Journal of Information Systems Engineering and Management
2025, 10(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Predicting Urban Expansion using Cellular Automata and Machine Learning: A Multi-Model Evaluation Framework

Ranu Lal Chouhan[1], Hardayal Singh Shekhawat[2]

[1,2]Engineering College Bikaner

Corresponding Author: Ranu Lal Chouhan

Email ID: chouhan.ranu@gmail.com[1], shekhawat.hardayal@gmail.com [2]

ORCiD: https://orcid.org/0009-0009-7133-9478[1], https://orcid.org/0009-0008-3127-8925[2]

ARTICLE INFO

Received: 16 Oct 2024

Revised: 07 Nov 2024

Accepted: 14 Nov 2024

ABSTRACT

Rapid urbanization has led to a significant land use land cover (LULC) evolution necessitating the robust predictive framework to guide sustainable urban planning. This study presents a comparative analysis of urban growth modeling using a hybrid cellular automata (CA) approach integrated with multiple machine learning techniques including random forest (RF), support vector machine (SVM) and Extreme Gradient Boosting (XGBoost). A tier-1 Indian city was selected as the study area (Jaipur city). Ancillary factors such as proximity to roads, rivers, elevation, slope and population density were incorporated to enhance the transition potential modeling. Model performance was evaluated using metrics such as Kappa coefficient, overall accuracy and F1 score. The results indicated that the CA-RF model outperformed the others in terms of both spatial realism and statistical accuracy, followed closed by CA-XGBoost. Spatial comparison of predicted outputs also revealed variations in growth direction, pattern compactness and built-up density. This research underscores the significance of combining CA with data-driven ML approaches for capturing complex urban dynamics and provides a decision-support framework for urban planners and policymakers.

Keywords: Machine Learning, Urban growth prediction, LULC, Urban planning.

1. INTRODUCTION

Urbanization, as a defining phenomenon of the 21st century, is reshaping landscapes across the world especially in countries like India. As the cities consistently approach towards expansion, the accommodation of economic as well as population domain continues to raise alarming concerns (Kodihal & Akhtar, 2023). This in turn increases the pressure on infrastructure, environment as well as natural resources. Prediction of spatial dynamics for the future creates a primitive baseline for a critical area of research in urban planning as well as geoinformatics (Luo et al., 2008). Accurate urban growth prediction can inform better land management decisions and also assist in mitigation of the environmental impacts. The convergence of geospatial technologies and machine learning algorithms has opened up new-spotlighted avenues for modeling urban expansion with high precision (Siddiqui et al., 2018).

Jaipur, the capital city of Rajasthan has undergone significant spatial transformation in the recent years. As a major economic as well as cultural hub, the city has experienced an incremented infrastructural development along with population influx and commercial activities (Jawaid et al., 2017). Residential expansion, industrial hotspot emergence, coupled with new urban-transport corridors has revamped the entire urban landscape of Jaipur city. The schematic representation of Jaipur city is shown in figure 1.

Journal of Information Systems Engineering and Management
2025, 10(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

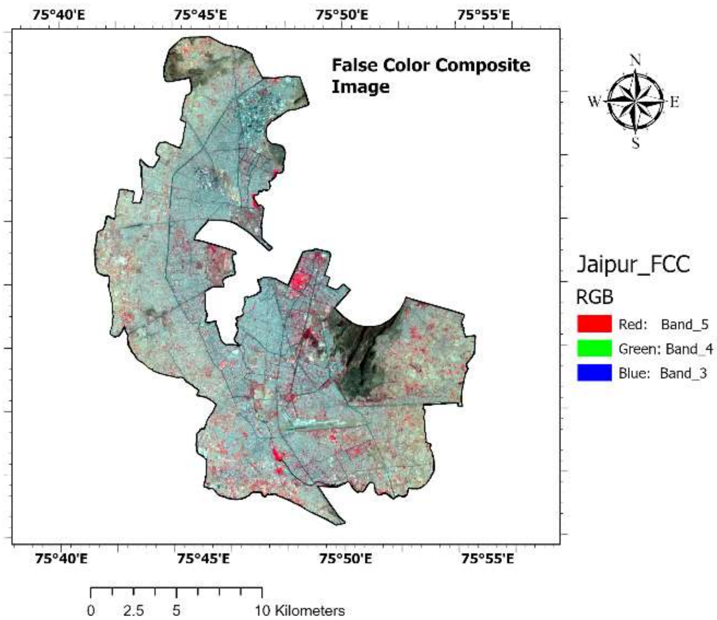


Fig 1: False color composite (fcc) image of Jaipur city

Analyzing the past-present combination of LULC patterns of Jaipur definitely provides valuable insights into the trends in urbanization. The period from 2011-2021 has witnessed substantial variations in built-up areas with agricultural lands and open spaces being transformed into residential-commercial zones (Deo et al., 2024). Through the utilization of CA models, this study aims to project Jaipur's urban sprawl/expansion by 2031 thus, offering an effective scientific basis for sustainable development strategies (Cheng & Masser, 2004).

