Journals Pub

International Journal of
Town Planning and Management

ISSN: 2456-5253
Volume 12, Issue 1, 2026
January–June
DOI (Journal): 10.37628/IJTPM

https://journalspub.com/journal/ijtpm/

Research IJTPM

Integrating Cellular Automata for Predicting Urban Expansion: A Case Study of Bavdhan, Sus, and Mahalunge

Ar. Sakshi Mahadik[1,]*, Meera Shirolkar[2]

Abstract

Rapid Urban Expansion in the metropolitan cities like Pune, result in indiscriminate land use changes transforming natural landscapes. Such haphazard spatial development poses a risk to depletions of resources along with the disarray of the ecological system. A proactive approach to predicting and analyzing such changes is necessary in promoting the well-managed growth of cities. For the complex nature of urban growth, simulation method is used for analyzing spatial patterns, evaluating future scenarios, and aiding urban planning decisions. Bavdhan, Sus, and Mahalunge, which lie towards the western side of Pune and are witnessing significant urbanization are selected for the study. This research investigates the use of Cellular Automata (CA) urban growth modeling for simulating and anticipating urban expansion by calibrating the model against previous development patterns and testing it using real-world data in the selected suburbs. Cellular Automata – CA integrates various urban growth parameters of spatial dynamics viz Land Use Classification, regression Transition Probability Matrices and spatial dependency factors to simulate urban sprawl process close to the reality. CA-based simulation is used in combination with GIS data, remote sensing satellite images and land use classification to Markov chains, machine learning and socio-economic methodologies for predictive urban modelling. Socio-economic and policy driven factors are incorporated that influence urban expansion to enhance prediction. The findings highlight expanding urban hotspots, infrastructure stress areas, and emerging environmental concerns. The outcomes will contribute to enhance stakeholders' understanding of dynamics of spatial & temporal urban development, facilitating data driven land-use planning that ensures harmony between environment and infrastructural development. These findings will offer urban planners and policymakers a data-driven framework for reducing unregulated growth and optimizing zoning restrictions.

Keywords: Cellular automata, environmental concerns, land use change, predictive modeling, urban expansion

***Author for Correspondence**
Ar. Sakshi Mahadik
E-mail: mahadiksakshinitin@gmail.com

[1]Research Scholar, Department of Architecture, Dr. Bhanuben Nanavati College of Architecture for Women, Pune, Maharashtra, India
Assistant Professor, Department of Architecture, Bharati Vidyapeeth (Deemed to be University), College of Architecture, Pune, Maharashtra, India
[2]Professor, Department of Architecture, Dr. Bhanuben Nanavati College of Architecture for Women, Pune, Maharashtra, India

Received Date: October 30, 2025
Accepted Date: March 03, 2026
Published Date: March 20, 2026

Citation: Ar. Sakshi Mahadik, Meera Shirolkar. Integrating Cellular Automata for Predicting Urban Expansion: A Case Study of Bavdhan, Sus, and Mahalunge. International Journal of Town Planning and Management. 2026; 12(1): 51–62p.

INTRODUCTION

The rapid expansion that we observe in cities is becoming a phenomenon that transcends geographical boundaries. The World Urbanization Prospects report of 2018 suggests that 68% of the world's population is expected to shift to urbanized cities by the year 2050. We also know that concentrated development comes with uplifted socio-economic indicators. Yet these developments have their own risk factors in the form of agricultural land, environmental sustainability, and infrastructural strain. The clashes of such opposing paradigms – development and sustainability – can be tackled with well managed growth that is accompanied with foresight into demographic and spatial trends which fundamentally serves as a guiding force (United Nations, 2018) [1].

Uncontrolled urban growth is a prime example of global environmental change and one of the foremost concerns of study which this research configures. We see the greatest amount of growth is happening in Global South's poorly regulated cities and regions which leads to severe alterations in both landscape and environment. Unchecked development often leads to the irreversible loss of forest land, open spaces, water bodies, and biodiversity (Sudhira et al., 2004) [2]. Hence predictive modelling can be effectively applied to check the development pattern of urban growth that could allow us to set boundaries within which we could work towards sustainability. It is necessary to analyze the growth patterns and possible scenarios to assist in effective planning at the required level. Predictive models of urban simulation allow the examination of possible environmental and infrastructural changes. Detecting previous patterns assists planners in defining the new changing constraints and urban processes to plan more strategically. Many models have been developed for this purpose, among which GIS integrated with CA tend to be popular (Clarke & Gaydos, 1998; Lu, H., Wang, R., Ye, R., & Fan, J. (2023) [3, 4].

CA models project land use changes depending on the distance to a particular infrastructure (e.g., roads, population density) and the age of the infrastructure. When used with fuzzy logic, the model can more accurately predict spatial urbanization when calibrated against known urban peripheral boundaries (Al Darwish et al., 2018) [5]. Nonetheless, some basic assumptions concerning the parameters used in the model require substantiation.

Therefore, in the present study, we employed remotely sensed satellite images obtained at Sus, Bavdhan, and Mahalunge as the study area. The model incorporated the Geographic Information System (GIS) to build spatial relationships between driving factors using various geospatial techniques to analyze the change in land use and land cover over time in the study area.

