Future Prediction of UrbanGrowth of Prayagraj

Mini Project
5th Semester
in
Information Technology

Submitted by
Hritik Kumar (IIT2021014)
Deepak Raj (IIT2021039)
Pratik Sharma (IIT2021026)
Mohit Kumar (IIT2021059)
Shresth Sharma (IIT2021089)
Under the guidance of
Dr. Triloki Pant

Department of Information Technology
Indian Institute of Information Technology, Allahabad 211015, (U.P.) INDIA

Indian Institute of Information Technology, Allahabad

Candidates Declaration

I hereby declare that the work presented in this report entitled **Future Prediction of Urban Growth of Prayagraj**, submitted towards the fulfillment of BACHELOR'S THESIS report of Information Technology at Indian Institute of Information Technology Allahabad is an authenticated record of our original work carried out under the guidance of Dr. Triloki Pant. Due acknowledgments have been made in the text to all other material used. The project was done in full compliance with the requirements and constraints of the prescribed curriculum.

Hritik Kumar(IIT2021014)

Date:
Place: Prayagraj

Indian Institute of Information Technology, Allahabad

Candidates Declaration

I hereby declare that the work presented in this report entitled **Future Prediction of Urban Growth of Prayagraj**, submitted towards the fulfillment of BACHELOR'S THESIS report of Information Technology at Indian Institute of Information Technology Allahabad is an authenticated record of our original work carried out under the guidance of Dr. Triloki Pant. Due acknowledgments have been made in the text to all other material used. The project was done in full compliance with the requirements and constraints of the prescribed curriculum.

Deepak Raj (IIT2021039)

Date:
Place: Prayagraj

Indian Institute of Information Technology, Allahabad

Candidates Declaration

I hereby declare that the work presented in this report entitled **Future Prediction of Urban Growth of Prayagraj**, submitted towards the fulfillment of BACHELOR'S THESIS report of Information Technology at Indian Institute of Information Technology Allahabad is an authenticated record of our original work carried out under the guidance of Dr. Triloki Pant. Due acknowledgments have been made in the text to all other material used. The project was done in full compliance with the requirements and constraints of the prescribed curriculum.

Pratik Sharma(IIT2021026)

Date:

Place: Prayagraj

Indian Institute of Information Technology, Allahabad

Candidates Declaration

I hereby declare that the work presented in this report entitled **Future Prediction of Urban Growth of Prayagraj**, submitted towards the fulfillment of BACHELOR'S THESIS report of Information Technology at Indian Institute of Information Technology Allahabad is an authenticated record of our original work carried out under the guidance of Dr. Triloki Pant. Due acknowledgments have been made in the text to all other material used. The project was done in full compliance with the requirements and constraints of the prescribed curriculum.

Mohit Kumar(IIT2021039)

Date:
Place: Prayagraj

Indian Institute of Information Technology, Allahabad

Candidates Declaration

I hereby declare that the work presented in this report entitled **Future Prediction of Urban Growth of Prayagraj**, submitted towards the fulfillment of BACHELOR'S THESIS report of Information Technology at Indian Institute of Information Technology Allahabad is an authenticated record of our original work carried out under the guidance of Dr. Triloki Pant. Due acknowledgments have been made in the text to all other material used. The project was done in full compliance with the requirements and constraints of the prescribed curriculum.

Shresth Sharma(IIT2021089)

Date:
Place: Prayagraj

Indian Institute of Information Technology, Allahabad

Certificate from Supervisor

This is to certify that the statement made by the candidate is correct to the best of my knowledge and belief. The project titled **Driving Factors & Future Prediction of Urban Growth of Prayagraj** is a record of candidates' work carried out by him under my guidance and supervision. I do hereby recommend that it should be accepted in the fulfillment of the requirements of the Bachelor's Thesis at IIIT Allahabad.

Dr. Triloki Pant
(On final examination and approval of the thesis)

Date:
Place: Prayagraj

Indian Institute of Information Technology, Allahabad

Certificate of Approval

The forgoing thesis is hereby approved as a creditable study carried out in the area of Information Technology and presented in a manner satisfactory to warrant its acceptance as a pre-requisite to the degree for which it has been submitted. It is understood that by this approval the undersigned does not necessarily endorse or approve any statement made, opinion expressed or the conclusion drawn therein, but approves the thesis only for the purpose for which it is submitted.

