Journal of the Geographical Association of Tanzania, Vol. 46, No. 2: 1–27

A Comparative Evaluation of the CA-Markov and Random Forest-Cellular Automata Models for Predicting Urban Sprawl in Dar es Salaam, Tanzania

Nyatabori Antonietta Kwame[*] *& Zakaria Ngereja*[§]

Abstract

Urban expansion is occurring rapidly in African cities, necessitating accurate predictive tools to guide sustainable urban planning. This paper draws data from a study that compared the performance of two spatial modelling approaches—Cellular Automata-Markov Chain (CA-Markov) and Random Forest-Cellular Automata (RF-CA)—in forecasting urban sprawl in Dar es Salaam, Tanzania. Multitemporal Landsat imagery (1995, 2004, 2014, and 2024) and key biophysical and socioeconomic drivers were used to model the past dynamics and predict the 2024 urban extent. The Random Forest-Cellular Automata (RF-CA) model integrates machine-learning-derived transition probabilities with a neighbourhood-based CA framework. In contrast, the CA-Markov model relies on historical transition probabilities calibrated from previous time steps. The results showed that RF-CA achieved a higher predictive performance (accuracy: 77.84%; Kappa: 0.67; Figure of Merit (FoM): 56.83%) than CA-Markov (accuracy: 69.38%; Kappa: 0.542; Figure of Merit (FoM): 51.33%), and more accurately captured the spatial complexity in peri-urban areas. These results imply that RF-CA is a more reliable model for predicting urban growth, especially in complex, rapidly changing peri-urban areas. Population density, elevation, and climate were found as the most influential factors in this study. These results highlight the value of hybrid machine-learning models for urban growth prediction in rapidly transforming cities and demonstrate their advantages over traditional Markov-based models. Furthermore, this paper generates empirical evidence on the applicability of RF-CA for urban growth modelling in data-scarce African contexts using geospatial and remote sensing methods.

Keywords: *Urban sprawl, CA-Markov, Random Forest - Cellular Automata, GIS and Remote Sensing, Sub-Saharan Africa*

1. Introduction

Urban sprawl has accelerated globally, driven by population growth, economic development, and increasing land demands. Between 1990 and 2014, global urban sprawl increased by approximately 95%; equivalent to 4% per year (Behnisch, Krüger & Jaeger, 2022). This rapid expansion has led to the conversion of agricultural land, proliferation of informal settlements, and heightened pressure on infrastructure, particularly in developing countries

[*]Department of Geospatial Sciences and Technology, Ardhi University, Dar es Salaam, Tanzania: nyataborikwame@gmail.com (Corresponding author)
[§] Department of Geospatial Sciences and Technology, Ardhi University, Dar es Salaam, Tanzania

© *Department of Geography, UDSM, 2026* ▪ https://doi.org/10.56279/jgat.v46i2.389

where planning systems often lag behind urbanisation (Seto, Güneralp & Hutyra, 2012). Therefore, accurately predicting future urban expansion patterns is essential for sustainable urban planning and management (Mnyali & Materu, 2021; Wang, Shi & Zhou, 2020).

Dar es Salaam, Tanzania's largest city and primary economic hub, exemplifies these challenges as one of Africa's fastest-growing cities. Its population surged from 4.3 million in 2012 to 5.3 million in 2022 (URT, 2022), thereby fuelling complex spatial dynamics characterised by informal settlement patterns and coastal vulnerability (Batty, Xie & Sun, 1999; Augustijn-Beckers, Flacke & Retsios, 2011).

Traditional land-use prediction methods struggle to capture heterogeneous and rapidly evolving land-use patterns (Jayabaskaran & Das, 2023). However, advances in geospatial analysis and machine learning have enhanced urban growth modelling. Cellular Automata (CA) integrated with a Markov Chain – known as CA-Markov is a widely used approach that combines historical transition probabilities with neighbourhood rules to simulate land-use changes (Sang et al., 2011; Hamad, Balzter & Kolo, 2018; Simon, Lyimo & Yamungu, 2023; Atef, Ahmed & Abdel-Maguid, 2024). However, the CA-Markov model assumes stationary historical transitions, thereby limiting its ability to represent dynamic nonlinear interactions in rapidly urbanising environments (Nugroho & Al-Sanjary, 2018).

To overcome these limitations, hybrid models incorporating machine learning—such as Cellular Automata (CA) and Artificial Neural Networks (ANNs) (Gharaibeh et al., 2020*)* and random forest-cellular automata (RF-CA) (Kamusoko & Gamba, 2015)—have emerged. The Random Forest—an ensemble method (Breiman, 2001)—excels at handling high-dimensional data, feature selection, and nonlinear relationships. When coupled with CA, RF-CA generates spatially explicit predictions by deriving transition probabilities from machine learning while retaining neighbourhood effects (Gharaibeh et al., 2020; Khawaldah, Farhan and Alzboun, 2020; Asif et al., 2023). Kamusoko and Gamba (2015) demonstrated the superiority of RF-CA over the Support Vector Machine Cellular Automata (SVM-CA) and the Logistic Regression Cellular Automata (LR-CA) in capturing heterogeneous urban patterns.

Nonetheless, there is limited empirical evidence comparing the predictive performance of RF-CA with the CA-Markov approach in rapidly urbanising cities of Sub-Saharan Africa despite these advances (Laros et al., 2014; Mberu, Béguy and Ezeh, 2017; Tumwesigye et al., 2023). Most RF-CA evaluations have focused on Asia and Europe, leaving a gap in understanding its performance in data-scarce and informally driven African contexts, such as Dar es Salaam.

Addressing this gap requires not only applying RF-CA in an African context, but also benchmarking it against the traditional CA-Markov approach—which is still widely used by planners in the region—to demonstrate whether, and by how much, the added complexity of the machine-learning hybrid model improves on established practice. Therefore, the research gap addressed in this study is the limited comparative evaluation of RF-CA and CA-Markov within the same urban context and evaluation framework.

The specific objectives were to: (i) identify the key drivers of urban expansion; (ii) analyse spatial and temporal growth dynamics over the past three decades; and (iii) evaluate the predictive performance of both models using multiple accuracy metrics. By applying RF-CA to Dar es Salaam—one of East Africa's fastest-growing cities, and (to our knowledge) an understudied context for this hybrid model—this study contributes to the limited evidence base on machine-learning integrated urban growth modelling in Sub-Saharan Africa.

2. Theoretical Framework

Urban sprawl does not occur as a single uniform event; it is a cumulative outcome of many local decisions that accumulate into identifiable area patterns. It is primarily driven by population growth, land demand, accessibility, environmental suitability, and existing built forms (Azabdaftari & Sunar, 2024; Fitawok & Minale, 2024; Li et al., 2025). This study draws on a complexity theory that frames cities as self-organising systems whose large-scale form emerges from simple and local rules of interaction (Spencer, 2009), which is in harmony with Tobler's (1970) principle that nearby locations influence one another more strongly than distant ones. Cellular automata (CA) operationalise this idea by modelling land as a grid of cells whose states evolve through neighbourhood-based transition rules (White & Engelen, 1993). Within a CA-Markov framework, CA supplies this spatial logic, whereas a Markov chain supplies the temporal logic using historically observed transition rates to project the extent of change likely to occur (Azabdaftari & Sunar, 2024).

However, traditional CA-Markov models are based on relatively simplistic rules that cannot effectively characterise the complex and context-dependent relationship between expansion and its drivers. By incorporating an ensemble of decision trees into the cellular automata framework, the random forest cellular automata (RF-CA) model solves this problem by estimating change probabilities using multiple interactive nonlinear predictors (Kamusoko & Gamba, 2015), given the potential of the Random Forest model to avoid overfitting and average predictions among resampled trees (Breiman, 2001), as well as the trend toward using machine-learning techniques in geospatial analysis (Shaamala et al., 2025). In both cases, this is possible because geospatial technologies can combine space and time into forecasts useful for

planning purposes (Musa, Hashim & Reba, 2017). Thus, comparing CA-Markov and RF-CA is relevant to understand which of the two approaches—rule-based or data-driven—provides the most accurate representation of urban dynamics in rapidly growing cities, such as Dar es Salaam.

3. Materials and Methods

3.1 Study Area

Dar es Salaam is the largest city and major port of Tanzania that borders the Indian Ocean, with an area of 1,590km^2. It lies between 6°36′S and 7°0′S latitude, and 39°0′E and 39°33′E longitude. The city is the country's main economic hub, and consists of five administrative districts: Ilala, Kinondoni, Temeke, Ubungo, and Kigamboni. According to the 2022 National Population and Housing Census, the population of Dar es Salaam increased from 4.3 million residents in 2012 to 5.3 million residents in 2022, making it one of the fastest-growing cities in East Africa. The city experiences a tropical coastal climate characterised by high humidity and warm temperatures throughout the year, with annual rainfall between 1,000 and 1,150mm. The temperature ranges from 25°C during the cooler months, to 38°C during the hottest months of the year. These demographic and climatic conditions, combined with ongoing urbanisation and informal settlement expansion, make Dar es Salaam a critical case study for examining urban sprawl, infrastructure challenges, and environmental management.

