



A machine learning approach to monitoring and forecasting spatio-temporal dynamics of land cover in Cox's Bazar district, Bangladesh from 2001 to 2019

Bishal Roy

Department of Geography and Environmental Science, Faculty of Life & Earth Sciences, Begum Rokeya University, Rangpur, 5404, Bangladesh

ARTICLE INFO

Keywords:

Random forest Algorithm
Remote sensing
GIS
Machine learning
Cox's Bazar

ABSTRACT

This study uses machine learning techniques to monitor the areal extent of land use land cover (LULC) classes in Cox's Bazar district, a well-known tourist destination in Bangladesh from 2001 to 2019. The main objective of this study is to quantify the changes in LULC classes during this period and forecasting the areal extents classes for 2019 and 2025. Satellite imagery of 2001, 2005, and 2010 from Landsat 4-5 TM and, 2015, and 2019 from Landsat 8 OLI/TIRS were used to classify the study area into four LULC classes (water bodies, urban areas, vegetation, and bare land) using Random Forest (RF) classifier. These classes were used to forecast areal extents of land cover using Cellular Automata Simulation (CAS) for 2025. The eleven LULC classes from MODIS land cover type products (IGBP scheme) were also forecasted using RF regressor to compare prediction accuracy. This study observed an increase in vegetation cover and urban settlements by 26.65% and 2.52%, respectively, with decreased water bodies and bare lands by 17.25% and 11.91%. The overall accuracy of the classification is between 89.04% to 95.24%; CAS forecast accuracy is 93.2% and RF regressor forecast accuracy ranges from 87% to 99.8%. Grid search with cross-validations was performed to boost the accuracy of both models. The results are consistent with overall global trends of LULC change.

1. Introduction

Historical comparison of Land Use Land Cover (LULC) change indicates the interaction between the environment and humans in that region, which is very important for natural resources management and sustainable development (Gillanders et al., 2008; Al Balasmeh and Karmaker, 2020). However, classifying a significant area based upon ground-based data is very time-consuming and may not be completely accurate (Alshari and Gawali, 2021). Remote sensing techniques in LULC classification significantly reduce the cost and time required for this task (Ghosh et al., 2014). Supervised learning methods, such as Artificial Neural Network, Support Vector Regression, Random Forest (RF), Maximum Likelihood, and many more can be trained to solve classification problems. The popularity of advanced machine learning and deep learning algorithms has increased rapidly; these techniques have significantly higher accuracies (van Leeuwen et al., 2020; Roy, 2021) and automate the classification process allowing for rapid change detection (Abdi, 2020). Machine learning allows for real-time classification of any area to expedite the analysis of LULC change. A dramatic increase in vegetation cover is observed worldwide, especially in the Indian region (Roy, 2021; Abdi, 2020; Chen et al., 2019; Bao et al., 2021;

Liu et al., 2018; Adepoju et al., 2019; Roy, 2020), while the population has also increased rapidly. All human activities are related to land, and change in human needs significantly modifies the land use characteristics (Bucala, 2014)

In the Chittagong division of southern Bangladesh, Cox's Bazar district is amongst the most popular tourist destinations for the last decade, attracting tourists worldwide (Ahmed, 2010). Cox's Bazar has seen a continuous increase in the number of tourists over the last decade (Ahmed, 2010; Roy, 2021), and tourism rapidly changes urban land usage (Boavida-Portugal et al., 2016). This area was also challenged with Rohingya infiltration, leading to a decrease in overall vegetation (Rashid et al., 2020). So, a comparison between LULC twenty years apart should visualize how the interaction between humans and the environment changed in this region. LULC classes were classified using RF classifier and landsat imagery and the scenario for 2025 was forecasted from this imagery using cellular automata simulation (CAS). However, the classified images can only identify visually dominant LULC classes due to the limited spatial resolution (Roy et al., 2014). The International Geosphere-Biosphere Programme (IGBP) classifies seventeen land cover types (FAO-FRA 2000) after multiple layers of processing. The higher consistencies of data (Loveland and Irons, 2016), a wide

E-mail address: prantoroy@gmail.com

<https://doi.org/10.1016/j.envc.2021.100237>

Received 7 June 2021; Received in revised form 4 August 2021; Accepted 4 August 2021

2667-0100/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

number of classes, and very high accuracies in classifying fragmented classes (Scepan and Estes, 2001) make this dataset suitable for analyzing segments of LULC classes over a long period. Forecasting LULC classes with high accuracies are crucial as it provides insights on future LULC. This data can also aid urban and transportation planning, as well as long-term disaster management. In this study, LULC class imagery and Moderate Resolution Imaging Spectroradiometer (MODIS) land cover type products (classified by IGBP scheme) were used to forecast class area extent. The temporal extent of this study is nineteen years, from 2001 to 2019. This study only uses data up to 2019 to discard any abrupt environmental changes caused by the Coronavirus and associated impacts of lockdown (Helm, 2020).

Selecting an optimum classification algorithm can boost classification accuracy significantly (Pan et al., 2020). Comparisons conclude that RF is the best algorithm for land use classification using satellite imagery (Talukdar et al., 2020). This study uses RF Classifier for land use classification, RF Regression for forecasting IGBP class area extents, and CA simulation to forecast LULC classes derived from RF classifications.

2. Data

2.1. Landsat imagery

As the study compares temporal images nineteen years apart, two datasets from these sources were selected. These images are freely available on United States Geological Survey Earth Explorer (www.earthexplorer.usgs.gov).

Landsat 5 (LS5) TM collects images in seven spectral bands, and all of their spatial resolutions are $30\text{m} \times 30\text{m}$. On the other hand, Landsat 8 (LS8) OLI collects images in eight spectral bands with the exact spatial resolution as LS5. Cloudless images for 2000, 2005, and 2010 from LS5 TM and images for 2015 and 2019 from LS8 OLI/TIRS were filtered by region using Google Earth Engine. All of the used images for classification are calibrated top-of-atmosphere (TOA) reflectance.

2.2. MODIS land cover type (MCD12Q1) v6 data product

MODIS land cover products provide LULC classification in six schemes (LP DAAC - MCD12Q1 2021). This data is derived from the supervised classification that undergoes post-processing to incorporate prior knowledge and improve the dataset. However, this dataset only provides data from 2000 to 2018 for this region. IGBP data were used as training and validation data for regression analysis and forecast estimated class area extent for the year 2019. The detailed class description is mentioned in Table 6.

2.3. Study area

The study area is Cox's Bazar district of Bangladesh (Fig. 1). This area is situated in the southeastern part of the country, along the coast of the Bay of Bengal. Cox's Bazar district is famous for being the longest uninterrupted sea beach anywhere in the world. This district has an area of $2,491.86\text{ km}^2$ and a population of 2,289,990 (Population and housing census 2011), meaning a population density of 919 people per square kilometer.

Fig 1 shows the location of our study area on a map. Our study area is defined as the land boundary of Cox's Bazar district, Bangladesh; So, all area extents were calculated inside that boundary to maintain consistency.

3. Methods

3.1. Image pre-processing

To ensure cloud-free optimal imagery for each year, Google Earth Engine filtered images by cloud cover and then selected the image with

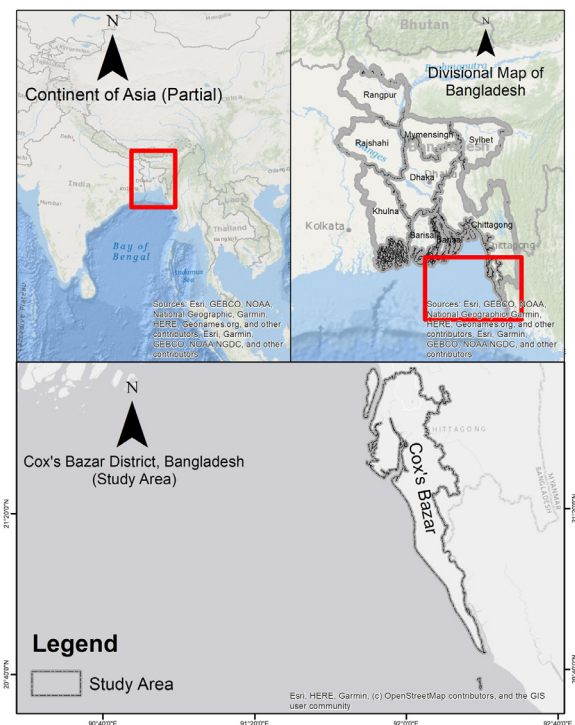


Fig. 1. Location of the study area.

the minor cloud cover. A small amount of cloud outside of the study area is present; this was ignored as it is outside of the study area. Multiple band combinations including, natural color, false color, and many more, were used to visualize features in the images.