Committee on final examination for the evaluation of thesis:

1. Dr. Triloki Pant
2. Prof. Uma Shankar Tiwari
3. Dr. Muneendra Ojha
4. Dr. Navjot Singh
5. Prof. Pritish Varadwaj

____
(Dean A & R)

Date:
Place: Prayagraj

Plagiarism Report

chapters/.png

Contents

Abstract

Urban Growth models have been used extensively to predict the Land Use and Cover Change (LUCC) to improve the quality of life and to know the impact ofthese changes. The rate of population growth in Prayagraj is increasing every year, so it is important to predict the changes for sustainable development. In this course of investigation, we are trying to find potential driving factors that affect urban growth using remote sensing data and these factors are used to predict the future LULC classes of Prayagraj for the year 2026.CA-Markov Model (LCM) is used in this study in the interest of the ability to handle the dynamic process of LULC. Through this study, out of the proximity factors considered it was found that Distance from Buildings and Roadways affect urban growth positively. It wasalso observed that urban growth has increased by 11.53% from the year 2018 to the year 2022.

Keywords: Landsat, Cellular Automata, CA-Markov Model, Land Use and Cover Change (LUCC), Logistic CA-Markov Model.

Chapter 1

Introduction

Urban Growth prediction plays an important role in planning a sustainable urban development especially in developing countries and cities. Prayagraj, our area of study is one of the developing cities in India. It's important to analyze previous predictions and predict for the future years, that could help governments in planning many Infrastructures. This could help plan a good efficient resource allocation like water and electricity, Infrastructure Development, Sustainable Environment, Disaster planning and Transportation Planning.

Urban Planning in an efficient way could generate good revenue and also can be used for cost reduction. Urban growth prediction can contribute to disaster risk reduction efforts. By identifying areas prone to natural hazards. Three rivers flow in Prayagraj and it's important to set up disaster units accordingly and also plan to save environmental reserves like forests. It significantly enhances designated areas as agricultural, industrial, and/or urban sectors within a region.An important issue for developing nations is the haphazard urbanization, which occurs lacking effective planning strategies in the majority of cases. This could generate a huge revenue for Real estate companies and have a better knowledge of the trends followed in that area.

To accommodate the increasing number of urban residents, a significant amount of expansion of existing urban areas and the development of new urban regions are needed. The process of rapid urbanization has brought forth a plethora of environmental and socioeconomic issues. These problems include urban heat islands, high levels of energy consumption, air pollution, public health concerns, deforestation, loss of biodiversity, and degradation of fertile agricultural lands.

Proper planning of urban development can enhance the quality of life. The process of urbanization has resulted in significant changes in land use and land cover, with vast tracts of farmland being converted to urban areas. To ensure the well-being of human society, it is crucial to plan urban systems efficiently and sustainably. This requires simultaneous monitoring of urban development progress and modeling to assess the impact of urbanization under various scenarios.

Over the years, the Urban growth models have evolved corresponding to the spatial and temporal dimensions. [1] Customarily to predict the Urban growth models, three types of urban growth models are used, specifically Land use/-transportation model, the Cellular Automata model and Agent based models are being used. [2] According to recent change detection using Landsat imagery,urban areas in China have multiplied by more than two times over the span of 20 years (1990–2010); a substantial portion of these urban areas were once agricultural land.

[3]A theoretical framework for modeling was put forward for comparison with the previous models. There are two distinct modeling methods: data-driven and

process-driven.[4] Recently evolved urban models took regional interactions and geographical variability into account. [5] Urban Growth models have proven to be effective in disaster situations in Malaysia, especially in the coastal region forregular monitoring of the land and to make informed decisions with the local government. Few case studies combine to use Land use Expansion Analysis Strategy and CA Model based on random patch seed for multi- classes for betterresults as this hybrid model can overcome the problem of landscape dynamic change.

Chapter 2

Related Work

[6] To predict the future urban growth for an Indian Metropolis city like Pune, Lakshmi N. Kantakumar et al. have considered driving factors. They have used data from Landsat 5 and landsat8 data set from 1992 to 2013. This study has concentrated on the driving factors which have affected the most in the previous developments and will be affecting more on future predictions. This study suggests that the relationship between propelling factors and urban growth is frequently modeled using a regression-based empirical model. Fifteen factors were used in this study and analyzed which driving factors are more impactful on Urban Growth. They have used Chi Square Test to determine the dependency of each driving factor and Multiple Logistic Regression to predict the future results. After considering fifteen driving factors, it founds out proximity factors and socio-economic factors are more urban growth driven factors.

[7] This study emphasizes the necessity of prediction validation. Land Change Modeler(LCM) has been used in a study conducted by Veluswamy Venkatramanan et al. to predict the urban growth in Delhi. This research seeks to ascertain the rate of geographical transformations, as well as their causes and effects, and to forecast urban development patterns in Delhi and its surround-

ings.Landsat satellite images of the years 2010,2020 are used to predict the changes for the year 2030.This study examined the environmental and socio-economic factors that have a direct impact on the pattern of urban expansion. Five Driving factors are considered, Cramers V coefficient is considered to assign a priority to each driving factor.For the Accuracy of the prediction, the value of kappa coefficient is considered.LULC changes provide a scientific foundation to address the implications of the reciprocal relationship between humans and the environment in the studied area.By comparing the changes and co-relating it with the driving factors, can state there could be many challenges for a sustainable development in future especially in cities like Delhi and Mumbai.

