Earth Systems and Environment
https://doi.org/10.1007/s41748-026-01126-2

ORIGINAL ARTICLE

Urban Land Use Land Cover Mapping Using Multi-Sensor Satellite Data and Machine Learning in Google Earth Engine: Toward Sustainable Planning

Adarsh Agrawal[1] · Priyamitra Munoth[1] · Vikas Poonia[1] · Lixin Wang[2]

Received: 10 November 2025 / Revised: 4 February 2026 / Accepted: 4 March 2026

Abstract

Urban land use/land cover (LULC) classification remains challenging due to high intra-class variability, spectral ambiguity, and mixed pixels in heterogeneous urban environments. For example, concrete and asphalt often exhibit spectral responses similar to barren leading to frequent misclassification in peri-urban transition zones. This study presents a comparative assessment of four widely used machine learning classifiers- Classification and Regression Tree (CART), Support Vector Machine (SVM), Random Forest (RF), and Gradient Tree Boosting (GTB) implemented on the Google Earth Engine platform for LULC mapping. Freely available multi-sensor satellite datasets from LISS-IV (5.8 m), Sentinel-2 (10 m), and Landsat-9 (30 m) were selected for their operational relevance and complementary resolutions, enabling systematic evaluation of resolution and spectral trade-offs in classification. Classifiers were trained using raw spectral bands and enhanced feature sets incorporating NDVI, NDWI, and sensor-specific band ratios (Red/Green and Green/Blue for Landsat-9 and Sentinel-2, Red/NIR and Green/NIR for LISS-IV). A standardized dataset of 100 training and 50 testing samples per class was used. Higher training proportions yielded only marginal accuracy gains, while lower proportions reduced classification stability. Results indicate that feature enhancement improved overall classification accuracy by approximately 2.4%-8.8% across sensors and classifiers. SVM, RF, and GTB consistently outperformed CART, with SVM applied to Landsat-9 achieving the highest accuracy (93.2%, Kappa = 0.915). The findings demonstrate the importance of integrating optimized feature engineering with robust classifiers to improve urban LULC mapping reliability. This study provides practical guidance for remote sensing analysis, urban planners and supports data-driven, sustainable urban development.

✉ Vikas Poonia
vgpcivilengineer@gmail.com

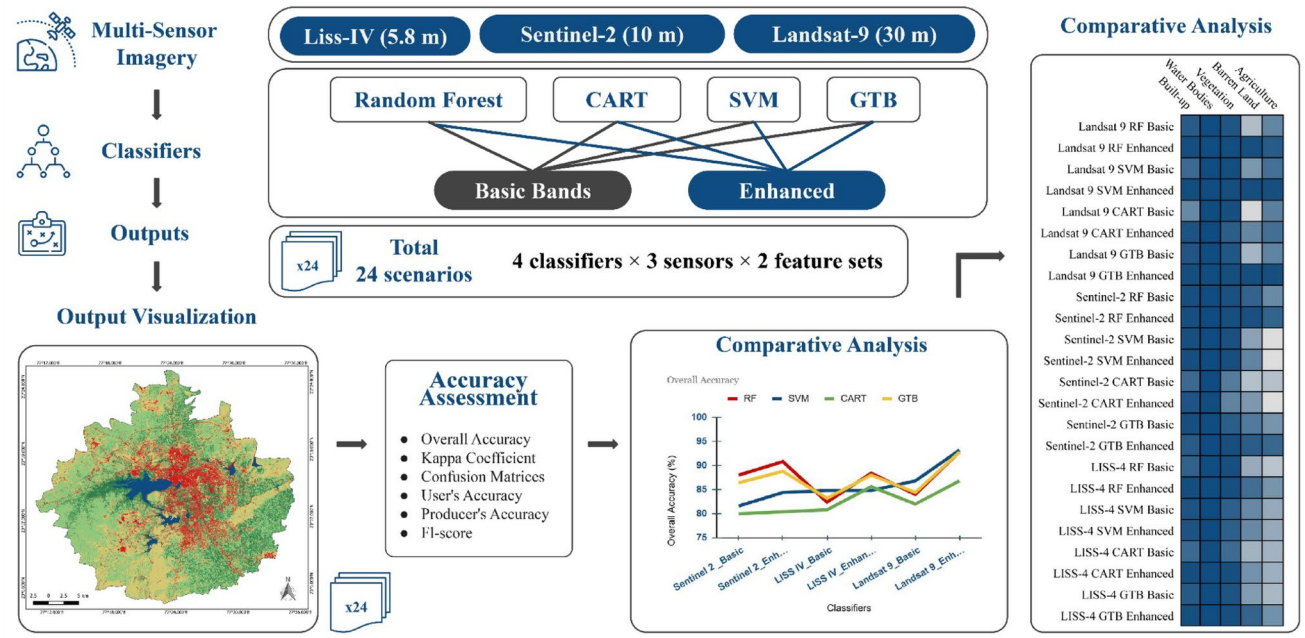
[1] Department of Civil Engineering, MANIT Bhopal, Bhopal 462003, Madhya Pradesh, India

[2] Department of Earth and Environmental Sciences, Indiana University Indianapolis, Indianapolis, IN 46202, USA

Published online: 09 April 2026

Graphical Abstract

This is a visual summary serves as a pivotal entry point into the research, offering a concise overview of the study's core findings and methodologies. This graphic abstract depicts a comparative analysis of machine learning classifiers and multi-sensor satellite imagery for Land Use/Land Cover (LULC) mapping utilizing Google Earth Engine (GEE). Three satellite datasets with varying spatial resolutions, i.e., LISS-IV (5.8 m), Sentinel-2 (10 m), and Landsat-9 (30 m), were examined. Four supervised machine learning algorithms, i.e., Random Forest (RF), Classification and Regression Tree (CART), Support Vector Machine (SVM), and Gradient Tree Boost (GTB), were employed under two feature conditions: basic spectral bands and enhanced features (comprising indices and band ratios). The experimental design generated 24 scenarios, each yielding classified outputs that were assessed through overall accuracy, kappa coefficient, confusion matrices, user accuracy, producer accuracy, and F1-score. The left panel illustrates sample classified maps, whereas the central plot displays variations in overall accuracy across classifiers, sensors, and feature sets. The heatmap on the right illustrates the F1 score, offering a comparative evaluation of classification performance in different LULC classes, including agriculture, vegetation, water bodies, built-up land, and bare land. This study examines the influence of sensor resolution, classifier selection, and feature enhancement on classification accuracy, offering insights for the advancement of land cover mapping and monitoring.



Highlights

- Enhanced spectral features, such as NDVI, NDWI, and band ratio (Red/Green and Green/Blue for Landsat-9 and Sentinel-2, Red/NIR and Green/NIR for LISS-IV), significantly improved classification accuracy by approximately 2.4% to 8.8% across all sensors.
- The Support Vector Machine (SVM) applied to Landsat-9 images achieved the best accuracy of 93.2% in classifying urban land cover.
- On Sentinel-2 and LISS-IV images, Random Forest (RF) and Gradient Tree Boosting (GTB) outperformed Classification and Regression Tree (CART) by approximately 5%–10% in overall accuracy.
- This study illustrates the impact of sensor resolution, classifier selection, and feature enhancement on classification results, providing insights for the advancement of LULC mapping and monitoring.

Keywords Google earth engine (GEE) · Machine learning (ML) · Multi-sensor satellite data · Spectral indices · Sustainable urban development

1 Introduction

Remote sensing technology is the main tool for mapping LULC, permitting repeated and synoptic observations of the Earth's surface. LULC classification traditionally was obtained through visual interpretation or basic spectral thresholding. More complex landscapes, specifically in cities, and large volumes of satellite imagery, however, require advanced techniques. Urban LULC classification is challenging in nature because of the intra-class variability of attributes within a limited spatial area, ambiguity among different classes across the spectral range (e.g., sparse vegetation and low-density urban, or asphalt and bare ground), and dominance by mixed pixels where a single pixel has more than one land cover category (Verma and Jana 2019). The evolution of machine learning (ML) algorithms has transformed remote sensing-based LULC classification through the provision of sophisticated functionality for processing high-dimensional, high-complexity satellite data (Pande et al. 2024). Many ML algorithms have been applied successfully (Amin et al. 2024). Another evolution of free and open-source satellite imagery and high cloud-computing environments, such as Google Earth Engine (GEE), has continued to democratize the analysis of remote sensing data (Boothroyd et al. 2021). GEE facilitates easy access to a large database of geospatial information and parallel processing at scale, greatly alleviating the computational burden for researchers and practitioners (Shafizadeh-Moghadam et al. 2021). The synergistic integration of big data, sophisticated algorithms, and cloud resources opens new avenues of research in large-scale LULC mapping using remote sensing data.

The literature consistently identifies findings related to the selection of satellite data and ML algorithms for attaining high-accuracy LULC classification across various geographical and thematic contexts. Random Forest (RF) and Support Vector Machine (SVM) are the most commonly tested non-parametric classifiers in terms of algorithms (Dabija et al. 2021; Mao et al. 2020). It is apparent that they perform more effectively than traditional statistical methods, such as Maximum Likelihood Classification (MLC), but the exact outcome depends on both the study area and data preparation (Singh and Pandey 2021). For instance, RF proved superior in complex urban land-use mapping in Hangzhou, China (Mao et al. 2020), arid urban classification in Casablanca (Ouchra et al. 2023), and tropical mangrove monitoring (Tan et al. 2021), often exhibiting robust performance across various geomorphologies. Conversely, SVM yielded the highest overall accuracy (OA) in a semi-arid Tunisian environment (Baccari et al. 2025), and when classifying Landsat 8 OLI data in Vietnam (Lan et al. 2025). For reliable large-scale and long-term monitoring, it is crucial to optimize both the temporal and spatial contexts. Research in intricate boreal landscapes demonstrates that incorporating multi-temporal Sentinel-2 scenes is necessary for accurately capturing the significant phenological variation needed to conduct precise LULC mapping (Abdi 2020). Additionally, applying temporal smoothing techniques to Landsat time series data can enhance map accuracy by reducing noise caused by seasonal fluctuations. These findings collectively underscore that optimal classification accuracy depends on enhancing the interplay among sensor resolution, feature augmentation, and the selected algorithm, often favoring ensemble or deep learning methodologies that can manage the intricacies of multi-source and multi-temporal data.

Since LULC maps are essential for understanding and planning various aspects of the Earth's surface, they play a pivotal role in city planning, environmental surveillance, natural resource management, and the study of climate change and hydro-climatic extremes such as droughts (Poonia et al. 2021b, c), flash droughts (Poonia et al. 2022, 2024), floods (Mukherjee et al. 2025), agricultural yield (Das et al. 2020; Poonia, et al. 2021a) and water scarcity (Poonia et al. 2025). The expansion of urbanization in cities worldwide necessitates the prompt and accurate collection of LULC data for sustainable growth and effective decision-making processes (Munoth and Goyal 2020; Pandey et al. 2021). Therefore, Bhopal, one of India's fast-growing urban areas in Madhya Pradesh, is a quintessential example of this requirement, overwhelmed by problems of urbanization, infrastructure growth, and the preservation of the natural environment, has been selected as the study area in this study. Despite numerous studies on LULC classification using various sensors and algorithms, a comparative systematic study diligently testing different machine learning classifiers using diverse spatial resolutions (i.e., LISS-IV, Sentinel-2, and Landsat-9) and spectral feature sets based on the same training and testing data set in a dynamically changing urban environment, such as Bhopal, is under reported. Such research is essential to determining the best sensor-classifier pairs that achieve the best accuracy for large-scale urban LULC mapping to enable actionable recommendations for remote sensing analysts and urban planners. Hence, this research aims to bridge this gap by conducting a thorough comparative study between four leading machine learning classifiers: Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Tree (CART), and Gradient Tree Boosting (GTB) for LULC mapping of the Bhopal planning boundary. The key novelty of this research lies in: (i) the cross-sensor benchmarking of high-resolution (LISS-IV), medium-resolution (Sentinel-2), and moderate-resolution (Landsat-9) imagery within the same complex urban environment; (ii) the explicit evaluation of basic versus enhanced spectral feature spaces to disentangle the relative contribution of

feature engineering from classifier choice; and (iii) This study demonstrates a comprehensive implementation of multi-sensor, multi-classifier LULC classification entirely within the Google Earth Engine cloud computing platform.

