Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 24 | No. 4 | Article ID B4308 | 2025

Original Research Paper

https://doi.org/10.46488/NEPT.2025.v24i04.B4308

Open Access Journal

Comparative Analysis of CART and Random Forest Classifiers for LULC Mapping: A Case Study of Brahmani-Baitarani River Basin, India

Sonali Kadam[1], Sangram Patil[2]†, Kavita Sawant[1], Sae Jamdade[1], Apurva Gadilkar[1], Chahal Ohri[1], Namrata Rathi[1] and Jotiram Gujar[3]

[1]Department of Computer Engineering, Bharati Vidyapeeth's College of Engineering for Women, Pune, Maharashtra, India

[2]Department of Civil Engineering, Bharati Vidyapeeth's College of Engineering, Lavale, Pune, Maharashtra, India

[3]Department of Chemical Engineering, Sinhgad College of Engineering, Pune, Maharashtra, India

†Corresponding author: Sangram Patil; patil.sangram@bharatividyapeeth.edu

Abbreviation: Nat. Env. & Poll. Technol.
Website: www.neptjournal.com

Received: 29-01-2025
Revised: 10-03-2025
Accepted: 13-03-2025

Key Words:

Remote sensing imagery
NDVI
NDWI
Supervised classification
Environmental monitoring

Citation for the Paper:

Kadam, S., Patil, S., Sawant, K., Jamdade, S., Gadilkar, A., Ohri, C., Rathi, N. and Gujar, J., 2025. Comparative analysis of CART and random forest classifiers for LULC mapping: A case study of Brahmani-Baitarani River Basin, India. *Nature Environment and Pollution Technology*, 24(4), B4308. https://doi.org/10.46488/NEPT.2025.v24i04.B4308

Note: From 2025, the journal has adopted the use of Article IDs in citations instead of traditional consecutive page numbers. Each article is now given individual page ranges starting from page 1.

ABSTRACT

Land Use and Land Cover (LULC) classification is essential for monitoring environmental changes, managing resources, and planning sustainable development. However, accurate classification remains challenging because of the diversity of landscapes and the computational demands of processing large datasets. Among various machine learning (ML) algorithms, such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Random Forest (RF), and Classification and Regression Trees (CART), RF and CART were chosen for this study because of their robustness, simplicity, and efficiency in handling complex LULC classification tasks. This study focuses on the Brahmani-Baitarani River Basin, a region known for its environmental significance and susceptibility to land-use changes. Using remote sensing data from Landsat 8, Landsat 9, and Sentinel-2 satellites, a comparative analysis of RF and CART was conducted to evaluate their LULC mapping performance. The datasets were processed and analyzed on the Google Earth Engine (GEE) platform using multi-temporal image data and advanced filtering techniques. The results revealed that RF consistently delivered higher classification accuracy than CART, making it a reliable choice for LULC studies in dynamic and heterogeneous landscapes. By integrating high-resolution satellite imagery with ML algorithms, this study provided detailed insights into the spatial distribution of land use across the Brahmani-Baitarani Basin. These findings have practical applications in urban planning, natural resource management, and environmental conservation, and offer valuable information for decision-makers and researchers working to address global environmental challenges.

INTRODUCTION

Land Use and Land Cover (LULC) is the manner in which humans use land and other features of the physical landscape. They form the basis for analyzing human socio-economic activities and their effects on Earth regarding the physical planning of towns, cities, and general natural resource management. LULC alterations have an impact on ecosystems, climate, and resources (Rong & Fu 2023). For example, an increase in the frequency of global flooding events changes LULC (Kadam et al. 2024). These systems are critical for standardizing, categorizing, and differentiating between the numerous forms of land that exist; as such, they are considered important tools of measurement in the context of environmental studies (Nedd et al. 2021). Some of the uses of mapping and spatial data are important in evaluating the land observations. Therefore, it is still difficult to properly assess classification systems, while their ability to follow the changes in the proposed land frequency remains high. Some of these drawbacks affect environmental monitoring, functioning, and the planning of environmental activities.

The Brahmani-Baitarani River Basin was chosen because of its distinct environmental problems and high variability of LULC, making it an important area for land cover classification research. The river ba-sin is subject to frequent floods, intensive land use change, and varied topography, mak-ing the area imperative for precise LULC mapping. The hydrological complexity of the ba-sin also creates difficulties in the study of land dynamics, necessitating strong classification methods. In addi-tion, efficient LULC analysis is crucial for flood risk management, agricultural planning, and ecologi-cal protection in this area.

Random Forest (RF) and Classification and Regression Trees (CART) have been extensively compared in LULC classification, but this study focuses on regional-specific analysis in a complex river basin while examining various sampling methods, including random sampling and stratified sampling. This study deviates from previous work by assessing how the algorithms succeed in handling uneven geographic terrains and sensitive hydrological areas. This research contributes fresh information about RF and CART performance in detecting seasonal and environmental impacts on land cover classification, despite limited investigation in past studies.

Satellite remote sensing (RS) is a central source of data essential for studying and mapping the Earth's surface. The availability and growing number of Remote Sensing data supported by improved and cheaper satellites, including radiometric, spectral, spatial, and temporal resolutions, allow users to work with large amounts of time-series data (Tassi &Vizzari 2020). However, this allows for greater accessibility owing to the ability to break down components and perform separate computational requests simultaneously; however, this comes at the cost of increased computational complexity and time. To help overcome these challenges, Google has developed one of the largest platforms based on the cloud, called the Google Earth Engine, which has demonstrated considerable results in recent years (Feng et al. 2022). Another is the online geospatial platform, which provides the utilization of a large amount of geospatial data as well as a set of powerful online tools for computational and visual analysis. Researchers can either preprocess or directly download multi-temporal image data that satisfy certain filtering conditions through the GEE platform (Velastegui-Montoya et al.2023). Subsequently, they can engage several machine learning algorithms to perform LULC classification and analysis online (Loukika et al. 2023).

The use of ML algorithms on remote sensing imagery for LULC classification has recently garnered significant interest. Unlike the limitations of human deciphering, AI techniques can effectively identify subtle patterns(Mahajan et al. 2024a, 2024b). ML techniques are now widely used in different sectors, such as remote sensing technology and smart agriculture. Supervised techniques encompass various algorithms, including Classification and Regression Trees (CART), support vector machines (SVM), Spectral Angle Mapper (SAM), Fuzzy Adaptive Resonance Theory-Supervised Predictive Mapping (Fuzzy ARTMAP), Random Forest (RF), Mahalanobis Distance (MD), Radial Basis Function (RBF), Decision Tree (DT), Multilayer Perceptron (MLP), Maximum Likelihood Classifier (MLC), Naive Bayes (NB), and Fuzzy Logic (Maxwell 2018).