[8] Kongming Li et al. predicted Land Use and Land Cover in 2030 using the Logistic-Cellular Automata-Markov chain(LCM). They have used data from 1980 to 2018 and used for exploring, analyzing the potential driving factors using binary logistic regression.Their area of study is Gansu Province, China and proposed a new and promising methodology for arid and semiarid regions.They have considered three type of factors for the research namely Natural factors, Proximity factors, Socio-economic factors. Out of these, Natural factors were more driving for growth, especially the areas featuring more farmlands. Socio-economic factors were not notable factors in this region. A little improvement on the model is needed because the accuracy is not good for water areas and built up areas. The model is validated after predicting the urban growth for 2015 and comparing it with the actual maps using kappa index.

Chapter 3

Study Area

This study area covers Prayagraj District, the most populous district in Uttar Pradesh in India. Allahabad has a humid subtropical climate, characterized by hot summers and cool winters. It covers an area of 5482 sq km and lies between 25° 31 ’21.55 ” N - 25°17’22.08”N latitudes , 81°39’17.3”E - 81°57’39.17”E longitudes. The annual average rainfall is 1042mm and with an average temperature of 25.8 deg Celsius.

A United Nations report states that Prayagraj’s urban population was around 5.9 million in 2014 and is expected to increase to 8.7 million by 2030. This

Figure 3.1: Area of Study

shows that the city has recently seen rapid population increase, which may have facilitated urban growth and densification. With these statistics it’s unambiguous that Prayagraj has and will undergo many changes in future. Furthermore, thereare many infrastructural developments and this could lead to urban growth.

Chapter 4

Background Model Description

4.1 Markov Chain Model

Markov Chain is a statistical model that is used to predict future states of a system purely based on the current state and the transition probability matrix. A Markov chain can be represented by an initial state, a state space, a transition matrix that describes the probability of each and every state to convert to any other state across the state space. It can be represented using the below equation

$$S(t+1) = P_{ij} \times S(t)$$

Where S(t+1) is the future state of the system, S(t) is the current state of the system and P_{ij} is the probability that land use type i will convert to land use type j. However Markov Chain can be used only to model temporal changes of a system, the spatial characteristics of a model cannot be simulated. Integrating Markov and the CA model can solve this problem.

4.2 Cellular Automata

Cellular automata (CA) is used in order to simulate spatially distributed phe- nomena and dynamics over time. It is made up of cells on a grid (usually a rectangle) that follow rules based on neighboring cells' states as they evolve through discrete time steps.

Each cell has a limited number of possible states, and predetermined rules are used to update each cell's state. Neighboring cells often in a local neighborhood have an impact on the status of a cell. At distinct time steps, cell states are updated simultaneously. CA models are distinguished by their local interac- tions and simplicity. From basic principles, CA may show complex emergent behavior and patterns. They have uses in many different disciplines, including physics, biology, and social sciences. CA have been used to the study of dy- namic system behavior, self-organization, and pattern generation. Boundary circumstances and starting settings might affect how CA behaves. CA provide a useful tool for comprehending and recreating complicated systems with the use of simple rules and neighborhood interactions.

Chapter 5

Proposed Methodology

There are four main steps in this study's framework:

1. Data collection and pre-processing
2. Classification into LULC land types
3. Future LULC prediction using integrated Markov-CA model

5.1 Data collection and Pre-processing

Many methods have been developed to use satellite imagery to track the size of urban areas, and remote sensing is a powerful tool for monitoring Land Use and Cover Change at all scales, from the local to the global.

For this study two kinds of data were used:

1. Geometrically corrected cloud-free multitemporal datasets of the mentioned study area were obtained from USGS (United States Geological

