Spatiotemporal Analysis of Multi-Decadal Land Use and Land Cover Dynamics Using GIS-Integrated Machine Learning Algorithms

Amitansu Pattanaik[1*] and Saurabh Srivastava [2]

[1*]Sharda School of Basic Sciences and Research, Sharda University, Agra,Uttar Pradesh 282007, India .
[2]University Institute of Engineering, Chandigarh University, Mohali, Punjab 140413,India .

*Corresponding author(s). E-mail(s): amitansu@yahoo.com ;
Contributing authors: srivastavasaurabh978@gmail.com;

Abstract

Monitoring land use and land cover (LULC) changes is crucial for comprehending ecological dynamics in sensitive ecosystems like the Eastern Sunderban Mangrove Forest. This study analyzes shifts in LULC from 2010 to 2024, utilizing satellite data from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 8. By integrating Geographic Information Systems (GIS) with machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF), we were able to detect substantial transformations in land use across the study area.The focus of our analysis was on changes in mangrove coverage, water bodies, and barren lands over the 14-year period. The performance of the machine learning models was evaluated based on their classification accuracy, with Random Forest achieving the highest accuracy at 92.5%, followed closely by SVM at 91.82% and KNN at 89.62%. These results illustrate the effectiveness of employing GIS-integrated machine learning for LULC analysis and underscore the increasing vulnerability of the Sunderban Mangroves due to anthropogenic pressures and climate change. By providing insights into these ongoing changes, this research contributes to the broader understanding of environmental management and conservation strategies in ecologically sensitive regions.

Keywords: Spatiotemporal, Land Cover, Ecological Vulnerability, Algorithms.

1 Introduction

Changes in Land Use and Land Cover (LULC) are critical in shaping environmental processes and maintaining ecological balance. Such transformations can have profound effects on biodiversity, water availability, and the overall health of ecosystems. It is especially important to monitor and understand LULC changes in sensitive ecosystems, such as mangrove forests, which act as essential buffers against climate change, provide habitats for various species, and sustain local communities' livelihoods. The Sunderban Mangrove Forest, spanning Eastern India and Bangladesh, ranks among the largest and most biodiverse mangrove systems worldwide. However, this unique ecosystem is increasingly threatened by human activities, climate change, and rising sea levels, which pose significant ecological and socio-economic challenges.Rising anthropogenic pressures, such as agricultural expansion, urban development, and industrial growth, have led to notable changes in the land cover of the Sunderban area. Recent progress in remote sensing technologies and Geographic Information Systems (GIS) has transformed our capability to monitor LULC changes across extensive spatial and temporal dimensions. Satellite imagery from tools like Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 8 provides valuable data for analyzing land cover dynamics, enabling researchers to observe LULC changes with high spatial resolution and accuracy, thereby offering critical insights into environmental changes in vulnerable areas.This study aims to perform a spatiotemporal analysis of LULC changes in the Eastern Sunderban Mangrove Forest from 2010 to 2024. Utilizing Landsat 7 ETM+ and Landsat 8 satellite data, we seek to evaluate the scale and trends of LULC changes in relation to both natural phenomena and human-induced factors. By applying machine learning algorithms—K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF)—we will gain insights into classifying land cover categories and assess the degree of mangrove forest degradation, urbanization, and land reclamation. Additionally, we will evaluate model performance using metrics such as precision, overall accuracy, Kappa coefficient, and recall to ensure reliable classification outcomes. The findings of this study will enhance our understanding of LULC dynamics in the Sunderban region and provide essential data to guide conservation efforts and sustainable land-use strategies.

2 Related Work

Debajit Deb's study [1] examines coastal land use and land cover (LULC) changes in the Indian Sundarbans using remotely sensed data. Deb identifies significant transformations in the region, including a shift from mangrove swamp ecosystems to agriculture and tourism-based developments. The study reports an increase in dense mangrove coverage, along with a reduction in open mangrove areas, highlighting crucial ecological changes. Deb stresses the need for sustainable management of mangrove ecosystems to prevent further degradation. He advocates for higher-resolution imaging, such as optical and RADAR-based technologies, to enhance accuracy in monitoring land cover changes. The study also recommends using air-borne hyperspectral data to improve species-level mapping and better understand mangrove dynamics in inaccessible regions of the Sundarbans. Shouraseni Roy et al. [2] conduct a comprehensive

study on coastal wetland transformations in the Sundarban Delta, focusing on LULC changes between 1999 and 2020. They uncover a distinct north-south divide, where mangroves dominate the southern regions, while pond aquaculture expands in the north. The study attributes these land-use patterns to local policies and human activities. Future projections indicate that warmer sea surface temperatures could enhance cyclone formation, posing a significant threat to the region's delicate ecosystem. Roy and colleagues emphasize the necessity of sustainable management practices to mitigate the impacts of climate change and human activities on the Sundarbans. They also suggest incorporating more detailed time series data in future research for a comprehensive analysis of land-use changes.Talukdar et al. [3] evaluate the accuracy of various machine-learning classifiers for LULC mapping using satellite data. Their findings demonstrate that the random forest (RF) algorithm achieves the highest accuracy (0.89), outperforming the support vector machine (SVM) algorithm (0.84). Conducted in a riparian landscape along the Ganga River, the study utilized Landsat 8 data for classification. The authors conclude that RF is the most suitable classifier for LULC modeling in highly dynamic charland areas and recommend future research to assess classifier performance in different morphoclimatic conditions.Junye Wang and colleagues [4] focus on the application of machine learning in modeling LULCC, exploring the challenges and limitations in predicting these processes. They employ artificial neural networks and deep learning algorithms to analyze and predict land use changes, emphasizing the importance of integrating environmental and socio-economic factors. The study also highlights the environmental impacts of land use changes, such as the urban heat island effect and carbon emissions, and calls for sustainable urban development practices.Andrea Vizzari [5] develops and tests an object-oriented classification approach by combining SNIC, GLCM, and machine learning algorithms to classify LULC using remote sensing data. The study tests this approach across different datasets, revealing better results on higher-resolution data, with SVM performing better on lower-resolution data. Vizzari's work highlights the computational demands of processing higher-resolution datasets but demonstrates the reliability of the proposed methodology.Jagannath Aryal's team from the University of Melbourne [6] introduces a novel approach for LULC classification by combining spatial, statistical, and index-based features. Their study, conducted in Melbourne, Australia, shows that the Random Forest algorithm outperforms others with an F1-score exceeding 0.99. The authors underscore the robustness of traditional machine learning (ML) algorithms compared to deep learning (DL)-based methods and emphasize the need for explainable schemes to improve ML performance.Денис Кривогуз [7] explores machine learning models for LULC classification in the Kerch Peninsula. The study reports that the deep neural network model achieved the highest accuracy of 96.2%, outperforming other models. Although limited by data sources, the study emphasizes the potential of using such models for environmental management and calls for future research incorporating field studies and ensemble methods to enhance accuracy.Prachi Chachondhia's team [8] evaluates the performance of machine learning algorithms for LULC classification using optical and microwave data in Doiwala City, India. Their findings indicate that fused datasets achieve better classification results, with overall

accuracy ranging from 85.71% to 93.43% using RF and SVM classifiers. The study suggests that combining data sources improves classification accuracy and recommends future exploration of fused datasets for LULC mapping.

