Topographical Feature Extraction Using Machine Learning Techniques from Sentinel-2A Imagery

Kuldeep Chaurasia, Baipureddy Neeraj, Dattu Burle, Vipul Kumar Mishra
Department of Computer Science Engineering
Bennett University, Greater Noida, INDIA
(kuldeep, BN7877, BD6043, vipul.mishra) @bennett.edu.in

***Abstract:* -** The advancement in the satellite technology has made it possible to easily and frequently obtain the satellite images of most of the regions on the Earth. The satellite data contains the abundant amount of information which can be very useful for variety of societal applications. However**,** manual identification of the land cover in a particular area is a very challenging and time-consuming task**.** In this manuscript, an attempt has been made to better extract the landcover types from Sentinel-2A imagery using the popular classifiers such as random forest, SVM, Naive Bayes, Decision Tree (CART). The manuscript also validates the results obtained by the used models by computing the performance metrics. The analysis reveals that random forest classifier outperforms the rest of the classification methods in terms of better accuracy of 95.67%. This automated approach can be applied to large sets of data, reducing the need for manual labeling.

Keywords: Remote sensing, Machine Learning, Feature extraction, Satellite Image, Landcover mapping

I. INTRODUCTION

The increasing availability of satellite sensors have augmented the need for development of the methods for efficient extraction of information from the satellite images. With the development in the satellite technology, huge amount of satellite data is captured everyday by various sensors and made available in public domain for research purpose. One of the such repository provide the Sentinel images of the Earth. Under the European Space Agency's Copernicus programme, Sentinel-2A optical imaging satellite was launched in 2015. The sentinel-2A satellite carries a wide swath high-resolution multispectral image sensor consisting of 13 spectral bands [2]. The satellite provides the images with the spatial resolution of 10m, 20m, and 60m. The image dataset is widely used for various applications including forest monitoring, land use mapping, disaster mapping & management etc. The Sentinel image dataset can be downloaded from Earth Explorer [3]. The Multi Spectral Instrument (MSI) is carried by the Sentinel-2 satellite with a swath width of 290 km. This instrument consists of 13 spectral bands in wavelength range which includes Visible, Near infrared and Short Wave Infrared Spectral (SWIR) bands. The spectral range of the bands lies between 443–2190 nm. The four spectral bands (visible and near-infrared) with 10 m spatial resolution, six spectral bands (red edge and shortwave infrared bands) with 20 m and three spectral bands (atmospheric correction bands) with 60 m spatial resolution are captured by MSI. Here the satellite image which has Band-3 (Green), Band-4 (red), and Band-8 (NIR), is used and mapped as blue green and red [4].

II. Related work

Koen et al. [1] proposed an approach for classifying the terrain features in Landsat images using Random Forests algorithm. In this paper they classified the Landsat images having spatial resolution of 30 m. The authors utilized the open street map to generate training data for the classifier and to mark the areas towards obtaining the statistics about the underlying terrain. Andualem et al. [9] carried out a research on Landuse change detection using remote sensing techniques. They adopted the Landsat images of year 2007 and 2018 of Amhara region to detect the change in agriculture, urban area, forest area etc. Borras et al. [10] carried out a research for land use classification using sentinal-2A satellite image. Author (s) utilised the True Colour Composite (TCC) image of the two different agricultural areas located in Valencia (Spain) and Buenos Aires (Argentina). Similar research using Sentinel images of Rome city was also carried out by Majidi et al. [11,12,13,14,15]. There are several methods which were implemented using Landsat imagery by various researchers can be found in literatures [16, 17, 18, 19, 20]. The objectives of the research work is to understand the general procedure of land cover classification from satellite images and to extract the various land cover features from the sentinel-2A satellite image using Supervised classification methods followed by the accuracy assessment of each method used in the analysis.

