

RESEARCH

Open Access



Characterizing land use/land cover change dynamics by an enhanced random forest machine learning model: a Google Earth Engine implementation

Chaitanya Baliram Pande^{1,2,5}, Aman Srivastava³, Kanak N. Moharir⁴, Neyara Radwan^{8,9}, Lariyah Mohd Sidek¹, Fahad Alshehri⁵, Subodh Chandra Pal⁶, Abebe Debele Tolche^{7*} and Mohamed Zhran¹⁰

Abstract

Land use and land cover (LULC) analysis is crucial for understanding societal development and assessing changes during the Anthropocene era. Conventional LULC mapping faces challenges in capturing changes under cloud cover and limited ground truth data. To enhance the accuracy and comprehensiveness of the descriptions of LULC changes, this investigation employed a combination of advanced techniques. Specifically, multitemporal 30 m resolution Landsat-8 satellite imagery was utilized, in addition to the cloud computing capabilities of the Google Earth Engine (GEE) platform. Additionally, the study incorporated the random forest (RF) algorithm. This study aimed to generate continuous LULC maps for 2014 and 2020 for the Shrirampur area of Maharashtra, India. A novel multiple composite RF approach based on LULC classification was utilized to generate the final LULC classification maps utilizing the RF-50 and RF-100 tree models. Both RF models utilized seven input bands (B1 to B7) as the dataset for LULC classification. By incorporating these bands, the models were able to influence the spectral information captured by each band to classify the LULC categories accurately. The inclusion of multiple bands enhanced the discrimination capabilities of the classifiers, increasing the comprehensiveness of the assessment of the LULC classes. The analysis indicated that RF-100 exhibited higher training and validation/testing accuracy for 2014 and 2020 (0.99 and 0.79/0.80, respectively). The study further revealed that agricultural land, built-up land, and water bodies have changed adequately and have undergone substantial variation among the LULC classes in the study area. Overall, this research provides novel insights into the application of machine learning (ML) models for LULC mapping and emphasizes the importance of selecting the optimal tree combination for enhancing the accuracy and reliability of LULC maps based on the GEE and different RF tree models. The present investigation further enabled the interpretation of pixel-level LULC interactions while improving image classification accuracy and suggested the best models for the classification of LULC maps through the identification of changes in LULC classes.

Keywords LULC, Change detection, GEE, Random forest, Remote sensing, Energy, SDG

*Correspondence:

Abebe Debele Tolche
abeberobe@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

The vegetation and waterbody maps are among LULC maps that provide a standard land surface outline [1–3]. The precise and timely collection of LULC information aids in gathering knowledge about human society's evolution and development. Subsequently, such investigations will allow active climate and environmental change modeling studies to develop, revealing the significance of climate change in society. Remote sensing (RS) imaging, GEE, and ML programming methods have been extensively used for classifying LULC at local and regional application scales [4–6]. However, the mid-1970s were a turning point for applying various interpretation approaches that aided in compiling LULC maps and change detection studies [7–9]. Owing to the importance of timely LULC information and advanced methodologies, the LULC variation mapping approach has evolved over the last five decades [10, 11]. One of the goals of scientific communities has been to generate spatially explicit LULC maps with the shortest possible temporal delay and best spatial resolution. The rapid development of RS technology has been accomplished to some extent compared with conventional surveying and mapping techniques [12–15]. These patients benefit from multiple observations and extensive coverage when suggesting high efficiency [16–18]. The most important aspect of modern-day planning is assessing LULC change and its implications as a baseline requirement for sustainable natural resource development, planning, and management [19, 20]. Many researchers have claimed that LULC significantly and consistently impacts the functioning of socioeconomic and environmental systems, with substantial trade-offs for sustainability, food security, biodiversity, and people's and ecosystems' socioeconomic vulnerability [21, 22]. Environmental factors such as altitude, geomorphology, hydrology, geological structure, soil type, slope, and technological, socioeconomic, and institutional characteristics are exclusively influenced by LULC classes [23–25]. The annual-scale LULC maps in the conterminous United States were created from 1973 to 2000. The dynamic patterns revealed that nearly 8.6% of the geographical area in the U.S. had experienced a change in LULC at least once during the analysis horizon [26]. Urban and rural population development, industrialization, climate change, and deforestation have all immediately impacted the LULC classes on the planet [27–29]. This has impacted various geo-environmental and ecosystem issues, including biodiversity, pollution, freshwater, the energy budget, and land use policy [30–32]. Moreover, LULC changes significantly affect climate, biogeochemical cycles, energy fluxes, and livelihoods [30, 33, 34].

To create long time-series LULC maps, remote sensing data are acquired and subsequently applied to unravel the intriguing layers of the Earth's land surface [35, 36]. Machine learning algorithms have played an essential role in LULC mapping and change analysis, particularly in optimizing the processing of large volumes of image stacks and mining the unique patterns and information associated with subsequent LULC modifications [37, 38]. Likewise, satellite images with improved resolutions have been more helpful in acquiring LULC maps with considerable knowledge and greater accuracy. Regional studies have employed a wide range of techniques, algorithms, and methods to accurately categorize LULC change maps [39, 40]. The utilization of Landsat images has proven valuable in providing a comprehensive overview of diverse landscapes [41, 42]. Additionally, object detection methods, machine learning classifiers, remote sensing, and geographic information system (GIS) approaches have been implemented to enhance the identification of LULC classes, leading to more accurate and reliable LULC maps [43]. Unmonitored gradation, supervised classification, GEE, machine-learning programming, and fuzzy grading are the most often used methodologies for image categorization via RS and GIS software [44, 45]. [46] prepared LULC variation maps based on parcels using a machine learning approach. [47] demonstrated future LULC changes and calibration via an impact assessment of LULC changes in two areas of Tehran, Iran. [48] used machine learning models such as support vector machines (SVMs) and artificial neural networks (ANNs) to extract better information and improve the accuracy of LULC change patterns in Kuwait. These methodologies, in fact, extend beyond LULC classification and have demonstrated successful applications in understanding a wide range of phenomena. These methodologies have been effectively applied in studying hydro-meteorological variations, climatological alterations, and the impacts of natural and anthropogenic disasters, such as floods and droughts. Their applications have provided valuable insights into sectors such as water resources and agriculture, among others [49–52]. To summarize, the studies mentioned above provide compelling evidence for the extensive application of GEE, ML, and other related platforms. These findings collectively highlight the robustness and effectiveness of these methodologies in various research domains.

With the advent of cloud computing approaches, several powerful platforms have been available for constructing LULC maps. Among these established platforms, the GEE has been used to conduct innovative research with greater ease in computational efficiency [53, 54]. The GEE has a significant volume of numerous long-term raw remote sensing observations and dynamic classification

algorithms, such as RF models [55, 56] and support vector machines (SVMs) [57, 58], which make it potentially effective at preparing continuous long-term LULC maps. Many studies have been piloted on the GEE and associated Google Earth platforms, such as global urban land clipped, surface water monitoring, land surface temperature, vegetation indices mapping, and LULC mapping with Landsat-8 images [59–63]. The combination of satellite data with indices and machine learning models for correcting demarcated surface water structures was based on the GEE for the Thoubal River Watershed, India [64]. LULC mapping is most important for understanding land changes and ecological systems; hence, all of these studies have produced rapid and precise results based on ML modeling via the GEE platform [65]. Object-oriented LULC extraction was performed using the Google Earth Engine platform with the integration of classifier models such as SNIC, GLCM, and ML [66]. The RF model was used to classify Zambian grasslands based on satellite data and feature importance [67]. The LULC mapping classification was used to integrate the pixel- and object-based approach from multi-temporal data, and this mapping included a random forest classifier [68]. Long-term spatiotemporal changes in surface water bodies in the Yellow River Basin from 1986 to 2020 were identified using GEE, ML and satellite data [69]. The effect of LULC variation on evapotranspiration and ecosystem system identification was studied using an ML model classifier via the GEE platform [70]

This study established a rigorous methodology for accurate LULC mapping and the identification of LULC classes utilizing the GEE platform. This study developed and compared the performances of two RF models, namely, RF-50 and RF-100, utilizing satellite data and training samples encompassing seven satellite bands for the Shrirampur area (in India) based on multi-temporal 30 m Landsat-8 satellite images. These models were instrumental in effectively classifying and generating LULC maps for the years 2014 and 2020. In addition, many regions across the globe are experiencing climate-related issues, yet the reasons behind the constant fluctuations in climate and pollution levels on the Earth's surface remain poorly understood. It is evident that various factors, including changes in LULC patterns, significantly impact climate and agricultural fields. In light of these circumstances, accurate and timely LULC maps can be developed with ML models. The present study employed the RF algorithm as mentioned before. The selection of these models allows the investigation of which model yields higher accuracy in classifying images for LULC mapping. This study aimed to gain valuable insights into the relationships among LULC patterns,

climate dynamics, and agricultural impacts by adopting these approaches.

The specific objectives and aims of this study were to (1) develop RF models with 50 and 100 trees to achieve accurate LULC mapping within the study area. By employing advanced ML, we aim to enhance the precision and reliability of LULC classification. (2) identify changes in LULC mapping by utilizing both the RF-50 and RF-100 models in conjunction with the GEE platform. This approach enables a comprehensive analysis of temporal variations in land use and land cover, facilitating a deeper understanding of LULC dynamics over time. (3) Analyze the effectiveness of the proposed algorithms in accurately identifying individual LULC classes and assessing the classification accuracy. By evaluating the performance of the RF models, we can determine their ability to distinguish different land cover categories, thereby providing insights into the strengths and limitations of the classification approach. (4) Additionally, we can compare and determine the optimal tree model and input parameters for LULC mapping using GEE and a machine learning approach. This comparative analysis will inform the selection of the most suitable model configuration and input parameters to achieve high-precision LULC mapping results. By identifying the optimal approach, we aim to improve the overall accuracy and reliability of the generated LULC maps. By achieving these objectives, the study aimed to offer practical recommendations for LULC mapping in the study area. The outcomes of this research can guide decision-makers in implementing efficient and timely management strategies based on up-to-date LULC information. A comparison between the RF-50 and RF-100 tree models will provide valuable insights for selecting the most suitable approach for accurate LULC mapping.

Materials and methods

Study area and database

The study area is located at 19.62 N, 74.66 E in the western Maharashtra, as shown in Fig. 1. The most significant land cover type is farmland. Large patches and plain topography characterize the agricultural area, and dry land is commonly cultivated with sugarcane, wheat, and onion in rotation of the study area. The mean temperature, wind speed, and humidity are 27 °C, 14 km/h, and 70%, respectively, for the study area. Sugarcane is commonly harvested in December and January, while sugarcane fields are sown between June and February. Wheat is generally harvested in March, and onion is grown between June and August. The annual rainfall is 400 to 550 mm, mostly confined to the monsoon season (June–October). The basin experiences highly undulating

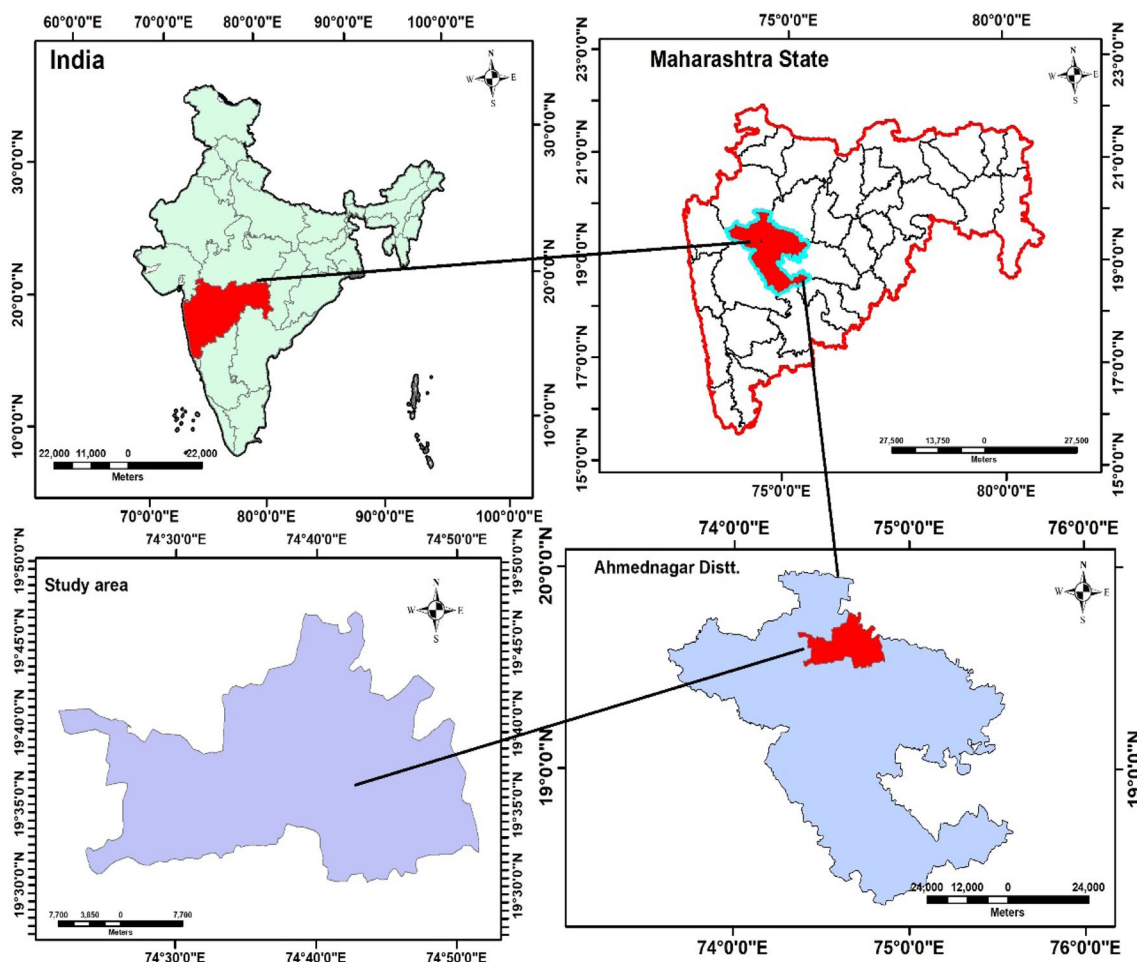


Fig. 1 Location map of the study area

topography, with the highest elevation being 541 m in the upland locations.

