Deforestation analysis using Random Forest and interactive supervised classification approach

Deforestation analysis using Random Forest and interactive supervised classification approach

Deforestation analysis using Random Forest and interactive supervised classification approach

Yogender[1], Sona Guliyeva[2], Elman Alaksarov[2], Arunima Singh[3], S.K.P. Kushwaha[4]

[1]Department of Earth Observation Science, Faculty ITC, University of Twente, Enschede, Netherlands
[2]Department of Earth Observation, Azercosmos, Baku, Azerbaijan
[3]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic
[4]Geomatics Group, Department of Civil Engineering, Indian Institute of Technology, Roorkee, Uttarakhand, India

*Corresponding author email: yogender105@gmail.com

Abstract. The increasing demand for forest resources leads to overexploitation to a greater extent, raising the alarm for the environmental consequences. The excessive use of forest resources results in deforestation, which needs to be addressed for maintaining the natural ecosystem balance. This research focuses on the deforestation analysis of a territory in the Tartar district in Azerbaijan. The changes in the forest area have been analyzed using the high-resolution Azersky satellite datasets for the three alternative years, 2017, 2019, and 2021. Two classification approaches, namely Random Forest (RF) and NDVI- based interactive supervised classification, were implemented for this purpose. The statistical analysis of the results indicates the gradual decrease in the forest area from the year 2017 to the year 2021, which has been evaluated by visual interpretation through the change maps of the forest area. From RF classification results, it has been observed that there has been an overall decrease of 9.5% from 2017 to 2021. Also, the NDVI-based interactive supervised classification approach indicates an overall deforestation rate of 4.79% from 2017 to 2021. This work shows that the forest area in the study region has considerably reduced over the years, and there is a need to closely monitor deforestation in the considered study area.

Keywords: *forest change analysis, image classification, random forest, NDVI-based interactive supervised classification*

1. Introduction

Deforestation is a matter of concern at a global level as it results in the loss of biodiversity and affects the lives of the people living in the ecosystem [1]. Frequent activities like the lopping of woods and illegal poaching are responsible for the degradation of forest areas [2]. The extent of deforestation is how the forest area has been affected while carrying out ecological activities [3]. Remote sensing techniques have been the economic source of high-quality datasets for earth surface monitoring over the past ten years [4]. Remote sensing techniques provide robust and time-efficient solutions for the large-

scale monitoring of the forests compared to the ground surveys [5]. Remote Sensing and GIS methods have been widely used to classify Land Use and Land Cover (LULC) with the aid of different types of datasets. Deforestation can be monitored and quantified with the satellite datasets from different timestamps, suitable to the scale of the analysis [6].

Among all categorization methods, supervised classification, unsupervised classification, random forest, and fuzzy classification are the most employed for image classification [7,8]. The rapid rate of ongoing developments has led to the overexploitation of natural resources resulting in deforestation, soil erosion, forest fires, etc. [9]. The primary factors accounting for deforestation are rapid socioeconomic developments, urbanization, expansion of agricultural land, and increasing demand for forest products which led to forest exploitation over the global scale [10]. The various organizations over different regions have constantly monitored the reduction in the forest area.

Out of numerous novel solutions for mapping and monitoring the forest land cover, a digital image classification approach was selected for this project. In this work, Azersky satellite data from 2017, 2019, and 2021 were used for the classification using the Normalized Differentiation Vegetation Index (NDVI) based on supervised interactive classification and the Random Forest (RF) approach [11]. As the work focuses on the forest area only, all the datasets were classified into two major classes, namely forest and non-forest. A supervised classification method was implemented with ten training samples for each class for both approaches. The change in the forest area has been calculated with statistical means and supported by the visual inspection of the maps showing the change detection over different time stamps.

2. Study Area and dataset description

The study area was a forest territory in the Tartar district in the Upper Karabakh region in Azerbaijan. The extent of the selected forest region is about 11377874 m^2 and is mainly contributed by vegetation, bare land, and a water body. The alarming rate of deforestation in the Tartar district around the selected territory was the key motivation for this work with this study area.

