

Fusion of Radar and Optical Data for Land Cover Classification Using Machine Learning Approach

David Nhemaphuki, Kiran Thapa Chetri, Sanjeevan Shrestha
davis10ge@gmail.com, thapachetrikiran@gmail.com, shr.sanjeevan@gmail.com

KEYWORDS

Google Earth Engine, Land cover, Random Forest, Sentinel, Synthetic Aperture Radar

ABSTRACT

This study evaluates the advantages of combining traditional space borne optical data with longer wavelengths of radar for land cover mapping. Land cover classification was carried out using Optical, radar data and combination of both for the Bardiya district using Random Forest algorithm. The fusion of optical and radar shows better land cover discrimination with 96.98% overall accuracy in compared to using radar data and optical data separately with overall accuracy of 69.2% and 95.89% respectively. Additionally, the qualitative result demonstrates that the combined utilization of optical and radar imagery yields useful land cover information over those obtained using either type of image on its own.

1. INTRODUCTION

Land cover information can provide a basis for formulating ecological protection measures and implementing sustainable development. Land Use Land Cover (LULC) dynamics serves as a crucial parameter in current strategies and policies for natural resource management and monitoring. Remote sensing technology has become the primary means of land cover information acquisition and has the advantages of high data acquisition and updating speed, wide range, economic convenience, and rich spatial information (William & Adriano, 2017). Land cover classification using traditional optical remote sensing approaches is not satisfactorily accurate since in certain conditions, similar spectral responses make classes difficult to separate like regions that are mixed, show irregular geometrical forms and change frequently (Soria-Ruiz, et. al.,2010). Moreover, optical

remote sensing applications are often limited by cloud cover. In particular, the all-weather capability of synthetic aperture radar (SAR) allows for more reliable earth observation data under cloud covered regions. However, the information content is different from the optical and sometimes difficult to interpret as it is unable to distinguish between the multiple targets. Since optical and radar images vary in spectral, spatial and temporal resolution, the combination of both offer a more complete perception of the target objects and more reliable results (Pohl & Van Genderen, 1998).

There have been reported improvements in crop classification when both visible infrared and active microwave sensor data are used together (Rosenthal, et. al., 1985) (Qinghua, et. al., 2001) (McNairn, et. al., 2002); advantages of combining multivariate visible, infrared, and SAR sensor datasets have

also been demonstrated (Brisco & Brown, 1995); improvement in accuracy land cover/use mapping was seen when radar data was combined with optical data (Soria-Ruiz, et. al., 2010) (Sim, et. al., 2013) (Joshi, et. al., 2016). Since, among other parameters, plant structure and water content influence the response data for SAR sensors, it is reasonable to assume that improved land-cover classifications of vegetated areas could be obtained using radar and optical data.

The specific objectives of this study are: (a) to classify a land cover map, using optical and radar images from Sentinel-1 and Sentinel-2 respectively; (b) to compare the accuracy of the classification while using radar, optical and both images. Overall, the purpose of the study is to determine whether SAR imagery, when combined with optical imagery, can provide acceptably accurate information about the distribution of important land-cover types, including agricultural, forest, water bodies, urban areas, and barren land and how the result would vary with either type of sensor separately.

2. MATERIALS AND METHODS

The study examines land cover classification and analysis using radar and optical data. To obtain a final land-cover map, the methodology followed these major processing steps: selection of study area, image preprocessing, preparing composite images, selection of samples, hyperparameter tuning for the Random Forest algorithm, land cover map preparation and finally the accuracy assessment.

2.1 Study Area

Bardiya lies in Province No. 5 in mid-western Nepal that covers 2025 square kilometers in area (Figure 1). Most of Bardiya is in the fertile Terai plains, covered with agricultural land and forest.

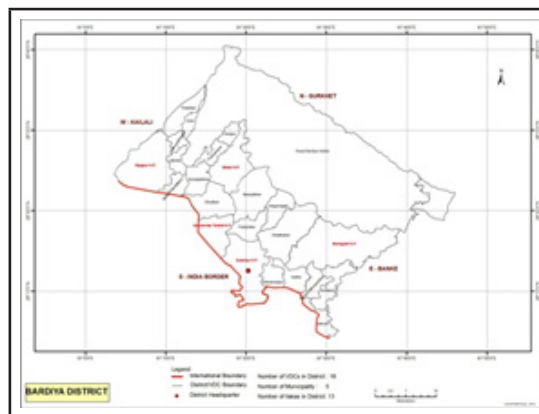


Figure 1: Study Area

(Source: LGCDP)

2.2 Data and Software

Table 1 and Table 2 present the satellite images used and the software used for different tasks in our study respectively.