Some of the machine learning methods are more accurate than others in classifying LULC. According to a review of previous studies, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and decision trees have the potential to conduct classification techniques. The results show that Random Forest (RF) performs better in general than other classification methods. Among all the analyzed machine learning approaches, RF and CART were identified as the best-performing algorithms for LULC. Classification that offers much more accuracy than those of other researchers (Carranza-García et al. 2019).

Many studies have been devoted to comparing various approaches to classification using machine learning techniques. These investigations are useful for providing information on the drawbacks and advantages of the approach in terms of the appearance of LULC maps. In the proposed research, the authors aim to conduct a comparative study focusing on the performance of RF and CART using two distinct datasets: Dynamic WorldCover and the European Space Agency (ESA). By evaluating the success of these techniques, we intend to help formulate future research on the use of LULC mapping.

MATERIALS AND METHODS

Study Area

The Brahmani Baitarani area, which is shown in Fig. 1, is located in eastern Odisha and between 83°55′ and 87°30′ east longitudes and 20°28′ and 23°38′ north latitudes. It starts from the highlands of Jharkhand and flows through Odisha to the Bay of Bengal. This can be written as the length of the river = 799 km and 355 km (Indraja et al. 2024). The Baitarani River is one of the six major rivers in Odisha, with an elevation range of 32 to 1024 m.

The major part of the region is covered with agricultural areas where the monsoon starts in June and extends till October. The southwest monsoon season (June to September) accounts for about 80% of the year-ly rainfall. The Baitarni sub-basin receives 1250-1500 mm, and the Brahmani sub-

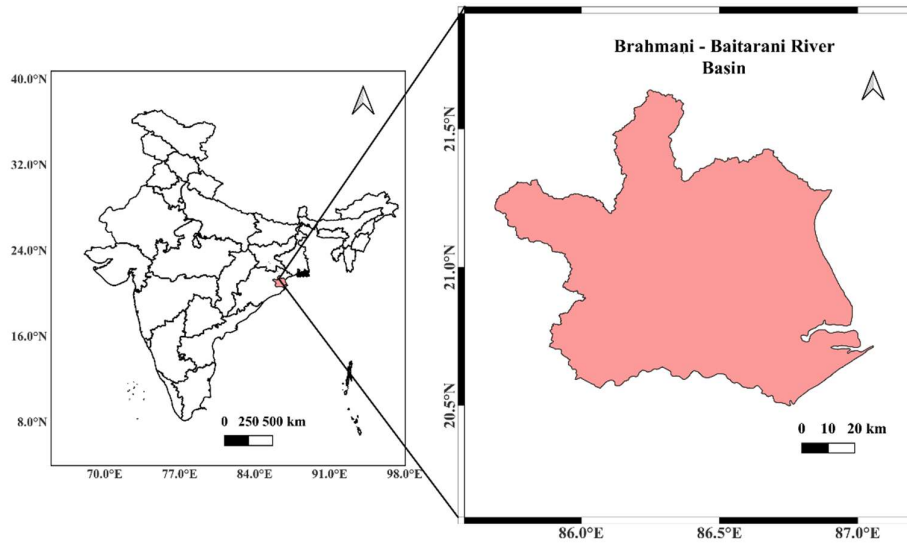


Fig. 1: Study Area.

basin receives 1250-1750 mm. Mean temperatures are 32°C (max) and 20°C (min). Recent land use and climate changes have led to frequent flooding, damaging infrastructure and agriculture (Swain et al. 2021).

Satellite Data

Landsat-8, launched in 2013, provides medium-resolution imagery (30 m) for land monitoring, urban mapping, and vegetation analysis. Landsat-9, launched in 2021, continues this mission with advanced technology, maintaining a 12-day revisit period and providing essential data for continued land cover monitoring. Sentinel-2 L1C images, starting from June 27, 2015, with a 2–5-day revisit period, are also pivotal in this process.

Dynamic World Cover and LULC Data. Dynamic World Cover offers near-real-time LULC data with a 10-meter resolution, categorizing land into nine classes with probabilities (Brown et al. 2022). Using the Cloud Displacement Index, Directional Distance Transform, and S2 Cloud Probability, less than 35% cloud masking was achieved. This project, inspired by the 2017 World Cover conference and launched by the ESA, produced a global 10-meter resolution land cover map incorporating data from both Sentinel-1 and Sentinel-2 satellites, focusing on 11 land cover types with over 75% accuracy (Zanaga et al. 2021, 2022).

Methodology

The methodologies employed in this study are described in detail in the subsequent sections. This study aids in classifying the classification made by two Machine

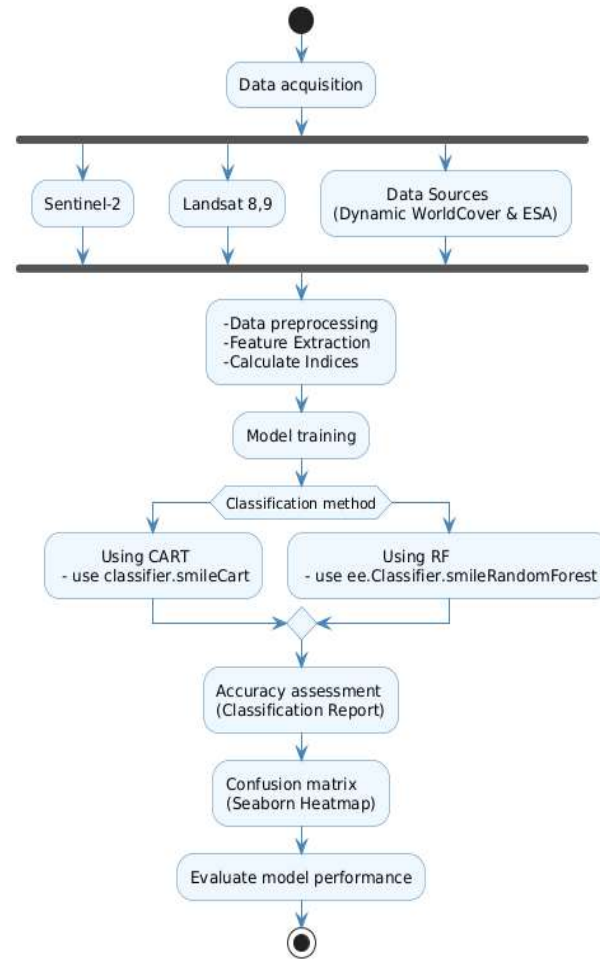


Fig. 2: Methodology.

