The Indian Geographical Journal
Volume 96 (2) December - 2021, pp 49-64
ISSN 0019-4824

SIMULATION OF URBAN SPRAWL USING GEO-SPATIAL ARTIFICIAL NEURAL NETWORKS AND CA-MARKOV CHAIN MODELS

Harshali Patil[1], Kanika Pillai[1], Sivakumar V[2*], Shivaji G Patil[3]

[1]COEP, Pune University, Pune, India
[2]Centre for Development of Advanced Computing (C-DAC), Pune, India
[3]Maharashtra Industrial Development Corporation, Pune, India
*Corresponding author: vsivakumar@cdac.in

Abstract

Land use/land cover (LULC) changes caused by human interferences have extensive consequences at the local and global levels. Rapid urbanization is mainly driven by many factors such as increasing migration to urban areas and rapidly increasing population, especially in developing countries like India. This study evaluates the changes in the urban area occurred due to human interferences, in Pune using multispectral satellite imagery. The Machine Learning (ML) method, which is available with Orfeo Toolbox open-source tool, has been used for image classification for the years 1990, 2000, 2010, and 2020. LANDSAT series datasets were used and classified into four classes, viz. urban cover, vegetation, water, and unclassified (vacant, barren land, etc.). The classified outputs were assessed for the accuracy using the open-source Semi-Automatic Classification (SCP) tool. In this study, urban expansion was analysed using certain proximity factors that cause for the growth pattern of Pune city. A hybrid simulation model like Logistic Regression-Markov-Chain and Artificial Neural Network (ANN) were used to predict the future urban sprawl for the years 2030, 2040, and 2050. The validation of output is carried out using actual and simulated outputs. The prediction model results show that 69.63% accuracy with ANN algorithm and 55.25% accuracy with Logistic Regression-Markov-Chain algorithm. The predictions show that the urban area is expected to grow manifolds to 735.52 sq. km in 2050 from 56.943 sq. km in 1990. The result indicates that the integration of remote sensing, GIS, and Artificial Intelligent models help the urban planners for appropriate planning and decision-making processes with most accurate information.

Keywords: GIS, Remote Sensing, Land Use / Land Cover, Logistic Regression-Markov-Chain, Artificial Neural Network, Urban Growth

Introduction

The term "land cover" refers to the man-made geophysical properties of the earth's surface, including the distribution of plants, structures, and other physical features [1]. Globally, a significant quantity of agricultural and forested land is being

converted into urban space to accommodate population growth, which immediately affects the composition of the land [2, 3]. This change in land use and land cover in an urban area is a gradual process. With every passing year, the population of a city increases; thus, increasing the demand for infrastructure and services, which initiates the city to expand to satisfy the increasing demand of growing population therein. Different factors like physical, human, and environmental aspects influence the expansion of an urban area. For sustainable development, decision-makers need systems to monitor and observe these change in land cover and accurately predict the future change in land cover [4]. Understanding the change in land cover and urban sprawl of a city would help the decision-makers in making more informed decisions for the future population of the city.

In recent years, Indian metropolises have been witnessing an increasing rate of urbanization. Agricultural or barren land in the peripheral areas of the city is being converted to urban land for the development of infrastructure and facilities. The traditional methods of physically mapping the existing situation are very time-consuming and expensive [5]. The credibility of such data has also been questioned and might affect prediction results [6]. Therefore, the use of remote sensing in the urban planning process along with Geo-Artificial Intelligence or Geospatial-AI (Geo-AI) would help to map such data with credibility and also to generate a more accurate analysis of the existing situation in a cost-effective and time-saving manner. The use of such techniques is slowly being adopted in urban planning procedures.

Artificial Intelligence (AI) is the ability of systems to perform a task that requires human intelligence. Geo-AI is a crucial component of spatial analysis in Geographical Information System (GIS). Geo-AI methods can provide accurate land cover classification reports and also predict the situation in the coming decades by training past and present data sets. With the help of advanced technology, multiple algorithms have evolved in Geo-AI which helps to get precise results and output whole predicting a change in land cover. Artificial Neural Networks (ANN) and Cellular Automata (CA) Markov Chain Models are two widely used models in monitoring and predicting land use change [7]. These methods have been used in order to accurately predict the land cover change in Pune city for the coming decades. The results of this study can be used by the civic bodies and decision makers to facilitate better decisions for the future of the city. In this paper, urban agglomeration in and around Pune City has been studied to predict the future urban cover using ANN and CA Markov Models. To predict the stages of urban growth in Pune over the next few decades, remotely sensed satellite data (LANDSAT series data) from the years 1990, 2000, 2010, and 2020 were employed.

Study Area

Pune is a metropolis in Maharashtra, India. It lies at an altitude of 560m above Mean Sea Level (MSL) on the Deccan plateau. At the centre of the city, the rivers Mula and Mutha converge. It is the eighth-most populated city in India with an estimated population of about 5.05 million as of 2019 (Areas under the jurisdiction of Pune Municipal Corporation (PMC) and Pimpri Chinchwad Municipal Corporation (PCMC) are included). Three

Cantonment Boards, viz. Pune Cantonment Board, Khadki Cantonment Board, and Dehuroad Cantonment Board, are in existence presently in and around Pune city. In the past three decades, the city has become one of the most leading IT centres in India. Due to the presence of many well-known educational institutions, it is also known as the "Oxford of the East". Additionally, the growth of satellite industrial cities like Pimpri-Chinchwad and Chakan has also contributed to an increase in urban cover around Pune City. Similarly, all the incoming highways and bypass roads from other major cities within the state to Pune city are densely populated on both sides due to urbanization process and these exhibit the signs of tremendous rate of development, which results in drastic changes in land cover on both sides of these highways and bypass roads. The population of the city is expected to increase as the city continues to be one of the most liveable cities in the country. This paper primarily examines the study area that sits between 18d34'31.6147"N to 18d83'13.6237"N latitude and 73d59'17.9289"E to 74d08'81.2515"E longitude, which covers an approximate total area of 2478 sq.km of Pune City (Figure 1).