The dataset for three different years was used for the deforestation analysis in this work. The multispectral datasets from the Azerksy satellite with a resolution of 6m for the years 2017, 2019, and 2021 were provided by the department of earth observation, Azercosmos, for this work. The dataset is shown in figure 1.

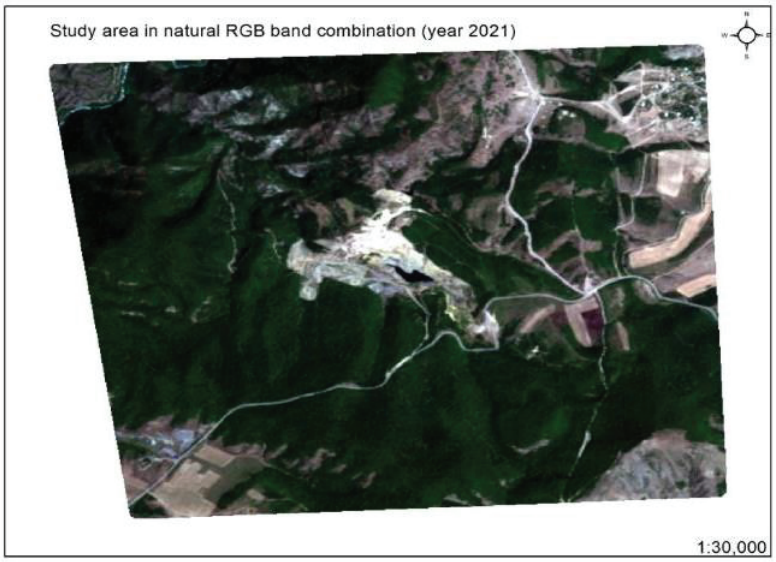


Figure 1: Study area dataset for the year 2021 shown in natural RGB band combination

3. Methodology workflow

3.1. Data pre-processing:

The raw dataset from Azersky for the study area had a black background that was contributed by pixel with zero values, and it was unnecessary for the classification purpose. In the initial processing step, the background was removed from all three datasets to maintain the homogeneity of the forest and non-forest pixels. This step was carried out in ArcMap software with the "*Copy raster*" tool.

3.2. Image classification with NDVI-aided interactive supervised and RF approach

NDVI was calculated for all the three-year datasets, and the image was re-classified based on a cut-off value of 0.6 for the high vegetation/forest class. All the pixels with NDVI values >0.6 were classified into the forest and other pixel values into non-forest. Training samples were collected from the datasets based on NDVI values for three corresponding datasets. For the supervised classification, twenty training samples for each class were considered with distribution throughout the study area, followed by interactive supervised classification in ArcMap software.

The supervised RF classification approach was selected because of its faster execution time and better accuracy as the second approach for classification in this work. The SNAP software was used for the execution because of the simplified user interface and faster processing. TiFF files from the pre-processing step were used as input in this step. Twenty training samples were collected for each forest and non-forest feature class based on the spectral profile of the datasets. Supervised RF classification was then implemented on all three datasets, and a classifier accuracy report was generated.

3.3. Forest area calculation

All the classified raster images are converted to Vector shapefiles using the "*raster to polygon*" data conversion tool in the ArcMap system toolbox. Since converted vector shapefiles are in the Global Coordinate System (GCS) WGS 1984 system, they need to be reprojected in the Projected Coordinate System (PCS) for the geometry and area calculations. The area was calculated using the "calculate geometry" tool from the attribute table in ArcMap software. Finally, all the attribute tables were exported as a .*txt* file in Microsoft Excel for the analysis and area calculations.