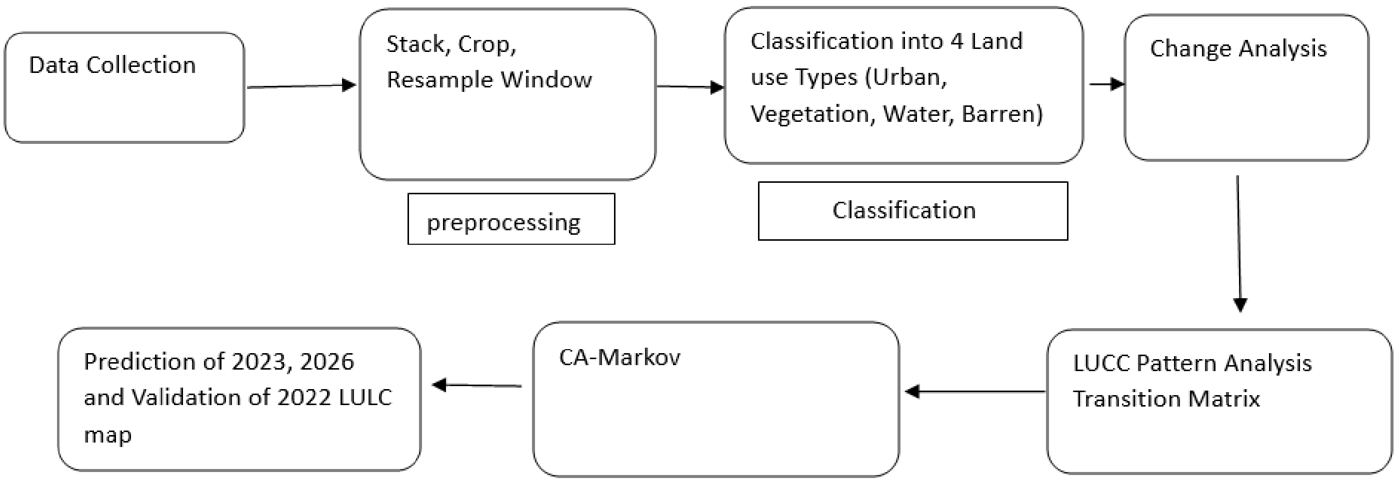

Figure 5.1: Methodology

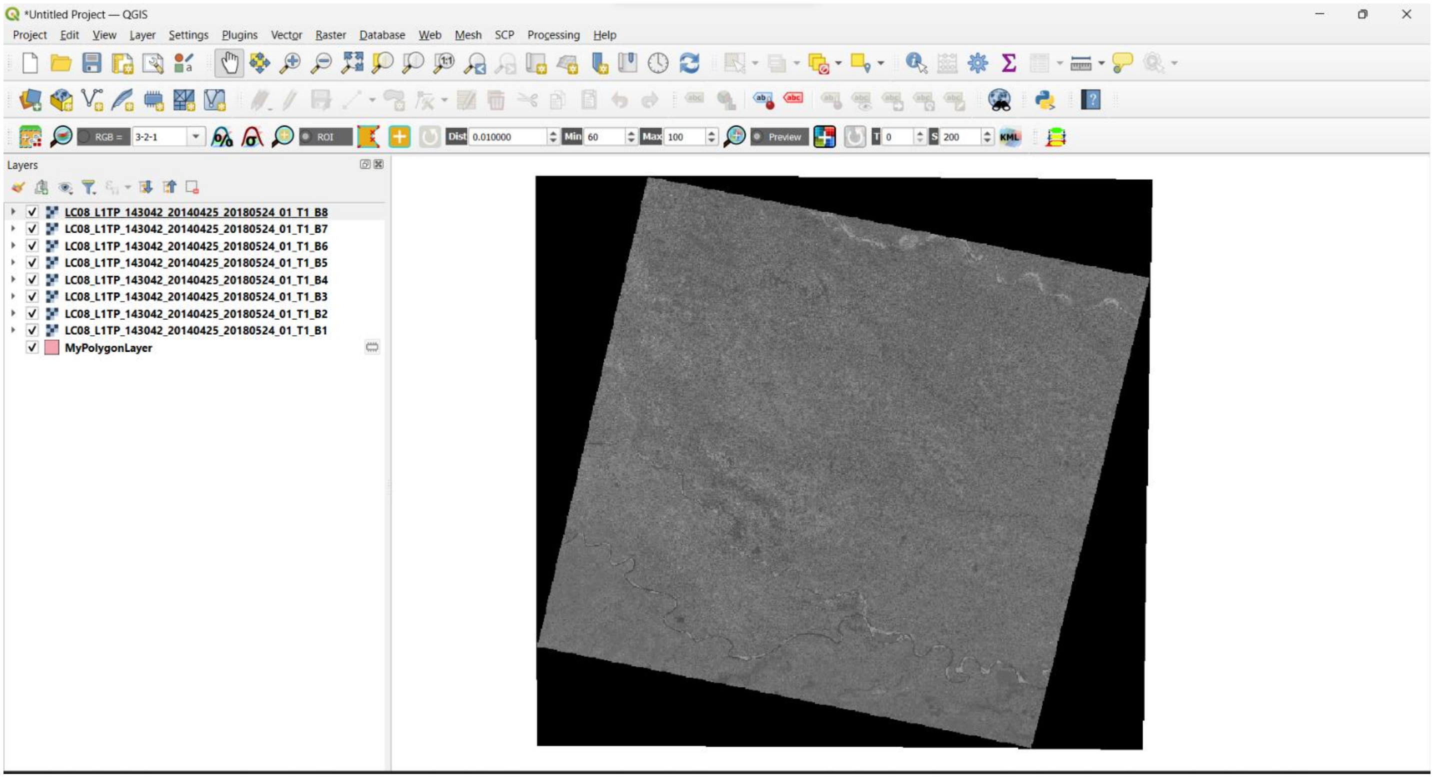

Figure 5.2: Prayagraj study area sourced from USGS

Survey). The study was done over Landsat 8 Level-2 satellite data which collects data using OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) sensors with a spatial resolution of 30x30 meters. The month of April was chosen specifically in order to avoid seasonal and cloud effects over the years 2013-2022. We are using 8 bands in this study. The obtained data were cropped to the required latitude-longitude extent and all the layers were stacked to obtain a composite image using SNAP software.

5.2 Classification into LULC land types

The different classes defined for this study can be referred from the table 5.1 Availing ground truth values manually by visiting the fields is out of the scope for this study. So instead, ground truth ROIs (Region of interest) were markedby examining false images created using various band combinations. The bandcombination Infrared (Band 5), Red (Band 4), Green (Band 3) is used to mark vegetation where plants are highlighted by dark green color due to the higherreflectance of Infrared radiation by plants.

Using the same band combination, barren land is also marked where it is high-lighted in brown. The band combination SWIR1 (Band 7), SWIR2 (Band 6), Red (Band 4) with a Bayesian filter is used to mark urban areas where the built up region is highlighted in purple. The same combination can be used to mark water and riverbed areas as they are highlighted in black and white

Table 5.1: Classes and their description

Class Name	Class Description
Built Up	Includes buildings, roads, and airports
Water	Includes river, lake, ponds
Barren	Includes harvested land, barren land Barren area surrounding rivers
Vegetation	Includes forest, trees, agricultural lands

Figure 5.4: Classification accuracy result of SVM for the year 2018

respectively. A minimum of 3000-4000 ROIs were marked for each class to give sufficient ground truth values for training the classification models.

Different classification techniques like SVM, Maximum Likelihood Classifier were performed on all the images and they were validated by separately creating test ROIs. Among the methods applied SVM was found to give better accuracy of around 95-98% for all.