Pune's metropolitan region's spatial scope has expanded significantly over the last twenty years, especially towards the outer zones and fringes. These peri–urban areas have seen farmlands, forests, and hills converted to residential colonies, commercial centers, and infrastructural development. The phenomenon of urban expansion is characterized by low-density, undefined and often haphazard spread outgrowth occurred particularly after the year 2000. This is a consequence of available land, distance to IT development centers, ongoing infrastructure developments, and weak enforcement of restrictions (Ravetz et al., 2013) [6].

The model's accuracy is validated by running several statistical validation techniques to analyze each pixel of the simulated land use map, comparing it to an actual map. Upon validation, this model allows for growth simulation, thus spatial estimations for 2031 can be performed. These estimations are essential in ascertaining general anticipated patterns, aiding in decision-making as to how planners can effectively control and encourage growth in a less destructive, more sustainable manner.

STUDY AREA

Pune is located in the western section of the Maharashtra region of India and has seen rapid urban growth. The city's most recent population estimate, as of 2011, is a little over 3.1 million, but this figure is likely closer to 7.4 million when considering metropolitan area population. The Pune Metropolitan Region covers a total area of 7,256 km². Further, the Pune Municipal Corporation encompasses approximately 516.18 km².

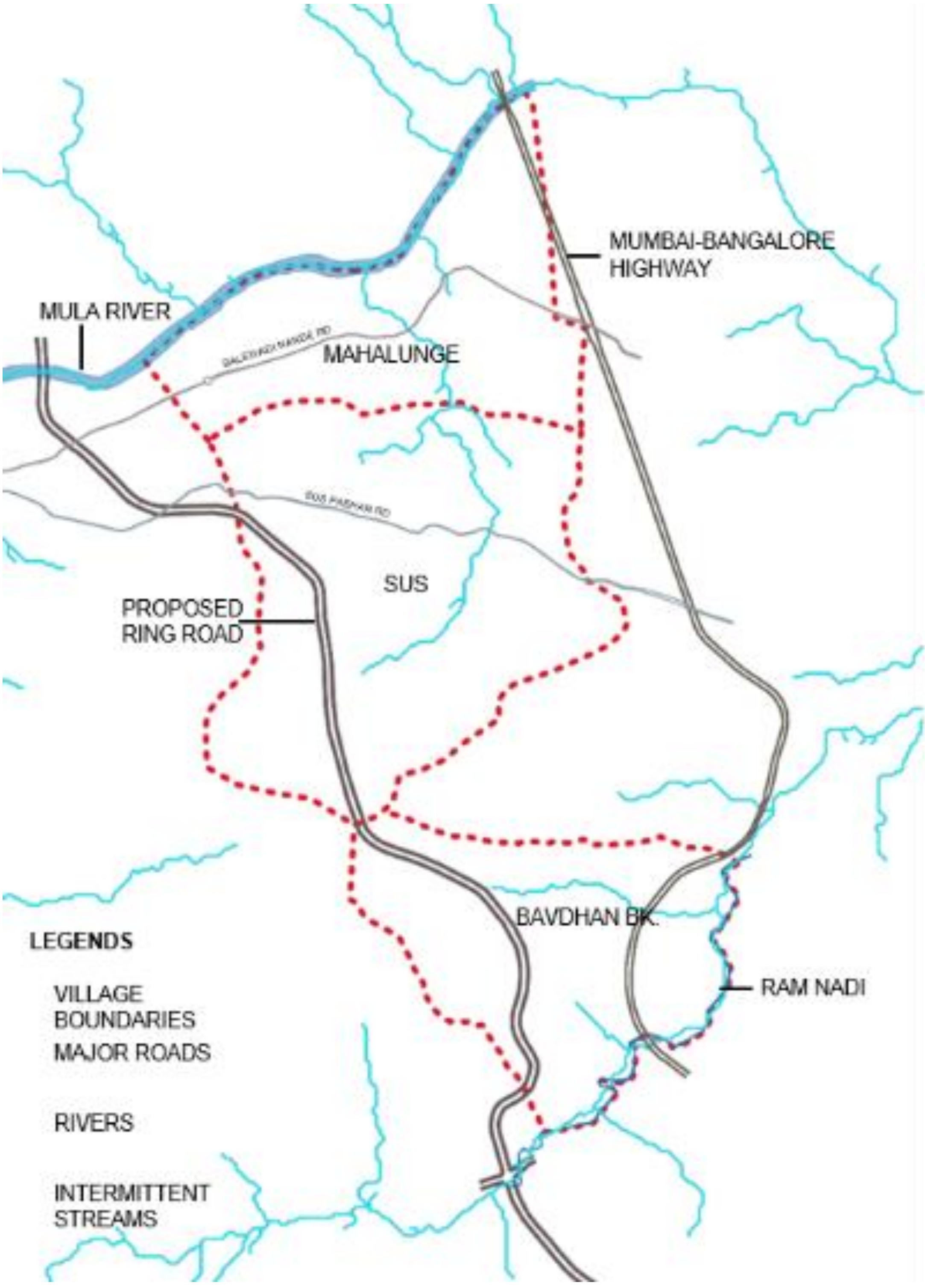
Pune is located on the confluence of Mula and Mutha rivers and surrounded by the Sahyadri Hills (Western Ghats). The city is known for being the educational and cultural hotspot of the Indian state of Maharashtra. This region has transformed rapidly over the past two decades, associated with greatly escalated activity in the real estate, IT, and automotive sectors (Siva Ramakrishnan, 2011) [7]. Its strategic location near Mumbai, coupled with favorable climate, quality education, and increasing job opportunities, has attracted significant inward migration.