Learning algorithms, random forest and CART, in Fig. 2.

Data Acquisition

The Sentinel 2 and Landsat data were used to classify Land Use and Land Cover in the regions of Brahamani and Baitarani. The underlying data were obtained from three different agencies to understand the differences in classification and datasets. Sentinel 2 data were collected from Dynamic WorldCover and the European Space Agency. The data consist of multispectral images, which undergo preprocessing techniques to remove unnecessary noise or artifacts, cloud masking, or any geometric or atmospheric correction. We extracted key spectral, textural, and contextual features from the segmented regions. As part of this process, various indices are calculated, such as the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-up Index (NDBI). The derived parameters can further help in understanding the specific land cover attributes (Zhao et al. 2024).

Pre-processing, Feature Extraction and Calculating Indices

Preprocessing involves many procedures to improve data quality before classification. First, Sentinel-2 and Landsat imagery were selected and are cropped based on a predefined shapefile for our region of interest (ROI). Cloud masking is the most important aspect of satellite data. Cloud masking is removed, which removes unwanted pixels. In this way, the accuracy of land cover classification can be substantially enhanced by removing harmful factors such as unreliable pixels caused by both clouds and tree foliage cover. It also guarantees that all spectral indices, such as the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Burn Ratio (NBR), are computed from correctly classified land surfaces and noise (clouds or cloud shadows) is excluded (Mateo-García et al. 2018).

The data is then further improved using geospatial formats, applying transformations, and exporting images for analysis. These methods collectively make LULC classification results more reliable by deleting atmospheric distortions and unwanted features from the images.

$$NDVI = (NIR - RED)/(NIR + RED)$$

$$NDWI = (GREEN - NIR)/(GREEN + NIR)$$

$$NBR = (NIR - SWIR)/(NIR + SWIR)$$

The vegetation index NDVI serves as a common tool that measures plant greenness through chlorophyll levels. The water content of a specific area was measured using NDWI (Ashok et al.2021). Distinct features within a 3-band satellite image of a basin were identified using the NDVI technique. NDVI evaluates vegetative cover by evaluating wavelengths in the near-infrared versus red bands, which produces values between -1 and 1(Gebeyehu et al. 2019). Vegetation indices derived from these satellite images have been used to assess vegetation cover, which is a biophysical indicator of soil erosion. The formula used combines the visible Red Band (RED) and near-infrared reflectance (NIR) to determine NDVI.

It is applied by calculating the difference between green and near-infrared reflectance using the NDWI technique to determine water bodies from within a 3-band satellite image. Areas with high moisture content are highlighted against the context of land features in this index. Likewise, the NBR index is used to monitor burned regions and vegetation health by comparing short-wave and near-infrared reflectance. The two indices derived from Sentinel 2 data improved the classification of water bodies and burned land, improving the accuracy of LULC analysis and environmental monitoring (Tempa et al. 2024).

Model Training

In this study, the training process began with the selection of training datasets derived from Sentinel-2 and Landsat 8/9 satellite imagery. Prior to utilization, the datasets underwent preprocessing to achieve quality and consistency standards. During model training and testing operations, the dataset is divided into two separate parts: training and testing samples. A random sampling and stratified sampling approach ensured this split. A distributed random value appears in each outcome column for every instance in the dataset. The distribution of data between training and testing emerges from the split ratio value, which functions as a floating-point number between 0 and 1. The split ratio establishes the specific portion of the data that contributes to the training process.

When preprocessing is complete, the system creates labeled training samples that identify unique LULC classes consisting of vegetation, water bodies, and constructed regions. The RF and CART systems used these selected samples to establish their models.

RF implements an ensemble approach, using numerous decision trees that combine predictions to drive better classification outcomes. CART implements a single decision tree approach that divides the dataset recursively based on the threshold of features to identify LULC types. Both classifiers leveraged the trained data of the extracted spectral and textural features alongside contextual features to optimally classify the different land cover types present in the study region.

Classification

CART: CART operates as a binary decision tree classifier, creating simple, logical if-then decisions. Research shows that people commonly use CART for remote sensing tasks, particularly land use and land cover mapping, as well as vegetation identification and land use update monitoring. Other ML algorithms often take time to normalize data, but CART does not require normalization. It evaluates input variables to determine which provides the high-est information gain, thus guiding node splits at each level (Shetty 2019). CART's robustness to noisy data and its ability to cope with split-outliers make the algorithm useful in a wider range of fields, as this algorithm adapts to some imperfections in data by adjusting splits to handle variance. This method is suitable for our study because high-resolution datasets are typically large, but the Cart soft-ware efficiently handles multiple datasets without affecting processing performance. Re-search comparing classifiers shows that the CART classifier delivers equally accurate, if not superior, results compared to other common classification algorithms. In this study, we utilized the "classifier.smil Cart" method for Land Use Land Cover classification.

Random Forest: Machine learning algorithms, such as Random Forest, have proven highly effective in analyzing intricate remote sensing datasets. Stratified forests have a key advantage in handling large datasets with high-dimensional features. By using multiple decision trees, random for-ests are more resistant to overfitting than individual decision trees, which contributes to their reliability, and they can classify complex patterns from satellite imagery (Mahendra et al. 2025). RF classifiers consist of multiple deci-sion trees, where the final classification result is obtained through a voting process among these trees. The RF approach involves two types of random selection. Firstly, it randomly helps in creat-ing subsets from the training dataset, representing a subtree that provides an individual classification result. The final result is then based on the aggregated votes from all these subtrees (Xie et al. 2019). Our study focused on land cover classification using ee.Classifier.smileRandomForest. This algo-rithm builds 100 decision trees for the classification task, enhancing the overall accuracy and robust-ness of the predictions.

This study area is particularly effective regionally because of the diverse land cover types ranging from urbanization to farmlands, rivers, forests, and water bodies. The multi-spectral images collected by Sentinel-2 and Landsat satellites, combined with Random Forests, have achieved a comprehensive RF classification that is detailed and accurate. It also makes it possible to map detailed land-cover change information through time series data analysis. This is particularly important for monitoring the impact of land-use change on water resources, agriculture, and the environment in this region.

Accuracy Assessment

In this research, accuracy assessment involves several key steps using machine learning tools and methodolo-gies. Specifically, the sklearn. The met-rics module is utilized to calculate and print the classification report, accuracy score, and confusion matrix. The classification_report function provides a detailed breakdown of recall, precision, and F1-score for each class. The accu-racy_score function computes the overall accuracy of the model, which is the ratio of correctly predicted instances to the total instances. The confusion matrix, displayed using Seaborn's heatmap, offers a visual representation of the true vs. predicted classifications, helping identify misclassifications.