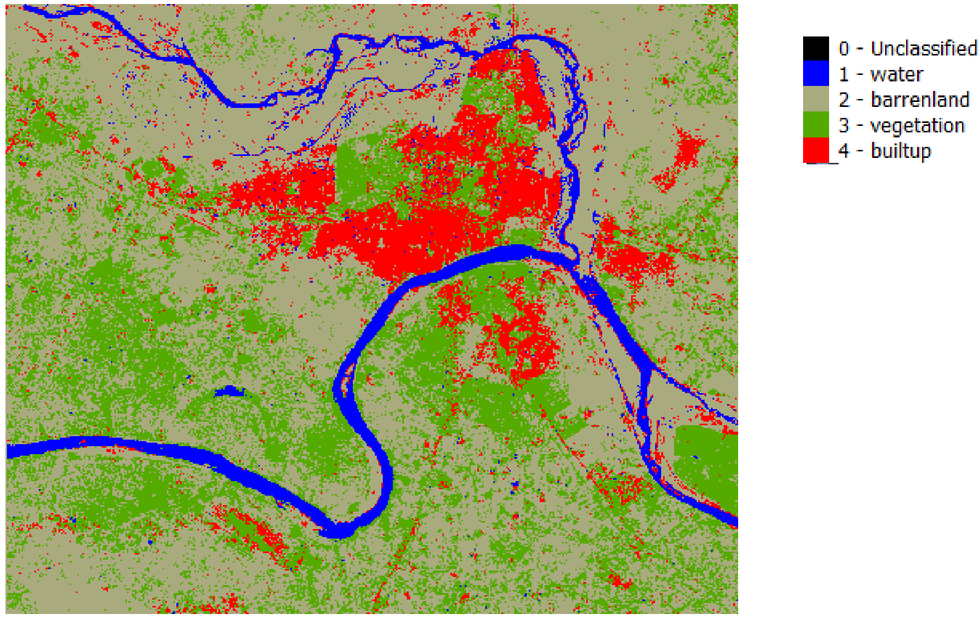

Figure 5.5: Classification Result for the year 2014 using SVM

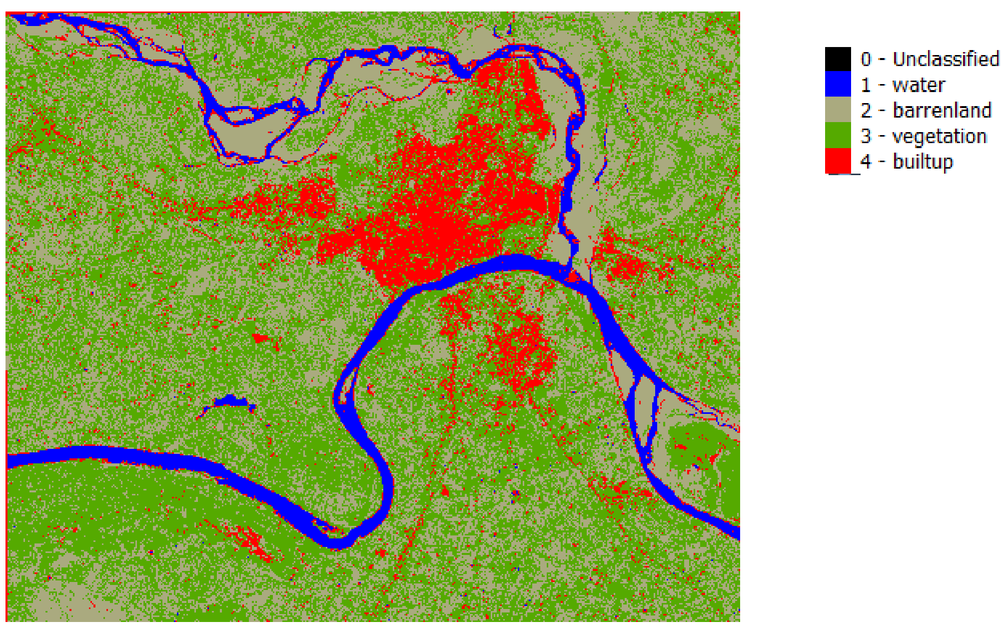

Figure 5.6: Classification Result for the year 2017 using SVM

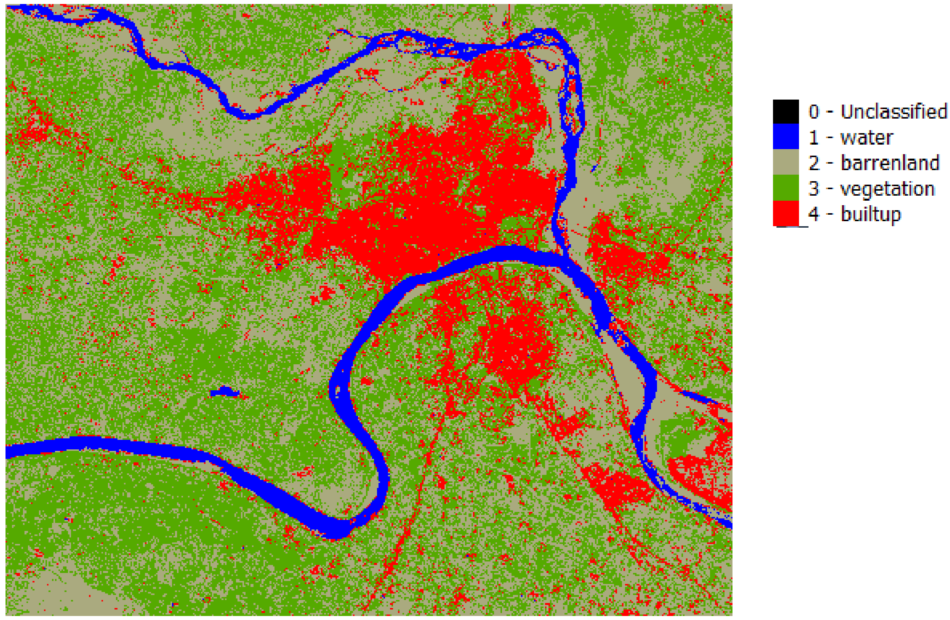

Figure 5.7: Classification Result for the year 2020 using SVM

5.3 Change Detection

Change detection is used to track how certain LULC classes have changed over time in terms of both their nature and their scope. The classified images were compared to estimate the net changes, the gains and losses in the LULC categories. The magnitude of change is a rate of expansion or reduction in the area of each class. A negative figure represents a loss in the region, while a positive figure represents a gain in the region.