Table 1: Satellite Images used in the study

S.N.	Satellite Imagery	No. of Bands	Date	Download Source
1	Sentinel-1 C-band	2	October 2020	Copernicus Open Access Hub
2	Sentinel-2 TOA	16		

Table 2: Software Used in the study

S.N.	Software	Usage
1	Google Earth	Sample Collection
2	Google Earth Engine (GEE)	Image Analysis and Classification
3	ArcGIS	Map Preparation

Ground Truth Collection: A total of 60 random samples or regions of interest (ROIs) for each land cover classes were collected. The samples were divided into training (40 samples), validating (10 samples) and testing datasets (10 samples).

2.3 Image Preprocessing

Preprocessing of the images is to be done for reducing any noises that may be present in the images before processing for analysis.

Speckle is a granular 'noise' that inherently exists in and degrades the quality of SAR images and it is reduced by applying spatial filters (Tadesse & Falconer, 2014). It applies a mathematical calculation on the pixel values within the moving window and replaces the central pixel with the new value. Applying speckle filtering reduces visual appearance of speckle and applies smoothing effect. A morphological mean filter was applied to each band with radius 50m.

Sentinel-2 TOA image was masked for cloud coverage area using the information contained in the 16th band QA60. Since Sentinel-2 TOA is reflectance data for the top of atmosphere it is already corrected from variance caused by sun angle, daytime etc and is ready for use.

2.4 Calculation of Normalized Difference Vegetation Index (NDVI)

The NDVI is a numerical indicator which uses visible and near-infrared waves as there is high interaction of energy in the visible and near infrared regions of the electromagnetic spectrum in vegetation. Generally, healthy vegetation absorbs most of the visible light and reflects a large portion of near-infrared light while unhealthy or sparse vegetation reflects more visible light and less near-infrared light. Bare soils on the other hand reflect moderately in both portion of the electromagnetic spectrum (Tadesse & Falconer, 2014). NDVI is calculated from these individual measurements as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

The value of NDVI ranges from -1 to +1 but in practice extreme negative values represent water, values around zero represent bare soil

and values over 0.6 represent dense green vegetation. NDVI can be used to distinguish the different land cover types in comparison to other indices which can distinguish a specific land cover e.g. Normalized difference water index (NDWI), Normalized difference built-up index (NDBI) etc.

2.5 Composite Images

Three sets of composite images were prepared for performing the classification. The first set contains the information of SAR image, second set contains Optical information and the last set contains the SAR and Optical information.

Table 3: Composite Images

S.N.	Composite Image	Bands
1	SAR Image	VV and VH
2	Optical Image	Band 2, Band 3, Band 4, Band 8 and NDVI
3	SAR and Optical Image	VV, VH, Band 2, Band 3, Band 4, Band 8 and NDVI

2.6 Land Cover Map Preparation

Preparation of Land cover map is done by Random Forest algorithm. It is an ensemble tool which takes a subset of observations and variables to build decision trees. It builds multiple such decision trees and amalgamates them together to get a more accurate and stable prediction. And using this method, studies have shown a significant improvement in classification accuracy for land cover classification. While using Random Forest algorithm for the classification it includes the following steps:

Hyperparameter tuning: It is the problem of choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process. Random Forest has two major parameters i.e. number of decision tree and number of feature considered by each tree when splitting the node. Tuning was performed

for the parameter number of decision trees to grow and accuracy was assessed for each image using the validation dataset as shown in Table 4.

Table 4: Hyperparameter Tuning

S.N.	No. of trees	SAR Image	Optical Image	SAR + Optical
1	1	54.10	93.00	93.90
2	5	64.40	93.00	96.60
3	10	67.00	93.80	96.90
4	15	67.80	93.60	96.60
5	20	69.70	93.80	96.40
6	25	69.00	93.60	96.50
7	40	69.50	93.50	96.20

For the number of trees the value 20 was selected as an optimal value as it provides good accuracy in all classified images. A default value was assigned for the parameter number of feature i.e. square root of the total number of features.

2.7 Classification

The Random Forest model was trained using the training sample with the parameters: value of 20 for number of trees and default value for the number of features. The model was used to classify the land cover type from the SAR image, Optical Image and SAR + Optical Image.

2.8 Accuracy Assessment

Accuracy of the classification of Land cover map was performed for all three land cover maps using the test dataset. A much better way to evaluate the performance of a classifier is to look at the confusion matrix. Confusion matrix was obtained and overall accuracy, kappa coefficient; users and producer's accuracy were derived.

Overall accuracy is the probability that an individual will be correctly classified by a test; that is, the sum of the true positives plus

true negatives divided by the total number of individuals tested. The Kappa coefficient is a statistical measure of inter-rater reliability or agreement that is used to assess qualitative documents and determine agreement between two rasters. Producer's Accuracy is the map accuracy from the point of view of the map maker. The producer's accuracy can be termed as recall or true positive rate or sensitivity. The User's Accuracy is the accuracy from the point of view of a map user, not the map maker. The user's accuracy can be termed as consumer's accuracy or precision or positive predictive value.