The formula for accuracy is expressed as:

Accuracy = Number of correct predictions/Total number of predictions

Additionally, the classifierMetrics function is designed to evaluate the performance of the mod-el, highlighting the effectiveness of the Random Forest classifier in this context. The function receives the true labels (y_test) and predicted labels (y_pred), printing out the classification metrics.

Confusion Matrix

The confusion matrix is plotted to visually assess how well the model is performing across different classes. This comprehensive assessment method ensures a robust evaluation of the classification results, providing insights into areas for potential improvement. Furthermore, the Random Forest classifier is trained and tested using a subset of the dataset, with cross-validation enhancing the reliability of the accuracy measurement.

The performance evaluation of classification decisions across different land cover categories is performed using the confusion matrix model. The Random Forest classifier exhibited high accuracy in particular land cover categories, but showed misidentification behavior in selected classes. The classification accuracy for Water and Bare Ground reached maximum levels because their spectral features enabled clear discrimination from other classes. The classification of Flooded Vegetation and Grassland suffers from major misinterpretation areas because their spectral signatures have similar characteristics. The spectral characteristics of crops align closely with those of Shrub & Scrub vegetation, thus leading to their classification overlap. Misclassification occurs because some images show a mixture of urban structures and vegetation elements. The classifier makes incorrect decisions

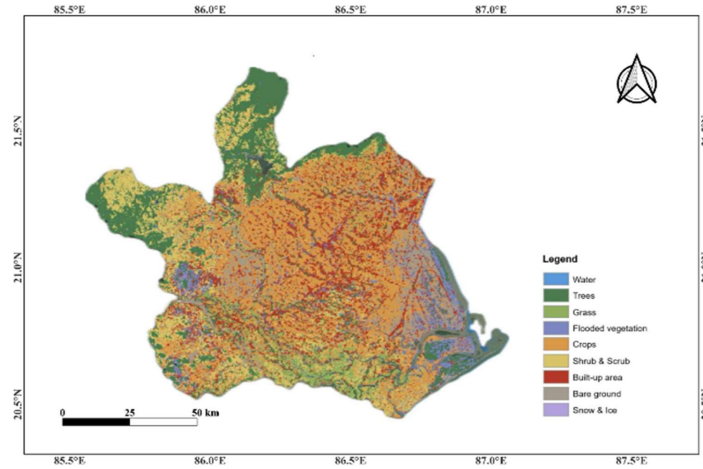
Table 1: Comparison of Area km^2 and Percentage of RF and CART.

Data set	Classes	RF		Cart	
Dynamic World Cover		Area km^2	Percentage	Area km^2	Percentage
	vegetation	609.96	6.47	775.87	8.23
	Built	1.228.10	13.02	1,217.78	12.91
	Barren	331.56	3.52	512.22	5.43
	Water	326.79	3.47	559.95	5.94

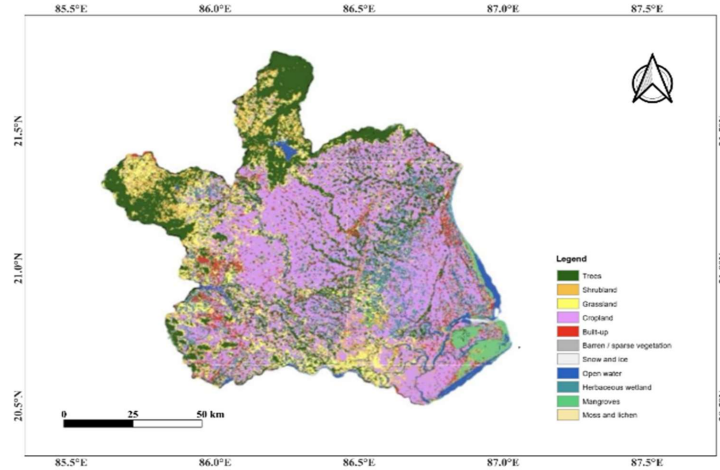
because land cover types share similar spectral information, and some pixels contain multiple classes as well as cyclic changes in vegetation and water levels.

Performance Evaluation

In our research, we perform a comprehensive accuracy assessment of the classification results through a se-ries of steps leveraging both GEE and the scikit-learn library. We use several key met-rics such as the classification report, accuracy score, class distribu-tion and confusion matrix to evaluate performance. The classification report provides recall, precision, and F1-score for each class, giving a detailed overview of the model's performance. The accuracy score is the ratio of correct-ly predicted instances to the total instances and provides a summary measure of the classifier's accuracy. The class distribution is presented



Dynamic World V1 2023



ESA WorldCover 10m v100/200 2023

Fig. 3a: Spatial distribution of LULC.

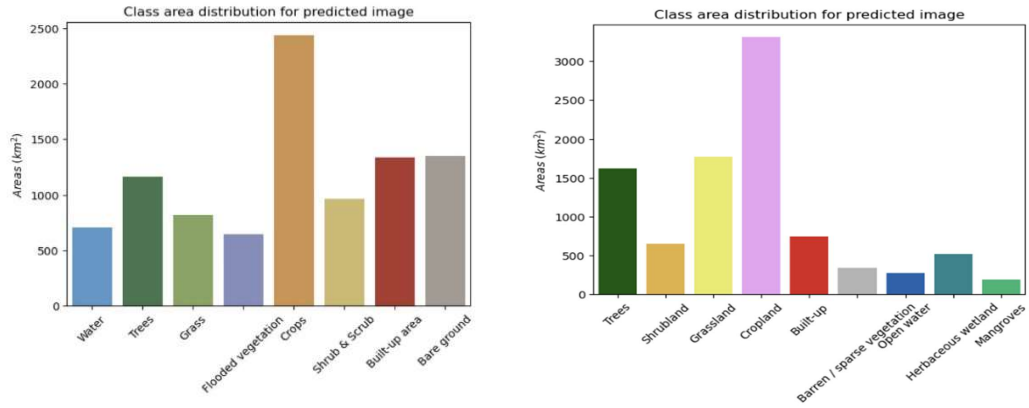


Fig. 3b: Class distribution of the LULC predicted image.