3.2. Random forest algorithm

RF (Breiman, 2001) uses the combination of multiple individual decision trees or estimators. RF can produce very high levels of accuracy and is a very flexible algorithm used across multiple sectors on a broad range of configurations, including vegetation-related studies (Roy, 2021; Zaimes et al., 2019). This high level of accuracy and versatility is possible as RF employs multiple decision trees and uses the average of those outcomes to improve accuracy and overfitting. RF is a supervised learning algorithm, which has two models,

1. RF Classifier (RFC), and
2. RF Regressor (RFR).

This study uses RFC for LULC classification of images and RFR to forecast IGBP Land Cover classes for 2019.

3.2.1. Random Forest classifier using google earth engine

Machine learning algorithms such as RF, support vector regression, neural networks, and many more are widely adopted for their classification accuracy and speed. In LULC image classification, an RF classifier categorizes each pixel based on its association with the training data. RF is an ensemble method that aggregates the results of multiple decision trees, creating a 'forest,' also known as the bagging method (Breiman, 2001). An RF classifier can model non-linear relationship variables (Ghosh et al., 2014; Breiman, 2001), allowing it to investigate data without a linear relationship. RF algorithms can be parallelized, speeding up the process of classification. RF also handles outliers by binning them while balancing between larger and smaller datasets to minimize errors.

This study utilizes an RF classifier as it has been used in multiple studies to classify LULC data (Rashid et al., 2020; Abdullah et al., 2018; Abdullah et al., 2018; Belgiu and Drăgu, 2016). Training the model was done using multiple training sample points for four classes, as mentioned in Table 3. RF classifier available in Google Earth Engine (GEE) was used to run the model (Google Earth Engine 2021). After the classification, the images were loaded into ESRI ArcMap 10.7 to quantify the classes and calculate geostatistics described in Section 3.4. Performance of the classification model was calculated by Error Matrix, for accuracy (Section 3.3.1), Confusion Matrix, for training performance (Section 3.3.2), and Kappa Coefficient, to determine producer's and user's accuracy (Section 3.3.4).

3.2.2. Random forest regression

The non-linear nature of this model is instrumental in the regression part as a large number of datasets may not have any linearity. Regardless of the high accuracy, RFR cannot extrapolate and will only produce results based on the training data.

In this study, RF Regression was used to forecast the area extent of LULC classes calculated by IGBP. These datasets (Land Cover Type) are available from 2001 to 2018 in the Land Processes Distributed Active Archive Center (LP DAAC), which were clipped to fit our region of interest (Google Earth Engine 2021). The calculation was done as described in Section 3.5. This processed dataset (from 2001 to 2016) was then used to train the model and then measure its predictive performance using datasets of 2017 and 2018. The performance metrics are as described in Section 3.3.3. The forecasting was done in Jupyter Notebook using Python programming language (Kluyver et al., 2016; Van Rossum and Drake Jr, 1995).

3.2.3. Grid search & hyperparameters calibration

Hyperparameters are properties or model parameters in a machine learning model that controls the supervised learning process and affects the accuracy. Using the correct combination of hyperparameters ensures high accuracy in classification and predictive modeling. The RF algorithm has nineteen available parameters that can be called to modify the learning process (Ho, 1995). These parameters can have from two (Boolean) up to hundreds of possible values (for integer and floats). So, a Grid search algorithm is used to determine the best combination of hyperparameters efficiently.

This study uses GridSearchCV with the 'neg_mean_squared_error' scoring metric to find the best hyperparameter combinations (Wu et al., 2019). To test the effectiveness of the machine learning models, cross Validation, a resampling procedure to test the model (if the dataset is limited), was utilized.

3.3. Classification and regression accuracy assessment

Both RF Classifier and Regressor are predictive models used to forecast data based on the training set. The reliability and effectiveness of the results are measured by the accuracy of the model. So, to ensure the high accuracy of the results, several performance matrices are used in this study. Error Matrix, Confusion Matrix, and Kappa Coefficient are used to check the classification accuracy by comparing it to a validation set. Multiple regression matrices check the predictive accuracy of data by dividing up the dataset into training and validation sets and comparing it to them later.

3.3.1. Error matrix

Error matrix consists of actual vs. referenced data in $n \times n$ format (Error Matrix 2021). The columns represent referenced values from the validation dataset, and the rows represent classification values from the RF model. For each year, the error matrix was calculated using

(ee.ImageCollection.errorMatrix | Google Earth Engine. 2021). Error matrix accuracy quantifies the level of accuracy on the model, which is shown in Table 2.

3.3.2. Confusion matrix

A confusion matrix is the most commonly used metric for evaluating the model's classification accuracy. Without accurate classifications, any conclusion derived from that dataset will not be valid. This performance matrix is beneficial when multiple classes are in the machine learning classification problem (Diez, 2018). The confusion matrix divides up the classification results into four categories, True Positive (TP) and True Negative (TN), which are accurate predictions; False Positive (FP), and False Negative (FN), which are errors in classification. The confusion matrix is beneficial for measuring accuracy (producer's, user's, and overall) (Trucco et al., 2019). In this study, a confusion matrix was used to calculate four accuracy matrices, which are:

1. Producer's Accuracy (PA): Measures how often a ground feature is correctly represented on the classified map. PA can be calculated by subtracting False Positive (omission error) from 100.

$$PA = \frac{\text{Correctly Classified Reference Points}}{\text{Total Reference Points in Study Area}}$$

2. User's Accuracy (UA): Measures how often a classified feature is present on the ground. UA can be calculated by subtracting False Negative (commission error) from 100.

$$UA = \frac{\text{Correctly Classified Points}}{\text{Total Classified Points in Study Area}}$$

3. Overall Training Accuracy (OTA): Measures what percentage of the total dataset was classified correctly. It is the most common indicator of training accuracy in a model.

$$OTA = \frac{\text{Correctly Classified Points}}{\text{Total Reference Points in Study Area}}$$

4. Kappa Coefficient: Cohen's Kappa is used for inter-rater reliability estimation, Kappa coefficient is widely used for testing classification performance (Al Sultan et al., 2008; Mahmon et al., 2015). This is discussed further in Section 3.3.4.

3.3.3. Regression accuracy matrices

The performance metrics were calculated by dividing the dataset into 90% training and 10% validation sets. Two performance metrics were used to determine the accuracy of the regression model.

1. Mean Squared Error (MSE): Is the distance between predicted and actual value, squared to remove any negative signs. MSE also gives more weight to distances (Sammut and Webb, 2010). Calculated using Eq. (1)

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

Where:

MSE = mean squared error
 n = number of data points
 Y_i = observed values
 \hat{Y}_i = predicted values

2. Mean Absolute Percentage Error (MAPE): Regression error is the difference between the observed value and the forecasted value. Percentage errors calculated without regards to signs are absolute percentage error, and the average of all these absolute percentage errors

