DETECTING LAND COVER CHANGES IN SITES HOSTING PROTECTED CULTURAL HERITAGE: A CASE EXAMPLE OF UNESCO PROTECTED SITE IN NALANDA, INDIA

Esha Semwal, Saumyata Srivastava, Chandan Grover, Hina Pande, Poonam S. Tiwari, S. Agrawal, S. Raghavendra

Indian Institute of Remote Sensing, Indian Space Research Organization, Dehradun
Email: esha.semwal1@gmail.com, saumyata.srivastava@gmail.com, chandangrover32@gmail.com, hina@iirs.gov.in, poonam@iirs.gov.in, shefali_a@iirs.gov.in

Abstract: India is home to several important archaeological sites of the world. The increasing pressure of population and anthropogenic factors have made it prone to rapid change due to human activities. This study discusses the changes in land use and land cover for the past decades in the area altering the rich cultural heritage in that region. The land cover and land use changes in terms of deforestation, urban growth and increase in agricultural activity have been evaluated in the present research. Some of these parameters have been quantified using remote sensing and GIS data by analyzing time series of images from 2009 to 2019. The area selected for study is Nalanda Mahavihara, a cultural heritage site that comes under Archeological Survey of India and UNESCO world heritage site.

A knowledge driven classification based on Support Vector Mechanics (SVM) and Classification and Regression Trees (CART) algorithms along with spectral indices was used to detect the current and historical changes of cultural heritage site. Landsat and Sentinel images of the years 2009, 2014 and 2019 have been processed to detect the current and historical changes of cultural heritage site and the area surrounding it. The classification scheme includes the following eight classes: (1) Built-up, (2) Water Bodies, (3) Agricultural Area, (4) Natural Vegetation, (5) Grassland, (6) Wetland, (7) Ponds and (8) Open Land, Roads, for the evaluation of the land cover changes over the period of time. From the past decades the human activity in form of urbanization and farming has increased significantly around the heritage site. The accuracy evaluation was carried out with the ground truth data from the present day survey and field work. The change analysis suggests the gradual and steady destruction of natural and cultural wealth of this area leading to complete fragility. The increased agricultural activity has led to the exposure of bricks of the ancient structures with in the archaeological site. Also, in some areas the mounds which could have been probable sites for further excavation have been covered by building activities. These human induced activities are destroying unique cultural heritage sites in this region.

KEYWORDS: Rule based classification, cultural heritage, time series analysis, change detection spectral indices.

1. INTRODUCTION

Land-use and land-cover (LULC) change is a prime focus area for the universal sector due to its serious influence on biological diversities, water resources, and climate variations(Vitousek 1994)(Houghton et al. 2012). LULC can be driven by multiple levels of collective factors like biophysical and social considerations. The former is concerned with the surface water bodies, underground water, geology, and meteorological conditions, and the later regards to changing aspects of the public, revenue, prosperity, technology, and political schemes (Agarwal et al. 2002)(Briassoulis n.d.). Both the factors fluctuate with the topography of the land by time. Thus, using multi-temporal dataset for analysing the change in land cover would be helpful to understand the social impact on the natural environment (Roy et al. 2015). Classifying the land cover on the basis of ground surveys would have been tedious as well as expensive because of the divergence in land use pattern at approximate every tens of meter distance. But with the ease in accessibility of free remotely sensed datasets, the mapping of land cover could be done much more easily and cost-effectively.

It is well-known that the selection of a suitable classifier as well as appropriate satellite bands is essential for improved classification accuracies(Lu and Weng 2007). Along with that, the various parameters that decide the threshold value for classification also have a prominent effect on the final outcome of the classification. In the same way, the inclusion of the NDVI which provides an index of vegetation, has been used for vegetation studies especially assessing the health of vegetation (Morawitz et al. 2006),with higher NDVI values indicating good healthy vegetation while lower NDVI values show deprived vegetation. Although the data mining approach when used along with the traditional digital image classification, provides the sense of environmental changes in the study area (Otukei and Blaschke 2010). (Archaeological Survey of India 2016) mention the overview of the land cover area in the region and its surrounding classes. It provides the idea of spectral signatures for each LULC classes. This study also helps in preserving a heritage site that was in existence from the fourth/fifth century (Rajani 2016).