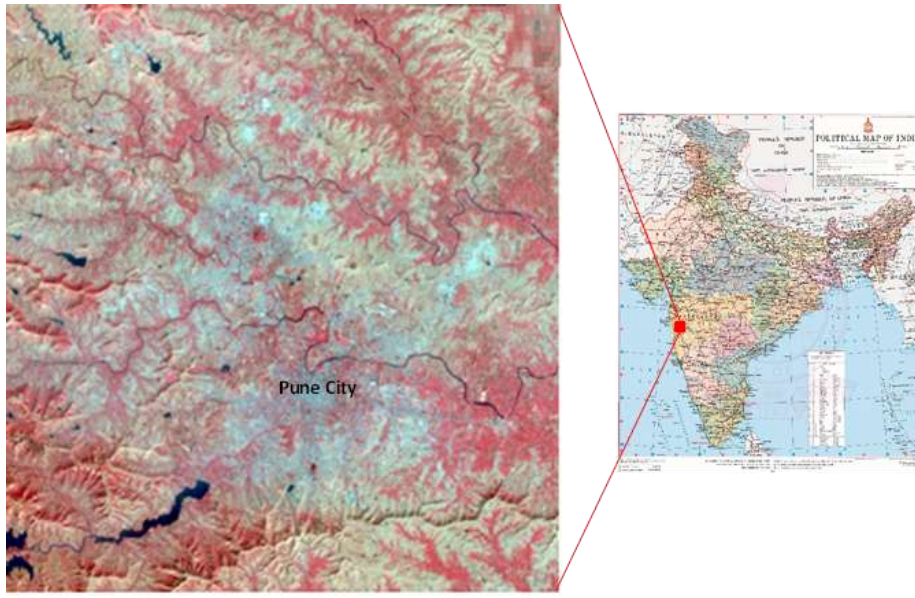

Figure 1. Location map of the study area (India map source: SOI)

Material and Methods

Figure 3 is a flow diagram of methodology of the study. The satellite images were obtained from USGS (earthexplorer.usgs.gov) for the years 1990, 2000, 2010, and 2020 (Figure 2). The details of the satellite data are listed in Table 1. All the satellite images of the recording years were downloaded and stacked. The upper left extent: 73d35'30.4512"E Longitude - 18d20'35.3796"N Latitude and lower right extent: 74d5'17.25"E Longitude - 18d49'52.9032"N Latitude were used for extracting study area and extracted all the images.

The satellite images of the respective four recording years were classified using Orfeo Toolbox, the Machine Learning Algorithm. The obtained classified result is assessed

using the Semi-Automatic Classification Plug-in and predicted and validated using ANN-Molusce and CA- Markov Chain Model. The software used to obtain the desired output were QGIS and Orfeo Tool Box (OTB) . The classified output is categorized into an urban area with built-up form, Vegetation, Waterbody, and Unclassified (bare land, fallow land, etc.). However, the study gives more emphasise to urban category. The details of the satellite images obtained and used are shown as Figure 2 and a flow diagram of methodology as (Figure 3) in the subsequent sections.

Table 1: Details of satellite images used in the study

Satellite	Spatial Resolution (m)	Date of Acquired
Landsat 5	30	05/05/1990
Landsat 5	30	29/03/2000
Landsat 8	30	26/04/2010
Landsat 8	30	07/05/2020

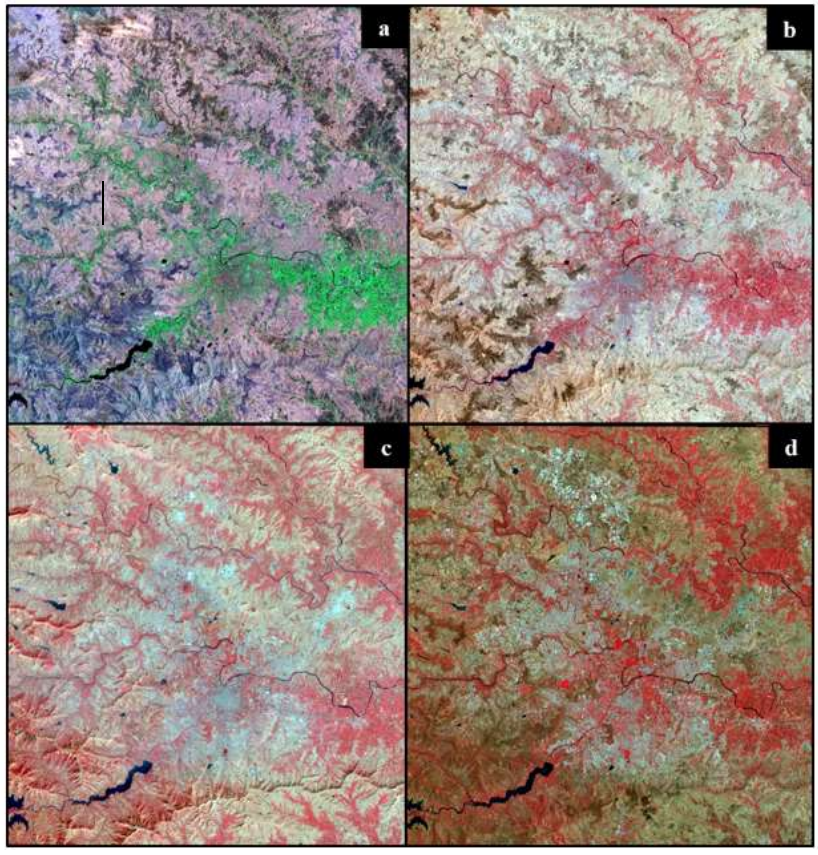

Figure 2. Landsat satellite image of the study area showing different year of acquisition: a) Year 1990 (RGB 345), b) Year 2000 (RGB 432), c) Year 2010 (RGB 432), d) Year 2020 (RGB 543)

