Journal of Autonomous Intelligence (2024) Volume 7 Issue 3
doi: 10.32629/jai.v7i3.1333

ORIGINAL RESEARCH ARTICLE

Identifying land use land cover dynamics using machine learning method and GIS approach in Karaivetti, Tamil Nadu

Thylashri Sivasubramaniyan*, Rajalakshmi Nagarnaidu RajaPerumal

Department of Computer Science and Engineering, Veltech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Avadi, Chennai 600062, TamilNadu, India

*** Corresponding author:** Thylashri Sivasubramaniyan, thylashri93@gmail.com

ABSTRACT

An important analytical tool for tracking, mapping, and quantifying changes in land use and land cover (LULC) across time serves as the use of machine learning techniques. The environment and human activities both have the potential to change how land is used and covered. Classifying LULC types at different spatial scales has been effectively achieved by models like classification and regression trees (CART), support vector machines (SVM), extreme gradient boosting (XGBoost), and random forests (RF). To prepare images from Landsat before sending and analysis for an aspect of our research, we employed the Google Earth Engine. High-resolution imagery from Google Earth images were used to assess each kind of method and field data collection. Utilizing Geographic Information System (GIS) techniques, LULC fluctuations between 2015 and 2020 were assessed. According to our results, XGBoost, SVM, and CART models proved superior by the RF model regarding categorization precision. Considering the data, we collected between 2015 and 2020, from 11.57 hectares (1.74%) in 2015 to 184.19 hectares (27.65%) in 2020, the barren land experienced the greatest variation, that made an immense effect. Utilizing the support of satellite imagery from the Karaivetti Wetland, our work combines novel GIS techniques and machine learning strategies to LULC monitoring. The created land cover maps provide a vital benchmark that will be useful to authorities in formulating policies, managing for sustainability, and keeping track of degradation.

Keywords: Geographic Information System (GIS); change detection; land use and land cover change (LULC); machine learning; Karaivetti

ARTICLE INFO

Received: 10 October 2023
Accepted: 15 November 2023
Available online: 10 January 2024

1. Introduction

The field of predicting variations in vegetation and land use through the integration of machine learning and GIS has seen a lot of effort recently. Understanding and projecting the dynamics of land use and land cover (LULC) are now essential due to the continuous impact of human activity on the environment. These specific types of data are essential for efficient land management, environmental conservation, and conservation efforts. Machine learning algorithms, combined with the spatial analysis capabilities of GIS, provide a powerful framework for analyzing and modelling these changes.

"Land use" refers to the human activities and practices conducted on a piece of land, such as agriculture, urban development, forestry, or transportation, while the term "Land cover" describes the natural and vegetative covering of the surface of the Earth, including any vegetation, water bodies, bare soil, or built-up regions. Variations in the way land is utilized and covered can have a significant impact on the environment's ecosystems, biodiversity, climate patterns, and socioeconomic systems.

Traditional methods for modelling land use and land cover changes relied heavily on expert knowledge, numerical analyses, then manual interpretation of aerial photographs or satellite imagery. However, these methods were often time consuming, subjective, and limited in their ability to handle complex spatial relationships and large datasets[1]. Machine learning, a subfield of artificial intelligence, has revolutionized the field by automating the process of pattern recognition, classification, and prediction.

By harnessing the power of machine learning algorithms, such as decision trees, random forests, support vector machines, or neural networks, researchers can analyze vast amounts of geospatial data and derive meaningful patterns and relationships[2]. GIS software complements machine learning by providing spatial analysis tools and a platform for integrating and visualizing various geospatial datasets[3].

The modelling process involves several stages, including data collection, preprocessing, feature extraction, training data preparation, algorithm selection, model training, validation, prediction, uncertainty analysis, and decision support. Through this iterative process, the models learn from historical land use and land cover patterns and their associated drivers, allowing them to make predictions and projections for the future.

We utilize the Google Earth Engine (GEE) cloud-based platform with a large data archive that allows for the analysis of environmental data at the planetary scale in ArcGIS to compare classification techniques to fill this research requirements. Our main contribution is to analyze LULC trends from 2015 to 2020 using the best accurate classifier[4]. The goals of this research are to evaluate how LULC has changed over the past half-decade and to use machine learning to improve the LULC categorization approach.

The increased levels of deforestation, agricultural growth, and environmental degradation associated with urbanization highlight the need for precise and current LULC maps. The problem is made more difficult by the environment, which modifies conditions in ways that require close scrutiny to identify possible harmful changes. This research is notable as it represents a novel attempt to evaluate the effectiveness of machine learning methods for LULC classification on a broader scale, covering vast regions.

We provide an in-depth analysis of the influence of machine learning on LULC assessments in the following sections, which also explore the study area details, classification techniques, and outcomes.

2. Literature review

By utilizing ANN, the research uncovers hidden relationships and provides a deeper understanding of how thermal attributes correlate with LULC modifications. The findings shed light on the thermal dynamics of different land cover types, offering valuable insights into the environmental impact of changing land use patterns[5]. An in-depth analysis of the changes in land cover and use between 2000 and 2018, using machine learning algorithms to predict future changes through 2050. This research offers predictions about the dynamics of land transitions.

The accuracy of classification approaches can be impacted by a variety of factors, with the choice of preparatory trials, the area under study diversity, technologies used and the number of characterizing classes[6]. Classifiers can be categorized into a variety of groups depending on the approach and technique employed[7]. With conventional visual interpretation and statistical techniques, LULC categorization is no longer accurate.

Sharifi[8] focuses on the significance of remote sensing is for tracking environmental changes and offers fascinating data about how natural occurrences affect the land. There have been numerous studies comparing classifiers in the last ten years, such as the comparison of SVM with RF for classifying land cover[9]. A great tool for editing huge remote sensing images to create land cover maps across vast areas is GEE. With this platform, users can use a web-based code editor known as an integrated development environment (IDE) to analyze all remotely sensed pictures without a local download. Our geospatial analysis is made more scalable and effective with the help of Tamiminia et al.[10] research on the effective handling of large-scale data using Google Earth Engine, providing a more thorough understanding of the region's changing land cover patterns.Users can thus quickly search, pick, and evaluate a big body of knowledge for a broad study field[11]. JavaScript is used to create client libraries, and Python is used to handle code modification.

