Spatio-Temporal Urban Land Use Change in Mumbai, India: Analysis and Prediction of 2030 Using Satellite Data and a Cellular Automata-Markov Chain Model

Kavathekar, V.,[1,2 *] Tripathy, A. K.[2] and Chettri, S. K.[1]
[1]Computer Science and Engineering, Computer Applications, Assam Don Bosco University, Assam, India
E-mail: vsengg27@gmail.com,* sarat.chettri@dbuniversity.ac.in
[2]Computer Engineering, Don Bosco Institute of Technology, Mumbai, India, E-mail: amiya@dbit.in
**Corresponding Author*

DOI: https://doi.org/10.52939/ijg.v21i5.4163

Abstract

Mumbai, India's economic hub, continues to grow due to its strategic location, robust infrastructure, and thriving economy. This rapid development, driven by nationwide migration, is transforming resource rich land into urban spaces. Understanding past land transformations and predicting future urban growth is crucial for sustainable planning. This study analyses Land Use and Land Cover (LULC) changes in Mumbai using multi-temporal Landsat data (1990, 2000, 2010, and 2020) and forecasts changes for 2030 and 2040. Supervised classification techniques Support Vector Machine (SVM), Random Forest (RF), and Maximum Likelihood Classifier (MLC) were applied. SVM outperformed by achieving 95.63% overall accuracy and a Kappa coefficient of 0.92, and thus used for future predictions. A Python-based Cellular Automata–Markov Chain (CA-MC) model was developed using multiple inputs to simulate future land transformations. The model was validated by predicting 2020 LULC and comparing it to actual classified data, achieving 95.24% accuracy. Results show a steady rise in urban land from 2000 to 2020, accompanied by notable declines in water bodies, water vegetation, dense vegetation, and barren land. Urban areas currently cover 45.83% of the 654.20 km² study area, projected to increase to 47.25% by 2030 and 55.84% by 2040. Between 2000 and 2020, water bodies declined by 16.80%, dense vegetation by 11.26%, and barren land by 37.29%. Urban expansion originated in South Mumbai and continues northward, converting natural landscapes into built-up areas. These insights support data-driven urban and environmental planning.

Keywords: Cellular Automata Markov Chain (CA-MC), Classification, Land Use Land Cover (LULC), Prediction, Supervised Classification, Support Vector Machine (SVM)

1. Introduction

Over recent decades, urbanization has surged, resulting in more than half of the global population residing in urban areas. This upward trend is projected to persist, with urban residents having increased from about one-third in 1950 to a forecasted two-thirds by 2050. Meanwhile, the world's population currently nearing 8 billion is expected to climb to approximately 9.7 billion by the middle of the century [1]. Most of this population is migrating from the regions like rural to urban in search of employment, causing a gradual increase in urban population density. Achieving sustainable progress requires effective administration of urban growth to develop sustainable cities worldwide [2].

The swift expansion of cities alongside increasing population growth have led to significant environmental degradation, causing issues such as land degradation, deteriorating water quality, deforestation, and various forms of pollution, including air, noise, dust, and thermal pollution. These pollutants have diverse effects on human health [3]. Additional challenges include managing solid and hazardous waste, addressing housing needs, and protecting human health [4]. A crucial factor contributing to environmental decline is the alteration of natural landscapes, especially the replacement of green spaces with urban developments. This change amplifies the urban heat island (UHI) phenomenon, increasing thermal discomfort within urban areas [5].

Understanding land cover referring to the Earth's physical surface features and land use how humans utilize and alter these areas is crucial for accurate environmental monitoring and related fields [6]. In LULC changes, the concept of land cover describes the land with natural resources including vegetation, water bodies, and forests. Whereas land uses refer to how humans utilize the land for activities like agriculture, conservation, construction, and industrialization [7]. Changes in LULC involve the transformation of land from one state to another. Analyzing LULC dynamics presents a significant challenge but plays an important role in supporting informed decision making [8].

With advancements in the space industry and the increasing accessibility of both free and commercial satellite imagery, it has become more feasible to analyze these images to study spatio-temporal variations in LULC, commonly known as land transformation. Conducting such analysis is crucial for tracking transformations in land surfaces over time, offering key insights into environmental changes, urban growth, deforestation, and other landscape alterations [9]. A key challenge in satellite imaging is image classification. A range of methods, such as ANN, machine learning, and deep learning, are employed to address this challenge. Remotely sensed images often exhibit intricate and diverse patterns, along with rich spatio-temporal spectral data, which necessitate advanced processing methods. These methods are extensively used in remote sensing for applications like LULC mapping, environmental data retrieval, data integration, resolution enhancement, and predictive analysis [10].

Various machine learning approaches, including both supervised and unsupervised techniques, are widely applied in LULC classification using remote sensing imagery. Popular unsupervised methods include K-means, ISODATA (Iterative Self-Organizing Data Analysis), Fuzzy C-means, and self-organizing maps (SOM). Among supervised approaches, commonly used classifiers are the maximum likelihood classifier (MLC), k-nearest neighbor (kNN), Mahalanobis distance, Parallelepiped, minimum distance classifier, support vector machines (SVM), and Random Forest (RF) [11] and [12]. As per the previous research for LULC classification SVM is one of the best methods used by many researchers [13]. After performing LULC classification, predicting LULC changes has garnered significant interest among researchers. Various approaches have been utilized to predict future changes in LULC, such as multi-layer perceptron's, linear regression techniques, cellular automata, Markov chain models, regression trees, and artificial neural networks. These techniques assist in examining spatial and temporal trends, leading to more accurate forecasts of land use dynamics.[14].

In India, various studies have utilized predictive modeling techniques to analyze LULC changes [15]. For instance, in Uttar Pradesh an artificial neural network (ANN) method integrated with QGIS revealed projected urban built-up area increases of 14.7% by 2019, 15.7% by 2022, and 18.68% by 2031[16]. In Tamil Nadu's Bhavani Basin, the QGIS MOLUSCE plugin and MLP-ANN predicted significant increases in grass land (20 km²) and built-up areas (10 km²) by 2025 and 2030 [17]. In Ranchi, the Markov transition matrix validated against the original LULC of 2015 [18]. Meanwhile, in Varanasi district, to predict LULC changes for the year 2015, three hybrid models stochastic Markov chain (ST-MC), cellular automata-Markov chain (CA-MC), and multi-layer perceptron-Markov chain (MLP-MC) were evaluated using data from 1988 and 2001[19].

In India, Mumbai is a highly urbanized and economically significant city. However, there has been a very less research is conducted on its urban dynamics in the past. Previous studies on the Mumbai Metropolitan Region (MMR) have utilized MLPNN-MCM models with TerrSet software to predict LULC changes [20]. However, these studies mainly focus on the broader MMR, which includes Mumbai city and its suburbs, without dedicated research specifically analysing LULC transformation in the city centre. Furthermore, studies indicate that the UHI effect has intensified in the past few decades due to transformation in LULC across MMR [5]. Future urbanization is expected to further modify the city's microclimate, potentially exacerbating the UHI effect and influencing local weather patterns [21].

Previous studies have largely overlooked the evolving dynamics of Mumbai's city center, leaving a notable gap in the research. Despite being one of India's most densely populated and economically vital cities, with over 20 million residents [22], most prediction models have concentrated on suburban areas [20]. In contrast, the city center already highly urbanized has been largely overlooked. Regions like Sanjay Gandhi National Park and the coastal zones are particularly vulnerable to future environmental pressures. Ensuring sustainable development in this core region, while preserving environmental quality, quality of life, and overall well-being, is essential. Yet, the specific patterns and challenges of central urbanization remain underexplored. Therefore, comprehensive analysis and accurate forecasting of urban growth in the city center are vital for effective planning and policy-making.