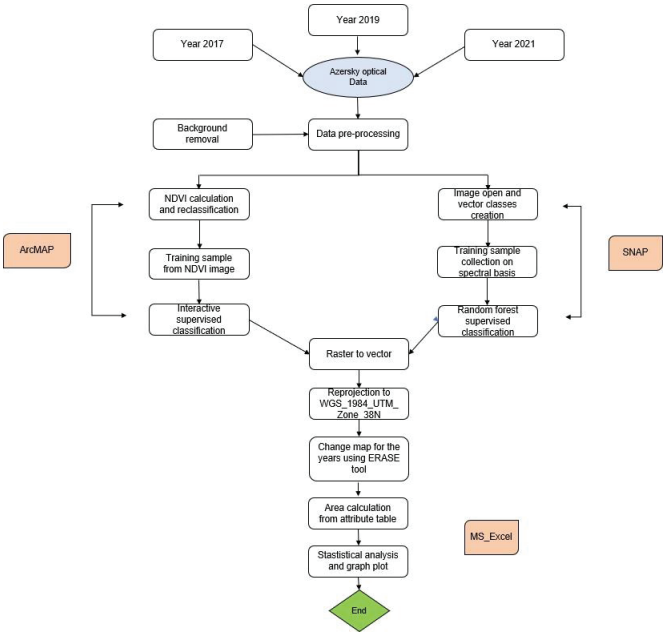


Figure 2: Methodology flowchart

4. Results

High-resolution Azersky datasets were classified into "Forest," and "Non-Forest" classes with NDVI-aided supervised classification and Random Forest classification. The classified image datasets are represented in figure 3.