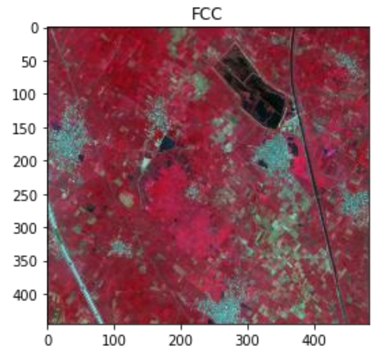
III. STUDY AREA AND DATA RESOURCES

The satellite image used here belong to Greater Noida region, of Uttar Pradesh state, India. The upper left and lower right scene coordinates are 77.547°E, 28.628°N and 77.596°E, 28.587°N respectively. The details of the satellite data used in the analysis is summarised in Table 1.

Table 1. Details of satellite data used

Satellite	Spatial Resolution (m)	Central Wavelength (µm)	Date of Acquisition
Sentinel-2A	10	Band-3 (Green) - 0.560 Band-4 (red) - 0.665 Band-8 (NIR) - 0.842	17/10/2019

The image contains various landcover features including cropland, bare-land, Built-up region, water bodies, roads etc.



(a)

(b)

Fig. 1. Input dataset (a) FCC (False Colour Composite) image (b) ROI (Region of Interest) Training Data.

Land use is one of the most vital factors which plays an important role in estimation of parameters such as evapotranspiration, runoff, and surface erosion in the watershed**.** The satellite dataset of the study area has been obtained from USGS Earth explorer portal (https://earthexplorer.usgs.gov). The dataset with spatial resolution of 10 m and cloud coverage of less than 10% has been used in the analysis. Training data is created using QGIS software.

IV. Methodology

It can be seen in the False Colour Composite (FCC) image that the red colour is of vegetation region; this is because of the NIR (Near Infrared) band as it has high reflectance for the vegetation region. The FCC clipped satellite image of Greater Noida region is shown in Fig. 1(a). The methodology of the current research work includes mainly four steps. Firstly, the FCC (false colour composite) clipped satellite image has been geo-referenced with reference to UTM zone 43. The training samples for the input dataset was generated using QGIS [5]. Here there are 28 training samples which have 5 classes in it i.e. Crop land, bare land, Water body, Built-up region, Road. The ROI Data is shown in Fig. 1(b). The Python programming environment has been utilised to implement the classification models. The python code after execution, classify the image into five different classes. The entire flow of the process has been summarised in Fig. 2. The proposed study implements the four classification methods viz. Naive Bayes (maximum likelihood), Decision Tree (CART), Support Vector Machine (SVM)[6,7, 8], and Random Forest [1].

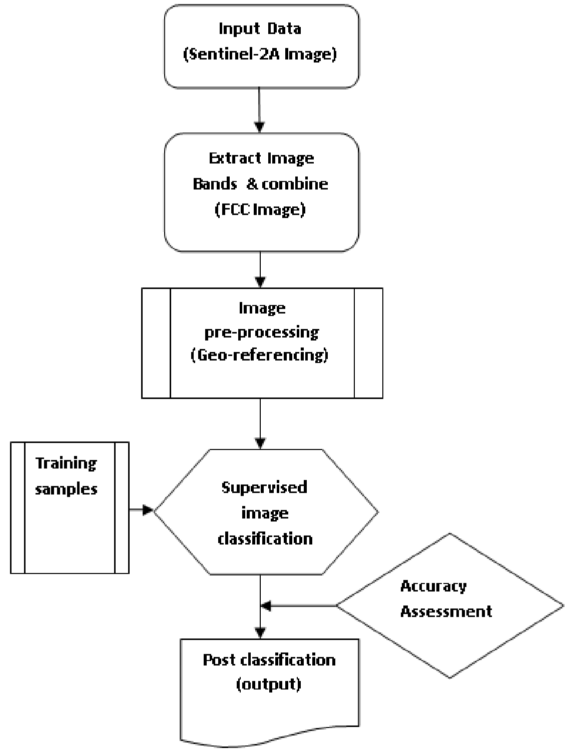


Fig. 2. Methodological flowchart

V. Results and Discussion

In this research work, the input satellite dataset has been used for land use detection for classifying the landcover features into 5 categories. Four types of classifiers are used to classify the satellite image and the best one is chosen after evaluating each classifier based on accuracy. The output image using the Naive Bayes classifier has been shown in Fig. 3.