Secondly, according to the studies, software tools such as QGIS or TerrSet plugins is commonly applied to predicting LULC changes. Most of the research have relied on existing software plugins to forecast future LULC transformations. As far as the authors are aware, this study is the first to design and apply a CA-MC model using multi-input variable specifically for the city center of Mumbai. This was achieved by developing a custom model using Python software, rather than relying on pre-existing tools.

This study sets out to accomplish the following primary objectives: (i) This study investigates changes in LULC between 2000 and 2020 using satellite imagery and supervised classification technique, (ii) to predict LULC trends for 2030 and 2040 by utilizing the CA-MC model, (iii) to evaluate spatial and temporal transformations over the last thirty decades and the anticipated transformation in the next decade, with a specific focus on urban expansion from 2000 to 2030 and 2040. The LULC prediction for 2030 and 2040 are intended to assist local authorities in formulating sustainable climate and environmental policies by providing valuable insights and effective urban and regional planning.

2. Study Area

Mumbai, the capital city of Maharashtra, was formerly known as Bombay. Figure 1 illustrates the geographical location of both Maharashtra and Mumbai. As India's most populous city and a key financial center, Mumbai forms the core of the Mumbai Metropolitan Region, which ranks sixth globally in population, accommodating over 23 million people under the jurisdiction of the Brihanmumbai Municipal Corporation [22] and [23]. Recognized in 2008, Mumbai earned the designation of an alpha world city [19]. The city is moderately warmer throughout the year, with significant humidity due to boundaries. Its tropical setting ensures minimal fluctuations in temperature across different seasons. The average temperature is around 27.2 °C [24]. Based on the Köppen–Geiger classification, the region experiences a warm and humid climate (Aw). Yearly rainfall ranges between 180 cm and 248 cm, with the southwest monsoon bringing most of the precipitation from June to September. Winter lasts from December to February, while summer occurs between March and May [25]. Mumbai, formed by seven islands, serves as India's powerhouse. It stands among the top ten global financial hubs [26] and [27]. The city accounts for a quarter of India’s industrial production, a significant share of maritime trade, and major capital transactions [28]. It hosts leading financial institutions, corporate headquarters, and premier research centers. Its dynamic economy contributes significantly to factory employment, tax revenues (income, entertainment, customs, and excise), foreign trade, and corporate taxes, making it a major force in the national economy [29].

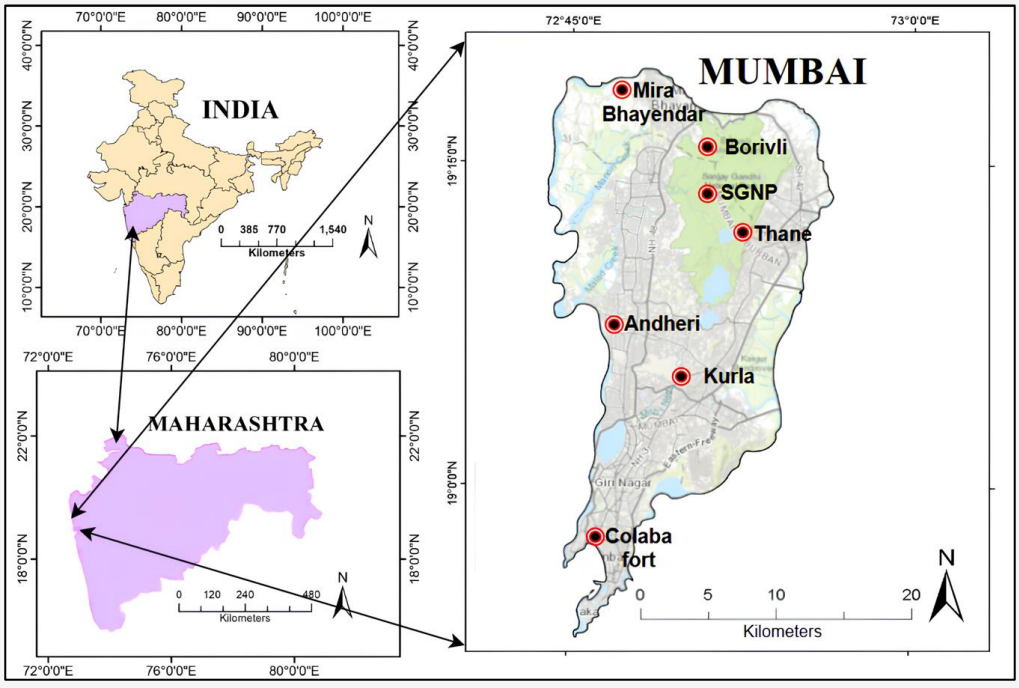

Figure 1: Geographical location of Maharashtra and Mumbai

This economic magnetism continues to attract migrants from across India raising critical concerns: Is there enough space to accommodate the growing influx? Will ecologically rich land be converted into residential areas to meet the demand? This research aims to examine the current rate of land use change and forecast upcoming patterns of urban development in Mumbai.

2.1 Data and Material

The dataset utilized in this research consists of satellite imagery, which is utilized to analyze spatiotemporal transformations in LULC. Satellite imagery from Landsat 5, Landsat 7, and Landsat 8 was obtained for the years 1990, 2000, 2010, and 2020 via the USGS Earth Explorer platform. Priority was given to images with minimal cloud cover, haze, and dust. To ensure optimal clarity and reduce atmospheric disturbances, only data captured during the summer months were utilized. A summary of the dataset is provided in Table 1.

3. Methodology

This study is organized into three key phases. The first phase focuses on developing a LULC classification system. The second phase involves analysing the classified data, and the third phase forecasts future LULC changes. Additionally, the study evaluates spatial patterns of urban expansion for the periods 2000–2030 and 2040. Figure 2 presents a block diagram outlining the LULC classification and prediction methodology, which consists of three main components: (3.1) Data Preprocessing, (3.2) Classification and Accuracy Assessment, and (3.3) Forecasting and Validation of Future LULC Trends. Each phase is detailed in the following sections.

3.1 Data Preprocessing

Satellite images sourced from the USGS Earth Explorer undergo preprocessing to improve their usability. Some key preprocessing techniques include geometric correction, radiometric correction, and atmospheric correction. Geometric correction is carried out using a toposheet published by the Survey of India to ensure accurate spatial alignment [30]. The images are then reprojected using the WGS 84 datum and the Universal Transverse Mercator (UTM) coordinate system in zone 43N. This correction process is essential for reducing various image distortions [31]. Once the image has been successfully preprocessed, the subsequent step involves classifying the image. A training dataset is then generated using field data points and Google Earth Engine [32]. The land is divided into five distinct categories: water, urban areas, vegetation, aquatic vegetation, and barren land. Table 2 provides further details on these categories. To prepare the training data, 200 sample points were collected from each category. A supervised classification approach using the SVM algorithm was implemented with ArcGIS 10.6 software.