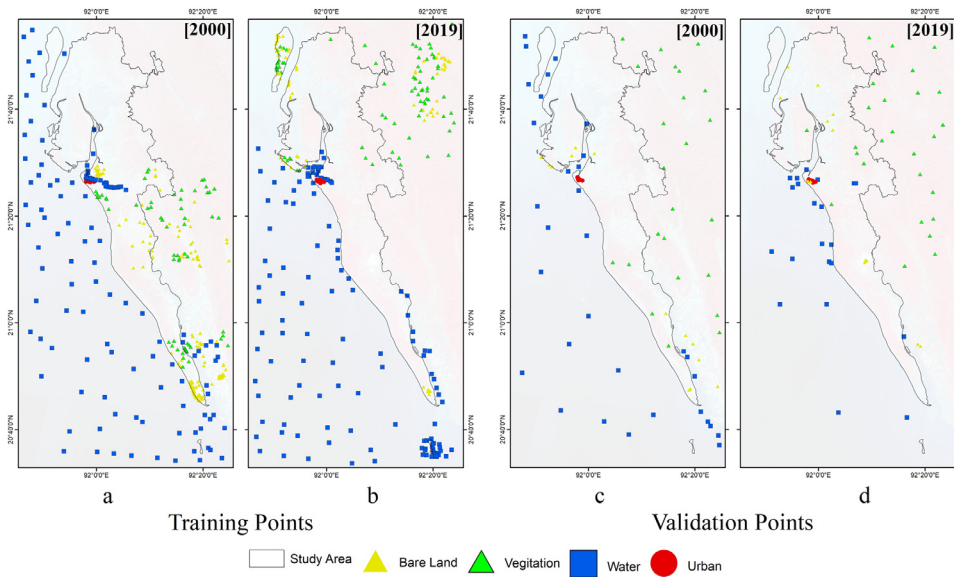


Fig. 2. Training (for RF classification) and validation (for accuracy test) sampling Point in the study area. Here (a) and (b) shows the location of training sample points for 2000 and 2019, respectively; (c) and (d) shows the location of validation sample points for 2000 and 2019, respectively. The base false-color composite was dimmed to 50% for better visualization.

are MAPE (Swamidass, 2000). MAPE provides a more straightforward interpretation of regression errors as they are expressed as a percentage. The formula for MAPE calculation is Eq. (2)

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \quad (2)$$

Where:

MAPE = Mean Absolute Percentage Error
 n = number of summation iteration
 A_i = Observed value
 F_i = Predicted value

3.3.4. Kappa coefficient

Cohen's Kappa Static is used to calculate the level of agreement between two values, considering that they will agree or disagree at some point (Viera et al., 2005). Kappa compensates for change agreement by estimating value from all confusion or error matrix components rather than entirely relying on the main diagonal (Raju et al., 2018).

Kappa Coefficient (k) is calculated in Eq. (3)

$$k = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r x_{i+} \cdot x_{+i}}{N^2 - \sum_{i=1}^r x_{i+} \cdot x_{+i}} \quad (3)$$

Where:

x_{ii} = sum of diagonal input of error matrix
 x_{i+} = sum of row i of the error matrix
 x_{+i} = sum of column i of the error matrix
 N = No. of elements in error

The output value of the Kappa coefficient (k) can be interpreted by

Table 1.

Table 1
 Explanation of Cohen's Kappa (McHugh, 2012).

Kappa	Agreement Level	Reliable Data
0- .20	None	0-4%
.21- .39	Minimal	4-15%
.40- .59	Weak	15-35%
.60- .79	Moderate	35-63%
.80- .90	Strong	64-81%
Above .90	Almost Perfect	82-100%

3.4. Cellular automata simulation and validation

Cellular automata is “a collection of cells on a grid of specified shape that evolves over time according to a set of rules driven by the state of the neighboring cells (Shi et al., 2017; Von Neumann; Singh and Gu, 2012; Wolfram, 2002). QGIS 2.18 and MOLUSCE plugin was used to simulate LULC results for 2019 and 2025. Eq. (4) shows the mathematical expression of the model,

$$S(t, t + 1) = f(S(t), N) \quad (4)$$

Where,

$S(t, t + 1)$ = system status, functioned by state of probability at any given time (N)

To simulate future LULC scenarios, LULC classes from 2000 and 2015 were considered as dependable variable. While, euclidian distance from roads, digital elevation, slope, and aspect were considered as independent variable. These variables are used to calculate pearson's correlation for each independent variable, and produce pixel-by-pixel changes table and map. In the next step, logistic regression was to predict transition potential (iterations = 100; neighborhood = 3px). The model simulates LULC map for 2019 by using the transition potential images in first iteration, and LULC map for 2025 in the second iteration.

CA simulation predicts LULC scenarios with high accuracy (Santé et al., 2010). The accuracy of the simulated scenarios were calculated using the validation method (Mansour et al., 2020; Nurwanda and Honjo, 2020; Al Kafy et al., 2020). The RFC derived image of 2019 was used to validate the simulate 2019 image for kappa coefficients ($K_{overall}$, K_{histo} , and K_{loc}), and percentage of correctness. The values show high levels of accuracy and strong agreement with K_{histo} = 0.93 and K_{loc} = 0.90. The results of overall kappa was 0.84, and percentage of correctness was 93.2%.

3.5. Projection and area calculation

The area calculation of two classified images (by RFC) and eighteen MODIS land cover type products were required for further analysis. All of these images were in the ‘World Geodetic System 1984 (WGS84)’ coordinate system (Agency et al., 2005), which is a Geographic Coordinate System (GCS). Nevertheless, for area calculations, those images must be re-projected into a Projected Coordinate System (PCS). All of these images were re-projected into Universal Transverse Mercator (UTM) ‘WGS_1984_UTM_Zone_46N’ projection system. This projection was selected as applicable between 90°E and 96°E, northern hemisphere

Table 2

Accuracies (User, producer, and overall) of L5 TM (2000, 2005, 2010) and L8 OLI (2015, 2019) images using RF classification technique.

LS	Year	Water		Urban		Vegetation		Bare Land		Accuracies		
		PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	OAT (%)	OAV (%)	K
L5 (TM)	2000	100	99.2	96.2	100	98.3	100	100	99.3	99.4	92.7	0.99
L5 (TM)	2005	99.1	99.2	96.6	100	98.3	100	100	99.4	99.4	95.2	0.99
L5 (TM)	2010	99.3	100	96.1	98.3	97.4	99.6	99.4	100	99.1	90.8	0.98
L8 (OLI)	2015	100	99.2	100	97.1	98.6	98.7	98.5	100	99.2	89.1	0.98
L8 (OLI)	2019	100	99.6	97.2	100	98.9	100	100	99.1	99.8	91.8	0.99

LS: Landsat; TM: Thematic Mapper; OLI: Optical Land Imager;

PA: Producer's accuracy; UA: user's accuracy; OA (T/V): overall accuracy (Training/ Validation).

between equator and 84°N, onshore and offshore, where our study area is located. After assigning proper projection, area calculation for each class was done using Eq. (5)

$$CA = \frac{(C \times X \times Y)}{1000000} \quad (5)$$

Where,

CA = Class Area (In Square Kilometers)

C = Cell Count

X = Cell Height

Y = Cell Width

4. Results

LULC classification results are visualized in Fig 3, while the statistics are shown in Table 4. Training and Validation sample sets are pointed in Fig 2, and the numbers of classwise sample points are in Table 3. Table 2 summarizes classwise accuracy results for all used classification accuracy matrices.

Regression analysis and forecasted class description with accuracy scores are in Table 6, while the classwise statistics are presented in Table 5.

These results are calculated in square kilometers and for the area inside the region of interest (ROI), as described in Section 2.3. The detailed findings of the study are discussed below,

4.1. Classification accuracy assessment

Accuracy assessment was done to evaluate the machine learning model's performance for this region of interest. Producer's and User's accuracy were calculated for each class, while the overall (training and validation) accuracy and kappa coefficient were calculated for each temporal set. Table 2 lists classwise and overall accuracy scores. The model was trained with 346 and 278 training sample points for 2000 and 2019, respectively. The producer's accuracy ranges from 91.1% to 100%; the user's accuracy is 99.1% to 100%. The overall training accuracy ranges from 99.1% to 99.8%, while the kappa coefficient ranges from 0.98 to 0.99.

Table 3 shows 79 and 77 validation sample points for 2000 and 2019, respectively, to calculate the validation accuracy of classified images. In Fig. 3, the classified image of 2000, yields 92.68% validation accuracy, and classified image for 2019, yields 91.78% validation accuracy. From Fig. 2, the locations of sampling points (training and validation) for 2000 and 2019 are visible. These points were selected to train the classifiers on a multitude of variants in the same class. The validation points were spaced out from training samples to calculate the actual accuracy score.

Table 3

Number of sample points used for training L5 TM (2000, 2005, 2010) and L8 OLI (2015, 2019) images for classification and validating results using error matrix.