Additionally, GEE is exceptional and in-demand due to its quick processing and simple utilization of Legion procedures, which open up RS tools to users of any ability, regardless of their experience. GEE has appeared in numerous academic publications in recent years. The MapReduce architecture is used by Google Earth Engine to divide large datasets into manageable chunks and distribute, process, and aggregate the data concurrently across several tools. Large amounts of data are efficiently managed with this technique[12]. Output datasets were assembled following the processing of the data as separate components. Landsat 8, MODIS, Sentinel 2 and many additional satellites are among the photos contained in GEE, particularly the nearly 40-years-old Landsat time series.

LULC change[13] in flood monitoring and various other applications, such as evidenced by the observation of coastline modifications and decreasing vegetation. The Google Earth Engine framework has not yet been used in any research to compare some of the models used for assessing differences in LULC. It also had various rates of urban sprawl modifications before and after 2016, which resulted in adjustments to the encroachments on agricultural fields.

A durable tool for managing Gerbig data is Google Earth Engine[14]. The authors conducted a full evaluation and statistical analysis, demonstrating its potential in various applications, including LULC detection. GEE cloud-based platform offers efficient data processing ability to monitor the environment on an enormous level.

Park et al.[13] discuss the significance of sensor technology and data collection methods for environmental studies, which can be applied to monitoring changes in land cover near water bodies. Mugo et al.[14] employ satellite remote sensing to quantify the Lake Victoria Basin's shifting use of land and vegetation cover insights into the trends and drivers of land cover changes trendy a critical region, emphasizing the value of remote sensing for long-term environmental studies.

Guerrero et al.[15] highlights the impacts of infrastructure development on land cover, demonstrating the relevance of remote sensing in monitoring environmental changes associated with human activities. Chen et al.[16] in Jiangle, China on the monitoring of land use changes through the use of remote sensing and GIS techniques provides a useful model for applying similar methodologies to analyze and predict land cover dynamics. Wahla et al.[17] evaluation of climatic variability through the use of spatiotemporal mapping techniques and machine learning models provides a possible way to evaluate the influence of changing climatic conditions on environment.

Nourani et al.[18] study of climate, land cover, and lake level changes using remotely sensed data and wavelet analysis provides a helpful framework for understand the complex interactions within elements of the environment. The land use change study by Temgoua et al.[19] in Cameroon provides a methodological model for evaluating variations in the ecosystems throughout the years. The mapping of mangroves in Thailand by Pimple et al.[20] using Landsat images and Google Earth Engine provides a precedent-based analytical framework for assessing sustainability transitions.

3. Materials and methods

3.1. Study area

Karaivetti was chosen as the study area due to its diverse ecological landscape, including wetlands, its status as a critical bird sanctuary, relevance to regional land use dynamics in South India, and the availability of historical geospatial data. Established on 5 April 1999, it ranks among the largest freshwater lakes in southern Tamil Nadu. This internationally recognized sanctuary, known as Karaivetti Wildlife Sanctuary, boasts a diverse ecosystem, providing a habitat for at least 188 recorded bird species. This comprehensive setting offers an opportunity to explore the applications of machine learning and GIS in monitoring land use and land cover dynamics, contributing to conservation efforts.

The Karaivetti Bird Sanctuary is a designated area known for its rich avian biodiversity and serves as a vital habitat for various bird species located in Ariyalur District, Tamil Nadu, South India the sanctuary spans across "454 ha." Geographically, the sanctuary is situated at approximately 10°58′13″ N latitude and 79°02′29″ E longitude. The coordinates are referenced to the WGS 1984 datum.

The study area of the Karaivetti Bird Sanctuary (**Figure 1**) is notable for its diverse landscape, which includes wetlands, marshes, lakes, and surrounding vegetation. It serves as a vital ecological haven, contributing significantly to avian biodiversity in the region and is characterized by its unique ecological features and natural resources.

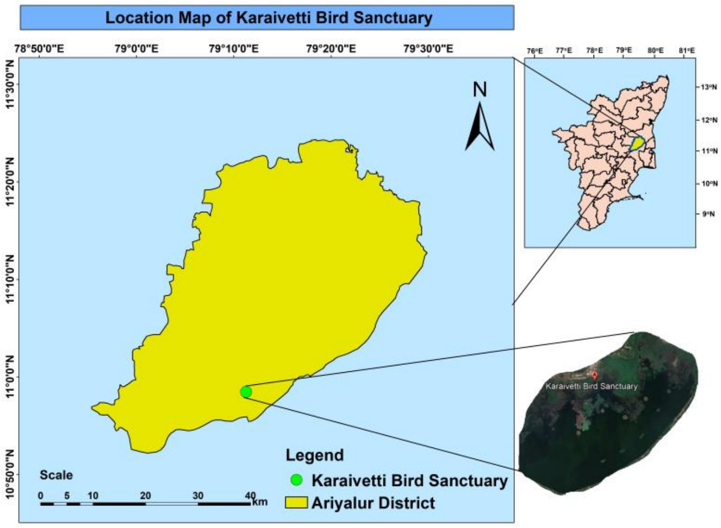

Figure 1. Location of Karaivetti Bird Sanctuary.

3.2. Datasets

3.2.1. Acquisition of land use and land cover (LULC) datasets and satellite imagery

For this research, we collected datasets of satellite imagery that covered the Karaivetti study area from 2015 to 2020. The Landsat 8 satellite imagery from the Karaivetti area was captured on 22 January 2015, and 2 October 2020, which correspond to Row 052 and Path 143. The images have a 30 m spatial resolution. Within the study area, these datasets provide comprehensive information on a range of land cover categories, such as forests, water bodies, greenery, barren land, and agricultural land.