In this study, the effect of multi-sensor satellite imagery at spatial resolutions of 5.8 m (LISS-IV), 10 m (Sentinel-2), and 30 m (Landsat-9) has been studied. Additionally, the performance of LULC classification using extended spectral feature sets, such as NDVI, NDWI, and band ratios, has been evaluated against the simple spectral approach with single spectral bands alone. A zone of uniform and steady ground truth is used as a standard for comparison for all experiments. Therefore, the specific objectives of this study are:

1. To comparatively evaluate the performance of RF, SVM, CART, and GTB classifiers in the context of LULC mapping in the Bhopal planning boundary.
2. To determine the impact of varying spatial resolutions (LISS-IV, Sentinel-2, and Landsat-9) on accuracy in classification.
3. To compare the effectiveness of enhanced feature sets over simple spectral bands in enhancing classification performance.
4. To identify the best machine learning classifier-sensor pair to produce accurate LULC mapping for urban planning purposes in Bhopal.

2 Study Area

The study region includes the Bhopal planning area (around 1005.64 km2) of the state of Madhya Pradesh, India. Geographically, Bhopal is situated around 23° 15′ 35.6″ N latitude and 77° 24′ 45.4″ E longitude as shown in Fig. 1. Bhopal is known for its distinctive landscape, represented by two noticeable lakes, the Upper Lake (Bada Talab) and Lower Lake (Chhota Talab), which constitute its core identity, source of water, and local ecosystem.

Bhopal has witnessed rapid and largely unplanned urbanization over the last few decades, driven by population growth and economic expansion (Barpete and Mehrotra 2023). Growth resulted in enormous LULC transformation, with conversion of agricultural land, barren land, and natural cover into constructed land (Sharma et al. 2024). These changes put tremendous pressure on natural resources, produce urban heat island effects, affect regional diversity, and put stress on existing infrastructure (Mehra and Swain

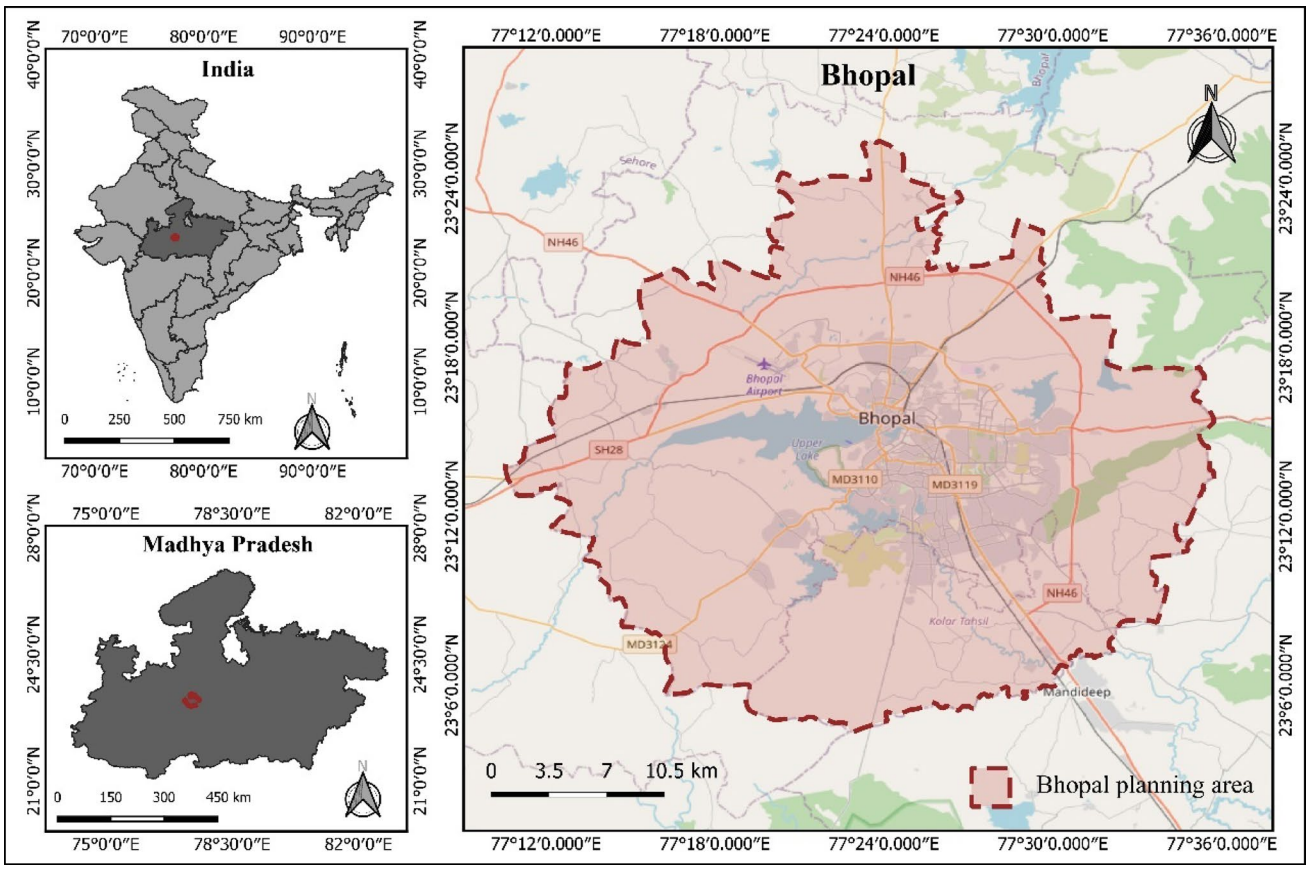


Fig. 1 Location map of the study area, Bhopal

2024). Individual environmental issues in Bhopal include water quality degradation in its lakes, caused by urban runoff, loss of open space, and rising impervious surfaces, which disrupt hydrological cycles. This dynamic nature of urban development and its related environmental consequences make Bhopal a good and suitable case study for the evaluation of different strategies for the preparation of LULC.

3 Data collection

3.1 Satellite Imagery Datasets

Multi-sensor satellite imagery from three satellites was used to compare the impacts of spectral depth and spatial resolution on the accuracy of LULC classification. All data acquisition and initial processing were done within the Google Earth Engine (GEE) platform, providing access to large archives and computational capacity. Table 1 provides a comprehensive list of the satellite imagery datasets used in the analysis, detailing their key specifications and sources.

3.2 Training and Testing Data

An even and useful training and testing dataset was created to train and test all the models of classifiers using various sensors and feature sets. The dataset contained 100 training points and 50 test points for each of the five LULC classes. This study employed a fixed 2:1 training-to-testing ratio (100:50 samples per class). While this ratio is standard in remote sensing literature and provided sufficient samples for robust model training (Belgiu and Drăguţ 2016). This ratio was selected as a compromise between (i) providing sufficient samples for robust classifier training and (ii) retaining an adequately sized, independent testing dataset for reliable accuracy assessment. In total, there were 500 training points and 250 test points in all the experiments. These points were gathered with utmost caution through a survey of India toposheets, utilizing high-resolution imagery sources, including Google Maps and Spectra Precision Mobile Mapper.

The multi-source strategy, combined with local topography and land use trend knowledge of the region in Bhopal, enabled the precise labeling of every point. To preclude bias, the points were selected purposively to cover the full spectral range for each class with a uniform spatial distribution across the study area and negligible over-sampling of specific sub-regions.

The training points were aggregated into a feature collection, while the testing points were aggregated into a separate feature collection. This study did not employ k-fold or spatial cross -validation because its primary objective was to conduct a controlled, relative comparison of classifier behavior across different sensors and feature sets, rather than to optimize or report the highest possible absolute classification accuracy. The same training and test data are used in all classification applications to guarantee an unbiased and equitable comparative evaluation of sensor and classifier performance.

4 Methodology

All image processing, feature extraction, classifier training, LULC classification, and accuracy assessment operations were conducted entirely within the Google Earth Engine (GEE) cloud-computing infrastructure using its JavaScript API. GEE's high-performance parallel processing and large data storage greatly facilitated conducting this comparative analysis at high resolution. The method follows some key steps presented in Fig. 2 and explained in the following sections.

4.1 Data Pre-processing

For Sentinel-2 and Landsat 9, raw image collections were pre-filtered by study area geometry (and a specific date range from March 1, 2025 to March 31, 2025) to maintain uniform seasonality. Landsat-9 Level-2 surface reflectance imagery, preprocessed using the LaSRC (Land Surface Reflectance Code) algorithm, was retrieved with a strict cloud cover threshold of less than 5%. LaSRC applies radiative transfer modeling using atmospheric and ancillary data to correct for molecular scattering, absorption, and aerosols,

Table 1 Satellite imagery datasets used in this study and their detailed information

Sensor	Spatial Resolution	Spectral Bands Used	Acquisition Date	Source
LISS-IV	5.8 m	Green (B2), Red (B3), NIR (B4)	14 March 2025	Bhunidhi Portal (https://bhoonidhi.nrsc.gov.in)
Sentinel-2	10 m	Blue (B2), Green (B3), Red (B4), NIR (B8)	1 March 2025, 6 March 2025, 11 March2025, 21 March 2025, 26 March 2025, and 28 March 2025	Copernicus Open Access Hub (https://www.copernicus.eu/en)
Landsat 9	30 m	Blue (SR_B2), Green (SR_B3), Red (SR_B4), NIR (SR_B5)	3 March2025, 12 March 2025, 19 March 2025, and 28 March 2025	USGS archive (https://earthexplorer.usgs.gov/)