LS	Year	Water		Urban		Vegetation		Bare Land		Validation Accuracy
		TP	VP	TP	VP	TP	VP	TP	VP	
L5 TM	2000	121	32	26	10	58	24	141	13	92.68%
L5 TM	2005	96	21	19	11	55	24	90	16	95.24%
L5 TM	2010	45	23	12	10	35	21	53	27	90.81%
L8 OLI	2015	98	21	16	10	78	31	54	17	89.04%
L8 OLI	2019	122	25	14	12	76	26	66	14	91.78%

TP: Number of Training Points; VP: Number of Validation Points

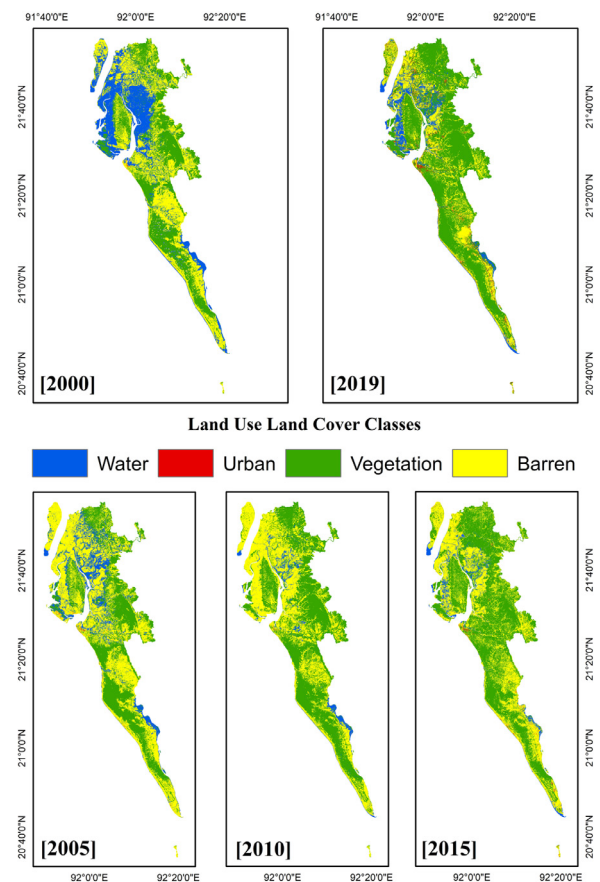


Fig. 3. Land use land cover classification results for 2000, 2005, 2010, 2015, and 2019 using RFC.

Table 4
LULC Classwise area extent derived from RF classification.

Class	2000		2005		2010		2015		2019	
	Area (KM ²)	%	Area (KM ²)	%	Area (KM ²)	%	Area (KM ²)	%	Area (KM ²)	%
Water	600.73	26.0	317.56	13.5	103.84	4.4	111.08	4.7	202.55	8.8
Urban	2.32	0.1	8.84	0.4	1.98	0.1	14.62	0.6	60.52	2.6
Vegetation	813.98	35.3	852.98	36.3	1183.7	50.4	1305.4	55.6	1428.9	62.0
Barren	890.49	38.6	1167.4	49.7	1057.3	45.1	915.68	39.0	615.56	26.7

Table 5
IGBP classwise area extent derived from MODIS MCD12Q1 v006 (2001 to 2018) and predicted area extent (2019) using RF regression.

Study Class	IGBP Class	2001				2019 (Predicted-RF)			
		Area (KM ²)	%	Total Area (KM ²)	%	Area (KM ²)	%	Total Area (KM ²)	%
Water	11	150.00	6.88	438.75	20.13	220.78	10.15	511.23	23.51
	17	288.75	13.25			290.45	13.36		
Urban	13	8.00	0.37	8.00	0.367	7.95	0.37	7.95	0.365
	02	152.00	6.97			15.61	0.72		
Vegetation	04	9.75	0.45	1677.50	76.98	5.22	0.24	1618.89	74.45
	08	205.25	9.42			216.15	9.94		
	09	53.00	2.43			65.46	3.01		
	10	117.25	5.38			253.99	11.68		
	12	1016.75	46.66			797.48	36.67		
	14	123.50	5.67			264.99	12.19		
	16	55.00	2.52			36.38	1.67		
Bare Land	16	55.00	2.52	55.00	2.52	36.38	1.67	36.38	1.67

Table 6
IGBP classification description and RF algorithm's forecast performance matrices.

Value	Plot	Description	MSE	MAPE
1	-	Evergreen Needle leaf Forests: dominated by evergreen conifer trees (canopy >2m and Tree cover>60 %)	-	-
2	Fig. 5 (a)	Evergreen Broadleaf Forests: dominated by evergreen broadleaf and palmate trees (canopy >2m and Tree cover>60 %)	0.00022	12.57%
3	-	Deciduous Needle leaf Forests: dominated by deciduous needle leaf (larch) trees (canopy >2m and Tree cover>60 %)	-	-
4	Fig. 5 (b)	Deciduous Broadleaf Forests: dominated by deciduous broadleaf (canopy >2m and Tree cover>60 %)	0.00033	2.43%
5	-	Mixed Forests: dominated by neither deciduous nor evergreen (40-60% of each) tree type (canopy >2m and Tree cover>60 %)	-	-
6	-	Closed Shrublands: dominated by woody perennials (1-2m height) (Tree cover>60 %)	-	-
7	-	Open Shrublands: dominated by woody perennials (1-2m height) (Tree cover 10-60 %)	-	-
8	Fig. 5 (c)	Woody Savannas: tree cover 30-60% (canopy >2m)	0.07882	8.18%
9	Fig. 5 (d)	Savannas: tree cover 10-30% (canopy >2m)	0.10276	12.64%
10	Fig. 5 (e)	Grasslands: dominated by herbaceous annuals (<2m)	0.00002	0.24%
11	Fig. 5 (f)	Permanent Wetlands: permanently inundated lands with 30-60% water cover and >10% vegetated cover	0.01594	4.45%
12	Fig. 5 (g)	Croplands: at least 60% of area is cultivated cropland	0.03119	5.31%
13	Fig. 5 (h)	Urban and Built-up Lands: at least 30% impervious surface area, including building materials, asphalt, and vehicles.	0.00160	1.38%
14	Fig. 5 (i)	Cropland/Natural Vegetation Mosaics: mosaics of small-scale cultivation 40-60% with natural tree, shrub, or herbaceous vegetation	0.00072	1.54%
15	-	Permanent Snow and Ice: at least 60% of the area is covered by snow and ice for at least 10months of the year	-	-
16	Fig. 5 (j)	Barren: at least 60% of the area is non-vegetated barren (sand, rock, soil) areas with less than 10%vegetation.	0.00281	4.86%
17	Fig. 5 (k)	Water Bodies: at least 60% of the area is covered by permanent water bodies	0.00698	0.22%

4.2. LULC classification and area extent

The LULC classifications were done using RF Classifier, which are shown in Fig. 3 using conventional colors. LULC classes of 2000 in Fig. 3 shows the dominant land use class as Bare Land, followed by water bodies and vegetation. The middle area and the coastal line of that ROI are dominated by bare land. Several landmasses, narrowly separated from the mainland by water, are also visible at the top of the ROI. The urban area is of such small scale that it is not visible separately from a distance.

LULC classes of 2009 in Fig. 3 shows significant changes in the LULC classes. The ROI is now heavily dominated by vegetation, followed by bare land and water bodies. However, bare lands are concentrated at the top of the ROI. There is also a concentration of bare land in the lower middle area. The coastal region on the bottom left consists of vegetation.

Images for 2005, 2010, and 2015 visualize the continuous increase of vegetation throughout the ROI. The waterbodies found the northwestern part of 2000's classified image changes from water body to barren land to vegetation. The steady increases in urban class up to 2015 is also apparent. The build-ups increased rapidly from 2015 to 2019.

LULC in the ROI has changed significantly during this period. The vegetation cover has increased 1.75 times, and the urban class has increased nearly 26 times. The urban expansion is staggering, while it only consists 2.62% of the total area. On the other hand, the bare land class has decreased by almost 30% compared to 2000. The water body in the ROI decreased by almost 65% compared to 2000. This high rate of land use pattern change indicates increased land utilization trends in this area.