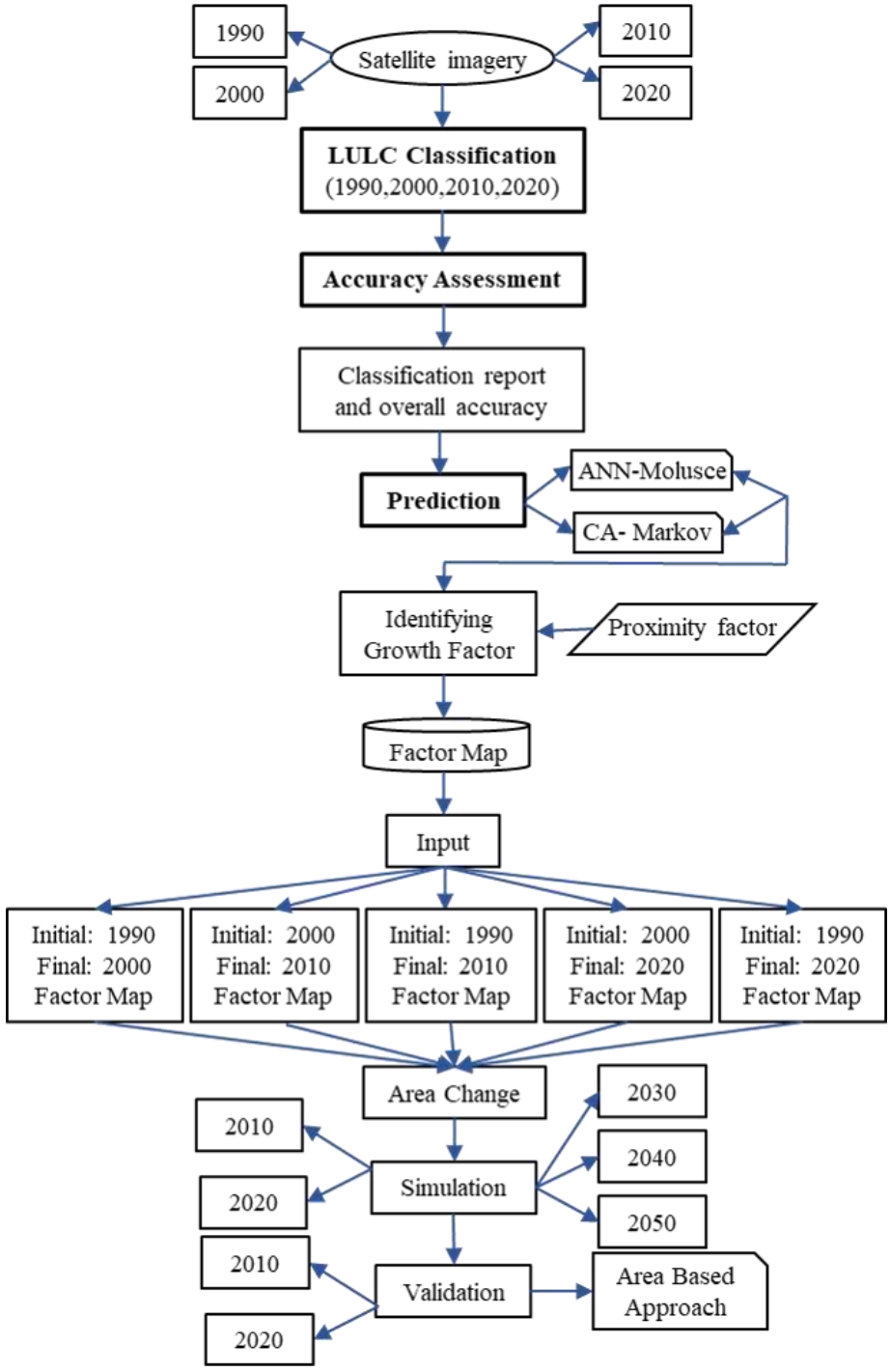

Figure 3. The flow-chart of methodology implemented for LULC classification and future prediction (simulation)

i) Image processing and Classification

Orfeo Tool Box (OTB) is open-source software for image processing. It aids in processing terabyte-scale high resolution optical, multispectral, hyperspectral, and radar pictures. OTB can be used for many different applications such as ortho-rectification or pan-sharpening, supervised or unsupervised classification, feature extraction, and SAR processing. For classification, OTB uses Monteverdi, a lightweight image rendering and processing tool. Monteverdi is a tool for fast visualization of processed results, which can display images in sensor geometry.

Through Image Statistics, Train Image Classifier, and Image classifier, image classification was carried out. It does not replace GIS software, suitable to edit, display, and relate different sources of geographic information both raster and vector. It uses Support Vector Machine (SVM) as a classifier due to its high-quality classification ability for classification and regression. In this method, to train the system, representative samples for each land cover class are selected. Then, "training sites" are applied to the entire image. The first step is to create training samples for each class. Following this, a signature file would be generated with all training samples' spectral information. Finally, the signature file is used to run a classification and obtain desired output. Object-based classification was carried out using Support Vector Machine (SVM) algorithm for Supervised Classification. Using OTB, LULC was carried out for 1990, 2000, 2010, and 2020 dataset. In this study, more emphasis is given for urban class extraction and accuracy assessment as it focuses more on urban area growth.

ii) Accuracy Assessment

QGIS based Semi-Automatic Classification Plug-in (SCP) allows access to the accuracy of the image classification result. SCP was used to process raster data, resulting in an automated workflow and simple land cover classification that local practitioners may easily access. It computes classification statistics, band set processing, ROI creation, and Accuracy assessment. Classified outputs of 1990, 2000, 2010 and, 2020 are assessed using SCP.