Several modeling approaches been employed in past studies to simulate and predict growth (urban dynamics) ranging from a plethora of statistical regression approaches to cellular automata (CA) (Vliet et al., 2009). Furthermore, ANNs have also been employed but still continues to steadily evolve. Among all the algorithms, CA approach has particularly proven to be a effective go-to technique for such simulations and spectrums. CA-based models consider the influence of neighborhood cells to replicate the real-world urban expansion behavior. However, a considerate limitation comes in the form of dependency on expert-defined transition rules which is suppressed by turning towards hybrid machine learning modeling approaches. In simpler terms, fusing CA with ML algorithms for transition modeling. Despite the increasing adoption of CA-ML hybrid models, there exists a strong void-scarce gap of comparative studies evaluating the scopes of multiple ML algorithms within a CA-based foundational framework (Gao & O'Neill, 2020).

In this study, the paramount objective comes in the form of comparative framework analysis for urban growth prediction for the city of Jaipur using CA coupled with three popular ML algorithms: SVM, RF and XGBoost. Landsat-based LULC classification from 2011-2021 were used as inputs with 2031 as the prediction target. The manuscript continues to contribute to the literature by offering a comparative analytical evaluation of different ML techniques for precise urban growth modeling in the spatial environment.

2. METHODOLOGICAL FRAMEWORK

The methodology integrated in this study involves a multi-stage process comprising of data collection and pre-processing, transition potential modeling using ML, urban growth through CA and performance evaluation of the hybrid models.

Journal of Information Systems Engineering and Management
2025, 10(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

2.1 Data Preparation and Pre-processing:

Landsat imageries of 2011 and 2021 were taken into account for generating the LULC layers. Supervised classification using SVM was applied on Landsat-8 OLI dataset which yielded four main classes: built-up, barren land, water and green area. These classified layers were reprojected to a common UTM zone and were further resampled into 30 m spatial resolution baseline. The LULC layers from 2011 (base layer) and 2021 (reference layer) were then integrated efficiently to generate transition matrices and calculate the change statistics.

2.2 Ancillary Data for Transition Potential Modeling:

To enhance the transition modeling, multiple spatial drivers influencing urban growth were considered. These include distance to roads, rivers, existing built-up, DEMs, slopes and population density. All these spatial drivers were normalized between the upper and lower caps (0 and 1) to ensure having a min-max normalization. Each cell in the raster is treated as an instance with class label "change/transition to built-up" (1) and/or "no change" (0), based on observed transitions from the years 2011 – 2021. This binary classification dataset was used to train machine learning models.

2.3 Transition Potential Modeling using Machine Learning and CA-based Prediction

Random Forest (RF) is an ensemble learning technique constructing a set of multi-decision trees during training and outputs the mean prediction/outputs of the individual trees. Furthermore, Support Vector Machine (SVM) constructs a hyper-plane in a high-dimensional space to classify data points. A radial basis function (RBF) kernel was also utilized to handle the non-linear separation (Ding & Wu, 2024). XGBoost on the other hand is a gradient boosting technique that combines multiple weak learners in an addictive manner. It in turn minimizes a regularized objective function.

The transition probability surfaces were integrated with a cellular automata (CA) model to simulate the spatial allocations of urban growth. The CA model considers both the transition potential of a cell as well as its neighborhood influence (Aburas et al., 2019). The model takes into account neighbor cell, its corresponding built-up cell and the surrounding number of neighboring cells (typically 8). The model then simulates the urban growth in iterations until the predicted built-up area matches the expected built-up growth from the historical trends and/or regression-based approaches (Tripathy & Kumar, 2019).

- Model Validation and Accuracy Assessment:

Predicted LULC maps for 2021 were compared with the actual classified maps to assess the -accuracy. The metrics utilized were "kappa coefficient", "F1-score" and "Overall accuracy". The models/metrics for quantitative comparison of CA-SVM, CA-Rf and CA-XGBoost.