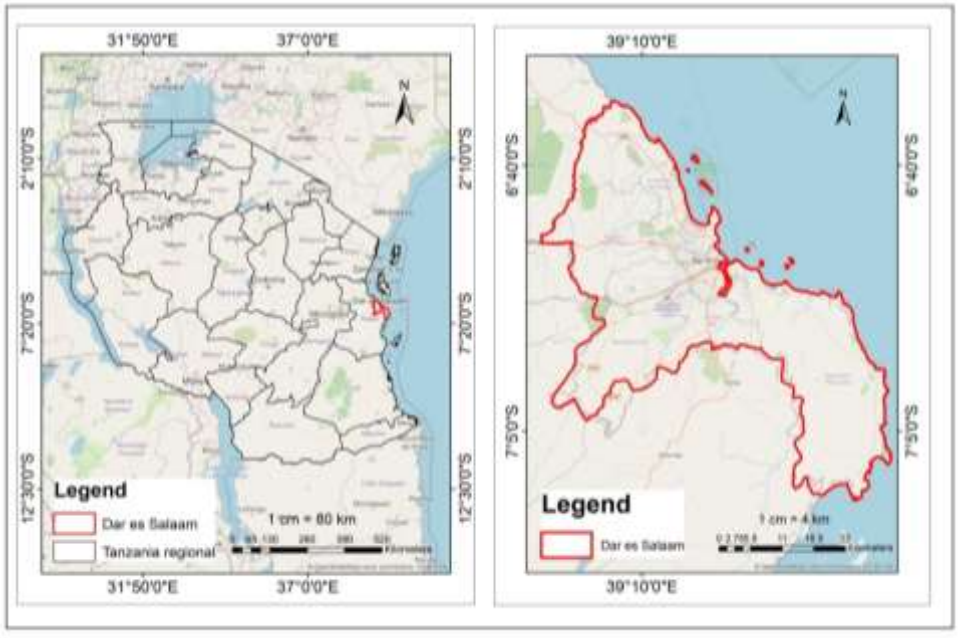


Figure 1: Location Map of Dar es Salaam

Source: Authors

3.2 Data Sources and Preprocessing

This study used multiple geospatial datasets (Table 1) covering the Dar es Salaam metropolitan area in 1995, 2004, 2014, and 2024 for the analysis. Landsat 5 TM and Landsat 8 OLI images, with a 30m resolution, were obtained from the USGS Earth Explorer platform for land cover classification. The topographical data, including elevation and slope, were derived from the Shuttle Radar Topography Mission; and processed in QGIS using the OpenTopography plugin. The accessibility and road network data were obtained from OpenStreetMap (OSM), and were used to generate the Euclidean distance to a road network. The population density data were retrieved from the Global Human Settlement (GHS) dataset (Schiavina et al., 2023).

Table 1: Data Used for Urban Sprawl Analysis in Dar es Salaam (1995–2024)

Data Type	Spatial Resolution	Temporal Coverage	Source	Description
Satellite Imagery (Landsat)	30m	1995, 2004, 2014 and 2024	Landsat 5 (TM), 8 (OLI) (USGS Earth explorer)	Land Use/Land Cover (LULC) classification
Tanzania Administrative Boundaries	vector	2022	Africa Geoportal	Determining the extent of the analysis
DEM (Digital Elevation Model) and slope	30m	static	SRTM (30m)	Indicates terrain steepness
Road Network Data	Vector	Static (2024)	OpenStreetMap (OSM) Geofabrik (website)	Assessing transportation impacts on urban expansion
Population Density	1km (resampled)	2004–2024	Global Human Settlement (GHS)	Understanding population-driven urban expansion
Climate Data	5km 4km (resampled)	2004–2024	CHIRPS (Rainfall) TERRA CLIMATE (Temperature)	Assessing climate-related urban growth constraints

The climatic variables were obtained directly from official data repositories, with rainfall derived from the CHIRPS Version 2.0 daily precipitation data at a spatial resolution of 0.05° (~5km). On the other hand, the maximum and minimum temperatures (Tmax and Tmin) data were obtained from the TerraClimate monthly climate dataset at a spatial resolution of approximately 1/24° (~4km) for the study years 1995, 2004, 2014, and 2024. The datasets were downloaded for the study area and processed by clipping them to the boundaries of Dar es Salaam. The CHIRPS daily precipitation was aggregated to annual rainfall totals (mm), while the monthly TerraClimate Tmax and

Tmin values were aggregated to annual temperature measures. The CHIRPS data were accessed from the Climate Hazard Centre data and TerraClimate repositories (Funk et al., 2015; Abatzoglou et al., 2018).

All datasets were projected to the coordinate reference system of the land cover data, and aligned to a common spatial grid before model development using R Studio (version 4.4.3). The 2004 land cover raster was used as the reference raster to standardise the spatial extent, resolution, and pixel alignment of the predictor layers using nearest-neighbour interpolation (NGB). Feature selection was based on the relevance of the features to urban expansion: population density and accessibility influence settlement distribution; slope affects buildable land; distance to roads indicates connectivity; and rainfall and temperature reflect environmental suitability for land conversion. This preprocessing workflow ensured that the multitemporal datasets were compatible with model calibration, simulation, and validation; following the methodological flowchart in **Error! Reference source not found.**.

The Landsat imagery comprised Collection 2 Level-2 Surface Reflectance scenes (Landsat 5 TM for 1995 and 2004; Landsat 8 OLI for 2014 and 2024), obtained from the USGS Earth Explorer platform (https://earthexplorer.usgs.gov), with scenes selected from the dry season (June–September) of each target year to minimise cloud cover. The Digital Elevation Model was the SRTM 1 Arc-Second Global (SRTMGL1 v3, 30m) product, accessed through the OpenTopography API (https://opentopography.org), and processed for slope derivation in QGIS. Road network vectors were obtained from the OpenStreetMap Tanzania extract distributed by Geofabrik (https://download.geofabrik.de). Population density was obtained from the Global Human Settlement Population Grid (https://human-settlement.emergency.copernicus.eu/download). The rainfall data were sourced from CHIRPS v2.0 (0.05° native resolution, resampled to 5km; Climate Hazards Centre (https://data.chc.ucsb.edu/products/CHIRPS-2.0)), while the temperature data were obtained from TerraClimate (4km native resolution; University of Idaho Climatology Lab (https://www.climatologylab. org/terraclimate.html)).

The land cover classification was performed using Google Earth Engine (GEE) with a Random Forest (RF) classifier. Four land cover classes were defined: water, vegetation, built-up land, and bare land. The training samples were generated through the visual interpretation of high-resolution basemaps in Google Earth Pro (for example, Google Imagery) with a set of reference samples. The RF classifier was trained using 500 trees, with 80% of the samples used for training and 20% for validation, and with a random stratified split for each year. The validation set comprised 120 reference points (distributed across water, vegetation, built-up, and bare land); and the full per-year and per-class confusion matrix are reported in the Appendix. These classified land cover maps served as