5.4 Future LULC prediction using Integrated Markov-CA model

By integrating Logistic Regression model and Markov CA model we are able to explain the reasons behind Urban growth using logistic regression model and also predict the future LUCC state using CA-Markov Model. Combining CA and Markov can efficiently model both temporal and spatial characteristics of LUCC. Such a model can be defined by the following equation.

$$S_{ij}(t+1) = f(S_{ij}^{t}, Q_{ij}^{t}, V)$$

Where t and t+1 are current and next time step, S_{ij}^{t}, $S_{ij}(t+1)$ respectively are the current and next state of the cell in row i and column j, Q_{ij}^{t} are the neighbors of the current state at time t and V are the suitability factors. Transition probability matrix is obtained from the Transition Potential Panel in Idrisi using Markov Chain as the model evaluator (see Fig.5.9). The resulting matrix consists of the probability of each class changing into every other class. In the same panel, the LULC image of the year 2018 is taken as the initial land cover image and the obtained transition probability matrix is used to predict the LULC classes for the year 2022.This predicted image of 2022 is compared with the original 2022 LULC image for validation. Using the same transition matrix, LULC classes for the year 2026 were predicted using the 2018 LULC map as the initial image.

	1	2	3	4
1	0.662016	0.178798	0.043933	0.115254
2	0.023434	0.477441	0.457354	0.041771
3	0.003326	0.257786	0.657423	0.081464
4	0.026223	0.138285	0.304014	0.531478

Transition matrix

1. Water
2. Barren land
3. Vegetation
4. Built-up

Figure 5.9: Transition Probability matrix

5.5 Software Used

For Cropping and Stacking Layers:

QGIS

Plugin:

1. *SCP: (Semi-Automatic Classification) Plugin that allows for the supervised classification of remote sensing images, providing tools for the download, the preprocessing and postprocessing of images.*
2. *MOLUSCE: Plugin provides a set of algorithms for land use change simulations such as ANN, LR, WoE, MCE. There is also validation using kappa statistics.*

Chapter 6

Results

6.1 LULC Change Analysis

Observing the LULC change between the years 2014 and 2017, we can infer that about 9.65% of Land underwent changes out of 290.12 sq.km total area. Among the five land use types, Urban, Riverbed, Barren increased by 11.7 sq.km, 9.71 sq.km and 22.21 sq.km respectively whereas Water and Vegetation decreased by 10.65 sq.km and 33 sq.km respectively typical of urbanization. The same results can also be observed from Fig.6.2. Detailed gains and losses in each land use type can be observed from Fig.6.3. Transition between each class to every other class can be observed from Fig.6.4. Transition between every class to urban specifically is also highlighted in Fig.6.5. The complex pattern of LULC change can be represented using a best fit polynomial of the third degree. Using the software, trends up to 9th degree can also be calculated but the time taken for the process increases significantly. Generally, a polynomial of 3rd degree is representative of the data and gives a broad overview. The result of this analysis can be found in Fig.6.6.

Class statistics

	Class color	2014	2017	Δ	2014 %	2017 %	Δ %
1	Water	39926700.00 sq. metre	40073400.00 sq. metre	146700.00 sq. metre	4.99083126146	5.00916873854	0.0183374770782
2	Barrenland	456021900.00 sq. metre	293100300.00 sq. metre	-162921600.00 sq. metre	57.0026662467	36.6374917031	-20.3651745435
3	Vegetation	219177900.00 sq. metre	380213100.00 sq. metre	161035200.00 sq. metre	27.3972032535	47.5265780918	20.1293748383
4	Builtup	84874500.00 sq. metre	86614200.00 sq. metre	1739700.00 sq. metre	10.6092992384	10.8267614665	0.217462228172

Figure 6.1: Change between all classes in square meters.