Yashon Ouma et al. [9] compare four machine learning algorithms (CART, RF, GTB, and SVM) for LULC classification using Landsat data from 1984 to 2020. Their results show that RF and SVM are most effective for mapping built-up areas, while GTB and SVM perform better for mapping water bodies. They propose integrating the best results from each classifier to enhance overall mapping accuracy.Anubhava Srivastava et al. [10] compare the performance of four machine learning algorithms (CART, RF, GTB, and SVM) for LULC classification near the Himalayan foothills using multi-temporal satellite data. RF and GTB show the highest accuracy rates, while SVM has comparatively lower performance. The study highlights the utility of error matrix metrics for accuracy assessment and recommends future research focusing on classifier performance across diverse regions.Sameer Mandal et al. [11] analyze LULC changes in the Pare River Basin, Arunachal Pradesh, India, using machine learning techniques. Their study projects future LULC scenarios and identifies driver variables, revealing a decrease in forest areas and an increase in built-up and cropland areas. Mandal's research provides insights for decision-makers to guide policy development toward sustainable land management practices.Andromachi Chatziantoniou et al. [12] utilize Sentinel-1 and Sentinel-2 data with machine learning classifiers to map wetlands. Their study achieves high classification accuracy by incorporating spectral and textural information, particularly for distinguishing vegetation types. They suggest that this methodology shows promise for long-term monitoring of wetlands and call for further validation across different wetland types.Ramesh Singh et al. [13] apply machine learning algorithms to LANDSAT images for mapping LULC in India. Their approach achieves an accuracy of 80-86%, demonstrating the potential of machine learning in improving mapping accuracy and consistency. The study concludes that these methods are cost-effective and scalable for other regions.Landa Sankarrao et al. [14] project a decrease in forest area by 9.02% and an increase in agricultural land by 8.74% for the Nagavali River Basin by 2030. They find that the hybrid MLP-MC-CA model performs best, with overall accuracy values greater than 80%. The study emphasizes the need for accurate land-use predictions for effective water resource planning.Abdulla Al Kafy et al. [15] focus on LULC and land surface temperature (LST) changes in Chattogram, Bangladesh. The study predicts a rise in LST due to LULC changes and underscores the importance of sustainable urban planning to mitigate urban heat islands. The authors highlight the role of cellular automata and artificial neural network models in simulating future scenarios.Hemant Pokhariya et al. [16] evaluate different classifiers for LULC mapping in Nainital, India, using Sentinel-2 data. Their findings show that RF achieves the highest accuracy, and the integration of spectral indices like NDVI and EVI further improves classification accuracy. The study stresses the importance of choosing suitable classifiers and indices for heterogeneous landscapes.

3 Study Area and Data

The research focuses on the Eastern Sunderban Mangrove Forest, a UNESCO World Heritage Site located within the delta region of the Sundarbans, which stretches across the southern part of West Bengal, India. This area is notable for its exceptional biodiversity and complex ecosystems, featuring vast mangrove forests, tidal waterways, and numerous islands. The Eastern Sunderban Mangrove Forest encompasses around 4,262 square kilometers, positioned between latitudes 21.5° N and 22.5° N, and longitudes 88.5° E and 89.5° E. The Bay of Bengal borders it to the south, significantly impacting the tidal dynamics and ecological characteristics of the area.Tracking land use and land cover (LULC) changes in the Eastern Sunderban Mangrove Forest is crucial for grasping the environmental dynamics and ecological well-being of this region. By examining LULC changes during the years 2010, 2017, and 2024, this research aims to shed light on the ongoing alterations affecting the mangrove ecosystem, providing valuable information for conservation and management efforts.

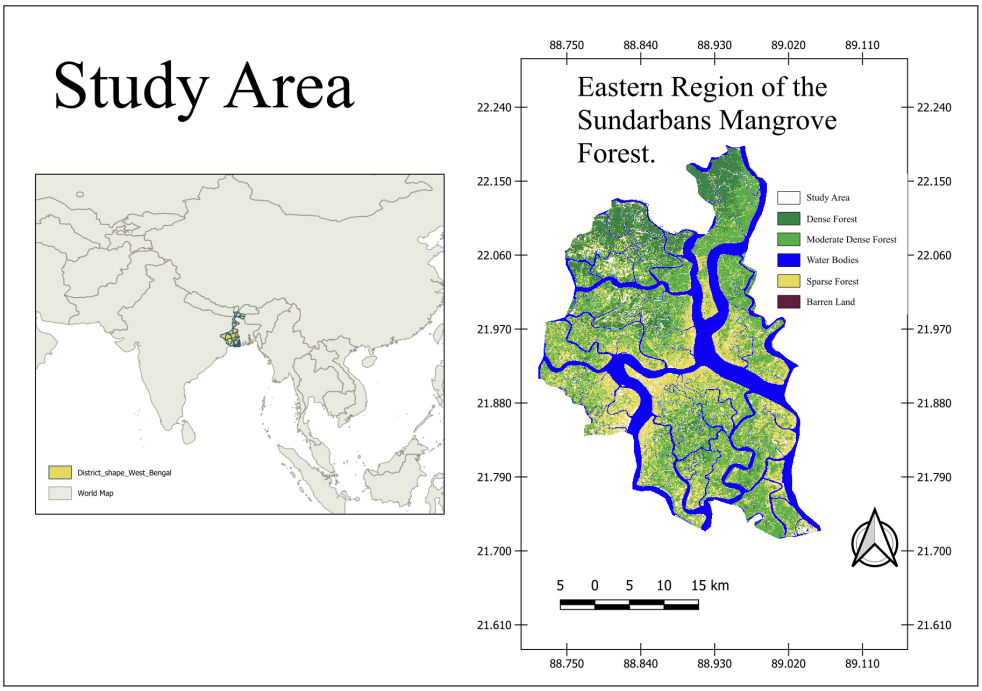

Fig. 1 Study Area Map of Easter Region of Sunderbans

Table 1 Classes and Descriptions

Classes	Description
Dense forest	Areas with high tree density, characterized by a closed canopy and rich biodiversity.
Moderate dense forest	Forests with a moderate number of trees and some gaps in the canopy, supporting a mix of species.
Water bodies	Includes rivers, lakes, ponds, and other aquatic environments, playing a crucial role in the ecosystem.
Sparse forest	Forested areas with low tree density, often featuring open spaces and a variety of undergrowth.
Barren land	Land that is devoid of vegetation, including exposed soil, rock, and areas affected by human activities.

The dataset employed in this research encompasses satellite imagery from the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 8 Operational Land Imager (OLI) sensors, focusing on three significant periods: 2010, 2017, and 2024. These specific years were selected to evaluate long-term changes in land use and land cover (LULC) within the ecologically critical Eastern Sunderban Mangrove Forest. The imagery was sourced from the United States Geological Survey (USGS) Earth Explorer platform, ensuring the acquisition of high-quality data appropriate for spatiotemporal analysis.To maintain the integrity and dependability of the analysis, only satellite images exhibiting cloud cover of 20% or less were utilized. This requirement was essential to reduce cloud interference and ensure visibility of key features such as mangrove forests, water bodies, and urban areas. A meticulous screening process was implemented using the cloud mask function in the Landsat Collection 1 Level-1 dataset available on the USGS Earth Explorer portal. The 2010 imagery was obtained from Landsat 7 ETM+, which provides multispectral data across seven spectral bands, encompassing visible light, near-infrared (NIR), and shortwave infrared (SWIR) sections of the electromagnetic spectrum. Additionally, this sensor features a panchromatic band with a spatial resolution of 15 meters, allowing for enhanced visual interpretation through high-resolution grayscale imagery.For the years 2017 and 2024, the research relied on Landsat 8 OLI imagery, which boasts significant enhancements compared to Landsat 7. Landsat 8 features 11 spectral bands with a higher radiometric resolution of 12 bits, facilitating better differentiation of surface characteristics. The OLI sensor captures data in visible, NIR, and SWIR wavelengths, incorporating additional bands for cirrus detection and two thermal infrared (TIR) bands. The spatial resolution of the multispectral bands remains at 30 meters, while the panchromatic band maintains a resolution of 15 meters, consistent with Landsat 7. Furthermore, Landsat 8 imagery provides an improved signal-to-noise ratio, contributing to the accuracy of LULC classification. Preprocessing of the images was conducted to eliminate cloud cover through the Fmask (Function of Mask) algorithm, which effectively removes cloud and shadow pixels from the dataset.