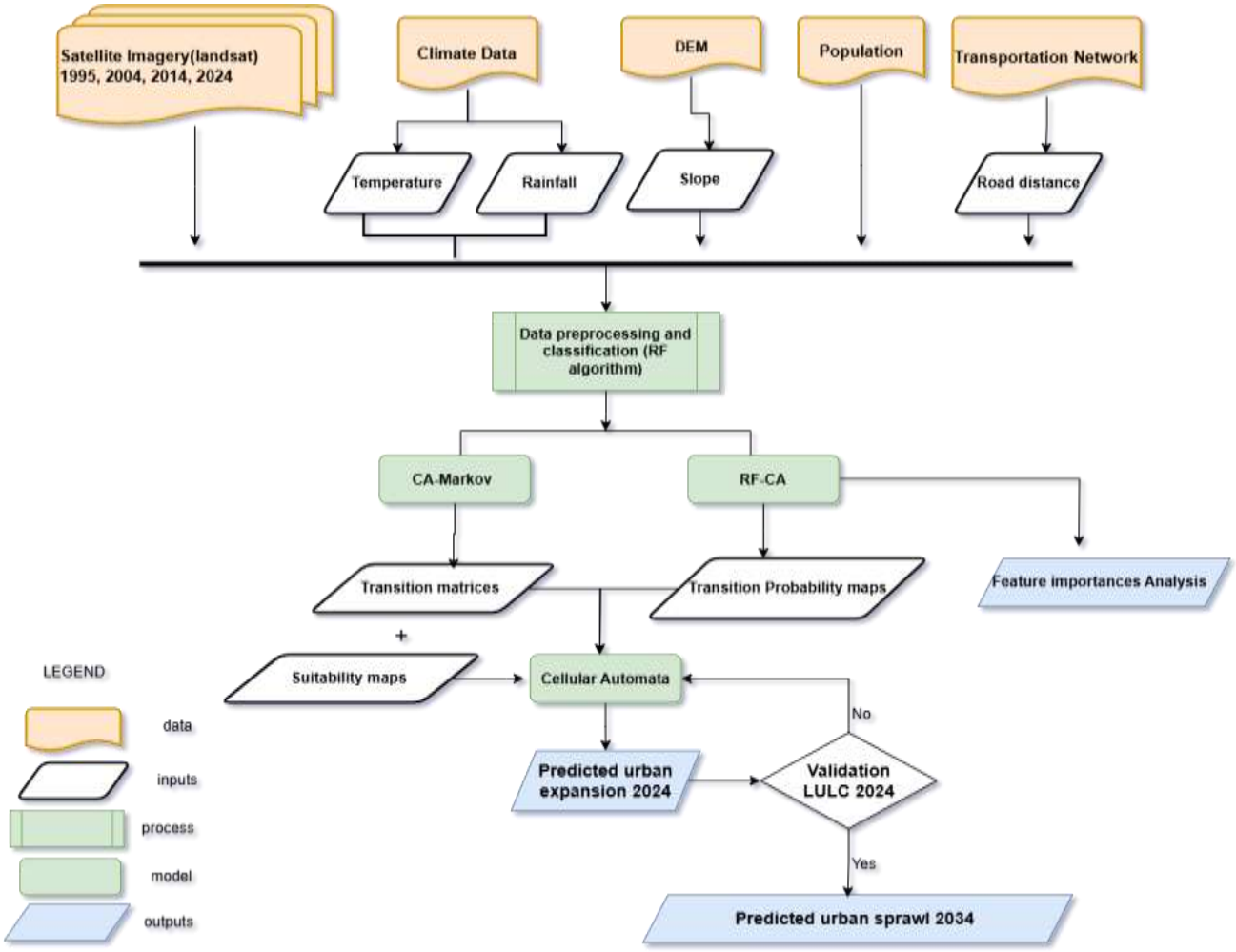


Figure 1: Methodological Flowchart for Urban Sprawl Prediction

3.3 Model Implementation

3.3.1 RF-CA Model

The RF-CA model architecture combines the Random Forest implementation of scikit-learn with a custom Cellular Automata module in Python. The model was implemented using Python 3.12 and JupyterLab (version 3.6.3). The Random Forest classifier was trained using predictor variables derived from the initial land cover state of 2004–2014. The predictor set consisted of ten variables: land cover, Moore neighbourhood, extended neighbourhood, slope, distance to roads, population density, DEM, mean temperature, rainfall, and distance to urban areas. Predictor variables were used in their native units without normalisation or rescaling since the Random Forest partitions are based on the relative ordering of feature values at each split rather than the absolute distance between observations. The dataset was divided into 70% training and 30% testing subsets using a random fixed seed of 42.

To address class imbalance, all observed urban-transition cells were retained, whereas non-transition non-urban cells were randomly sampled at a maximum ratio of 3:1. The resulting dataset was divided into 70% training and 30% testing subsets using a random seed of 42. Unlike the land cover classification Random Forest in Section 3.2, the RF-CA transition model was trained on a balanced sample of pixels using a 70/30 training-testing split.

To minimise temporal information leakage, predictor variables were derived from the initial state of each historical calibration interval, whereas the subsequent land cover was used only to construct the transition target. For the 2004–2014 calibration, predictor information from 2014 was not used to train the Random Forest, but was used only to define the binary transition target.

For a given cell i, the probability of transitioning to urban class $P_{i,t}$ was calculated as the mean prediction across all trees in a forest:

$$P_{i,t} = \frac{1}{N_{trees}} \sum_{k=1}^{N_{trees}} I\left(h_k\left(X_{i,t}\right)\right) \qquad (1)$$

Where:

$P_{i,t}$ = predicted probability of urban transition for pixel i at time t

$h_k\left(X_{i,t}\right)$ = prediction of the k-th decision tree for the feature vector $X_{i,t}$ of pixel i

N_{trees} = total number of trees in the forest.

I = indicator function (1 if the tree predicts "urban", 0 otherwise)

$X_{i,t} = \left[LC_{i.t}, N_{i,t}^{(3\times3)}, N_{i,t}^{(5\times5)}, S_i, R_i, DEM_i, POP_{i,t}, Rain_{i,t}, Du_{i,t}, T_{i,t} \right]$

Where:

$LC_{i.t}$ = land cover class at pixel i at time t
$N_{i,t}^{(3\times3)}$ = number of urban cells within the 3×3 neighbourhood surrounding pixel i at time t (Moore neighbourhood)
$N_{i,t}^{(5\times5)}$ = number of urban cells within the 5×5 neighbourhood surrounding pixel i at time t (extended neighbourhood)
S_i = slope at pixel i
R_i = distance from pixel i to the major road
DEM_i = digital elevation at pixel i
$POP_{i,t}$ = population density at pixel i at time t
$Rain_{i,t}$ = Rainfall at pixel i at time t
$Du_{i,t}$ = distance from pixel i to existing urban areas at time t
$T_{i,t}$ = mean temperature of pixel i at time t

The resulting transition probabilities were integrated into a CA simulation using an annual iterative allocation procedure. The CA simulation was run iteratively at an annual time-step for ten iterations. At each iteration, 3×3 Moore and 5×5 extended windows were recalculated from the updated land cover state to represent the local and broader spatial influences of existing and newly urbanised cells, respectively. The formula used was:

$$S_t = Top_{N_t}(P_{i,t}) \qquad (2)$$

Where:

S_t = set of pixels selected for conversion to the urban class at iteration t
Top_{N_t} = selection of the N_t cells with the highest predicted transition probabilities
$P_{i,t}$ = predicted probability of urban transition
N_t = number of cells to be converted during iteration t

The RF transition probabilities $P_{i,t}$ were passed to the CA allocation procedure, which ranks all non-urban N_t cells, and selects the cells with the highest probabilities as per the formula:

$$N_t = [0.01N_{nu,t}] \qquad (3)$$

Where:

N_t = number of cells to be converted during iteration t
0.01 = annual transition rate (1%)
$N_{nu,t}$ = number of non-urban cells

The number of cells converted during each iteration was determined using a fixed transition rate of 1% of the currently non-urban cells. After each

allocation step, the land cover was updated, and the 3×3 and 5×5 neighbourhood measures were recalculated annually until the specific target year was reached.

The simulated 2024 urban extent was compared with an independently classified 2024 land cover map derived from Landsat satellite images. For the 2024–2034 simulation, the 2024 predictor layer was held constant throughout the CA iteration, while the land cover state and neighbourhood windows were updated annually. Subsequently, the 2034 RF-CA output represented a conditional projection of urban expansion based on the 2024 predictor variables.

3.3.2 CA-Markov Model

For the CA-Markov implementation in TerrSet (v19.0.6), the transition potential analysis focused on the 2004–2014 period as the calibration interval, with the CA simulation run iteratively at an annual time step of 10 iterations per period to capture yearly urban growth dynamics, excluding earlier transitions (1995– 2004).

According to Wang, Munkhnasan and Lee (2021) and Simon, Lyimo and Yamungu (2023), the Markov Chain model calculates a transition probability matrix where each element P_{ij} represents the probability that a pixel of class i will change to class j over a specified time interval, as follows:

$$P_{ij} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \cdots \\ P_{n1} & P_{n2} & \cdots & P_{nm} \end{bmatrix} \qquad (4)$$

The model was then calculated as follows:

$$LS_{(t+1)} = P_{ij} \times LS_t \qquad (5)$$

Where:

$LS_{(t+1)}$ represents Land Cover at time $(t + 1)$

LS_t represents Land Cover at time t

The Markov chain model used the 2004 Land Cover as the starting image and the 2014 Land Cover as the ending image. The transition probability matrix was used alongside a set of suitability maps derived from driving factors—such as distance to roads, slope, elevation, and population density—using the Multi-Criteria Evaluation (MCE) module in TerrSet; and was then used in the CA simulation to predict the 2024 urban extent. For future projections, the transition probabilities from 2014 to 2024 were used to project the conditional 2034 scenario, while the same MCE-derived suitability maps and weights were held constant.