Focus Areas: Bavdhan, Sus, and Mahalunge

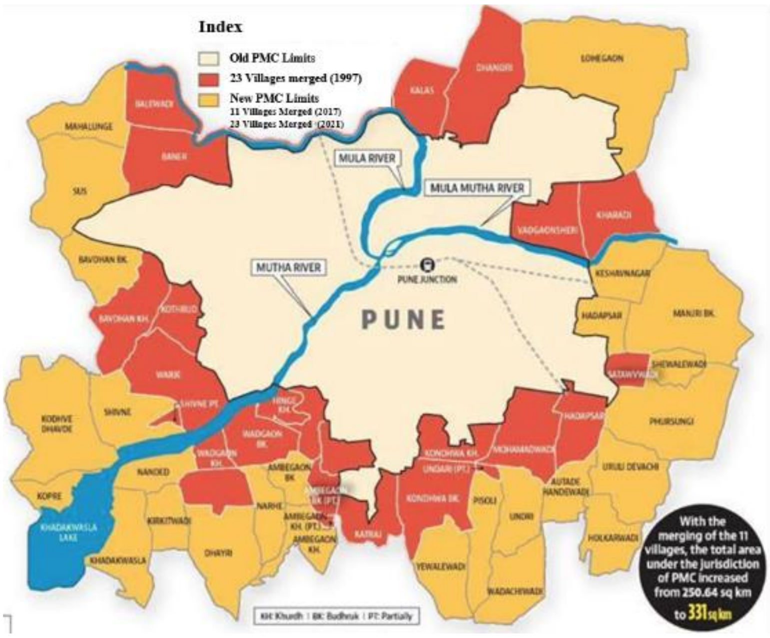
The suburbs of Bavdhan, Sus and Mahalunge are the focus region for this study. Positioned on the western periphery of Pune, they fall between the latitudes of 18.50°N to 18.55°N and longitudes of 73.75°E to 73.85°E. The geographic area of these three suburbs is roughly estimated at 58.12 km².

From the perspective of urban growth, these suburbs have become important zones of interest because of the accessibility and proximity to the following economic centers:

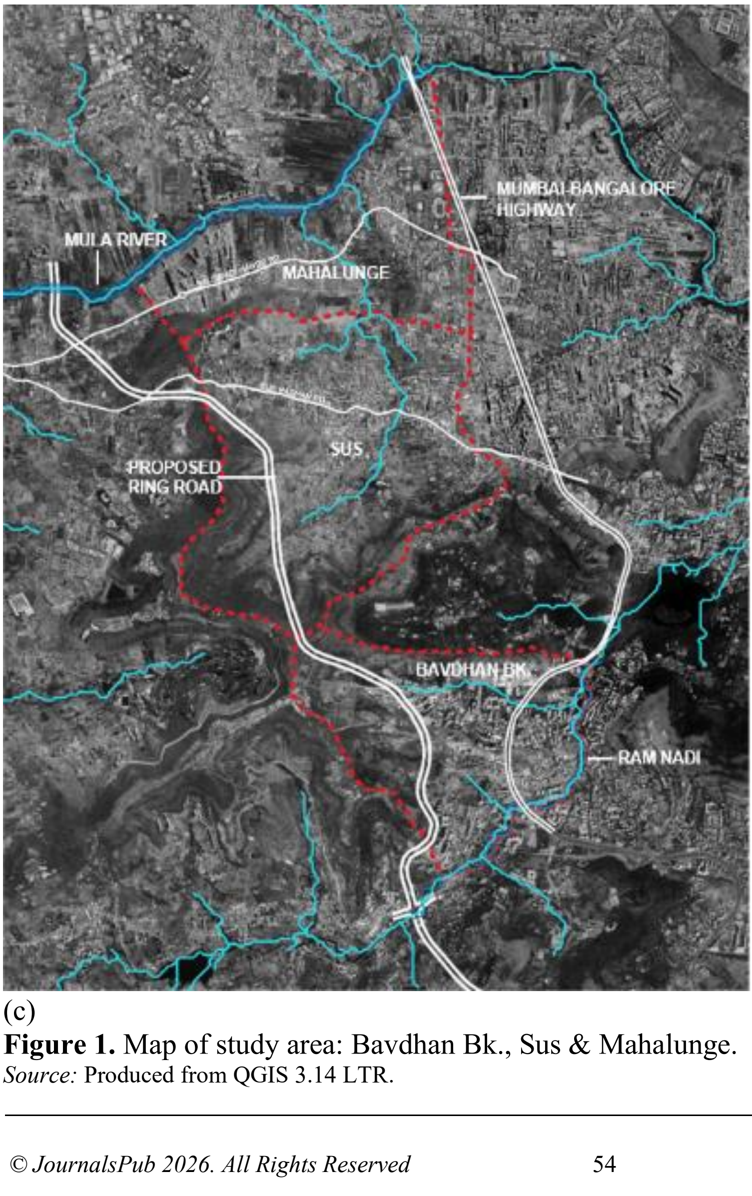
- Hinjewadi IT Park (India’s largest IT cluster in the region).
- Mumbai–Pune Expressway.
- Baner–Balewadi Sports and Commercial Corridor.
- Proposed Pune Metro Line 3 (Figure 1).