To identify the land use changes for the past decades in the study area, LULC classification was performed using multi-temporal remotely sensed images of Sentinel and Landsat. The analysis work has been done in Google Earth Engine (GEE). GEE is developed by Google as a free cloud computing platform that is used for big Earth observation data management and analysis. GEE is easily available to the research community for LULC change classification, image processing, town planning, and weather analysis (Amani et al. 2020) by making use of various satellite datasets such as Landsat, Sentinel, and Modis. Suitable classification algorithms are required to derive valid information from satellite data (Lu and Weng 2007). In past few years various classification methods have been established such as Maximum Likelihood Classification (MLC), K-Means, , Minimum Distance to Means,(Anon n.d.) Classification and Regression Trees (CART), Random Forest (RF), Support Vector Machines (SVM) (Breiman et al. 1984)(Mathur and Foody 2008) .In the current study, SVM and CART techniques were applied for classification. SVM and CART are both supervised and non-parametric classification models. SVM was proposed by Vapnik and Chervonenkis in 1971. The SVM algorithm finds a hyper-plane that distinctly separates the data points of different classes(Mathur and Foody 2008). CART was first introduced by Leo Breiman, Jerome Friedman, Richard Olshen and Charles Stone in 1984. CART is the modified version of Decision Trees model which works on both classification and regression trees procedures (Breiman et al. 1984). The main objectives of this study can be summarised in two points: 1. Identifying the change in land cover in the region which could have been potential sites for further excavation from the LULC map developed using long-term time-series satellite images. 2. A comparative study of different classification algorithms (SVM and CART) for high and medium resolution satellite imagery and their accuracy analysis.

2. DESCRIPTION OF STUDY AREA

The study area is a UNESCO World heritage site, Nalanda Mahavihara. It is located in the state of Bihar, in north-eastern India. The geographical coordinates of the study area are: 25^{o} 08’7.36” N and 85^{o}26’32.68” E with maximum average elevation of 68m above sea level, shown in fig. 1. The heritage site is spread over an area of 23 hectares with a buffer zone of 57.88 hectares (Anon n.d.). It is an ancient revered Buddhist monastery that also served as a well-known centre of learning before it was destroyed and transformed into ruins by Bakhtiyar Khalji in 1200 CE. The site remained unattended by the research community until the 19th century. However, in the 20th century, the site was surveyed and preliminary excavations were conducted by the Archaeological Survey of India (ASI). The ruins founded until now includes eleven monasteries, six major brick temples, stupas, chaityas, viharas, and shrines. Archaeologist stated that only 10% of the ruins have been excavated so far and the actual mahavihara possibly extended upto much larger area. The area around the heritage site is occupied with agricultural land, seasonal and non-seasonal water bodies, roads and sparsely populated small settlements. The uneven topography of Nalanda proves to be unfavourable for agriculture, yet the majority of population is involved in farming (Anon n.d.).The major crops in this area are rice, wheat, maize, pulses, potato, fruits and vegetables. Nalanda has hot climate in summers and cool in winter with an average annual rainfall of 120cm. The major rivers flowing through Nalanda are Phalgu, Mohane, Jirayan, and Kumbhari (Anon n.d.).