Remote sensing database

The 30-m Landsat-8 top-of-atmosphere (TOA) multi-temporal satellite images were used for LULC mapping for 2014–2020; these images were subsequently accumulated and systematically processed on the GEE platform [71]. The revisit phase of Landsat-8 satellites can reach 16 days, and after one month, two datasets are available as time series observations. Landsat-8 TOA images of 12 varying spectral and thermal bands, with spatial resolutions varying from 30 m to 60 or 100 m, individually serve different purposes. Considering that is not all Landsat-8 satellite images have been geometrically and atmospherically rectified [72], the Landsat-8 TOA reflectance datasets under Level 1-C were used in the land use mapping for the period 2014–2020. Furthermore, due to cloud cover, air pollution, and seasonal, intermittent rainfall, this study used <10% clouds to filter the accessible

data through the downloading data gateway. All Landsat-8 satellite images were acquired for LULC extraction during 2014–2020 using 50 and 100 tree RF models. The satellite images were obtained and processed using the GEE platform and developed algorithm, while the data availability were checked by the Google Earth code editor platform (Table 1) [73].

Table 1 Specification of the satellite products used

Sensor	Landsat-8
Provider	U.S. Geological Survey (USGS)
Resolution	30 m
Data Acquisition	2014 and 2020
Data source	https://developers.google.com/earth-engine/datasets

Training and validation datasets

This study identified and collected random field datasets for each LULC class with some selected data points using the visual interpretation method from Google Earth images. To ensure the accuracy and robustness of the image classification, 800 samples were collected for each year, as shown in Fig. 2. The collected sample data were divided into 70% and 30% for the training and validation periods, respectively, for both the 50 and 100 tree classification models. These datasets have been used in the classification models for 50 and 100 tree RF models using the GEE interface.

Image classification protocols

The GEE is a cloud-based and web-based graphical user interface (GUI) platform. The GEE serves as a repository for a vast collection of remote sensing datasets. The GEE combines a multipetabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities [35, 36]. Developers use GEE to detect changes and map trends and quantify differences on the Earth's surface. The public data archive includes more than thirty years of historical imagery and scientific datasets that are updated and expanded daily. GEE can enable processing of large datasets using the JavaScript code editor platform, which shortens the correction and processing of satellite datasets [74]. In this research, the LULCs were classified using the RF-50 and RF-100 tree models in the GEE platform, which included processing, atmospheric correction, cloud-free image creation, missing and gap data filling, training, testing, and confusion matrix preparation. In this study, we compared 50 and 100 RF tree models, which yielded more accurate LULC mapping based on the machine learning approach and GEE platform. Cloud-free satellite data are available for the selected dates and subsequently estimated to improve the discriminatory ability of various land cover patterns. During image processing, composites are formed and

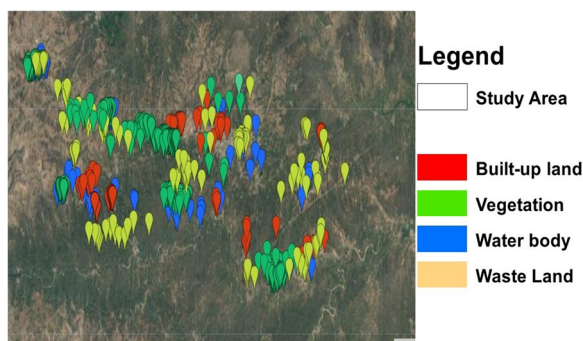


Fig. 2 Training points (field location-based) in the study area

used for land cover data to avoid missing information and cloudy images, and land cover maps are derived using the GEE platform. We have been selected a two-year dataset for the periods 2014 and 2020 for the preparation of LULC maps and subsequently used it to analyze the spatiotemporal distribution of the LULC maps using the RF-50 and RF-100 tree models and observed the LULC changes in the study area.

The classification procedure was separated into three steps: data collection, segmentation, and classification. Two approaches characterize the conventional LULC classification method: unsupervised and supervised classifications, wherein the former is based on a radiometric resolution-based algorithm-specific classification. It was later characterized by ground truthing-based user-defined classification. In the present study, the supervised classification algorithm was chosen to evaluate the potential of existing image classification approaches. During classification, ground truthing was performed using the GEE, wherein at least 200 training samples were chosen for individual LULC classes. To apply the RF model, two basic parameters are needed: the number of trees (ntree) for building an entire forest and the number of selected features utilized for node splitting. The selected study locations inside the study area chosen for analysis were identified on the basis of satellite images, as presented in Fig. 3. LULC change detection was performed using RF-50 and 100 tree models, and the model that exhibited better results is recommended for effective LULC mapping at the location of interest. A flowchart of the adopted novel methodology is shown in Fig. 4, Table 2a to d and Table 3 a to d.

Random forest machine learning model

It is a bootstrap resampling-based combinatorial classifier that extracts numerous subsets of training samples from the original data to generate multiple subdatasets [75, 76]. Furthermore, various iterations of the decision process are performed using the previously generated datasets to train the model. The decision trees are treated as classification trees operated based on a set of binary rules to estimate a target value. The RF algorithm computes response variables such as the land cover class by generating numerous decision trees (in the order of 100 s). Subsequently, each object to be modeled is attached below the decision tree. The response from individual decision trees was evaluated, and the most predicted class was assigned as the targeted LULC class. Hence, the efficiency of the RF-based model is highly dependent on the mode of creation of the decision trees.

In general, the random selection process in the tree formation stage of the RF algorithm is executed in two steps. In the first step of tree formation, random selection is

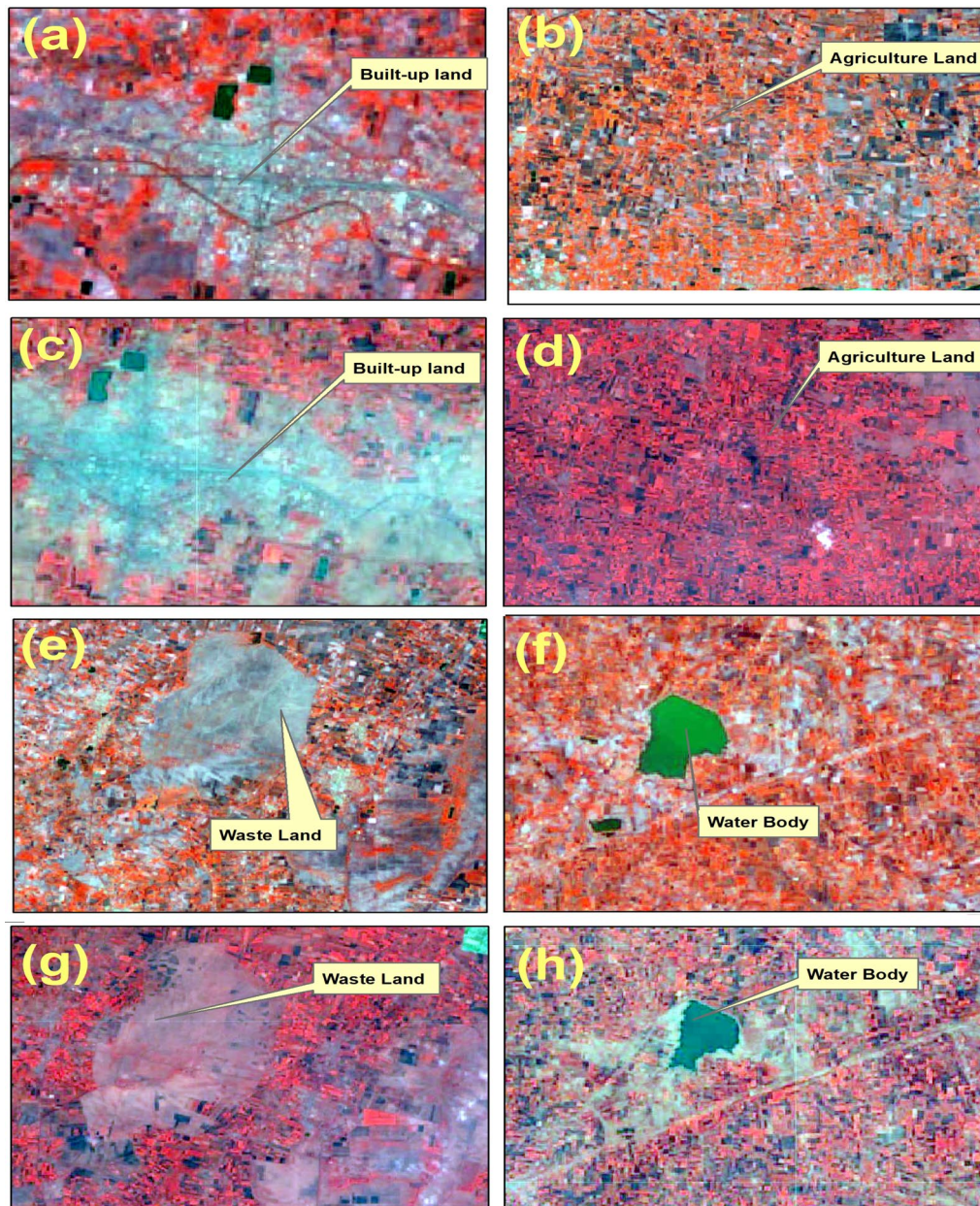


Fig. 3 The selected sites of the LULC classes from 2014 to 2020: **a** built-up land of 2014, **b** agricultural land of 2014, **c** built-up land of 2020, **d** agricultural land of 2020, **e** waste land of 2014, **f** water body of 2014, **g** waste land of 2020, and **h** water body of 2020

accompanied by the replacement of data from the training sample. For an individual tree, a subset of the training data is used for decision-making, and remaining data, known as the “out-of-bag (OOB)” sample, are used for testing the model accuracy. The second step uses the binary rule to determine the split condition at each node. The splitting rule primarily includes the maximum information gain, the maximum information gain rate, and the minimum Gini index. In the process of partitioning, the

purity of the node increases, i.e., the sample contained by the node is included in the same category. When a large number of trees are generated, voting for the most popular class is performed; hence, this approach is treated as the RF technique.