In the initial stage, the Classification Report tool in the Semi-Automatic Classification Plug-in is used to generate a report on the area covered by each class. Data is provided in the form of number of pixels, area (in sq. m) and the percentage out of the total area. The Classification reports confirm the observations made from the Image classification results. In this study, multiple data source like open street map, google imagery, town planning maps, NRSC LULC map and published maps from journal / book are used for validating results through online. The Band Set Tool in SCP is used to create a Band set comprising the classification raster that is utilised as an input by other tools for accuracy assessment. Multiple Regions of Interest (ROI) has to be generated which would be used to compare the accuracy of the classified image. The Multiple ROI Creation tool in SCP generates multiple sample pixels which then are assigned a class by comparing to the

initial raster image. This random sample is then used in the accuracy tool to assess the accuracy of the classification. The final step of accuracy assessment is to determine overall accuracy and kappa statistics. The key benefit of this method is that it allows one to evaluate accuracy across homogeneous surfaces created automatically from ROIs rather than on single spots only.

iii) Prediction

a) Artificial Neural Network (ANN) -MOLUSCE

MOLUSCE refers to the Model for Land-use Change Evaluation. It analyses, models, and simulates land-use changes. It employs an algorithm that can be developed for LULC analysis, urban analysis, and forest application. It is well suited to analyse land use and forest cover changes, model land-use transition potential, simulate future LULC change and, validate the predicted output with actual LULC [8]. The study uses input as five sets: 1990-2000, 2000-2010, 1990-2010, 2000-2020 and 1990-2020. Hence, the prediction was carried out for 2010, 2020, 2030, 2040 and 2050 respectively and validation was carried out for 2010 and 2020 using both software and an area-based approach as the actual LULC of them is available. Transition Potential Modelling is carried out using ANN, which is the Geo-AI model of neural networks that can predict and provide solutions to complex problems. The ANN method is faster compared to traditional techniques and can solve multiple problems. Due to these advantages, ANN methods have been used in numerous real-time applications [9].

b) Input

Two recording year LULC maps are added to the MOLUSCE Plug-in together with previously processed variables' factors for LULC changes. Roads are one of the contributing factors to urban growth in the study area, therefore, major highways, including the Mumbai-Pune Expressway are used in computing the factors. Distance from the road was calculated to determine the growth factor using proximity Plug-in in QGIS. As the Classified LULC for the recording years 2010 and 2020 are available, the simulated output of 2010 and 2020 is validated with actual classified results.

Growth driver - proximity factor: In order to create a layer with distance-from-nearest-road as a habitat, distance from Road GIS layers employs QGIS for processing spatial information. This layer allows us to generate a probability of occupancy layer and extract distance-from-nearest-road for each location in order to model data. Using Distance from Road in QGIS, spatial variable, raster road tiff file is created. Analysing closeness is done using the spatial analyst tool. The distance toolset includes tools that either assign each cell to the nearest feature or build a raster indicating each cell's distance from a group of features. Tools for measuring distance also determine the quickest route through a corridor between two points to reduce two different sets of expenses. For overlay analyses, distance from any entity is used as inputs. The spatial variable used in this study is the

raster roadway data which is obtained as distance from the road using Proximity Plug-in in QGIS.

c) Land use change

This stage computes transition between the first recording year of LULC and the second recording year of LULC. In this research, five datasets are considered having initial and final recording years. The value of this LULC change in the form of change map is mapped, which would then be used for the fourth stage. The outcomes of the steps are achieved using Transition Probability and Transition Suitability maps. The distribution of each LULC class was projected based on the transition probability p_{ij} between two LULC classes (i and j). p_{ij} was determined over a specific period from time t to time t+1 is presented as Equation (1).

$$p_{ij} = \frac{n_{ij}}{n_i} \quad (1)$$

$$\sum_{j=1}^{k} p_{ij} = 1$$

Where,

n_i = the total number of pixels of class i transformed over the transition period,

n_{ij} = the number of pixels transformed from class i to j,

k = the number of LULC classes.

The distribution of each LULC class at time t+1(M_{t+1}) was projected forward using the LULC distribution at the beginning time t (M_t) and the transition probability matrix P;

$$P * M_t = M_{(t+1)}$$

d) Transition Potential Modelling

Artificial Neural Network (Multi-Layer Perception), Weights of Evidence, Multi Criteria Evaluation, and Logistic Regression are the four techniques available through MOLUSCE plug-in for predicting LULC change. This study utilised LULC prediction model using an artificial neural network (multi-layer perception) technique. After processing resultant image, the outcome would display the kappa validation value. By using training data and sampling, this method examined the altered pixels between two LULC maps as well as spatial factors. The random sampling mode was specified. Additionally, the model must include the maximum iteration and neighbouring pixels.

