Study of Land Use and Land Cover Change Detection Using Machine Learning on GEE of Chandigarh, India

Amandeep Kaur[1, a], *Gurwinder Singh[2, b], Amit Jain[3, c], Beena Kapadia[4, d]

[1]Uiversity Institute of Computing, Chandigarh University, Gharuan, India

[2]Department of Computer Application, Chandigarh School of Business, Division of Research & Innovation, Chandigarh Group of Colleges, Jhanjeri, India

[3]Department of Computer Application, Chandigarh School of Business, Jhanjeri, India

[4]KES' Shroff College, Kandivali West, Mumbai, India

[a]uic.amandeepkaur@gmail.com, [b]gurwindersingh.balahara@gmail.com, [c]amit_jainci@yahoo.com, beena.v.kapaida@gmail.com[d]

Abstract: Satellite imagery has proven its skills in the field of evaluating and supervising land use and land cover (LULC) for better eco-friendly management. High-resolution and high-quality datasets can improve LULC classification when implemented with various Machine Learning (ML) and Deep Learning (DL) models. DL requires high-end computation facilities but can give better accuracy results, whereas ML algorithms require learning from humans to make decisions, but its implementation is easy. In this study, the Sentinel-2 satellite imagery dataset is used to study the land statistics of Union Territory, i.e., Chandigarh. The major classes of this study are urban, water bodies, forests, and bare land. In this paper, LULC classification is analyzed using Random Forest (RF). RF model gave an overall accuracy of 96.6%. All the results have proved that RF is delivering the best accuracy among all the other models. This research has a broad spectrum of applications, such as monitoring and mapping land use and land cover areas using ML algorithms.

Keywords: Sentinel-2, Machine Learning, Random Forest, Support Vector Machine, Land use, Land cover, Remote Sensing

1. Introduction:

Land is a fundamental and vital resource of the world and acts as a foundation of human survival and development[1]. As life is growing at a very fast pace, a lot of factors significantly impact the way land is being used. It is apparent that due to urbanization, there is a considerable downfall in agricultural activities, yield production, weather disturbances, pollution, etc. As the population grows, it is becoming very crucial to study the land statistics assessment in order to maintain the supply chain[2]. The notable transition in land use land cover (LULC)can be monitored with remote sensing. Landsat-8, Landsat-9, Sentinel-2, and MODIS are the categories of remote sensing approach, i.e., Optical Remote Sensing dataset. LULC is a combination of words used as substitutes, although land use refers to the way land is purposely used, such as agriculture, residential, wildlife, etc. In contrast, land cover is what is actually found on the land, such as built-up areas, wetlands, vegetation areas, grasslands, water bodies, and fallow land [3].

Analyzing and detecting land change in LULC analysis can be helpful for improved organization and better land[4]management. In the past, various techniques have evolved to assess remote sensing data. However, ML and DL have proved dominant in significantly generating accurate observations of change in LULC.

N. Vyas et al. (eds.), *Proceedings of the International Conference on Advancements in Computing Technologies and Artificial Intelligence (COMPUTATIA 2025)*, Advances in Intelligent Systems Research 189,
https://doi.org/10.2991/978-94-6463-700-7_4

Widely accepted Supervised ML proved to be the best classification technique for extracting LULC features; however, unsupervised ML still needs improvement, whereas supervised DL approaches, C-NN (Convolution Neural Network) and R-CNN (Regions with Convolution Neural Network), submit evidence of giving accurate results as compared to Supervised ML approaches [15]. In this study, Supervised ML approaches such as RF are used to access the LULC statistics.