3.4 Parameter Calibration and Selection

3.4.1 CA-Markov Model

The CA-Markov model was parameterised in TerrSet (v19.0.6) using a 10-year calibration/prediction interval, represented by ten annual CA iterations. The 2004–2014 land-cover transition was used for model calibration and validation of the 2024 urban extent, and the 2014–2024 transition was subsequently used to derive the transition information for the 2034 scenario. Transition quantities were estimated using the Markov Chain component, and spatial allocation was controlled using suitability maps generated through Multi-Criteria Evaluation based on distance to roads, slope, elevation, and population density. The CA component then spatially allocates the projected transitions using the configured neighbourhood and suitability information.

3.4.2 RF-CA Model

A 3×3 Moore neighbourhood and a 5×5 extended neighbourhood measure were recalculated at each CA iteration and supplied as separate predictors to the Random Forest model. The Random Forest consisted of 100 trees with a fixed random seed of 42, and the transition dataset used a 70/30 training-testing split. Class imbalance was addressed as follows: all non-urban cells were sampled at a maximum ratio of 3:1. The CA component used an annual time step and a transition rate of 1%, whereby eligible non-urban cells were ranked according to their RF-derived transition probabilities, and the cells with the highest probabilities were converted to the urban class.

Table 2: RF-CA Neighbourhood Window Comparison

Configuration	Cells	Approx. coverage	Role in this study
3×3	8 neighbours	90m×90m	Local neighbourhood predictor
5×5	24 neighbours	150m×150m	Extended neighbourhood predictor

3.5 Model Validation and Accuracy Assessment

The predictive performance of each model was evaluated by comparing the simulated 2024 land cover maps with an independently classified reference map for 2024. The accuracy of the model was assessed using several complementary metrics. The overall accuracy was computed as the ratio of the correctly classified pixels to the total number of pixels in all land cover classes. Kappa simulation was used to measure the level of agreement between the predicted and reference maps. To specifically evaluate the urban class performance, an urban-specific F1-score was calculated to balance precision and recall, as defined below:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (6)$$

Furthermore, the Figure of Merit (FoM) was used to assess the accuracy of the predicted land-use change from non-urban to urban areas between 2014 and 2024. The FoM quantifies the proportion of correctly predicted changes relative to the total observed and predicted changes, which is calculated as:

$$FoM = \frac{H}{H + M + F} \qquad (7)$$

Where:

H represents hits (correctly predicted changes)
M denotes misses (observed changes that are not predicted)
F denotes false alarms (predicted changes that are not observed).

Only pixels that transitioned to urban areas in the reference map were included in the FoM calculations. In addition, a 2×2 was constructed for each model, from which the user's and producer's accuracies for the urban class were derived, alongside quantity and allocation disagreement, following Pontius and Millones (2011). Quantity disagreement is the absolute difference between the predicted and observed urban areas attributable to over-/or under-estimation of the total quantity disagreement is the area misclassified due to spatial mismatch despite the correct overall quantity. In contrast, allocation disagreement is the area misclassified due to spatial mismatch despite the correct overall quantity.

Spatial patterns of agreement and disagreement were further examined using side-by-side comparisons of predicted and observed urban extents (Figures 5 and 6), while transition matrices summarised class-to-class land cover changes between 2004 and 2014, and 2014 and 2024 (Figures 7 and 8). These analyses provided a robust assessment of each model's ability to capture the spatial complexity of urban expansion, particularly in peri-urban areas.

4. Results

4.1 Land Cover Classification Accuracy and Urban Expansion Trends

Random Forest classification achieved high accuracy in distinguishing land cover classes across all study years, with the overall accuracies ranging from 83.33% to 87.5%; and the kappa values ranging from 0.765 to 0.824. The full per-year confusion matrices, sample sizes, and user/producer accuracies are reported in the Appendix. This reliable classification provides a robust foundation for analysing urban sprawl in cities. The urban extent increased substantially from 147.40km² (in 1995) to 505.01 km² (in 2024), a gain of 357.62km² (Figures 3 and 4). The expansion occurred in three distinct phases: between 1995 and 2004 the urban land increased by 29.56km² (20.05%); between 2004 and 2014 the growth was more rapid, adding 152.20km² (86.01%); and expansion continued strongly between 2014 and 2024 with a 175.86km² (53.43%) expansion, marking the largest absolute increase in the study period. Most of this growth occurred at the expense of vegetation and bare land, whereas the water area remained relatively stable.

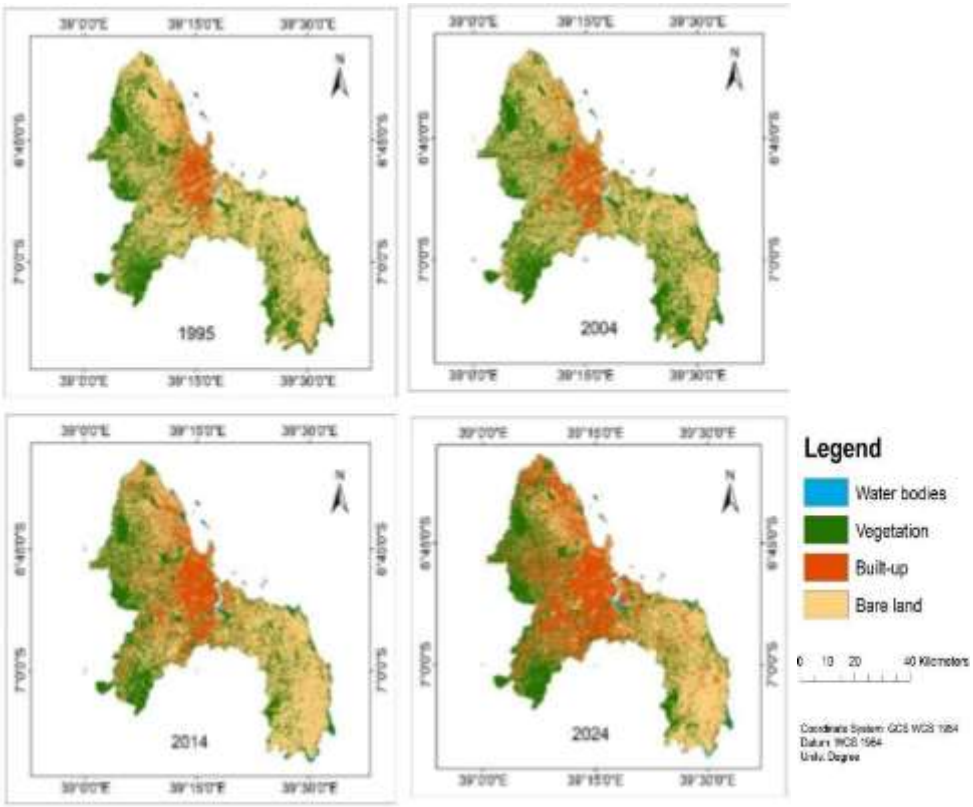


Figure 2: Land Cover Classification (1995 to 2024) Using the Random Forest (RF) Algorithm

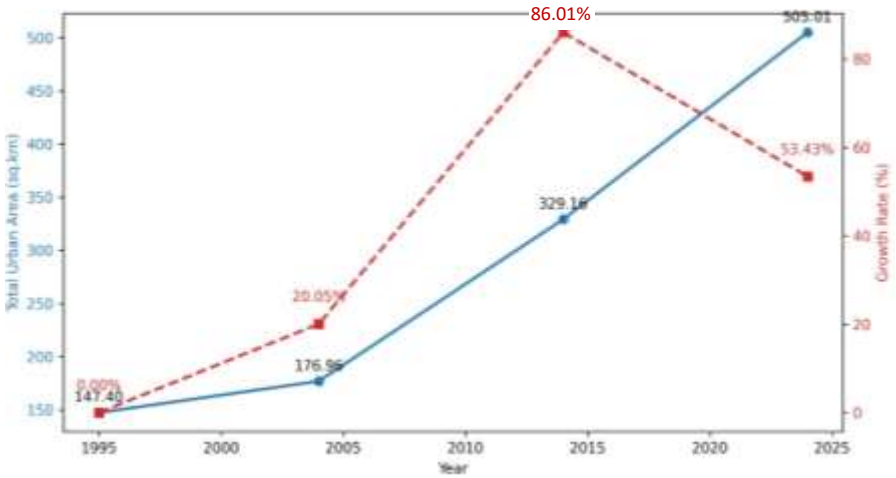


Figure 3: Trends in Urban Expansion and Growth Rate in Dar es Salaam (1995–2024)

Sources: Authors

4.2 Model Performance Comparison

When comparing the predictive models, both CA-Markov and RF-CA captured the general trend of outward expansion but differed in spatial detail and accuracy. The CA-Markov model projected an urban area of 457.30km² by 2024, whereas the RF-CA model projected 441.64km², both of which underestimate the actual size of 505.01km². However, RF-CA produced more accurate results, correctly predicting 343.05km² of urban area compared with 326.43km² for CA-Markov. It also reduced errors, with lower under-prediction (161.96 vs 178.59km²) and over-prediction (98.59 vs 130.88km²) compared with CA-Markov.