ANN is a component of computational intelligence techniques. The utility is justified by the fact that it can be used to address issues wherein a lot of input data is dealt, and where the solution technique is ambiguous or challenging, to employ. It is exceedingly challenging to apply many analysis parameters linked to terrain or land cover. Because

artificial neural networks implement the principles of fuzzy logic, which is feasible to assess the suitability of a terrain in a continuous range, such as between 0 and 1. The interactions between neurons and the alteration of the weight connections connecting them make up the artificial neural network's essential component. This adjustment is dependent on the supplied information and the anticipated.

e) Simulation

Cellular automata simulation is used to carry out the third year LULC change prediction process because the prior kappa value complies with the assessment standard. The simulation iteration count should be set to 1. Taking into consideration the input dataset, simulation maps for the recording years 2010, 2020, 2030, 2040 and 2050 are predicted.

f) Validation

The validation is processed using the reference map (the third year LULC as it actually occurred), referred to the simulated map (the third year LULC prediction), and computation of the overall Kappa value. Then it is to the cellular automata simulation stage to predict future LULC after the total kappa value satisfies the assessment requirement. The year of prediction is determined by the difference between the final year and the initial year and the following year of prediction can be calculated by adding the difference to the final year.

The Kappa coefficient [12] is calculated using equation (2).

$$\text{Kappa} = \frac{p_o - p_e}{1 - p_e} \quad (2)$$

Where,

p_o is the proportion of observed agreements, and p_e is the proportion of agreements expected by chance.

$$p_o = \sum_{c=1}^{i} p_{ij}$$

$$p_e = \sum_{c=1}^{i} p_i T_i T_j$$

Where,

p_{ij}= the i-th and j-th cell of the contingency table,

p_i T_i= the sum of all cells in i-th row,

$p_j T_j$= the sum of all cells in j-th column,

c = the count of the raster category.

iv) Cellular Automata (CA) - Markov Chain Model

The MC model is a stochastic process model that illustrates the concept of changing from one state to another. It has a well-coded algorithm, namely the transition probability matrix [13]. The probabilities of transitioning between these time steps are calculated in the MC model using two LULC of different periods as input. This model has been used to predict urban sprawl and land-use change [13–15], however the MC model is used in this study to confirm the effectiveness. Similarly, with the same input set, prediction and validation is done for recording years using Markov Chain Model. Both the ANN-MOLUSCE and CA- Markov use the same methodology to predict and validate the output. But ANN- MOLUSCE uses the TIFF data file type, whereas CA- Markov Chain Model uses the RST data file type as input and output. The specific formula is presented as equation (3).

Mathematically CA can be expressed as [11],

$$(\{S_{t+1}\} = f(\{S_t\} * \{I_t^h\} * \{V\}) \quad (3)$$

Where,

$\{S_{t+1}\}$=the state of a cell in the CA at time (t +1),

$\{S_t\}$ = the state of a cell in the CA at time (t),

$\{I_t^h\}$ = the neighbourhood,

{V} = the suitability of a cell for urban growth,

f () = the transition rules,

t = the time steps in temporal space,

h = the neighbourhood size.

Results and Discussion

The LULC classification is performed for four recording years as shown in the Figure 4-a, b, c, d using Orfeo Toolbox. Classification report of four recording years is shown in the Figure 5. Accuracy of the output is assessed using Semi-automatic Plug-in and the overall accuracy is shown in the Figure 6. Likewise, using ANN- MOLUSCE and CA- Markov Chain Model, transition matrix is obtained for all the five input datasets, which are shown in the Table 2.

The transition matrix is obtained for the recording years given in the input dataset through ANN- Molusce and CA-Markov. From 1990 to 2020, urban areas and vegetation have increased by 20% and 17%, respectively. The water body and unclassified landform have been decreased, probably due to urbanisation by 0.5% and 38%, respectively. The transition matrix obtained through ANN- MOLUSCE is more efficient and accurate than CA-Markov due to its high probability outcomes. Growth of IT Parks, industrial areas, educational institutes contribute to the development of the city which needs to accommodate the ever-increasing, pulled and migrated population. The increase in built-up area in the city results in rapid urbanisation.

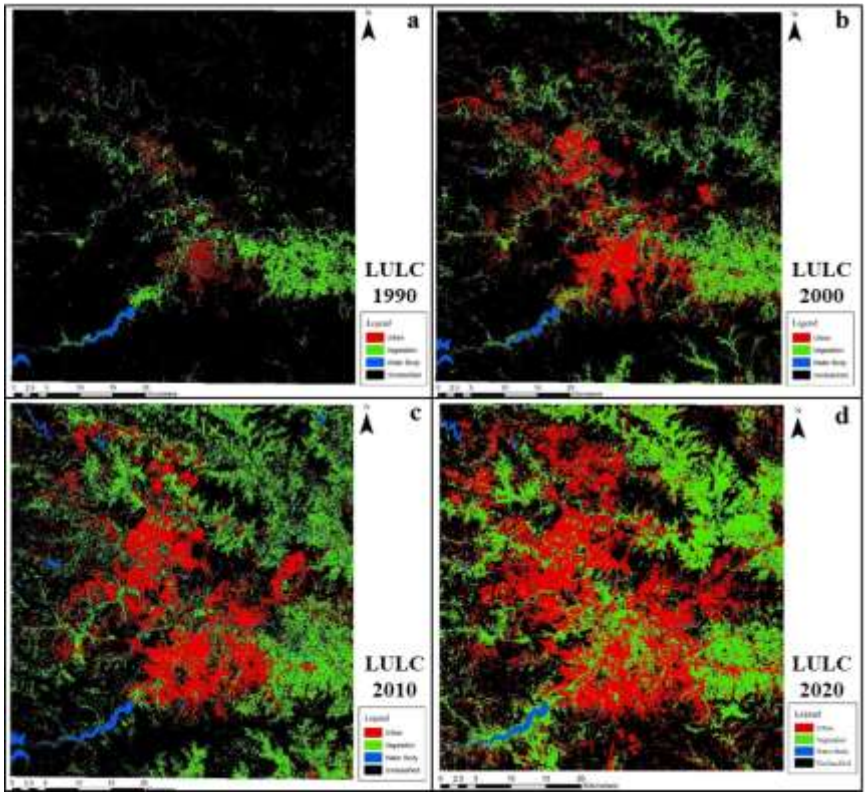

Figure 4. LULC classification map of the study area. Landsat satellite imageries are used to derive the LUCL: a) 1990, b) 2000, c) 2010, d) 2020