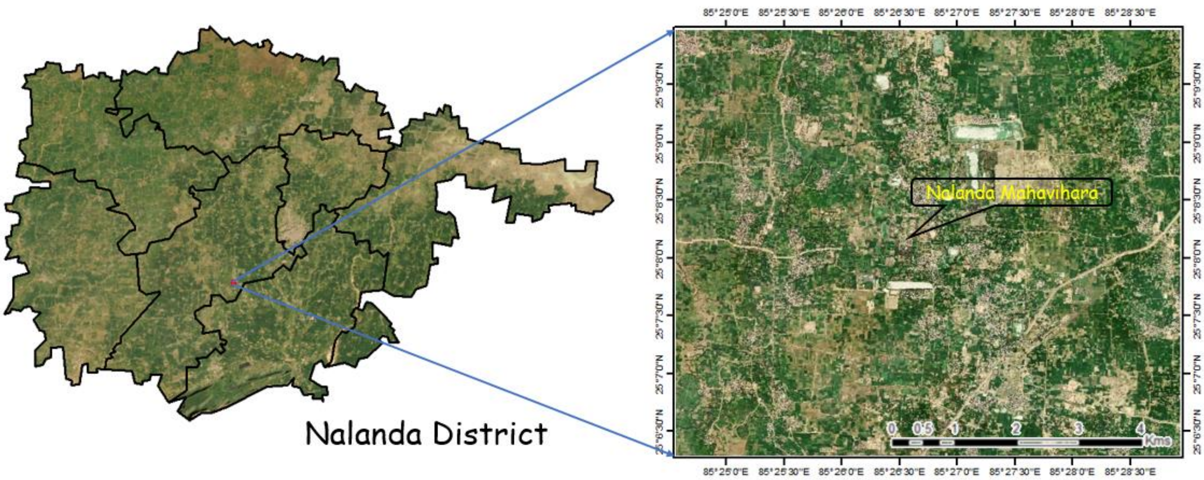

Figure 1 Location of Study Area

3. DATA USED

The description of multi-temporal satellite data used for current work is given in table below

Table 1 Dataset used

Satellite	Spatial Resolution	Spectral Resolution	Date of Pass
Sentinel 2	10m	Band 2: 0.496 - 0.492 µm Band 3: 0.560 - 0.559 µm Band 4: 0.664 - 0.665 µm Band 8: 0.835 - 0.833 µm	3-Jan-2019
Landsat 8	30m	Band 2: 0.450 - 0.515 µm Band 3: 0.525 - 0.600 µm Band 4: 0.630 - 0.680 µm Band 5: 0.845 - 0.885 µm	6-Feb-2014
Landsat 5	30m	Band 1: 0.45 - 0.52 µm Band 2: 0.52 - 0.60 µm Band 3: 0.63 - 0.69 µm Band 4: 0.76 - 0.90 µm	16-Jan-2009

The scenes from Landsat 7 had data gaps caused by Scan Line Corrector (SLC) failure(Hossain et al. 2015). So we have used Landsat 5 data instead of Landsat 7 for 2009.

4. METHODOLOGY

The first step of the classification method was to develop the classification algorithm based on our objective. Fig.2 shows the workflow adopted for this study. The land cover classification method adopted for this paper consist of 8 classes named built-up area, water bodies, agricultural land, natural vegetation, grassland, wetlands, pond and rest are included in open land/roads class. Next step was to decide the suitable image for time series classification. As Nalanda completely lies in the Sub Tropical region of Temperate zone and its climate type is Humid Sub Tropical, due to which post- monsoon and winter season lasts from end of November till February of next year. Therefore in order to extract maximum information from Water bodies (like wetlands, ponds, rivers) and in order to catch natural vegetation at its maximum growth we chose the data with minimum cloud cover in between January and March. It would avoid the mixing of pixels of dried vegetation and agricultural land just after sowing. After the determination of the appropriate bands, a classification was performed using CART and SVM. Same number and order of bands were used in classification with the purpose of removing the biasness caused by different bands. The classification was carried out on Google Earth Engine.