3.2.2. Ground truth data collection

Ground truth data were gathered to supplement and validate the information obtained from satellites. The various land cover categories and their precise locations within the Karaivetti study area are described in detail by this data.

3.2.3. Calculation of remote sensing indicators

Using the collected satellite images, we computed key remote sensing indicators, such as the normalized difference vegetation index (NDVI). Within the research area, the NDVI values were utilized to gather significant data about the condition of the vegetation and other aspects of the land cover.

3.2.4. Use of GIS and machine learning techniques

We used advanced GIS modelling and machine learning techniques to analyze changes in land cover and use in Karaivetti. This required the use of data fusion, change detection techniques, spatial analysis, and classification algorithms. These methods were applied to comprehend the types and causes of changes in land use in the research area.

3.3. Proposed approach

The proposed approach is shown in **Figure 2** consists of the following four basic actions to achieve the stated goals. These tasks include classifying data accurately using multiple classification algorithms, assessing the accuracy of these algorithms to identify the most effective one for the task at hand, creating LULC maps, and monitoring variations in the spatial and temporal relationships within the dataset. By employing a range of algorithms and conducting rigorous accuracy assessments, the research aims to generate reliable LULC maps and effectively detect changes over time, providing revealing details about the changing conditions of the research area.

Creating training datasets, classifying the images, and evaluating the accuracy are the three stages were employed. Greeneries, barren land and water bodies are three of the LULC classes we first defined in our study area before creating training datasets. To adapt to the different forms of landscape cover for the years 2015 and 2020, we found to choose these three LULC classes.

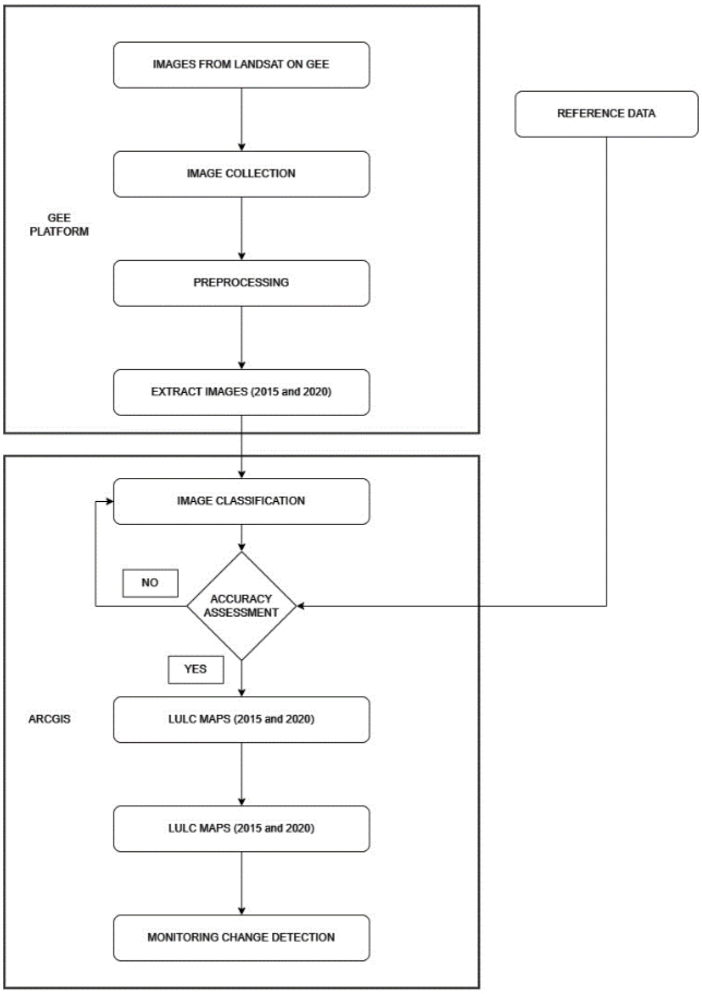

Figure 2. The proposed approach of theoretical framework.

3.4. Image acquisition and preprocessing

In the pursuit of identifying land use and land cover dynamics using machine learning and GIS techniques, the process begins with the acquisition of high-resolution satellite imagery, meticulously selected to cover the study area over different time intervals. These images undergo a series of critical preprocessing steps to ensure data accuracy and quality. Pixel values are normalized using radiometric calibration, and then any atmospheric effects that could skew surface reflectance values are carefully removed using advanced atmospheric correction methods. Pixels are then carefully aligned for accurate comparative analysis by carefully registering the images to a shared spatial reference system. By excluding undesired artefacts like clouds and shadows, masks help to maximize focus on areas of interest while reducing the possibility of mistakes. Following this, image enhancement techniques may be applied to improve visual clarity, and change detection is performed to identify areas of land cover transformation. Extracted features are then used to train machine learning models, and the results are integrated into a GIS framework for spatial and temporal analysis. This rigorous image acquisition and preprocessing phase lays the foundation for precise land use and land cover dynamics assessment and informed decision-making in land management and conservation.

3.5. Image classification

The random forest algorithm is a prevailing machine learning technique commonly used for image classification tasks. It combines multiple decision trees to form an ensemble model that can accurately classify images. For optimal variation and prevent over fitting, the method operates through choosing an arbitrary amount of the training data as well as features for every decision tree. During prediction, each decision tree independently assigns class labels to input images, and the final prediction is determined through a voting mechanism. The random forest algorithm is renowned for its reliability, capacity for managing large-scale data, and durability against excessive fit. It is widely used in image classification due to its accuracy and capability to handle complex relationships within image data.

One of the supervised learning techniques used to address a variety of regression and classification problems is SVM. In spectral-radiometric characteristics that are retrieved using trained samples and decision limits to categorize various groups. SVM uses support vectors as its training samples since they specify the hyper plane's margin. Each training sample's kernel weight is generated by SVM using kernel functions, and the kernel size has an impact on functional similarity. Two hyper parameters are required to build an SVM model: C and gamma. A kernel's radius is controlled by gamma, while the degree of model fitting is controlled by C. Although ArcGIS Pro has an SVM classifier, the SVM algorithm's C and gamma options are not available.