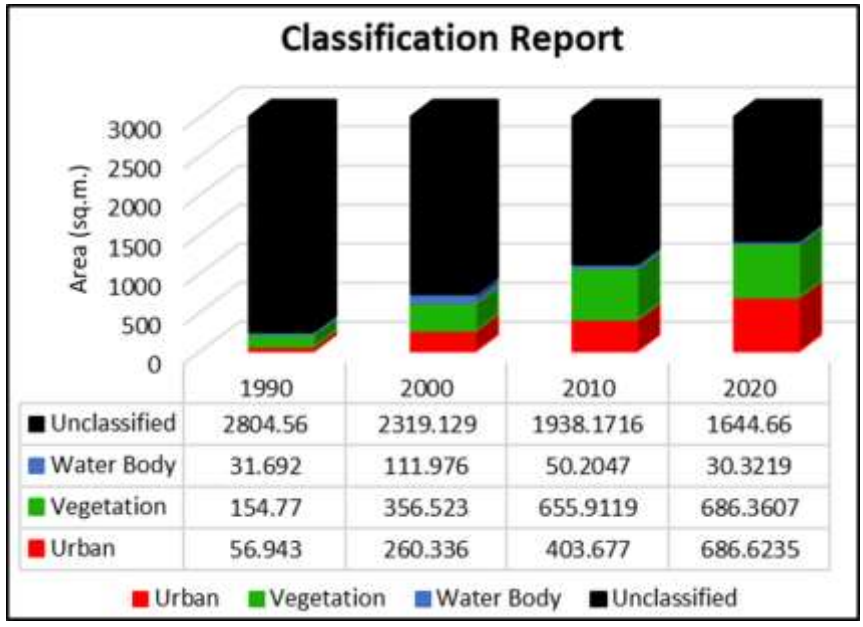

Figure 5. Comparation of LULC classification output with different year

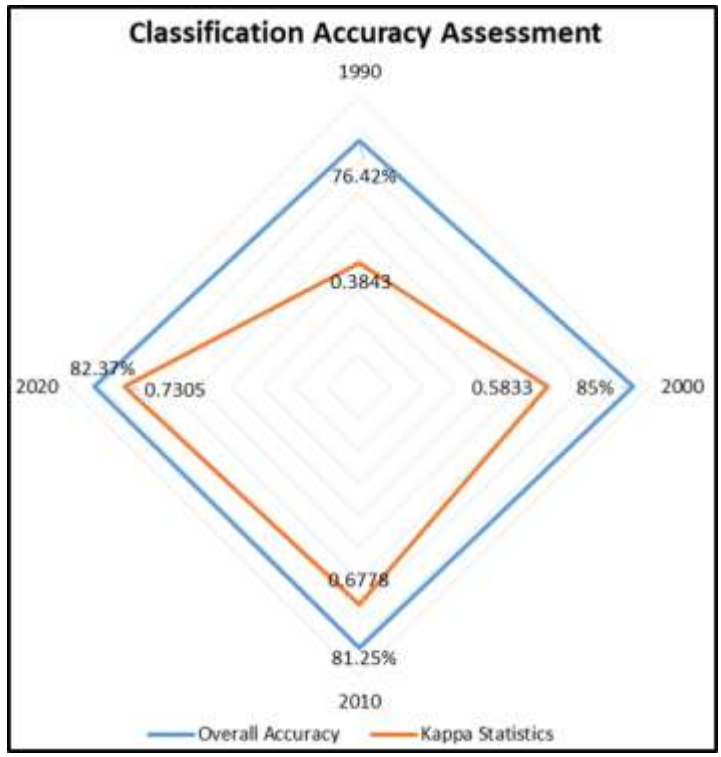

Figure 6. LULC classification accuracy assessment shown in the Radar chart. Blue and orange line show overall accuracy and Kappa statistics, respectively

Simulation is carried out for the future recording years 2010-50 and validation for 2010 and 2020 is carried out since the actual LULC is available for same recording years. Figure 7-a, b, c, d, e show the predicted LULC from 2010 to 2050, respectively. The predicted LULC changes (2010, 2020, 2030, 2040, and 2050) provide average Kappa overall accuracy values of 50%. According to the actual LULC, between 1990 and 2020, the built-up area witnessed the highest change, while the barren region observed the biggest change in a drop. During 1990-2020, barren land is transformed into the built-up landform. Also, further validation was carried out for simulated 2010 and 2020, as the actual LULC for both recording years is available. As per the simulated map, the urban class is supposed to expand by 21% in 2030, 23% in 2040, and 26% by 2050 from 1% in 1990. In 2030, the urban expansion is supposed to be blooming along the National Highways, State Highways and around the Chakan industrial belt. In, 2040 and 2050, the same areas are probably strengthened with the core area developing in radially outward. Pune City Mission plans to develop the city in terms of transportation, education, and economy. Hence, the additional pulled and migrated population can be accommodated along with the roadway infrastructure for easy access, decreasing the substance on the core area. The simulated outcomes for future years are predicted and validation was carried out (Table 3, 4).