In the literature, various remote sensing datasets have been implemented, including the Sentinel dataset and Landsat dataset, for land cover mapping of land from different parts of the world [5] and ML models, as shown in Table 1. Moreover, it was investigated that the DLCNN algorithm works best for a time series of images that can easily classify the level of water inside the croplands[6]. RF, i.e., the supervised ML approach, plays best in mapping the land cover of maize cropland in Nigeria via pixel-based image classification[2]. Plants reflect color due to chlorophyll, which is used to distinguish healthy crops from stressed crops. For this RF, SVM models of ML are employed[7]. Normalized Difference Vegetation Index (NDVI) was used to collect the plant's leaf properties. Normalized Difference Water Index (NDWI) is used to identify and supervise the water bodies and the soil's water content. In this study, Chandigarh has been selected as ROI (Region of Interest) from date January 01, 2018, to detect the different classes as urban, water area, forest area, and bare land. In this paper, the Sentinel-2 dataset is used to determine the change using ML classification algorithms.

Table 1. Schema of collection of papers includes Number of classes, Accuracy, Dataset, classification accuracy

Author	Classes	Accuracy	Satellite Dataset	Approach
[8]	1	97%	Sentinel-1/Sentinel-2	RF, SVM, CART
[7]	7	93.50%	Landsat 8	ML,SVM, RF
[9]	6	97.8%, 96.2%	Sentinel-2	U-Net, RF
[10]	4	88%	Landsat 8	RF
[6]	19	91%	Sentinel-2	CNN
[4]	4	80%	Landsat	MLA
[5]	15	90.13%, 90.18%	Sentinel-2	CNN, R-CNN
[11]	5	87% Landsat-8, 92%Sentinel-2	Landsat-8/sentinel-2 (GEE)	SVM, RF, ML, CART

Note: RF: Random Forest; SVM: Support Vector Machine; CART: Classification and Regression Tree; ML: Machine Learning, U-Net: U-shaped convolutional neural network; CNN: Convolutional Neural Network; MLA: Maximum Likelihood Algorithm; R-CNN: Region-based Convolutional Neural Network; GEE:

In the present study, LULC change has been implemented on Sentinel-2 images on land. Supervised models of ML are used to conduct this study. The primary aims of this study are: a (a) implementation of ML Models for understanding the LULC on Region of Interest, i.e., area of Chandigarh; (b) Major land use change detection of areas, i.e., urban, water, forest, and bare land; (c) analysis of the accuracy of models used. This study can help better manage land, i.e., natural resources. This study area is selected due to the increase in urbanization and drastic changes in temperature.

2. **Study Area and Satellite Dataset:**

This section mainly focused on the Region of Interest (ROI) area selected for conduct in this study from Latitude and Longitude 30°45′22″ N to 76°42′18″E and 30°43′27″N to 76°50′56″E, i.e., Chandigarh which is a union territory (India). These geographical coordinates basically acquires114 km² as shown in Figure 1. Chandigarh has served as the capital of the states of India, i.e., Punjab and Haryana, since 1966. Chandigarh is known for its state-of-the-art architecture, urban design, and natural beauty, which was designed by architect Le Corbusier in the 1950s. Initially, Chandigarh was planned for a 5-lakh population, but as per the census of 2024, the population was increased to 1,240,000, i.e., an increase of 148%. This study includes major class area categories such as urban, water, forest, and bare land. The ROI imagery was acquired on a winter day with minimal clouds, i.e., 01-Jan-2018 and 01-Jan-2023 from(https://earthexplorer.usgs.gov/) and processed using QGIS. Mostly during this season, the atmosphere is clear and cloud-free. The spatial resolution of Sentinel-2 is between 61 pixels to 76 pixels with 13 spectral bands.