Table 1: Dataset Summary

Sr. No.	Data used	Acquisition Dates	Purpose	Spatial Resolution	Source
1	Landsat 5TM	22/03/1990	To retrieve LULC Map	30 m	https://earthexplorer.usgs.gov/
2	Landsat 5TM	26/10/2010	To retrieve LULC Map	30 m	https://earthexplorer.usgs.gov/
3	Landsat 7 ETM+	25/02/2000	To retrieve LULC Map	30 m	https://earthexplorer.usgs.gov/
4	Landsat 8 OLI/TIRS)	24/2/2020	To retrieve LULC Map	30 m	https://earthexplorer.usgs.gov/
5	Shuttle Radar Topography	03/2014	For Exploratory Variables	30 m	https://www.earthdata.nasa.gov/data/instruments/srtm
6	Google Earth Maps	1990-2020	For Accuracy Assessment		https://earth.google.com/web/

Table 2: The details of the classes categories

Sr. No.	LULC Classes	Class Categories
1	Water	Ocean, river, lake, dam
2	Urban	Industrial zones, transportation hubs, residential sectors, and various other developed areas
3	Vegetation	Forest, Cropland, Pasture
4	Water Vegetation	Mangroves, creek, and salt land
5	Barren	Open land

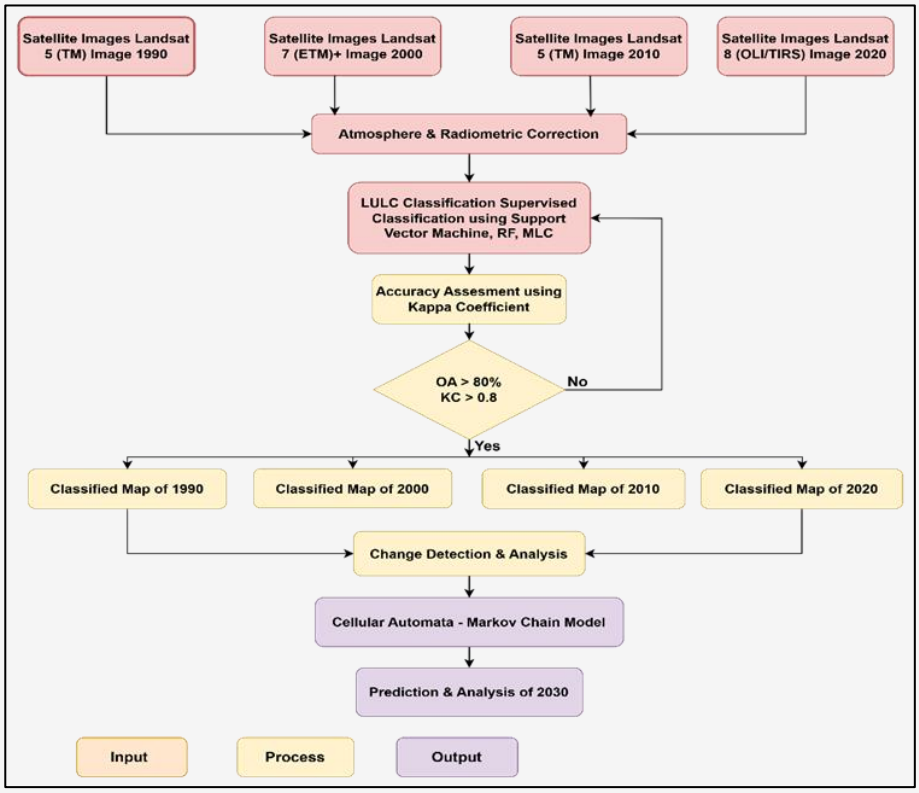

Figure 2: Block diagram of LULC prediction

3.2 Land Cover Classification and Validation

Ensuring the accuracy a classified image is an essential part of this procedure. If the accuracy reaches the necessary threshold, it confirms the correct classification of the satellite image. This evaluation involves aligning the classified image with specific reference data or authentic ground truth information. Reference data can be acquired from multiple sources, such as field surveys, high-resolution satellite imagery, topographic maps, and previously categorized images.

As part of this analysis, an accuracy assessment was conducted to assess the reliability of the results, utilizing 100 independent reference points that were distinct from the training dataset. These reference points were obtained through the Google Earth Engine and field surveys conducted by the Greater Mumbai Corporation Authority [33]. The gathered reference points were subsequently analyzed alongside the classified maps, resulting in the formation of an error matrix (also known as a confusion matrix). This matrix evaluates accuracy by considering user accuracy, producer accuracy, and overall accuracy. User accuracy reflects commission error, indicating the likelihood that the classified map correctly represents real-world features.

Producer accuracy assesses omission error, demonstrating the extent to which actual ground features are accurately mapped. Additionally, the nonparametric kappa coefficient (k) was calculated to further evaluate classification accuracy. This coefficient indicates the proportion of accurately classified pixels relative to the total number of pixels analyzed [34]. If $k \geq 0.8$, it suggests excellent concurrence, while if it falls between 0.4 and 0.6, it indicates satisfactory concurrence, and if $k \leq 0.4$, it denotes inadequate concurrence [35]. The subsequent Equation 1 was employed to compute the kappa coefficient index [34].

$$K = \frac{n\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r}\left(X_i + X_{+i}\right)}{n^2 - \sum_{i=1}^{r}\left(X_i + X_{+i}\right)}$$

Equation 1

In this context, n denotes the total number of observations, while r refers to the total number of rows within the matrix. The term Xii represents the number of observations at the intersection of row I and column i, while $X+i$ indicates the sum of values in column i. Moreover, $Xi+$ signifies the total of the values in row i.

3.3 Forecasting and Validating Future LULC Trends

The LULC transitions pattern observed between 2000–2010 and 2010–2020 serves as a basis for forecasting future LULC changes in 2030. The Land Transition Model (LTM) was developed to obtained a projected LULC map for 2030 and 2040. This model integrates the CA-MC approach to simulate future LULC dynamics. Several key exploratory factors, such as slope, elevation, proximity to transportation networks, and distance from settlements, were taken into account, as they significantly influence LULC transformations. Figure 3 presents the exploratory variables utilized for future prediction the LULC distribution in 2030 and 2040. Markov chain analysis is widely applied in LULC [36] and [37].