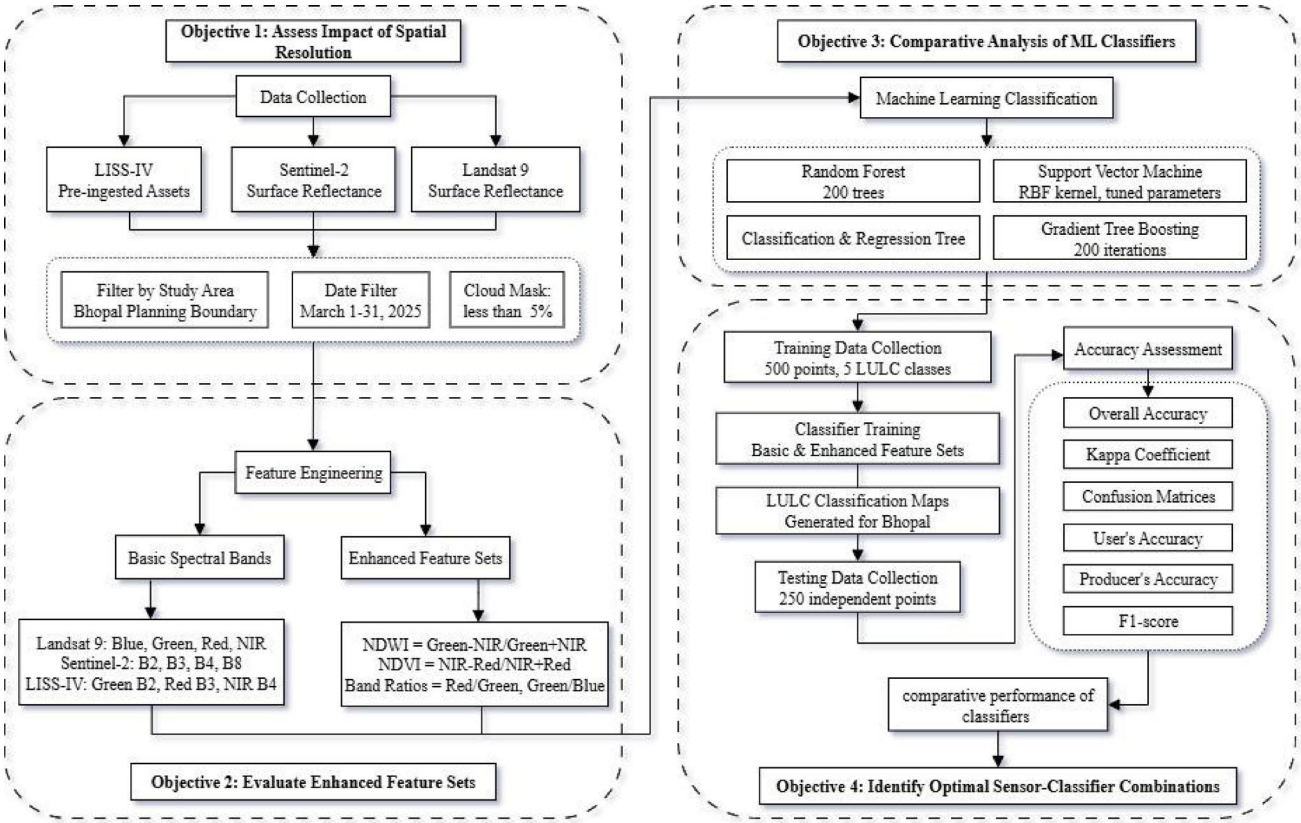


Fig. 2 Flowchart of the land use/land cover classification methodology

thereby converting top-of-atmosphere (TOA) reflectance into bottom-of-atmosphere (BOA) surface reflectance (Vermote et al. 2018). Along with this, Sentinel-2 Level-2 A surface reflectance imagery, corrected through the Sen2Cor processor, applying a cloud filter of less than 5% to ensure high-quality inputs. Sen2Cor performs atmospheric, terrain, and cirrus correction to derive BOA reflectance from Level-1 C TOA imagery (Main-Knorn et al. 2017). A median composite was then obtained from the filtered set of images to reduce residual cloud influence and normalize atmospheric variation. Landsat 9 Surface Reflectance bands SR_B2, SR_B3, SR_B4, and SR_B5 were scaled by a factor of 0.0000275 and offset by −0.2 in order to transform into true reflectance values (U.S. Geological Survey 2025). Sentinel-2 Surface Reflectance bands B2, B3, B4, and B8 were divided by 10,000 to get reflectance (European Space Agency 2015).

LISS-IV imagery was first downloaded from the Bhoonidhi portal (ISRO's geo-platform) and subsequently ingested into Google Earth Engine (GEE). Since LISS-IV products are distributed as standard georeferenced imagery without atmospheric correction, a Dark Object Subtraction (DOS) method (Chavez 1988) was implemented to minimize haze effects and path radiance. In this approach, a 1% percentile of the darkest pixels in each band within the study area was assumed to represent dark objects (such as deep water or shadows), and this value was subtracted from the entire image to approximate BOA reflectance. This percentile value is chosen to capture atmospheric scattering effects while avoiding bias from noise or extreme outliers in the scene. A single atmospherically corrected high-resolution scene acquired on March 14, 2025, was used directly, without compositing. All the preprocessed images were finally clipped to the specified Bhopal planning area.

4.2 Feature Engineering

To measure the effect of spectral ion in aggregate, classifications were performed using two different sets of features, namely the First Basic Spectral Bands, which included the fundamental spectral bands provided by all sensors, as shown in Table 2. Second, Enhanced Feature Sets, which integrated several widely used spectral indices and band ratios, were designed to enhance the discriminability of various LULC classes. The spectral indices used were Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and the Band Ratios (Table 3).

For Landsat-9 and Sentinel-2, two band ratios were calculated: Red/Green and Green/Blue. Since LISS-IV does

Table 2 Basic spectral bands and their corresponding spectral ranges used for classification across different sensors

Sensor	Band Name	Spectral Range (μm)
Landsat-9	B2 – Blue	0.45–0.51
	B3 – Green	0.53–0.59
	B4 – Red	0.64–0.67
	B5 – NIR	0.85–0.88
Sentinel-2	B2 – Blue	0.46–0.52
	B3 – Green	0.54–0.58
.	B4 – Red	0.65–0.68
	B8 – NIR	0.84–0.88
LISS-IV	B2 – Green	0.52–0.59
	B3 – Red	0.62–0.68
	B4 – NIR	0.77–0.86

Table 3 Enhanced spectral bands used for classification across sensors

Sensor	NDVI	NDWI	Band Ratio
Landsat-9	(B5 − B4) / (B5 + B4)	(B3 − B5) / (B3 + B5)	B4 / B3 and B3 / B2
Sentinel-2	(B8 − B4) / (B8 + B4)	(B3 − B8) / (B3 + B8)	B4/ B3 and B3 / B2
LISS-IV	(B4 − B3) / (B4 + B3)	(B2 − B4) / (B2 + B4)	(B3 / B4 and B2 / B4)

Table 4 LULC Class description used in this study

LULC Class	LULC Class Name	Description
Class 0	Built-up	All impervious human-made surfaces, including structures and transport networks, indicating urban development.
Class 1	Water Bodies	Semi-permanent and permanent water features (lakes, streams, ponds).
Class 2	Vegetation	Semi-natural and natural green cover (parks, forests, grasslands).
Class 3	Barren Land	Rocks, Exposed soil and sand; includes fallow land and construction sites.
Class 4	Agriculture	Actively cultivated lands (irrigated, rain-fed, fallow fields), representing rural/peri-urban farming.

not have a band in the blue (0.45–0.51 μm) range, two band ratios (Red/NIR and Green/NIR) were calculated for the LISS-IV sensor. These ratios have been shown to effectively reduce redundancy between adjacent spectral bands while amplifying subtle variations in reflectance, particularly for separating spectrally similar classes such as urban areas and barren land. By normalizing spectral differences, band ratios also help to mitigate the influence of atmospheric and illumination conditions, thereby improving classification robustness. These indices and ratios were subsequently calculated for the imagery of each sensor and utilized as additional bands to produce the "enhanced" image stacks.

Five distinct LULC classes, namely Built-up, Water bodies, Vegetation, Barren land, and Agriculture, have been identified as discernible within the Bhopal boundary, based on resources such as Esri (2021), which were defined for classification (Table 4).

4.3 Machine Learning Classifiers

Machine learning (ML) classifiers play a crucial role in remote sensing because they can handle high-dimensional spectral data, capture nonlinear relationships, and adapt well to diverse land cover types (Maxwell et al. 2018). Therefore, these methods are also known as non-parametric methods. Non-parametric methods are especially valuable since they do not assume any predefined data distribution, allowing them to manage complex feature spaces more effectively than traditional parametric approaches such as Maximum Likelihood Classification (MLC) (Mancino et al. 2023). This characteristic makes these classifiers particularly effective for remote sensing, where class boundaries are often irregular and heterogeneous, and the high dimensionality of spectral and index-based feature spaces poses challenges for parametric methods (Abdi 2020).

4.3.1 Random Forest (RF) & Gradient Tree Boosting (GTB) and Classification & Regression Tree (CART) Classifiers

Decision trees (iterations), one of the ML classifiers, form the conceptual basis for several of these methods. A decision tree is a supervised learning technique that recursively partitions the dataset into subsets based on feature values, thereby generating a hierarchical structure where internal nodes represent conditions (e.g., NDVI > 0.3) and terminal nodes correspond to classes (e.g., water or vegetation) (Sharma et al. 2013). While decision trees are easy to interpret and computationally efficient, their tendency to overfit when used individually limits their effectiveness. To overcome this, ensemble learning techniques have been developed that aggregate the outputs of multiple trees. Bagging, as employed by Random Forest, generates multiple trees using bootstrapped samples and combines their predictions, while boosting, as implemented in Gradient Tree Boosting, sequentially corrects the errors of previous trees to produce a strong predictive model (Sagi and Rokach 2018). Both strategies improve accuracy, stability, and generalization as compared to single-tree approaches. An ensemble learning method that builds multiple decision trees, favored for high accuracy and robustness in remote sensing (Pal 2005). GTB is an advanced ensemble learning technique that builds a strong predictive model from a series of weaker decision tree models. It operates by iteratively correcting the errors made by previous models in the sequence, with each new tree focusing on the residuals of the prior ones, leading to high accuracy and competitive performance in various remote sensing applications (Rizkallah 2025).

In this study, The hyperparameters adopted for the machine learning classifiers were selected to balance classification stability, computational efficiency, and fairness of comparison across sensors and feature sets. RF was implemented with the number of trees explicitly set to 200. This value is a commonly used parameter that provides a good balance between achieving high accuracy and maintaining computational efficiency for this type of problem. GTB was also implemented, with the number of trees set to 200, a standard practice for achieving robust performance without incurring excessive computational costs. Empirical evidence from prior studies and preliminary trial runs indicate that increasing the number of trees beyond ~150–300 yields diminishing returns, with only marginal improvements in overall accuracy (typically < 1%), while substantially increasing computational cost (Belgiu and Drăguţ 2016; Rodriguez-Galiano et al. 2012). Conversely, reducing the number of trees to around 100 can lead to slightly higher variance in predictions and reduced robustness, particularly for spectrally complex classes such as built-up and barren land. Increasing the ensemble size to 300 trees was found to produce negligible accuracy gains relative to 200 trees, while increasing execution time and memory usage within the cloud-computing environment(Gislason et al. 2006).

Another decision tree learning technique, CART, recursively partitions the dataset based on feature values to create a tree-like model. While simpler than more complex ensemble methods, CART offers the advantages of interpretability (the decision-making process is easy to follow) and computational efficiency, making it a foundational and widely used algorithm in remote sensing classification for baseline comparisons (Maxwell et al. 2018). In this study, default parameters were used for the implementation of CART. By default, the parameters include: (1) Max depth = unlimited (the tree can keep splitting until pure or until it hits other stopping rules), (2) Min leaf population = 1 (a leaf node can contain as few as one sample), (3) Number of trees = 1 (a single CART tree is grown, unlike ensemble methods such as Random Forest), (4) Splitting criterion = Gini impurity, which decides the best feature to split on, (5) No pruning = the tree is fully grown unless limited by training data.

4.3.2 Support Vector Machine (SVM)

Another ML classifier, i.e., Support Vector Machines (Cortes et al. 1995), represents a different non-parametric strategy that relies on kernel functions to transform data into higher-dimensional spaces where complex class boundaries can be separated linearly. A widely used example in remote sensing is the Radial Basis Function (RBF) kernel (Pal and Mather 2005). The Radial Basis Function (RBF) kernel, frequently adopted in classification problems, is particularly effective in reshaping data so that a linear boundary can separate complex patterns. The idea is illustrated in Fig. S1 (a) (b). Points that form concentric circles in two dimensions remain inseparable by a line, but once lifted into a higher dimension, a flat hyperplane can distinguish the classes. This trick allows SVMs to handle nonlinear problems without explicitly calculating the transformation, which is what makes them both powerful and efficient.