A decision tree-based method is the classification and regression trees (CART) algorithm. used for image classification. It recursively partitions the feature space based on selected features and threshold values, aiming to minimize impurity and increase class label homogeneity within subsets. The algorithm constructs a tree-like model where each leaf node represents a class label assignment. CART is known for its simplicity, interpretability, and ability to handle categorical and continuous features. In image classification, it can be applied to pixels or segments, using their features to predict class labels for new data points. However, caution should be exercised to prevent overfitting and ensure robust model generalization.

XGBoost, primarily designed for structured data, can be adapted for machine learning-based image classification in LULC analysis. In this approach, the key lies in feature engineering, where relevant features are extracted from images, such as texture, color histograms, or spectral indices. These engineered features are then used to represent each image as a feature vector. XGBoost is configured with appropriate hyperparameters and trained on this feature-engineered dataset to learn patterns and relationships between the features and land cover classes. The model's performance is validated, and hyperparameters may be fine-tuned for optimization. While XGBoost can be a viable option for image classification, given their ability to automatically learn features directly from pixel values. Nonetheless, XGBoost can still offer a valuable alternative, particularly in scenarios with limited labeled data or computational resources.

3.6. Normalized difference vegetation index (NDVI)

The normalized difference vegetation index (NDVI) plays a pivotal role in land use and land cover (LULC) analysis. The normalized difference vegetation index (NDVI) for the years 2015 and 2020 in the Karaivetti Bird Sanctuary was calculated to assess the vegetation dynamics and changes over the five-year period. To compute the NDVI, satellite imagery data capturing the near infrared band values (NIR) and red (R) band values were used. Equation (1) illustrates the NDVI calculation technique as:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{1}$$

It serves as a critical indicator for detecting and characterizing vegetation in satellite imagery and remote sensing data. NDVI is computed using reflectance values in the near-infrared (NIR) and red spectra, and its values range from −1 to 1. High NDVI values typically signify healthy vegetation, aiding in the identification and classification of land cover types like forests, croplands, and wetlands. Additionally, NDVI is instrumental in monitoring vegetation health, tracking seasonal changes, and assessing the environmental impact of human activities on the landscape, making it a fundamental tool in LULC analysis and ecological studies.

By comparing the NDVI values between 2015 and 2020, it is possible to observe changes in vegetation patterns and trends. Higher NDVI values indicate denser and healthier vegetation, while lower values suggest sparse or stressed vegetation. This information can be valuable for monitoring and managing the Karaivetti Bird Sanctuary, helping to understand the impact of environmental factors and human activities on its vegetation cover over the five-year period.

3.7. LULC classification

In this research, predictor features were used to build a brief model of class labels using supervised learning techniques. Contrary to unsupervised approaches, this method needs input from skilled, qualified specialists. For effective model training, it is crucial to choose training samples, and for our research, we used Google Earth to hand select each example. As shown in **Table 1**, the procedures used towards categorize Landsat images classified as LULC for the years 2015 and 2020 of water, greeneries and barren land. Different training patterns were found for each class employing visual perception, and ground truth points were used to confirm the resulting maps.

Table 1. Classes offered by LULC in Karaivetti.

S. No	Name of the class	Explanation
1	Greeneries	Vegetation
2	Barren land	Sandy and abandoned land
3	Water	Water body

A fingerprint files including the multiple variables for every LULC was created depending on data from training to choose the best classifier in ArcGIS. Enhancing the accuracy and precision of LULC classification can be achieved through the expansion of training sample sizes and the adoption of innovative machine learning techniques. We systematically collected training samples for each year to facilitate controlled classification processes.

3.8. Evaluation of accuracy

We used a variety of machine learning models to generate and validate LULC maps using two distinct approaches. The confusionMatrix() function was used in the first method to calculate and compare the classification accuracy for each of the four models (CART, SVM, XGBOOST, and RF) using the test and validation datasets. We calculated the producer's accuracy and the overall accuracy (OA), with the OA values being verified by statistically significant tests. To determine the total proportion of LULC classes that were appropriately categorized, the OA divides the entire quantity of pixels in the dataset by the number of correctly classified land cover pixels.

A measure used to evaluate the accuracy of specific land cover classes during a land use and land cover map phase is the producer accuracy. It is determined by splitting the total number of pixels that fall under a certain land cover class in the data being used with the quantity of pixels that have been correctly identified for that class. Any misclassified pixels are termed "errors of omission" in producer's accuracy assessment.

On the other hand, user accuracy assesses a LULC map's dependability in terms of how closely it matches actual findings. It is calculated through splitting the total number of pixels grouped in a given land use region with the amount of pixel properly recognized within that class. This metric provides insight into how accurately the generated map represents the real-world conditions. Misclassified pixels are also known as errors of omission in user accuracy.

To calculate the Kappa coefficient for the provided values, you would typically need the confusion matrices or contingency tables for each combination of LULC class and classification algorithm. The Kappa value assesses the agreement between the observed group and the expected classification beyond what would be expected by chance.

Here the formula for calculating Kappa coefficient shown in Equation (2):

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{2}$$

where: k is the Kappa coefficient; p_0 is the overall accuracy; p_e is the expected agreement.

The kappa coefficient quantifies the level of agreement, in terms of percentage, within a created land cover map, across the evaluation and testing information. The possibility that the test data and approval data in the surface coverage analysis are accurate is evaluated process will exhibit a reasonably close alignment. The overall accuracy has a strong correlation with the kappa coefficient. These accuracy ratings often indicate how well an accurate divided land cover map corresponds to actuality. They give a reliable method for verifying land cover because they have been utilized for validating land cover maps created at various geographic scales. We produced LULC maps from the CART, SVM, XGBOOST, and RF accuracy assessments by forecasting model outcomes using the subset of data from the research region.