Table 2 Landsat 8 Band Characteristics

Band Number	Description	Spatial Resolution (m)	Band Width
1	Coastal/Aerosols	30	20
2	Blue	30	65
3	Green	30	30
4	Red	30	40
5	Near Infrared (NIR)	30	85
6	Shortwave Infrared (SWIR1)	30	140
7	Shortwave Infrared (SWIR2)	30	180
8	Panchromatic	15	115
9	Cirrus	30	20
10	Thermal Infrared 1 (TIRS1)	100	10000
11	Thermal Infrared 2 (TIRS2)	100	10500

Table 3 Landsat 7 ETM+ Band Characteristics

Band Number	Description	Spatial Resolution (m)	Band Width
1	Blue	30	60
2	Green	30	70
3	Red	30	60
4	Near Infrared (NIR)	30	120
5	Shortwave Infrared (SWIR1)	30	140
6	Thermal	60	10500
7	Shortwave Infrared (SWIR2)	30	200
8	Panchromatic	15	180

4 Methodology

The methodology utilized in this research encompasses satellite image processing, machine learning classification, and Geographic Information System (GIS) integration to analyze multi-temporal satellite imagery from the years 2010, 2017, and 2024. Satellite images from Landsat 7 and Landsat 8 for these years were subjected to processing and classification through machine learning algorithms.This integrated approach enabled the identification of notable transformations in vegetation, water bodies, forest coverage, and barren land across the specified time intervals. The processed imagery was imported into a GIS framework, where geo-referencing and accuracy assessments were conducted. This integration allowed for the overlay of various datasets, execution of spatial queries, and performance of change detection analyses to quantify LULC transformations over time.The resultant LULC maps for 2010, 2017, and 2024 are vital for understanding trends related to urbanization, deforestation, and other ecological changes. By merging GIS capabilities with machine learning outputs, the methodology enhances the clarity of results, providing a visual representation of spatial distributions and evolving trends. This approach aligns with contemporary research practices in environmental monitoring and land use studies .