The effectiveness of SVMs, however, is not solely determined by the choice of kernel but also by the careful tuning of model hyperparameters. Settings such as gamma (γ) and cost (C) strongly influence classification performance by shaping the decision boundary and controlling the trade-off between generalization and error minimization (Dabija et al. 2021). As shown in Fig. S2, lower gamma (γ) values produce smoother, more generalized boundaries, whereas higher values result in complex, localized boundaries that may risk overfitting. Cost (C), on the other hand, regulates the balance between maximizing the margin and minimizing classification errors; a smaller C allows for a wider margin with some misclassification tolerance, while a larger C enforces stricter separation with narrower margins. These parameters are typically fine-tuned through trial-and-error or optimization strategies.

In this study, the SVM was implemented with a Radial Basis Function (RBF) kernel. Two key hyperparameters, 'gamma' and 'cost' were carefully tuned using a trial-and-error approach. After preliminary testing, optimal parameter values were selected for each sensor: for Landsat-9, γ = 0.5 and C = 100; for Sentinel-2, γ = 0.5 and C = 50; and for LISS-IV, γ = 0.001 and C = 200. These configurations provided a reasonable trade-off between accuracy and generalization, ensuring stable classification performance across different datasets.

All four classifiers used are non-parametric. CART is a simpler, non-parametric method with clear decision rules, while RF and GTB build upon this tree-based method to create ensemble methods that are much better at generalizing and being robust (Zurqani 2025). SVM is not tree-based, but it is still a non-parametric method because it uses kernel functions to set class boundaries, rather than predefined probability distributions (Blanco et al. 2023). The combination of these classifiers provides a wide range of non-parametric approaches, which enables a strong comparison across sensors and feature sets. Therefore, in this study, four non-parametric classifiers i.e. Random Forest (RF), Gradient Tree Boosting (GTB), Classification & Regression Tree (CART), and Support Vector Machine (SVM), were selected to compare classification performance across multiple sensors and feature sets. These algorithms represent distinct methodological strategies, together providing

a robust framework for land cover classification in remote sensing.

4.4 Classification Workflow

A total of 24 distinct classification experiments were performed for each of the three sensors, using both the basic feature set and the enhanced feature set. Feature values for the base spectral bands and the improved feature sets were taken from the respective preprocessed images for every 500 training points This was accomplished using the Sample Regions function in GEE, with the sampling scale determined by the native sensor resolution. All the ML classifiers were trained on the respective basic and enhanced feature set with the training Data (a total of 500 points in 5 classes). The trained classifiers were then applied to their respective full images (basic and enhanced band stacks) to obtain the entire LULC classification maps of the Bhopal planning boundary. Features were also extracted for testing, and 250 independent test sites were selected using Sample Regions at the appropriate sensor resolutions. The synthesized LULC maps were tested against the independent test data. For every classification, the approximated class labels for the test sites were compared with their corresponding actual reference class labels to calculate an error matrix, and the accuracy assessment was performed.

4.5 Accuracy Assessment

All LULC classes were checked for accuracy using the standard metrics derived from the confusion matrix. Overall accuracy (OA), the Kappa coefficient, and F1-score are commonly used metrics to evaluate the reliability of LULC classifications. Overall accuracy measures the proportion of correctly classified pixels relative to the total number of reference pixels, providing a straightforward but general assessment of classifier performance (Islami et al. 2022). The Kappa coefficient accounts for chance agreement, offering a more robust evaluation of classification reliability by considering the expected accuracy of a random classification (Shao et al. 2019). The F1-score, which harmonizes both omission and commission errors, delivers a class-level measure of accuracy that balances precision and recall, and is particularly informative for classes that are difficult to discriminate (Chicco and Jurman 2020). The F1-score provides an unbiased measure of the accuracy of a classifier on a single specific class and can be used to quantify performance on potentially unbalanced datasets.

While these metrics provide valuable summaries of classification performance, a detailed examination of these errors helps identify systematic weaknesses, such as confusion between spectrally similar classes (e.g., vegetation and agriculture). By analysing both errors, researchers and planners can better interpret classification outcomes and refine strategies to improve accuracy in future studies.

5 Results

This section presents the quantitative and qualitative results of the LULC classification experiments, detailing the performance of each machine learning classifier across different satellite sensors and feature sets.

5.1 Visual Classification Maps

The classified LULC maps for the Bhopal planning area clearly present the visual distinction with sensor resolution and features used. Figure 3 represents LULC maps produced using Landsat-9, similarly, Fig. S3 and Fig. S4 represents LULC maps produced from Sentinel-2 and LISS-IV data, each using both basic and enhanced feature sets symbolizing the classification result for the four machine learning classifiers: GTB, CART, SVM, and RF. This enables direct visual comparison of how each classifier performs on the same imagery and feature set.

A visual inspection of the LULC maps generated from Landsat-9 data (Fig. 3) reveals notable distinctions in classification patterns, particularly for the built-up and barren land categories. All four models effectively and consistently delineated the water bodies. The ‘RF’ (Fig. 3a) and ‘GTB’ (Fig. 3d) models present a more balanced and visually plausible depiction. They strike a middle ground between the extremes of ‘CART’ and ‘SVM’, showcasing a well-defined urban core without significant over-or under-representation, and a more reasonable distribution of vegetation, barren, and agricultural classes in the surrounding landscape. However, the ‘SVM’ model (Fig. 3b) appears to be more conservative in its classification of urban areas, showing a more fragmented and smaller built-up core. This underestimation corresponds with a significant overestimation of barren land, which is visibly more widespread in the ‘SVM’ map, particularly in the peri-urban fringes where it likely misclassifies sparse residential areas and fallow lands. In stark contrast, the ‘CART’ model (Fig. 3c) exhibits a clear tendency to overestimate the built-up area, resulting in a larger and more contiguous urban footprint compared to the other classifiers. Overall, the visual evidence suggests that the ‘RF’ and ‘GTB’ classifiers produced the most coherent and realistic LULC maps for the Bhopal planning area.

The integration of enhanced features markedly improved the performance of all four classifiers, yielding LULC maps Fig. 3 with greater visual clarity and fewer speckling artifacts. A comparative visual analysis shows that the models

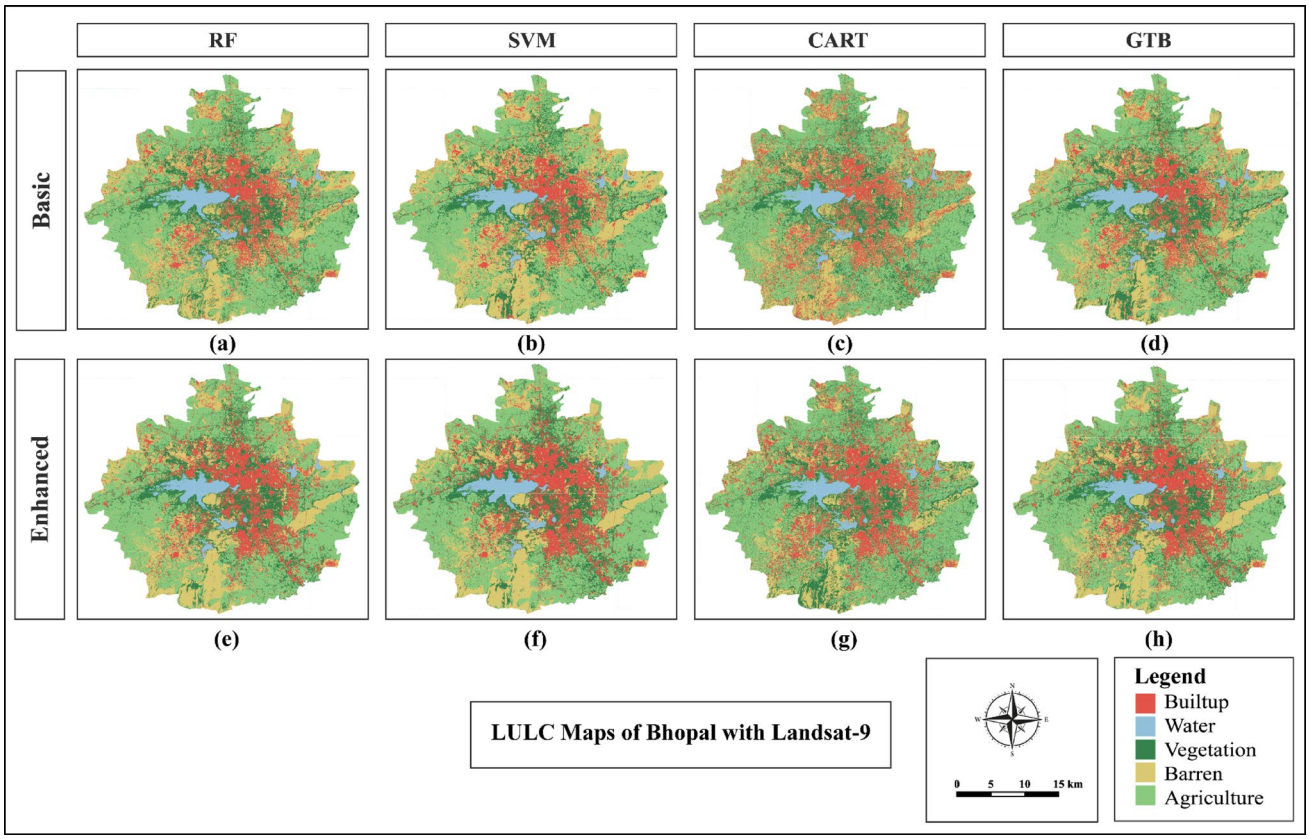


Fig. 3 LULC Maps of Bhopal with Landsat-9, Basic Features- (**a**) RF, (**b**) SVM, (**c**) CART, (**d**) GTB and Enhanced Features (**e**) RF, (**f**) SVM, (**g**) CART, (**h**) GTB

now produce a more consolidated and extensive built-up area, particularly evident in the 'SVM' (Fig. 3f) and 'CART' (Fig. 3g) classifications, which depict the larger urban footprints. Despite this general trend, key differences remain. The 'CART' model (Fig. 3g) is characterized by a significantly larger classification of vegetation, which appears to encroach upon areas classified as barren land by the other models. In contrast, the 'RF' model (Fig. 3h) presents a more conservative estimate of the built-up area, while simultaneously classifying a larger extent of barren land, especially in the transition zones around the city. The 'SVM' (Fig. 3f) and 'GTB' (Fig. 3h) models produced the most visually balanced and coherent results. Although the 'SVM' map shows larger water bodies and a slightly more expansive urban core, both models demonstrate a refined ability to distinguish between complex land covers, an observation strongly corroborated by their high-performance metrics. Ultimately, 'GTB' and 'SVM' appear to generate the most reliable LULC representations when leveraging the enhanced feature set.

Similarly, LULC maps generated from Sentinel-2 data Fig. S3 exhibit a noticeable "salt-and-pepper" effect, indicating challenges in class separability. The visual differences among the classifiers are pronounced. The 'GTB' and 'RF' model offer a more visually balanced and plausible classification (Fig. S3 a & d). While still containing some noise, they avoid the extreme biases of 'SVM' model (Fig. S3 b) and 'CART'(Fig. S3 c), presenting a more moderate and credible distribution of all LULC classes. The 'SVM' model shows a dramatic misclassification between barren land and agriculture. It produces a map with a vast and disproportionate amount of barren land, while the agriculture class is visibly diminished and sparsely represented, suggesting significant spectral confusion. Conversely the 'CART' model demonstrates a clear overestimation of the built-up class, creating an exaggerated and overly agglomerated urban core that is inconsistent with the other models., Overall, 'RF' and 'GTB' appear to handle the spectral complexity of the Sentinel-2 data more effectively than the other two models when using with only basic features.