Classification report:

	precision	recall	f1-score	support
Water	0.71	0.69	0.70	291
Trees	0.58	0.57	0.58	330
Grass	0.55	0.60	0.58	286
Flooded vegetation	0.53	0.56	0.54	311
Crops	0.40	0.36	0.38	330
Shrub & Scrub	0.41	0.45	0.43	301
Built-up area	0.46	0.41	0.44	295
Bare ground	0.65	0.67	0.66	299
Snow & Ice	0.00	0.00	0.00	0
accuracy			0.54	2443
macro avg	0.48	0.48	0.48	2443
weighted avg	0.54	0.54	0.54	2443

Accuracy score: 0.54
[0, 1, 2, 3, 4, 5, 6, 7, 8]

Dynamic World V1 2023

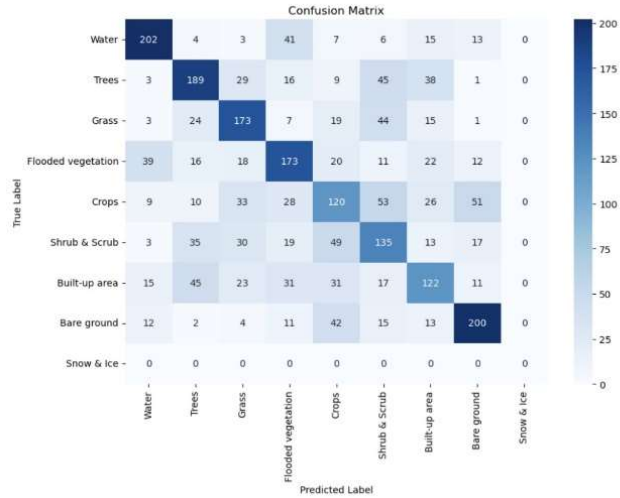
Classification report:

	precision	recall	f1-score	support
Trees	0.64	0.65	0.64	296
Shrubland	0.56	0.52	0.54	304
Grassland	0.43	0.47	0.45	311
Cropland	0.64	0.58	0.61	305
Built-up	0.61	0.59	0.60	319
Barren / sparse vegetation	0.57	0.55	0.56	294
Snow and ice	0.00	0.00	0.00	0
Open water	0.79	0.80	0.79	304
Herbaceous wetland	0.55	0.55	0.55	300
Mangroves	0.74	0.85	0.79	269
Moss and lichen	0.00	0.00	0.00	0
accuracy			0.61	2702
macro avg	0.50	0.50	0.50	2702
weighted avg	0.61	0.61	0.61	2702

Accuracy score: 0.61
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

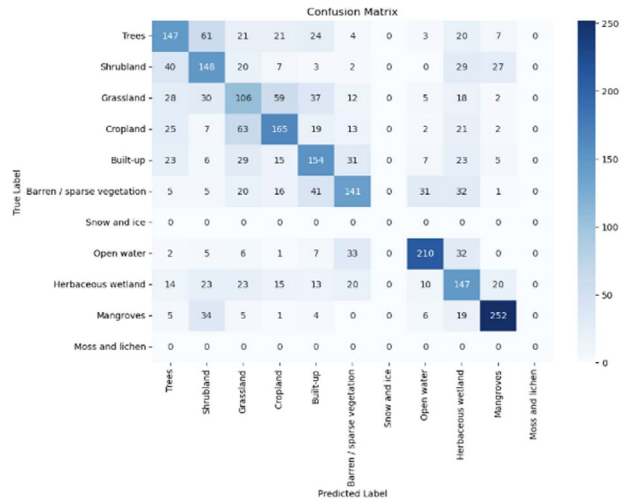
ESA WorldCover 10m v100/200 2023

Fig. 3c: Classification Report on LULC data prediction.



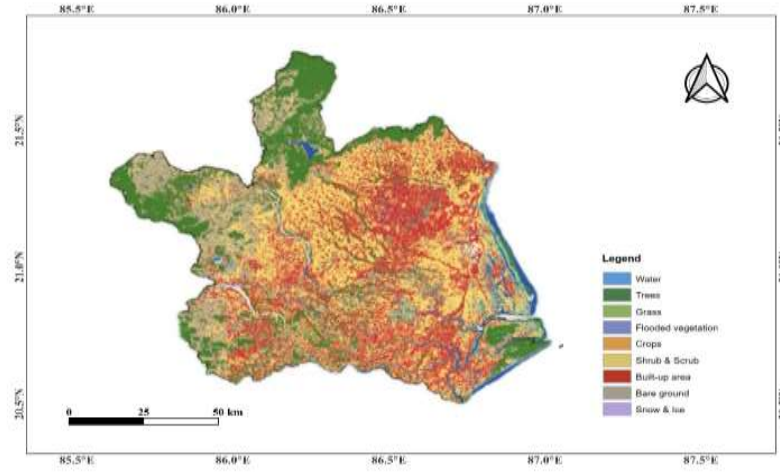
Dynamic World V1

Figure Cont....

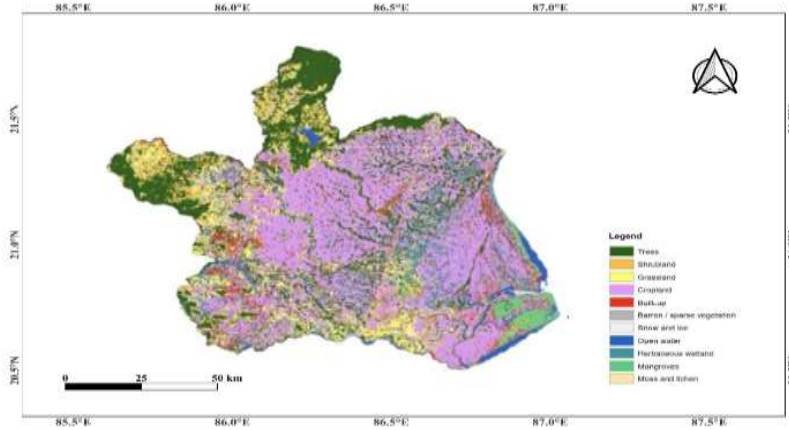


ESA WorldCover 10m v100/200

Fig. 3d: Confusion matrix of LULC data prediction.