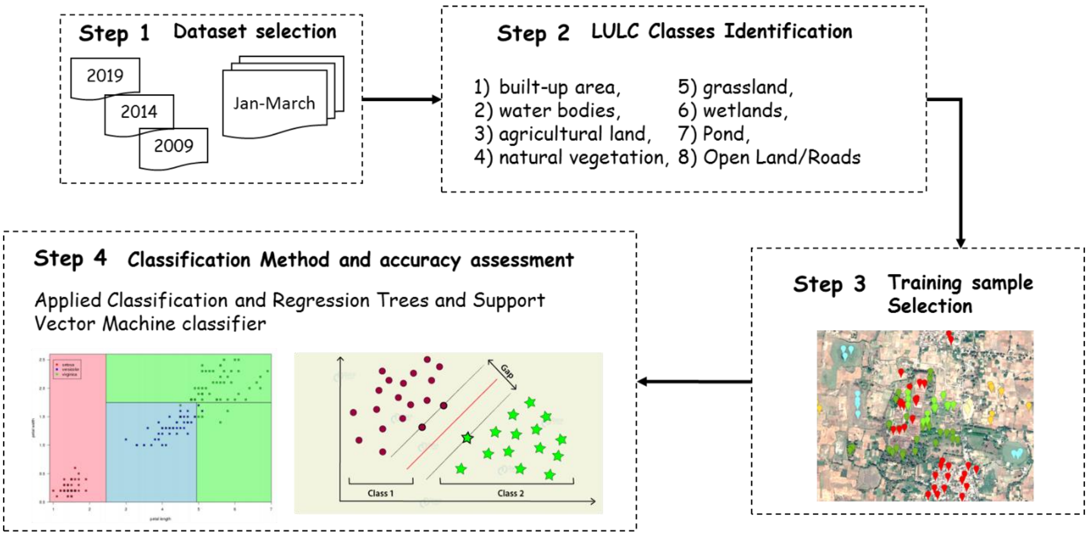

Figure 2 Workflow for LULC methodology

Next step was to delineate training data based on analyst's prior knowledge of the study area. Further the description of topology of study area was illustrated in the literature by ASI. (Archaeological Survey of India 2016). To improve the dataset of training and validation data, the NDVI parameter is used to demarcate the land based on threshold value, later on, vegetation class is further sub-classed into natural vegetation, agricultural land and grassland classes (Table 2). Further NDWI is used to demarcate the water bodies that was again sub-classed into ponds, wetlands and other water bodies.

Vegetation Index Name	Landsat 4-5 Images	Landsat 8 Images
NDVI	$\frac{Band\ 4\ (NIR) - Band\ 3 (Red)}{Band\ 4\ (NIR) + Band\ 3 (Red)}$	$\frac{Band\ 5\ (NIR) - Band\ 4\ (Red)}{Band\ 5\ (NIR) + Band\ 4\ (Red)}$
NDWI	$\frac{Band\ 2\ (Green) - Band\ 4 (NIR)}{Band\ 2\ (Green) + Band\ 4 (NIR)}$	$\frac{Band\ 3\ (Green) - Band\ 5\ (NIR)}{Band\ 3\ (Green) + Band\ 5\ (NIR)}$