The application of enhanced features to the Sentinel-2 imagery results in a transformative improvement in LULC classification, producing maps Fig. S3 with significantly reduced noise and much sharper, more coherent class boundaries. The 'RF' and 'GTB' (Fig. S3 e & h) emerge as the top performers, yielding visually impressive and well-balanced maps. They successfully delineate a realistic built-up extent while maintaining a clear and accurate distinction between the often-confused classes of vegetation

and agriculture classes. In contrast, the other two models display specific biases. Both 'SVM' model (Fig. S3 f) and 'CART' model (Fig. S3 g) exhibit a major weakness in separating vegetative classes; they both classify a very large area as vegetation at the expense of the agriculture class, a visual artifact that is confirmed by their weaker performance metrics for these categories. The 'CART' continues its trend of creating the most expansive urban footprint, suggesting an overestimation of the built-up area. Furthermore, Therefore, while all models benefit from the enhanced features, 'RF' and GTB provide a more reliable and accurate representation of the landscape.

Similarly, LULC classifications derived from LISS-IV imagery (Fig. S4) are characterized by a high degree of spectral confusion, resulting in maps with significant "salt-and-pepper" noise and highly divergent outcomes among the classifiers. The 'RF' (Fig. S4 a) and 'GTB' (Fig. S4 d) models offer a more moderate and visually plausible result. Although they are not immune to the salt-and-pepper effect, they provide a much more balanced representation of the urban extent and the surrounding landscape as compare to 'SVM' models (Fig. S4 b) and 'CART' (Fig. S4 c), making them the more reliable classifiers for this particular dataset. The most striking feature of this set is the extreme and opposing performance of the 'SVM' models (Fig. S4 b) and 'CART' (Fig. S4 c). The 'SVM' model exhibits an aggressive overestimation, producing a vast and sprawling built-up classification that appears to have incorrectly absorbed large portions of other classes, particularly barren land. The 'SVM' map also shows a uniquely large vegetation footprint. In complete contrast the 'CART' model drastically underestimates the built-up area, rendering it as a sparse and fragmented collection of pixels that fails to capture the city's contiguous urban core.

The inclusion of enhanced features significantly improves the quality of LULC classification for the LISS-IV data (Fig. S4), transforming the noisy and fragmented results of the basic features into visually coherent and well-defined maps. All four models show substantial improvement, but a clear hierarchy in performance emerges. The 'RF' (Fig. S4 e) and 'GTB' model (Fig. S4 h) deliver excellent, highly plausible results. The 'SVM' model (Fig. S4 f), although visually improved, remains an outlier. It retains its systemic bias, generating the largest built-up area and the most extensive vegetation cover, a visual over-representation that corresponds with lower accuracy scores for those classes. The 'CART' model (Fig. S4 g), which previously struggled, now produces the most accurate classification for the built-up category, presenting a compact and precise urban form. They depict a slightly more extensive built-up area than 'CART' while demonstrating a superior ability to accurately classify vegetation and barren land. In conclusion, for the LISS-IV sensor, the enhanced features enable 'CART', 'RF', and 'GTB' to produce highly reliable and accurate LULC maps, with 'SVM' being a less dependable option due to its persistent classification biases.

5.2 Overall and Class-Level Accuracy Assessment

The quantitative accuracy assessment indicates that enhanced spectral feature sets consistently improved classification performance across all sensors and classifiers (Figs. 4 and 5). Overall accuracy increased by approximately 2.4% to 8.8% following feature enhancement, highlighting the effectiveness of integrating spectral indices and band ratios. Among the tested models, SVM with enhanced features applied to Landsat-9 achieved the highest overall accuracy (93.2%) and Kappa coefficient (0.915), followed closely by RF and GTB. For Sentinel-2 and LISS-IV datasets, RF and GTB with enhanced features consistently outperformed CART and SVM in most cases.

Beyond overall accuracy metrics, the confusion matrices reveal systematic patterns in class-wise classification performance across all sensors (Fig. 6, Fig. S5, and Fig. S6). Water Bodies (Class 1) and Vegetation (Class 2) consistently achieved the highest classification accuracy, reflecting their distinct and stable spectral signatures.

In contrast, persistent misclassification was observed between Built-up (Class 0) and Barren Land (Class 3), and to a lesser extent Agriculture (Class 4), indicating strong spectral overlap in peri-urban and transitional zones. These errors were most pronounced under basic feature configurations, highlighting the limitations of raw spectral bands in complex urban environments.

The incorporation of enhanced spectral features substantially improved class separability by reducing confusion among urban, agricultural, and barren land classes. This improvement was most evident in Landsat-9 SVM and Sentinel-2 and LISS-IV RF and GTB models, which exhibited consistently low off-diagonal errors. The superior performance of these configurations reflects the combined effects of feature enhancement and ensemble or kernel-based learning strategies in mitigating spectral ambiguity.

Overall, the confusion matrix analysis confirms that classification reliability in heterogeneous urban landscapes is strongly dependent on both feature design and classifier robustness, with enhanced feature–classifier combinations providing the most stable and accurate results.

5.3 Class-Specific Accuracies

To provide a detailed understanding of classifier performance at the individual class level, Producer's Accuracy (PA), User's Accuracy (UA), and F1-Score were calculated

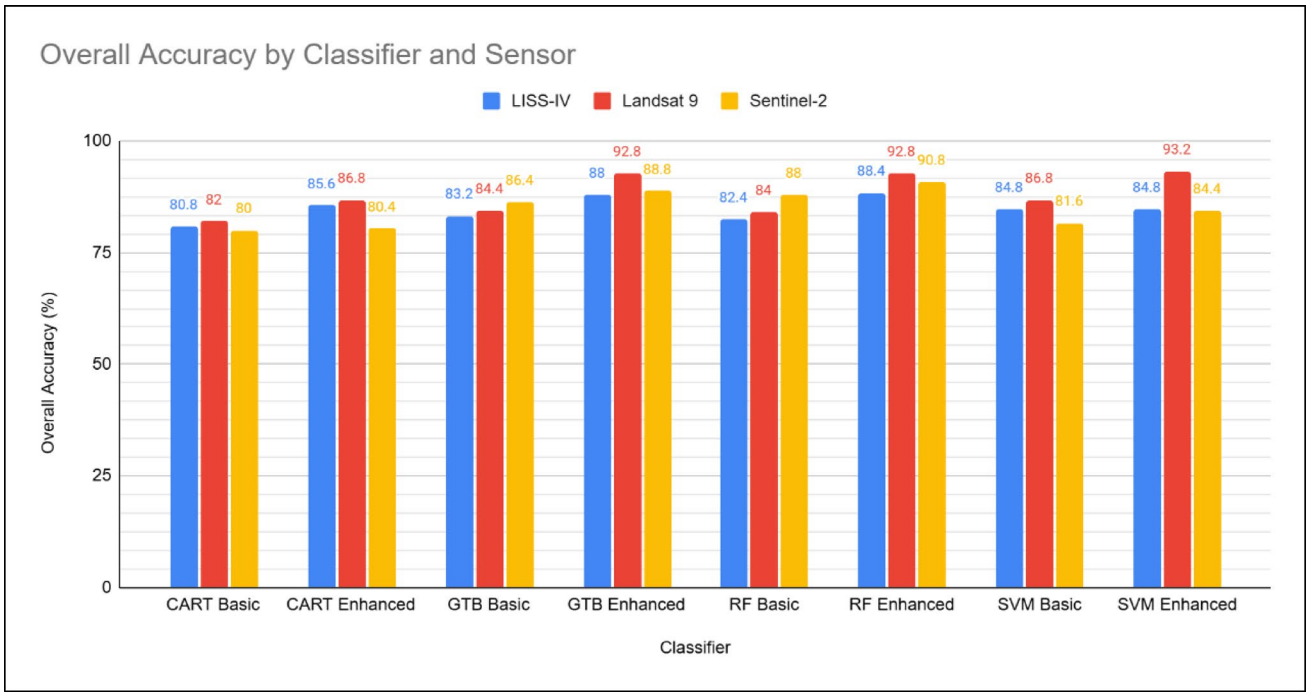


Fig. 4 Overall Accuracy by Classifier and Sensor

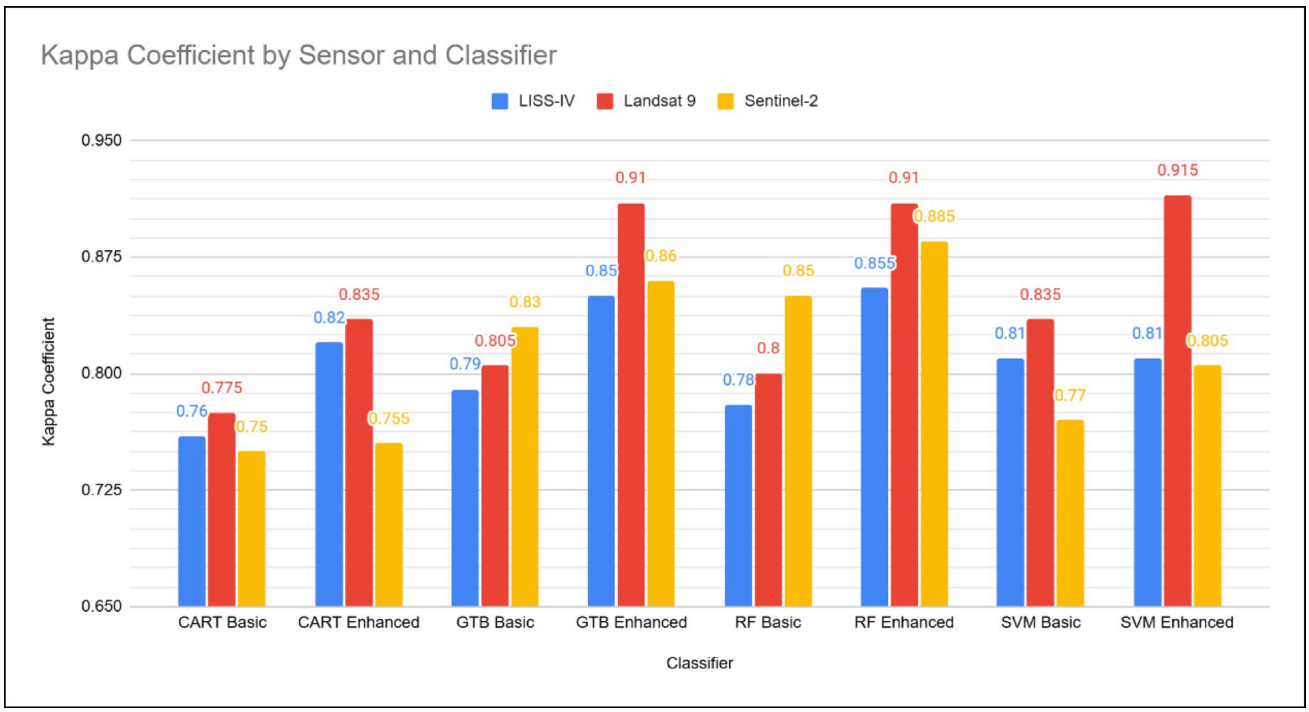


Fig. 5 Kappa Coefficient by Classifier and Sensor

for all models and classes (Figs. 7–9). PA highlights omission errors by quantifying the proportion of reference pixels correctly identified, while UA addresses commission errors, and the F1-score harmonizes both metrics for a comprehensive view of classification reliability. Across all sensors, enhanced spectral features substantially improved classification outcomes (Nicolau et al. 2024). Water Bodies and Vegetation consistently demonstrated high accuracies due to their distinct spectral reflectance, achieving near-perfect PAs (0.94-1.00), UAs (0.98-1.00), and F1-scores