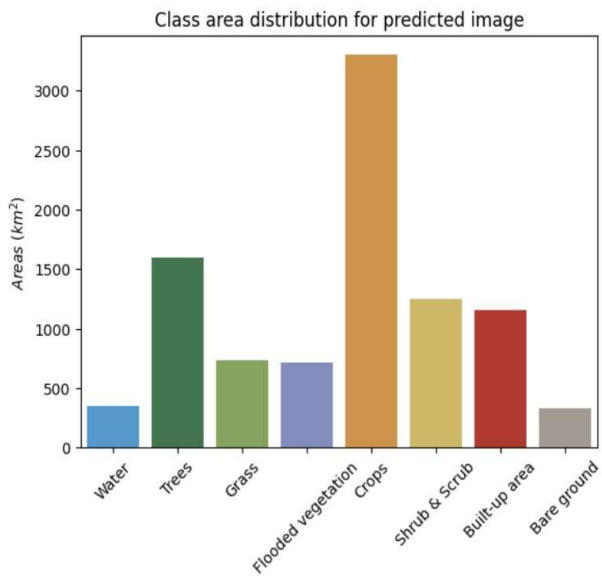


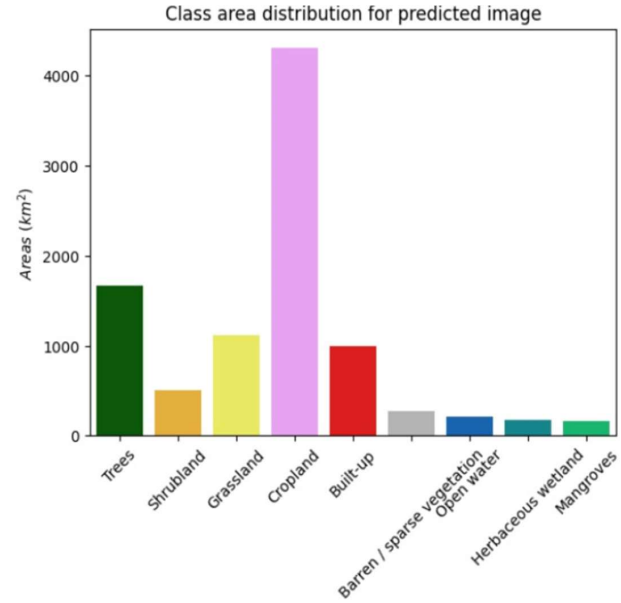
Dynamic World V1



ESA WorldCover 10m v100/200

Fig. 4a: Spatial distribution of LULC.





Dynamic World V1 ESA WorldCover 10m v100/200

Fig. 4b: Class distribution for the LULC predicted image.

Classification report:

	precision	recall	f1-score	support
Water	0.85	0.69	0.76	291
Trees	0.72	0.68	0.70	330
Grass	0.66	0.77	0.71	286
Flooded vegetation	0.62	0.74	0.67	311
Crops	0.63	0.54	0.58	330
Shrub & Scrub	0.58	0.52	0.55	301
Built-up area	0.59	0.58	0.59	295
Bare ground	0.70	0.84	0.77	299
Snow & Ice	0.00	0.00	0.00	0
accuracy			0.67	2443
macro avg	0.60	0.59	0.59	2443
weighted avg	0.67	0.67	0.67	2443

Accuracy score: 0.67
[0, 1, 2, 3, 4, 5, 6, 7, 8]

Dynamic World V1

Classification report:

	precision	recall	f1-score	support
Trees	0.73	0.74	0.74	296
Shrubland	0.65	0.63	0.64	304
Grassland	0.62	0.51	0.56	311
Cropland	0.73	0.68	0.70	305
Built-up	0.66	0.75	0.70	319
Barren / sparse vegetation	0.61	0.78	0.68	294
Snow and ice	0.00	0.00	0.00	0
Open water	0.90	0.77	0.83	304
Herbaceous wetland	0.67	0.60	0.63	300
Mangroves	0.81	0.90	0.85	269
Moss and lichen	0.00	0.00	0.00	0
accuracy			0.70	2702
macro avg	0.58	0.58	0.58	2702
weighted avg	0.71	0.70	0.70	2702

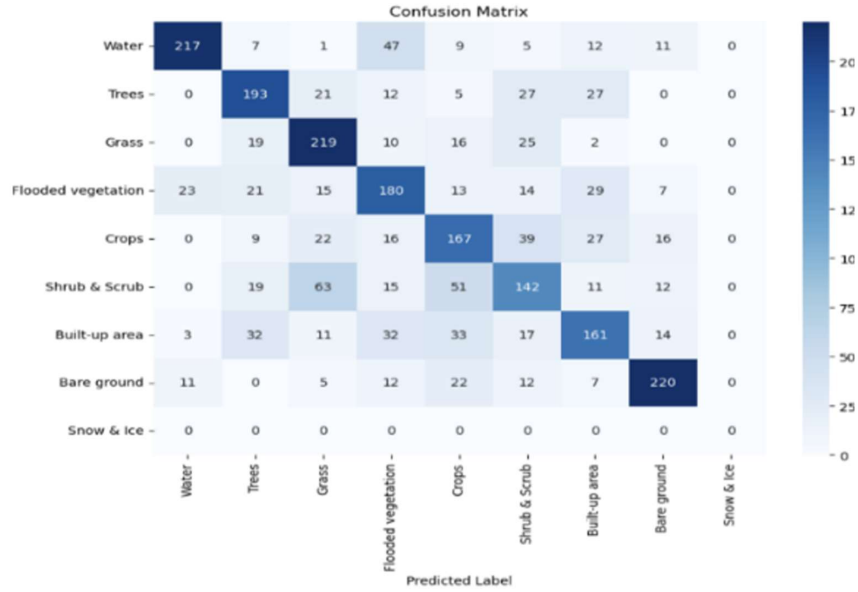
Accuracy score: 0.70
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

ESA WorldCover 10m v100/200

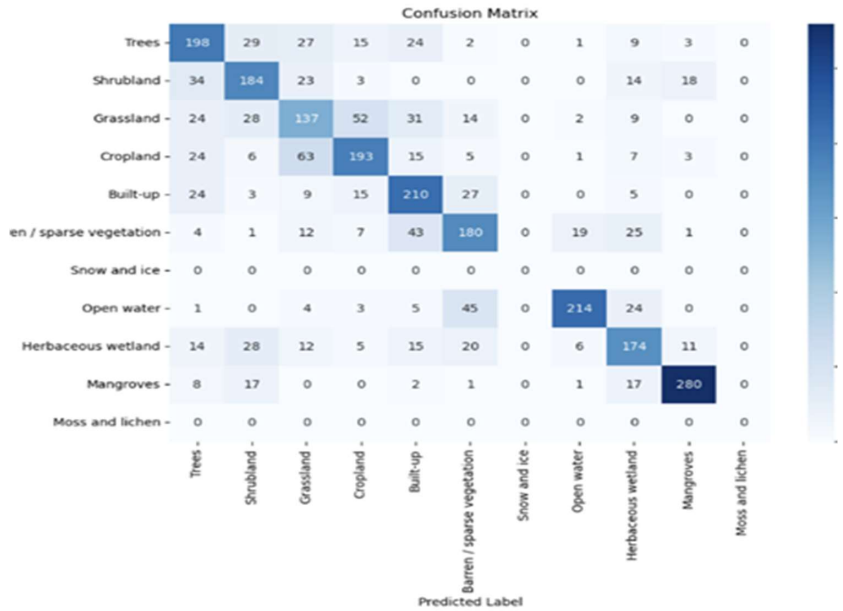
Fig. 4c: Classification Report on LULC data prediction.

in Table 1. The confusion matrix visually represents the true versus predicted classifications, highlighting areas of misclassification and allowing us to evaluate the classifier's performance comprehensively.