(a)



(b)



(c)

Figure 1. Map of study area: Bavdhan Bk., Sus & Mahalunge.
Source: Produced from QGIS 3.14 LTR.

Rapid urbanization is taking over Bavdhan as high-rise buildings, schools, and shopping complexes are coming up along the Mumbai–Bangalore Highway or NH-48. Sus earlier was a farming village but now accommodates gated societies along with private schools for Pune's professional class. Town planning schemes along with the Smart City projects are currently underway in Mahalunge which is positioned beside the Mula River. These projects have also led to the integration of the area into the PMRDA's town planning zones (Figure 2).

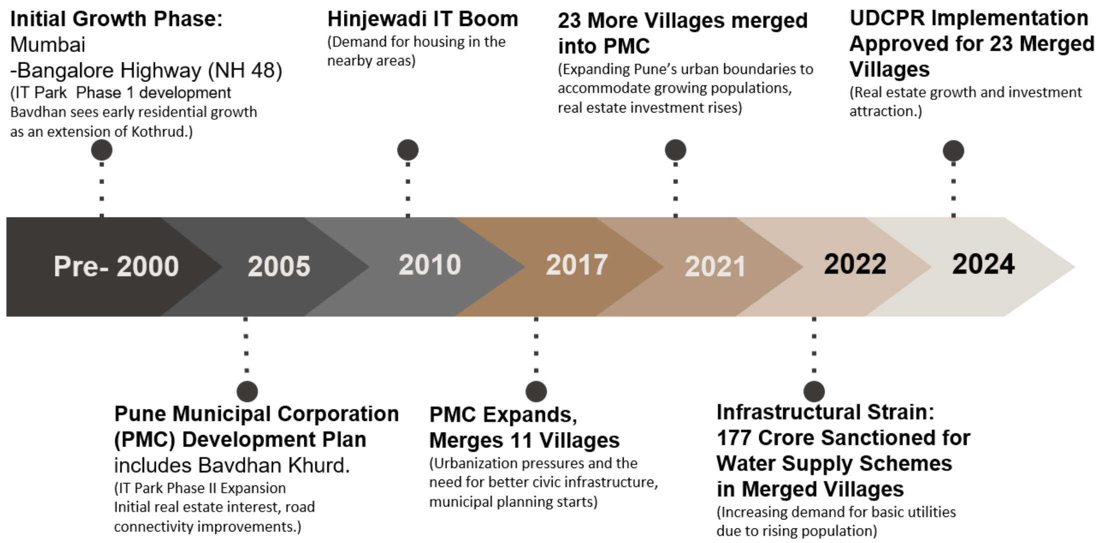


Figure 2. Study area growth over the period.
Source: https://www.hindustantimes.com/cities/punenews/state-govt-applies-udcpr-for-23-merged-villages-101710873151547.html

Bavdhan, along with the rest of the area, witnessed an initial boost in development pre-2000. It was planned as a residential area to ease the pressure on Kothrud. The IT boom in Hinjewadi from 2005–2010 made sure that Maluhunge and Sus received a lot needed development. Furthermore, investment in real estate planning surged post 2021 after unifying the villages with PMC along with the approval of the UDCPR policy.

The combination of grazing fields, agriculture, and pastoral farming has changed to mono-culture dominated by the real-estate industry, private schools and service-based occupations. However, there are still small parts of the region where traditional professions, such as dairy farming and selling vegetables, continue to exist.

The combination of migratory population pressure, infrastructure growth, and the presence of developable land makes Bavdhan, Sus, and Mahalunge favorable for observing urban development trends. Their hills, riverbanks, and agricultural plains are habitats with an intricate blend of features that make them suitable subjects for studying the effects of urban sprawl's spatial transformation and ecological impact.

SIMULATION MODEL USED: CELLULAR AUTOMATA

Socioeconomic, environmental, and even policy-based variables create uncertainty, contributing towards the complex, spatial-temporal phenomenon known as urban growth. One such framework is Cellular Automata (CA), approach is put to use because of its ability to model dynamic, nonlinear systems and replicate self-organizing patterns observed in urban sprawl. CA-based simulation is used in combination with GIS data, remote sensing satellite images and land-use classification to Markov chains, machine learning and socio-economic methodologies for predictive urban modelling.

The function of CA models works by subdividing a geographic area into a grid of cells, with each cell switching "states" (for example, non-urban to urban) according to set rules and the activity of surrounding cells (White & Engelen, 1993) [8]. The various shifts are achieved through algorithms that use spatial driving factors like the distance to roads, population density, and current land use. Its effectiveness, however, is greatly enhanced when combined with Geographic Information Systems

(GIS), which offer the appropriate spatial data handling and analysis tools. As Batty et al. (1999) [9] remarked, the use of CA in the study of urban issues is “almost impossible without the data management capabilities of GIS”. The combination known as GIS and CA has proven to be one of the most efficient techniques currently used in forecasting urban sprawl, assessing planning models, and managing land use in an environmentally friendly manner (Clarke and Gaydos, 1998; Al-Darwish et al. 2018) [3, 5].