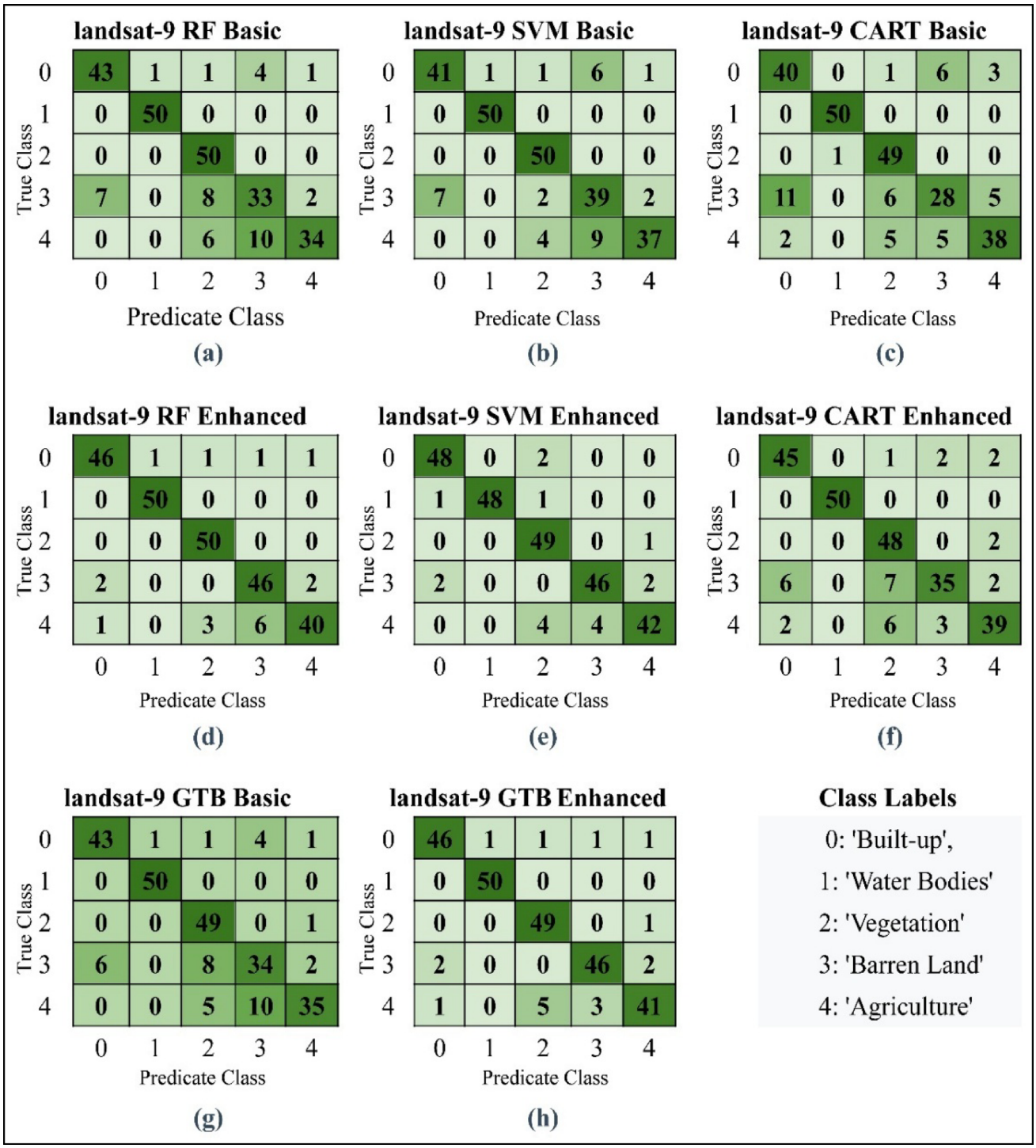


Fig. 6 Confusion Matrix for Landsat 9 LULC Classification, (**a**) RF basic, (**b**) SVM basic, (**c**) CART basic, (**d**) RF enhance, (**e**) SVM enhance, (**f**) CART enhance, (**g**) GTB basic, (**h**) GTB enhance

(0.97-1.00) across all sensors and classifiers. For example, Landsat-9 'SVM' Enhanced achieved 96% PA and 100% UA for Water Bodies (F1 = 0.98), while Sentinel-2 'RF' Enhanced reached 100% for both metrics (F1 = 1.00). Built-up areas and Barren Land showed greater variability due to spectral similarity, but improved markedly with enhanced features. For Landsat-9, Built-up PA increased from 0.82 ('SVM' Basic) to 0.96 ('SVM' Enhanced), with F1-scores rising from 0.83–0.87 to 0.93–0.95. Barren Land exhibited the most dramatic gains, improving from PA values of 0.56 ('CART' Basic) to 0.92 ('SVM' and 'GTB' Enhanced), with F1-scores jumping from 0.63–0.75 to 0.89–0.92. Sentinel-2

Producer Accuracy of Models Across LULC Classes

	Built-up	Water Bodies	Vegetation	Barren Land	Agriculture
Landsat 9 RF Basic	0.86	1	1	0.66	0.68
Landsat 9 RF Enhanced	0.92	1	1	0.92	0.8
Landsat 9 SVM Basic	0.82	1	1	0.78	0.74
Landsat 9 SVM Enhanced	0.96	0.96	0.98	0.92	0.84
Landsat 9 CART Basic	0.8	1	0.98	0.56	0.76
Landsat 9 CART Enhanced	0.9	1	0.96	0.7	0.78
Landsat 9 GTB Basic	0.86	1	0.98	0.68	0.7
Landsat 9 GTB Enhanced	0.92	1	0.98	0.92	0.86
Sentinel-2 RF Basic	0.92	1	0.94	0.86	0.68
Sentinel-2 RF Enhanced	0.9	1	0.94	0.94	0.76
Sentinel-2 SVM Basic	0.9	1	0.94	0.88	0.36
Sentinel-2 SVM Enhanced	0.9	1	0.96	0.9	0.46
Sentinel-2 CART Basic	0.86	1	0.88	0.62	0.64
Sentinel-2 CART Enhanced	0.88	1	0.84	0.8	0.5
Sentinel-2 GTB Basic	0.92	1	0.94	0.82	0.64
Sentinel-2 GTB Enhanced	0.86	1	0.96	0.9	0.78
LISS-4 RF Basic	0.78	1	0.98	0.82	0.54
LISS-4 RF Enhanced	0.9	1	0.98	0.9	0.64
LISS-4 SVM Basic	0.9	1	0.98	0.76	0.6
LISS-4 SVM Enhanced	0.9	1	0.98	0.76	0.6
LISS-4 CART Basic	0.8	0.96	0.94	0.74	0.6
LISS-4 CART Enhanced	0.92	0.96	0.96	0.82	0.62
LISS-4 GTB Basic	0.82	1	0.96	0.8	0.6
LISS-4 GTB Enhanced	0.88	1	0.98	0.92	0.64

Fig. 7 Producer Accuracy of Models Across LULC Classes

User Accuracy of Models Across LULC Classes

	Built-up	Water Bodies	Vegetation	Barren Land	Agriculture
Landsat 9 RF Basic	0.86	0.98	0.77	0.70	0.92
Landsat 9 RF Enhanced	0.94	0.98	0.93	0.87	0.93
Landsat 9 SVM Basic	0.85	0.98	0.87	0.72	0.92
Landsat 9 SVM Enhanced	0.94	1.00	0.88	0.92	0.93
Landsat 9 CART Basic	0.75	0.98	0.80	0.72	0.83
Landsat 9 CART Enhanced	0.85	1.00	0.77	0.88	0.87
Landsat 9 GTB Basic	0.88	0.98	0.78	0.71	0.90
Landsat 9 GTB Enhanced	0.94	0.98	0.89	0.92	0.91
Sentinel-2 RF Basic	0.88	0.98	0.82	0.83	0.89
Sentinel-2 RF Enhanced	0.90	1.00	0.89	0.82	0.95
Sentinel-2 SVM Basic	0.88	1.00	0.82	0.62	0.86
Sentinel-2 SVM Enhanced	0.90	1.00	0.81	0.69	0.88
Sentinel-2 CART Basic	0.81	1.00	0.73	0.74	0.71
Sentinel-2 CART Enhanced	0.86	0.94	0.74	0.70	0.78
Sentinel-2 GTB Basic	0.88	0.98	0.78	0.79	0.91
Sentinel-2 GTB Enhanced	0.86	1.00	0.87	0.82	0.91
LISS-4 RF Basic	0.92	0.98	0.80	0.63	0.87
LISS-4 RF Enhanced	0.96	0.98	0.85	0.76	0.91
LISS-4 SVM Basic	0.85	1.00	0.75	0.79	0.88
LISS-4 SVM Enhanced	0.85	1.00	0.75	0.79	0.88
LISS-4 CART Basic	0.87	0.98	0.75	0.66	0.83
LISS-4 CART Enhanced	0.96	1.00	0.83	0.71	0.82
LISS-4 GTB Basic	0.91	0.98	0.80	0.69	0.81
LISS-4 GTB Enhanced	0.94	0.98	0.82	0.78	0.94

Fig. 8 User Accuracy of Models Across LULC Classes

F1 Score of Models Across LULC Classes

	Built-up	Water Bodies	Vegetation	Barren Land	Agriculture
Landsat 9 RF Basic	0.86	0.99	0.87	0.68	0.78
Landsat 9 RF Enhanced	0.93	0.99	0.96	0.89	0.86
Landsat 9 SVM Basic	0.83	0.99	0.93	0.75	0.82
Landsat 9 SVM Enhanced	0.95	0.98	0.93	0.92	0.88
Landsat 9 CART Basic	0.77	0.99	0.88	0.63	0.79
Landsat 9 CART Enhanced	0.87	1.00	0.85	0.78	0.82
Landsat 9 GTB Basic	0.87	0.99	0.87	0.69	0.79
Landsat 9 GTB Enhanced	0.93	0.99	0.93	0.92	0.88
Sentinel-2 RF Basic	0.90	0.99	0.88	0.84	0.77
Sentinel-2 RF Enhanced	0.90	1.00	0.91	0.88	0.84
Sentinel-2 SVM Basic	0.89	1.00	0.88	0.73	0.51
Sentinel-2 SVM Enhanced	0.90	1.00	0.88	0.78	0.61
Sentinel-2 CART Basic	0.83	1.00	0.80	0.67	0.67
Sentinel-2 CART Enhanced	0.87	0.97	0.79	0.75	0.61
Sentinel-2 GTB Basic	0.90	0.99	0.85	0.80	0.75
Sentinel-2 GTB Enhanced	0.86	1.00	0.91	0.86	0.84
LISS-4 RF Basic	0.85	0.99	0.88	0.71	0.67
LISS-4 RF Enhanced	0.93	0.99	0.91	0.83	0.75
LISS-4 SVM Basic	0.87	1.00	0.85	0.78	0.71
LISS-4 SVM Enhanced	0.87	1.00	0.85	0.78	0.71
LISS-4 CART Basic	0.83	0.97	0.83	0.70	0.70
LISS-4 CART Enhanced	0.94	0.98	0.89	0.76	0.70
LISS-4 GTB Basic	0.86	0.99	0.87	0.74	0.69
LISS-4 GTB Enhanced	0.91	0.99	0.89	0.84	0.76

Fig. 9 F1-Scores of Models Across LULC Classes

and LISS-IV showed similar patterns, with 'RF' and 'GTB' Enhanced models consistently outperforming basic classifiers for these classes. Agriculture remained the most challenging class across all sensors, characterized by lower PAs but comparatively higher UAs, indicating omission errors. Landsat-9 achieved the best results with 'GTB' Enhanced (PA = 0.86, UA = 0.93, F1 = 0.88), while Sentinel-2 showed poor performance under 'SVM' models (PA = 0.36–0.46) but substantial improvement with 'RF' Enhanced (PA = 0.76, UA = 0.95, F1 = 0.84–0.95). LISS-IV, despite its finer spatial resolution, showed only moderate Agriculture classification (PA = 0.54–0.64, UA = 0.91, F1 = 0.67–0.76), even with enhanced features. LISS-IV excelled in Built-up mapping due to its higher spatial resolution, with 'CART' Enhanced achieving PA = 0.92, UA = 0.96, and F1 = 0.94. However, this advantage did not extend uniformly to all classes, as Agriculture and Barren Land remained weaker compared to the other sensors. Overall, the results confirm that enhanced spectral indices significantly improve classification reliability, with the choice of sensor and classifier influencing performance depending on the target land cover class.