The RF classifier consisted of a collection of decision tree classifiers, as given below [77]:

$$\{h(x, \Theta_k) \quad k = 1, \dots\} \tag{1}$$

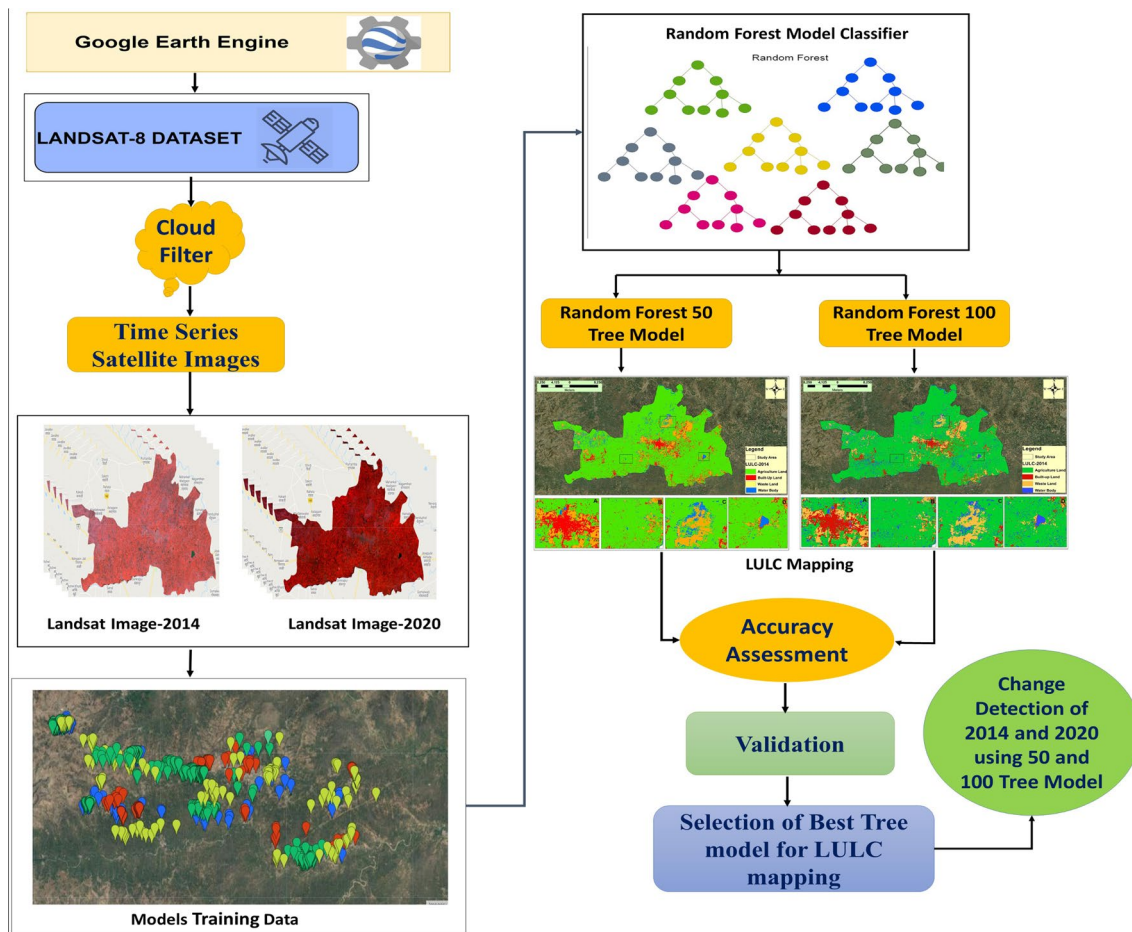


Fig. 4 Flowchart of the methodology framework of the study

where x is the input vector and $\{\Theta_k\}$ is the independent identically distributed random vector.

The OOB elements are used to estimate the model performance; this metric is termed an unbiased estimator of generalization error and is given as follows:

$$V(x_i) = \frac{1}{N} \sum_{t=1}^{t=N} (e_t^i - e_t) \tag{2}$$

where $V(x_i)$ is the mean decrease in the accuracy score, e_t is the OOB error in each decision tree of the RF, and e_t^i is the new i th OOB error computed by altering the values of variable x_i .

The present study adopted two RF classification models considering the models' varying decision degrees. The first model consists of 50 decision trees, and the second model consists of 100 decision trees, referred to as RF-50 and RF-100. The training of the supervised classification algorithm was performed individually by the two abovementioned models. Finally, the land cover

classification for the two analysis years, 2014 and 2020, was achieved using the two classification algorithms. The overall methodology framework adopted in the present study is described in Fig. 4.

Accuracy assessment for LULC maps of classification

An accuracy assessment of any prepared LULC map is a standard statistical procedure that cannot be avoided for obtaining a perfect LULC scenario for the region of interest. The steps followed to accomplish the accuracy assessment are detailed as follows:

1. Initially, 50 random points were generated within the spatial extent of interest with the 'create random sample' tool of ArcGIS 10.3.
2. Furthermore, the corresponding LULC classes at predefined random locations were extracted and updated in the attribute table belonging to the "random point raster".

Table 2 Resubstitution error and confusion matrix for the random forest (RF) 50 model

(a) Resubstitution error matrix for 2014				
Land-use type	Agriculture land	Built-up land	Wasteland	Waterbody
Agriculture land	55	1	4	0
Built-up land	0	78	0	0
Wasteland	0	1	295	1
Waterbody	1	0	1	62
Training Overall accuracy=0.98				
(b) Confusion matrix for 2014				
Land-use type	Agriculture land	Built-up land	Wasteland	Waterbody
Agriculture land	11	0	13	3
Built-up land	1	17	7	3
Wasteland	3	3	109	5
Waterbody	1	2	4	19
Testing (Validation) accuracy = 0.78				
(c) Resubstitution error matrix for 2020				
Land-use type	Agriculture land	Built-up land	Wasteland	Waterbody
Agriculture land	50	0	2	0
Built-up land	0	76	2	0
Wasteland	0	0	309	0
Waterbody	0	0	3	60
Training Overall accuracy=0.98				
(d) Confusion matrix for 2020				
Land-use type	Agriculture land	Built-up land	Wasteland	Waterbody
Agriculture land	16	4	14	1
Built-up land	1	27	0	0
Wasteland	2	1	101	4
Waterbody	0	1	8	18
Testing (Validation) accuracy = 0.82				

3. A random point raster file was opened on Google Earth Pro., and individual random points were compared with the ground truth LULC classes derived from Google Earth Pro.
4. The corresponding match between the ground truth LULC class and the image-derived class is assigned a true value.
5. Finally, the overall accuracy was estimated as follows:

adapted to increase the confidence in the classification accuracy. This approach ensures confidence in the class division within the map. It provides information about the actual accuracy obtained and the agreement expected by chance. Different ranges of the kappa coefficient indicate that a map with a degree of accuracy in the range of 0.85–0.99 is treated as a better model. This approach could be adopted for future LULC preparation. The kappa coefficient is calculated as follows:

$$Accuracy_{Overall} = \frac{\text{Total sum of all the correctly accessed points}}{\text{Total number of random points taken}} \times 100 \tag{3}$$

A generalized discrete multivariate technique, i.e., the kappa coefficient-based accuracy approach, is further

$$Kappa\ coefficient = \frac{\text{Observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}} \tag{4}$$

Table 3 Resubstitution error and confusion matrix for Random Forest (RF)-100 model

(a) Resubstitution error matrix for 2014					
Land-use type	Agriculture land	Built-up land	Wasteland	Waterbody	
Agriculture land	66	0	2	0	
Built-up land	0	89	0	1	
Wasteland	0	0	309	0	
Waterbody	1	0	1	102	
Training Overall accuracy=0.99					
(b) Confusion matrix for 2014					
Land-use type	Agriculture land	Built-up land	Wasteland	Waterbody	
Agriculture land	16	0	14	2	
Built-up land	1	22	8	9	
Wasteland	6	2	118	7	
Waterbody	5	2	10	29	
Testing (Validation) accuracy = 0.74					
(c) Resubstitution error matrix for 2020					
Land-use type	Agriculture land	Built-up land	Wasteland	Waterbody	
Agriculture land	68	0	1	0	
Built-up land	0	92	0	1	
Wasteland	0	0	312	0	
Waterbody	0	0	3	103	
Training Overall accuracy=0.99					
(d) Confusion matrix for 2020					
Land-use type	Agriculture land	Built-up land	Wasteland	Waterbody	
Agriculture land	16	2	11	2	
Built-up land	1	31	3	2	
Wasteland	2	3	121	4	
Waterbody	1	0	19	23	
Testing (Validation) accuracy = 0.79					

where *Observedaccuracy* is the accuracy obtained from the total accuracy and *chanceagreement* is the sum of the products of the row and column totals for each class.

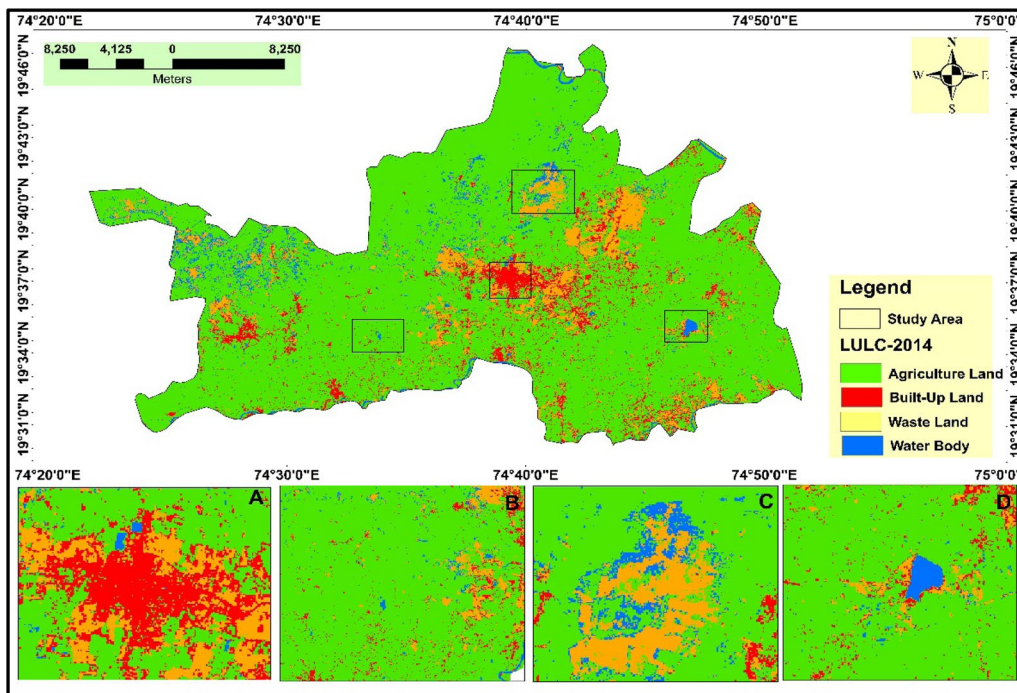
The points that needed to be validated were generated in the ESRI ArcGIS 10.3 software using the stratified random sampling method of randomly creating control points. The control points of the respective LULC classes were established in the same way as the training polygons were created. The control points were uploaded to the GEE, where the inbuilt classifier package was used to validate the classification. Specifically, the classifier.confusionMatrix() function and the errorMatrix() function of the GEE constructed the confusion matrix and overall accuracy of classification, followed by the ConfusionMatrix.kappa() function for the estimation of the kappa index. The kappa index was estimated for each combination of input parameters, and the combination

corresponding to the highest kappa index was selected to determine the overall accuracy.

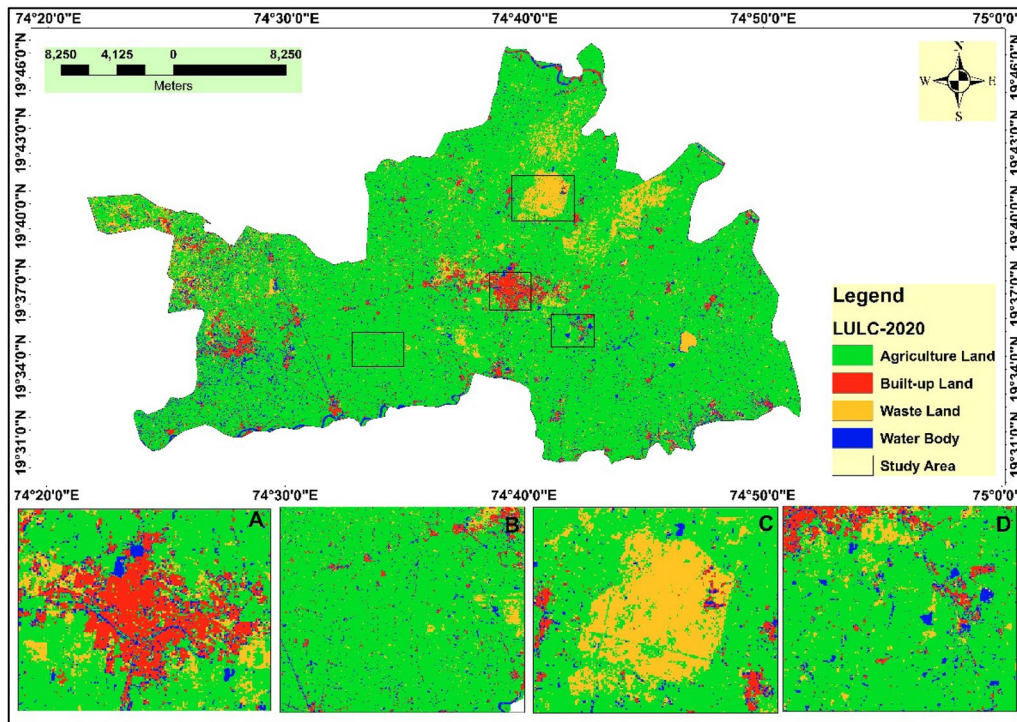
Results

LULC thematic (spatial) maps and accuracy assessment for the maximum likelihood, RF-50, and RF-100 tree models

The LULC thematic maps of the Shrirampur area for 2014 and 2020 obtained using the RF-50 tree model and RF-100 tree model are presented in Figs. 5 and 6, respectively. The four LULC classes, agricultural land, built-up land, wasteland, and land covered with water bodies, were identified as the essential classes and are the focus of the present investigation. Spatially, the central region of the study site broadly represents built-up land (zoomed region marked as “B”), which is dominantly surrounded by wasteland (region “C”). While the water bodies are limited in the built-up region, their availability is