The scikit-learn functions classification_report, accuracy_score, and confusion_matrix were used to compute the classification metrics. Seaborn: heatmap to plot the confusion matrix gives an intuitive sense of classifier performance. By performing these extensive calculations of accuracy, we will ultimately obtain a well-rounded view of how well our approach in this specific case worked, as well as information on any concrete areas we can improve upon.



Dynamic World V1



ESA WorldCover 10m v100/200

Fig. 4d: Confusion matrix of LULC data prediction.

RESULTS

CART

The CART algorithm was employed to categorize the study area into four major categories: vegetation, barren land, built-up areas, and water bodies. According to the classification results, water 710.07 km^2, trees 1,162.23 km^2, grass land 818.87 km^2, flooded vegetation 647.74 km^2, crops 2,433.10 km^2, shrub and scrub 965.40 km^2, built-up area 1,338.98 km^2, and bare ground 1,353.42 km^2. These results were

obtained using the Dynamic World Cover dataset with stratified sampling over 6 months. Similarly, comparable results were observed using the ESA dataset, as illustrated in Figs. 3a, b, c, d.

Random Forest Algorithm

Figs. 4a, b, c, and d depict the outcomes of a Random Forest classifier, utiliz-ing stratified sampling of Dynamic World Cover and ESA data over a 6-month period with monthly classification. The land cover distribution in the Brahmani Baitarani basin is as follows: 330.36 km² of water, 1,193.78 km² of trees, 795.17 km² of grass, 637.55 km² of flooded vegetation, 2,908.39 km² of crops, 582.44 km² of shrub and crops, 1,570.72 km² of built-up areas, and 1,411.40 km² of bare ground. Similar results were obtained using the ESA dataset.

Result Validation

For LULC classification, RF is preferred over CART because RF has more accuracy and robustness than CART (Talukdar et al. 2020). In the Brahmani-Baitarani Basin, RF outperformed the accuracy of CART. The Dynamic World Cover dataset classified 559 km² of water, 2,522 km² of crops, and 1,217 km² of built-up areas, with RF remaining consistent over a six-month period. This ensemble method mitigates overfitting by averaging predictions over many decision trees, yielding more stable classifications, particularly in high-dimensional datasets such as satellite imagery (Ren et al. 2024). The CART algorithm provided a satisfying accuracy, yet in our research, it displayed limitations common to decision trees, including susceptibility to small changes in the training data and overfitting. In contrast, RF showed the ability to cope with noisy data and classified more complex land covers, such as flooded vegetation and shrub areas (trees 1397 km², flooded vegetation 775.87 km²), with higher accuracy. The performance of CART was limited by its lack of ensemble learning, which potentially renders RF a valid algorithm for mapping land use/land cover (LULC) in our study area. These findings demonstrate the significant benefits of employing RF for more accurate and consistent LULC classification, especially for diverse types of land.

DISCUSSION

The outcome of this research study demonstrates that Random Forest (RF) and Classification and Regression Trees (CART), which are ML algorithms, excel in achieving precise LULC classification. RF demonstrated robustness in classification in terms of accuracy. The reasonable performance of CART was overshadowed by its propensity for overfitting, which damaged the classification accuracy results. This study used Sentinel-2 and Landsat 8/9 datasets alongside Dynamic WorldCover and ESA WorldCover, which have proven effective for classification purposes. The utilization of the Google Earth Engine (GEE) simplifies data analysis by providing smooth nonlinear processing for large-scale remote sensing image data. The preprocessing stage that involved cloud masking alongside geometric corrections and spectral index calculations of NDVI, NDWI, and NDBI proved vital for achieving superior classification results.

The main result of this study demonstrates that RF-based LULC classification can serve practical environmental objectives, including environmental monitoring, urban planning, flood risk assessment, and sustainable land management. Water resource management and disaster preparedness decisions benefit from precise LULC classification within the Brahmani-Baitarani River Basin because the region undergoes several land-cover changes through monsoon-driven flooding. In addition, RF can offer the government more effective zoning regulation, deforestation, and agricultural land use tracking. Scientists can address issues of habitat destruction and loss of biodiversity by using data collected via RF-based classifications, as urban planners can integrate the information to implement resilient infrastructure and encourage green spaces.

A comparison with other studies proves that RF achieves better performance than CART in LULC, corroborating similar findings in Sentinel and Landsat studies. The classification accuracy of our study is fairly close to that reported previously, which reinforces that RF is a robust algorithm for large-scale mapping. The performance difference observed is primarily due to the RF's capability to deal with high-dimensional data while reducing variance through ensemble learning and the singularity of a CART tree that becomes much more sensitive to noise. Nevertheless, one of the drawbacks of Random Forest is its dependency on high-quality training data and sensitivity towards class imbalance, making it a better algorithm to use with continuous improvement of the dataset, as well as hybrid modeling for better reliability.

CONCLUSIONS

The investigation carried out in the Brahmani Baitarani area emphasizes the necessity of combining higher-order remote sensing approaches for the assessment of LULC classes. Our research demonstrated the effectiveness of categorizing heterogeneous landscapes using Landsat and Sentinel-2 satellite images, providing a better understanding of vegetation extent, built-up areas, water bodies, and barren land in the study area. Therefore, the findings show that Random Forest has performed much better than the CART

algorithm. This superiority could be attributed to the fact that RF uses an ensemble method of classifying features that helps it avoid overfitting, while simultaneously improving the level of classification accuracy, especially in diverse and disaggregated land cover classes. The classification report, confusion matrix, and other measures used in this chapter also provide a more refined accuracy estimation than the overall accuracy measure, which confirms the ability of RF to handle high-dimensional data, such as satellite imagery data. In conclusion, this study provides a sound platform for environmental monitoring of the region as well as the application of sustainable land management practices in the Brahmani–Baitarani area.

ACKNOWLEDGMENTS

The Authors thank the National Institute of Hydrology (NIH) for their guidance and support.

