-crop classification for an Indian region. Key conclusions are: 1 Optical and radar imagery are complementary to each other, 2. Features derived from these satellites, when used jointly results in better classification accuracy. The fidelity of radar data and therefore accuracy also depends on the crop cycle. 3. SAR based features are better for paddy. 4. The selection of features for classification and corresponding improvements in the results are found to be consistent with other published studies. 5. Crops having limited ground reference data showed good classification, but need more ground reference data for these crops to improve classifier confidence.

5. REFERENCES

[1] Jordi Inglada, Arthur Vincent, Marcela Arias, and Claire Marais-Sicre, "Improved early crop type identification by joint use of high temporal resolution sar and optical image time series," *Remote Sensing*, vol. 8, no. 5, pp. 362, 2016.

[2] Amanda Veloso, St"é"phane Mermoz, Alexandre Bouvet, Thuy Le Toan, Milena Planells, Jean-Fran""¸c""ois Dejoux, and Eric Ceschia, "Understanding the temporal behavior of crops using sentinel-1 and sentinel-2-like data for agricultural applications," *Remote Sensing of Environment*, vol. 199, pp. 415–426, 2017.

[3] Sylvain Ferrant, Adrien Selles, Michel Le Page, Pierre-Alexis Herrault, Charlotte Pelletier, Ahmad Al-Bitar, Stéphane Mermoz, Simon Gascoin, Alexandre Bouvet, Mehdi Saqalli, et al., "Detection of irrigated crops from sentinel-1 and sentinel-2 data to estimate seasonal groundwater use in south india," *Remote Sensing*, vol. 9, no. 11, pp. 1119, 2017.

[4] Pradeep Kumar, Dileep Kumar Gupta, Varun Narayan Mishra, and Rajendra Prasad, "Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using liss iv data," *International Journal of Remote Sensing*, vol. 36, no. 6, pp. 1604–1617, 2015.

[5] Jitendra Singh, Umamaheswari Devi, Jagabondhu Hazra, and Shivkumar Kalyanaraman, "Crop-identification using sentinel-1 and sentinel-2 data for indian region," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 5312–5314.

Table 3. Confusion matrix using NDVI data as feature. Overall accuracy 82.1%, $\kappa = 0.75$.

Crop	Sugarcane	Paddy	Millet	Pigeon Pea	Sorghum	Non-Cropland	Built-Up	Precision(%)
Sugarcane	32	3	0	1	0	2	0	84.2
Paddy	3	7	2	0	0	0	1	53.8
Millet	0	3	5	0	2	0	1	45.4
Pigeon Pea	0	0	0	0	0	0	0	-
Sorghum	0	0	0	1	3	0	0	75.0
Non-Cropland	0	0	0	0	0	5	0	100
Built-Up	0	0	0	0	1	0	40	97.5
Recall (%)	91.4	53.8	71.4	50.0	50.0	71.4	95.2	

Table 4. Confusion matrix using VV data as feature. Overall accuracy 74.1%, $\kappa = 0.63$.

Crop	Sugarcane	Paddy	Millet	Pigeon Pea	Sorghum	Non-Cropland	Built-Up	Precision(%)
Sugarcane	33	3	1	2	4	5	2	66.0
Paddy	0	4	1	0	0	0	0	80.0
Millet	0	4	5	0	2	0	0	45.4
Pigeon Pea	0	0	0	0	0	0	0	-
Sorghum	0	0	0	0	0	0	1	0.0
Non-Cropland	0	0	0	0	0	2	0	100
Built-Up	2	2	0	0	0	0	39	90.6
Recall (%)	89.1	30.7	71.4	0.0	0.0	28.5	92.8	

Table 5. Confusion matrix using VH data as feature. Overall accuracy 83.9%, $\kappa = 0.77$.

Crop	Sugarcane	Paddy	Millet	Pigeon Pea	Sorghum	Non-Cropland	Built-Up	Precision(%)
Sugarcane	32	2	1	1	1	2	1	80.0
Paddy	1	8	0	0	0	0	0	88.8
Millet	0	1	6	0	3	0	0	60.0
Pigeon Pea	0	0	0	1	0	0	0	100
Sorghum	0	1	0	0	2	0	0	66.6
Non-Cropland	0	0	0	0	0	5	1	83.3
Built-Up	2	1	0	0	0	0	40	93.0
Recall (%)	91.4	61.5	85.7	50.0	33.3	71.4	95.2	

Table 6. Confusion matrix using VV and VH data as features. Overall accuracy 84.82%, $\kappa = 0.79$.

Crop	Sugarcane	Paddy	Millet	Pigeon Pea	Sorghum	Non-Cropland	Built-Up	Precision(%)
Sugarcane	33	3	1	1	2	0	0	82.5
Paddy	1	5	0	0	0	0	0	83.3
Millet	0	4	6	0	3	0	0	46.1
Pigeon Pea	0	0	0	1	0	0	0	100
Sorghum	0	1	0	0	1	0	0	50.0
Non-Cropland	0	0	0	0	0	7	0	100
Built-Up	1	0	0	0	0	0	42	97.6
Recall (%)	97.0	38.4	85.7	50.0	16.6	100	100	

Table 7. Confusion matrix using NDVI, VV and VH data as features. Overall accuracy 89.2%, $\kappa = 0.85$.

Crop	Sugarcane	Paddy	Millet	Pigeon Pea	Sorghum	Non-Cropland	Built-Up	Precision(%)
Sugarcane	34	2	1	1	0	1	0	87.1
Paddy	1	8	0	0	0	0	0	88.8
Millet	0	3	6	0	2	0	0	54.5
Pigeon Pea	0	0	0	0	0	0	0	-
Sorghum	0	0	0	1	4	0	0	80.0
Non-Cropland	0	0	0	0	0	6	0	100
Built-Up	0	0	0	0	0	0	42	100
Recall (%)	97.1	72.7	85.7	50.0	66.6	85.7	100	

Table 8. Confusion matrix using NDVI data as feature. Overall accuracy 91.8%, $\kappa = 0.86$.

Crop	Sugarcane	Wheat	Mustard	Precision(%)
Sugarcane	28	3	0	90.3
Wheat	0	21	1	95.4
Mustard	1	0	7	87.5
Recall(%)	96.5	87.5	87.5	

Table 9. Confusion matrix using VV data as feature. Overall accuracy 85.2%, $\kappa = 0.75$.

Crop	Sugarcane	Wheat	Mustard	Precision(%)
Sugarcane	24	2	1	88.8
Wheat	5	22	1	78.5
Mustard	0	0	6	100
Recall(%)	82.7	91.6	75	

Table 10. Confusion matrix using VH data as feature. Overall accuracy 88.5%, $\kappa = 0.80$.

Crop	Sugarcane	Wheat	Mustard	Precision(%)
Sugarcane	27	2	2	87.0
Wheat	2	22	1	88.0
Mustard	0	0	5	100
Recall(%)	93.1	91.6	62.5	

Table 11. Confusion matrix using VV and VH data as features. Overall accuracy 91.8%, $\kappa = 0.86$.

Crop	Sugarcane	Wheat	Mustard	Precision(%)
Sugarcane	28	2	1	90.3
Wheat	1	22	1	91.6
Mustard	0	0	6	100
Recall(%)	96.5	91.6	75	

Table 12. Confusion matrix using NDVI, VV and VH data as features. Overall accuracy 96.7%, $\kappa = 0.94$.

Crop	Sugarcane	Wheat	Mustard	Precision(%)
Sugarcane	29	1	0	96.6
Wheat	0	23	1	95.8
Mustard	0	0	7	100
Recall(%)	100	95.8	87.5	