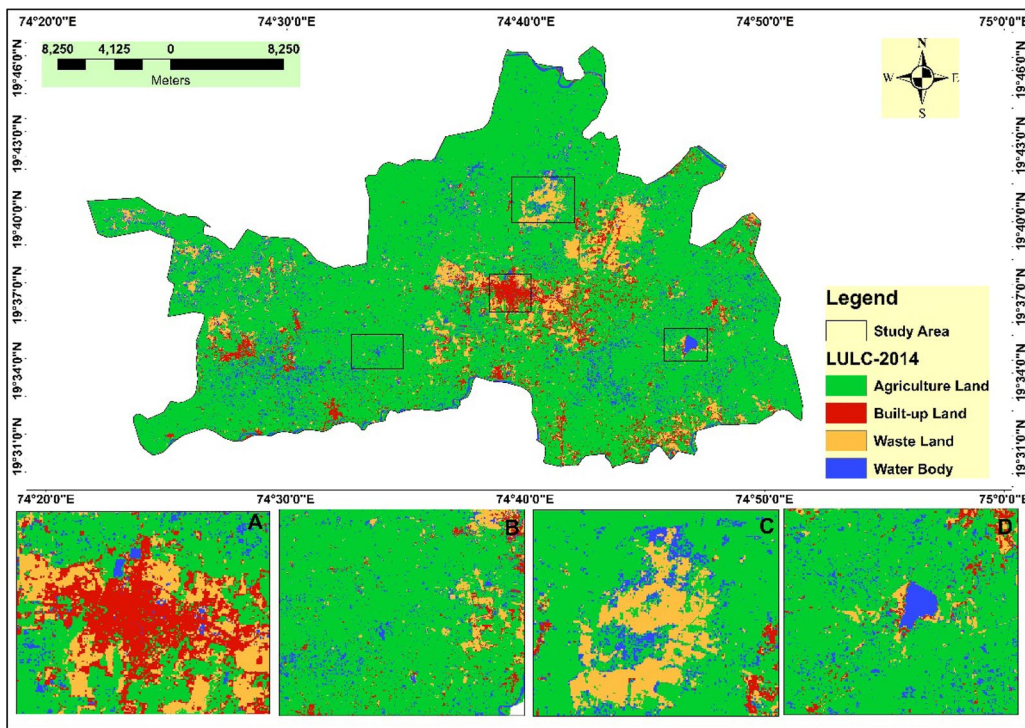


(a)

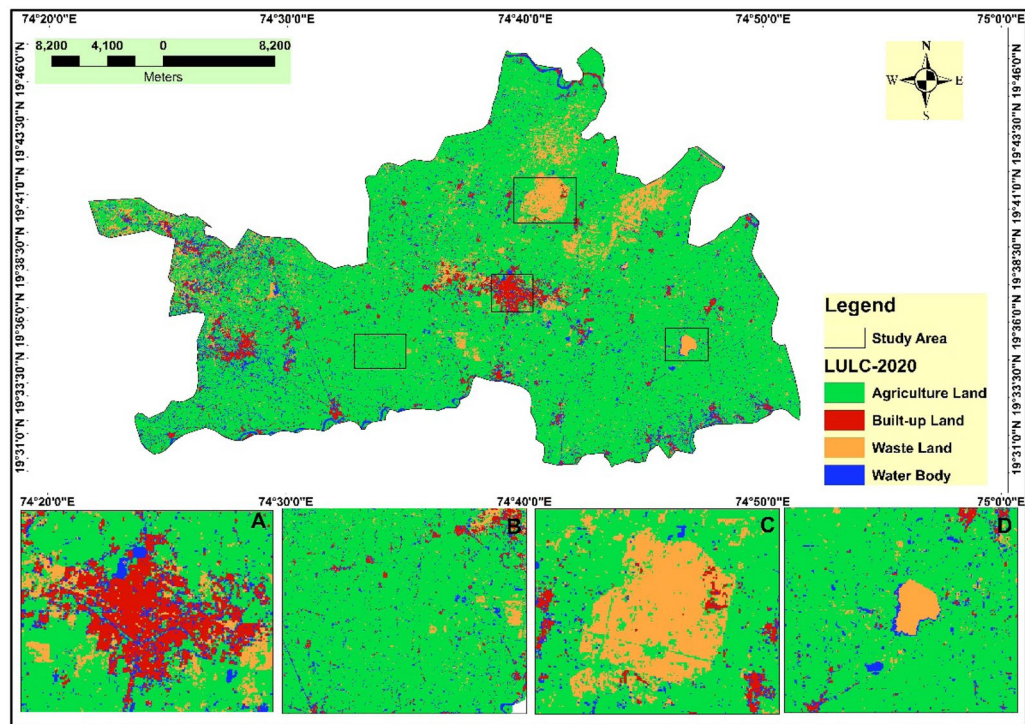


(b)

Fig. 5 LUC thematic map for the Shirampur region in Maharashtra for **a** 2014 and **b** 2020; developed using the RF-50 tree model



(a)



(b)

Fig. 6 LUC thematic map for the Shrirampur region in Maharashtra for **a** 2014 and **b** 2020; developed using the RF-100 tree model

localized primarily in the North and North–West regions (regions “C” and “D”), with one major water source in the East (region “D”). Agricultural practices are the primary occupation of the study site. They are located at the peripheries of the study boundary (region “A”) in close vicinity with large numbers of micro- and medium-scale water bodies. Qualitatively, significant changes occurred across 2014 and 2020 in the LULC types described above, as evident from the spatial plots; however, the RF-50 Tree and RF-100 Tree models depicted the alterations differently. Therefore, a quantitative analysis was conducted and is presented in Sects. “Advantages, limitations, and future research on the GEE and RF models” to ascertain the essential changes in the LULC classes. This aspect motivated the study to assess each model for the same study site to quantify the models’ suitability.

The RF models developed for the present study site were assessed for accuracy and suitability for the Shrirampur study site. This was performed by developing a resubstitution error matrix and confusion matrix for each LULC type in 2014 and 2020. Table 2a, b, c, d show the matrixes for the RF-50 tree model and RF-100 tree model, respectively. For the RF-50 Tree model, the findings indicated 98% accuracy under overall training for both 2014 and 2020, while 78% and 82% accuracy were recorded under the validation phase for the same years (Fig. 7). Conversely, for the RF-100 Tree model, the findings indicated 99% accuracy under overall training for both 2014 and 2020, while 74% and 79% accuracy were recorded under the validation phase for the same years (Fig. 7b; Table 3a, b, c, d). The developed RF models and derived high-accuracy LULC maps were inferred to be highly suitable for conducting spatiotemporal investigations and pattern analysis at the present study site.

Analysis of RF-50 tree model-estimated spatiotemporal changes and patterns of LULC

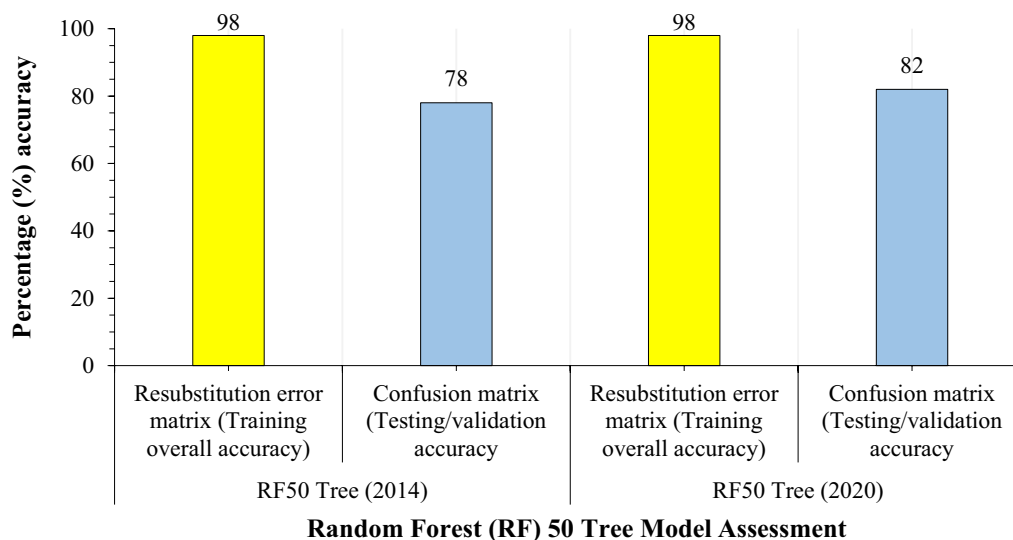
To quantitatively assess and compare the spatial and temporal changes in different LULC types in the Shrirampur area, the regions estimated from the RF-50 tree model and the cumulative ratio for each class for 2014 and 2020 were computed and are presented in Fig. 8. It can be inferred that agricultural land cover is the most dominant land use, with an area proportion of 84%. This was followed by area cover under the wasteland occupying 7% of the study site. In general, agricultural land use coupled with wasteland remained less altered in view of individual area proportion alterations between 2014 and 2020. In contrast, the area proportions of built-up land and water bodies underwent significant changes across the same years. For example, the area proportion of built-up land decreased from 6% in 2014 to 4% in 2020, whereas

the area proportion of water bodies increased from 3% in 2014 to 5% in 2020.

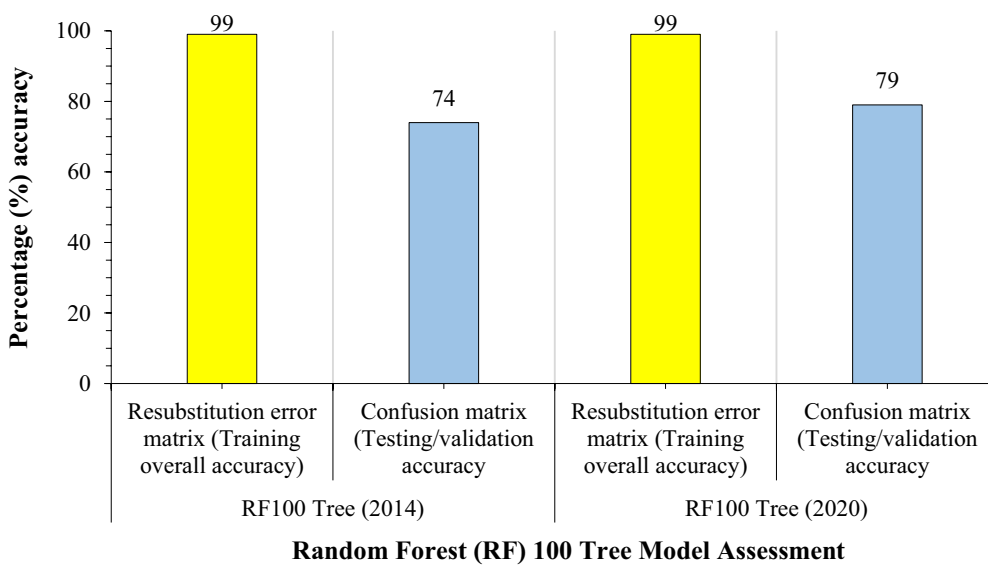
Changes in different LULC types are more apparent in Fig. 10, where a drastic positive shift in the area covered with water bodies, as high as 42% (from 25 km² to 36 km²), was observed. Conversely, a drastic negative change in area cover under built-up conditions was observed, reaching 26% (from 47 km² to 35 km²). A comparison of the percentage (%) change between RF-50 Tree and RF-100 Tree revealed that the area of built-up land decreased, as shown in Fig. 9. We compared the results of both models for built-up land and found that the RF-100 tree model extracted less built-up land because of the decrease in built-up land, as shown in Fig. 9. In this paper, we found that the RF-50 tree better classified built-up land than did the RF-100 tree, and the RF-100 tree had a slightly lower performance in classifying built-up land overall. In contrast, the field investigation and qualitative interviews revealed rising households (built-up land) and a reduced water quantity across the study site. To argue, the increase in the area covered by water bodies and decrease in built-up land area are opposite to the prevailing (general) trends in the Shrirampur region. Despite these facts, the RF-50 Tree model yielded contradicting results (at least in the case of built-up spaces and water bodies). The aforesaid findings suggested redoing the analysis using the RF-100 Tree model to ascertain the suitability of a specific RF model in coherence with the study site conditions. Moreover, it is important to highlight here that merely applying an RF model to a study site, just because its accuracy assessment is appropriate (as observed for the RF-50 Tree model in the present case), is insufficient to determine the different degrees of alternation in LULC. There is a need to develop multiple models of such types (for example, the two models developed for the present study area, viz., the RF-50 Tree and RF-100 Tree models) to compare their findings and, if possible, coupled with field investigations (ground truthing) to validate the results.