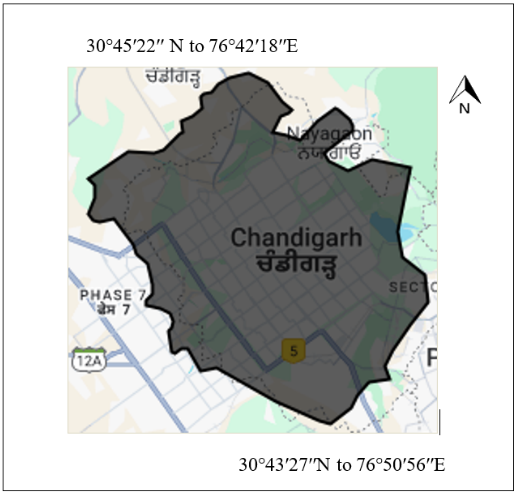

Figure 1. Region of Study: Chandigarh map (India)

3. Methodology

The proposed methodology is working as per the flowchart shown in Figure 2. It involves the three basic steps: (a) Extraction of ROI map and Pre-processing of Sentinel-2 data acquired for ROI, including cloud masking, radiometric calibration, and atmospheric correction; (b) Implementation of RF of ML; 80% of data is training data:

3.1 ROI Extraction and Preprocessing of the Input Data

QGIS was used to extract the ROI from a large shapefile. Shapefile was submitted to QGIS, from which ROI was selected from the attribute table, and the selected region was clipped and converted to shapefile, which was then used for land change detection using GEE. To preprocess, harmonized sentinel-2image data of 2018 and 2023 with surface reflectance was filtered and selected where cloud visibility was less than 1.

Images were processed using the cloud mask function in GEE, from which 504 training data was prepared. Training data was classified using ML-supervised models for both years. At last, change detection was analyzed using classified results for the years 2018 and 2023. In this study, the Google Earth Engine (GEE) tool was used to preprocess satellite images and obtain quality images. After obtaining the images, classes and bands were determined. After the creation of training data, the classifier is trained and applied to generate a visualization of selected satellite imagery.

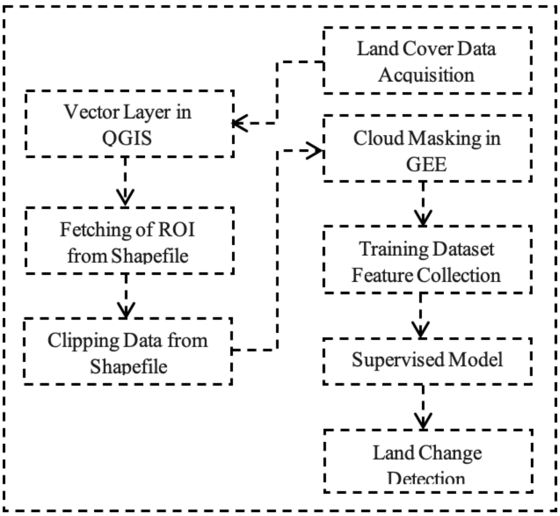

Figure 2. Workflow chart from data acquisition through QGIS to classification.

3.2 Input Selection

Within the scope of remote sensing, ML and DL have shown their potential to improve image classification[12]. ML and DL have the ability to process multi-dimensional data and better map and differentiate the different classes. It is crucial to select the accurate input variables to generate the optimal result and save computational resources[13]. ML and DL-based classifiers are not only limited to LULC but extend to other applications like construction and demolition waste[14], boreal landscape [15] , river systems, etc. Similar to ML, DL employs NN to extract features from the dataset and learn from that dataset[16]. Generally, it can be labeled as Supervised, Unsupervised, and Reinforcement Learning. RF is an ML approach that is based on a collection of decision trees to analyze the dataset and generate desired accurate results[8]. Whereas SVM primarily focuses on a set of support vectors or training data that are at the optimal closeness to boundaries between two classes [17]. GEE can classify input data based on Pixel-Based Classification and Object-Based Classification [9]. In total, 504 training samples were employed for the classification process. For each and every class category, multiple training samples have been selected. After training the model, visual classified maps are generated, defining each category of class.

4. Result and analysis

Figure 3 (a-b) shows the classified maps of urban area, water area, forest area, and bare land, respectively. ML has been implemented to classify each class category [15] to generate the land cover areas. GEE-obtained thematic maps additionally show the land change in selected class categories. Table 2 shows the statistical result graphs. Recall rate, precision, loss and accuracy, and performance of ML and DL models. After classification, the result shows that due to urbanization and due to an increase in commercialization, an average of 2.66% of land converted to urban land from forest land is shown in Figure 4.