NOMENCLATURE

ANN: Artificial Neural Networks

CART: Classification and Regression Trees

CNN: Convolutional Neural Network

DT: Decision Tree

ESA: European Space Agency

GEE: Google Earth Engine

LULC: Land Use Land Cover

MD: Mahalanobis Distance

ML: Machine Learning

MLC: Maximum Likelihood Classifier

MLP: Multilayer Perceptron

NB: Naive Bayes

NBR: Normalized Burn Ratio

NDBI: Normalized Difference Built-up Index

NDVI: Normalized Difference Vegetation Index

NDWI: Normalized Difference Water Index

NIR: Near Infrared Reflectance

RBF: Radial Basis Function

RED: Red Band

RF: Random Forest

RS: Remote Sensing

SAM: Spectral Angle Mapper

SVM: Support Vector Machine

REFERENCES

Ashok, A., Rani, H.P. and Jayakumar, K.V. 2021. Monitoring of dynamic wetland changes using NDVI and NDWI-based Landsat imagery. *Remote Sensing Applications: Society and Environment*, *23*, p.100547. [DOI]

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S. and Schwehr, K., 2022. Dynamic World, Near-real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1), pp.251. [DOI]

Carranza-García, M., García-Gutiérrez, J. and Riquelme, J.C., 2019. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing,* 11(3), pp.274. [DOI]

Feng, S., Li, W., Xu, J., Liang, T., Ma, X., Wang, W. and Yu, H., 2022. Land use/land cover mapping based on GEE for the monitoring of changes in ecosystem types in the upper Yellow River basin over the Tibetan Plateau. *Remote Sensing,* 14(21), pp.5361. [DOI]

Gebeyehu, A.E., Chunju, Z. and Yihong, Z., 2019. Assessment and mapping of land use change by remote sensing and GIS: A case study of Abaya Chamo Sub-basin, Ethiopia. *Nature Environment and Pollution Technology*, 18(2), pp.549-554.

Indraja, G., Aashi, A. and Vema, V.K., 2024. Spatial and temporal classification and prediction of LULC in Brahmani and Baitarni basin using integrated cellular automata models. *Environmental Monitoring and Assessment*, 196(2), pp.117. [DOI]

Kadam, S., Kadam, A., Devale, P., Bandgar, A., Manepatil, R., Kale, R., Gujar, J., Bundele, C. and Chavan, T., 2024, February. Improving Earth Observations by correlating Multiple Satellite Data: A Comparative Analysis of Landsat, MODIS and Sentinel Satellite Data for Flood Mapping. In *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom*, 1421, pp. 1581-1587). [DOI]

Loukika, K.N., Keesara, V.R. and Sridhar, V., 2021. Analysis of land use and land cover using machine learning algorithms on Google Earth Engine for Munneru River Basin, India. *Sustainability*, 13(24), pp.13758. [DOI]

Mahajan, M., Kadam, S., Kulkarni, V., Gujar, J., Naik, S., Bibikar, S., Ochani, A. and Pratap, S., 2024a. ECG signal classification via ensemble learning: addressing intra- and inter-patient variations. *International Journal of Information Technology*, *16*(8), pp.4931-4939. [DOI]

Mahajan, M., Kadam, S., Kulkarni, V., Gujar, J., Naik, S., Bibikar, S., Ochani, A. and Pratap, S., 2024b. A machine learning framework for the classification of ECG Signals. In *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 270-264). IEEE. [DOI]

Mahendra, H., Pushpalatha, V., Rekha, V., Sharmila, N., Kumar, D.M., Pavithra, G.S., Basavaraju, N.M. and Mallikarjunaswamy, S., 2025. LULC classification for change detection analysis of remotely sensed data using a machine learning-based random forest classifier. *Remote Sensing,* 71, p.451. [DOI]

Mateo-García, G., Gómez-Chova, L., Amorós-López, J., Muñoz-Marí, J. and Camps-Valls, G., 2018. Multitemporal cloud masking in the Google Earth Engine. Remote Sensing, 10(7), pp.1079. [DOI]

Maxwell, A.E., Warner, T.A. and Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), pp.2784-2817. [DOI]

Nedd, R., Light, K., Owens, M., James, N., Johnson, E. and Anandhi, A., 2021. A synthesis of land use/land cover studies: Definitions, classification systems, meta-studies, challenges and knowledge gaps on a global landscape. *Land*, 10(9), pp.994. [DOI]

Ren, H., Pang, B., Bai, P., Zhao, G., Liu, S., Liu, Y. and Li, M., 2024. Flood susceptibility assessment with random sampling strategy in ensemble learning (RF and XGBoost). *Remote Sensing*, 16(2), pp.320. [DOI]

Rong, C. and Fu, W., 2023. A Comprehensive Review of Land Use and Land Cover Change Based on Knowledge Graph and Bibliometric Analyses. Land, 12(8), pp.1573. [DOI]

Shetty, S., 2019. Analysis of machine learning classifiers for LULC classification on Google Earth Engine. Master's thesis. University of Twente

Swain, S.S., Mishra, A., Chatterjee, C. and Sahoo, B., 2021. Climate-changed versus land-use altered streamflow: A relative contribution assessment using three complementary approaches at a decadal time-spell. *Journal of Hydrology*, 596, p.126064. [DOI]

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.A. and Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations: A review. *Remote Sensing*, 12(7), p.1135. [DOI]

Tassi, A. and Vizzari, M., 2020. Object-oriented LULC classification in Google Earth Engine combining SNIC, GLCM, and machine learning algorithms. *Remote Sens*ing, 12(22), pp.1 – 17. [DOI]

Tempa, K., Ilunga, M., Agarwal, A. and Tashi, 2024. Utilizing Sentinel-2 Satellite Imagery for LULC and NDVI change dynamics for Gelephu, Bhutan. *Applied Sciences,* 14(4), pp.1578. [DOI]

Velastegui-Montoya, A., Montalván-Burbano, N., Carrión-Mero, P., Rivera-Torres, H., Sadeck, L. and Adami, M., 2023. Google Earth Engine: A global analysis and future trends. *Remote Sensing*, 15(14), pp.3675. [DOI]

Xie, Z., Chen, Y., Lu, D., Li, G. and Chen, E., 2019. Classification of land cover, forest, and tree species classes with ZiYuan-3 multispectral and stereo data. *Remote Sensing*, 11(2), pp.164. [DOI]

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S. and Cartus, O., 2021. ESA WorldCover 10 m 2020 v100 (Version v100) [Data set]. Zenodo. [DOI]

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S. and Lesiv, M., 2022. ESA WorldCover 10 m 2021 v200 (Version v200) [Data set]. [DOI]

Zhao, Z., Islam, F., Waseem, L.A., Tariq, A., Nawaz, M., Islam, I.U., Bibi, T., Rehman, N.U., Ahmad, W., Aslam, R.W. and Raza, D., 2024. Comparison of three machine learning algorithms using Google Earth Engine for land use land cover classification. *Rangeland Ecology & Management*, 92, pp.129-137. [DOI]