Table 2 Representing vegetation index formula

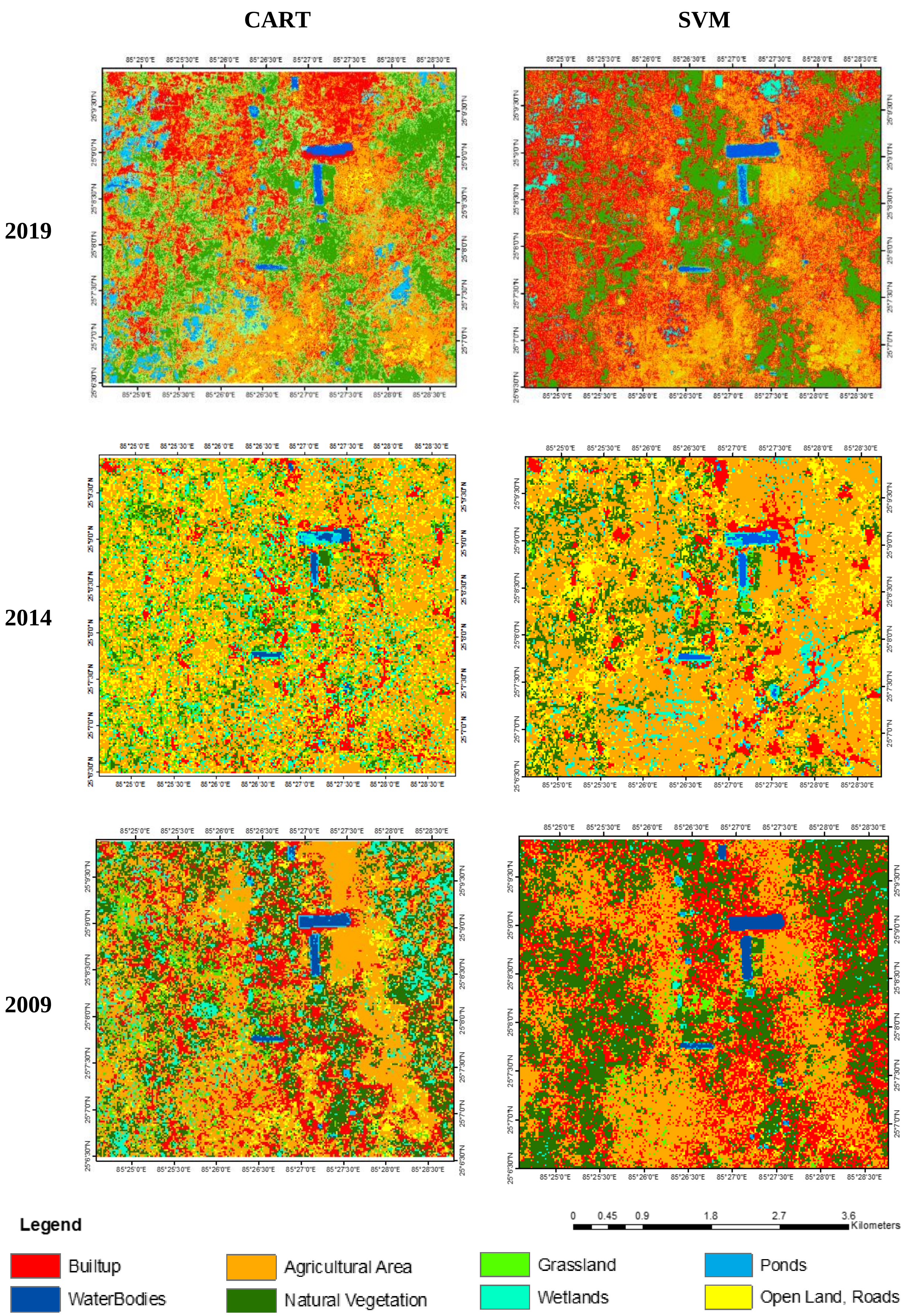

Figure 3 Result representing time series analysis in CART and SVM classifier for 3 years

5. RESULTS AND DISCUSSION

The error matrix and overall classification accuracy was evaluated using the confusion matrix. A separate but similar data set was used as a validation dataset for accuracy assessment in all cases.

Table 3 Overall accuracy from CART and SVM classifier

Year		**2019**		**2014**		**2009**
Method	CART	SVM	CART	SVM	CART	SVM
Overall accuracy	0.82	0.83	0.79	0.73	0.87	0.81

The classification accuracy was increased from CART to SVM when the classifier was applied on the high resolution data i.e. Sentinel 2A for 2019 year. It was also observed, as represented in Fig. 4 that the misclassified pixel of agricultural land class was correctly classified in SVM classifier. It could be interpreted from the classification error matrix that CART classifier puzzles between pixels of built-up area, agricultural land and open land in Sentinel 2A data. Around 14% of built-up area pixels are misclassified in agricultural land using CART classifier while most of them are correctly recognized in agricultural land area using SVM classifier.