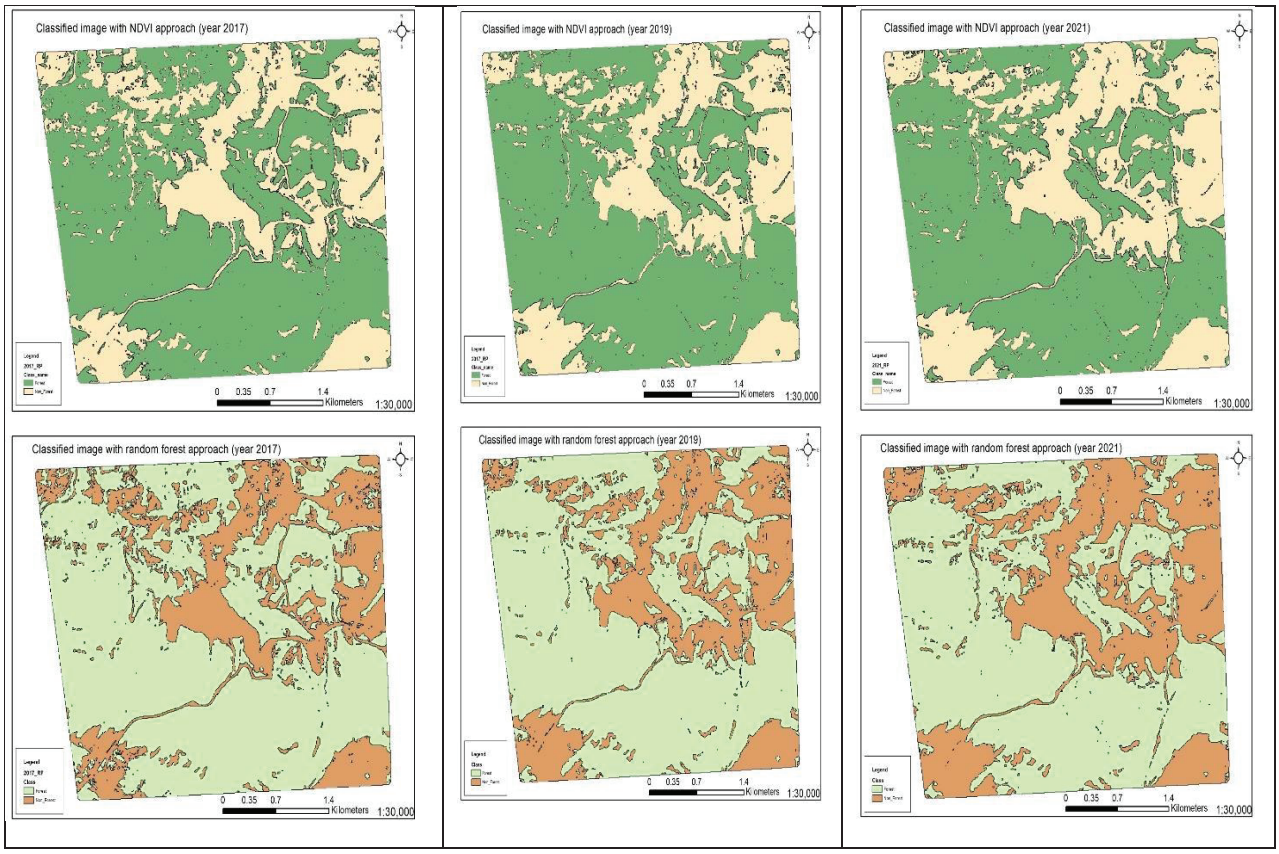
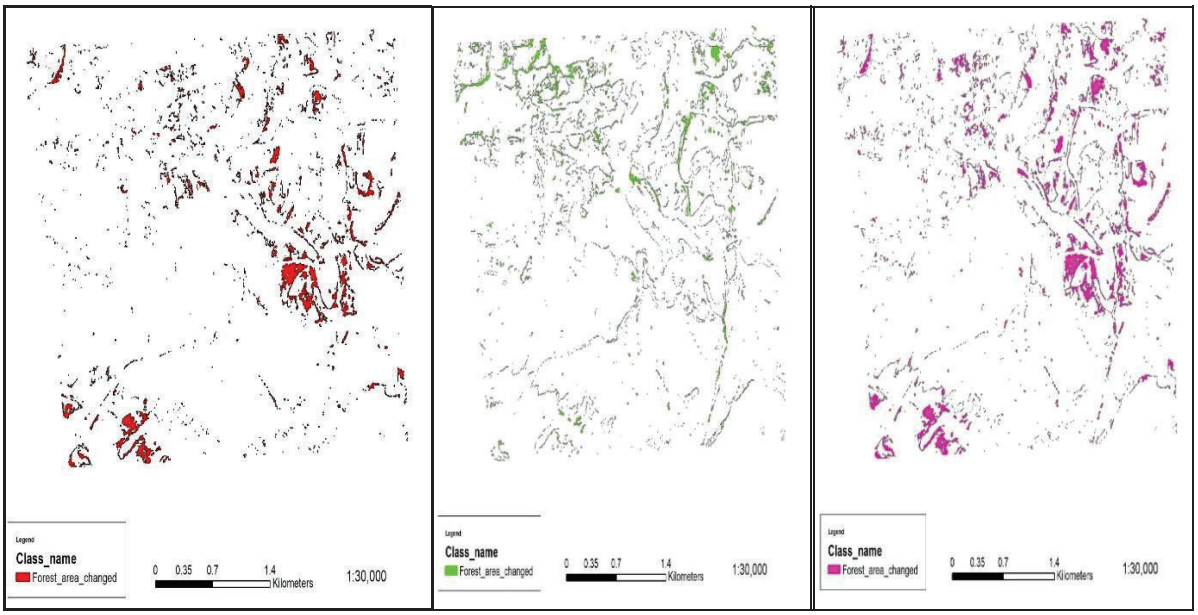


Figure 3: Classified datasets with NDVI-based supervised interactive classification (above) and random forest classification (below)

The change in forest area was calculated for all the classified datasets, and net change was analyzed from the year 2017 to the year 2021. The forest area was calculated from the attribute tables of the classified datasets, and deforestation was evident from the results of both classification approaches. Some misclassified pixels in both approaches might contribute to other classes. The forest area results from table 1 indicate the deforestation in the selected area at a rate of 3.67% and 9.50% with interactive supervised and random forest classification approaches, respectively.