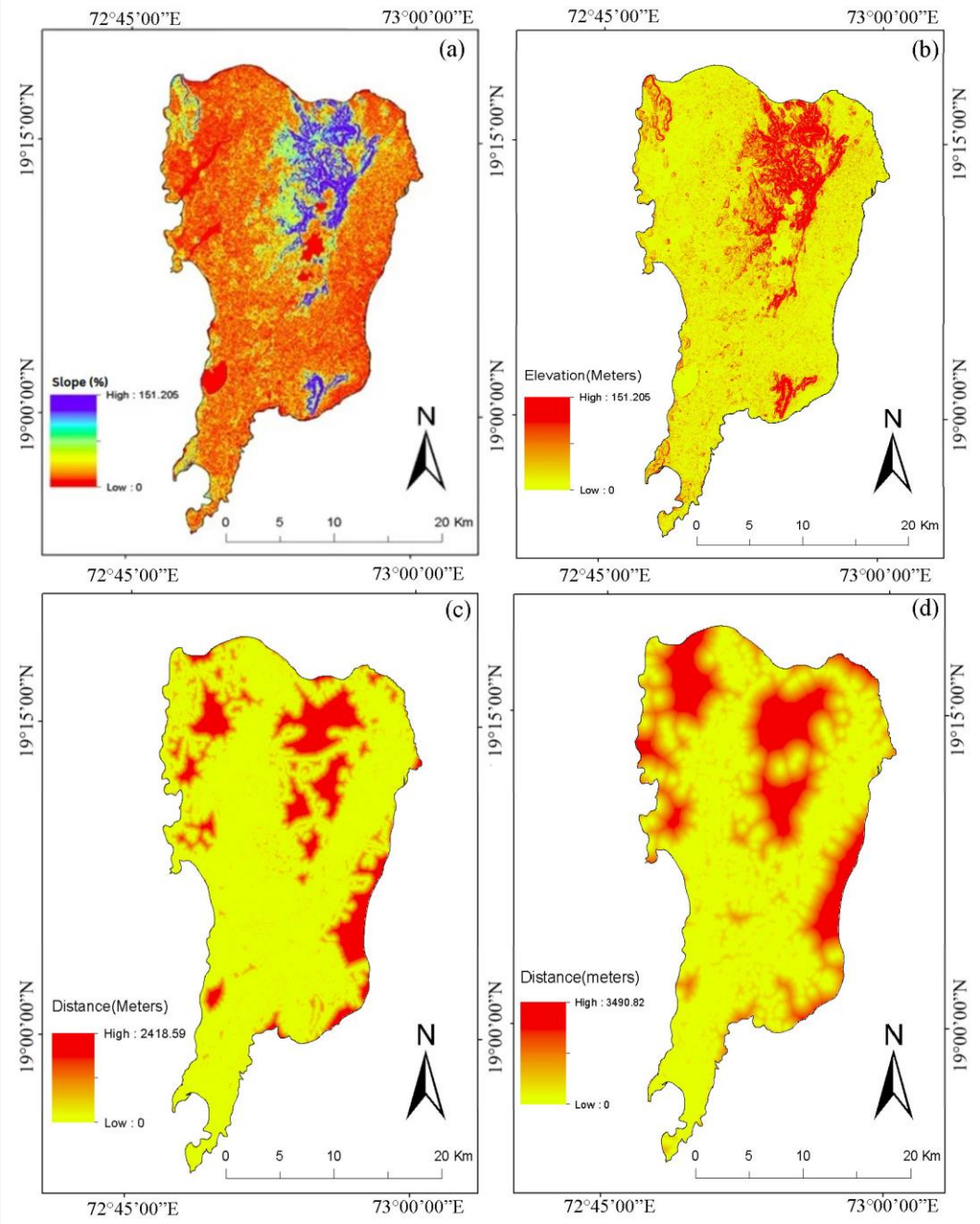

Figure 3: Explanatory variable: (a) Slope, (b) Elevation, (c) Transportation, (d) Settlement

This LULC projection studies, enabling the estimation of transition probabilities from historical imagery to forecast future land use and cover change (LUCC) trends based on specific time intervals [38]. The core operation of the Markov chain involves the creation of both a transfer matrix and a probability transition matrix, which are used to predict future trends in LULC changes. The Markov chain model is represented by a collection of states, denoted as *s = {s0, s1, s2, ..., sn}*. In the context of this study, the present state, labeled as *St*, transitions to a future state *Sj* according to the associated transition probabilities *Pij*. Therefore, the system's future state *St+1* is determined by its preceding state St, as described in Equations 2 and 3 [39]:

$$P_{ij} = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}$$

Equation 2

$$0 < P_{ij} < 1 \text{ and } \sum_{i=1}^{n} P_{oj} = 1,\ i, j = 1, 2, 3, \dots, n$$

Equation 3

In the Markov model, *P* refers to the probability matrix, with P_{ij} indicating the probability of moving from the current state *i* to the next state *j* in the following time period. The variable S represents the LULC condition, and *t* and *t+1* correspond to different time intervals, as shown in Equation (3) [40]. CA is a dynamic, bottom-up modeling technique that integrates both spatial and temporal aspects, providing a structured framework for simulation. This method effectively captures the complexities of space and time, despite their discrete nature. The CA model plays an essential role in LULC change studies by enabling the simulation of spatial and dynamic processes, making it a valuable tool for LULC prediction [39].

The fundamental components of a CA model include cells, cell space, neighboring cells, transition rules, and time. The neighborhood configuration is defined using a particular filter in which cells closer to the center are assigned higher weights than those farther away [41]:

$$S_{t+1} = f(S_t, N)$$

Equation 4

In Equation 4, *S* denotes the set of states for the finite cells; *t* and *t+1* refer to the earlier and later years, respectively; *N* represents the neighboring cells, and *f* signifies the conversion rule for the local space.

4. Results and Discussion

4.1 LULC Classification and Evaluation of its Accuracy

Supervised learning algorithms, a subset of machine learning techniques, were utilized for LULC classification. Among various available methods, including SVM, RF, and MLC, the SVM algorithm was chosen for image classification. Figure 4 illustrates the classified images related to the years 1990, 2000, 2010, and 2020. After classification, an accuracy assessment was conducted to evaluate the reliability of the results. This assessment employed statistical measures such as the Kappa Coefficient, which quantifies the agreement between two sets of labels. Additional accuracy indicators are summarized in Table 3. The findings revealed overall accuracies for the years 1990, 2000, 2010, and 2020 stood at 91.82%, 88.73%, 94.07%, and 95.63%, respectively, surpassing the USGS benchmark of 85%. In comparison, the RF algorithm yielded an accuracy of 86.24%, while the MLC Classification method achieved 81.36%, both of which fell short of the performance demonstrated by the SVM model. This further justified the selection of SVM as the most suitable method for LULC classification in this study.

Table 3: Evaluation of the accuracy in LULC classification

Year	Accuracy (%)	Water	Urban	Vegetation	Water Vegetation	Bareen	OA (%)			Kappa Coefficient		
							SVM	RF	MLC	SVM	RF	MLC
1990	UA	89.66	93.92	99.66	98.95	94.09	95.13	89.45	85.24	0.93	0.86	0.82
	PA	98.33	87.66	99.00	95.00	95.66						
2000	UA	95.68	97.59	87.71	83.10	82.97	88.73	84.54	84.12	0.86	0.83	0.89
	PA	88.67	81.00	88.00	98.33	87.67						
2010	UA	93.13	91.42	95.17	99.32	91.41	94.07	86.78	83.82	0.93	0.81	0.86
	PA	99.33	92.33	92.00	98.00	88.67						
2020	UA	83.33	97.91	88.97	93.21	92.82	95.63	84.36	81.20	0.92	0.82	0.80
	PA	94.59	97.37	92.25	93.97	90.32						