4. Results

4.1. Land cover change analysis

The establishment of LULC serves as the primary basis for grouping it, with levels being separated into many categories according to the degree of precision and intent required. The extraction of LULC from different image datasets can be achieved using various methodologies. However, this process is often complex and impractical. A straightforward and effective approach for detecting LULC change in any area is to compare multiple datasets obtained from different satellites and captured on various dates, as demonstrated in some existing methodologies. The qualitative and quantitative aspects of the transition during the years 2015 and 2020 were examined in this study were determined by comparing the LULC maps generated by the best classifier. Notably, the top classifier in **Table 2** exhibited variations in the three selected land cover classes. It represents the modifications to land usage and vegetation categories among 2015 and 2020 in a specific area.

Table 2. Land cover and land use (LULC) changes from 2015 to 2020.

LULC categories	LULC changes (in hectares)	
	2015	**2020**
Greeneries	390.013	230.970
Water	262.251	251.057
Barren land	11.588	184.018

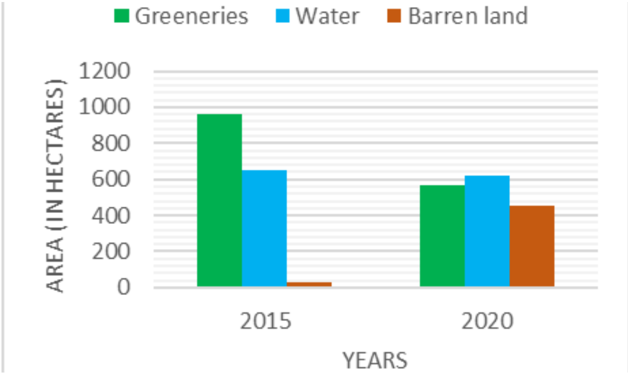

Figure 3. Distribution of LULC in the study area from 2015 to 2020.

In **Figure 3** distribution of LULC in the study area from 2015 to 2020 wherein 2015, the area covered by greeneries was 390.013 hectares, but by 2020, it decreased to 230.970 hectares, indicating a decline in vegetation such as forests, grasslands, or agricultural fields. The water area decreased slightly from 262.251 hectares in 2015 to 251.057 hectares in 2020, suggesting a minor reduction in water bodies like rivers, lakes, and ponds. On the other hand, barren land saw a significant expansion, increasing from 11.588 hectares in 2015 to 184.018 hectares in 2020. This indicates a notable growth in non-vegetated or sparsely vegetated areas such as rocky or sandy terrain. The changes in these LULC categories reflect shifts in the landscape and can be influenced by factors like urbanization, agricultural practices, and natural processes. Further analysis is needed to understand the drivers and implications of these changes within the specific study area.

4.2. Monitoring change detection

Land transformations and transitions in LULC classes between 2015 and 2020 are shown in **Figure 4**. A LULC map visually represents the various land cover categories and land use practices within a defined geographic region. It is created through the classification and mapping of satellite imagery. The map depicts the study region contains a wide variety of land cover types. These categories can include natural features like greeneries, water bodies, and barren land as well as human-influenced land use activities such as urban areas, agricultural fields, transportation networks, and industrial zones. The boundaries and spatial extent of each class are delineated based on the analysis of the satellite imagery or data used in the classification process.

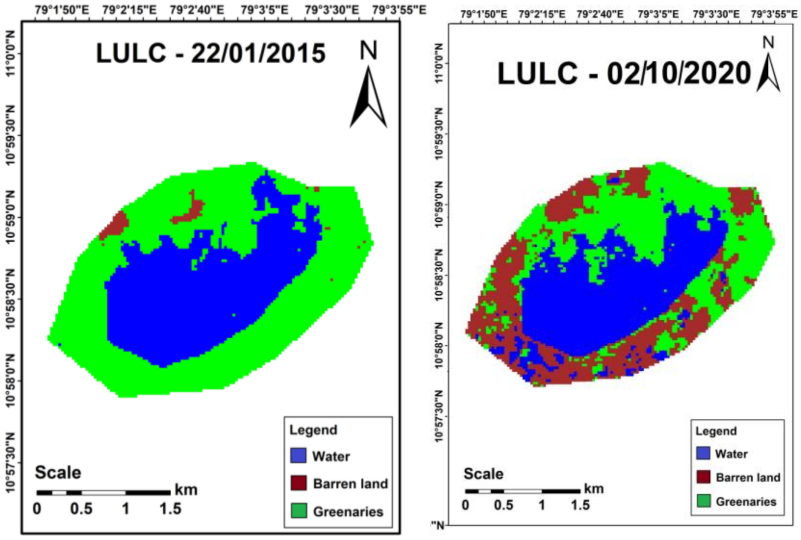

Figure 4. Changes in geographic distribution from 2015 to 2020.

The dynamics of land use and land cover (LULC) between 2015 and 2020 show notable patterns that can be divided into three categories: barren land (BL), water (WA), and Greenaries (GR). By 2020, the Greenaries would have shrunk to 231.03 hectares (34.68%) from its 389.82 hectares in 2015, as well as 58.73% of the overall area. Comparably, in 2015 the water accounted for 39.53% (262.37 hectares), but by 2020 it had dropped to 37.68% (251.00 hectares). Notably, from 11.57 hectares (1.74%) in 2015 to 184.19 hectares (27.65%) in 2020, the barren land increased dramatically. The study area increased slightly overall between 2015 and 2020, from 663.76 hectares to 666.23 hectares. This change highlights the complex interactions between generated and ecological influences that shaped the landscape during the span of the five years.

4.3. Labelling of images and performance evaluation

As illustrated in **Figure 5** and explained in **Table 3**, the dataset was utilized in this study to test supervised classification methods. To determine each classifier's accuracy, we employed a validation dataset that was distinct from the set used for training. It is important to note that these performance metrics are based on the specific dataset and evaluation criteria used, and further analysis and validation may be necessary for a comprehensive evaluation of the models performance.