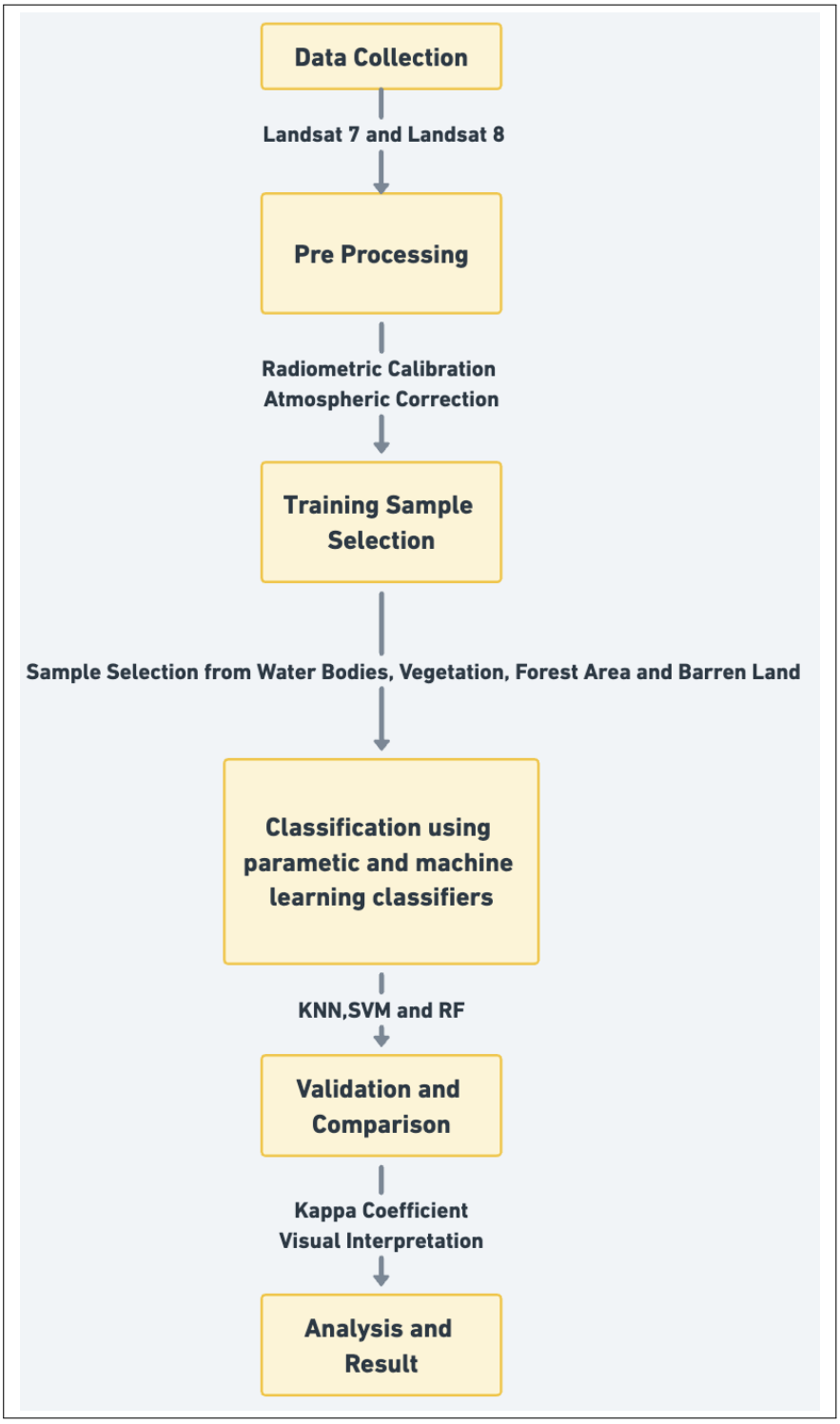

Fig. 2 The land use land cover (LULC) mapping process is depicted in a flowchart

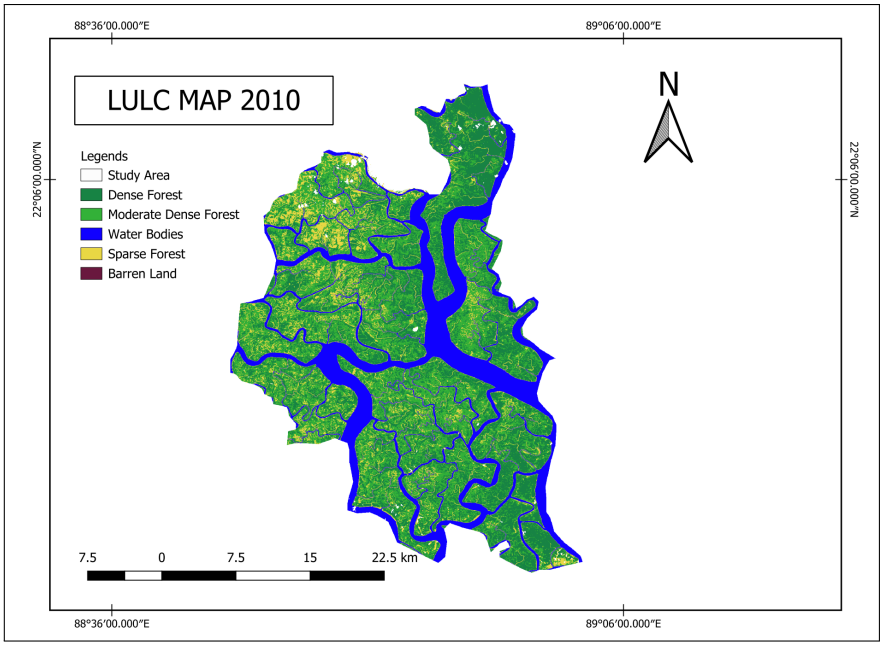

Fig. 3 The land use land cover (LULC) mapping of Eastern Region of Sunderbans in 2010

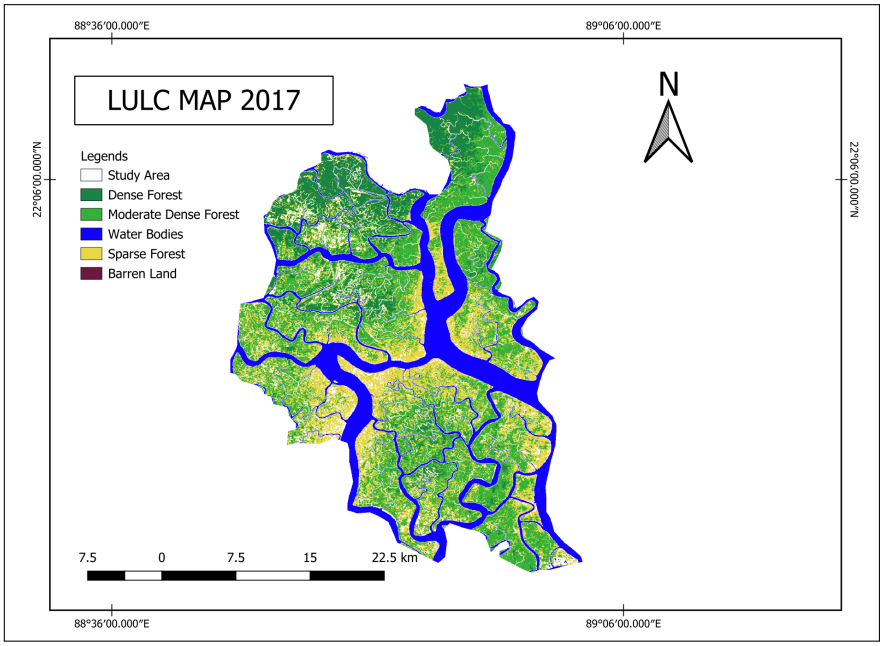

Fig. 4 The land use land cover (LULC) mapping of Eastern Region of Sunderbans in 2017

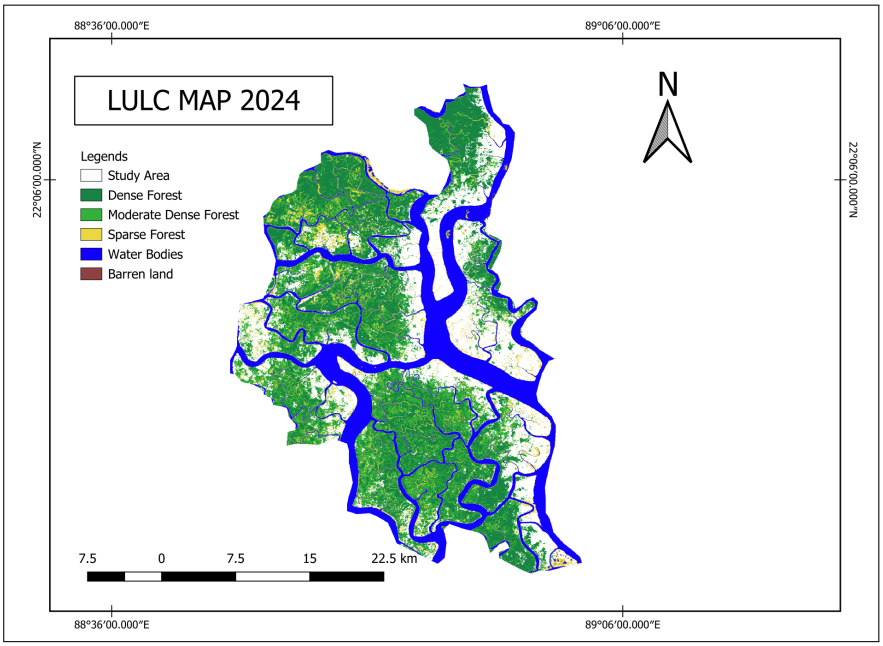

Fig. 5 The land use land cover (LULC) mapping of Eastern Region of Sunderbans in 2024

4.1 Machine Learning Algorithms

Machine learning (ML) methods are increasingly utilized for classifying land use and land cover (LULC), which is crucial for interpreting satellite imagery and various geospatial datasets to categorize different land types. This classification is vital for tracking environmental shifts, urban development, deforestation, and agricultural trends, yielding valuable insights into ecosystem dynamics and aiding informed decision-making in land management and conservation efforts.The advent of high-resolution satellite images from sources such as Landsat 7 and Landsat 8, alongside extensive geospatial datasets, has greatly improved the effectiveness of machine learning in this area. These data sources provide essential information about land surface features, enabling more precise and detailed evaluations of land use patterns over time. The following machine learning algorithms will be utilized in this research:

1. **K-Nearest Neighbors (KNN):** KNN [3],[17] utilised to classify land cover types in a remote sensing context and achieved promising results, demonstrating the algorithm's ability to distinguish between various land classes effectively.Incorporating [18]) features, such as texture and spectral indices, enhanced the KNN model's performance, enabling it to distinguish between urban and non-urban land covers more accurately. KNN's reliance on local information allows it to adapt well to the inherent variability found in remote sensing data.KNN [19] is particularly effective in classifying densely vegetated areas, which are typically challenging due to the similarities in spectral characteristics among different vegetation types.

   The KNN algorithm for Land Use/Land Cover detection can be expressed as:

$$\hat{y} = \text{argmax}_{c \in C} \sum_{i=1}^{k} \mathbb{I}(y_i = c) \cdot \exp\left(-\frac{D_{\text{weighted}}(\mathbf{x}, \mathbf{x}_i)}{2\sigma^2}\right) \cdot w_i$$

Where: -$\hat{y}$ is the predicted LULC class label, - C is the set of LULC classes, - k is the number of nearest neighbors, - y_i is the class label of the i-th neighbor, - $\mathbb{I}$ is an indicator function, -σ is a bandwidth parameter.

The weighted distance is computed as:

$$D_{\text{weighted}}(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^{m} w_j \cdot (f_j(\mathbf{x}) - f_j(\mathbf{x}_i))^2}$$

Where: - m is the number of features; - $f_j(\mathbf{x})$ and $f_j(\mathbf{x}_i)$ are the j-th features, - w_j is the weight of the j-th feature.