The results of CA simulation for modelling future LULC scenarios are in Table 7. The results are also visualized in fig 7. The accuracy and validation scores are in Table 8.



Fig. 4. The heat map of the correlation coefficients of IGBP Classes (Separated & Combined) from 2001 to 2018.

4.3. Forecast accuracy assessment

The forecasting accuracy of a regression model is crucial for understanding how closely the model's output will be observed. As described in Section 3.3.3, I calculated two performance metrics for the regression model. These results are listed in Table 6. MSE for the model ranges from 0.00002 to 0.10276. The MSE scores are normalized, and can be interpreted independently without context.

MAPE expresses percentage errors, so the performance is comparable between datasets. MAPE values nine classes are from 0.2% to 8.2%, interpreting to model accuracy between 91.8% and 99.8%. However, three classes (2 and 9) have lower accuracy rates compared to others. Their MAPE value is 12.6%, interpreting to model accuracy of 87.4%. In summary, this model has above 91% accuracy for nine out of eleven classes and no accuracy below 87%.

LULC modelling for future scenarios validation percentage of correctness is 93.2%, interpreting to a high accuracy forecast using the simulated models. However, the simulation performs poorly in forecasting water, urban, and barren land classes; while predicting vegetation with very high accuracy (96.9%). Prediction accuracy and validation scores are mentioned in Table 8.

4.4. Changes in IGBP class extent and forecast

Data from MODIS Land Cover Type products classified according to the IGBP scheme is quantified in Table 5. The data from this classification does not show any significant changes in this area. The majority of the changes are minor, ranging from 0.001% for the urban class to 3.38% for water bodies. However, there are significant changes in subclasses of the data (IGBP classes). The changes in vegetation cover types are up to 9.99% of the entire study area. The majority of the loss in vegetation class is from IGBP class 02 and 12, which are Evergreen Broadleaf Forests and Croplands, respectively. The water body loss is from IGBP Class 11, a permanent wetland rather than a wholly submerged water body. The urban and barren land class shows insignificant changes.

Forecasting results are visualized as plots in Fig. 5. The correlation matrix in Fig. 4 shows a statistically significant correlation between these classes.

LULC forecast results and observed values are visualized in Fig 7. The simulation predicts increases in vegetation and urban buildups while decreases in waterbodies and barren land. This result is consistent with the trends observed by classified images.

5. Discussions

LULC classifications (Section 4.2) for the ROI in the two temporal extents reveal fascinating usage of the land class. Overall, the total area of vegetation class increased 1.75 times. This result supports previous studies on land cover, that vegetation has increased dramatically over the last decade (Chen et al., 2019; Roy, 2020; Dunne, 2019; Wu et al., 2020). Some areas in the top middle part of the ROI, in which land-masses seem to be narrowly separated from the coast, have been reconnected by a new barren land, mostly sandy bar. There are no build-ups around that sandy bar, suggesting natural sediment deposition; possibly due to recurrent floods and high river flow (Ashworth et al., 2000). The vegetation cover has increased from the coast in the left bottom of the ROI. A significant concentration of bare land is present in the bottom middle part of the picture, the Rohingya Refugee Camps location (Rashid et al., 2020). These camps have repurposed the surrounding vegetations to camps.

Correlation matrices in Fig. 4 show the correlations between all of the IGBP Classes; The correlation significance is denoted in the matrices. Fig 4(a) shows the correlations between all present IGBP classes in the ROI. Class 11: Permanent wetlands show very high positive correlations with vegetation classes (09, 10, and 14), high negative correlations with vegetation classes (02 and 12), and urban class 16. This mixed Correlation with multiple vegetation classes indicates that wetlands are correlated to selected vegetation types only, all of which are strong correlations. Robust correlations with bare land are expected as wetland becomes bare land when dried, and bare land becomes wetlands when submerged into water. The water bodies show negative correla-

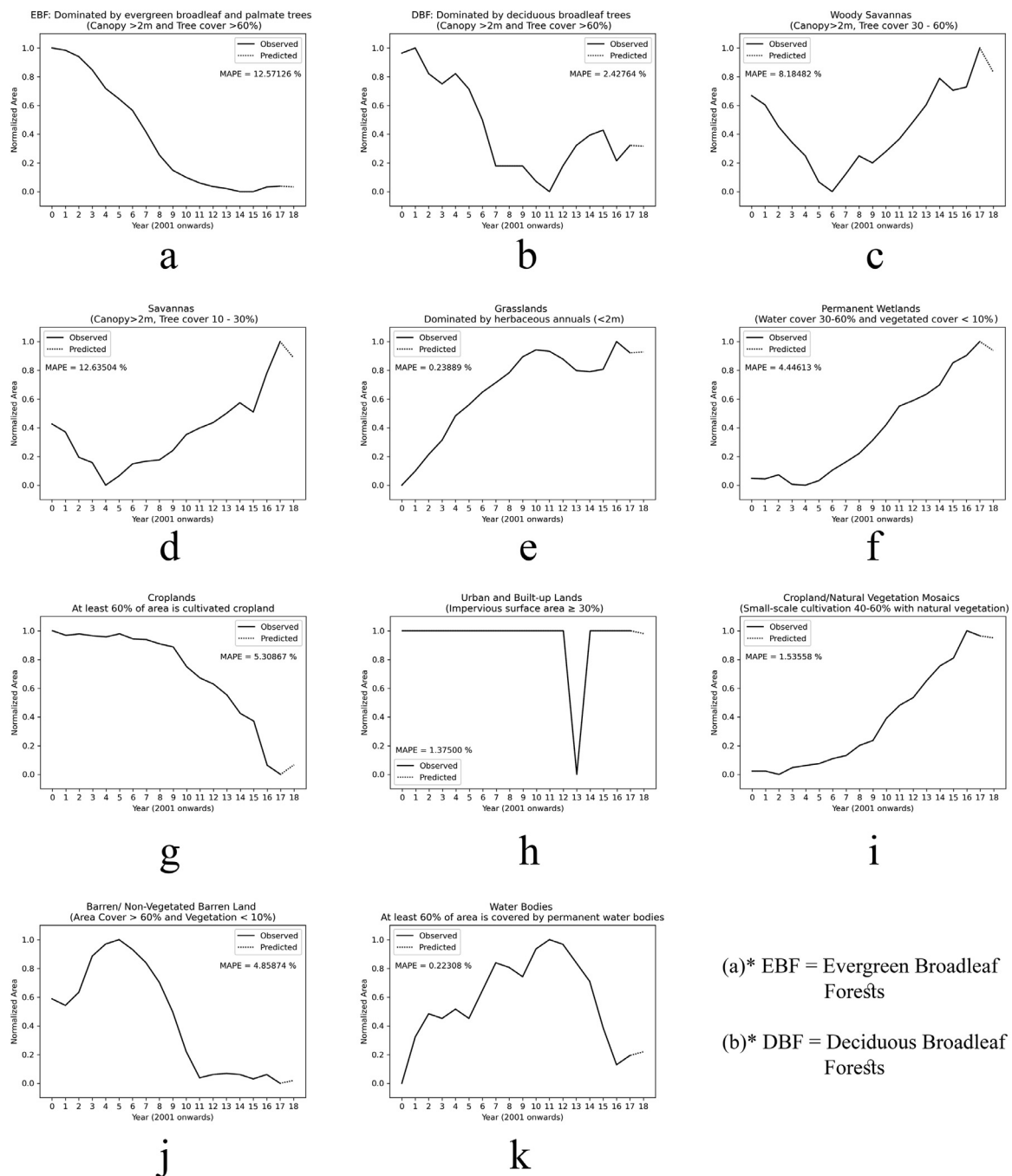


Fig. 5. Annual IGBP classification area (Square Kilometers) in the Study Area from 2001 to 2018; Class area for 2019 forecasted using RF algorithm.

tions with other vegetation types but are moderately correlated with class 10, grasslands. Urban build-ups in this region have no significant correlations with any other classes.

The vegetation classes in this region show a high correlation between these classes. These correlations are shown in Fig. 4(a); as this study only analyses all of the vegetation covers as one class, these individual class correlations require further studies.