Table 3. Transitions and conversions between LULC classes between 2015 and 2020.

LULC category	LULC change			
	2015		2020	
	Area		Area	
	(In hectares)	%	(In hectares)	%
GR	389.8219432	58.73	231.0334	34.68
WA	262.3694377	39.53	251.0018	37.68
BL	11.56963876	1.74	184.1923	27.65
Total	663.7610197	100	666.2275	100

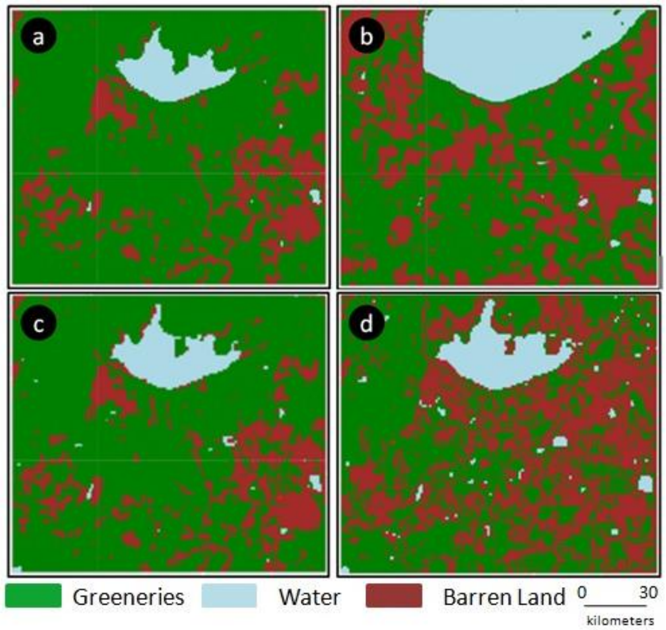

Figure 5. Comparative assessment of the classification of image algorithms, **(a)** support vector machines; **(b)** classification and regression trees; **(c)** extreme gradient boosting; **(d)** random forests.

In **Table 4** we assessed the classification performance of four machine learning algorithms regression trees, support vector machines (SVM), extreme gradient boosting (XGBoost), and random forest (RF)—for three distinct LULC categories: water, greeneries, and barren land. Notably, random forest consistently outperformed the other algorithms, exhibiting the highest accuracy across a LULC classes, producer's accuracy values ranged from 95.0% to 99.5%, while overall accuracy values ranged from 90.3% to 92.3%. These results underline the efficacy of Random Forest for accurate LULC classification, making it the preferred choice among the tested algorithms.

Table 4. Accuracy assessment of supervised classification.

LULC classes	**Classification algorithms**							
	CART		**XGBOOST**		**SVM**		**RF**	
	OA (%)	**PA (%)**	**OA (%)**	**PA (%)**	**OA (%)**	**PA (%)**	**OA (%)**	**PA (%)**
GR	88.9	92.8	90.2	94.2	89.3	93.5	91.0	95.0
WA	89.7	95.5	91.2	97.1	90.1	96.3	92.3	98.0
BL	89.2	98.1	89.8	99.0	89.5	98.5	90.3	99.5

In the evaluation of classification algorithms—CART, XGBoost, SVM, and random forest (RF)—we assessed their performance across three distinct LULC classes: greeneries (GR), water (WA), and barren land (BL). In **Table 5** the Kappa coefficients were computed to gauge the agreement between the observed classification results and the expected classification results, considering both overall accuracy and producer's accuracy.

Table 5. Accuracy assessment of supervised classification.

Algorithms	**GR Kappa**	**WA Kappa**	**BL Kappa**
CART	0.810	0.828	0.823
XGBOOST	0.853	0.867	0.861
SVM	0.836	0.850	0.844
RF	0.895	0.908	0.902

For the greeneries class (GR), the Kappa coefficients indicate the level of agreement between the classifiers and the true classifications. Among the algorithms, random forest (RF) demonstrated the strongest concordance, with Kappa values spanning from 0.810 to 0.895, indicating a significant level of agreement. Similarly, for the water class (WA), the Kappa coefficients demonstrated the algorithms' performance in accurately classifying water bodies. Once again, random forest (RF) displayed the highest agreement, with Kappa values ranging from 0.828 to 0.908. Lastly, for the barren land class (BL), the Kappa coefficients showcased the algorithms' effectiveness in classifying barren land areas. Random forest (RF) maintained its lead with Kappa values ranging from 0.823 to 0.902, indicating strong agreement with the true classifications.

These Kappa coefficient values provide insights into the reliability and accuracy of the classification algorithms in delineating LULC classes within the study area, **Figure 6** illustrates a comparison between the different machine learning models kappa indices, with Random Forest consistently demonstrating the highest agreement.

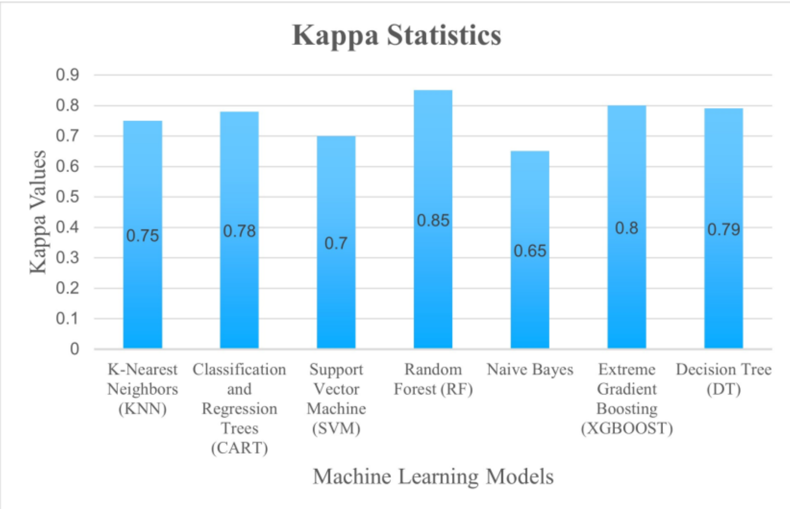

Figure 6. Comparison between the different machine learning models kappa indices.