Table 3: Cross Tabulation of Correctly Predicted, Missed and Falsely Predicted Urban Areas for CA-Markov and RF-CA Models

Model	Correctly Predicted (TP)	Missed Urban (FN)	False Urban (FP)	User's/Producer's Accuracy
CA-Markov (km²)	326.43	178.59	130.88	71.4%/64.6%
RF-CA (km²)	343.05	161.96	98.59	77.7%/67.9%

Table 4 indicates that although RF-CA slightly underestimated the total urban quantity by a larger margin than CA-Markov, it produced substantially less spatial misallocation: meaning its errors were more a matter of predicting the wrong exact location than the wrong overall amount of growth, a distinction that the raw accuracy and Kappa alone do not reveal.

Table 4: Quantity and Allocation Disagreement

Model	Quantity Disagreement (km²)	Allocation Disagreement (km²)	Total Disagreement (km²)	Main Errors Characteristic
CA-Markov	47.71	261.76	309.47	High spatial misallocation
RF-CA	63.37	197.18	260.55	Higher quantity error but lower spatial misallocation

Table 5 shows that the RF-CA model consistently outperformed the CA-Markov model across all evaluation metrics. It achieved 77.84% accuracy, Kappa 0.67, F1-score 72.5%, and FoM 56.83%; whereas CA-Markov achieved 69.38% accuracy, Kappa 0.542, F1-score 67.8%, and FoM 51.33%. These results demonstrate that RF-CA not only reduces prediction errors, but also captures the spatial complexity of urban growth more effectively than other models.

Table 5: Performance Comparison Between RF-CA and CA-Markov Model

Metric	RF-CA	CA-Markov
Accuracy	77.84%	69.38%
Figure of Merit (FoM)	56.83%	51.33%
Kappa Simulation	0.67	0.542
F1 Score	72.5%	67.8%

4.3 Spatial Patterns of Predicted Urban Growth

The RF-CA model predicted urban sprawl in Dar es Salaam for 2024 (Figure 5), showing edge expansion and infill development mainly around existing built-up areas such as Ubungo, Kimara, and Mbezi. Growth is influenced by population pressure and proximity to major roads, resulting in a visually contiguous pattern of urban growth around these nodes. This description is based on a visual inspection of the predicted map rather than quantified landscape metrics; and is, therefore, presented as a qualitative observation only. Figure 5 shows a side-by-side comparison of the projected (left) and actual (right) land cover in 2024.

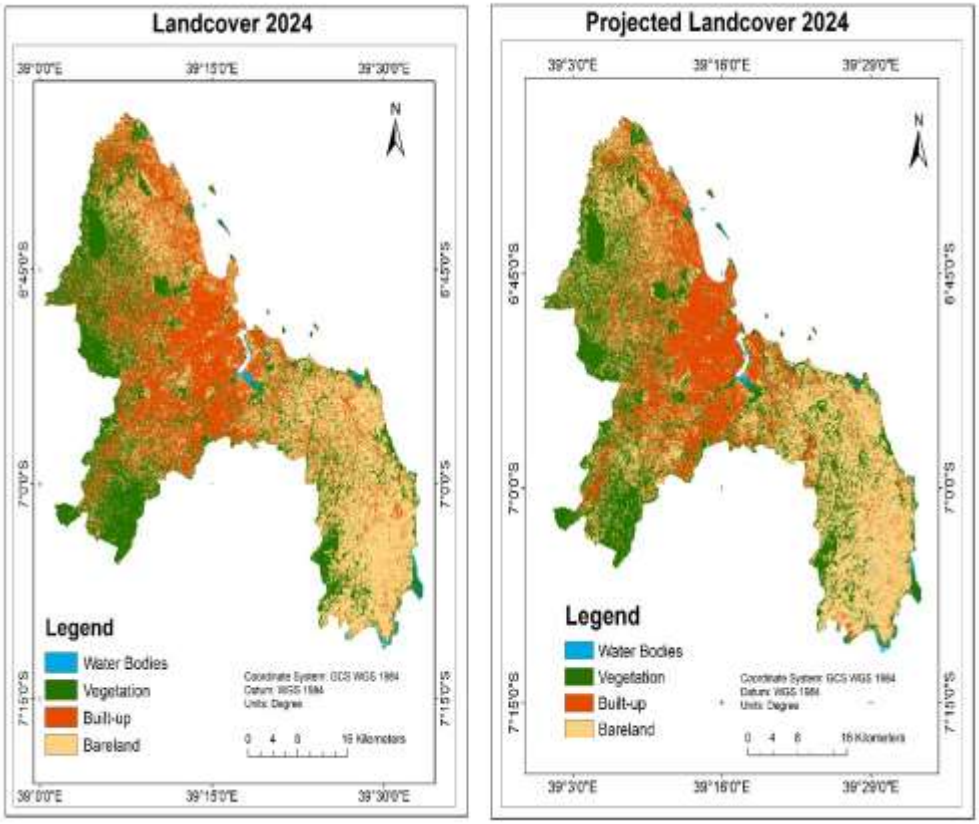


Figure 4: Projected Land Cover versus Actual Land Cover in 2024 by Random Forest – Cellular Automata (RF-CA) model.
Source: Authors

Figure 6 shows a side-by-side comparison of projected and actual land cover. The CA-Markov model's 2024 prediction shows urban expansion primarily over bare land and vegetation, driven by historical land cover transitions and spatial factors such as slope and road distance. The expansion was concentrated near roads and in low-slope areas, resulting in a more linear and constrained growth pattern than the RF-CA model predictions.

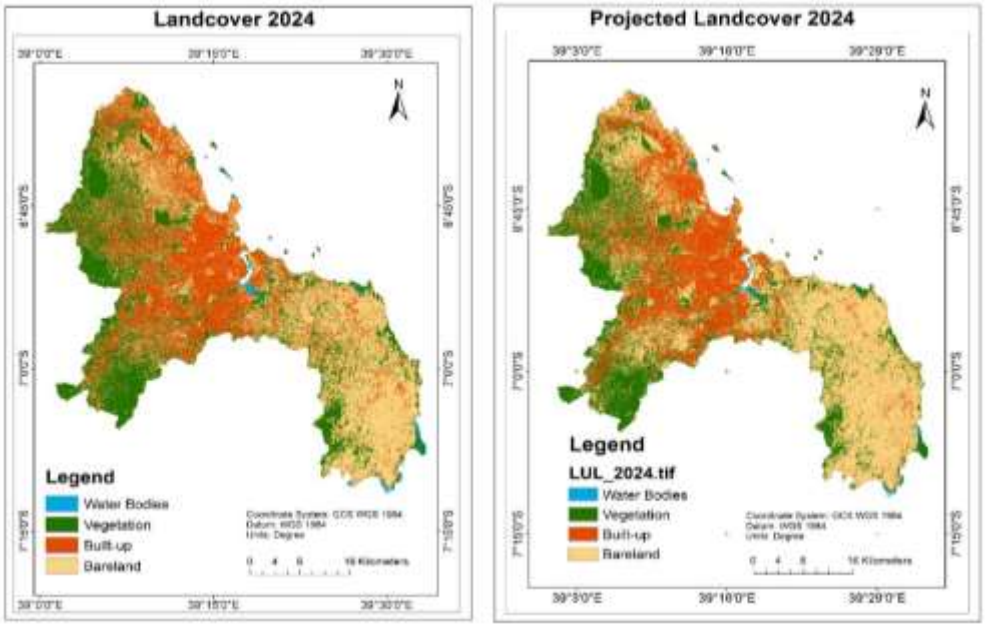
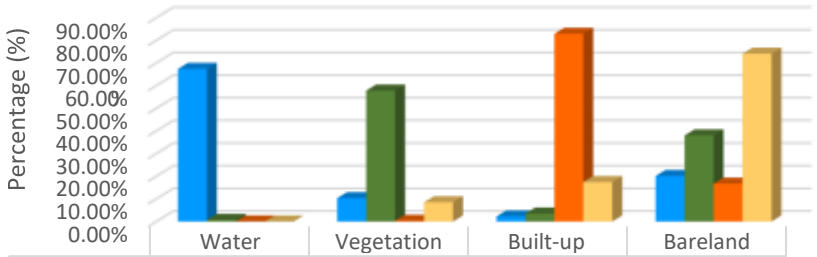


Figure 5: Projected Land Cover versus Actual Land Cover of 2024 by (CA-Markov) Model

Source: Author)

4.4 Transition Probabilities and Driving Factors

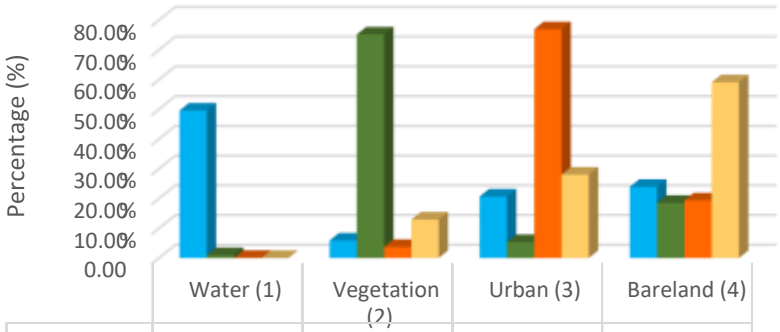
A transition probability analysis illuminated the key land conversion processes. The 2004–2014 period, as shown in Figure 7, showed a vegetation-to-urban transition of 3.5%, and a bare land-to-urban change of 17.4%. These rates increased to 5.3% and 28.0%, respectively, in the 2014–2024 projection (Figure 8).