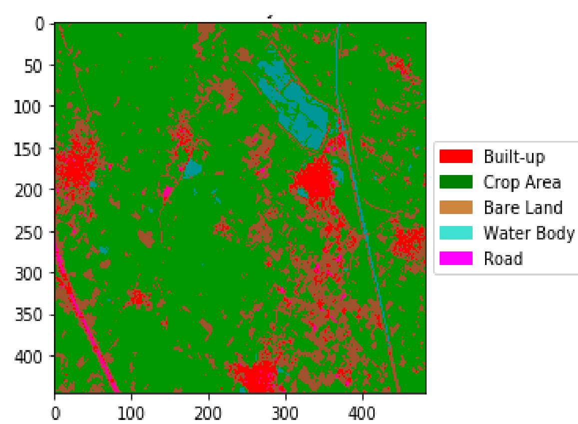


Fig. 3. Output image after using Naive Bayes classifier

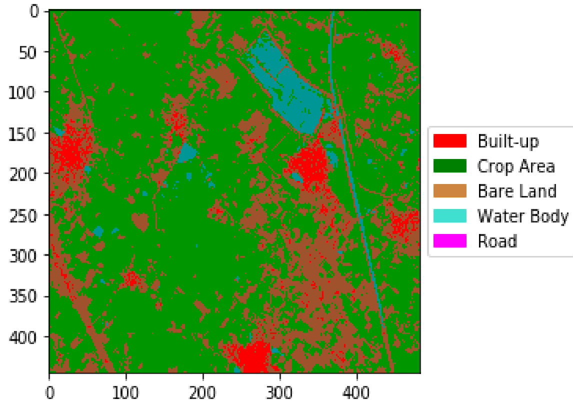


Fig. 4. Output image after using Decision Tree Classifier (CART)

The output image using the Decision Tree Classifier has been shown in Fig. 4. The decision tree classifier has better classified the land features such as bare land and water body when compared to output image with Naive Bayes classifier. The absence of the pink color pixels in Fig. 4 indicates that Decision Tree Classifier is unable to classify the Road feature which is clearly picked up by the Naive Bayes classifier in Fig. 3. A support vector machine classifier has also been implemented to extract the landcover features into five classes. The output image using the SVM classifier has been shown in Fig. 5.

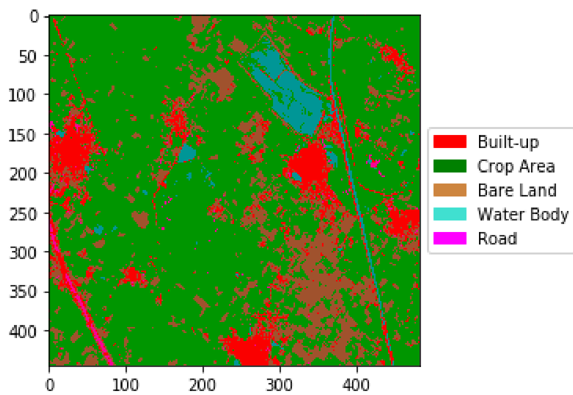


Fig. 5. Output image after using the SVM classifier

The comparison of Fig. 3, Fig. 4 and Fig. 5 reveals that there are more samples classified as Built-up region in Fig.5 and also there are samples which are classified as Road than Fig. 3, Fig. 4, So SVM has performed better than the decision tree and Naïve Bayes as visualized from the output images. Random Forest is an extension over bagging where in each classifier in the ensemble is a decision tree classifier obtained by randomly selecting the attributes at each node to determine the split. The most popular class is obtained based on the votes from each tree during the classification [14]. Random forest with 500 trees has been found as the optimal parameter to be used in our model. The output image using the Random Forest classifier has been shown in Fig. 6.