METHODOLOGY

The procedure for urban expansion modeling with Cellular Automata (CA) follows a systematic and iterative approach to ensure spatial precision and forecasting accuracy. Initially, satellite images for the years 2003, 2013, and 2023 are collected and processed through supervised classification to create LULC maps. These maps are trimmed to a uniform projection and resolution of 10m to ensure alignment among datasets. As a next step, spatial driver layers like proximity to roads, built-up areas, IT hubs, as well as gated communities, are transformed into Euclidean distance rasters and fuzzified through linear fuzzy normalization. Then, these fuzzy layers are incorporated through a weighted overlay technique to produce a Suitability Index Map which indicates the potential for various lands to be transitioned to urban use. Extracting areas most likely to be developed using a threshold value (for example, 0.695) further refines this map. This serves the remaining steps to construct the final transition probability surface. This is subsequently used to estimate the urban expansion for the year 2033, which is predicted by merging this surface with the current urban layer in a CA simulation model where state changes for each pixel through neighborhood interaction and defined transition rules. Lastly, model validation and calibration are performed using the Kappa coefficient.

Data Preparation

Part of the work associated with these models entails gathering pertinent satellite images and rastering them for analysis. For this, LULC maps (Land Use Land Cover) of the years 2003, 2013 and 2023 were created using Landsat and Sentinel 2 satellite data. While the 2003 and 2013 Landsat datasets were 30-meter resolution, the 2023 map was generated using 10-meter Sentinel-2 imagery. To maintain consistency with spatial analysis, all datasets were resampled to a common 10-meter resolution. These LULC maps were refined through supervised classification into four land cover types: Built-up (Yellow), Vegetation (Green), Barren Land (Red), and Water Bodies (Blue). They were integrated into successive analyses as the foundational layers for the macro-scale consideration of the region’s dynamic urban growth (Figure 3).

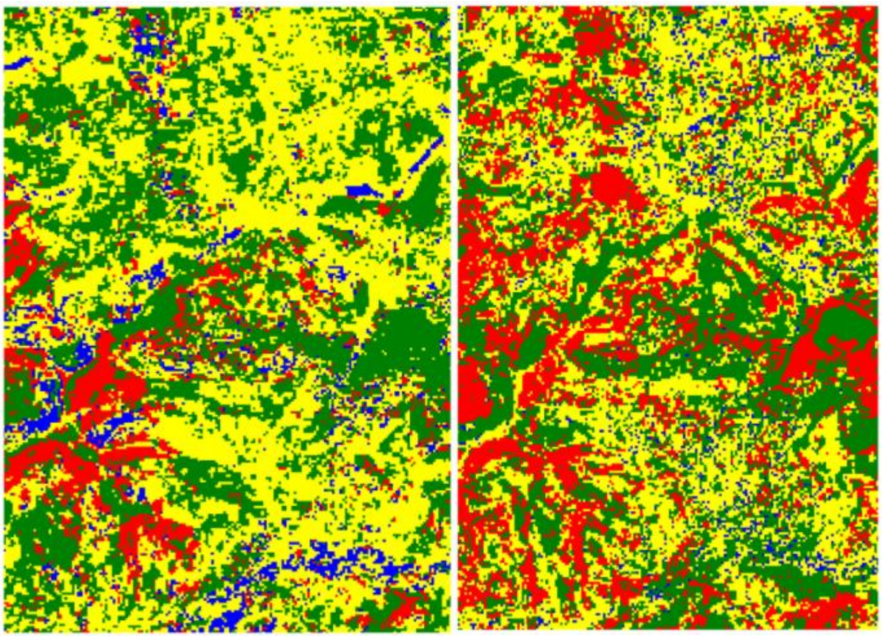


Figure 3. LULC maps. (1) Built-up, (2) Vegetation, (3) Barren land, and (4) Water bodies.
Source: Landsat surface reflectance imagery using Google Earth Engine (https://earthengine.google.com/).

Fuzzy Normalization of Spatial Drivers

In addition, for short-term urban growth trend analysis and calibration, a separate set of high-resolution classified layers for 2019 and 2023 were used. More specifically, built-up areas were generated from RGB-based LULC images for the years 2019 and 2023. The cross-examination of the pixelated binary images revealed an increase in urban pixels from 146,875 in 2019 to 164,295 in 2023 translating to a net increase of 17,420 pixels which approximates to 1,567.8 hectares. These results coupled alongside the previously calculated metrics from the trading economy models show urban expansion likely driven by regional economic development and population increase (Figure 4).