The bare land has positive correlations with classes 02, 04, 12 (classes in forests and croplands). This land class is mainly cleared out and repurposed for agricultural or economic usage (Landsberg and Gower, 1997; Gibbs et al., 2010). Classes 11 (water), 08, 09, 10, and 14 (vegetation) show negative correlations with bare land, as drying up wetlands and repurposing vegetation classes generally result in new bare lands.

The combined class correlations in Fig. 4(b) show the correlations between the LULC classes from IGBP data. Water bodies have a robust negative correlation with vegetation and bare lands. This Correlation is most likely indicative that wetlands decrease as vegetation and bare land increase and vice versa. This phenomenon is due to reclamation of waterbodies and barren land to agricultural land as the demand for food production is increasing (India: In Uttar Pradesh, farmers green their barren lands 2021). There is also a strong positive correlation between vegetation and bare land; no anthropogenic activity defines this relationship. There are no significant correlations with any class with urban class. LULC classification supports the correlations between vegetation and water bodies. Nevertheless, this classification does not support the Correlation between any other classes. However, as correlations are based upon historical data, the correla-

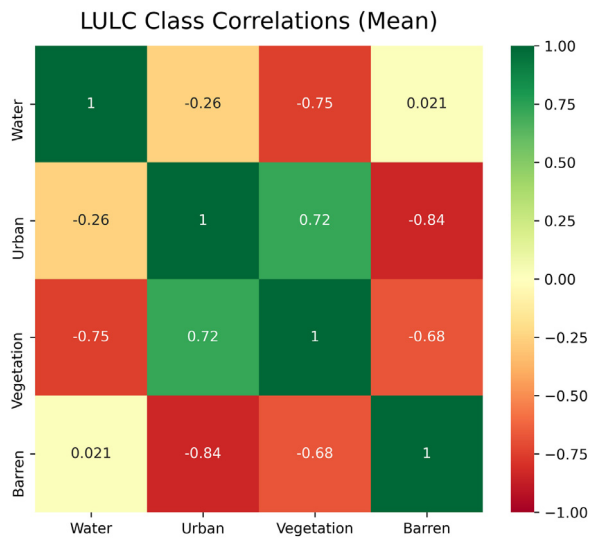


Fig. 6. The heat map of the correlation coefficients of LULC classes derived from Landsat imagery using RFC.

tions may be valid for the whole temporal extent rather than any specific year.

Fig. 5 visualizes the forecasting results of the regression model. These plots are based on normalized data of each class from MODIS IGBP classification. Min-max normalization was performed to preserve the relationship between each variable while making the datasets comparable in scale (Han et al., 2012). The class description for each plot is in Table 6. These plots can be used to understand the trends in that specific class area dynamics. Fig 5 (a) and (j) show dramatic downward trends for evergreen broadleaf forest, and barren land respectively. While Fig 5-d

Table 7

LULC classwise predicted area extent (2019 and 2025) using CA simulation.

LULC Class	2019 (Predicted)		2025 (Predicted)	
	Area (KM ²)	%	Area (KM ²)	%
Water	113.1183	4.82	96.0516	4.09
Urban	11.7252	0.50	88.8822	3.79
Vegetation	1434.3075	61.12	1796.1057	76.53
Barren	787.662	33.56	365.7735	15.59

(savannas), 5-e (grasslands) 5-f (permanent wetlands), and 5-i (croplands/natural vegetation mosaics) show rapid upwards trends. Fig. 5-b (deciduous broadleaf forest), 5-g (croplands), and 5-k (water bodies) show non-absolute downwards trends. Fig. 5-c (woody savannas) show a mixed upwards trend in data, and Fig. 5-h (urban build-ups) provides no understanding of the data. The model predicts changes in trends; woody savannas, savannas, and permanent wetlands are forecasted not to follow the upward trend and drop in value. Croplands, barren land, and water bodies are forecasted to increase in value instead of the downward trend. All other classes are forecasted to have consistent values.

Fig. 7 visualizes observed (2000–2019) and simulated (2019–2025) values for land cover extent. The simulated results are consistent with the trends observed in classified images in Fig. 3. However, as these simulations predict data based on logistic regression trendline, sudden changes in the class extent cannot be forecasted accurately. Fig. 6 shows the correlations between classes based on observed values (2000 to 2019). Waterbodies and barren land have very strong negative correlations; this is due to land reclamation for agricultural purposes (India: In Uttar Pradesh, farmers green their barren lands 2021), and natural causes (Ashworth et al., 2000). Classified images in fig 3 visualize these changes, from water body to barren land to vegetation. Urban class shows strong positive correlations with vegetation class. This is

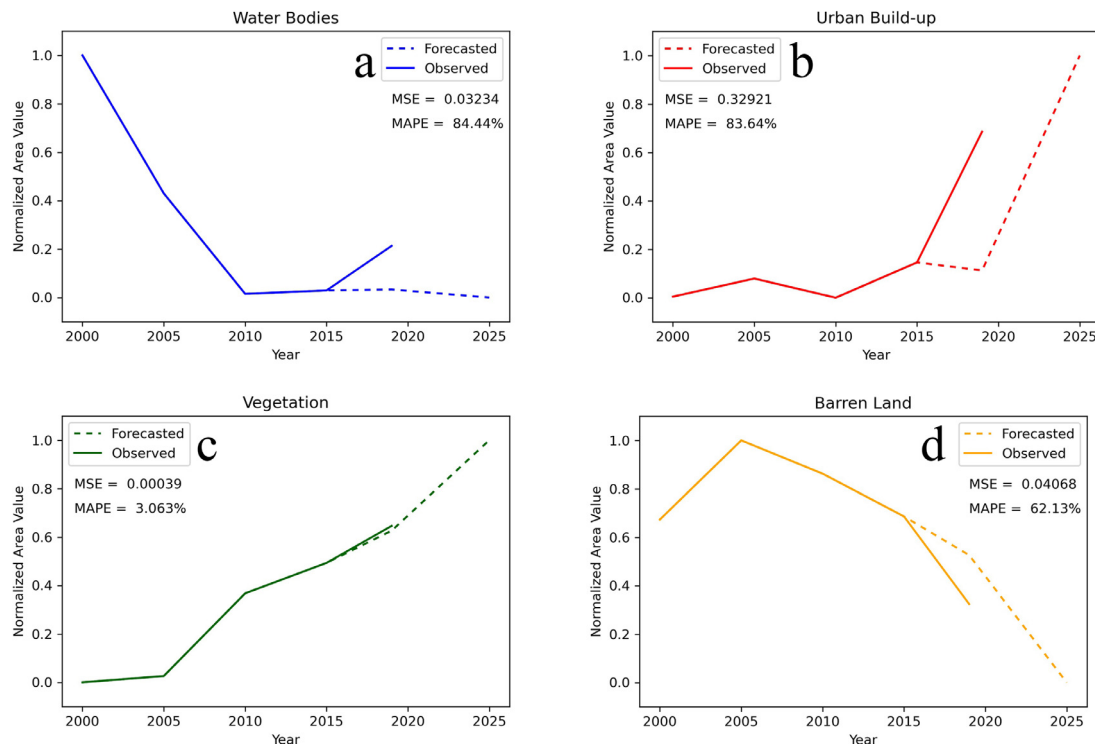


Fig. 7. LULC classification area (normalized) in the Study Area for corresponding years; Class area values for 2019 and 2025 forecasted using cellular automata approach.

Table 8
Cellular automata simulation performance metrics.

Class	Accuracy		Validation			
	MSE	MAPE	K_{loc}	K_{histo}	$K_{overall}$	% Of Correctness
Water	0.03234	84.44%	0.90270	0.93162	0.84098	93.19193
Urban	0.32921	83.64%				
Vegetation	0.00039	3.063%				
Barren	0.04068	62.13%				

also due to repurposing lands for agricultural purposes (Carlet et al., 2017).

This study area was selected as tourism, and the refugee crisis has dramatically affected the land usage of this region; nevertheless, despite all of that, the worldwide trend of rapid vegetation increase is visible here as well. However, while the LULC classification concludes and confirms results, IGBP classifications are inconsistent with the findings and show no dramatic changes. In this study, previous research and field visits were used to understand the reasons behind land class extent change and aid the training of classification algorithms.