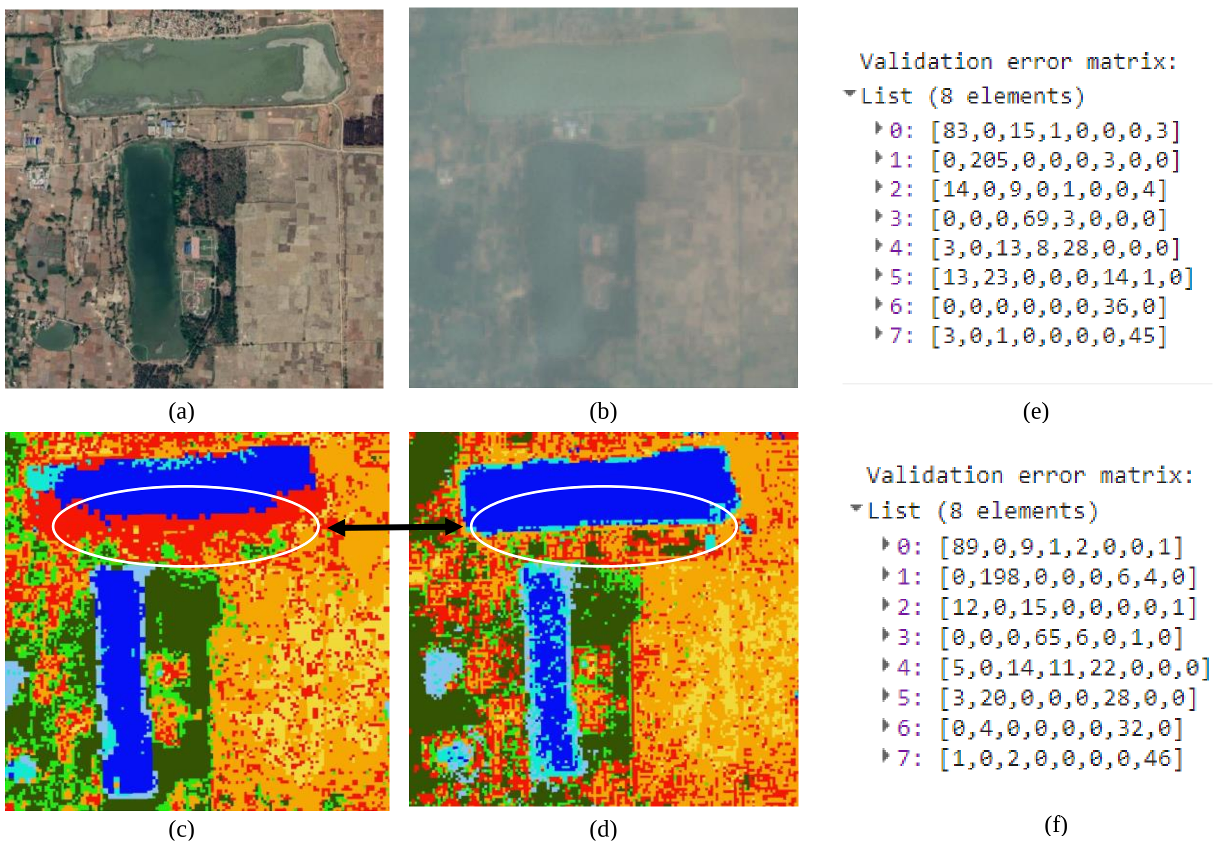

Figure 4 (a) Google Earth Image of the area; (b) Sentinel 2A imagery; (c) CART classified image; (d) SVM classified Image; (e) Confusion matrix of CART classified image; (f) Confusion matrix of SVM classified image (ref: 0- built-up area, 1- water bodies, 2- agricultural land, 3- natural vegetation, 4- grassland, 5- wetlands, 6- pond and 7- open land/roads class)

While in Landsat 5 of 2009, it was observed that CART classifier gives comparatively good result as compared to SVM. As interpreted from Fig. 5 open land was misclassified in SVM classifier as a built-up area. But overall high accuracies were obtained for both of the techniques.