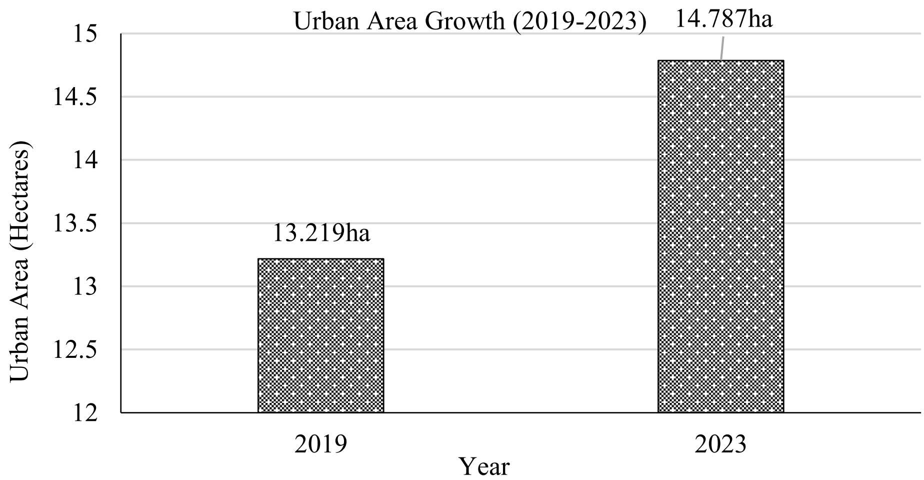


Figure 4. Urban Area Growth between 2019 and 2023.

To generate a transition potential surface suitable for simulation, the spatial driving factors were normalized using a linear fuzzy transformation. The fuzzy membership function employed was a linear decay function calculated using:

$$\text{Fuzzy Output} = 1 - \left(\frac{ED\ Value}{D_{max}}\right)$$

This converts raw Euclidean Distance (ED) rasters into standardized values between 0 and 1, where 1 indicates maximum suitability (closest proximity to the feature). The spatial drivers normalized include: distance to roads, distance to existing buildings, IT companies, educational institutions, and gated communities. These were rasterized and aligned to a uniform spatial extent and resolution of 10 meters (Figure 5).

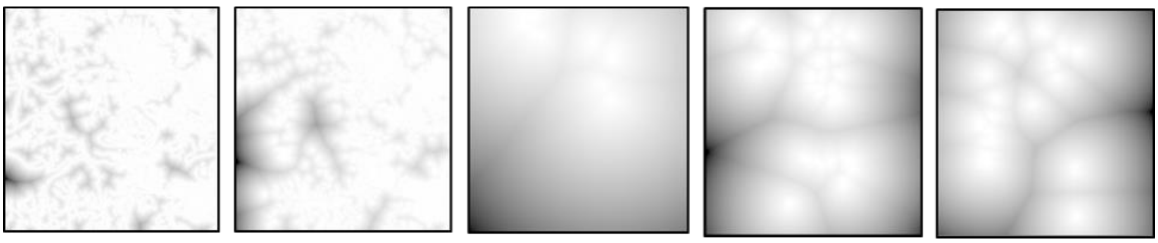


Figure 5. Fuzzy normalized spatial driver maps for urban growth modeling.
Note: Brighter areas represent higher suitability for urban growth.
Source: Processed by author using QGIS based fuzzy normalization using QGIS 3.14 LTR.

Weighted Overlay for Suitability Index

Following the procedure of normalizing each spatial driver using fuzzy linear decay logic, the Suitability Index raster was created using weighted overlay analysis. The included layers were distance rasters to: roads, existing built-up areas, IT companies, gated communities, and institutional proximity. These fuzzy layers were re-aligned and resampled to 10 m resolution to preserve spatial agreement (Figure 6).

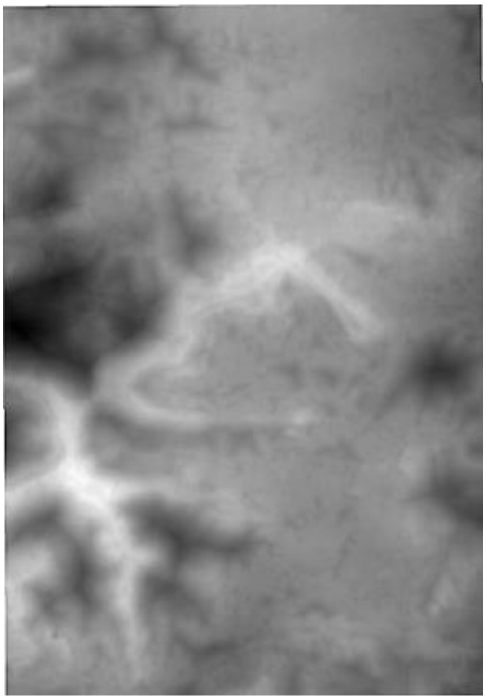


Figure 6. Weighted Suitability Index Map.
Note: Brighter areas represent higher overall suitability for urban growth.
Source: Generated by author in QGIS 3.14 LTR.

The final set of weights assigned to each layer was:

- Fuzzy Roads: 0.25.
- Fuzzy Buildings: 0.20.
- Fuzzy IT Companies: 0.20.
- Fuzzy Institutions: 0.15.
- Fuzzy Gated Communities: 0.20.
- The weighted sum equation used in the raster calculator was:

$$\text{Transition Potential} = \sum_{i=1}^{n} \text{Fuzzy Layer}_i \times \text{Weight}_i.$$

This raster output, subsequently named, was instrumental in pinpointing regions with maximum potential for future urban expansion.

Thresholding

Objective extraction of the regions most suited for urbanization was performed by applying a thresholding technique to the calculated Suitability Index (Figure 7). Statistically analyzing the rasters suggests several values between 0 and 1 (after normalization). Multiple threshold values were trialed, and the best performing found was 0.695. This proved close alignment with modelled expectations for 2033 while allowing for the spatial distribution consistency.