Analysis of RF-100 tree model-estimated spatiotemporal changes and patterns of LULC

This section aimed to quantitatively assess and compare the spatial and temporal changes in different LULC types in Shrirampur Province via area estimations generated by the RF-100 tree model. These parameters were computed statistically (the cumulative ratios for each class for 2014 and 2020 were computed) and are presented in Fig. 10. Like in the RF-50 Tree model, in the RF-100 Tree model, agricultural land cover was the most common land use, with an area proportion of nearly 84% (a slight increase of 1% between 2014 and 2020 was observed). This was followed by area cover under the wasteland occupying 7%



(a)

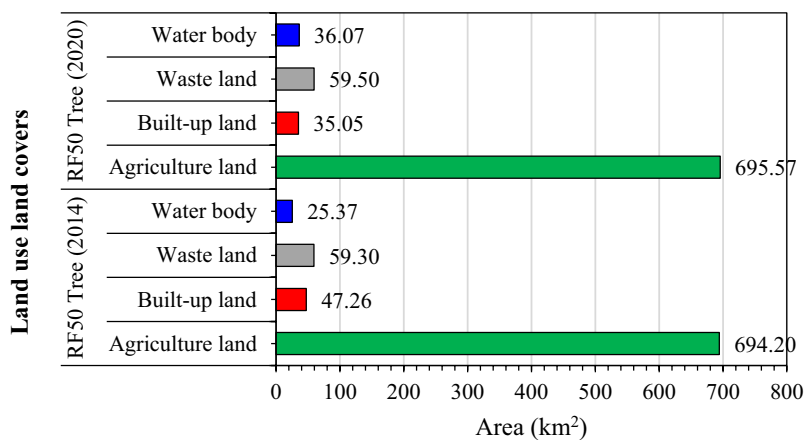


(b)

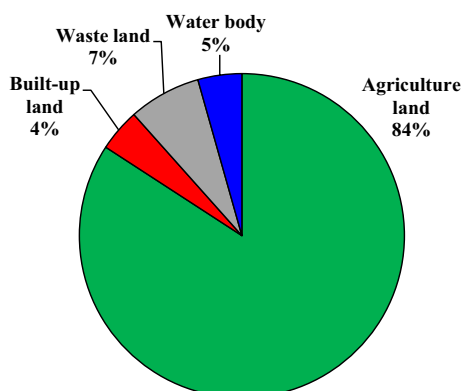
Fig. 7 Accuracy assessment of random forest (RF) models, viz., **a** for the RF-50 tree model and **b** RF-100 tree model, using the resubstitution error matrix and confusion matrix

of the study site, as observed in the RF-50 Tree model. In general, agricultural land use coupled with wasteland also remained less impacted in view of individual area proportion alterations between 2014 and 2020. The area under built-up land was also observed to be more or less the same during the study period (4%); however, water bodies changed across the same years, decreasing from 6% in 2014 to 5% in 2020 (against the increasing trend, as indicated by the RF-50 Tree model).

Changes in different LULC types are more apparent in Fig. 11, where a drastic positive shift in area cover under built-up land, as high as 24% (from 34 km² to 42 km²), was observed. However, a moderate negative change in the area covered with water bodies, which was 6% (from 46 km² to 56 km²), was observed. These findings are in agreement with the observations drawn from field investigations and qualitative interviews and contrary to the findings of the RF-50 Tree model. Importantly, even though both the RF-50 and RF-100 models satisfied the

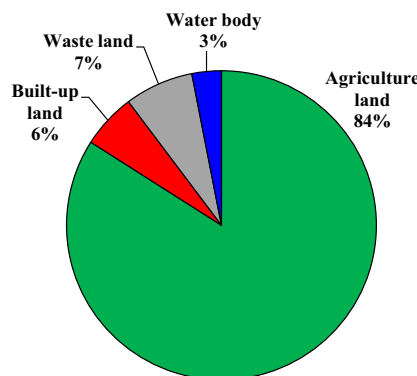


(a)



Percentage (%) area in 2020 using RF50 Tree

(b)



Percentage (%) area in 2014 using RF50 Tree

(c)

Fig. 8 Land use/land cover (LULC) change analysis using the random forest (RF)-50 tree model for 2014 and 2020; **a** showing LULC changes at a glance for both study periods and **b** and **c** showing the area proportion of each LULC type in the respective study periods for the Shirampur area of Maharashtra

accuracy assessment criteria, their area demarcation ability and area proportion were observed to vary, thereby resulting in different results for the same study site. In the case of the former, the trend of LULC alterations for water bodies and built-up land was observed to increase and decrease, respectively; however, the trend almost reversed when it was subjected to the latter. For the other two land use types, viz., wasteland and agriculture, their performances were similar. There is a need to quantify the difference in performance between these two models (done in the next section). Furthermore, developing a model for multiple study sites may provide contradictory results whose validation could not be performed unless a multiple model development approach is considered (as happened in the present case with the RF-50 Tree and RF-100 Tree models). Hence, the findings and inferences suggest considering a multiple model development

approach for deciding the suitability of RF models alongside conducting accuracy assessments.

Discussion

Typewise comparisons between LULC classes generated by the RF-50 tree and RF-100 tree models

The present section describes the varying results yielded by the RF-50 and RF-100 models for the Shirampur study area. The objective here is to quantify the suitability of these two models by comparing their area estimations. Figure 12 shows the spatial plots for different LULC types developed using the RF-50 and RF-100 models for 2014 and 2020. These two models provide insight into the visual interpretation of alterations in LULC. LULC classes such as agricultural land, built-up land, wasteland, and water bodies were estimated by the RF-50 and RF-100 models, as shown in Fig. 13. In general, the findings for each LULC type indicated a substantial difference in area

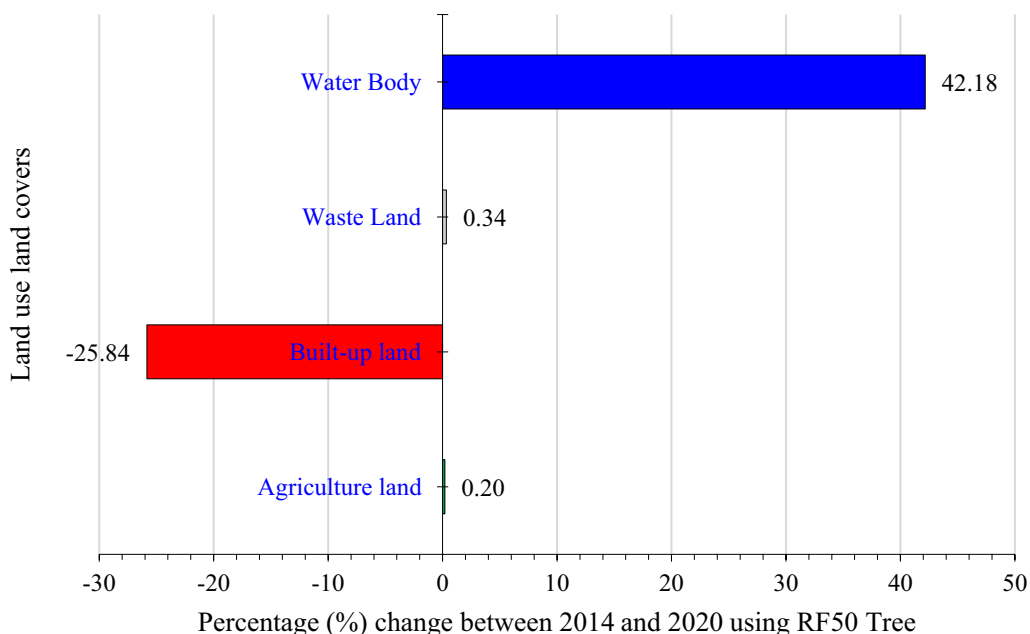


Fig. 9 Comparative analysis of LULC alterations between 2014 and 2020 for the random forest (RF)-50 tree model developed for the Shirampur area of Maharashtra

estimations for 2014 and 2020, especially for water bodies and built-up land (as was also identified in the previous sections). For example, in the case of water bodies, the area estimated by the RF-50 Tree model for 2014 was 25 km², while that estimated by the RF-100 Tree model was 46 km². A large positive difference (meaning that the estimated area increased from the RF-50 Tree model to the RF-100 Tree model) was observed for a total of 81% of the water bodies in 2014. However, the area estimation for water bodies in 2020 was positive at 20%. Similarly, in the case of built-up areas, a large positive difference of 27% was observed for 2014, and a large negative difference (meaning that the estimated area decreased from the RF-50 Tree model to the RF-100 Tree model) of 21% was observed for 2020. The difference in the area estimations for water bodies and built-up land was significant for both 2014 and 2020. However, it would be interesting to observe the periodwise variation in the case of the other two land use types to ascertain the aforementioned phenomena for these two models.

According to Fig. 13a, c, it can be inferred that the differences in the estimated area cover for wasteland and agricultural land using the RF-50 Tree and RF-100 Tree models, respectively, exhibited negative differences of 5% and 0.7% for 2014 and a mere 2% and 6% for 2020. In light of the evidence shown, this study revealed that both the RF-50 Tree and RF-100 Tree models are suitable for estimating the area covered by agricultural land and wasteland and thus can be employed for future land-use

investigations in the present study area. However, when these models are subjected to estimating the areal cover under built-up spaces and water bodies, the RF-100 Tree model performed better than the RF-50 Tree model. The reason may be attributed to the LULC characteristics, as dominant land uses (agricultural land and wasteland) were estimated better by both models than was recessive land use (built-up and water bodies). In addition, the better fit of the RF-100 Tree model can also be attributed to its slightly better accuracy than that of the RF-50 Tree model. Since the RF-100 Tree model corresponded well with the realistic trend of a declining cover of water bodies coupled with an increasing cover of built-up land (against the findings of the RF-50 Tree model), this study suggested the application of the RF-100 Tree model for future investigations.

Advantages, limitations, and future research on the GEE and RF models

The present study adopted the RF-based classification algorithm for characterizing the long-term change in LULC in the corresponding study region, which has the following advantages over the existing approaches. The ability of RF to process large amounts of data upon excluding feature selection helps in balancing the classification error, thereby providing a greater degree of accuracy in classification. Since the RF-based methodology requires minimal manual intervention, the RF design becomes simpler and requires less effort from

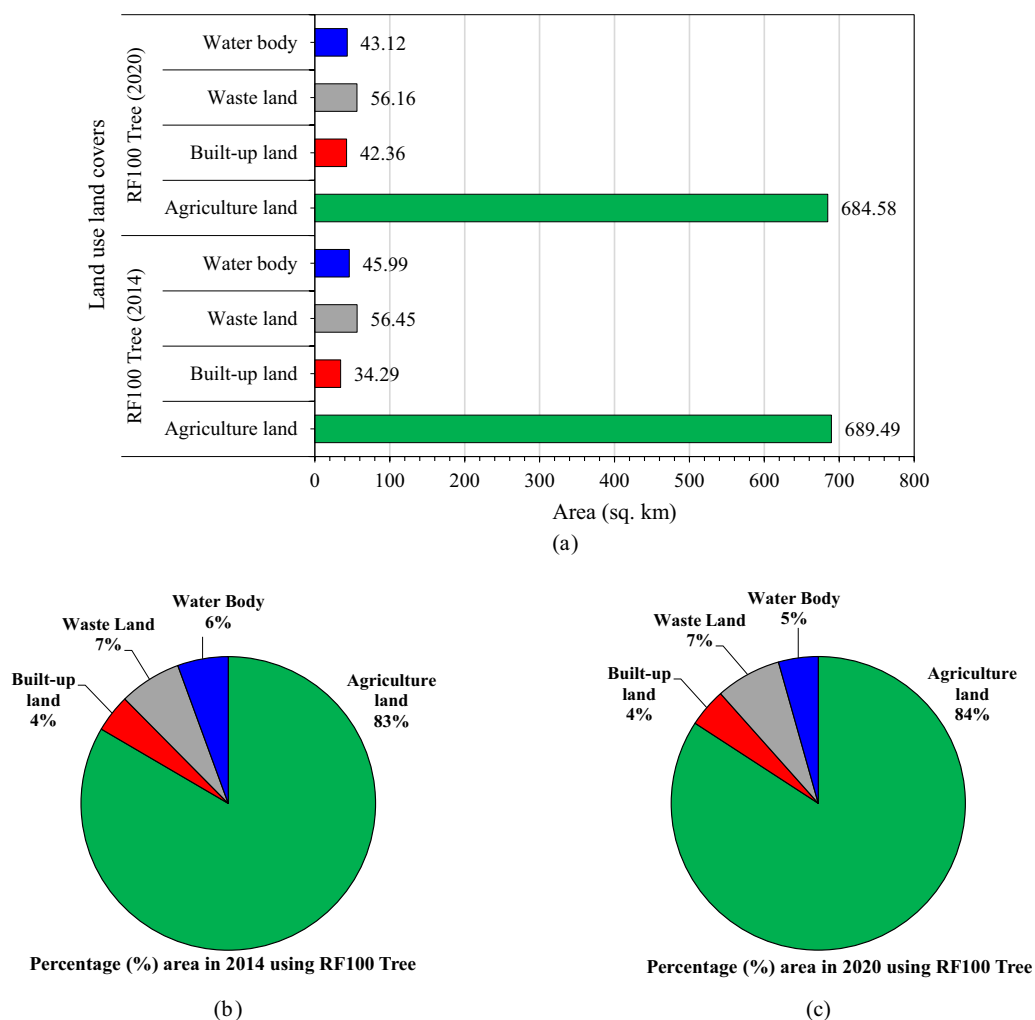


Fig. 10 Land use/land cover (LULC) change analysis using the random forest (RF)-100 tree model for 2014 and 2020; **a** LULC changes at a glance for both study periods and **b** and **c** the area proportions of each LULC type in the respective study periods for the Shirampur area of Maharashtra