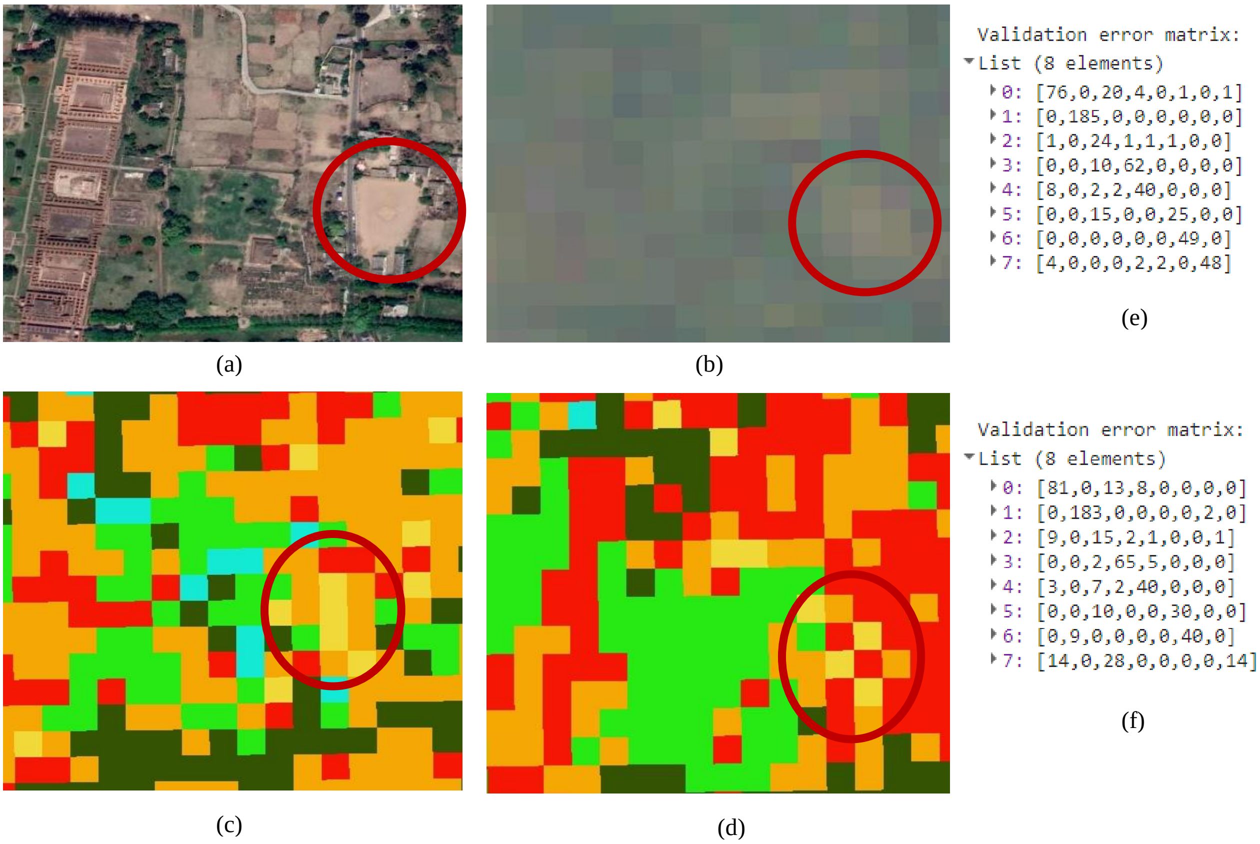

Figure 5 (a) Google Earth Image of the area; (b) Landsat 5 imagery; (c) CART classified image; (d) SVM classified Image; (e) Confusion matrix of CART classified image; (f) Confusion matrix of SVM classified image (ref: 0- built-up area, 1- water bodies, 2- agricultural land, 3- natural vegetation, 4- grassland, 5- wetlands, 6- pond and 7- open land/roads class)

It was also observed that most of the agricultural land in the north and west of the heritage site has been converted into built-up area since the past decades. While natural vegetation have been chopped off to build habitation for the people. This increase in habitation would adversely affect the health of preserved heritage site, unless it is not kept in check.

7. REFERENCES

- Agarwal, Chetan, Glen M. Green, J. Morgan Grove, Tom P. Evans, and Charles M. Schweik. 2002. *A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice*. Vol. 297.
- Amani, Meisam, Arsalan Ghorbanian, Seyed Ali Ahmadi, Mohammad Kakooei, Armin Moghimi, S. Mohammad Mirmazloumi, Sayyed Hamed Alizadeh Moghaddam,

Sahel Mahdavi, Masoud Ghahremanloo, Saeid Parsian, Qiusheng Wu, and Brian Brisco. 2020. “Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13(September):5326–50.