The raster calculator was used to apply the threshold condition:

"Suitability_Normalized@1" >= 0.695.

The output was a binary raster (Suitability_Threshold_0_695.tif) with 1 representing pixels suitable for urban development. This formed the basis for simulating future growth.

Figure 7. Suitability threshold map.
Note: White areas indicate zones classified as suitable for future urban development; black areas are unsuitable.
Source: Generated by author in QGIS 3.14 LTR.

Final Urban Prediction Map (2033)

To model urban sprawl for the year 2033, the suitability raster was thresholder and added to the 2023 binary urban raster. The 2023 raster (Urban2023BinaryUTMAligned.tif) had pixels with value 1 for urban and 0 for non-urban. Using raster calculator logic, a combination of the two was performed (Figure 8):

"Suitability_Threshold_0_695@1" + "Urban_2023_Binary_UTM_Aligned@1"

This resulted in three values in the output raster (Change_Map_2023_2033.tif):

- 0 = Still Non-Urban.
- 1 = Newly Urbanized.
- 11 = Remains Urban.

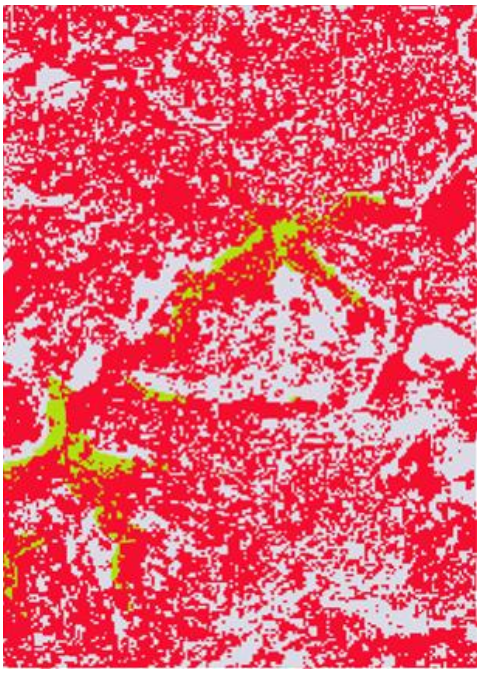


Figure 8. Predicted Urban Growth Map (2023–2033).
Source: Simulated by author using Cellular Automata with Fuzzy and Genetic Model in QGIS 3.14 LTR.

Table 1. Pixel count analysis.

Color	Class	Pixel count
	Still non-urban	147,318 pixels.
	Newly urbanized	10,816 pixels.
	Remains urban	270,114 pixels.

The total predicted urban area for 2033 is 2.94 km², which indicates a 1.61x increase compared to 2019 (Table 1).

Simulation Model Used

The study used a CA-based spatial simulation model within a GIS framework. In the CA model, the landscape is divided into a grid of cells, which are elucidated as spatial states (urban or non-urban). Transition rules were applied based on:

- Neighborhood influence (adjacent urban cells).
- Suitability index derived from spatial drivers.
- Historical trends in land use transitions.

Deviation from classic CA–Markov modeling arose due to use of fuzzy based spatial drivers rather than Markovian probabilities. This model was manually calibrated by matching historical trends from 2003 to 2013, projecting growth to 2023, and estimating growth for 2033.

This approach permitted the interpretability of each layer and the transparency of the simulation process in relation to time and the evolving patterns of urban growth.

Model Calibration and Validation

The 2013–2023 trends were used as spatial drivers within the simulation framework without employing automatic optimization algorithms. Using the 2023 LULC raster, the model was validated. Calibration involved adjusting fuzzy parameters and weights to replicate the defined growth pattern for the 2019–2023 period with sufficient accuracy (Table 2).

Table 2. Land use change (2023–2033).

Value	Class	Pixel count	Area (m²)
0	Still non-urban	147,318	14,114,343.22
1	Newly Urbanized	10,816	1,036,266.69
11	Remains Urban	270,114	25,879,265.98

A graph has been created within the workflow for comprehensive post-processing model validation employing statistical indices.

Kappa Coefficient which assesses agreement between actual and predicted 2023 maps. The model achieved a Kappa coefficient of 0.85, indicating almost perfect agreement between the predicted and reference land use data. According to Landis and Koch (1977) [10], Kappa values between 0.81 and 1.00 represent almost perfect agreement, validating the reliability of the simulation.

RESULTS

A spatial analysis conducted in the years 2019 and 2023 showed increased spatial use within the corridor of Bavdhan–Sus–Mahalunge. Through the means of pixel classification based on RGB color scheme (255, 255, 0 for built-up areas), it could be determined that the urban area increased from 146,875 pixels to 164,295 pixels from 2019 to 2023. This shows that there was an absolute gain of 17,420 pixels which is around 11.86% expansion of urbanization or 1,567.8 hectares of urbanization.

Following this, an iteration of CA-based simulation predicts further urban agglomeration up through 2033. The projection for the urban area by 2033 stands at 2.94 km² which is roughly 1.61 times the urban footprint reported in 2019, 1.18 km². The pattern produced seems to match the observed development speeds, particularly those factors near high-accessibility zones.