**UA is user accuracy, PA is producer accuracy, OA is overall accuracy

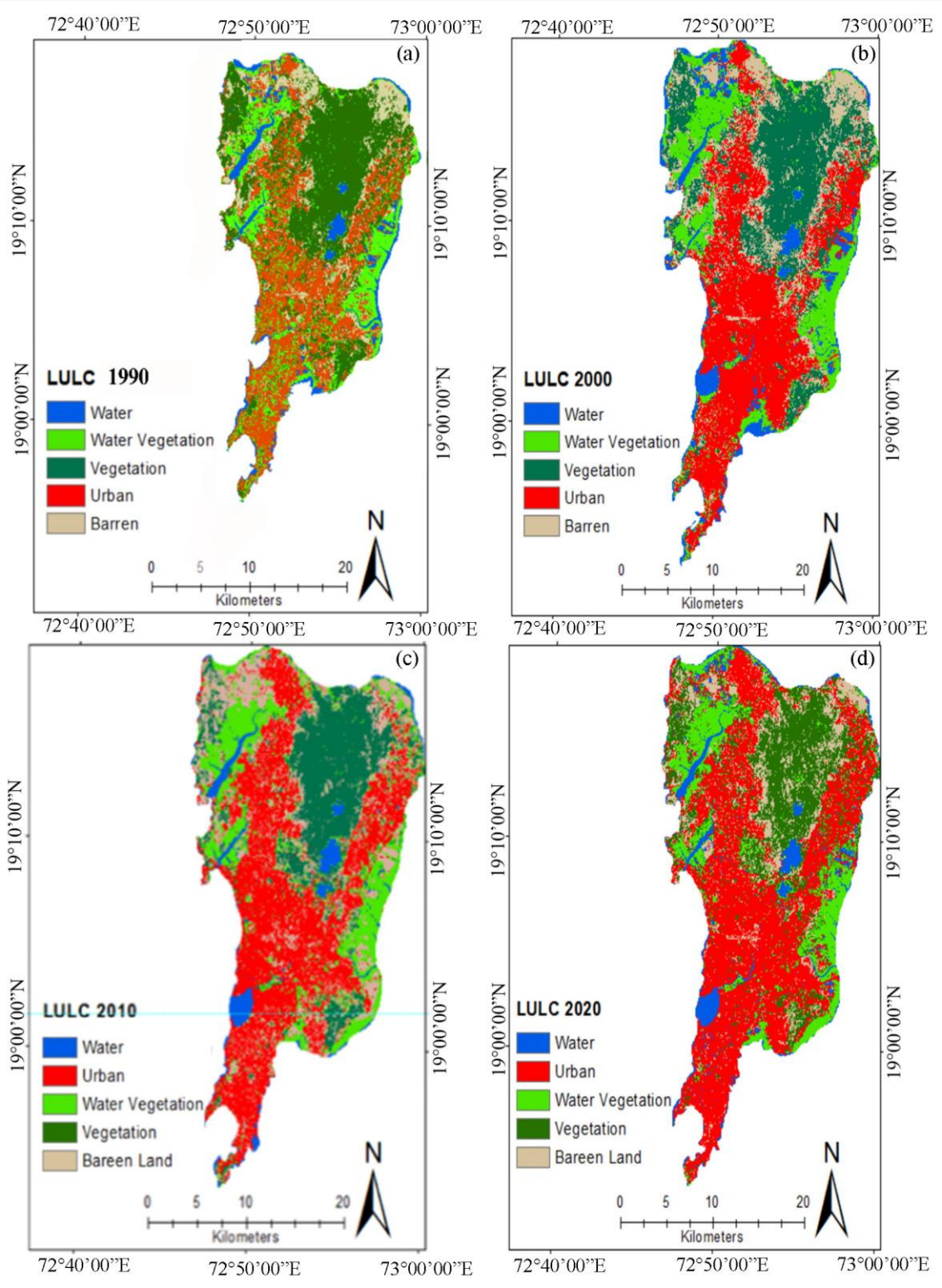

Figure 4: Classified images of LULC: (a)1990, (b) 2000, (c) 2010, (d) 2020

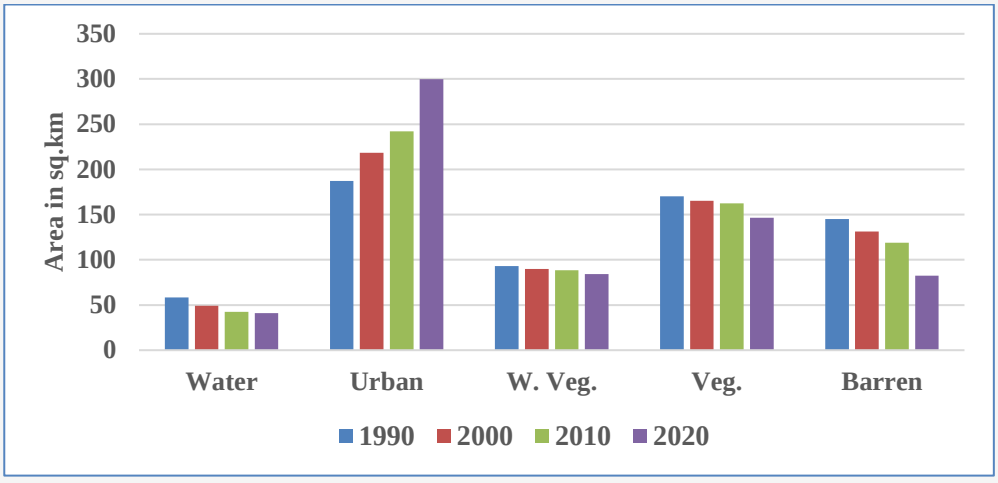

Figure 5: LULC variation from 1990-2020

Table 4: Area of each LULC class in percentage

Year	1990		2000		2010		2020	
LULC Classes	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)
Water	58.27	8.93	49.28	7.53	42.24	6.45	41.00	6.26
Urban	187.23	28.62	218.41	33.38	242.22	37.02	299.80	45.83
W. Veg.	92.94	14.20	89.94	13.74	88.55	13.53	84.30	12.88
Veg.	170.34	26.03	165.27	25.26	162.42	24.82	146.66	22.42
Barren	145.21	22.19	131.28	20.06	118.77	18.15	82.32	12.58

Table 4 and Figure 5 illustrate the spatial changes in LULC across the years 1990, 2000, 2010, and 2020, emphasizing the decadal shifts within the MMR. The findings indicate notable land use transformations over the last 30 years within the 654.20 sq. km study area. Between 1990 and 2000, urbanization began in South Mumbai, particularly in areas such as Colaba, Fort, and Girgaon. During this period, many industries and textile mills, which were once the backbone of Mumbai's economy, began to decline. These areas underwent significant redevelopment, with mill lands being converted into commercial and residential infrastructure. As a result, people began moving into these redeveloped neighborhoods, drawn by improved amenities and job opportunities. From 2000 to 2010, Mumbai further strengthened its position as a major economic hub, this led to significant migration from various regions of India. The growing population placed increased burden on the city's infrastructure and resources. Affordable housing initiatives and improved transportation infrastructure, such as rail and road connectivity, encouraged both residential and commercial expansion into new areas. During the decade from 2010 to 2020, Mumbai experienced its peak phase of urbanization. Major infrastructure projects such as Mumbai Metro Phase 1, the Eastern Freeway, the Bandra-Worli Sea Link, and the Monorail played a key role in facilitating urban expansion and renewal. However, the city also faced growing challenges, including overcrowding.

As a result of MMR exponential growth there is 37.26% increase in urban land from 2000 to 2020, marking a substantial transformation over two decades. In contrast, other land categories consistently declined. Barren land experienced the sharpest decrease, followed by vegetation and water vegetation (including coastal and mangrove areas), which showed a continual reduction from 1990 to 2010, with further decline in 2020. Water bodies also witnessed a slight decline, decreasing from 7.53% in 2000 to 6.26% in 2020. The results strongly suggest rapid urban expansion, primarily replacing natural and undeveloped lands. The most notable transformation was the shift from barren terrain to urban zones, highlighting the environmental costs linked to fast-paced urbanization in the MMR.

4.2 Evaluation of Alteration Detection

Table 5 and Figure 6 provides an overview of the changes in terms of percentage and area over three-time intervals: 2000-2010, 2010-2020, and 2000-2020.

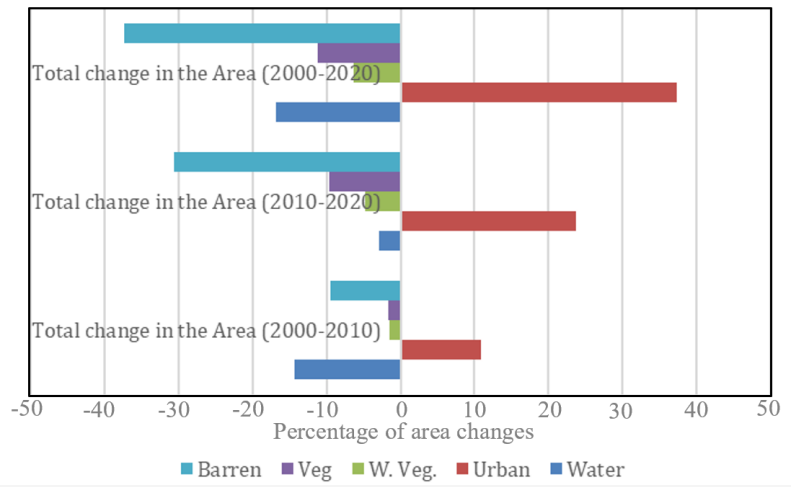

Figure 6: Decadal LULC change detection

Table 5: LULC change detection for the year 2000-2010, 2010-2020, and 2000-2020

LULC Classes	2000-2010		2010-2020		2000-2020	
	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)
Water	-7.04	-14.28%	-1.24	-2.93	-8.28	-16.80
Urban	+23.81	+10.90%	+57.58	+23.77	+81.39	+37.26
Water Vegetation	-1.39	-1.54%	-4.25	-4.79	-5.64	-6.27
Vegetation	-2.85	-1.72%	-15.76	-9.70	-18.61	-11.26
Barren	-12.51	-9.52%	-36.45	-30.68	-48.96	-37.29

During 2000-2010, urban land increased by 10.90%, while barren land, vegetation, water vegetation, and water bodies decreased by 9.52%, 1.72%, 1.54%, and 14.28%, respectively. From 2010–2020, urban land saw a further increase of 23.77%, whereas barren land, vegetation, water vegetation, and water bodies declined by 30.68%, 9.70%, 4.79%, and 2.93%, respectively. Over the two decades from 2000–2020, urban land expanded by 37.26%, while barren land decreased by 37.29%, vegetation by 11.26%, water vegetation by 6.27%, and water bodies by 16.80%.