3. RESULT AND DISCUSSION

Land cover maps for the month of October 2020 using SAR image, Optical Image and SAR + Optical image were classified using Random Forest as shown in Figures 2, 3 and 4 respectively.

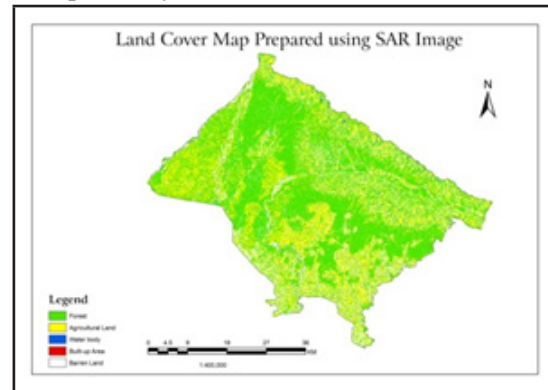


Figure 2: Land cover map (SAR)

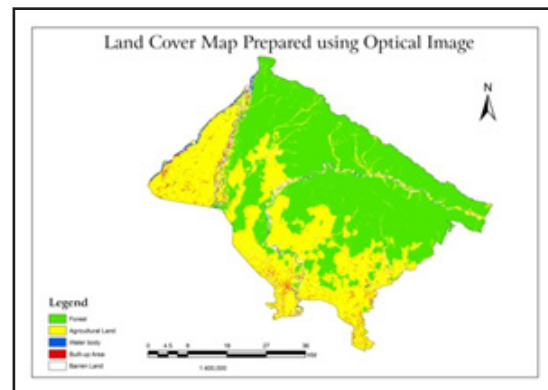


Figure 3: Land cover map (Optical)

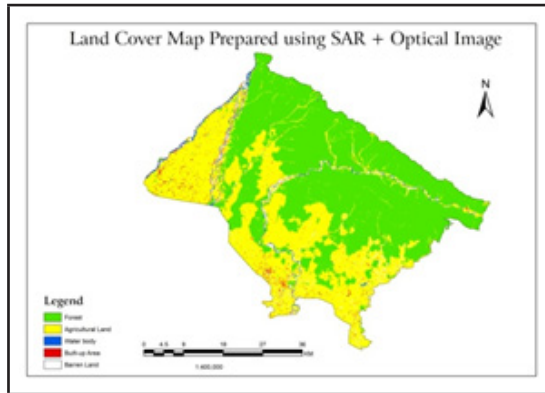


Figure 4: Land cover map (SAR + Optical)

The accuracy of land cover map obtained by using SAR image was 69.29%, Optical image was 95.89% and SAR + Optical Image was 96.98%. Land cover map using SAR image had the least accuracy while there is significant increase in accuracy of the land cover maps obtained using Optical and SAR + Optical image. An increment about 1% in accuracy was observed while using SAR image with Optical image than the Optical image only.

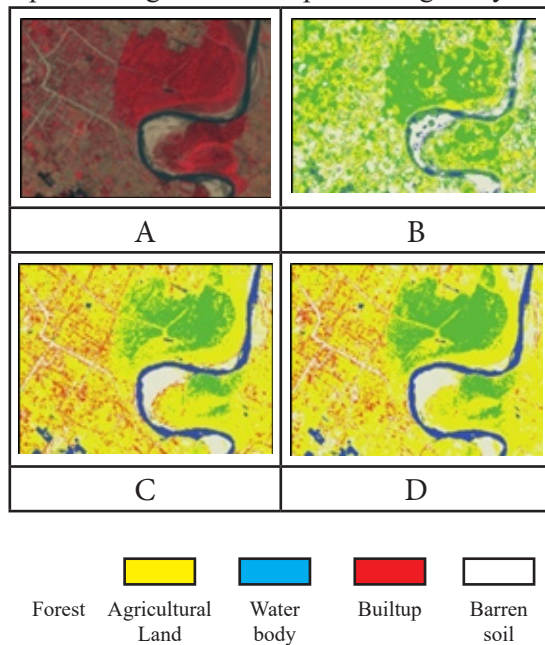


Figure 5: A portion of the study area is shown which covers the built-up area and Babai River. A shows false color composite image

where vegetation is represented by red color, water body by dark pixels and bare soil by light pixels; B, C and D represents the classified land cover map using SAR image, Optical image and SAR and optical image respectively.

Using SAR data only the features like forest and water body were distinguishable but other classes were not. Optical data significantly improved the land cover map than the SAR as it was able to distinguish the built-up area, barren land and agricultural area as well. With the addition of SAR data to the Optical data an improved land cover map was obtained as pixels misclassified as built-up area decreased. The producer's and user's accuracy was calculated for the land cover map obtained using SAR + Optical images (Figure 5).