Table 2: Transition Matrix for the year 1990-2000, 2000-2010, 1990-2010, 2000-2020 and 1990-2020

Class Transition	Method / Algorithm	Transition Year				
		1990-2000	2000-2010	1990-2010	2000-2020	1990-2020
Urban - Urban	ANN	0.605	0.633	0.588	0.656	0.614
	CA - Markov	0.511	0.538	0.545	0.633	0.457
Urban - Vegetation	ANN	0.156	0.147	0.193	0.214	0.283
	CA - Markov	0.135	0.185	0.143	0.199	0.237
Urban - Water Body	ANN	0.011	0.030	0.018	0.011	0.006
	CA - Markov	0.001	0.038	0.016	0.010	0.006
Urban - Unclassified	ANN	0.228	0.190	0.200	0.119	0.097
	CA - Markov	0.200	0.239	0.196	0.099	0.079
Vegetation - Urban	ANN	0.148	0.074	0.155	0.195	0.266
	CA - Markov	0.115	0.093	0.116	0.157	0.228
Vegetation - Vegetation	ANN	0.625	0.632	0.535	0.647	0.612
	CA - Markov	0.430	0.537	0.264	0.623	0.569
Vegetation - Water Body	ANN	0.013	0.013	0.031	0.007	0.015
	CA - Markov	0.007	0.016	0.031	0.007	0.013
Vegetation - Unclassified	ANN	0.214	0.282	0.279	0.151	0.107
	CA - Markov	0.196	0.355	0.247	0.124	0.075
Water Body - Urban	ANN	0.191	0.003	0.123	0.091	0.165
	CA - Markov	0.149	0.003	0.095	0.093	0.097
Water Body - Vegetation	ANN	0.194	0.005	0.176	0.223	0.348
	CA - Markov	0.118	0.005	0.149	0.199	0.299
Water Body - Water Body	ANN	0.483	0.247	0.554	0.660	0.430
	CA - Markov	0.325	0.210	0.539	0.612	0.387
Water Body - Unclassified	ANN	0.132	0.745	0.147	0.026	0.057
	CA - Markov	0.110	0.782	0.129	0.016	0.060
Unclassified - Urban	ANN	0.075	0.091	0.132	0.206	0.236
	CA - Markov	0.046	0.130	0.103	0.147	0.198
Unclassified - Vegetation	ANN	0.088	0.169	0.202	0.186	0.220
	CA - Markov	0.059	0.239	0.165	0.160	0.199
Unclassified - Water Body	ANN	0.004	0.004	0.010	0.003	0.006
	CA - Markov	0.001	0.006	0.009	0.002	0.005
Unclassified - Unclassified	ANN	0.832	0.735	0.656	0.605	0.538
	CA - Markov	0.458	0.625	0.348	0.365	0.457

Table 3: Validation for the year 2010 using an area-based approach

LULC Class	Method / Algorithm	Validation for the year 2010				Validation for the year 2020			
		Actual	Simulated	Validated	% Difference	Actual	Simulated	Validated	% Difference
Urban	ANN	403.68	266.08	-137.6	34.08%	746.8	405.89	-340.9	45.64%
	CA - Markov		220.67	-183	45.33%		382.51	-364.28	48.77%
Vegetation	ANN	655.91	355.05	-300.85	45.86%	746.53	656.6	-90.44	12.11%
	CA - Markov		280.71	-375.19	57.20%		580.94	-165.59	22.18%
Water Body	ANN	50.2	111.84	61.63	55.11%	90.5	50.39	-40.1	44.31%
	CA - Markov		83.81	33.61	66.95%		42.97	-47.52	52.51%
Un-classified	ANN	1938.2	2315	376.82	16.27%	1464.1	1935.58	471.46	32.20%
	CA - Markov		2462.75	5245.84	27.06%		2041.52	577.4	39.43%
% Correctness		ANN= 62.63% CA - Markov= 50.87%				ANN= 71.75% CA - Markov= 59.28%			

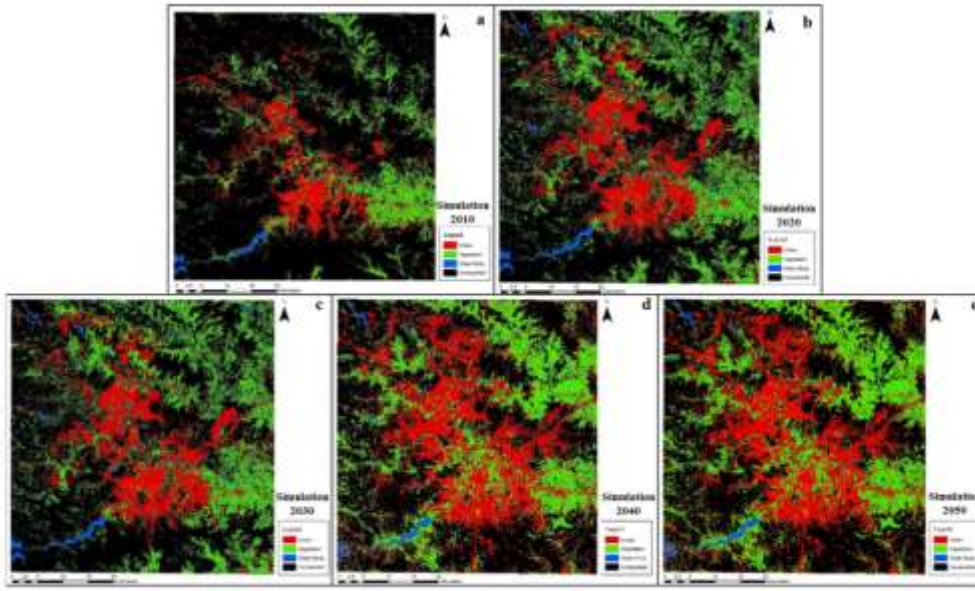

Figure 7. Predicted (simulated) LULC change map, derived through ANN and CA-Markov models using previous year data, a, b, c, d and e represents predicted LULC: a) 2010, b) 2020, c) 2030, d) 2040, e) 2050