4.3 Prediction and Validation of 2020

The CA-Markov model implemented using Python 3 within the Jupiter Notebook environment. For model validation, classified images from the years 1990, 2000, and 2010 were utilized as input data, while the 2020 land cover was simulated. The predicted 2020 image was then evaluated against the actual classified 2020 image using the chi-square test. Table 6 presents the actual and simulated 2020 results, demonstrating that the simulated and actual distributions are statistically significant. Accuracy assessment was also conducted using the Kappa coefficient. For the predicted LULC map of 2020, accuracy assessment using standard metrics showed an overall accuracy (OA) of 92.26% and a Kappa coefficient of 0.91, demonstrating a high level of agreement and validating the reliability of the prediction model. The Urban and Barren land cover classes exhibit relatively high classification errors, both exceeding 10%. Notably, the Barren class is significantly overestimated. In contrast, the Water, Wetland Vegetation (W. Veg.), and Vegetation (Veg.) classes fall within or near the acceptable error threshold of 10%, with W. Veg. showing particularly minimal deviation. The model underestimated Water coverage by approximately 5.26%, which remains within an acceptable margin for LULC simulation. This relatively small error suggests that the model is well-calibrated for identifying stable water bodies, reflecting a reasonably accurate simulation overall. Urban Clas is underestimation may be due to rapid urban sprawl that is not fully captured by the current model. To improve accuracy, incorporating more recent urban transition rules or dynamic expansion drivers such as proximity to major roads, economic hubs, or population growth centers could enhance predictive performance. Barren overestimation might stem from confusion between temporary features such as dry agricultural fields or recently harvested land and truly barren areas. Integrating better temporal indicators, such as seasonal vegetation indices or multi-temporal imagery, may help distinguish between transient land states and permanent barren land.

Table 6: Comparison between actual and simulated areas of 2020

LULC Class	Percentage of simulated area for 2020; E	Percentage of actual area for 2020; O	Error [O-E]×100/O (%)
Water	5.94	6.27	5.26
Urban	39.27	45.87	14.38
Water Vegetation	13.11	12.88	-1.78
Vegetation	24.31	22.42	-8.42
Barren	17.37	12.56	-38.29
Total	100	100	

Table 7: The percentage of LULC projected for the year 2030

LULC Classes	Area (%)				
	2000	2010	2020	2030	2040
Water	7.53	6.45	6.26	5.56	5.99
Urban	33.38	37.02	45.83	47.25	55.84
Water Vegetation	13.74	13.53	12.88	12.73	10.30
Vegetation	25.26	24.82	22.42	20.19	16.51
Barren	20.06	18.15	12.58	13.64	11.37

Table 8: Analysis of LULC changes from 2000 to 2040

LULC Class	Total change (%)				
	2000 - 2020	2000 - 2030	2000 - 2040	2020 - 2030	2020 - 2040
Water	-16.80	-26.16	-20.45	-11.18	-4.31
Urban	+37.26	+41.55	+67.28	3.09	+21.84
Water Vegetation	-6.27	-7.35	-24.81	-1.16	-19.79
Vegetation	-11.26	-20.07	-34.63	-9.94	-26.36
Barren	-37.29	-32.00	-43.32	-8.42	-9.61

4.4 Prediction of LULC 2030 and 2040

The CA-MC Model was developed in Python to predict future LULC changes for 2030. The model processes input data from three images taken at 10-year intervals (2000, 2010, and 2020) to forecast changes. Results, in Table 7 and Figure 7. The predicted results are illustrated in Figure 8 and Figure 9, indicate that by 2030, urban land is expected to increase by 47.25%. Meanwhile, other land categories are expected to change as follows: barren land (13.64%), vegetation (20.19%), water vegetation (12.73%), and water bodies (5.56%).

4.5 LULC Change Detection and Analysis

The spatiotemporal data analysis for 2000, 2010, 2020, and the predicted 2030 and 2040 is summarized in Table 8. The analysis covers three intervals: 2000–2020 (last 20 years), 2000–2030 (last 30 years), and 2020-2030 (recent decade), 2000-2040(40 years). Observations indicate significant decadal land cover changes, with barren land being predominantly converted into urban areas. This transformation is especially evident in regions such as South West Byculla, Andheri city center, Chandivali, South East Ghatkopar, Vikhroli, North East Thane, Manpada, North West Charcop, and Mira Road all of which have become almost entirely urbanized. According to the prediction for the year 2030, there is a drastic decline in vegetation cover. This degradation is primarily attributed to infrastructure developments, including the Mumbai Metro project and other expansion plans led by the MMRDA [28]. The only significant patch of vegetation remaining in Mumbai is the Sanjay Gandhi National Park (SGNP), which is a designated reserve zone. However, the boundaries of SGNP are under constant threat due to urban expansion and may be encroached upon for infrastructure projects such as roads, bridges, and metro lines. By 2040, Mumbai is projected to be almost entirely urbanized. As a result, new residents and development pressures are expected to shift towards suburban areas such as Navi Mumbai, Greater Thane, Panvel, and Vasai. Additionally, the coastal regions of Mumbai are at risk due to increasing urban pressure and developmental activities.