	Water	Vegetation	Built-up	Bareland
Water	67.40%	10.30%	2.20%	20.10%
Vegetation	0.70%	57.80%	3.50%	38.00%
Built-up	0.00%	0.30%	82.90%	16.70%
Bareland	0.00%	8.50%	17.40%	74.10%

Figure 6: Land Cover Transition Matrix from 2004–2014.

Source: Authors



	Water (1)	Vegetation (2)	Urban (3)	Bareland (4)
Water (1)	49.57%	5.85%	20.66%	23.92%
Vegetation (2)	0.97%	75.14%	5.34%	18.55%
Urban (3)	0.02%	3.71%	76.90%	19.37%
Bareland (4)	0.01%	12.91%	28.03%	59.06%

Figure 7: Land Cover Transition Matrix from 2014–2024.

Source: Authors

In addition to its superior predictive performance, the RF-CA model provides insights into the main drivers of sprawl through a feature importance analysis as shown in Figure 9.

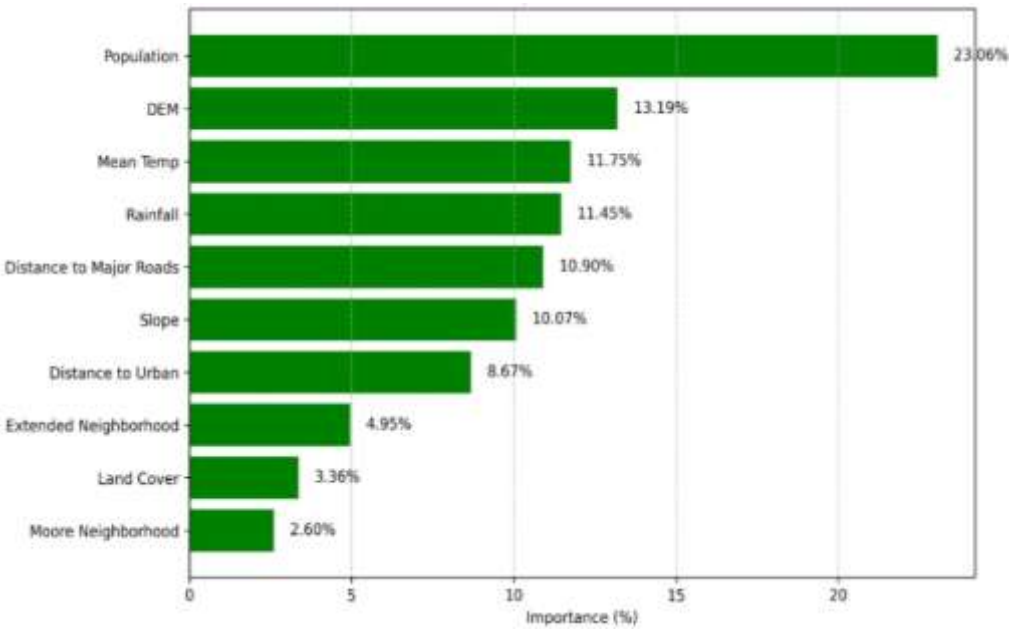


Figure 8: Feature Importance (%) Ranking for Urban Sprawl Prediction Using the RF-CA Model

Source: Authors

Figure 9 shows that population density (23.06%) was the dominant factor in urban sprawl; followed by elevation (13.19%), temperature (11.75%), rainfall (11.45%), and distance to roads (10.90%). Other contributors included slope (10.07%), and distance to existing urban areas (8.67%). The neighbourhood effects contributed comparatively less to the trained model, with the extended (5×5) neighbourhood accounting for 4.95%, and the Moore (3×3) neighbourhood accounting for 2.60%. Both neighbourhood variables were retained as separate predictors in the RF-CA simulation (see Section 3.3).

The RF importance was assessed using impurity-based feature importance provided by scikit-learn (feature importance), which represents the mean decrease in impurity contributed by each predictor across the trees in the fitted forest. These values represent model associations; and should not be interpreted as evidence of the causal effects of population, rainfall, or any other predictor of urban growth.

4.5 Future Projections to 2034

The 2034 outputs shown in Figure 10 are conditional scenarios because neither model incorporates future changes in population, climate, or road infrastructure. Future projections to 2034, as shown in Figure 10, indicate continued outward expansion under both models, but with distinct spatial patterns.

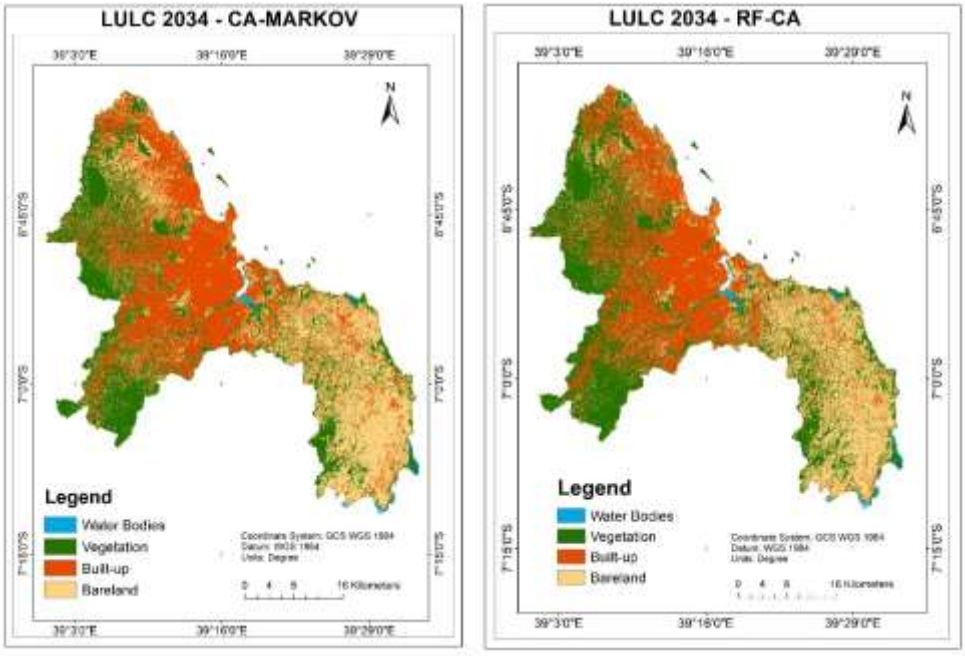


Figure 9: Land Cover 2034 showing Expansion of Urban Pattern Using Both CA -MARKOV and RF-CA Approaches

Source: Authors

As per Figure 10, the CA-Markov model suggested radial growth with some leapfrog development into isolated areas. In contrast, the RF-CA model indicated a more compact and organised expansion characterised by edge growth and infill development along the transport corridors. Overall, the results demonstrate that Dar es Salaam has undergone rapid urban expansion over the past three decades, with machine-learning integrated approaches—such as RF-CA—providing more accurate and spatially realistic predictions than traditional models, such as CA-Markov.

5. Discussion

5.1 Model Performance and Comparative Advantages

A comparative analysis of the CA-Markov and RF-CA models revealed significant differences in their predictive capabilities for forecasting urban sprawl in Dar es Salaam, Tanzania. The results indicate that the hybrid RF-CA model outperformed the traditional CA-Markov model, achieving higher accuracy (77.84% vs 69.38%), Figure of Merit (FoM: 56.83% vs 51.33%), and Kappa coefficient (0.67 vs 0.542). However, the magnitude of improvement should be studied alongside the study's single-city scope discussed in Section 5.4.

This improvement aligns with the findings of previous studies that employed hybrid RF-based models. For instance, Kamusoko and Gamba (2015) applied the RF-CA model to binary urban predictions; and reported a Kappa of 0.51 and FoM of 47%, which outperformed Support Vector Machine-Cellular Automata (SVM-CA; Kappa 0.39, FoM 39%) and Logistic Regression-Cellular Automata (LR-CA; Kappa 0.22, FoM 6%), even in simplified two-class setups.