6. Conclusion

Bangladesh is the world's 8th most populated country and the 94th in terms of land area; the rapid and dynamic demands for lands shape the land usage around the country. Cox's Bazar is also one of the most popular tourist destinations in this country. So, understanding the land usage in this region will allow for better planning in the future. This study used multi-spectral satellite images with high accuracy machine learning classifiers to produce LULC classification maps spanning over twenty years, which allowed for comparison, and analysis. This study also utilized an RF regression model, and CA simulations to forecast how the land cover extent will change based on historical data. Both LULC classification and IGBP classes indicate a rapid increase in vegetation and urban classes while the water bodies and bare land decreased. The overall classification findings are consistent with the worldwide results of increased vegetation in the larger region (Chen et al., 2019; Roy, 2020; Dunne, 2019; Wu et al., 2020), along with deforested bare land at the location of Rohingya camp (Rashid et al., 2020). The dramatic decrease in bare land can be correlated to increased human economic activities due to increased population (Roy et al., 2021; Garg, 2017). This large population results in increased water demand (Boretti and Rosa, 2019), but this study reaffirms that the water bodies are shrinking, creating even more shortages. Climate change and human activities are both responsible for the demise of water bodies at an alarming rate (Stout, 1990; Okello et al., 2015). The CA simulation forecasts how the areal extent may change in the future.

The accuracy of this study can be significantly increased by using more advanced machine learning algorithms such as Artificial Neural Networks. This study is limited by the available high-resolution historical satellite imagery and climatic data. Integration of local knowledge for classification and atmospheric or biological factors for forecasting may improve the accuracy significantly. Separate studies in urban and forest parts of the study are also required for understanding the actual factors behind these dynamics. The Correlation found between sub-classes of vegetation needs to be studied further to understand why different types of vegetation correlated with the same class oppositely. The results and findings of this study may help future sustainable planning and conservation of nature.

Declaration of Competing Interest

The author states that there is no conflict of interest.

References

- Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sens.* 57 (1), 1–20. doi:10.1080/15481603.2019.1650447.
- Abdullah, H.M., Islam, I., Miah, M.G., Ahmed, Z., 2018. Quantifying the spatiotemporal patterns of forest degradation in a fragmented, rapidly urbanizing landscape: a case study of Gazipur, Bangladesh. *Remote Sens. Appl. Soc. Environ.* 13, 457–465. doi:10.1016/j.rsase.2019.01.002, August/Jan. 2019.
- Adepoju, K., Adelabu, S., Fashae, O., 2019. Vegetation response to recent trends in climate and Landuse dynamics in a typical humid and dry tropical region under global change. *Adv. Meteorol.* 2019, 1–15. doi:10.1155/2019/4946127, Dec.
- Agency, N. G., et al., "World geodetic system 1984," pp. 0–2, 2005.
- Ahamed, S.S., 2010. Impact of tourism in Cox's Bazar, Bangladesh. *North* 92.
- Al Balasmeh, O.I., Karmaker, T., 2020. Effect of temperature and precipitation on the vegetation dynamics of high and moderate altitude natural forests in India. *J. Indian Soc. Remote Sens.* 48 (1), 121–144. doi:10.1007/s12524-019-01065-8.
- Al Kafy, A., Rahman, M.S., Faisal, A.A., Hasan, M.M., Islam, M., 2020. Modelling future land use land cover changes and their impacts on land surface temperatures in Rajshahi, Bangladesh. *Remote Sens. Appl. Soc. Environ.* 18, 100314. doi:10.1016/j.rsase.2020.100314.
- Al Sultan, S., Lim, H.S., Mat Jafri, M.Z., Abdullah, K., Mohd Saleh, N., 2008. Innovative satellite image map of R. Alkhafra Area, Saudi Arabia using high resolution image. In: 21st Congr. Int. Soc. Photogramm. Remote Sensing, ISPRS, 37, pp. 1355–1357 2008.
- Alshari, E.A., Gawali, B.W., 2021. Development of classification system for LULC using remote sensing and GIS. *Glob. Transitions Proc.* 2 (1), 8–17. doi:10.1016/j.gltp.2021.01.002, Jun.
- Ashworth, P., Best, J., Roden, J., Bristow, C., Klaassen, G., 2000. Morphological evolution and dynamics of a large, sand braid-bar, Jamuna River, Bangladesh. *Sedimentology* 47, 533–555. doi:10.1046/j.1365-3091.2000.00305.x, Jun.
- Bao, Z., et al., 2021. The sensitivity of vegetation cover to climate change in multiple climatic zones using machine learning algorithms. *Ecol. Indic.* 124, 107443. doi:10.1016/j.ecolind.2021.107443, May.
- Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. doi:10.1016/j.isprsjprs.2016.01.011, Elsevier B.V. Apr. 01.
- Boavida-Portugal, I., Rocha, J., Ferreira, C.C., 2016. Exploring the impacts of future tourism development on land use/cover changes. *Appl. Geogr.* 77, 82–91. doi:10.1016/j.apgeog.2016.10.009, Dec.
- Boretti, A., Rosa, L., 2019. Reassessing the projections of the World Water Development Report. *npj Clean Water* 2 (1), 15. doi:10.1038/s41545-019-0039-9.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. doi:10.1023/A:1010933404324.
- Bucala, A., 2014. The impact of human activities on land use and land cover changes and environmental processes in the Gorce Mountains (Western Polish Carpathians) in the past 50 years. *J. Environ. Manage.* 138, 4–14. doi:10.1016/j.jenvman.2014.01.036, Jun.
- Carlet, F., Schilling, J., Heckert, M., 2017. Greening U.S. legacy cities: urban agriculture as a strategy for reclaiming vacant land. *Agroecol. Sustain. Food Syst.* 41 (8), 887–906. doi:10.1080/21683565.2017.1311288, Sep.
- Chen, C., et al., 2019. China and India lead in greening of the world through land-use management. *Nat. Sustain.* 2 (2), 122–129. doi:10.1038/s41893-019-0220-7.
- Diez, P., 2018. Introduction. In: *Smart Wheelchairs and Brain-computer Interfaces: Mobile Assistive Technologies*. Elsevier, pp. 1–21.
- Dunne, D., 2019. One third of the world's new vegetation is in India and China, data shows | World Economic Forum. Carbon Brief. (accessed May 27, 2021) <https://www.weforum.org/agenda/2019/02/one-third-of-world-s-new-vegetation-in-china-and-india-satellite-data-shows>.
- "ee.ImageCollection.errorMatrix | Google Earth Engine." <https://developers.google.com/earth-engine/apidocs/ee-imagecollection-errormatrix> (accessed May 25, 2021).
- "Error Matrix." https://gis.humboldt.edu/OLM/Courses/GSP_216_Online/lesson6-2/error-matrix.html (accessed May 25, 2021).
- FAO-FRA, "Forest cover mapping & monitoring with Noaa-Avhr & other coarse spatial," 2000.
- Garg, S., "Impact of overpopulation on land use pattern," 2017, pp. 137–154.
- Ghosh, A., Sharma, R., Joshi, P.K., 2014. Random forest classification of urban landscape using Landsat archive and ancillary data: combining seasonal maps with decision level fusion. *Appl. Geogr.* 48, 31–41. doi:10.1016/j.apgeog.2014.01.003, Mar.
- Gibbs, H.K., et al., 2010. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc. Natl. Acad. Sci. USA* 107 (38), 16732–16737. doi:10.1073/pnas.0910275107, Sep.