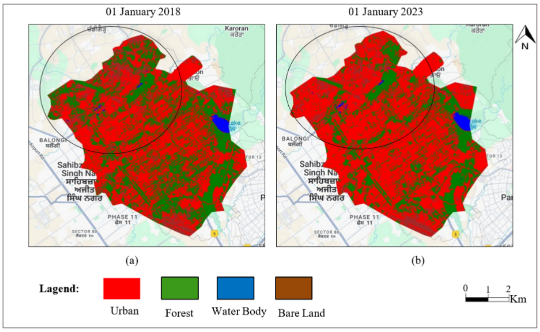

Figure 3 Comparative LULC analysis for 2018 and 2023, showcasing the classification results from the RF-ML Model

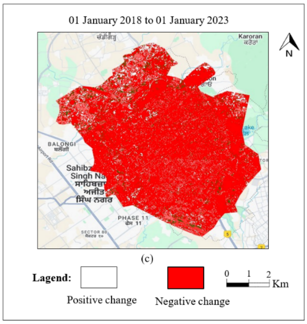

Figure 4 Total Land Change

Table 2. Parameters for Assessing Accuracy

Classifier	Classes	RT (%)	CT (%)	NC (%)	PA (%)	UA (%)	KC
RF	Urban	73.8	72.2	79.4	99.9	98.2	0.96
	Forest	10.4	12.6	9.8	98.6	96.1	0.98
	Water body	7.2	6.8	5.6	97	98.3	0.97
	Bare Land	8.6	8.4	5.3	100	97.4	0.95
OA = 96.6%; OK = 0.9871							

Note: RT: Reference total; CT (%): Classified total; NC (%): Number of correct; PA (%): Producer's accuracy; UA (%): User's Accuracy; KC: Kappa coefficient; OA: Overall accuracy; OK: Overall kappa.

Conclusion:

This paper uses GEE to analyze and evaluate the different class categories, i.e., urban areas, water bodies, forest areas, and bare land. This study was conducted using sentinel-2 satellite imagery. Different widely known ML Supervised models were implemented. After analyzing the results, it has been concluded that the ML model, i.e., performed with an accuracy of 96.6%. Although it fetched the land change features very well, many challenges still exist in analyzing the agricultural land and very small water bodies. Future research could improve the accuracy by implementing the improved algorithms for detecting and mapping cropland and non-cropland. Improved algorithms can be implemented to evaluate the percentage of land changed from agricultural land to urban area, bare land to urban area, and change of water area to urban area. Additionally, future studies can rely on exploring more ML and DL algorithms to evaluate the land change of the different class categories with high-dimensional datasets and high-resolution data.

References

[1] O. R. Abd El-Kawy, J. K. Rød, H. A. Ismail, and A. S. Suliman, "Land use and land cover change detection in the western Nile delta of Egypt using remote sensing data," *Appl. Geogr.*, vol. 31, no. 2, pp. 483–494, Apr. 2011, doi: 10.1016/j.apgeog.2010.10.012.

[2] G. A. Abubakar *et al.*, "Mapping Maize Cropland and Land Cover in Semi-Arid Region in Northern Nigeria Using Machine Learning and Google Earth Engine," *Remote Sens.*, vol. 15, no. 11, Jun. 2023, doi: 10.3390/rs15112835.

[3] A. I. Al-Taei, A. A. Alesheikh, and A. Darvishi Boloorani, "Land Use/Land Cover Change Analysis Using Multi-Temporal Remote Sensing Data: A Case Study of Tigris and Euphrates Rivers Basin," *Land*, vol. 12, no. 5, May 2023, doi: 10.3390/land12051101.