Further spatial and quantitative comparisons of the 2024 predicted values highlight these advantages. The actual observed urban area was 505.01km². The CA-Markov model predicted 457.30km², underestimating the urban area by approximately 47.71km²; whereas the RF-CA model predicted 441.64km², a slightly lower estimate, but with superior overall spatial accuracy. Notably, the RF-CA model correctly identified 343.05km² of urban land compared to 326.43km² identified by the CA-Markov model.

Such deviations suggest that the CA-Markov model may underestimate the rate of land transformation because it relies primarily on historical transition probabilities, and does not fully incorporate recent socioeconomic drivers or dynamic feedback mechanisms into its calculations. In contrast, the RF-CA model, which integrates updated geospatial and temporal variables, captures the nonlinear and rapid nature of contemporary urban growth in Dar es Salaam more effectively. This underscores the value of hybrid modelling approaches for rapidly urbanising areas.

5.2 Key Drivers of Urban Sprawl

A feature importance analysis using the RF-CA model provided valuable insights into the primary drivers of urban expansion in Dar es Salaam. Population density was the most influential factor, accounting for 23.06% of the impurity feature importance. This indicates a strong association between population density and the probability of urban transition. This finding aligns with urban growth theories, and reflects the city's sustained population growth, which has more than tripled since the late 1980s and had reached approximately 5.3 million by 2022 (URT, 2022).

Topographic constraints also played a significant role, with an elevation of 13.19%, indicating that urban development was directed toward flatter and more accessible areas. Additionally, climate variables contributed to the model's prediction of urban transition—with temperature and rainfall contributing 11.75% and 11.45%, respectively—highlighting the influence of environmental factors on urban planning decisions in tropical regions, such as Dar es Salaam.

Comparisons with previous studies further contextualise this rapid growth. Bhanjee and Zhang (2018) documented that built-up areas more than doubled from 255.06km² in 1990 to 532.13km² in 2014 (an expansion of 108.6%). Similarly, the present study's observed urban area of approximately 505.01km² by 2024 compared with the earlier projections by Simon, Lyimo and Yamungu (2023), who forecasted a built-up area of only 530km² by 2050 using a CA-Markov model calibrated from 1995 to 2022. These findings emphasise the accelerated pace of urban expansion in Dar es Salaam beyond historical trends, reinforcing the need for models that incorporate dynamic drivers.

5.3 Implications for Urban Planning and Policy

The rapid and sustained urban growth revealed by the RF-CA model has significant implications for urban planning and policy in Dar es Salaam. The identification of key drivers—particularly population density, topography, and climate—suggests that planning strategies should prioritise managing demographic pressures, protecting vulnerable low-elevation areas, and incorporating climate-resilient designs.

The superior performance of the RF-CA model and its ability to capture nonlinear growth patterns highlight the importance of adopting hybrid machine-learning-integrated approaches, and regularly updating datasets. This is particularly crucial in fast-growing cities where traditional models, such as the CA-Markov model, may fail to account for recent acceleration of sprawls. Proactive policies informed by advanced simulations can support sustainable land use allocation, infrastructure development, and environmental conservation to mitigate uncontrolled expansion.

5.4 Limitations and Future Research Directions

The study that generated the data for this paper relied on standard accuracy metrics (overall accuracy, Kappa, FoM, and F1-score) for model evaluation and 30-m Landsat imagery, which limited the detection of fine-scale changes, such as informal settlements. Resampling population and climate datasets may introduce uncertainty, and the CA-Markov model's assumption of temporal stationarity is problematic in rapidly changing urban contexts. Socioeconomic and policy-driven factors have not been fully incorporated into this model. Accuracy assessment also relied on a random pixel-based training/testing split, which does not guarantee spatial independence between samples, given the spatial autocorrelation among neighbouring pixels, potentially leading to modest overestimation of the reported accuracy, Kappa, and F1 score values. Additionally, feature importance rankings from the RF-CA model reflect impurity-based rankings from a single fitted model, and their stability across random-forest seeds and sensitivity to correlations among predictors were not evaluated; hence, a permutation-based analysis would strengthen future interpretation of driver effects.

Moreover, future research could strengthen validation by employing per-class confusion matrices, spatial error maps, and null model comparisons to identify prediction errors better. Incorporating dynamic socioeconomic variables, extending projections to longer horizons, and exploring ensemble modelling techniques would further refine forecasts in similar rapidly urbanising settings. Additionally, detailed investigations—using advanced spatial analytics, landscape metrics, and multitemporal data—into specific sprawl morphologies—such as leapfrog development, edge expansion, and infill patterns—could elucidate the drivers of discontinuous growth and governance influences on urban form. Also, integrating quantitative analyses with qualitative stakeholder insights and machine learning for automated pattern classification provides targeted strategies to complement broader modelling approaches.

6. Conclusion

This study provided a comprehensive comparison of the CA-Markov and RF-CA models for predicting urban sprawl in Dar es Salaam, Tanzania, using multitemporal Landsat imagery and key biophysical and socioeconomic drivers from 1995 to 2024. The RF-CA model results demonstrate a better predictive performance, which achieved higher predictive accuracy (77.84% vs 69.38%), Kappa coefficient (0.67 vs 0.542), and Figure of Merit (56.83% vs 51.33%) than the traditional CA-Markov approach. The RF-CA model captured the nonlinear spatial complexities of rapid urban expansion more effectively, particularly in peri-urban areas; and identified population density, elevation, and climatic factors as the primary drivers of growth.

These findings underscore the value of integrating machine-learning with cellular automata frameworks for urban growth modelling in data-scarce and rapidly urbanising African cities, where traditional stationary models are often inadequate. For urban planners and policymakers in Dar es Salaam, adopting RF-CA can enable more reliable forecasts to guide proactive infrastructure planning, zoning regulations, and targeted interventions in high-growth zones—such as Ubungo, Kimara, and Mbezi—thereby curbing uncontrolled sprawl and informal settlements. Environmental agencies can leverage these predictions to monitor and mitigate ecological impacts, including vegetation loss and wetland encroachment.

More broadly, this study highlights the potential of hybrid geospatial-machine-learning approaches to support sustainable urban development amid Africa's unprecedented urbanisation, thereby offering a methodological template that can be tested and adapted for other fast-growing SSA cities facing similar data and planning constraints.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. & Hegewisch, K. C. (2018). TerraClimate, A High-resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958–2015. *Scientific Data*, 5(1): 1–12. https://doi.org/10.1038/sdata.2017.191.

Asif, M. et al. (2023). Modelling of Land Use and Land Cover Changes and Prediction Using CA-Markov and Random Forest. *Geocarto International*, 38(1). https://doi.org/10.1080/10106049.2023.2210532.

Atef, I., Ahmed, W. & Abdel-Maguid, R.H. (2024). Future Land Use Land Cover Changes in El-Fayoum Governorate: A Simulation Study Using Satellite Data and CA-Markov Model. *Stochastic Environmental Research and Risk Assessment*, 38(2). https://doi.org/10.1007/s00477-023-02592-0.

Augustijn-Beckers, E.W., Flacke, J. & Retsios, B. (2011). Simulating Informal Settlement Growth in Dar es Salaam, Tanzania: An Agent-based Housing Model. *Computers, Environment and Urban Systems*, 35(2). https://doi.org/10.1016/j.compenvurbsys.2011.01.001.

Azabdaftari, A. & Sunar, F. (2024). Predicting Urban Tomorrow: CA-Markov and District Evolution. *Earth Science Informatics*, 17(4). https://doi.org/10.1007/s12145-024-01340-4.

Batty, M., Xie, Y. & Sun, Z. (1999). The Dynamics of Urban Sprawl. *Centre for Advanced Spatial Analysis*, 15.

Behnisch, M., Krüger, T. & Jaeger, J.A.G. (2022). Rapid Rise in Urban Sprawl: Global Hotspots and Trends Since 1990. *PLOS Sustainability and Transformation*, 1(11). https://doi.org/10.1371/journal.pstr.0000034.

Bhanjee, S. & Zhang, C.H. (2018). Mapping Latest Patterns of Urban Sprawl in Dar es Salaam, Tanzania. *Papers in Applied Geography*, 4(3). https://doi.org/10.1080/23754931.2018.1471413.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1). https://doi.org/10.1023/A:1010933404324.

Fitawok, M.B. & Minale, A.S. (2024). A Review of the Application and Implications of Cellular Automata-based Urban Growth Model in Africa. *South African Geographical Journal*, 106(3). https://doi.org/10.1080/03736245.2024.2308909.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D. H., Verdin, J. P., Shukla, S., Husak, G. J., Rowland, J., Harrison, L., Hoell, A. & Michaelsen, J. (2015). The Climate Hazards Infrared Precipitation with Stations - A New Environmental Record for Monitoring Extremes. *Scientific Data*, 2. https://doi.org/10.1038/sdata.2015.66.