the modeler. The inherent potential of RF for the characterization of a variety of data assists in evaluating the importance of individual land use classes by estimating the generalization error, and the overall accuracy of classification has improved. The enhanced computational speed of the RF algorithm enhances the parallelization in the building of decision trees for classification and subsequently reduces the time required for classification. In general, the RF algorithm behaves more as an ensemble tree approach than does conventional approaches, viz., boosting and stacking; hence, the RF algorithm becomes overly sensitive to the quality of training data, and instability persists with slight alterations in the training samples. This investigation provided a better understanding of several important characteristics. The utilization of GEE has revolutionized the process of generating LULC maps in a highly efficient and organized manner, thereby

surpassing the limitations of previous approaches [78]. In the past, generating LULC maps involved cumbersome tasks such as processing satellite data using local computers. This process was not only challenging but also inefficient, particularly when dealing with image processing tasks such as cloud removal, atmospheric correction, and mosaicking. However, with the introduction of GEE, these challenges have been effectively addressed. The GEE offers a cloud-based platform with powerful computational capabilities that streamline the generation of LULC maps. This platform facilitates seamless data processing, including automated cloud removal, atmospheric correction, and mosaicking. As a result, the entire workflow has become significantly more efficient, allowing for faster and more accurate LULC mapping.

Moreover, the motivation behind using RF-50 and RF-100 trees, despite the default mtree value being 500,

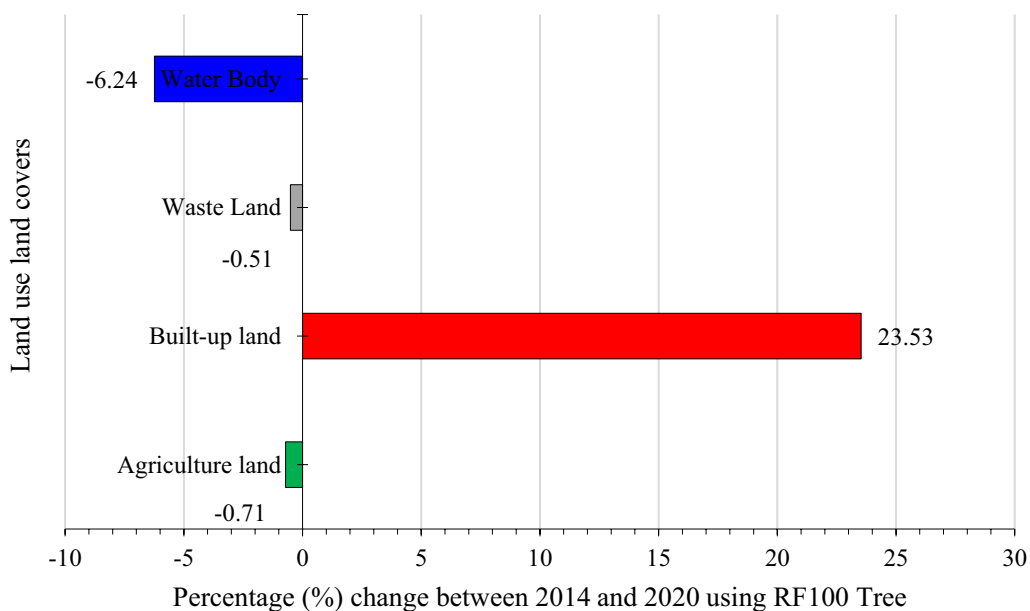


Fig. 11 Comparative analysis of LULC alterations between 2014 and 2020 for the random forest (RF)-100 tree model developed for the Shirimpur area of Maharashtra

is to explore the performance and accuracy of the RF algorithm with different numbers of trees in the context of this study. When the default value of *mtree* is set to 500, this does not necessarily imply that this is the optimal or most effective number of trees for all cases. By selecting RF-50 and RF-100 tree models, the study could investigate the impact of reducing or increasing the number of trees on the classification accuracy and performance of the RF algorithm in the specific application of LULC mapping. This approach allowed for a comparative analysis and determination of whether a lower or higher number of trees can yield improved results for the given study area and data. This approach thus provides valuable insights into the performance characteristics of the RF algorithm with varying tree numbers, surpassing the default setting of *mtree*.

As a complete large dataset cloud-based platform for resource combination, GEE is a satellite data processing and analysis platform that can aid in large storage and robust calculations [78]. Users can quickly access interactive satellite images and algorithms to conduct research-related scientific investigations and display spatial and nonspatial data [79]. The GEE platform is currently the best option for data collection and includes a set of geospatial datasets and satellite images from remote sensing satellites that cover more than 40 years, with some vector data also presented in the GEE platform [78]. While GEE has been used extensively to create maps of different types of land use and land cover [80], investigations of the estimation of

soil qualities rarely employ it. As a result, additional research and development are required for the implementation of GEE in the random forest model. As a result, implementing GEE can significantly increase researcher productivity through the use of machine learning models [71], and GEE has been regularly used in scientific analysis in several domains at the local, regional, and international levels [80–84].

The present study acknowledges certain limitations that need to be addressed for a more robust understanding. First, the RF-50 and RF-100 tree classifier models were selected for the Landsat-8 satellite datasets, but it was observed that some bare soil and built-up land pixels exhibited similar spectral characteristics. This similarity resulted in mixed classes during the LULC mapping process, impacting the accuracy of the RF model [1]. Second, the absence of field data and object data points hindered the proper identification of LULC classes based on the RF models. Specifically, the water body and wasteland classes were not effectively classified using the RF-50 tree model on the GEE platform. It was observed that increasing the number of trees in the RF model could improve land use classification and change detection maps. Future studies could explore the integration of popular remote sensing indices, such as the normalized difference vegetation index (NDVI), normalized water index (NWI), and spectral vegetation index (SVI), to enhance the overall performance of image classification. Considering

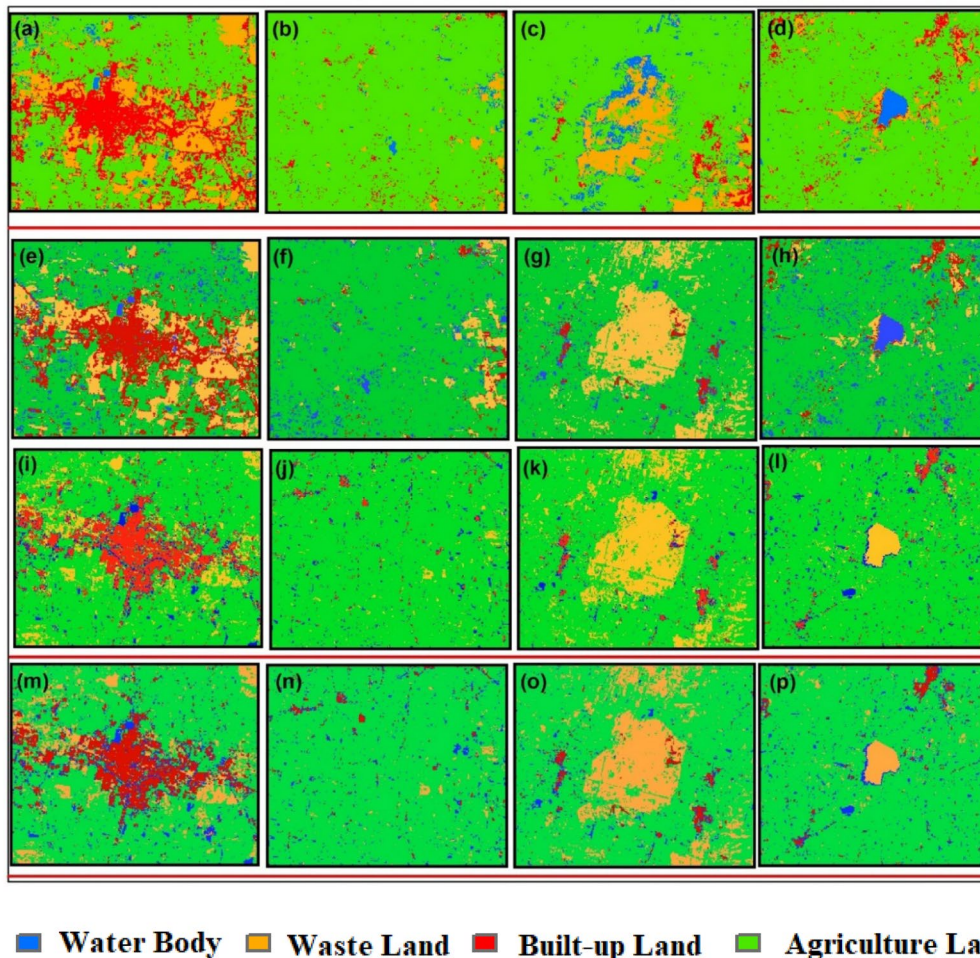


Fig. 12 Land use/land cover (LULC; spatial plot) thematic map for the different types of LULC in the Shirampur region in Maharashtra; **a** built-up land, **b** agricultural land, **c** waste land and **d** water body data shown in 2014 using the RF-50 tree model; **e** built-up land, **f** agricultural land, **g** waste land, and **h** water body data shown in 2020 using the RF-50 tree model; **i** built-up land, **j** agricultural land, **k** waste land, and **l** water body data shown in 2014 using the RF-100 tree model; and **m** built-up land, **n** agricultural land, **o** waste land, and **p** water body data shown in 2020 using the RF-100 tree model

these limitations, future research endeavors will focus on developing and testing more comprehensive and suitable machine learning models based on GEE. This approach aims to incorporate precise land use and land cover mapping methodologies and image classification techniques to enhance accuracy. By leveraging the GEE cloud-based platform and minimizing the reliance on field information, the objective is to generate more precise LULC maps.

Conclusions

This study aimed to analyze the spatiotemporal variation in LULC classes over two distinct analysis periods. Additionally, the performance of the RF machine learning model-based classification algorithm was evaluated with 50 and 100 trees for accurately identifying individual LULC classes. While numerous machine learning

models and classification tools are available, they often produce errors and incorrect maps within short periods. To address this issue, this study developed RF models with 50 and 100 trees using seven input bands as a variables used in the GEE platform. The primary objective was to determine which model performed better when applied to satellite data from 2014 to 2020. By comparing the performances of RF-100 on 2014 and 2020 satellite data and achieving similar accuracies, this model has demonstrated its ability to leverage time series datasets. This capacity allows for the creation of LULC maps spanning 30 years within a minimal timeframe. The findings of this study hold substantial significance for sustainable development, as they enable the efficient monitoring of LULC changes and their associated climate effects on the Earth's land surface. These results contribute to a comprehensive understanding of the dynamic relationships