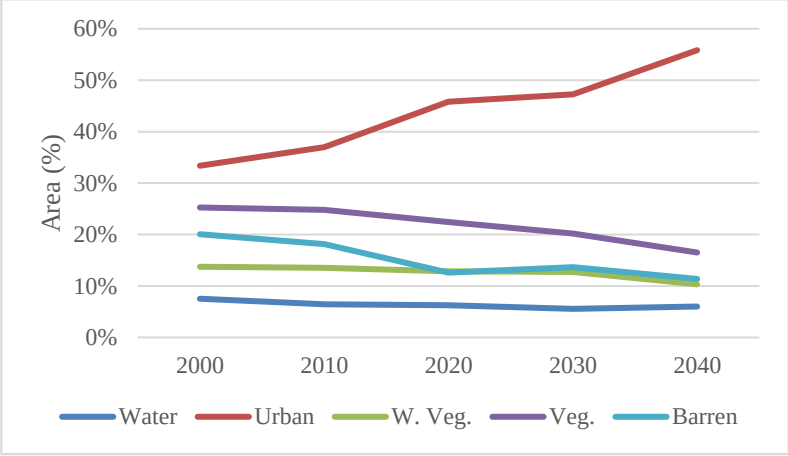

Figure 7: Graphical representation of 2000 to 2040

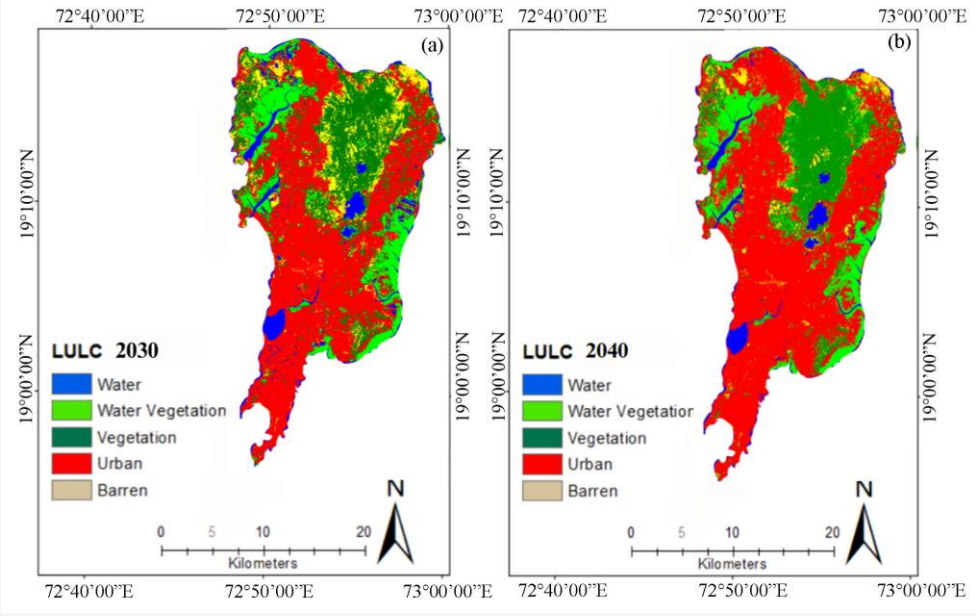

Figure 8: Predicted LULC using Python based CA-Markov: (a) 2030 and (b) 2040

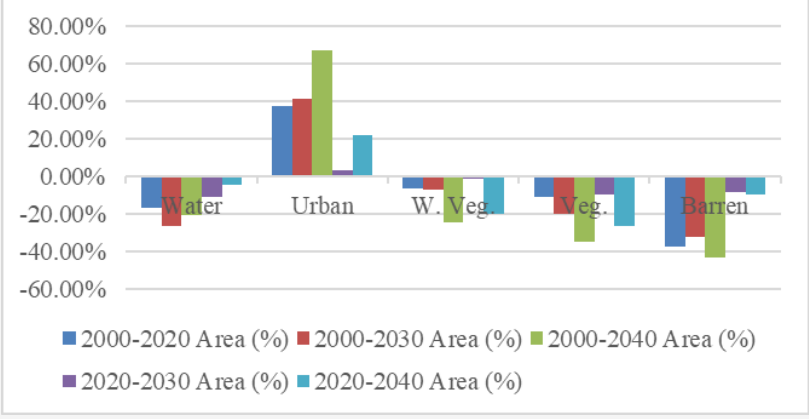

Figure 9: Decadal changes in LULC

These findings highlight significant land-use changes in the Mumbai region. As of 2020, Mumbai, being a major economic hub, is already highly urbanized. Due to the scarcity of available space, opportunities for further urban expansion are limited. Alarmingly, vegetation areas, particularly those surrounding Sanjay Gandhi National Park, are being converted into barren land. Immediate measures are crucial to protect the park and its surroundings, as further degradation may result in barren land is transformed into urban land in the near future.

5. Conclusion

This research examines historical and projected trends in LULC changes, concentrating on transitions that have occurred over the past 20 years and forecasting shifts for the next two decades. Historical analysis utilized Landsat satellite images from the years 2000, 2010, and 2020, while future scenarios for 2030 and 2040 were predicted based on this data.

Among various methods tested, the SVM supervised classification technique proved to be the most effective for the dataset. Key driving factors including slope, elevation, proximity to roads, and distance from settlements were incorporated into the classification process, as these play a significant role in influencing urban growth. To simulate future LULC patterns, the CA-MC model was employed. The results were assessed at ten-year intervals, and the accuracy was evaluated by comparing the simulated 2020 LULC map with actual observations from the same year, showing minimal differences. Forecasted maps for 2030 and 2040 suggest ongoing land use transformations, marked by continued urban sprawl and expansion. Overall, the study underscores major LULC changes, predominantly driven by accelerated urbanization and industrial growth. These findings emphasize the urgent need for urban planners and policymakers in Mumbai to address the emerging challenges posed by rapid development.

Nevertheless, some limitations exist. The analysis was confined to Landsat imagery, and incorporating data from additional satellite sources could enhance classification accuracy. Moreover, the study considered only four influencing factors; integrating additional parameters such as industrial zones, commercial centers, or population density might offer a more detailed insight into LULC dynamics. Future research may also explore the use of advanced deep learning models to further refine LULC classification and prediction.

References

[1] United Nation, World, Department of Economics and Social Affairs, (2022). World Population Prospects 2022 Summary of Result, UN DESA/POP/2021/TR/NO. 3 [Online]. Available: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf. [Accessed: Mar. 6, 2024]

[2] Mazumdar, D., (1987). Chapter 28, Rural-Urban Migration in Developing Countries. *Handbook of Regional and Urban Economics*, Vol. 2, 1097-1128. https://doi.org/10.1016/S1574-0080(87)80014-7.

[3] Kavathekar, V., Tripathy, A. K., Chettri, S. K. and Bhanage, V., (2024), Assessment and Prediction of Urban Pollutants and its Influence on Human Health using Deep Learning Algorithm, *IEEE 9th International Conference for Convergence in Technology (I2CT)*, IEEE, 1–7. https://doi.org/10.1109/I2CT61223.2024.10543725.

[4] Liang, L., Wang, Z. and Li, J., (2019). The Effect of Urbanization on Environmental Pollution in Rapidly Developing Urban Agglomerations. *Journal of Clean Production*, Vol. 237. https://doi.org/10.1016/j.jclepro.2019.117649.

[5] Kavathekar, V., Tripathy, A. K., Chettri, S. K. and Bhanage, V., (2024). Evaluation of Land Use Land Cover Dynamics and Urban Heat Island Effects Over Mumbai Metropolitan Region, India. *International Journal of Environmental Science and Technology*, https://doi.org/10.1007/s13762-024-06266-4.

[6] Anthony, T., Shohan, A., Oludare, A., Alsulamy, S., Al Kafy, A. and Khedher, K. M., (2024). Spatial Analysis of Land Cover Changes for Detecting Environmental Degradation and Promoting Sustainability. *Kuwait Journal of Science*, Vol. 51(2). https://doi.org/10.1016/j.kjs.2024.100197.

[7] Matlhodi, B., Kenabatho, P. K., Parida, B. P. and Maphanyane, J. G., (2019). Evaluating Land Use and Land Cover Change in the Gaborone Dam Catchment, Botswana, from 1984-2015 Using GIS and Remote Sensing. *Sustainability*, Vol. 11(19). https://doi.org/10.3390/su11195174.

[8] Cihlar, J., (2000). Land Cover Mapping of Large Areas from Satellites: Status and Research Priorities, *International Journal of Remote Sensing*, Vol. 21(6-7), 1093-1114, https://doi,org/10.1080/014311600210092.

[9] Mortoja, Md. G. and Yigitcanlar, T., (2020). Local Drivers of Anthropogenic Climate Change: Quantifying the Impact through a Remote Sensing Approach in Brisbane. *Remote Sensing*, Vol. 12(14). https://doi.org/10.3390/rs12142270.

[10] Yuan, Q., Shen, H. and Li, T., (2020), Deep Learning in Environmental Remote Sensing: Achievements and Challenges. *Remote Sensing Environment*, Vol. 241. https://doi.org/10.1016/j.rse.2020.111716.

[11] Alshari, E. A. and Gawali, B. W., (2021). Development of Classification System for LULC Using Remote Sensing and GIS. *Global Transitions Proceedings*, Vol. 2(1), 8–17, https://doi.org/10.1016/j.gltp.2021.01.002.

[12] MohanRajan, S. N., Loganathan, A. and Manoharan, P., (2020). Survey on Land Use/Land Cover (LU/LC) Change Analysis in Remote Sensing and GIS Environment: Techniques and Challenges, *Environmental Science and Pollution Research*, Vol. 27(24), 29900–29926. https://doi.org/10.1007/s11356-020-09091-7.

[13] Deilmai, B. R., Bin Ahmad, B. and Zabihi, H., (2014), Comparison of Two Classification Methods (MLC And SVM) to Extract Land Use and Land Cover in Johor Malaysia. *IOP Conference Serve Earth Environment Science*, Vol. 20. https://doi.org/10.1088/1755-1315/20/1/012052.