Table 1: Forest area calculated from classification with Random Forest and NDVI-based interactive supervised approach

Year	Random Forest classification Approach	NDVI-based interactive supervised classification approach
	Forest area (m^2)	Forest area (m^2)
2017	7940616	7291340
2019	7732537	7206421
2021	7185180	6941613
Net change	**9.50%**	**4.79%**



a) b) c)

Figure 4: Change in area maps a) years 2017-19 b) years 2019-21 c) years 2017-21 with random forest classification approach

Deforestation was observed in the selected region from the statistical analysis of the classified datasets throughout the years 2017 to 2021. Overall decrease of 349,727 m^2 forest area was observed with the analysis of classification results from NDVI-aided interactive supervised classification approach, whereas with Random Forest approach, a decrease of 755,436 m^2 was observed. The net change is different for both approaches because of different classification algorithms and accuracy. The visual interpretation from the change maps in figure 4 also supports the deforestation from 2017 to 2021 in the selected region.

5. Conclusion

The statistical analysis of the results indicates the decrease in the forest area from the year 2017 to the year 2021, which is supported by the visual interpretation through the change map of the forest area. From random forest classification results, it has been observed that there has been a reduction in forest area by 2.7% between years 2017-19 followed by a further fall by 7.07% between years 2019-21 with an overall decrease of 9.5% from years 2017 to the year 2021. The interactive supervised classification approach results also indicate deforestation with a reduction in forest area by 1.17% between years 2017-19 followed by further decrease in forest area by 3.81% during years 2019-21 and a net decrease of 4.79% during the years 2017-21. The results from random forest classification can be accepted for this work as it is based on many self-learning decision trees and is more robust than a single decision tree.

Further, the datasets can be classified into more feature classes like bare land, road, built up, etc., to analyze the causes of deforestation and sustainable planning and development in the Tartar district in Azerbaijan. Also, the Unmanned Aerial Vehicles (UAVs) can be employed for the forest monitoring and to analyze the deforestation situation in real-time.

6. References

[1] Dlamini W M 2016 Analysis of deforestation patterns and drivers in Swaziland using efficient Bayesian multivariate classifiers *Model. Earth Syst. Environ.* **2** 1–14

[2] Singh A and Kushwaha S K P 2021 Forest Degradation Assessment Using UAV Optical

Photogrammetry and SAR Data *J. Indian Soc. Remote Sens.* **49** 559–67
[3] Ervin J 2003 Protected Area Assessments in Perspective *Bioscience* **53** 819–22
[4] Velasco Gomez M D, Beuchle R, Shimabukuro Y, Grecchi R, Simonetti D, Eva H D and Achard F 2015 A long-term perspective on deforestation rates in the Brazilian Amazon *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XL-7/W3** 539–44
[5] Singh A, Kushwah S K P and Kumar S 2020 Backscatter and coherence analysis using space borne C-band data for forest characterization *J. Geomatics* **14** 39–48
[6] Phua M H, Tsuyuki S, Furuya N and Lee J S 2008 Detecting deforestation with a spectral change detection approach using multitemporal Landsat data: A case study of Kinabalu Park, Sabah, Malaysia *J. Environ. Manage.* **88** 784–95
[7] Lu D, Mausel P, Brondízio E and Moran E 2004 Change detection techniques *Int. J. Remote Sens.* **25** 2365–401
[8] Zhang S, Zhang S and Zhang J 2000 A study on wetland classification model of remote sensing in the Sangjiang plain *Chinese Geogr. Sci.* **10** 68–73
[9] Kalaba F K, Quinn C H, Dougill A J and Vinya R 2013 Floristic composition, species diversity and carbon storage in charcoal and agriculture fallows and management implications in Miombo woodlands of Zambia *For. Ecol. Manage.* **304** 99–109
[10] Yang R, Luo Y, Yang K, Hong L and Zhou X 2019 Analysis of Forest Deforestation and its Driving Factors in Myanmar from 1988 to 2017 *Sustainability* **11**
[11] Fawagreh K, Gaber M M and Elyan E 2014 Random forests: from early developments to recent advancements *Syst. Sci. \& Control Eng.* **2** 602–9