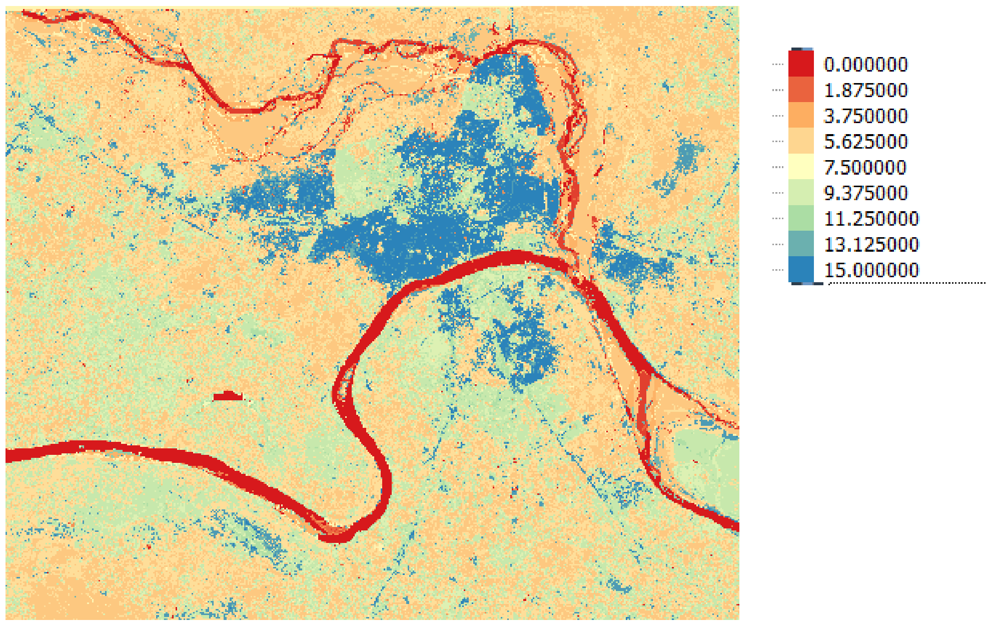

Figure 6.4: Transition between each class to all other classes

6.2 Model Validation

Model validation is a crucial step since it assesses a model's predictive capabilities and reliability of results.This can help us understand the precise depiction of the truth as we are comparing it with the real time results.This helps us gain confidence in model to apprehend the spatial and temporal dynamics of the landscape map. It's substantial to identify the uncertainty in the model and improve the model, as this has many real world use cases. We have used three way cross tabulation for validating the LULC map of year 2022.

6.3 Future Prediction

Historical rates of change and transition potentials were utilized to forecast the situation for a certain time. Future predictions can be used for policy making, growth impact on environment, weather forecasting, resource management and useful for data driven decision making. They make it possible to plan ahead, improve sustainability, and foster adaptability to potential changes in landscape.

For predicting the future land use for 2026, we used previous rates of change and transition potentials. Extracted from the year 2022 and predicted using the CA Markov model for the year 2026. From the predicted map of 2026, we can infer that the Built Up area will increase notably near the Airport region of the Prayagraj. A few patches of vegetation have become barren areas. The decline in the water region is expected to convert into riverbed. The urban area is expected to increase extensively in 2026 is explained by this predicted drop in vegetation and barren land.

On the west side of Prayagraj, the vegetation is expected to increase more than the other parts.Built up area near the Airport region, near Naini region, Civil lines and Jhusi will increase.Vegetation near the Civil Lines region have more odds to decrease.

Chapter 7

Conclusions and scope for future work

The research illustrates that by combining remote sensing, Geographic Information System (GIS), and Logistic Markov-Cellular Automata (LCM) with multi-temporal data, we can enhance our comprehension and evaluation of Land Use and Land Cover (LULC) dynamics. This integrated approach also enables us to predict future patterns and trends in urban expansion.

In this study, we analyzed the LULC pattern for the years 2013 to 2022 in Prayagraj region.We mainly focussed on the years 2014,2018 and 2022.We have analyzed the driving mechanism based on multiple logistic regression model.We took proximity factors into consideration, and predicted the future land use and land cover conditions for the year 2026 using CA-Markov model.
We can conclude that Distance from buildings and Distance from important Roads play an important role for the development of an area, whereas Distance from railways and water shows a negative impact. Concentration of built up regions is expected to increase as we go away from rivers, and the results verify the same. Coming to the Future works, we have considered only proximity factors for predicting urban growth. We can consider more driving factors like Natural factors, Socio-economic factors and Legal Factors.This method could give a holistic result.

Chapter 8

References

[1] Xuecao Li, Peng Gong, Urban growth models: progress and perspective,Science Bulletin, Volume 61, Issue 21, 2016,Pages 1637-1650,ISSN 2095-9273

[2] Wang L, Li CC, Ying Q et al (2012) China’s urban expansion from 1990 to 2010 determined with satellite remote sensing. Chin Sci Bull 57:2802–2812

[3] A Achmad et al 2018 IOP Conf. Ser.: Earth Environ. Sci. 126 012166 DOI 10.1088/1755-1315/126/1/012166

[4] Xu, Linfeng, Xuan Liu, De Tong, Zhixin Liu, Lirong Yin, and Wenfeng Zheng. 2022. ”Forecasting Urban Land Use Change Based on Cellular Automata and the PLUS Model” Land 11, no. 5: 652. https://doi.org/10.3390/land11050652

[5] Alsharif, A.A.A., Pradhan, B., 2013. Urban Sprawl Analysis of Tripoli Metropolitan City (Libya) Using Remote Sensing Data and Multivariate Logistic Regression Model. J. Indian Soc. Remote Sens. 42, 149–163. doi:10.1007/s12524-013-0299-7

[6] Kantakumar, Lakshmi Kumar, Shamita Schneider, Karl. (2017). IDENTI-