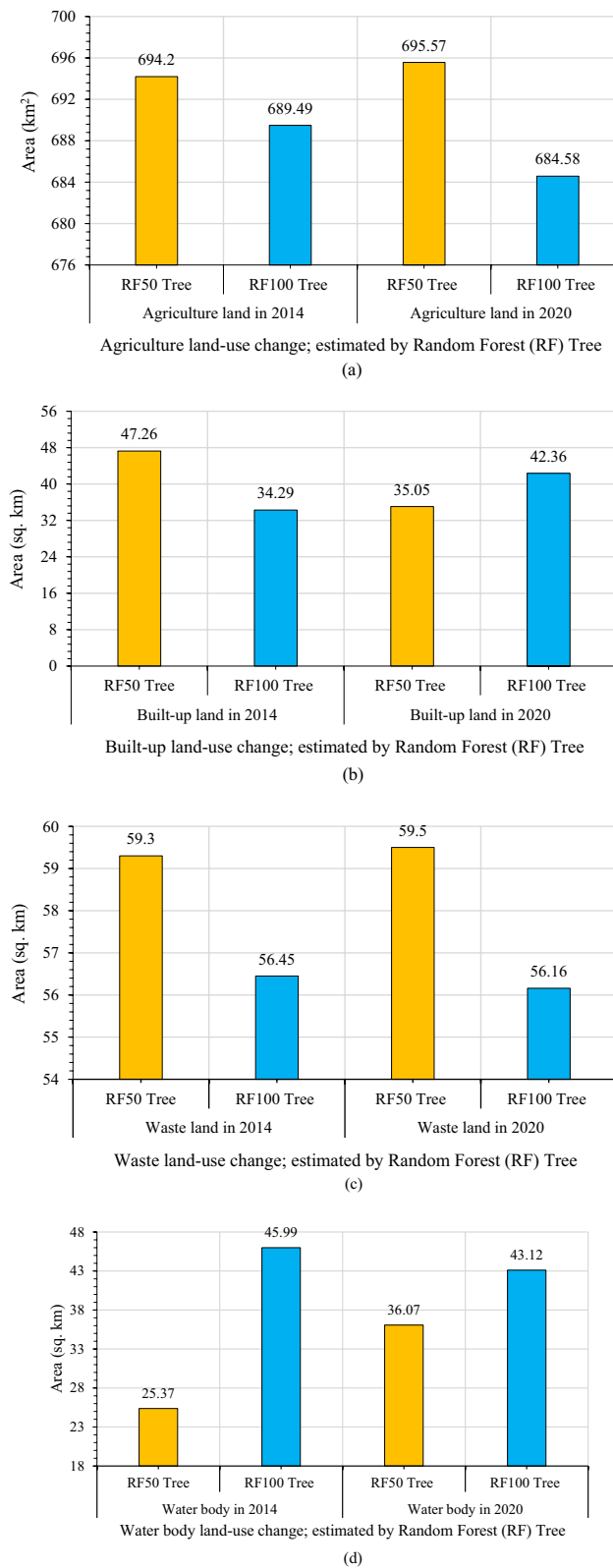


Fig. 13 LUC typewise comparative analysis between RF-50 Tree and RF-100 Tree models for 2014 and 2020 for **a** agricultural land, **b** built-up land, **c** wasteland, and **d** water bodies

among land use, climate, and environmental sustainability. The following conclusions can be drawn from the present study:

1. While the classification accuracy of the RF-50 model demonstrated significant improvement throughout the entire analysis period, reaching 98%, it is important to critically examine specific findings within the results. One noteworthy observation is the seemingly disproportionate 2% increase in waterbodies over a six-year analysis period, which warrants further investigation and consideration. In addition, another notable observation is the lack of variation in the percentage of agricultural land throughout the analysis period. These findings raise questions regarding the stability of agricultural practices and land use patterns within the studied area during the designated timeframe.
2. The RF-100 classification algorithm exhibited exceptional performance during the analysis period, achieving an impressive classification accuracy of 99%. This algorithm effectively captured the transitions between wasteland and built-up land, providing valuable insights into the dynamics of these land cover classes. The study emphasized the robust performance of the RF-100 classification algorithm, as demonstrated by its high accuracy and ability to capture fine-scale transitions, underscoring its suitability for comprehensive LULC mapping applications. These findings provide a strong foundation for future research and potential implementation in large-scale mapping endeavors.
3. The proposed machine learning-based classification algorithms demonstrated improved performance compared to the existing classification approaches. However, slight disagreement was observed between the classifications of bare land and built-up land pixels. To address this issue, the present study urged the incorporation of commonly used remote sensing-based indices, such as the normalized difference vegetation index (NDVI), normalized water index (NWI), and spectral vegetation index (SVI), to enhance the accuracy and performance of image classification. This suggestion aligns with established practices in the field of remote sensing and machine learning, where the integration of spectral indices has proven effective in improving the performance of image classification algorithms. By leveraging these widely used indices, we can leverage additional information and improve the overall robustness and accuracy of our classification model. The utilization of the proposed 50 and 100 trees based on a machine learning model significant improvements in

the identification and segregation of distinct LULC classes. Through rigorous evaluation, we assessed the accuracy of the RF-50 and 100 tree models and determined their effectiveness in accurately identifying LULC classes using the GEE platform and satellite data. This methodology has broader applicability beyond specific study areas, as it can be adopted in various global regions that encompass diverse LULC classes. By incorporating a greater number of training samples, a substantial enhancement in classification accuracy is anticipated. Increasing the training sample size enables machine learning models to learn more comprehensively from a diverse range of examples, leading to improved discrimination and classification of LULC classes.

Acknowledgements

This work was supported by Tenaga Nasional Berhad (TNB) and Universiti Tenaga Nasional (UNITEN) through the BOLD Refresh Postdoctoral Fellowships under the project code of J510050002-IC-6 BOLDREFRESH2025-Centre of Excellence. The authors would like to express sincere gratitude to Al Maarefa University, Riyadh, Saudi Arabia, for supporting this research. The authors extend their appreciation to Abdullah Alrushaid Chair for Earth Science Remote Sensing Research for funding.

Author contributions

Chaitanya Baliram Pande: Conceptualization, Idea of topic, Writing—original draft, Supervision, Formal Analysis, developed algorithms, Data Interpretations and satellite data processing, Development of Methodology, Software, Analyzing and Drafting results and analysis, Statistical Analysis, Field data collection, Random Forest Model development, Creation and Interactive visualization of figures and methodology chart, Writing—review and editing., **Aman Srivastava:** Analyzing and Drafting results and discussions, Writing—review and editing, creation and Interactive visualization of charts of results, **Kanak N. Moharir:** Formal Analysis, Investigation, Field data collection, Investigation, Survey, and Writing—review and editing, **Neyara Radwan:** Writing review and editing, Formal analysis and Investigation, **Fahad Alshehri:** Writing—original draft, Formal analysis, Writing—review and editing, **Lariyah Mohd Sidek:** Supervision, Writing—original draft, Writing—review & editing, **Subodh Chandra Pal:** Writing—original draft, Writing—review & editing, **Abebe Debele Tolche:** Writing—review & editing, formal analysis, Investigation. **Mohamed Zhran:** Writing—original draft, Writing—review and editing.

Funding

Not applicable.

Data availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Declarations

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

Author details

¹Institute of Energy Infrastructure, Universiti Tenaga Nasional, 43000 Kajang, Malaysia. ²New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Nasiriyah 64001, Iraq. ³Department of Civil Engineering, Indian Institute of Technology (IIT) Kharagpur, Kharagpur 721302, West Bengal, India. ⁴Department of Remote Sensing,

Banasthali University, Jaipur, India. ⁵Abdullah Alrushaid Chair for Earth Science Remote Sensing Research, Geology and Geophysics Department, King Saud University, 11451 Riyadh, Saudi Arabia. ⁶Department of Geography, The University of Burdwan, Purbas, Bardhaman, West Bengal 713104, India. ⁷School of Water Resources and Environmental Engineering, Haramaya Institute of Technology, Haramaya University, Dire Dawa, Ethiopia. ⁸Industrial Engineering Department, College of Applied Sciences, Al Maarefa University, 71666, 13713 Diriyah, Saudi Arabia. ⁹Mechanical Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt. ¹⁰Public Works Engineering Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt.

Received: 9 November 2023 Accepted: 3 April 2024

Published online: 24 April 2024

References

- Pande CB (2022) Land use/land cover and change detection mapping in Rahuri watershed area (MS), India using the google earth engine and machine learning approach. *Geocarto Int* 37(26):13860–13880. <https://doi.org/10.1080/10106049.2022.2086622>
- Wen Z, Wang Q, Ma Y, Jacinthe PA, Liu G, Li S et al (2024) Remote estimates of suspended particulate matter in global lakes using machine learning models. *Int Soil Water Conserv Res* 12(1):200–216. <https://doi.org/10.1016/j.iswcr.2023.07.002>
- Huang J, Ma H, Sedano F, Lewis P, Liang S, Wu Q et al (2019) Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST-PROSAIL model. *Eur J Agron* 102:1–13. <https://doi.org/10.1016/j.eja.2018.10.008>
- Yuan F, Sawaya KE, Loeffelholz B, Bauer ME (2005) Land cover classification and change analysis of the twin cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. *Remote Sens Environ* 98:317–328
- Srivastava A, Chinnasamy P (2021) Investigating impact of land-use and land cover changes on hydroecological balance using GIS: insights from IIT Bombay. *India SN Appl Sci* 3(3):343. <https://doi.org/10.1007/s42452-021-04328-7>
- Harmay NSM et al (2021) Urban heat Island associated with land use/land cover and climate variations in Melbourne. *Aust Sustainable Cities Soc* 69(6):102861
- Huang J, Gómez-Dans JL, Huang H, Ma H, Wu Q, Lewis PE et al (2019) Assimilation of remote sensing into crop growth models: current status and perspectives. *Agric For Meteorol* 276–277. <https://doi.org/10.1016/j.agrformet.2019.06.008>
- Jiang C, Wang Y, Yang Z, Zhao Y (2023) Do adaptive policy adjustments deliver ecosystem-agriculture-economy co-benefits in land degradation neutrality efforts? Evidence from southeast coast of China. *Environ Monit Assess* 195(10):1215. <https://doi.org/10.1007/s10661-023-11821-6>
- Hu Q, Zhao Y, Hu X, Qi J, Suo L, Pan Y et al (2022) Effect of saline land reclamation by constructing the “Raised Field -Shallow Trench” pattern on agroecosystems in Yellow River Delta. *Agric Water Manag* 261:107345. <https://doi.org/10.1016/j.agwat.2021.107345>
- Sekertekin A, Marangoz AM, Akcin H (2017) Pixel based classification analysis of land use land cover using sentinel-2 and landsat-8 data. *Int Arch Photogramm Remote Sens Spatial Inf Sci*. <https://doi.org/10.5194/isprs-archives-XLII-4-W6-91-2017>
- Wu T, Luo J, Fang J, Ma J, Song X (2018) Unsupervised object-based change detection via a Weibull mixture model-based binarization for high-resolution remote sensing images. *IEEE Geosci Remote Sens Lett* 15:63–67
- Yin L, Wang L, Li T, Lu S, Yin Z, Liu X et al (2023) U-Net-STN: a Novel end-to-end Lake boundary prediction model. *Land* 12(8):1602. <https://doi.org/10.3390/land12081602>
- Yin L, Wang L, Li T, Lu S, Tian J, Yin Z et al (2023) U-Net-LSTM: time series-enhanced lake boundary prediction model. *Land* 12(10):1859. <https://doi.org/10.3390/land12101859>
- Shen J, Sheng H, Wang S, Cong R, Yang D et al (2023) Blockchain-based distributed multi-agent reinforcement learning for collaborative multi-object tracking framework. *IEEE Trans Comput*. <https://doi.org/10.1109/TC.2023.3343102>
- Cui Z, Sheng H, Yang D, Wang S, Chen R et al (2024) Light field depth estimation for non-lambertian objects via adaptive cross operator. *IEEE Trans Circuits and Sys Video Technol* 34(2):1199–1211. <https://doi.org/10.1109/TCSVT.2023.3292884>
- Pande Chaitanya B, Kanak N, Moharir SFR (2018) Study of Land Use Classification in the Arid Region using Multispectral Satellite Images. *Appl Water Sci* 8(5):1–11
- Du C, Bai X, Li Y, Tan Q, Zhao C, Luo G et al (2024) Storage, form, and influencing factors of karst inorganic carbon in a carbonate area in China. *Sci China Earth Sci*. <https://doi.org/10.1007/s11430-023-1249-9>
- He M, Dong J, Jin Z, Liu C, Xiao J, Zhang F et al (2021) Pedogenic processes in loess-paleosol sediments: clues from Li isotopes of leachate in Luochuan loess. *Geochimica et Cosmochimica Acta* 299:151–162. <https://doi.org/10.1016/j.gca.2021.02.021>
- Srivastava A, Chinnasamy P (2023) Watershed development interventions for rural water safety, security, and sustainability in semiarid region of Western-India. *Environ Dev Sustain*. <https://doi.org/10.1007/s10668-023-03387-7>
- Read JM, Lam NSN (2002) Spatial methods for characterizing land cover and detecting land-cover changes for the tropics. *Int J Remote Sens* 23(12):2457–2474
- Qiu S, Yang H, Zhang S, Huang S, Zhao S, Xu X et al (2023) Carbon storage in an arable soil combining field measurements, aggregate turnover modeling and climate scenarios. *CATENA* 220:106708. <https://doi.org/10.1016/j.catena.2022.106708>
- Xu Z, Li X, Li J, Xue Y, Jiang S, Liu L et al (2022) Characteristics of source rocks and genetic origins of natural gas in deep formations, gudian depression, Songliao Basin, NE China. *ACS Earth Space Chem* 6(7):1750–1771. <https://doi.org/10.1021/acsearthspacechem.2c00065>
- Rai SC, Sharma E, Sundriyal RC (1994) Conservation in the Sikkim Himalaya: traditional knowledge and land-use of the Mamlay watershed. *Environ Conserv* 21(1):30–34
- Du W, Wang G (2013) Intra-event spatial correlations for cumulative absolute velocity, Arias intensity, and spectral accelerations based on regional site conditions. *Bull Seismol Soc Am* 103(2A):1117–1129. <https://doi.org/10.1785/0120120185>
- Du W, Wang G (2014) Fully probabilistic seismic displacement analysis of spatially distributed slopes using spatially correlated vector intensity measures. *Earthq Eng Struct Dyn* 43(5):661–679. <https://doi.org/10.1002/eqe.2365>
- Sleeter BM, Sohl TL, Loveland TR, Auch RF, Acevedo W, Drummond MA, Saylor KL, Stehman SV (2013) Land-cover change in the conterminous United States from 1973 to 2000. *Glob Environ Chang* 23:733–748
- Dash SS, Sena DR, Mandal U, Kumar A, Kumar G, Mishra PK, Rawat M (2021) A hydrological modeling-based approach for vulnerable area identification under changing climate scenarios. *J Water Climate Change* 12(2):433–452
- Voogt JA, Oke TR (2003) Thermal remote sensing of urban climates. *Remote Sens Environ* 86(3):370–384. [https://doi.org/10.1016/s0034-4257\(03\)00079-8](https://doi.org/10.1016/s0034-4257(03)00079-8)
- Lambin EF (1999) Monitoring forest degradation in tropical regions by remote sensing: some methodological issues. *Global Ecol Biogeography* 8(3–4):191–198
- Gashaw T, Tulu T, Argaw M, Worqlul AW (2018) Modeling the hydrological impacts of land use/land cover changes in the Andassa watershed, Blue Nile basin. *Ethiopia Sci Total Environ* 619(620):1394–1408
- Zhou G, Xu C, Zhang H, Zhou X, Zhao D, Wu G et al (2022) PMT gain self-adjustment system for high-accuracy echo signal detection. *Int J Remote Sens* 43(19–24):7213–7235. <https://doi.org/10.1080/01431161.2022.2155089>
- Zhou G, Zhou X, Li W, Zhao D, Song B, Xu C et al (2022) Development of a lightweight single-band bathymetric LiDAR. *Remote Sens* 14(22):5880. <https://doi.org/10.3390/rs14225880>
- Park BSK (2021) Temperature trend analysis associated with land-cover changes using time-series data (1980–2019) from 38 weather stations in South Korea. *Sustain Cities Soc* 65(2):102615
- Zhou G, Zhang H, Xu C, Zhou X, Liu Z, Zhao D et al (2023) A real-time data acquisition system for single-band bathymetric LiDAR. *IEEE Trans Geosci Remote Sens* 61. <https://doi.org/10.1109/TGRS.2023.3282624>
- Pande CB, Moharir KN (2023) Application of hyperspectral remote sensing role in precision farming and sustainable agriculture under climate