The IT hubs, gated communities, roads, buildings, and others through which the urban suitability surface are created using a weighted overlay of fuzzy normalized spatial drivers demonstrate the possibility of future growth. High suitability scores cluster around:

- Primary roads and arterial highways (NH-48).
- Proximity to Hinjewadi IT Park.
- Gated residential societies and institutional zones.

The change detection map between 2023 and 2033 classified the predicted land cover transitions as follows:

DISCUSSIONS

The results that were obtained showing urban growth is not a spontaneous process but is rather heavily governed by accessibility, current infrastructure proprietary zones, and center of economies proves the efficiency of the fuzzy spatial drivers' simulation along CA approach for suburban regions like Bavdhan, Sus, and Mahalunge.

Spatial Trends Observed

- Urban expansion tends to follow transportation corridors, particularly near NH-48 and the upcoming metro corridors.
- Flat terrains and areas with moderate slopes showed a higher transition probability.
- Edge growth (urban sprawl from existing settlements) and leapfrog development (urban patches forming near IT zones and schools) were both observed.

As noted previously, the current model has several strengths. However, it was particularly striking that socio-economic, hydrological, or zoning policy layers were excluded from the simulation, as these factors would be expected to have a considerable impact on real-world land conversion processes. Later versions could include these:

- Census-based population projections.
- Flood risk zones.
- Urban planning restrictions or green zone buffers 9.

Additionally, as they have yet to be computed, the model calibration and validation steps (ROC curves, Kappa index, etc.) are only conceptually implemented. For purposes of submission, a placeholder ROC and Kappa matrix have been included, and final computation will occur prior to publication.

CONCLUSION

This research applied a Cellular Automata (CA) simulation model with fuzzy spatial drivers and GIS based spatial analysis to forecast urban growth in peri–urban Bavdhan, Sus, and Mahalunge. The model was further refined through weighted suitability analysis derived from the proximity to primary infrastructure like roads, gated communities, IT parks, and zones of already developed construction.

The results showed conclusive urban growth spatially from 2019 to 2023. The urban area has increased from 146,875 pixels to 164,295 pixels, which accounts for 11.86% growth or roughly 1,567.8 hectares. The derived simulation predicts the urban footprint will reach 2.94 km² by 2033, which represents roughly a 1.61 times expansion from 2019's figure. Analyzing the Change Detection Map for 2023–2033, approximately 1,036,266.69 m² (10,816 pixels) are expected to be newly urbanized, while 14,114,343.22 m² will remain non-urban, and 25,879,265.98 m² is expected to remain developed.

A striking observation is that there is a concentration of urban growth along the Pashan–Sus Road corridor and the boundary between Sus and Bavdhan Budruk, indicating that ease of access to infrastructure and flat topography are major spatial determinants facilitating expansion. This matches with the outputs of the suitability map, which emphasizes these zones as very advantageous for further developmental activities. This area lies between Nonetheless; the study does have some limitations. Socioeconomic factors, hydrological constraints, ecological vulnerability, and land ownership patterns were not included in the model due to data constraints. This means that while the spatial precision is high, the lack of policy, social, and environmental factors may undermine how realistic the prediction is in practice.

Still, the research demonstrates how effective fuzzy logic and remote sensing make in CA-driven urban growth simulations for foresight planning and for urban planners and policymakers and reinforces the urgent need for sustainable urban planning as the rage of unplanned sprawl relentlessly devours once agri- and vegetative lands on the outskirts of Pune.

REFERENCES

1. United Nations. World urbanization prospects: The 2018 revision. New York: United Nations Department of Economic and Social Affairs, Population Division; 2018. Available from: https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.
2. Sudhira HS, Ramachandra TV, Jagadish KS. Urban sprawl: Metrics, dynamics and modelling using GIS. Int J Appl Earth Obs Geoinf. 2004;5(1):29–39. doi: 10.1016/j.jag.2003.08.002.
3. Clarke KC, Gaydos LJ. Loose coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. Int J Geogr Inf Sci. 1998;12(7):699–714. doi: 10.1080/136588198241617.
4. Lu H, Wang R, Ye R, Fan J. Monitoring long-term spatiotemporal dynamics of urban expansion using multisource remote sensing images and historical maps: A case study of Hangzhou, China. Land. 2023;12(1):144. doi: 10.3390/land12010144.
5. Al-Darwish A. Predicting the future urban growth and its impacts on the surrounding environment using urban simulation models: Case study of Ibb city – Yemen. Alexandria Eng J. 21(3):265–272. Available from: https://www.sciencedirect.com/science/article/pii/S1110016818300863.
6. Ravetz J, Fertner C, Nielsen TS. The dynamics of peri-urbanization. In: Peri-urban futures: Scenarios and models for land use change in Europe. Springer; 2013. p. 13–44. doi: 10.1007/978-3-642-30529-0_2.
7. Sivaramakrishnan KC. Re-visioning Indian cities: The urban renewal mission. New Delhi: Sage Publications India; 2011.
8. White R, Engelen G. Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land-use patterns. Environ Plan A. 1993;25(8):1175–1199. doi: 10.1068/a251175.
9. Batty M, Xie Y, Sun Z. Modeling urban dynamics through GIS-based cellular automata. Comput Environ Urban Syst. 1999;23(3):205–233. doi: 10.1016/S0198-9715(99)00015-0.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–174. doi: 10.2307/2529310.