2. **Support Vector Machine:** Support Vector Machine (SVM) has become a popular choice for land use and land cover (LULC) classification in remote sensing due to its ability to handle high-dimensional data and effectively model complex decision boundaries. For instance, Rana et al. [20] demonstrated the efficacy of SVM in distinguishing various land cover types using satellite imagery, achieving superior classification accuracy compared to traditional methods. The algorithm's effectiveness lies in its kernel trick, which allows it to transform input data into higher-dimensional spaces, thereby facilitating the separation of non-linearly separable classes.Further studies, such as Prasad et al. [21] , highlighted SVM's robustness in urban land cover classification. By incorporating features such as spectral information and texture metrics, SVM was able to improve classification performance, particularly in complex urban environments where land cover types are closely related. The authors noted that SVM's reliance on a limited number of support vectors contributes to its ability to generalize well, even in the presence of noisy data.Additionally, Adam et al. [22]explored the application of SVM for classifying forest types, revealing that the algorithm effectively managed the spectral similarities often encountered in densely vegetated areas. Their findings suggested that parameter optimization, particularly the choice of kernel and regularization parameters, significantly influenced SVM's classification accuracy.

   The decision function for SVM can be represented as:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right)$$

   where: - $f(x)$ is the output class label for the input feature vector x, - α_i are the Lagrange multipliers, - y_i are the class labels, - $K(x_i, x)$ is the kernel function that measures the similarity between the support vectors x_i and the input vector x, - b is the bias term.

   This formula illustrates how SVM classifies land cover types by determining the optimal separating hyperplane in a transformed feature space.

3. **Random Forest:** Random Forest (RF) has emerged as a powerful method for land use and land cover (LULC) classification in remote sensing applications. For instance, Hayas et al. [23] employed RF to classify various land cover types using high-resolution satellite imagery, achieving high classification accuracy due to the algorithm's ability to manage the complexities of heterogeneous landscapes effectively. The ensemble nature of RF, which aggregates predictions from multiple decision trees, enhances its robustness against overfitting, particularly in cases with noisy data.Amini et al. [24],highlighted that incorporating various features, including texture and spectral attributes, significantly improved RF's performance in urban area classification. This ability to integrate diverse data types allows RF to differentiate between closely related land classes, such as built-up areas and vegetation, which often present classification challenges.Kavzoglu et al. [25] found RF particularly effective in classifying complex ecosystems like mangroves, where the algorithm excelled in distinguishing subtle variations in land cover. The prediction of a Random Forest can be expressed as:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} f_t(\mathbf{x})$$

where: - $\hat{y}$ is the predicted output, -T is the total number of trees in the forest, -$f_t(\mathbf{x})$ is the prediction from the t-th decision tree.

Each f_t can be represented by the recursive partitioning function:

$$f_t(\mathbf{x}) = \begin{cases} \text{class}_j & \text{if } \mathbf{x} \in R_j \\ \text{undefined} & \text{otherwise} \end{cases}$$

where: -R_j is the region corresponding to leaf node j in tree t.

LULC Change Detection Algorithm Using Machine Learning

This algorithm integrates three machine learning approaches—K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM)—to enhance Land Use and Land Cover (LULC) change detection over time. Initially, remote sensing data for two distinct time periods is collected and preprocessed to ensure consistency. Relevant features, including spectral indices and texture attributes, are extracted from the datasets. Each method employs its unique classification mechanism: KNN identifies classes by evaluating distances to nearest neighbors, Random Forest aggregates predictions from multiple decision trees trained on bootstrapped samples, and SVM optimizes a hyperplane to separate classes. The predicted labels for both time periods are compared to generate a change detection map, indicating areas of significant LULC change.

Algorithm 1 LULC Change Detection Using KNN, Random Forest, and SVM (Part 1)

Input:

$\mathcal{X}_{t1}$: Feature set for time 1 (e.g., spectral bands, textures)
$\mathcal{X}_{t2}$: Feature set for time 2
C: Set of LULC classes
k: Number of nearest neighbors

Output:

Change map indicating LULC changes between time 1 and time 2

Step 1: Data Preparation

- Collect remote sensing data for two different time periods (t_1 and t_2).
- Preprocess the data to remove noise and correct for atmospheric effects.
- Align and normalize the datasets for consistent analysis.

Step 2: Feature Extraction

- Extract relevant features from both datasets, including spectral indices (e.g., NDVI, NDBI) and texture features.
- Formulate the feature vectors: $\mathbf{x}_i^{t1}$ for time 1 and $\mathbf{x}_j^{t2}$ for time 2.

Step 3: KNN Classification for Time 1

for each pixel $\mathbf{x}_i^{t1}$ in $\mathcal{X}_{t1}$ **do**

Calculate distances: $d(\mathbf{x}_i^{t1}, \mathbf{x}_j^{t1})$ for $j = 1$ to N.
Select k nearest neighbors based on minimum distance.
Predict class label:

$$\hat{y}_i^{t1} = \text{argmax}_{c \in C} \sum_{m=1}^{k} \mathbb{I}(y_m^{t1} = c)$$

end for

Step 4: Random Forest Classification for Time 2

for each pixel $\mathbf{x}_j^{t2}$ in $\mathcal{X}_{t2}$ **do**

Predict LULC class using Random Forest:

$$\hat{y}_j^{t2} = H(\mathbf{x}_j^{t2})$$

where H is the ensemble of decision trees.

end for

Step 5: SVM Classification

for each pixel i in the dataset **do**

Predict LULC class using SVM:

$$\hat{y}_i = \text{SVM}(\mathbf{x}_i)$$

end for

Step 6: Change Detection Analysis

- Compare predicted class labels $\hat{y}_i^{t1}$, $\hat{y}_j^{t2}$, and $\hat{y}_i$ for each corresponding pixel.
- Generate a change map:

$$\text{ChangeMap}(i) = \begin{cases} 1 & \text{if } \hat{y}_i^{t1} \neq \hat{y}_j^{t2} \text{ or } \hat{y}_i \neq \hat{y}_j \\ 0 & \text{if } \hat{y}_i^{t1} = \hat{y}_j^{t2} \text{ and } \hat{y}_i = \hat{y}_j \end{cases}$$

Step 7: Output Results

- Visualize the change map using appropriate mapping techniques.
- Analyze the results to assess the nature and extent of LULC changes.

4.2 Model Evaluation Metrics

1. **Kappa Coefficient:** The Kappa Coefficient (Cohen's Kappa) is a statistic that measures the agreement between two raters (or classification models) while considering the possibility of agreement occurring by chance. In LULC detection, it helps evaluate how well the classifier's predictions match the ground truth, adjusted for random chance.

   The Kappa statistic for LULC detection can be expressed as:

$$\kappa = \frac{\left(\sum_{i=1}^{C} TP_i\right) - \left(\sum_{j=1}^{C} \frac{(TP_j+FP_j)(TP_j+FN_j)}{N}\right)}{1 - \left(\sum_{j=1}^{C} \frac{(TP_j+FP_j)(TP_j+FN_j)}{N}\right)}$$

   Where: TP_i is the True Positives for class i, FP_j is the False Positives for class j, FN_j is the False Negatives for class j, C is the Total number of classes, N is the Total number of samples.