[14] Thammaboribal, P., and TRIPATHI, N. (2024). Predicting Land Use and Land Cover Changes in Pathumthani, Thailand: A Comprehensive Analysis from 2013 to 2023 Using Landsat Satellite Imagery and CA-ANN Algorithm, with Projections for 2028 and 2038. *International Journal of Geoinformatics*, Vol. 20(5), 13–27. https://doi.org/10.52939/ijg.v20i5.3225.

[15] Gaur, S. and Singh, R., (2023), A Comprehensive Review on Land Use/Land Cover (LULC) Change Modeling for Urban Development: Current Status and Future Prospects. *Sustainability*, Vol. 15(2). https://doi.org/10.3390/su15020903.

[16] Somvanshi, S. S., Bhalla, O., Kunwar, P., Singh, M. and Singh, P., (2020). Monitoring Spatial LULC Changes and its Growth Prediction Based on Statistical Models and Earth Observation Datasets of Gautam Budh Nagar, Uttar Pradesh, India. *Environmental Development Sustainability*, Vol. 22(2), 1073–1091, https://doi.org/ 10.1007/s10668-018-0234-8.

[17] Kamaraj, M. and Rangarajan, S., (2022), Predicting the Future Land Use and Land Cover Changes for Bhavani Basin, Tamil Nadu, India, using QGIS MOLUSCE Plugin. *Environmental Science and Pollution Research*, Vol. 29(57), 86337–86348. https://doi.org/10.1007/s11356-021-17904-6.

[18] Ahmad, F., Goparaju, L. and Qayum, A., (2017). LULC Analysis of Urban Spaces Using Markov Chain Predictive Model at Ranchi in India. *Spatial Information Research*, Vol. 25(3), 351–359. https://doi.org/10.1007/s41324-017-0102-x.

[19] Mishra, V. N., Rai, P. K., Prasad, R., Punia, M. and Nistor, M. M., (2018). Prediction of Spatio-Temporal Land Use/Land Cover Dynamics in Rapidly Developing Varanasi District of Uttar Pradesh, India, Using Geospatial Approach: A Comparison of Hybrid Models. *Applied Geomatics*, Vol. 10(3), 257–276. https://doi.org/10.1007/s12518-018-0223-5.

[20] Vinayak, B., Lee, H. S. and Gedem, S., (2021), Prediction of Land Use and Land Cover Changes in Mumbai City, India, Using Remote Sensing Data and a Multilayer Perceptron Neural Network-Based Markov Chain Model. *Sustainability*, Vol. 13(2). https://doi.org/10.3390/su13020471.

[21] Vinayak, B., Lee, H. S., Gedam, S. and Latha, R., (2022), Impacts of Future Urbanization on Urban Microclimate and Thermal Comfort Over the Mumbai Metropolitan Region, India. *Sustainable Cities and Society*, Vol. 79. https://doi.org/10. 1016/j.scs.2022.103703.

[22] The World City Report 2020, (2022). *The Value of Sustainable Urbanization*. Available: https://unhabitat.org/sites/default/files/2020/10/wcr_2020_report.pdf. [Accessed: Ma. 22, 2024].

[23] Urban Agglomerations Census 2011, Cities Population One Lakh and Above (2011), [Online]. Available: https://www.census2011.co.in/urban agglomeration. [Accesed: Feb. 4, 2025].

[24] India Meteorological Department, Extremes of Temperature & Rainfall for Indian Stations (Up To 2012) Government of India Ministry of Earth Sciences India Meteorological Department Issued by Climate Services Division Office of The Additional Director General of Meteorology (Research) India Meteorological Department Pune-411005. [Online], Available: https://www.imdpune.gov.in/library/public/EXTREMES%20OF%20TEMPERATURE%20and%20RAINFALL%20upto%202012.pdf. [Accessed: Mar. 25, 2024].

[25] Central Pollution Control Board, (2024). [Online], Available: https://cpcb.nic.in/ [Accessed: Dec. 25,2024].

[26] Ramachandra, T. V., Bharath, H. A. and Sowmyashree, M. V., (2021). Urban Footprint of Mumbai - The Commercial Capital of India. *Journal of Urban and Regional Analysis*, Vol. 6(1). https://doi.org/ 10.37043/Jura.2014.6.1.5.

[27] Chatterjee, A., Mondal, M., Dhar, J. and Nazir, Md. I., (2023), *Positioning of Mumbai in the Indo-Pacific Megacity System: Economic and Transport Opportunities*. 399–425. https://doi.org/10.1007/978-981-99-6218-17.

[28] Regional Plan Mumbai Metropolitan Region Development Authority, (2024). [Online], Available: https://mmrda.maharashtra.gov.in/planning/regional-plan/final-rp-for-mmr. [Accessed: Feb. 27, 2024]

[29] Indian Institute for Human Settlements. (2011). Urban India 2011: Evidence. [Online]. Available: https://iihs.co.in/wp-content/uploads/2013/12/IUC-Book.pdf. [Accessed: Feb. 27, 2024].

[30] Survey of India, (2024). [Online], Available: https://onlinemaps.surveyofindia.gov.in/. [Accessed: Mar. 20, 2024].

[31] Michael Wulder, S. E. F., (2003). *Remote Sensing of Forest Environments*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-0306-4.

[32] Aji, A., Husna, V., and Purnama, S. (2024). Multi-Temporal Data for Land Use Change Analysis Using a Machine Learning Approach (Google Earth Engine). *International Journal of Geoinformatics*, Vol. 20(4), 19–28. https://doi.org/10.52939/ijg.v20i4.3145

[33] Mendiratta, P. and Gedam, S., (2018). Assessment of Urban Growth Dynamics in Mumbai Metropolitan Region, India Using Object-Based Image Analysis for Medium-Resolution Data. *Applied Geography*, Vol. 98, 110–120. https://doi.org/10.1016/j.apgeog.2018.05.017.

[34] Cohen, J., (1960). A Coefficient of Agreement for Nominal Scales. *Educational Psychological Measurements*, Vol. 20(1), 37–46. https://doi.org/10.1177/001316446002000104.

[35] Foody, G. M., (2004). Thematic Map Comparison. *Photogrammetric Engineering Remote Sensing*, Vol. 70(5), 627-633. https://doi.org/10.14358/PERS.70.5.627.

[36] Faichia, C., Tong, Z. and Zang, J., (2020). Using RS Data-Based CA–Markov Model for Dynamic Simulation of Historical and Future LUCC in Vientiane, Laos. *Sustainability*, Vol. 12(20). https://doi.org/10.3390/su12208410.

[37] Hsu, L. C., (2003). Applying the Grey Prediction Model to the Global Integrated Circuit Industry. *Technological Forecasting and Social Change*, Vol. 70(6), 563–574. https://doi.org/10.1016/S0040-1625(02)00195-6.

[38] Murugesan, S., Zhang, J. and Vittal, V., (2012). Finite State Markov Chain Model for Wind Generation Forecast: A Data-Driven Spatiotemporal Approach. *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*, 1–8. https://doi.org/10.1109/ISGT.2012.6175764.

[39] Wang, S. W., Gebru, B. M., Lamchin, M., Kayastha, R. B. and Lee, W. K., (2020). Land Use and Land Cover Change Detection and Prediction in the Kathmandu District of Nepal Using Remote Sensing and GIS. *Sustainability*, Vol. 12(9). https://doi.org/10.3390/su12093925

[40] Veldkamp, A. and Lambin, E. F., (2001). Predicting Land-Use Change. *Agricultural Ecosystem and Environment*, Vol. 85(1–3), 1–6. https://doi.org/10.1016/S0167-8809(01)00199-2.

[41] Subedi, P., Subedi, K. and Thapa, B., (2013). Application of a Hybrid Cellular Automaton – Markov (CA-Markov) Model in Land-Use Change Prediction: A Case Study of Saddle Creek Drainage Basin, Florida. *Applied Ecology and Environmental Sciences*, Vol. 1(6), 126–132. https://doi.org10.12691/aees-1-6-5.