6 Discussion

This study presents a comparative analysis of the four machine learning classifiers ('RF', 'SVM', 'CART', 'GTB') and three multi-sensor satellite data sets in the classification of LULC for the dynamic Bhopal planning boundary. The systematic methodology with shared training and testing datasets, along with comparing low-level and high-level spectral features, is helpful in deriving the best recommendations for mapping urban LULC.

6.1 Impact of Feature Engineering

Our findings clearly shows that the use of enhanced feature sets (NDVI, NDWI, and band ratios) considerably enhanced LULC classification accuracy in all sensors and classifiers. This result is consistent with similar other studies that have reported the significance of feature engineering in remote sensing. Spectral indices such as NDVI and NDWI are constructed to highlight particular biophysical characteristics (health of vegetation, water content) which in the original spectral bands might not be conspicuous. The selected band ratios also helped by accentuating distinctive spectral responses, which proved useful in distinguishing spectrally matching urban structures like built-up land and bare land. This adding discriminative power eventually carries forward to greater classification accuracies and enhanced LULC maps. Overall improvement for all the sensor resolutions of the enhanced feature sets also proves their universal applicability for LULC classification.

6.2 Classifier Performance Comparison

Among all the machine learning algorithms tried in this study, 'RF' and 'GTB' were better than 'CART' for all the sensors. On Sentinel-2, 'RF' Enhanced (90.8% OA) and 'GTB' Enhanced (88.8% OA) were superior to 'SVM' and 'CART' by a significant margin. On LISS-IV, 'RF' Enhanced (88.4% OA) and 'GTB' Enhanced (88.0% OA) were superior to all the others. But with Landsat-9, 'SVM' with Enhanced features was found to yield highest overall accuracy (93.2%) as well as Kappa (0.915), with 'RF' and 'GTB' being a close second (92.8% OA, 0.91 Kappa both). Better performance of ensemble methods ('RF', 'GTB') and best-tuned 'SVM' can be accredited because of their capability to deal with complex, high-dimensional, and usually noisy remote sensing data. 'RF', through generating many decision trees and voting their predictions by a process known as bagging, significantly minimizes overfitting and enhances generalization performance by minimizing variance. Likewise, the sequential boosting technique employed by 'GTB', where each new tree continues to refine the errors of the older ones by concentrating on misclassified samples, results in a very strong and accurate model. These multifaceted approaches are especially suited to handle multicollinearity among features and their natural robustness to outliers, which are common problems for spectral data.

Conversely, 'CART' as a standalone decision tree model typically consisted of the lowest accuracy compared to all the sensors (e.g., Sentinel-2 Basic 80.0%, LISS-IV Basic 80.8%, Landsat-9 Basic 82.0%). This comparatively weaker performance can be attributed to the inherent tendency of single decision trees to overfit training data and their limited ability to generalize complex, non-linear class boundaries in high-dimensional feature spaces. like they do with ensemble methods. 'SVM', theoretically strong and lauded for providing consistent performance on the majority of classification problems, performed erratically compared to 'RF' and 'GTB' based on the sensor. Its good performance with Landsat-9 indicates that its hyperparameter search for this particular dataset was outstanding, while with LISS-IV and Sentinel-2, 'RF' and 'GTB' maintained a narrow edge overall performance. 'SVM's sensitivity as a hyperparameter to 'gamma' and 'cost' translates to the fact that poor tuning can have a substantial impact on its performance. Although an initial tuning has already been done, finding the absolute optimum setting for 'SVM' for all the various datasets might not always be feasible without much computational power for exhaustive search approaches .

6.3 Impact of Source and Resolution of Imagery

The effect of spatial resolution and spectral quality on LULC classification accuracy was evident across the datasets. Among the tested combinations, Landsat-9, 'SVM' Enhanced achieved the highest overall accuracy of 93.2%, followed by Sentinel-2 'RF' Enhanced at 90.8%, and LISS-IV 'RF' Enhanced at 88.4%. Although LISS-IV provides a finer native resolution of 5.8 m, this did not translate to the highest classification performance. This result should be interpreted as context-specific. Superior performance of Landsat-9 with 'SVM' Enhanced in the Bhopal case can be attributed to several interacting factors. First, Landsat-9 provides spectrally stable and radiometrically consistent surface reflectance data, which, when combined with enhanced indices and band ratios, produces a feature space well suited to SVM's margin-based optimization. Second, the relatively moderate spatial resolution (30 m) reduces extreme intra-class variability and noise common in very high-resolution imagery, enabling SVM to construct cleaner decision boundaries for dominant urban classes. Third, careful but restrained tuning of SVM hyperparameters allowed effective separation of spectrally overlapping classes such as built-up and barren land in this specific landscape.

While fine spatial resolution can reduce mixed-pixel effects and better capture small urban features, on the other hand spectral quality, calibration, and feature selection play a dominant role in classifier performance. Landsat-9 and Sentinel-2, both atmospherically corrected (using LaSRC and Sen2Cor, respectively), provided high-quality reflectance, enabling more accurate urban LULC classification despite slightly coarser resolution. LISS-IV imagery, atmospherically corrected via Dark Object Subtraction (DOS) at the 1percentile, still performed competitively for fine-scale urban mapping, but spectral limitations and reduced feature richness likely constrained its classification accuracy. Landsat-9 imagery used Blue, Green, Red, and NIR bands, which are well-documented for discriminating vegetation, water, and built-up areas. The enhanced feature set (NDVI, NDWI, band ratios B4/B3 and B3/B2) further amplifies contrasts between vegetation, soil, water, and built-up surfaces. LISS-IV, with only three bands (B2–B4), provides fewer opportunities for effective spectral indices, reducing class separability. Ensemble classifiers like 'RF' and 'GTB' benefit from "cleaner" training signatures provided by Landsat's aggregated pixels. High-resolution imagery such as Sentinel-2 and LISS-IV may capture more spectral variability (including noise and sub-pixel heterogeneity), which can reduce the stability of classifiers like 'SVM' or 'CART', lowering overall accuracy.

Overall, these results emphasize that sensor selection for urban LULC mapping should balance spatial resolution with spectral quality, atmospheric correction, and feature design, while carefully matching classifiers to data characteristics to optimize performance.

6.4 Limitations of the Study

This research is carefully crafted comparative framework, comparing several sensors, classifiers, and feature sets over a uniform dataset in robust GEE framework. Systematic design ensures validity and comparability of results. Addition of a specific urban planning boundary offers highly pertinent results for regional development and governance efforts. Use of the GEE further promotes reproducibility and scalability of the methodology.

Nevertheless, the research is not without some limitations. Manual tuning through initial preliminary selection was involved in selecting the 'SVM' hyperparameters; more precise optimization could possibly identify an even greater maximum 'SVM' performance for all sensors. Although the LISS-IV data were in fact acquired at 5.8 m, being pre-ingested as independent assets with limited insight into its intrinsic atmospheric correction and composite assembly within the GEE script. Additionally, only one date of acquisition in March 2025 was accounted for by the study. Integration with multi-temporal imagery and time-series analysis could possibly enhance discrimination among some LULC classes, particularly those seasonally dependent (i.e., agriculture), and make change detection analysis easier.

The results should be interpreted within the geographical and environmental context of the study area, which represents a mid-sized, inland Indian city characterized by a mix of built-up areas, vegetation, agriculture, barren land, and inland water bodies under a tropical monsoon climate.

Although the relative trends observed in this study—such as the consistent improvement achieved through enhanced feature sets and the robust performance of ensemble classifiers—are likely to hold across many urban environments, the absolute accuracy values and optimal sensor–classifier combinations should be considered context-dependent. Factors such as local climate, seasonal variability, dominant construction materials, landscape fragmentation, and the proportion of mixed pixels can significantly influence outcomes.

7 Conclusions

This study successfully conducted a comprehensive comparative analysis of machine learning classifiers and multi-sensor satellite imagery for LULC classification over the Bhopal planning boundary using Google Earth Engine. The findings of this study confirm that the strategic integration

of enhanced spectral feature sets (NDVI, NDWI, and band ratios) significantly boosts classification accuracy across all evaluated sensors and machine learning algorithms. Among the tested classifiers, 'SVM' achieved the highest overall accuracy for Landsat-9 (93.2%), while 'RF' and 'GTB' consistently delivered superior performance for Sentinel-2 and LISS-4, attributable to their ensemble learning capabilities and robustness. The study also highlighted the critical role of spatial resolution, with higher-resolution imagery generally yielding more accurate LULC maps for the heterogeneous urban landscape of Bhopal.

Beyond reporting classification accuracies, this study makes three substantive contributions to urban remote sensing research. First, it demonstrates that classifier performance rankings are not universal, but are strongly conditioned by sensor resolution and feature space, with SVM outperforming ensemble methods only under specific spectral–spatial configurations. Second, it empirically establishes that feature enhancement can compensate for coarser spatial resolution, enabling moderate-resolution sensors such as Landsat-9 to rival or exceed higher-resolution datasets when paired with optimal classifiers. Third, the study provides a methodologically rigorous and fully reproducible multi-sensor benchmarking framework implemented entirely on Google Earth Engine, which can be readily transferred to other cities with similar urban environments.

This research finds operational utility for urban planners and remote sensing analysts involved in the LULC mapping process, especially for cities like Bhopal and other similar cities that are rapidly expanding. In reality, Random Forest, Gradient Tree Boosting, and SVM optimally fine-tuned with enhanced feature sets prove to perform better-a good method for the difficult tasks of accurate LULC mapping in dynamically expanding cities.

For detailed urban planning and precise feature identification, such as delineating specific land parcels for zoning or infrastructure projects, a combination of higher-resolution imagery like LISS-IV (5.8 m) or Sentinel-2 (10 m) and robust machine learning classifiers (RF, GTB, or optimized SVM) with an enhanced feature space is strongly recommended. As an example, the accuracy obtained for Built-up class with Landsat-9 'SVM' Enhanced (96% producer's accuracy, 94.12% user's accuracy) and LISS-IV 'RF' Enhanced (90% producer's accuracy, 95.7% user's accuracy) can be used in monitoring accurate urban sprawl and informing smart growth policies. Enhanced discrimination of Barren Land from Built-up land, especially with increased resolution, can be used in the planning of areas for future development or rehabilitation.