2. **Accuracy:** Accuracy is the proportion of correctly classified instances (both true positives and true negatives) out of the total number of instances. It's a general measure of how often the classifier is correct.

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} w_i \cdot TP_i}{\sum_{i=1}^{N} (TP_i + FP_i + FN_i)}$$

   Where: N is the number of classes, TP_i is the true positives for class i, FP_i is the false positives for class i, FN_i is the false negatives for class i, w_i is the weight for class i, representing its importance.

3. **Recall (Sensitivity):** Recall measures the classifier's ability to correctly identify positive instances. It's particularly useful when missing positive cases is more critical.

$$\text{Recall} = \frac{\sum_{i=1}^{N} w_i \cdot \frac{TP_i}{TP_i+FN_i}}{\sum_{i=1}^{N} w_i}$$

4. **Precision:** Precision is the proportion of correctly predicted positive instances out of all predicted positives. It reflects how many of the predicted positives are actually relevant, helping in scenarios where minimizing false positives is important.

$$\text{Precision} = \frac{\sum_{i=1}^{N} w_i \cdot \frac{TP_i}{TP_i + FP_i}}{\sum_{i=1}^{N} w_i}$$

Where: N is the number of classes, TP_i is the true positives for class i, FP_i is the false positives for class i, w_i is the weight for class i, representing its importance.

5 Result and Analysis

Tables and figures illustrating the findings of the spatiotemporal analysis of land use and land cover dynamics are presented below. Various satellite datasets and parameters have been utilized for evaluation. The assessment criteria are based on classification accuracy and model performance.

Table 4 Performance Metrics for KNN

KNN	**Precision**	**Recall**
Dense Forest	91.96	96
Moderate Dense Forest	92.38	89.3
Water Bodies	86	85.3
Sparse Forest	83.88	87.8
Barren Land	85.76	90.68
	OA= 89.62	Kappa= 0.856

Table 5 Performance Metrics for SVM

SVM	**Precision**	**Recall**
Dense Forest	92.01	96.65
Moderate Dense Forest	93.33	90.5
Water Bodies	87.43	89
Sparse Forest	86.86	91.7
Barren Land	87.57	91.23
	OA=91.82	Kappa=0.879

Table 6 Performance Metrics for RF

RF	**Precision**	**Recall**
Dense Forest	92.57	96
Moderate Dense Forest	96.06	93.4
Water Bodies	87.44	87
Sparse Forest	88.3	92.42
Barren Land	90.39	93.67
	OA= 92.50	Kappa= 0.899

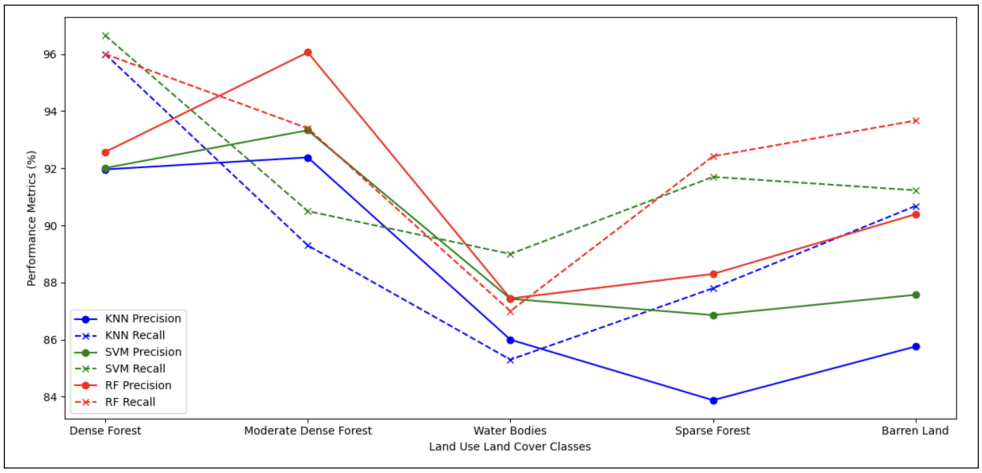

Fig. 6 Visualisation of Performance Metrics

The area of LULC classes detected in the Eastern Region of the Sunderban mangrove forest for the years 2010, 2017, and 2024.

KNN Classification For the KNN classification, the area of Dense Forest declined from 2450 km^2 in 2010 to 2300 km^2 in 2017, then increased to 2500 km^2 in 2024. Moderate Dense Forest experienced a decrease from 780 km^2 in 2010 to 700 km^2 in 2017, followed by an increase to 740 km^2 in 2024. The area of Sparse Forest rose from 620 km^2 in 2010 to 670 km^2 in 2017, and remained stable at 620 km^2 in 2024. Water Bodies dropped from 390 km^2 in 2010 to 340 km^2 in 2017, then increased to 360 km^2 in 2024. Barren Land increased from 660 km^2 in 2010 to 780 km^2 in 2017, and then slightly decreased to 680 km^2 in 2024. As with the other classifications, the total area consistently remained at 5000 km^2 throughout these years.

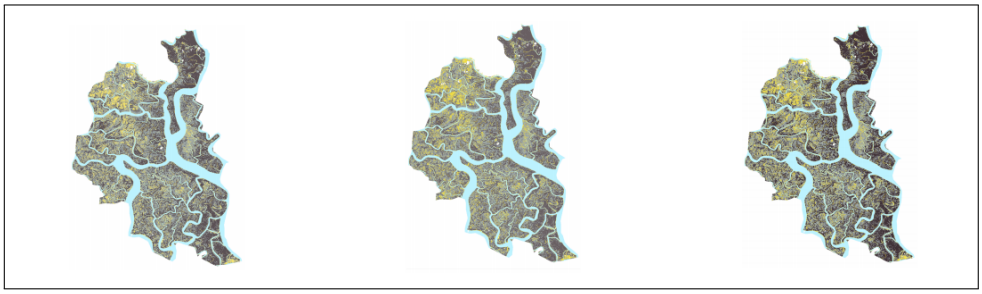

Fig. 7 LULC Mapping through KNN in respective years

SVM Classification In the SVM classification, the area of Dense Forest decreased from 2500 km^2 in 2010 to 2400 km^2 in 2017, before increasing again to 2550 km^2 in 2024. The area of Moderate Dense Forest declined from 800 km^2 in 2010 to 700 km^2 in 2017, with a slight recovery to 750 km^2 by 2024. Sparse Forest area increased from 600 km^2 in 2010 to 650 km^2 in 2017, but dropped to 550 km^2 in 2024. Water Bodies

also saw a decline from 400 km^2 in 2010 to 350 km^2 in 2017, before recovering to 370 km^2 in 2024. Barren Land increased from 600 km^2 in 2010 to 700 km^2 in 2017, and slightly decreased to 680 km^2 in 2024. The total area for all classes remained consistent at 5000 km^2 throughout the years.

Fig. 8 LULC Mapping through SVM in respective years

Random Forest (RF) Classification In the Random Forest classification, the area of Dense Forest showed a slight decrease from 2550 km^2 in 2010 to 2450 km^2 in 2017, followed by an increase to 2600 km^2 in 2024. The area of Moderate Dense Forest decreased from 750 km^2 in 2010 to 720 km^2 in 2017, and then rose to 800 km^2 in 2024. The Sparse Forest area slightly decreased from 580 km^2 in 2010 to 640 km^2 in 2017, before falling back to 600 km^2 in 2024. Water Bodies decreased from 410 km^2 in 2010 to 360 km^2 in 2017, and increased to 380 km^2 in 2024. Barren Land rose from 610 km^2 in 2010 to 730 km^2 in 2017, then decreased to 640 km^2 in 2024. The total area remained stable at 5000 km^2 across all years.