5. Discussion

5.1. Assessment and evaluation of ML algorithms for classification

In this study, we analyzed and compared generated by several algorithms for machine learning is land cover. Our results show that all four machine learning (ML) classification models used in this instance (CART, SVM, XGBOOST, along with RF) remain reliable techniques that have the potential to reduce classification. However, it is important to note that our failure to take into account any potential nonlinearities that can result from the dynamics of the ecosystem in this area; these issues could be handled in future research applying models. However, four algorithms employed in this study have demonstrated the capacity to achieve a notably high level of classification accuracy and as a result, they may be able to outperform other supervised classifiers, as demonstrated by previous research.

When employed for analyzing particular entities, the random forest model outperformed the other three models combined, obtaining an accuracy of more than 90%, according to additional research. These results clearly show that the best technique for mapping land use and land cover is random forest. When compared to different machine learning methods, random forest consistently performs better. As a result, random forest is the generally accepted best method for LULC mapping, which is consistent with the findings of our research.

The RF technique leverages utilizing Landsat regions to divide assessment collections homogeneous subcategories, which are subsequently employed to construct individual decision trees. This approach effectively addresses challenges the drawbacks of using publicly downloadable, low-resolution satellite imagery. Within this ensemble technique, the model autonomously selects the most suitable decision trees for generating maps of the landscape produced using individual pixels analysis. The RF model emerges as a potent tool for predicting LULC with substantial computational capabilities, making it a priority in LULC mapping efforts. Despite this evidence, it's worth noting that the SVM model remains the most widely adopted classification method for monitoring LULC changes and their temporal evolution using Landsat imagery.

Through a comparative analysis of our categories of habitat using actual worldwide land surface components, we successfully validated the accuracy of our RF results. Although this method is reliable, a different validation strategy can involve comparing the results to datasets that have been locally categorized or ground-truthed data. A more accurate method of identifying land use and land cover categories can be found by comparing the usage of worldwide land coverage materials in legitimacy with neighborhood identification based on topography and the creation of existence maps. For our field of research, there are not any ground-truthed or regional datasets

5.2. Land cover variations from 2015 and 2020

Our examination of the land use and land cover (LULC) classifications generated by the RF model between 2015 and 2020 unveiled notable alterations in land cover. At this time, with an increase in croplands and populated areas and a corresponding decline in wooded regions. It is significant that the region's dense forest area shrank from an area in 2020 that was equal to a wooded area in 2015 to a smaller area in 2016. Considering the research area's frequent exposure to elevated seasonal temperatures and limited precipitation, the observed reduction in water bodies could conceivably be linked to shifts in climate patterns.

Rising temperatures brought on by climate change have been linked to higher rates of evaporation, which could lead to smaller lakes and different patterns of surface runoff. Furthermore, in addition to the effects of climate change, flood control strategies might also be a factor in the depletion of water supplies in highland regions. The results of our research provide important baseline information that is needed for the development of environmental protection plans, policy, urban planning, and evaluation of logging and farming operations in the Karaivetti area.

Our techniques, especially the novel application of Landsat imagery, indicate a major advancement in the use of machine learning algorithms for efficient land use and land cover (LULC) monitoring. Our research provides to the understanding of dynamic land cover patterns by focusing on the effects of human interventions and climate change. We promote further evaluation of these mapping methods, especially in forested areas and other areas that have been neglected in the past by satellite imagery studies. This method improves our understanding of LULC dynamics and offers insightful management advice for changing land use in environmentally sensitive areas. As a result, our work expands the scope of remote sensing research and develops the application of machine learning algorithms, resulting in advancements in the area.

6. Conclusion

Global climate change is largely attributable to human activity-induced changes in land use and land cover patterns, most notably deforestation and urbanization. The government has started a number of projects to evaluate spatiotemporal changes in an effort to prevent uncontrolled urban growth on cultivated land. In order to track changes in land utilization over time and space, this research made use of machine learning and GIS technologies. By using Google Earth Engine to process images and assessing the performance of various classifiers, we presented a novel method for detecting LULC changes.

According to our research, when compared to support vector machine (SVM), classification and regression trees (CART), and XGBoost, the Random forest (RF) classification approach had the highest accuracy in classifying LULC. The region's agricultural land has been affected by notable changes LULC including desertification and urbanization. Over a given time period, agricultural fields have gradually decreased and become more integrated into urban areas.

Moreover, the LULC map can support various applications and decision-making processes. It helps urban planners in assessing land use patterns, identifying suitable locations for infrastructure development, and managing urban growth. Environmental scientists can utilize the map to study habitat fragmentation, monitor changes in natural resources, and assess the consequences of changing vegetation cover on biodiversity. The geographical representation can be useful for controlling land use, managing disasters, and performing other regional studies.

Our approach has a number of advantages over other approaches, such as processing large datasets more efficiently and finding the best algorithm to detect LULC changes by utilizing GEE. To increase the level of classification accuracy in LULC mapping, we suggest incorporating deep learning methods. Our method has a number of advantages over earlier ones, including the ability to process huge datasets more quickly and effectively, the ability to find the utmost precise algorithm for detecting modifications to LULC and exploiting GEE are proposed. The paper also makes suggestions for further research, such as sustaining the integration of deep learning techniques to improve classification accuracy using LULC.

Finally, our research offers convincing evidence of the effects of LULC modifications on land use and climate change. After a thorough analysis, the most accurate method for classifying LULC changes was found to be the RF classification method. A few restrictions must be noted regarding this research. Initially, the results of the analysis were not entirely comprehensive due to gaps in the dataset, especially during specific time periods. A more detailed review of some variables was not feasible due to time constraints, which might have limited the depth of our conclusions. And last, even though the selected machine learning model showed good accuracy, it might not have taken into consideration every detail in the complex patterns of land use. Additional research addressing these weaknesses can provide a greater understanding of the problem. The future research methods include the integration of GEE and the possible application of deep learning techniques. Authorities and land management experts can benefit greatly from these findings, which highlight the need for sustainable practices to reduce the negative climate effects caused by land use and transformation.