3. RESULTS AND DISCUSSION

The overall potential of a technique comes in the form of high accuracy-precision along with validated simulations for the adapted technique (Kim et al., 2022). Transition spatial map shows the efficacy of the preferred technique based on which the prediction model gets executed/chosen.

3.1Accuracy Assessment and Comparative Metrics

CA-SVM recorded the highest kappa-coefficient and F1-score, indicating a better match with actual built-up areas (Wahyudi & Liu, 2016). CA-RF had the lowest quantity disagreement but at the expense of spatial accuracy. Furthermore, CA-XGBoost showed a balanced performance but suffered from slightly longer execution times (computational processing time) as well as the moderate allocation issues. Table 1 shows the comparative metrics assessment.

Journal of Information Systems Engineering and Management
2025, 10(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

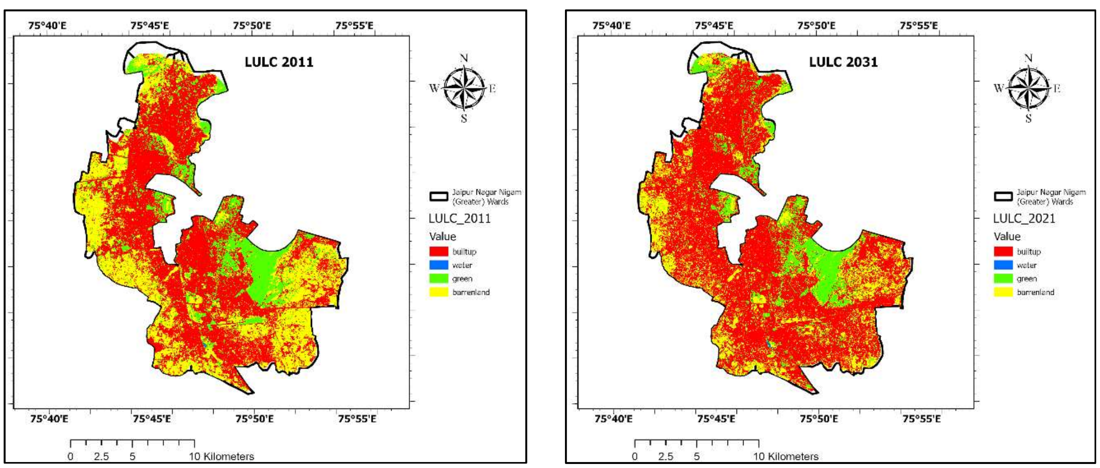
Table 1: Comparative metrics assessment

Metric	CA-SVM	CA-RF	CA-XGBoost
Overall accuracy (OA)	**89.3**	85.6	86.2
Kappa coefficient (k)	**0.842**	0.791	0.802
F1 score (built-up)	**0.901**	0.862	0.877
Quantity disagreement (%)	5.1	**4.3**	4.8
Allocation disagreement (%)	**3.8**	5.7	5.2
Processing time (mins)	14	**11**	17

The CA-SVM model successfully captured the radial growth pattern observed in Jaipur over the last decade. High-fidelity predictions were witnessed along Ajmer Road, Tonk road and emerging growth centres near the central business hubs of the city. Furthermore, CA-RF and CA-XGBoost tended to misclassify small patches/voids of agricultural land as built-up, specifically near the vicinities of water bodies and low-elevation regions.

3.2 Spatial Portrayal and Landscape Metrics Comparison

The classified LULC maps for 2011 and 2021 (reference layers) are portrayed in figure 2.



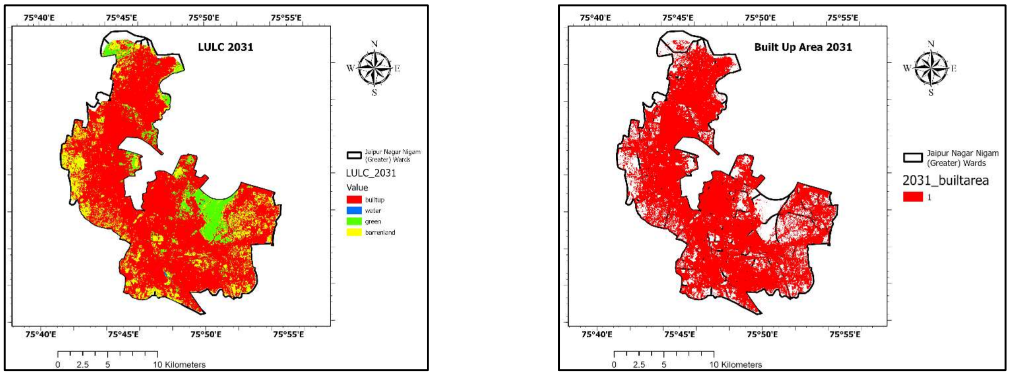
(a) (b)