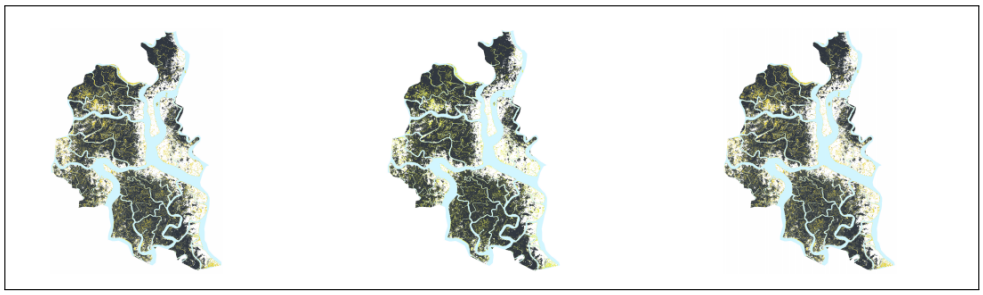

Fig. 9 LULC Mapping through RF in respective years

6 Discussion

In our study, Random Forest (RF) emerged as the best-performing model in comparison to K-Nearest Neighbor (KNN) and Support Vector Machine (SVM), demonstrating superior precision, recall, overall accuracy (OA), and Kappa values. These findings align with extensive literature that underscores RF's capacity for handling complex and high-dimensional datasets due to its ensemble learning approach and random subset selection (Belgiu and Drăgut, [26]; Gislason et al., [27]). RF is particularly effective in reducing overfitting and accommodating noisy data, which is often encountered in land use and land cover (LULC) mapping applications (Rodriguez-Galiano et al.,[28]). Its ability to generate diverse decision trees for classification allows RF to handle the heterogeneity of land cover classes with greater accuracy than single decision tree-based models (Mellor et al., [29]; Pelletier et al., [30]). In addition, RF's ensemble approach leverages multiple weak learners to strengthen model prediction, providing it with an edge in accuracy, particularly in environmental data analysis where errors and misclassifications can arise from the noisy or incomplete data common in freely available satellite imagery like Landsat (Friedl and Sulla-Menashe [31]).Contrary to RF's consistent performance, KNN tends to struggle in high-dimensional settings, often underperforming in datasets that involve complex class boundaries. This limitation is well documented in other studies, where KNN's accuracy tends to decline with increasing dimensionality, making it less suitable for LULC classifications that require detailed, pixel-based precision (Heydari and Mountrakis [32]). Moreover, KNN's sensitivity to the choice of the k-value further complicates its application, as improper selection can result in suboptimal classifications, especially in heterogeneous environments (Pouteaua et al.[33]). Despite these challenges, KNN has demonstrated better performance in less complex classifications, such as in agricultural or homogenous landscapes, which aligns with findings from other geographic contexts (Gong et al. [34]).SVM, while known for its robustness and ability to handle nonlinear data using kernel functions, also showed lower performance compared to RF in our study, largely due to the potential for overfitting in high-dimensional spaces and challenges in selecting optimal kernel parameters (Pelletier et al. [35]. Despite this, SVM remains a frequently used tool for LULC classification, primarily in applications where boundary precision between classes is crucial, though it lacks the computational efficiency of RF in large-scale data environments (Khatami and Mountrakis, [36]). This observation supports the findings of other researchers who have noted that SVM may not scale as well with large datasets or work as effectively in regions where training data is sparse or noisy.

The RF model's adaptability and computational efficiency, especially in dealing with fragmented or missing data, make it particularly useful in tropical and subtropical regions. Such environments are prone to incomplete or low-resolution datasets, often obtained from freely available satellite imagery sources like Landsat (Gislason et al. [27]; Belgiu and Drăgut, [26]). The high dimensionality and varying spectral properties of these datasets make them challenging for simpler models like KNN or SVM to manage effectively (Friedl and Sulla-Menashe [31]). Our results support the assertion that RF is the most consistent and reliable machine learning algorithm for such geospatial applications, especially in mapping LULC, which is corroborated by

studies focusing on global forest and crop cover (Hansen et al. [37]) Potapov et al. [38]).It's noteworthy that while RF achieved the best performance metrics in this study, some literature suggests that KNN can outperform RF under certain conditions. For instance, in homogenous landscapes or specific LULC classifications, KNN has been shown to yield better results than RF. Such discrepancies underline the importance of context in model selection and suggest that while RF is generally superior, there may be scenarios where simpler models like KNN or more specialized models like SVM could be more suitable, depending on the dataset characteristics and classification task. Although this validation method provides a robust comparison, it's important to acknowledge the limitations inherent in using global-scale products for local-level LULC assessments. Misclassification of certain land cover features has been reported in these products, particularly in heterogeneous regions like tropical forests [39] , which could lead to slight deviations in accuracy . An alternative approach would be to validate with ground-truthed data, though such datasets were not available for this study area. However, despite these constraints, the RF model's ability to handle such validation protocols, coupled with its robust performance, affirms its utility in LULC mapping across diverse geographic contexts. Overall, our findings provide further evidence of RF's efficacy in reducing classification uncertainties in complex LULC tasks, particularly in underrepresented tropical regions like equatorial Africa [40], which are often characterized by fragmented and incomplete datasets. These results could have important implications for future research, as they demonstrate the potential for RF to enhance the accuracy of land cover classifications, informing policy development and sustainable land use planning initiatives.

7 Conclusion

This research provides a comprehensive spatiotemporal analysis of land use and land cover (LULC) changes in the Eastern Sunderban Mangrove Forest between 2010 and 2024. Significant changes in land cover, especially a decline in mangrove forests and an expansion of barren land and urbanized regions, were observed, highlighting the growing vulnerability of this ecosystem to both human activities and environmental stressors. Integrating GIS with machine learning models—namely K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF)—proved effective in monitoring these changes. Among the models, Random Forest displayed the highest accuracy at 92.50%, followed by SVM at 91.82%, and KNN at 89.62%. This study reinforces the potential of using GIS-enhanced machine learning techniques to efficiently track LULC dynamics.The study lays the groundwork for further investigation, particularly by introducing more advanced machine learning methods, such as deep learning models like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which could enhance the accuracy of LULC classification. Increasing the temporal resolution of the satellite data would allow for real-time analysis and monitoring of ongoing changes in sensitive areas like the Sunderbans.Further, investigating the socio-economic drivers behind these LULC changes can provide insights into human influences on land cover transformation, such as urban expansion, agriculture, and

population growth. This would offer actionable insights for policy development, allowing for more precise intervention strategies to conserve and restore the ecosystem. Going forward, efforts should focus on sustainable land management and conservation strategies, including reforestation and mangrove restoration. International collaborations, supported by frameworks like the United Nations Sustainable Development Goals (SDGs) and the Ramsar Convention on Wetlands, are essential to protecting the Sunderban Mangroves. Additionally, AI-based predictive models can offer valuable insights into the future impact of climate change on this region, aiding in adaptive management strategies.Ultimately, this research highlights the critical need for more sustainable approaches to land management in the Sunderban Mangrove region. Advanced methodologies and stronger conservation efforts will be pivotal in safeguarding this unique ecosystem against further degradation and ensuring its resilience to future environmental and human-induced challenges.