Gharaibeh, A. A., Al-Shboul, A. A., Al-Kofahi, S. & Al-Zu'bi, M. (2020). Improving Land-use Change by Integrating ANN with Cellular Automata-Markov Chain Model. *Heliyon*, 6(9). https://doi.org/10.1016/j.heliyon.2020.e05092.

Hamad, R., Balzter, H. & Kolo, K. (2018). Predicting Land Use/Land Cover Changes Using a CA-Markov Model Under Two Different Scenarios. *Sustainability (Switzerland)*, 10(10). https://doi.org/10.3390/su10103421.

Jayabaskaran, M. & Das, B. (2023). Land Use Land Cover (LULC) Dynamics by CA-ANN and CA-Markov Model Approaches: A Case Study of Ranipet Town, India. *Nature Environment and Pollution Technology*, 22(3). https://doi.org/10.46488/NEPT.2023.v22i03.013.

Kamusoko, C. & Gamba, J. (2015). Simulating Urban Growth Using a Random Forest-cellular Automaton (RF-CA) Model. *ISPRS International Journal of Geo-Information*, 4(2). https://doi.org/10.3390/ijgi4020447.

Khawaldah, H.A., Farhan, I. & Alzboun, N.M. (2020). Simulation and Prediction of Land Use and Land Cover Change Using GIS, Remote Sensing and CA-Markov Model. *Global Journal of Environmental Science and Management*, 6(2). https://doi.org/10.22034/gjesm.2020.02.07.

Laros, M., & Jones, F. (Eds.). (2014). *The State of African Cities 2014: Re-imagining Sustainable Urban Transitions*. UN-Habitat.

Li, M., Cao, Y., Dai, J., Song, J., & Liang, M. (2025). A Comprehensive Review of Urban Expansion and its Driving Factors. *Land*, 14(8). https://doi.org/10.3390/land14081534.

Mberu, B., Béguy, D. & Ezeh, A.C. (2017). Internal Migration and Slums in sub-Saharan Africa. *Africa's Population: In Search of a Demographic Dividend*. https://doi.org/10.1007/978-3-319-46889-1_20.

Mnyali, E.T. & Materu, S.F. (2021). Analysis of the Current and Future Land Use/Land Cover Changes in Peri-Urban Areas of Dar es Salaam City, Tanzania using Remote Sensing and GIS Techniques. *Tanzania Journal of Science*, 47(5). https://doi.org/10.4314/tjs.v47i5.12.

Musa, S.I., Hashim, M. & Reba, M.N.M. (2017). A Review of Geospatial-based Urban Growth Models and Modelling Initiatives. *Geocarto International*. https://doi.org/10.1080/10106049.2016.1213891.

Nugroho, F. & Al-Sanjary, O.I. (2018). A Review of Simulation Urban Growth Model. *International Journal of Engineering and Technology (UAE)*, 7(4). https://doi.org/10.14419/ijet.v7i4.11.20681.

Pontius, R.G. & Millones, M. (2011). Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431161.2011.552923.

Sang, L. et al. (2011). Simulation of Land Use Spatial Pattern of Towns and Villages Based on CA-Markov Model. *Mathematical and Computer Modelling*, 54(3–4). https://doi. org/10.1016/j.mcm.2010.11.019.

Schiavina, M., Freire, S., Carioli, A. & MacManus, K. (2023). *GHS-POP R2023A: GHS Population Grid Multitemporal (1975–2030)* [Data set]. European Commission, Joint Research Centre. doi.org.

Seto, K.C., Güneralp, B. & Hutyra, L.R. (2012). Global Forecasts of Urban Expansion to 2030 and Direct Impacts On biodiversity and Carbon Pools. *Proceedings of the National Academy of Sciences of the United States of America*, 109(40). https://doi.org/10.1073/pnas.1211658109.

Shaamala, A., Yigitcanlar, T., Nili, A. & Nyandega, D. (2025). Machine Learning Applications for Urban Geospatial Analysis: A Review of Urban and Environmental Studies. *Cities*, 165. https://doi.org/10.1016/j.cities.2025.106139.

Simon, O., Lyimo, J. & Yamungu, N. (2023). Land Use and Cover Change in Dar es Salaam Metropolitan City: Satellite Data and CA-Markov Chain Analysis. *GeoJournal*, 88(6). https://doi.org/10.1007/s10708-023-10960-0.

Spencer, D. (2009). Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. *The Journal of Architecture*, 14(3). https://doi.org/10.1080/13602360903028044.

Tobler, Waldo. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography (Supplement: Proceedings, International Geographical Union, Commission on Quantitative Methods. *Economic Geography* 11(46).

Tumwesigye, S., Vanmaercke, M., Hemerijckx, L.-M., Opio, A., Poesen, J., Twongyirwe, R. & Van Rompaey, A. (2023). Spatial Patterns of Urbanisation in Sub-Saharan Africa: A Case Study of Uganda. *Development Southern Africa*, 40(1): 1–21. https://doi.org/10. 1080/0376835X.2021.1932426.

United Republic of Tanzania (URT). (2022). *The 2022 Population and Housing Census: Administrative Units Population Distribution Report*, *National Population and House Census of Tanzania.* National Bureau of Statistics, Dar es Salaam, Tanzania.

Wang, S.W., Munkhnasan, L. & Lee, W.K. (2021). Land Use and Land Cover Change Detection and Prediction in Bhutan's High-altitude City of Thimphu, Using Cellular Automata and Markov Chain. *Environmental Challenges*, 2. https://doi.org/10.1016/j.envc.2020.100017.

Wang, X., Shi, R. & Zhou, Y. (2020). Dynamics of Urban Sprawl and Sustainable Development in China. *Socio-Economic Planning Sciences*, 70. https://doi.org/10.1016/j.seps.2019.100736.

White, R. & Engelen, G. (1993). Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-use Patterns. *Environment & Planning A*, 25(8). https://doi.org/10.1068/a251175.

Appendix

Confusion Matrices

The Appendix reports the year-specific error (confusion) matrices underlying the Random Forest land cover classification reported in Section 3.2. A set of reference points was interpreted from high-resolution basemap imagery and compared with the classified map. Producer's accuracy (omission error) is read along each row, and user's accuracy (commission error) is read along each column. The overall accuracy and kappa coefficient computed from each matrix are given below the corresponding table.

Table A1: 1995 RF Land-Cover Classification

Reference \ Predicted	**Water**	**Vegetation**	**Built-up**	**Bare land**	**Row Total**	**Producer's Accuracy**
Water	9	1	0	0	10	90.0%
Vegetation	1	40	2	3	46	87.0%
Built-up	2	4	36	2	44	81.8%
Bare land	0	1	1	18	20	90.0%
Column Total	**12**	**46**	**39**	**23**	**120**	
User's Accuracy	**75.0%**	**87.0%**	**92.3%**	**78.3%**		

Overall Accuracy = 85.83%
Kappa = 0.796

Table A2: 2004 Rf Land-Cover Classification

Reference\ Predicted	**Water**	**Vegetation**	**Built-up**	**Bare land**	**Row Total**	**Producer's Accuracy**
Water	11	3	0	0	14	78.6%
Vegetation	3	39	3	2	47	83.0%
Built-up	0	1	35	1	37	94.6%
Bare land	0	5	1	16	22	72.7%
Column Total	**14**	**48**	**39**	**19**	**120**	
User's Accuracy	**78.6%**	**81.3%**	**89.7%**	**84.2%**		

Overall Accuracy = 84.17%
Kappa = 0.774

Table A3: 2014 Rf Land-Cover Classification

Reference \ Predicted	Water	Vegetation	Built-up	Bare land	Row Total	Producer's Accuracy
Water	13	0	0	0	13	100.0%
Vegetation	4	33	4	3	44	75.0%
Built-up	3	5	37	1	46	80.4%
Bare land	0	0	0	17	17	100.0%
Column Total	**20**	**38**	**41**	**21**	**120**	
User's Accuracy	**65.0%**	**86.8%**	**90.2%**	**81.0%**		

Overall Accuracy = 83.33%
Kappa = 0.76

Table A4: 2024 Rf Land-Cover Classification

Reference \ Predicted	Water	Vegetation	Built-up	Bare land	Row Total	Producer's Accuracy
Water	14	1	0	0	15	93.3%
Vegetation	0	39	3	1	43	90.7%
Built-up	2	4	34	4	44	77.3%
Bare land	0	0	0	18	18	100.0%
Column Total	**16**	**44**	**37**	**23**	**120**	
User's Accuracy	**87.5%**	**88.6%**	**91.9%**	**78.3%**		

Overall Accuracy = 87.50%
Kappa = 0.824