- Gillanders, S.N., Coops, N.C., Wulder, M.A., Gergel, S.E., Nelson, T., 2008. Multitemporal remote sensing of landscape dynamics and pattern change: describing natural and anthropogenic trends. *Progr. Phys. Geogr.* 32 (5), 503–528. doi:10.1177/0309133308098363, SAGE Publications Sage UK Oct. 01.
- Google Earth Engine, “Supervised Classification,” 2021. <https://developers.google.com/earth-engine/guides/classification>.
- Han, J., Kamber, M., Pei, J., 2012. 3 - Data preprocessing. In: Han, J., Kamber, M. (Eds.), *The Morgan Kaufmann Series in Data Management Systems*. Morgan Kaufmann, Boston, pp. 83–124. J. B. T.-D. M. (Third E. Pei, Eds).
- Helm, D., 2020. The environmental impacts of the coronavirus. *Environ. Resour. Econ.* 76 (1), 21–38. doi:10.1007/s10640-020-00426-z.
- Ho, T.K., 1995. Random decision forests. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1, pp. 278–282.
- “India: In Uttar Pradesh, farmers green their barren lands.” (accessed Aug. 03, 2021) <https://www.worldbank.org/en/news/feature/2006/10/18/india-in-uttar-pradesh-farmers-green-their-barren-lands> (accessed Aug. 03, 2021).
- Kluyver, T., et al., 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pp. 87–90.
- Landsberg, J.J., Gower, S.T., 1997. 2 - Forest biomes of the World. In: Landsberg, J.J., of P. E. to F. M. Gower, S.T.B.T.-A. (Eds.), *Physiological Ecology*. Academic Press, San Diego, pp. 19–51e.
- Liu, H., Zheng, L., Yin, S., 2018. Multi-perspective analysis of vegetation cover changes and driving factors of long time series based on climate and terrain data in Hanjiang River Basin, China. *Arab. J. Geosci.* 11 (17), 509. doi:10.1007/s12517-018-3756-3.
- Loveland, T.R., Irons, J.R., 2016. Landsat 8: the plans, the reality, and the legacy. *Remote Sens. Environ.* 185, 1–6. doi:10.1016/j.rse.2016.07.033.
- “LP DAAC - MCD12Q1.” <https://lpdaac.usgs.gov/products/mcd12q1v006/> (accessed May 26, 2021).
- Mahmon, N.A., Ya'acob, N., Yusof, A.L., 2015. Differences of image classification techniques for land use and land cover classification. *IEEE Control Syst. Soc.* 6–8.
- Mansour, S., Al-Belushi, M., Al-Awadhi, T., 2020. Monitoring land use and land cover changes in the mountainous cities of Oman using GIS and CA-Markov modelling techniques. *Land Use Policy* 91, 104414. doi:10.1016/j.landusepol.2019.104414.
- McHugh, M.L., 2012. Interrater reliability: the kappa statistic. *Biochem. Medica* 22 (3), 276–282. doi:10.11613/BM.2012.031.
- Nurwanda, A., Honjo, T., 2020. The prediction of city expansion and land surface temperature in Bogor City, Indonesia. *Sustain. Cities Soc.* 52, 101772. doi:10.1016/j.scs.2019.101772.
- Okello, C., Tomasello, B., Greggio, N., Wambiji, N., Antonellini, M., 2015. Impact of population growth and climate change on the freshwater resources of Lamu Island, Kenya. *Water* 7 (3). doi:10.3390/w7031264.
- Pan, L., Gu, L., Ren, R. and Yang, S. “Land cover classification based on machine learning using UAV multi-spectral images,” Aug. 2020, vol. 11501, p. 50, doi: 10.1117/12.2566128.
- “Population and housing census,” 2011.
- Raju, R.S., Sudarsana Raju, G. and Rajasekhar, M. “Land use/land cover change detection analysis using supervised classification, remote sensing and GIS In Mandavi river basin, YSR Kadapa District, Andhra Pradesh, India,” vol. 9, no. 3, pp. 46–54, 2018.
- Rashid, K.J., Hoque, M.A., Esha, T.A., Rahman, M.A., Paul, A., 2020. Spatiotemporal changes of vegetation and land surface temperature in the refugee camps and its surrounding areas of Bangladesh after the Rohingya influx from Myanmar. *Environ. Dev. Sustain.* (0123456789) doi:10.1007/s10668-020-00733-x.
- Roy, D.P., et al., 2014. Landsat-8: science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172. doi:10.1016/j.rse.2014.02.001.
- Roy, B., Bari, E., Nipa, N.J., Ani, S.A., 2021. Comparison of temporal changes in urban settlements and land surface temperature in Rangpur and Gazipur Sadar, Bangladesh after the establishment of city corporation. *Remote Sens. Appl. Soc. Environ.* 23, 100587. doi:10.1016/j.rsase.2021.100587, Aug.
- Roy, B. “Assessment of vegetation health in Saint Martin’s Island, Bangladesh using remote sensing & GIS,” vol. 7, pp. 89–93, 2020, doi:10.6084/m9.figshare.12674426.
- Roy, B., 2021. Optimum machine learning algorithm selection for forecasting vegetation indices: MODIS NDVI & EVI. *Remote Sens. Appl. Soc. Environ.* 23, 100582. doi:10.1016/j.rsase.2021.100582, Aug.
- Roy, B. “Land use & socio-environmental studies of saint Martin’s Island in 2020,” Roy, B. (2021). *Land Use & Socio-Environmental Studies Of Saint Martin’s Island in 2020*. <https://doi.org/10.6084/m9.figshare.14376347>, 2021.
- Sammur, C., Webb, G.I., 2010. Mean squared error. In: *Encyclopedia of Machine Learning*. Springer US, Boston, MA, p. 653.
- Santé, I., García, A.M., Miranda, D., Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. *Landscape Urban Plan.* 96 (2), 108–122. doi:10.1016/j.landurbplan.2010.03.001.
- Scepan, J., Estes, J.E., 2001. Thematic validation of global land cover data sets-procedures and interpretation methods. In: *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, 3, pp. 1119–1121. doi:10.1109/IGARSS.2001.976765 vol.3.
- Shi, Z., Fonseca, J.A., Schlueter, A., 2017. A review of simulation-based urban form generation and optimization for energy-driven urban design. *Build. Environ.* 121, 119–129. doi:10.1016/j.buildenv.2017.05.006.
- Singh, V., Gu, N., 2012. Towards an integrated generative design framework. *Des. Stud.* 33 (2), 185–207. doi:10.1016/j.destud.2011.06.001.
- Stout, G.E., 1990. Climate and water. *Eos, Trans. Am. Geophys. Union* 71 (12), 339. doi:10.1029/90EO00112.
- Swamidass, P. M. Ed., “MAPE (mean absolute percentage error) Mean absolute percentage error (Mape),” in *Encyclopedia of Production and Manufacturing Management*, Boston, MA: Springer US, 2000, p. 462.
- Talukdar, S., et al., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—a review. *Remote Sens.* 12 (7). doi:10.3390/rs12071135.
- Trucco, E., et al., 2019. Validation. In: *Computational Retinal Image Analysis*. Elsevier, pp. 157–170.
- van Leeuwen, B., Tobak, Z., Kovács, F., 2020. Machine learning techniques for land use/land cover classification of medium resolution optical satellite imagery focusing on temporary inundated areas. *J. Environ. Geogr.* 13 (1–2), 43–52. doi:10.2478/jen-geo-2020-0005, Apr.
- Van Rossum, G., Drake Jr, F.L., 1995. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.
- Viera, A.J., Garrett, J.M., Viera, Anthony J., Garrett, Joanne M., 2005. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37(5):360–63. *Fam. Med.* 37 (5), 360–363. 2005, [Online]. Available http://www1.cs.columbia.edu/~julian/courses/CS6998/Interrater_agreement.Kappa_statistic.pdf.
- Von Neumann, J. “THE GENERAL AND LOGICAL THEORY OF AUTOMATA.”
- Wolfram, S., 2002. *A New Kind of Science*, 5. Wolfram media, Champaign, IL.
- Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., Deng, S.-H., 2019. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* 17 (1), 26–40. doi:10.11989/JEST.1674-862X.80904120.
- Wu, J., et al., 2020. Monitoring tropical forest degradation and restoration with satellite remote sensing: a test using Sabah Biodiversity Experiment. *Adv. Ecol. Res.* 62, 117–146 Academic Press Inc.
- Zaimes, G.N., Gounaridis, D., Symeonakis, E., 2019. Assessing the impact of dams on riparian and deltaic vegetation using remotely-sensed vegetation indices and Random Forests modelling. *Ecol. Indic.* 103, 630–641. doi:10.1016/j.ecolind.2019.04.047.