Table 4: Validation for the year 2010 and 2020 using software

		2010	2020
% Correctness	ANN	64.18%	69.63%
	CA - Markov	48.86%	55.25%
K value	ANN	0.6122	0.6511
	CA - Markov	0.392	0.4438

The percentage correctness for the years 2010 and 2020 is validated using an area-based approach and software. Table 4 shows the output obtained by both the methods, which is almost the nearby values of one another and indicates that the validation is appropriate. But, as the techniques used, MOLUSCE (ANN based) is more accurate and efficient than Markov model, because percentage correctness using MOLUSCE by both methods is greater than 60% as compared to Markov. In addition, K value is satisfying the value obtained using MOLUSCE than Markov Chain Model.

In the study, the classification is obtained using the Machine Learning Algorithm is not only accurate than traditional superlative method but also the most appropriate tool for predicting land use land cover changes and identifying the trends with different logical conditions. Hence, the predicted and validated outcomes of ANN are more efficient than the Markov Model in terms of probability statistics and percent correctness. The major transition in the future indicates a significant growth of urban land cover due to a rapid increase in population and tremendous growth of industrial areas all around Pune city. Over recent decades, the growth has been more towards the west direction due to: proximity to Mumbai and connectivity with the JNPT Port, near Nhava Shewa; bigger scope for opportunities of employment and businesses; more suitable and better living standards in comparison to other similar tier cities in the country; highly conducive environment for faster and ease of doing business; the large-scale availability of abundant water sources in the study region; and conducive weather conditions.

Conclusion

Geospatial technologies along with artificial intelligence methods demonstrate that urban growth is expected to be occurred along Pune-Bengaluru National Highway, Pune-Ahmednagar State Highway and Pune-Solapur-Hyderabad National Highway, which are proposed to be connected by Ring Road for smoother and efficient traffic and transportation to avoid city traffic congestion. The proposed construction of roads will have bigger impact on the development in the ribbon strips along the roads. Considering the potential of the recent advanced artificial intelligence and geospatial technologies, it is observed that it can detect and predict all the requisite details of the various changes occurring in the land use and land cover. Hence, it is recommended to the administrators, planners and policy/decision makers to make use of Geo-AI technology for urban planning studies, analysis, and decision-making process. The limitation of this study is that the difference between simulated and validated models' results are huge, so "distance to road" factor was only considered. Therefore, other proximity factors like industrial location, rivers, waterbodies and hills should also be included in future research.

Acknowledgement

The authors wish to thank editors and anonymous reviewers for their comments and suggestions that helped to improve the paper. In addition, they thank Prof. Isha Panse, COEP and C-DAC, Pune, for providing facilities and support to complete the work successfully.

References

1 Rawat, J.S. and Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India; The Egyptian Journal of Remote Sensing and Space Science, 18(1):77-84.

2 Li, L.; Lu, D. And Kuang, W. (2016). Examining Urban Impervious Surface Distribution and Its Dynamic Change in Hangzhou Metropolis. Remote Sens-Basel, 8(3).

3 Basommi, L. P.; Guan, Q.; Cheng, D. and Singh, S. K. (2016). Dynamics of land use change in a mining area: a case study of Nadowli District, Ghana; J Mt Sci. 13(4), 633-42.

4 Marwa W.; Halmy, A.; Gessler, P. E.; Hicke, J. A. and Boshra, B. S. (2015). Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA; M.W.A. Applied Geography, 63.

5 Sivakumar, V. (2014). Urban mapping and growth prediction using remote sensing and GIS techniques, Pune, India; The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; Volume XL-8.

6 Maktav, D.; Erbek, F. S. and Jurgen, S. (2005). Remote sensing of urban areas. International Journal of Remote Sensing, vol. 26, no. 4, 655-659.

7 Liping, C.; Yujun, S.and Saeed, S. (2004). Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques: A case study of a hilly area, Jiangle, China; PLoS ONE 13(7): e0200493.

8 Gismondi, M.; Kamusoko, C.; Furuya, T.; Tomimura, S. and Maya, M. (2014). MOLUSE: An open-source land use Change analyst for QGIS, Documentary on Molusce Plug-in, 20, 612-614.

9 Xia, L. and Yeh, A.G. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS, International Journal of Geography Information Science, 16, 323-343.

10 Siddiqui, A.; Siddiqui, A.; Maithani, S.; Jha, A. K.; Kumar, P. and Srivastav, S.K. (2018). Urban growth dynamics of an Indian metropolitan using CA Markov and Logistic Regression, The Egyptian Journal of Remote Sensing and Space Sciences, 21, 229-236.

11 Saputra, M. H. and Lee, H. S. (2019). Prediction of Land Use and Land Cover Changes for North Sumatra, Indonesia, Using an Artificial-Neural-Network-Based Cellular Automaton, Article in Sustainability, 11, 1-16.

12 Arsanjani, J. J.; Helbichb, M.; Kainza, W. and Bolooranic, A. D. (2013). Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion, International Journal of Applied Earth Observation and Geoinformation, 21, 265-275.

13 opez, E.; Bocco, G.; Mendoza, M. And Duhau, E. (2001). Predicting land cover and landuse change in the urban fringe: a case in Morelia city, Mexico, Landscape and Urban Planning, 55, 271-285.

14 Arsanjani, J. J., Kainz, W. And Mousiv and, A. J. (2011). Tracking dynamic land-use change using spatially explicit Markov Chain based on cellular automata: the case of Tehran, International Journal of Image and Data Fusion, 2(4), 329-345.