- Anon. n.d. “Archaeological Site of Nalanda Mahavihara at Nalanda, Bihar.” Retrieved October 12, 2020a (https://whc.unesco.org/en/list/1502/).
- Anon. n.d. “Geography of Nalanda, Topography and Climate of Nalanda.” Retrieved October 16, 2020b (https://www.nalandaonline.in/city-guide/geography-of-nalanda).
- Anon. n.d. “Remote Sensing and Image Interpretation - Thomas Lillesand, Ralph W. Kiefer, Jonathan Chipman - Google Books.” Retrieved October 16, 2020c (https://books.google.co.in/books?hl=en&lr=&id=AFHDCAAAQBAJ&oi=fnd&pg=PA91&dq=Lillesand,+T.M.,+Kiefer,+R.W.,+1999.+Remote+Sensing+and+Image+Interpretation.+John+Wiley+and+Sons+Ltd.,+New+York.&ots=0Dnml0bDar&sig=G6Exo61Fd3cu17JlIq5MgM1WEqQ#v=onepage&q&f=false).
- Archaeological Survey of India. 2016. “Nomination Dossier: Property-Excavated Remains of Nalanda Mahavihara.”
- Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Wadsworth & Brooks/Cole Advanced Books & Software.
- Briassoulis, Helen. n.d. “M ES PL C E O – PL C E O –.” I.
- Hossain, Mohammad Shawkat, Japar Sidik Bujang, Muta Harah Zakaria, and Mazlan Hashim. 2015. “Assessment of the Impact of Landsat 7 Scan Line Corrector Data Gaps on Sungai Pulai Estuary Seagrass Mapping.” *Applied Geomatics* 7(3):189–202.
- Houghton, R. A., J. I. House, J. Pongratz, G. R. van der Werf, R. S. DeFries, M. C. Hansen, C. Le Quéré, and N. Ramankutty. 2012. “Carbon Emissions from Land Use and Land-Cover Change.” *Biogeosciences* 9(12):5125–42.
- Lu, D. and Q. Weng. 2007. “A Survey of Image Classification Methods and Techniques for Improving Classification Performance.” *International Journal of Remote Sensing* 28(5):823–70.
- Mathur, A. and G. M. Foody. 2008. “Multiclass and Binary SVM Classification: Implications for Training and Classification Users.” *IEEE Geoscience and Remote Sensing Letters* 5(2):241–45.
- Morawitz, Dana F., Tina M. Blewett, Alex Cohen, and Marina Alberti. 2006. “Using NDVI to Assess Vegetative Land Cover Change in Central Puget Sound.” *Environmental Monitoring and Assessment* 114(1–3):85–106.
- Otukei, J. R. and T. Blaschke. 2010. “Land Cover Change Assessment Using Decision Trees, Support Vector Machines and Maximum Likelihood Classification Algorithms.” *International Journal of Applied Earth Observation and Geoinformation*

12(SUPPL. 1):27–31.

- Rajani, M. B. 2016. “The Expanse of Archaeological Remains at Nalanda: A Study Using Remote Sensing and GIS.” *Archives of Asian Art* 66(1):1–23.
- Roy, Parth S., Arijit Roy, Pawan K. Joshi, Manish P. Kale, Vijay K. Srivastava, Sushil K. Srivastava, Ravi S. Dwevidi, Chitiz Joshi, Mukunda D. Behera, Prasanth Meiyappan, Yeshu Sharma, Atul K. Jain, Jamuna S. Singh, Yajnaseni Palchowdhuri, Reshma M. Ramachandran, Bhavani Pinjarla, V. Chakravarthi, Nani Babu, Mahalakshmi S. Gowsalya, Praveen Thiruvengadam, Mrinalni Kotteeswaran, Vishnu Priya, Krishna Murthy V. N. Yelishetty, Sandeep Maithani, Gautam Talukdar, Indranil Mondal, Krishnan S. Rajan, Prasad S. Narendra, Sushmita Biswal, Anusheema Chakraborty, Hitendra Padalia, Manoj Chavan, Satish N. Pardeshi, Swapnil A. Chaudhari, Arur Anand, Anjana Vyas, Mruthyunjaya K. Reddy, M. Ramalingam, R. Manonmani, Pritiranjan Behera, Pulakesh Das, Poonam Tripathi, Shafique Matin, Mohammed L. Khan, Om P. Tripathi, Jyotihman Deka, Prasanna Kumar, and Deepak Kushwaha. 2015. “Development of Decadal (1985-1995-2005) Land Use and Land Cover Database for India.” *Remote Sensing* 7(3):2401–30.
- Vitousek, Peter. 1994. “Vitousekgw.Pdf.” *Ecology* 75(7):1861–76.