Nevertheless, for regional-scale analysis over large areas, change detection of big areas, or where high-resolution data is not available, or when it is computationally costly, Landsat-9 (30 m) and Sentinel-2 (10 m) are a good compromise between spatial resolution and data availability, particularly with enhanced features and advanced classifiers. The proven effectiveness of GEE for such comparative analysis is validating its potential as a valuable tool for operational LULC mapping and monitoring to facilitate swift evaluation and decision-making toward sustainable urbanization for Bhopal and other such cities.

Future work could expand this research by exploring more advanced deep learning techniques for LULC classification, integrating additional geospatial data sources such as Synthetic Aperture Radar (SAR) or LiDAR, and conducting multi-temporal analyses to capture LULC dynamics and change detection more effectively. Furthermore, conducting statistical significance tests on classifier performance differences would add further rigor to comparative analyses. Further investigation into systematic hyperparameter optimization for all classifiers across different datasets could also yield additional performance gains. future research could also explore the sensitivity of classifier performance to different data split ratios (e.g., 70:30, 80:20) through cross-validation approaches. Such analysis would provide insights into the minimum training data requirements for each classifier and help optimize data collection efforts for operational LULC mapping.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41748-026-01126-2.

Acknowledgements The authors gratefully acknowledge the support provided by Maulana Azad National Institute of Technology (MANIT) Bhopal for access to research infrastructure and computational facility.

Author Contributions All authors contributed to the study. **Mr. Adarsh Agrawal**, prepared the material, collected data analyzed and wrote the first draft of the manuscript. **Dr. Priyamitra Munoth** contributed to concept & design, analyzing tools and methodology, provided inputs in writing the manuscript and editing. **Dr. Vikas Poonia** contributed to concept & design, analyzing tools and methodology, writing the manuscript and editing. **Prof. Lixin Wang** contributed to writing the manuscript and editing. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Competing interests The authors declare that they have no conflicting interests.

References

Abdi AM (2020) Land cover and land use classification performance of machine learning algorithms in a boreal landscape using

Sentinel-2 data. GIScience Remote Sens 57(1):1–20. https://doi.org/10.1080/15481603.2019.1650447
Amin G et al (2024) Assessment of Machine Learning Algorithms for Land Cover Classification in a Complex Mountainous Landscape. J Geovisualization Spat Anal 8(2):34. https://doi.org/10.1007/s41651-024-00195-z
Baccari N et al (2025) Assessment of Machine Learning Techniques in Mapping Land Use/Land Cover Changes in a Semi-Arid Environment. Earth Syst Environ 9(2):519–539. https://doi.org/10.1007/s41748-024-00562-2
Barpete K, Mehrotra S (2023) Climate-Informed Planning through Mapping of Urban Thermal Load and Cooling Potential: Case of Tropical City of Bhopal. J Indian Soc Remote Sens 51(7):1375–1391. https://doi.org/10.1007/s12524-023-01710-3
Belgiu M, Drăguț L (2016) Random forest in remote sensing: A review of applications and future directions. ISPRS J Photogrammetry Remote Sens 114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011
Blanco V, Japón A, Puerto J (2023) Multiclass optimal classification trees with SVM-splits. Mach Learn 112(12):4905–4928. https://doi.org/10.1007/s10994-023-06366-1
Boothroyd RJ et al (2021) Applications of Google Earth Engine in fluvial geomorphology for detecting river channel change. WIREs Water 8(1). https://doi.org/10.1002/wat2.1496
Chavez PS (1988) An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sens Environ 24(3):459–479. https://doi.org/10.1016/0034-4257(88)90019-3
Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21(1):6. https://doi.org/10.1186/s12864-019-6413-7
Cortes C, Vapnik V, Saitta L (1995) Support-Vector Networks Editor, Machine Leaming. Kluwer Academic
Das J, Poonia V, Jha S, Goyal MK (2020) Understanding the climate change impact on crop yield overEastern Himalayan Region : ascertaining GCM and scenario uncertainty Understanding the climate change impact on crop yield over EasternHimalayan Region : ascertaining GCM and scenario uncertainty. https://doi.org/10.1007/s00704-020-03332-y
Dabija A et al (2021) Comparison of support vector machines and random forests for corine land cover mapping. Remote Sens 13(4):1–35. https://doi.org/10.3390/rs13040777
European Space Agency (ESA) (2015) Sentinel-2 User Handbook
Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. Pattern Recognit Lett 27(4):294–300. https://doi.org/10.1016/j.patrec.2005.08.011
Islami FA et al (2022) Accuracy Assessment of Land Use Change Analysis Using Google Earth in Sadar Watershed Mojokerto Regency, IOP Conference Series: Earth and Environmental Science, 950(1), 012091. https://doi.org/10.1088/1755-1315/950/1/012091
Lan LT, Vinh TQ, Giang (2025) A Comparison among Different Machine Learning Algorithms in Land Cover Classification Based on the Google Earth Engine Platform: The Case Study of Hung Yen Province, Vietnam. J Environ Earth Sci 7(1):132–139. https://doi.org/10.30564/jees.v7i1.6652
Main-Knorn M et al (2017) Sen2Cor for Sentinel-2, in, Image and Signal Processing for Remote Sensing XXIII, Proceedings of SPIE, Vol. 10427, 1042704. https://doi.org/10.1117/12.2278218
Mancino G et al (2023) Comparison between Parametric and Non-Parametric Supervised Land Cover Classifications of Sentinel-2 MSI and Landsat-8 OLI Data. Geographies 3(1):82–109. https://doi.org/10.3390/geographies3010005
Mao W et al (2020) Comparison of machine-learning methods for urban land-use mapping in hangzhou city, china. Remote Sens 12(17):1–17. https://doi.org/10.3390/rs12172817
Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. Int J Remote Sens 39(9):2784–2817. https://doi.org/10.1080/01431161.2018.1433343
Mehra N, Swain JB (2024) Assessment of land use land cover change and its effects using artificial neural network-based cellular automation. J Eng Appl Sci 71(1):70. https://doi.org/10.1186/s44147-024-00402-0
Mukherjee A, Poonia V, Swarnkar S (2025) Assessing compound flood drivers in Peninsular India: multivariate copula-based approach. J Environ Manag 395:127875. https://doi.org/10.1016/j.jenvman.2025.127875
Munoth P, Goyal R (2020) Impacts of land use land cover change on runoff and sediment yield of Upper Tapi River Sub-Basin, India. Int J River Basin Manage 18(2):177–189. https://doi.org/10.1080/15715124.2019.1613413
Nicolau PA et al (2024) Accuracy Assessment: Quantifying Classification Quality. in Cloud-Based Remote Sensing with Google Earth Engine. Springer International Publishing, Cham, pp 135–145. https://doi.org/10.1007/978-3-031-26588-4_7.
Ouchra H, Belangour A, Erraissi A (2023) Comparison of Machine Learning Methods for Satellite Image Classification: A Case Study of Casablanca Using Landsat Imagery and Google Earth Engine. J Environ Earth Sci 5(2):118–134. https://doi.org/10.30564/jees.v5i2.5928
Pal M (2005) Random forest classifier for remote sensing classification. Int J Remote Sens 26(1):217–222. https://doi.org/10.1080/01431160412331269698
Pande CB et al (2024) Characterizing land use/land cover change dynamics by an enhanced random forest machine learning model: a Google Earth Engine implementation. Environ Sci Europe 36(1):84. https://doi.org/10.1186/s12302-024-00901-0
Pandey PC et al (2021) Land use/land cover in view of earth observation: data sources, input dimensions, and classifiers—a review of the state of the art. Geocarto Int 36(9):957–988. https://doi.org/10.1080/10106049.2019.1629647
Poonia V, Das J, Goyal MK (2021a) Impact ofclimate change on crop water and irrigation requirements over eastern Himalayan region. Stochastic Environ Res RiskAssess 6(Ipcc 2014). https://doi.org/10.1007/s00477-020-01942-6
Poonia V, Goyal MK, Gupta BB, Gupta AK, Jha S, Das J(2021b) Drought occurrence in Different River Basins of India and blockchain technology based framework for disaster management. J Cleaner Product 312:127737. https://doi.org/10.1016/j.jclepro.2021.127737
Poonia V, Jha S, Goyal MK (2021c) Copula basedanalysis of meteorological, hydrological and agricultural drought characteristics across Indian river basins. Int J Climatolvol 7091. https://doi.org/10.1002/joc.7091
Poonia V, Jha S, Srinivas VV, Wang L (2024) Spatiotemporal characteristics and triggers offlash droughts across all the river basins in India. J Hydrometeorol 25(9):1357–1369. https://doi.org/10.1175/JHM-D-23-0080.1
Poonia V, Goyal MK, Jha S, Dubey S (2022) Terrestrial ecosystem response to flash droughts over India. J Hydrol 605:127402. https://doi.org/10.1016/j.jhydrol.2021.127402
Poonia V, Mukherjee A, Dev A, Somil S (2025) Clustering-based assessment oflong-term surface water scarcity and per capita vulnerability in arid India
Rizkallah LW (2025) Enhancing the performance of gradient boosting trees on regression problems. J Big Data 12(1):35. https://doi.org/10.1186/s40537-025-01071-3

Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J Photogrammetry Remote Sens 67:93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Sagi O, Rokach L (2018) Ensemble learning: A survey. WIREs Data Min Knowl Discov 8(4). https://doi.org/10.1002/widm.1249

Shafizadeh-Moghadam H et al (2021) Google Earth Engine for large-scale land use and land cover mapping: an object-based classification approach using spectral, textural and topographical factors. GIScience Remote Sens 58(6):914–928. https://doi.org/10.1080/15481603.2021.1947623

Shao G, Tang L, Liao J (2019) Overselling overall map accuracy misinforms about research reliability. Landscape Ecol 34(11):2487–2492. https://doi.org/10.1007/s10980-019-00916-6

Sharma R, Ghosh A and Joshi PK (2013) Decision tree approach for classification of remotely sensed satellite data using open source support. J Earth Syst Sci 122(5):1237–1247. https://doi.org/10.1007/s12040-013-0339-2

Sharma M, Kumar V, Kumar S (2024) A systematic review of urban sprawl and land use/land cover change studies in India. Sustainable Environ 10(1). https://doi.org/10.1080/27658511.2024.2331269

Singh G, Pandey A (2021) Evaluation of classification algorithms for land use land cover mapping in the snow-fed Alaknanda River Basin of the Northwest Himalayan Region. Appl Geomatics 13(4):863–875. https://doi.org/10.1007/s12518-021-00401-3

Tan J et al (2021) MLAs land cover mapping performance across varying geomorphology with Landsat OLI-8 and minimum human intervention. Ecol Inf 61. https://doi.org/10.1016/j.ecoinf.2021.101227

U.S. Geological Survey (USGS) (2025) How do I use a scale factor with Landsat Level-2 science products? U.S. Geological Survey (USGS)

Verma D, Jana A (2019) LULC classification methodology based on simple Convolutional Neural Network to map complex urban forms at finer scale: Evidence from Mumbai. https://doi.org/10.48550/arxiv.1909.09774. https://doi.org/

Vermote E et al (2018) LaSRC (Land Surface Reflectance Code): Overview, application and validation using MODIS, VIIRS, LANDSAT and Sentinel 2 data's, in IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 8173–8176. https://doi.org/10.1109/IGARSS.2018.8517622

Wainer J, Fonseca (2020) How to tune the RBF SVM hyperparameters? An empirical evaluation of 18 search algorithms. http://arxiv.org/abs/2008.11655

Zurqani HA (2025) A multi-source approach combining GEDI LiDAR, satellite data, and machine learning algorithms for estimating forest aboveground biomass on Google Earth Engine platform. Ecol Inf 86:103052. https://doi.org/10.1016/j.ecoinf.2025.103052