[4] A. F. Koko, Z. Han, Y. Wu, G. A. Abubakar, and M. Bello, "Spatiotemporal Land Use/Land Cover Mapping and Prediction Based on Hybrid Modeling Approach: A Case Study of Kano Metropolis, Nigeria (2020–2050)," *Remote Sens.*, vol. 14, no. 23, Dec. 2022, doi: 10.3390/rs14236083.

[5] E. Papadopoulou, G. Mallinis, S. Siachalou, N. Koutsias, A. C. Thanopoulos, and G. Tsaklidis, "Agricultural Land Cover Mapping through Two Deep Learning Models in the Framework of EU's CAP Activities Using Sentinel-2 Multitemporal Imagery," *Remote Sens.*, vol. 15, no. 19, Oct. 2023, doi: 10.3390/rs15194657.

[6] A. M. Simón Sánchez, J. González-Piqueras, L. de la Ossa, and A. Calera, "Convolutional Neural Networks for Agricultural Land Use Classification from Sentinel-2 Image Time Series," *Remote Sens.*, vol. 14, no. 21, Nov. 2022, doi: 10.3390/rs14215373.

[7] P. Dash, S. L. Sanders, P. Parajuli, and Y. Ouyang, "Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data in an Agricultural Watershed," *Remote Sens.*, vol. 15, no. 16, Aug. 2023, doi: 10.3390/rs15164020.

[8] P. A. Palanisamy, K. Jain, and S. Bonafoni, "Machine Learning Classifier Evaluation for Different Input Combinations: A Case Study with Landsat 9 and Sentinel-2 Data," *Remote Sens.*, vol. 15, no. 13, Jul. 2023, doi: 10.3390/rs15133241.

[9] G. Singh, S. Singh, G. Sethi, and V. Sood, "Deep Learning in the Mapping of Agricultural Land Use Using Sentinel-2 Satellite Data," *Geographies*, vol. 2, no. 4, pp. 691–700, Nov. 2022, doi: 10.3390/geographies2040042.

[10] S. E. Pizarro, N. G. Pricope, D. Vargas-Machuca, O. Huanca, and J. Ñaupari, "Mapping Land Cover Types for Highland Andean Ecosystems in Peru Using Google Earth Engine," *Remote Sens.*, vol. 14, no. 7, 2022, doi: 10.3390/rs14071562.

[11] S. Basheer *et al.*, "Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques," *Remote Sens.*, vol. 14, no. 19, Oct. 2022, doi: 10.3390/rs14194978.

[12] A. Shukla and K. Jain, "Automatic extraction of urban land information from unmanned aerial vehicle (UAV) data," *Earth Sci. Informatics*, vol. 13, no. 4, pp. 1225–1236, Dec. 2020, doi: 10.1007/s12145-020-00498-x.

[13] C. Deng and Z. Zhu, "Continuous subpixel monitoring of urban impervious surface using Landsat time series," *Remote Sens. Environ.*, vol. 238, p. 110929, Mar. 2020, doi: 10.1016/j.rse.2018.10.011.

[14] L. Zhou *et al.*, "Machine learning comparison and parameter setting methods for the detection of dump sites

for construction and demolition waste using the google earth engine," *Remote Sens.*, vol. 13, no. 4, pp. 1–17, 2021, doi: 10.3390/rs13040787.
[15] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIScience Remote Sens.*, vol. 57, no. 1, pp. 1–20, Jan. 2020, doi: 10.1080/15481603.2019.1650447.
[16] J. M. de Souza, P. Morgado, E. M. da Costa, and L. F. de N. Vianna, "Modeling of Land Use and Land Cover (LULC) Change Based on Artificial Neural Networks for the Chapecó River Ecological Corridor, Santa Catarina/Brazil," *Sustainability*, vol. 14, no. 7, p. 4038, Mar. 2022, doi: 10.3390/su14074038.
[17] N. Nuthammachot and M. Zeeshan, "Comparative study of multiple algorithms classification for land use and land cover change detection and its impact on local climate of Mardan," vol. 18, no. June 2024, 2025.