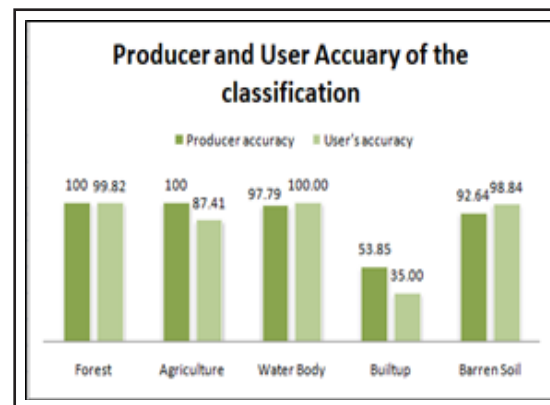


Figure 6: Producers and Users accuracy

Producer's and user's accuracy for each land cover classes were high except for the built-up class. The producer's accuracy was 53.85% while the user's accuracy was 35.00% only, which shows there is high percentage of misclassification in built-up class. About 31% of built-up was misclassified as barren land and about 15% as agricultural area while 65% of area classified as built-up area represents barren land which shows that there is high percentage of misclassification between built-up area and barren land classes.

4. CONCLUSION

This paper focuses on using combined SAR and optical image for land cover classification using Random Forest algorithm and assesses the accuracy of land cover maps prepared using SAR image, Optical image and combination of both images.

The accuracy obtained from land cover map using SAR image was the least while accuracy obtained using optical and combination of both images were found to be high. Also, there was an increment in the accuracy of the land cover map obtained by using optical and SAR image than the optical image only which shows that an improved land cover classification can be obtained providing detailed land cover classes.

REFERENCES

- Ali, J., Khan, R., Ahmad, N., Maqsood, I., (2012). *Random forests and decision trees*.
- Brisco, B. And Brown, R., (1995). *Multidate SAR/TM synergism for crop classification in western Canada. Photogrammetric Engineering and Remote Sensing*, 61, pp. 1009–1014.
- Grinand, C., Rakotomalala, F., Gond, V., Vaudry, R., Bernoux, M., Vieilledent, G., (2013). *Estimating deforestation in tropical humid and dry forests in Madagascar from 2000 to 2010 using multi-date Landsat satellite images and the random forests classifier.*, 139, 68–80.
- Horning, N., (2010). *Random Forests: An algorithm for image classification and generation of continuous fields data sets. In Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*.
- Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., ... & Waske, B. (2016). A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, 8(1), 70.
- McNairn, H., Ellis, J., Van Der Sanden, J., Hirose, T. And Brown, R.J., (2002). *Providing crop information using Radarsat-1 and satellite optical imagery. International Journal of Remote Sensing*, 23, pp. 851–870.
- Model, T. and Srivastava, T., (2020). *Random Forest Parameter Tuning | Tuning Random Forest*.
- Pohl, C. and Van Genderen, J.L., (1998). *Multisensor image fusion in remote sensing: Concepts, methods and applications. International Journal of Remote Sensing*, 19, pp. 823–854.
- Qinghua, Y., Zhizhong, L., Hongbin, F., Maocheng, Y., Zhigans, S., Shuenxi, L. And Wenzhang, Q., (2001). *Application Of Radar Image In Land Use Dynamic Monitoring In Beijing*.
- Rosenthal, W.D., Blanchard, B.J. and Blanchard, A.J., (1985). *Visible/infrared microwave agriculture classification, biomass and plant height algorithms. IEEE Transactions on Geoscience and Remote Sensing*, 23, pp. 84–90.
- Sahebjalal, E., & Dashtekian, K., (2013). Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods. *African Journal of Agricultural Research*, 8(37), 4614-4622.

Sim, C. K., Abdullah, K., Jafri, M. Z. M., & San Lim, H. (2013). Combination of radar and optical remote sensing data for land cover/use mapping. In *2013 IEEE International Conference on Space Science and Communication (IconSpace)* (pp. 224-227). IEEE.

Soria-Ruiz, J., Fernandez-Ordonez, Y. and Woodhouse, I., (2010). *Land-cover classification using radar and optical images: a case study in Central Mexico*.

Tadesse, H. & Falconer, A., (2014). *Land Cover Classification And Analysis Using Radar And Landsat Data In North Central Ethiopia*.

William E., Adriano C., (2017). *Introduction to Satellite Remote Sensing*



Author's Information

Name	: David Nhemaphuki
Academic Qualification	: Qualification: Bachelor in Geomatics Engineering
Current Organization	: Survey Department of Nepal
Current Designation	: Survey Officer
Work Experience	: 5 years
Published paper/article	: None
E-mail	: davis10ge@gmail.com