- change: a review. In: *Climate change impacts on natural resources, ecosystems and agricultural systems*. Springer Climate. Springer, Cham. https://doi.org/10.1007/978-3-031-19059-9_21
36. Pande CB, Moharir KN, Varade AM, Abdo HM, Mulla S (2023) Intertwined impacts of urbanization and land cover change on urban climate and agriculture in Aurangabad city (MS), India using google earth engine platform. *J Clean Prod* 422:138541. <https://doi.org/10.1016/j.jclepro.2023.138541>
 37. Dhanuka A, Srivastava A, Khadke L, Kushwaha NL (2023) Smart geometric design of sing HTML programming for sustainable and climate resilient cities. In: Chatterjee U, Bandyopadhyay N, Setiawati MD, Sarkar S (eds) *Urban Commons, Future Smart Cities and Sustainability*. Springer Geography, Cham
 38. Mohammad RR (2021) Forecasting land-use changes in Mashhad Metropolitan area using cellular automata and markov chain model for 2016–2030. *Sustain Cities Soc* 64(1):102548
 39. Pande CB (2020) Thematic Mapping for Watershed Development. In: *Sustainable watershed development*. SpringerBriefs in Water Science and Technology. Springer, Cham. https://doi.org/10.1007/978-3-030-47244-3_3
 40. Pande CB, Moharir KN, Singh SK, Varade AM, Elbeltagi A, Khadri SFR, Choudhari P (2021) Estimation of crop and forest biomass resources in a semi-arid region using satellite data and GIS. *J Saudi Soc Agric Sci* 20(5):302–311. <https://doi.org/10.1016/j.jssas.2021.03.002>
 41. Ozesmi SL, Bauer ME (2002) Satellite remote sensing of wetlands. *Wetl Ecol Manage* 10:381–402
 42. Punia M, Joshi PK, Porwal MC (2011) Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data. *Expert Syst Appl* 38(5):5577–5583
 43. Rawat JS, Kumar M (2015) Monitoring land use/cover change using remote sensing and GIS techniques: a case study of Hawalbagh block, district Almora, Uttarakhand, India. *Egypt J Remote Sens Space Sci* 18:77–84
 44. Lu D, Mauseil P, Brondizio E, Moran E (2004) Change detection techniques. *Int J Remote Sens* 25:2365–2407
 45. Boori MS, Voženílek V, Choudhary K (2015) Land use/cover disturbance due to tourism in Jeseníky Mountain, Czech Republic: a remote sensing and GIS based approach. *Egypt J Remote Sens Space Sci* 18:17–26
 46. Tepe E, Safikhani A (2023) Spatiotemporal modeling of parcel-level land-use changes using machine learning methods. *Sustain Cities Soc* 90(3):104390
 47. Sobhani P et al (2021) Simulation and impact assessment of future land use and land cover changes in two protected areas in Tehran. *Iran Sustainable Cities Society* 75(12):103296
 48. AlDousari AE et al (2022) Modeling the impacts of land use/land cover changing pattern on urban thermal characteristics in Kuwait. *Sustain Cities Soc* 86(11):104107
 49. Elbeltagi A, Srivastava A, Kushwaha NL, Juhász C, Tamás J, Nagy A (2023) Meteorological data fusion approach for modeling crop water productivity based on ensemble machine learning. *Water* 15(1):30. <https://doi.org/10.3390/w15010030>
 50. Elbeltagi A, Srivastava A, Al-Saeedi AH, Raza A, Abd-Elaty I, El-Rawy M (2023) Forecasting long-series daily reference evapotranspiration based on best subset regression and machine learning in Egypt. *Water* 15(6):1149. <https://doi.org/10.3390/w15061149>
 51. Elbeltagi A, Srivastava A, Deng J, Li Z, Raza A, Khadke L, Yu Z, El-Rawy M (2023) Forecasting vapor pressure deficit for agricultural water management using machine learning in semiarid environments. *Agric Water Manag* 283:108302. <https://doi.org/10.1016/j.agwat.2023.108302>
 52. Pande CB, Al-Ansari N, Kushwaha NL, Srivastava A, Noor R, Kumar M et al (2022) Forecasting of SPI and meteorological drought based on the artificial neural network and M5P model tree. *Land* 11(11):2040. <https://doi.org/10.3390/land11112040>
 53. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens Environ* 202:18–27
 54. Tamiminia H, Salehi B, Mahdianpari M, Quackenbush L, Adeli S, Brisco B (2020) Google Earth Engine for geo-big data applications: a meta-analysis and systematic review. *ISPRS J Photogramm Remote Sens* 164:152–170
 55. Rodriguez-Galiano VF, Chica-Olmo M, Abarca-Hernandez F, Atkinson PM (2012) Random forest classification of Mediterranean land cover using multiseasonal imagery and multiseasonal texture. *Remote Sens Environ* 121:93–107
 56. Belgiu M, Dragut L (2016) Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* 114:24–31
 57. Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: A review. *ISPRS J Photogramm Remote Sens* 66:247–259
 58. Naboureh A, Ebrahimi H, Azadbakht M, Bian J, Amani M (2020) An ensemble method to handle the class imbalance problem in land cover mapping using google earth engine. *Remote Sens* 12:3484
 59. Petropoulos GP, Arvanitis K, Sigrimis N (2012) Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Exp. Syst Appl* 39(2012):3800–3809
 60. Pekel J-F, Cottam A, Gorelick N, Belward AS (2016) High-resolution mapping of global surface water and its long-term changes. *Nature* 540:418
 61. Forstmaier A, Shekhar A, Chen J (2020) Mapping of eucalyptus in natura (2000) areas using sentinel 2 imagery and artificial neural networks. *Remote Sens* 12:2176
 62. Srivastava A, Chinnasamy P (2021) Water management using traditional tank cascade systems: a case study of semiarid region of Southern India. *SN Appl Sci* 3:281. <https://doi.org/10.1007/s42452-021-04232-0>
 63. Chinnasamy P, Srivastava A (2021) Revival of traditional cascade tanks for achieving climate resilience in drylands of South India. *Frontiers Water* 3:639637. <https://doi.org/10.3389/frwa.2021.639637>
 64. Rahaman MH, Masroor RM, Sajjad H (2023) Integrating remote sensing derived indices and machine learning algorithms for precise extraction of small surface water bodies in the lower Thoubal river watershed India. *J Clean Product* 422:138563. <https://doi.org/10.1016/j.jclepro.2023.138563>
 65. Aryal J, Sitaula C, Frery AC (2023) Land use and land cover (LULC) performance modeling using machine learning algorithms: a case study of the city of Melbourne. *Australia Sci Rep* 13:13510. <https://doi.org/10.1038/s41598-023-40564-0>
 66. Tassi A, Vizzari M (2020) Object-oriented LULC classification in google earth engine combining Snic, Glcm, and machine learning algorithms. *Remote Sensing* 12(22):3776. <https://doi.org/10.3390/rs12223776>
 67. Zhao Y et al (2022) Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period. *Ecol Ind* 135:108529
 68. Bui DH, Muçi L (2021) From land cover map to land use map: a combined pixel-based and object-based approach using multitemporal landsat data, a random forest classifier, and decision rules. *Remote Sens* 13:1700
 69. Yangchengsi Z et al (2022) Long-term detection and spatiotemporal variation analysis of open-surface water bodies in the Yellow River Basin from 1986 to 2020. *Sci Total Environ* 845(1):157152. <https://doi.org/10.1016/j.scitotenv.2022.157152>
 70. Pande B et al (2024) Impact of land use/land cover changes on evapotranspiration and model accuracy using google earth engine and classification and regression tree modeling. *Geomat Nat Haz Risk* 15(1):1–29
 71. Kumar L, Mutanga O (2018) Google earth engine applications since inception: usage, trends, and potential. *Remote Sens* 10:1509
 72. Santhosh Kumar B, Rajesh J, Pande CB, Varade A (2023) Land use and cover variations and problems associated with coastal climate in a part of southern tamil nadu, india, using remote sensing and GIS approach. In: Pande CB, Moharir KN, Singh SK, Pham QB, Elbeltagi A (eds) *Climate change impacts on natural resources, ecosystems and agricultural systems*. Springer Climate, Springer, Cham
 73. Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
 74. Dong J, Xiao X, Menarguez MA, Zhang G, Qin Y, Thau D, Biradar C, Moore B (2016) Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sens Environ* 185:142–154
 75. Pande CB, Moharir KN, Khadri SFR (2021) Assessment of land-use and land-cover changes in Pangari watershed area (MS), India, based on the remote sensing and GIS techniques. *Appl Water Sci* 11:96. <https://doi.org/10.1007/s13201-021-01425-1>
 76. Breiman L (1996) *Bagging Predictors* (Vol. 24). Berkeley: Kluwer Academic Publishers. Accessed <https://link.springer.com/content/pdf/10.1023%2FA%3A1018054314350.pdf>.

77. Foody GM, Arora MK (1997) An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *Int J Remote Sens* 18:799–810. <https://doi.org/10.1080/014311697218764>
78. Huang H, Chen Y, Clinton N, Wang J, Wang X, Liu C, Gong P, Yang J, Bai Y, Zheng Y et al (2017) Mapping major land cover dynamics in Beijing using all landsat images in google earth engine. *Remote Sens Environ* 202:166–176
79. Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. *Pattern Recogn Lett* 27(4):294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
80. Gulhane VA, Rode SV et al (2022) Correlation analysis of soil nutrients and prediction model through iso cluster unsupervised classification with multispectral data. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-022-13276-2>
81. Zhang M, Huang H, Li Z, Hackman KO, Liu C, Andriamiarisoa RL, Rahevivo TNAN, Li Y, Gong P (2020) Automatic high-resolution land cover production in madagascar using sentinel-2 time series, tile-based image classification and google earth engine. *Remote Sens* 12:3663
82. Pande C, Moharir K (2014) Analysis of land use/land cover changes using remote sensing data and GIS techniques of Patur Taluka, Maharashtra, India. *Int J Pure Appl Res Eng Technol* 2(12):85–92
83. Orimoloye IR, Olusola AO, Belle JA et al (2022) Drought disaster monitoring and land use dynamics: identification of drought drivers using regression-based algorithms. *Nat Hazards* 112:1085–1106. <https://doi.org/10.1007/s11069-022-05219-9>
84. Kandekar VU, Pande CB, Rajesh J et al (2021) Surface water dynamics analysis based on sentinel imagery and google earth engine platform: a case study of Jayakwadi dam. *Sustain Water Resour Manag* 7:44. <https://doi.org/10.1007/s40899-021-00527-7>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.