Fig 2: LULC (a) 2011; (b) 2021

A clear visualization difference is already visible and evident where the 2021 map portrays more red pixels (built-up increase) as compared to the prevalence of them in 2011 map. This portrays synchronous coherence with the ongoing uncertain urban sprawl specially confining towards the built-up sprawl getting worse year-by-year. Based on the aforementioned reference layers, the LULC for 2031 was predicted as shown in figure 3. Also, the corresponding built-up dynamics is also shown to portray the prevalence of the built-up areas in the predicted and/or projected land use land cover dynamics.

Journal of Information Systems Engineering and Management
2025, 10(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



(a) (b)

Fig 3: (a) Predicted/Projected LULC 2031; (b) Built-up dynamics 2031

The main emphasis comes in the form of LULC transition, where the dynamics of class-to-class conversion is indicated. Figure 4 portrays the LULC transition dynamics for the year 2031 projected through CA-SVM approach.

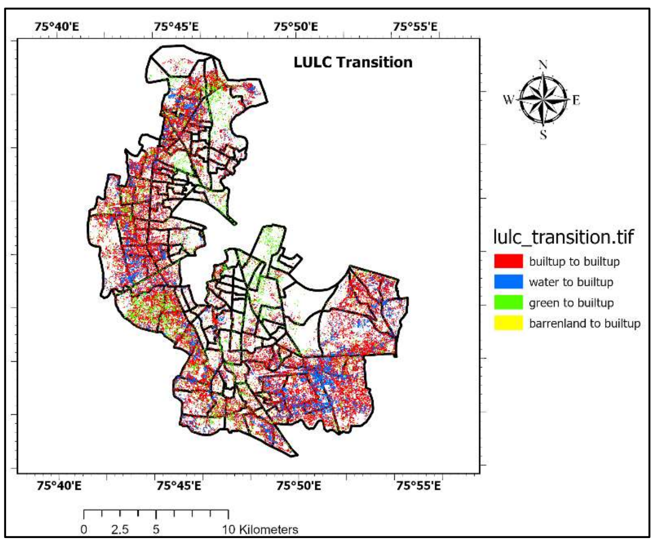


Fig 4: LULC transition dynamics for 2031

As evident from figure 4 layer, built-up to built-up (vertical growth) shows maximum accountability followed by barrenland/green areas to built-up (lateral growth). This portrays a sense of mixed-cluster urban growth where the overall significance needs to be intervened in the form of preserving the green and barrenlands where maximum policy implications can be executed. Table 2 shows the landscape metrics.

Table 2: Landscape metrics comparison

Metric	CA-SVM	CA-RF	CA-XGBoost
Patch Density (PD)	**3.4**	5.1	4.7
Mean Patch Size (MPS)	**4.72ha**	3.18 ha	3.89 ha
Aggregation Index (AI)	**82.7**	74.2	77.5

The assessment of landscape metrics was done by key metrics like patch density, mean patch size and the aggregation index. CA-SVM model showed the best optimum performance yielding more aggregated and less fragmented growth aligning with the observed patterns of densification. Furthermore, the RF model predicted higher fragmentation and thus, overestimating the built-up sprawl.

3.3 Why CA-SVM outperformed the other models?

The CA-SVM model yielded the best performance (optimum and efficient), however, it is to be deciphered on why it outperformed the likes of CA-RF and CA-XGBoost. The reasons are:

- Kerned-based flexibility of SVM allows it to model complex non-linear transitions far-better than the tree-based methods (RF) in urban environmental schematics.
- CA-SVM effectively and efficiently captured the urban growth gradients influenced by elevation, aspect, slope, proximity to arterial roads and historical urban cores.
- The smaller allocation disagreement suggests CA-SVM model to be more accurate for spatially unmixing the new built-up area pixels.
- Although CA-XGBoost has advanced regularization, it overfits in the medium-scale cities like Jaipur and thus, the input data in such cases is modest.