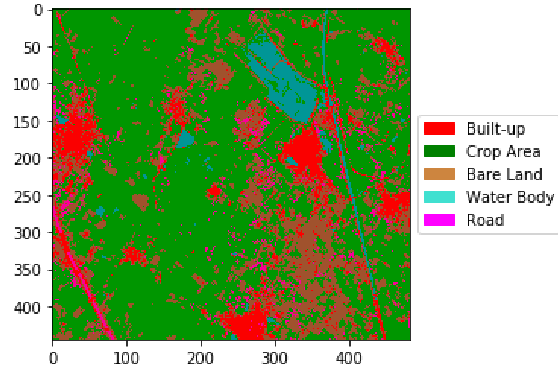


Fig . 6. Output image after using Random Forest classifier

Table 2. Details of Accuracy Assessment

S.No.	Model	Parameters used	Accuracy
1	Naive Bayes	GaussianNB, priors= none, Var_smoothing= 1e-9	**0.7993**
2	Decision Tree (CART)	Criterion = Gini, Random_state = 100 max_depth=3, min_samples_leaf = 5	**0.8335**
3	SVM	C=1, Kernel=linear, tolerance=0.001	**0.8960**
4	Random Forest	No. of trees = 500 oob score = True	**0.9567**

Table 2. summarizes the accuracy of all the four methods along with the details of the parameters used with the various models. we can see that the **Random Forest** classifier has highest accuracy (OOB (out of bag accuracy score) of 0.9567) when compared to other classification methods. Random forest performed well because it employs many decisions' trees for classification. Bagging and feature randomness parameters are utilized by the classifier when building each individual tree to try to create an uncorrelated forest of trees. The prediction of the created forest by committee is more accurate than that of any individual tree. The classification accuracy (in %) obtained using various models is shown in Fig. 7.

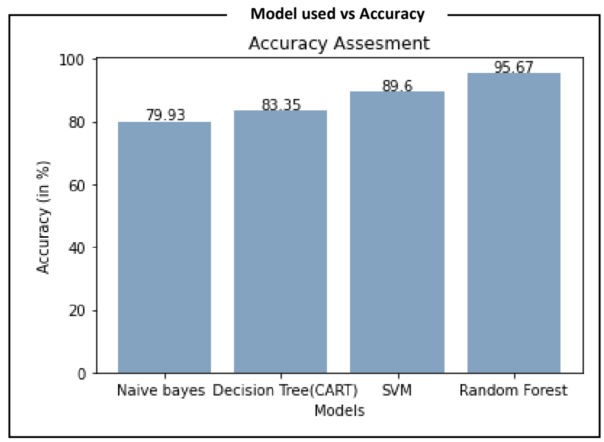


Fig. 7. Models and their classification accuracy

VI. CONCLUSIONS

We have presented a way to classify the landcover types in Sentinal-2A satellite imagery. The F**CC** image was generated and used as input in the classification process due to its high reflectance to vegetation region. The research work implements the four types of classifiers to classify the input image into 5 classes which enable us to extract the bare lands, crop lands, water bodies, built-up regions and road. And among the four classifiers, the **Random Forest** classifier outperformed with an accuracy of **95.67%**. This automated approach can be applied to large sets of data towards reducing the need for manual extraction of the landcover features**.** The future work may include the extraction of crop types including wheat paddy etc. in the crop region by incorporating the multi temporal dataset. The use of hybrid machine learning methods can also be adopted towards further improvement in the classification accuracy.

REFERENCES

[1] Koen Basten, Fabian Gieseke, Arjen de Vries "Classifying Landsat Terrain Images via Random Forests" Bachelor thesis Computer Science, Radboud University, August 15, 2016.

[2] Sentinel-2A Specification https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/sentinel-2a/. Accessed :- 28/10/2019.

[3] Sentinel-2A dataset Available :- https://earthexplorer.usgs.gov . Accessed :- 2/11/2019.

[4] Why False color composite (FCC) ? Available :- http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson3-1/composites.html Accessed :- 2/11/2019.

[5] QGIS Software Available :- https://qgis.org/en/site/forusers/download.html Accessed :- 2/11/2019.

[6] Weston, Jason, and Chris Watkins."Multi-class support vector machines". Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May, 1998.