References

[1] Datta D, Deb S. Analysis of coastal land use/land cover changes in the Indian Sunderbans using remotely sensed data. Geo-spatial Information Science. 2012;15(4):241–250.

[2] Roy SS, Ghosh T, Hewavithana DK. Transformation of coastal wetlands in the Sundarban Delta (1999–2020). Environmental monitoring and assessment. 2024;196(8):758.

[3] Talukdar S, Singha P, Mahato S, Pal S, Liou YA, Rahman A. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. Remote sensing. 2020;12(7):1135.

[4] Wang J, Bretz M, Dewan MAA, Delavar MA. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. Science of The Total Environment. 2022;822:153559.

[5] Tassi A, Vizzari M. Object-oriented lulc classification in google earth engine combining snic, glcm, and machine learning algorithms. Remote Sensing. 2020;12(22):3776.

[6] Aryal J, Sitaula C, Frery AC. Land use and land cover (LULC) performance modeling using machine learning algorithms: a case study of the city of Melbourne, Australia. Scientific Reports. 2023;13(1):13510.

[7] Krivoguz D, Chernyi SG, Zinchenko E, Silkin A, Zinchenko A. Using Landsat-5 for accurate historical LULC classification: A comparison of machine learning models. Data. 2023;8(9):138.

[8] Chachondhia P, Shakya A, Kumar G. Performance evaluation of machine learning algorithms using optical and microwave data for LULC classification. Remote Sensing Applications: Society and Environment. 2021;23:100599.

[9] Ouma Y, Nkwae B, Moalafhi D, Odirile P, Parida B, Anderson G, et al. Comparison of machine learning classifiers for multitemporal and multisensor mapping of urban LULC features. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2022;43:681–689.

[10] Srivastava A, Bharadwaj S, Dubey R, Sharma V, Biswas S. Mapping vegetation and measuring the performance of machine learning algorithm in lulc classification in the large area using Sentinel-2 and Landsat-8 datasets of dehradun as a test case. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2022;43:529–535.

[11] Mandal S, Bandyopadhyay A, Bhadra A. Dynamics and future prediction of LULC on Pare River basin of Arunachal Pradesh using machine learning techniques. Environmental Monitoring and Assessment. 2023;195(6):709.

[12] Chatziantoniou A, Petropoulos GP, Psomiadis E. Co-Orbital Sentinel 1 and 2 for LULC mapping with emphasis on wetlands in a mediterranean setting based on machine learning. Remote Sensing. 2017;9(12):1259.

[13] Singh RK, Singh P, Drews M, Kumar P, Singh H, Gupta AK, et al. A machine learning-based classification of LANDSAT images to map land use and land cover of India. Remote Sensing Applications: Society and Environment. 2021;24:100624.

[14] Sankarrao L, Ghose DK, Rathinsamy M. Predicting land-use change: Intercomparison of different hybrid machine learning models. Environmental Modelling & Software. 2021;145:105207.

[15] Kafy AA, Shuvo RM, Naim MNH, Sikdar MS, Chowdhury RR, Islam MA, et al. Remote sensing approach to simulate the land use/land cover and seasonal land surface temperature change using machine learning algorithms in a fastest-growing megacity of Bangladesh. Remote Sensing Applications: Society and Environment. 2021;21:100463.

[16] Pokhariya HS, Singh D, Prakash R. Evaluation of different machine learning algorithms for LULC classification in heterogeneous landscape by using remote sensing and GIS techniques. Engineering Research Express. 2023;5(4):045052.

[17] Kalpana YB, Nandhagopal S. LULC image classifications using K-means clustering and KNN algorithm. Dynamic Systems and Applications. 2021;30(10):1640–1652.

[18] Zhang S, Li X, Zong M, Zhu X, Wang R. Efficient kNN Classification With Different Numbers of Nearest Neighbors. IEEE Transactions on Neural Networks and Learning Systems. 2018;29(5):1774–1785. https://doi.org/10.1109/TNNLS.2017.2673241.

[19] Neyns R, Canters F. Mapping of urban vegetation with high-resolution remote sensing: A review. Remote sensing. 2022;14(4):1031.

[20] Rana VK, Suryanarayana TMV. Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands. Remote Sensing Applications: Society and Environment. 2020;19:100351.

[21] Prasad S, Savithri TS, Krishna IVM. Comparison of accuracy measures for RS image classification using SVM and ANN classifiers. International Journal of Electrical and Computer Engineering. 2017;7(3):1180.

[22] Adam E, Mutanga O, Odindi J, Abdel-Rahman EM. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. International Journal of Remote Sensing. 2014;35(10):3440–3458.

[23] Hayes MM, Miller SN, Murphy MA. High-resolution landcover classification using Random Forest. Remote sensing letters. 2014;5(2):112–121.

[24] Amini S, Saber M, Rabiei-Dastjerdi H, Homayouni S. Urban land use and land cover change analysis using random forest classification of landsat time series. Remote Sensing. 2022;14(11):2654.

[25] Kavzoglu T, Bilucan F. Effects of auxiliary and ancillary data on LULC classification in a heterogeneous environment using optimized random forest algorithm. Earth Science Informatics. 2023;16(1):415–435.

[26] Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing. 2016;114:24–31.

[27] Gislason PO, Benediktsson JA, Sveinsson JR. Random forests for land cover classification. Pattern recognition letters. 2006;27(4):294–300.

[28] Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS journal of photogrammetry and remote sensing. 2012;67:93–104.

[29] Mellor A, Haywood A, Stone C, Jones S. The performance of random forests in an operational setting for large area sclerophyll forest classification. Remote Sensing. 2013;5(6):2838–2856.

[30] Tariq A, Yan J, Gagnon AS, Riaz Khan M, Mumtaz F. Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest. Geo-Spatial Information Science.

2023;26(3):302–320.

[31] Sulla-Menashe D, Gray JM, Abercrombie SP, Friedl MA. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. Remote sensing of environment. 2019;222:183–194.

[32] Heydari SS, Mountrakis G. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. Remote Sensing of Environment. 2018;204:648–658.

[33] Pouteau R, Meyer J, Fourdrigniez M, Taputuarai R, Larrue S. Novel ecosystems in the Pacific Islands: Assessing loss, fragmentation and alteration of native forests by invasive alien plants on the island of Moorea (French Polynesia). Biodiversity and Societies in the Pacific Islands. 2013;p. 19–33.

[34] Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, et al. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. International journal of remote sensing. 2013;34(7):2607–2654.

[35] Pelletier C, Valero S, Inglada J, Champion N, Dedieu G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. Remote Sensing of Environment. 2016;187:156–168.

[36] Khatami R, Mountrakis G, Stehman SV. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. Remote sensing of environment. 2016;177:89–100.

[37] Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, et al. High-resolution global maps of 21st-century forest cover change. science. 2013;342(6160):850–853.

[38] Potapov P, Hansen MC, Kommareddy I, Kommareddy A, Turubanova S, Pickens A, et al. Landsat analysis ready data for global land cover and land cover change mapping. Remote Sensing. 2020;12(3):426.

[39] Potapov P, Li X, Hernandez-Serna A, Tyukavina A, Hansen MC, Kommareddy A, et al. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment. 2021;253:112165.

[40] Potapov AM, Dupérré N, Jochum M, Dreczko K, Klarner B, Barnes AD, et al. Functional losses in ground spider communities due to habitat structure degradation under tropical land-use change. Ecology. 2020;101(3):e02957.