4. CONCLUSION AND FUTURE SCOPE

The study presents a comparative analysis of comparative analysis of machine learning based models for urban growth prediction integrated with cellular automata (CA). The overall aim is to simulate the spatial expansion of Jaipur city for the year 2031. Utilizing the classified LULC layers from 2011 and 2021, the models incorporated a variety of spatial and socio-environmental drivers like distance to roads, elevation, slope and aspects to estimate the transition potential for urban growth.

Three frameworks: CA-SVM, CA-RF and CA-XGBoost were taken into account for this and were evaluated. The CA-SVM approach emerged as the most effective and go-to technique outperforming the others in terms of overall accuracy (89%), kappa coefficient (0.842) and F1-score (0.901). Furthermore, it also portrayed low allocation disagreement and higher aggregation index unlike the rest two models which failed in this regard in a humongous scale. The findings not only advance the methodological framework for urban growth modeling but also offers practical implications for policy-makers as well as urban-planners. While the proposed approach yielded promising results, several enhancements can be explored in the future studies to further refine the methodology in terms of integrating socio-economic dynamics, higher-resolution data (hyperspectral) and deep learning integration (CNN).

REFERENCES

[1] Aburas, M. M., Ahamad, M. S. S., & Omar, N. Q. (2019). Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: A review. *Environmental Monitoring and Assessment*, *191*(4), 205. https://doi.org/10.1007/s10661-019-7330-6

[2] Cheng, J., & Masser, I. (2004). Understanding Spatial and Temporal Processes of Urban Growth: Cellular Automata Modelling. *Environment and Planning B: Planning and Design*, *31*(2), 167–194. https://doi.org/10.1068/b2975

[3] Deo, P., Siddiqui, M. A., Ramiz, M., Siddiqui, L., Naqvi, H. R., Shakeel, A., & Dwivedi, D. (2024). Measuring the spatial dynamics of urban sprawl in Jaipur City. *GeoJournal*, *89*(3), 108. https://doi.org/10.1007/s10708-024-11090-x

[4] Ding, Y., & Wu, K. (2024). USIR-Net: Sand-dust image restoration based on unsupervised learning. *Machine Vision and Applications*, *35*(3), 45. https://doi.org/10.1007/s00138-024-01528-0

[5] Gao, J., & O'Neill, B. C. (2020). Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nature Communications*, *11*(1), 2302. https://doi.org/10.1038/s41467-020-15788-7

[6] Jawaid, M. F., Sharma, M., Pipralia, S., & Kumar, A. (2017). City profile: Jaipur. *Cities*, *68*, 63–81. https://doi.org/10.1016/j.cities.2017.05.006

[7] Kim, Y., Safikhani, A., & Tepe, E. (2022). Machine learning application to spatio-temporal modeling of urban growth. *Computers, Environment and Urban Systems*, *94*, 101801. https://doi.org/10.1016/j.compenvurbsys.2022.101801

[8] Kodihal, S., & Akhtar, M. P. (2023). A sustainable approach to study and prediction of urban footprints of Jaipur and its growth direction using satellite based-data analysis. *Materials Today: Proceedings*, S2214785323033333. https://doi.org/10.1016/j.matpr.2023.05.693

[9] Luo, J., Yu, D., & Xin, M. (2008). Modeling Urban Growth Using GIS and Remote Sensing. *GIScience & Remote Sensing*, *45*(4), 426–442. https://doi.org/10.2747/1548-1603.45.4.426

[10] Siddiqui, A., Siddiqui, A., Maithani, S., Jha, A. K., Kumar, P., & 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 Science*, *21*(3), 229–236. https://doi.org/10.1016/j.ejrs.2017.11.006

[11] Tripathy, P., & Kumar, A. (2019). Monitoring and modelling spatio-temporal urban growth of Delhi using Cellular Automata and geoinformatics. *Cities*, *90*, 52–63. https://doi.org/10.1016/j.cities.2019.01.021

[12] Vliet, J. V., White, R., & Dragicevic, S. (2009). Modeling urban growth using a variable grid cellular automaton. *Computers, Environment and Urban Systems*, *33*(1), 35–43. https://doi.org/10.1016/j.compenvurbsys.2008.06.006

[13] Wahyudi, A., & Liu, Y. (2016). Cellular Automata for Urban Growth Modelling: A Review on Factors defining Transition Rules. *International Review for Spatial Planning and Sustainable Development*, *4*(2), 60–75. https://doi.org/10.14246/irspsd.4.2_60