[7] Crammer, Koby, and Yoram Singer. "On the algorithmic implementation of multiclass kernel-based vector machines." The Journal of Machine Learning Research 2 (2002): 265-292.

[8] Tsochantaridis, Ioannis, et al. "Large margin methods for structured and interdependent output variables." Journal of Machine Learning Research 6.2 (2006): 1453.

[9] Andualem, Tesfa Gebrie, Gizew Belay, and Adebabay Guadie. "Land use change detection using remote sensing technology." *J. Earth Sci. Clim. Chang* 9 (2018): 1-6. doi: 10.4172/2157-7617.1000496

[10] BORRÀS, J. et al. Land use classification from Sentinel-2 imagery. **Revista de Teledetección**, [S.l.], n. 48, p. 55-66, june 2017. ISSN 1988-8740. Available at: <https://polipapers.upv.es/index.php/raet/article/view/7133>. Date accessed: 24 dec. 2019. doi:https://doi.org/10.4995/raet.2017.7133.

[11] M. Majidi Nezhad, et al. "land cover classification using sentinel-2 images" a case study in city of rome, Rome, Italy – April, 2019 Paper No. ICEPTP 158 DOI: 10.11159/iceptp19.158.

[12] M. Cavur, et al. "LAND USE AND LAND COVER CLASSIFICATION OF SENTINEL 2-A": ST PETERSBURG case study, 16–17 Sept. 2019, Warsaw, Poland, Available at: https://doi.org/10.5194/isprs-archives-XLII-1-W2-13-2019.

[13] Hongrui Zheng et al. "Performance Evaluation of Downscaling Sentinel-2 Imagery for Land Use and Land Cover Classification by Spectral-Spatial Features" *Remote Sens*. **2017**, *9*(12), 1274; Available at : https://doi.org/10.3390/rs9121274.

[14] Phan thanh noi, et al. "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery" *Sensors* **2018**, *18*(1), 18; Available at : https://doi.org/10.3390/s18010018.

[15] Wahyu catur wibowo, et al."Classification of Land Cover from Sentinel-2 Imagery Using Supervised Classification Technique (Preliminary Study)" 3-5 september 2018, Bina Nusantara University, Jakarta, Indonesia.DO - 10.1109/ICIMTech.2018.8528122. Availabe at : : https://www.researchgate.net/publication/328908507_Classification_of_Land_Cover_from_Sentinel2_Imagery_Using_Supervised_Classification_Technique_Preliminary_Study

[16] Alam, A., Bhat, M.S. & Maheen,"Using landsat satellite data for assesing the land use and land cover change in Kashmir valley" M. GeoJournal (2019).Available at : https://doi.org/10.1007/s10708-019-10037-x

[17] Eldiiar Duulatov, et al., "Detecting land use/land cover change using landsat imagery" Jumgal district, Kyrgyzstan JO - International Journal of Geoinformatics 12(4):1-7 · December 2016, Available at : https://www.researchgate.net/publication/316550592_Detecting_land_useland_cover_change_using_landsat_imagery_Jumgal_district_Kyrgyzstan

[18] Mustafa guler, et al., "Using landsat data to determine land use/land cover changes in Samsun, Turke",DOI : 10.1007/s10661-006-9270-1, Environmental monitoring and assessment 127(1-3):155-67 · May 2007, Available at : https://www.researchgate.net/publication/6869811_Using_landsat_data_to_determine_land_useland_cover_changes_in_Samsun_Turkey

[19] W. Muttitanon & N. K. Tripathi Corresponding author (2005) Land use/land cover changes in the coastal zone of Ban Don Bay, Thailand using Landsat 5 TM data, International Journal of Remote Sensing, 26:11, 2311-2323, DOI: 10.1080/0143116051233132666

[20] Tengyun Hu, et al., "Mapping Urban Land Use by Using Landsat Images and Open Social Data" : 17 February 2016,Remote Sens. 2016, 8, 151; doi:10.3390/rs8020151, Available at : file:///C:/Users/AAO%20ZP%20VSP/Downloads/remotesensing-08-00151.pdf