Author contributions

Conceptualization, TS and RNR; methodology, TS; software, TS; validation, TS and RNR; formal analysis, TS; investigation, RNR; resources, TS; data curation, TS; writing—original draft preparation, TS; writing—review and editing, TS; visualization, RNR; supervision, RNR; project administration, RNR. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Kulkarni K, Vijaya PA. Using Combination Technique for Land Cover Classification of Optical Multispectral Images. International Journal of Applied Geospatial Research. 2021, 12(4): 22-39. doi: 10.4018/ijagr.2021100102
2. Gibril MBA, Idrees MO, Yao K, et al. Integrative image segmentation optimization and machine learning approach for high quality land-use and land-cover mapping using multisource remote sensing data. Journal of Applied Remote Sensing. 2018, 12(01): 1. doi: 10.1117/1.jrs.12.016036
3. Butt A, Shabbir R, Ahmad SS, et al. Land use change mapping and analysis using Remote Sensing and GIS: A case study of Simly watershed, Islamabad, Pakistan. The Egyptian Journal of Remote Sensing and Space Science. 2015, 18(2): 251-259. doi: 10.1016/j.ejrs.2015.07.003
4. Khachoo YH, Cutugno M, Robustelli U, et al. Machine Learning for Quantification of Land Transitions in Italy Between 2000 and 2018 and Prediction for 2050. 2022 IEEE International Workshop on Metrology for the Sea,

Learning to Measure Sea Health Parameters (MetroSea); 3 October 2022. doi: 10.1109/metrosea55331.2022.9950871
5. Khachoo YH, Cutugno M, Robustelli U, et al. Unveiling the Dynamics of Thermal Characteristics Related to LULC Changes via ANN. Sensors. 2023, 23(15): 7013. doi: 10.3390/s23157013
6. Hamad R. An Assessment of Artificial Neural Networks, Support Vector Machines and Decision Trees for Land Cover Classification Using Sentinel-2A Data. Applied Ecology and Environmental Sciences. 2020, 8(6): 459-464. doi: 10.12691/aees-8-6-18
7. Tewabe D, Fentahun T. Assessing land use and land cover change detection using remote sensing in the Lake Tana Basin, Northwest Ethiopia. Cogent Environmental Science. 2020, 6(1). doi: 10.1080/23311843.2020.1778998
8. Sharifi A. Development of a method for flood detection based on Sentinel-1 images and classifier algorithms. Water and Environment Journal. 2021, 35(3): 924-929. doi: 10.1111/wej.12681
9. Huang X, Wang Y, Li J, et al. High-resolution urban land-cover mapping and landscape analysis of the 42 major cities in China using ZY-3 satellite images. Science Bulletin. 2020, 65(12): 1039-1048. doi: 10.1016/j.scib.2020.03.003
10. Tamiminia H, Salehi B, Mahdianpari M, et al. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. ISPRS Journal of Photogrammetry and Remote Sensing. 2020, 164: 152-170. doi: 10.1016/j.isprsjprs.2020.04.001
11. Asokan A, Anitha J. Change detection techniques for remote sensing applications: a survey. Earth Science Informatics. 2019, 12(2): 143-160. doi: 10.1007/s12145-019-00380-5
12. Mishra M, Santos CAG, da Silva RM, et al. Monitoring vegetation loss and shoreline change due to tropical cyclone Fani using Landsat imageries in Balukhand-Konark Wildlife Sanctuary, India. Journal of Coastal Conservation. 2021, 25(6). doi: 10.1007/s11852-021-00840-5
13. Park J, Kim KT, Lee WH. Recent Advances in Information and Communications Technology (ICT) and Sensor Technology for Monitoring Water Quality. Water. 2020, 12(2): 510. doi: 10.3390/w12020510
14. Mugo R, Waswa R, Nyaga JW, et al. Quantifying Land Use Land Cover Changes in the Lake Victoria Basin Using Satellite Remote Sensing: The Trends and Drivers between 1985 and 2014. Remote Sensing. 2020, 12(17): 2829. doi: 10.3390/rs12172829
15. Guerrero JVR, Escobar-Silva EV, Chaves MED, et al. Assessing Land Use and Land Cover Changes in the Direct Influence Zone of the Braço Norte Hydropower Complex, Brazilian Amazonia. Forests. 2020, 11(9): 988. doi: 10.3390/f11090988
16. Chen L, Sun Y, Saeed S. Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China. PLOS ONE. 2018, 13(7): e0200493. doi: 10.1371/journal.pone.0200493
17. Wahla SS, Kazmi JH, Sharifi A, et al. Assessing spatio-temporal mapping and monitoring of climatic variability using SPEI and RF machine learning models. Geocarto International. 2022, 37(27): 14963-14982. doi: 10.1080/10106049.2022.2093411
18. Nourani V, Tootoonchi R, Andaryani S. Investigation of climate, land cover, and lake level pattern changes and interactions using remotely sensed data and wavelet analysis. Ecological Informatics. 2021, 64: 101330. doi: 10.1016/j.ecoinf.2021.101330
19. Temgoua LF, Ajonina G, Woyu HB. Land Use and Land Cover Change Analysis in Ajei Upland Watershed Community Forest, North West Region, Cameroon. Journal of Geoscience and Environment Protection. 2018, 6(9): 83-99. doi: 10.4236/gep.2018.69007
20. Pimple U, Simonetti D, Sitthi A, et al. Google Earth Engine Based Three Decadal Landsat Imagery Analysis for Mapping of Mangrove Forests and Its Surroundings in the Trat Province of Thailand. Journal of Computer and Communications. 2018, 6(1): 247-264. doi: 10.4236/jcc.2018.61025