

Article

Land-Use Change Prediction in Dam Catchment Using Logistic Regression-CA, ANN-CA and Random Forest Regression and Implications for Sustainable Land–Water Nexus

Yashon O. Ouma ^{1,*} , Boipuso Nkwae ¹, Phillimon Odirile ¹ , Ditiro B. Moalafhi ² , George Anderson ³, Bhagabat Parida ⁴ and Jiaguo Qi ⁵ 

¹ Department of Civil Engineering, University of Botswana, Gaborone Private Bag UB0061, Botswana; nkwaeb@ub.ac.bw (B.N.); odirilep@ub.ac.bw (P.O.)

² Faculty of Natural Resources, BUAN, Gaborone Private Bag 0027, Botswana; dmoalafhi@buan.ac.bw

³ Department of Computer Science, University of Botswana, Gaborone Private Bag UB0061, Botswana; andersong@ub.ac.bw

⁴ Department of Civil and Environmental Engineering, BIUST, Palapye Private Bag 16, Botswana; paridab@biust.ac.bw

⁵ Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI 48824, USA; qi@msu.edu

* Correspondence: oumay@ub.ac.bw

Abstract: For sustainable water resource management within dam catchments, accurate knowledge of land-use and land-cover change (LULCC) and the relationships with dam water variability is necessary. To improve LULCC prediction, this study proposes the use of a random forest regression (RFR) model, in comparison with logistic regression–cellular automata (LR-CA) and artificial neural network–cellular automata (ANN-CA), for the prediction of LULCC (2019–2030) in the Gaborone dam catchment (Botswana). RFR is proposed as it is able to capture the existing and potential interactions between the LULC intensity and their nonlinear interactions with the change-driving factors. For LULCC forecasting, the driving factors comprised physiographic variables (elevation, slope and aspect) and proximity-neighborhood factors (distances to water bodies, roads and urban areas). In simulating the historical LULC (1986–2019) at 5-year time steps, RFR outperformed ANN-CA and LR-CA models with respective percentage accuracies of 84.9%, 62.1% and 60.7%. Using the RFR model, the predicted LULCCs were determined as vegetation (−8.9%), bare soil (+8.9%), built-up (+2.49%) and cropland (−2.8%), with water bodies exhibiting insignificant change. The correlation between land use (built-up areas) and water depicted an increasing population against decreasing dam water capacity. The study approach has the potential for deriving the catchment land–water nexus, which can aid in the formulation of sustainable catchment monitoring and development strategies.

Keywords: land-use land-cover (LULC) change; logistic regression; artificial neural network; cellular automata; random forest regression; sustainable land–water nexus



Citation: Ouma, Y.O.; Nkwae, B.; Odirile, P.; Moalafhi, D.B.; Anderson, G.; Parida, B.; Qi, J. Land-Use Change Prediction in Dam Catchment Using Logistic Regression-CA, ANN-CA and Random Forest Regression and Implications for Sustainable Land–Water Nexus. *Sustainability* **2024**, *16*, 1699. <https://doi.org/10.3390/su16041699>

Academic Editor: Hariklia D. Skilodimou

Received: 6 January 2024

Revised: 21 January 2024

Accepted: 5 February 2024

Published: 19 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this Anthropocene era where human activities are forcing significant changes in the environment, rising incomes together with the growing demands for goods and services, including food, energy, water and land, have resulted in increased pressures on natural resources and ecosystems, leading to their overexploitation and degradation. This predicament is further fueled by climate change. For sustainable development, systematic integrated solutions are necessary. This can be achieved in part through nexus-based cross-sectoral links between resources for increased and efficient exploitation of resources at local, national and global scales. For catchments, especially dam water catchments, the determination of accurate links between land use and water resources is very significant [1,2].

Focusing on the sustainability assessment of a land–water nexus within catchment ecosystems, it is necessary to apply an integrated approach for accurate identification of land-use and land-cover changes, future predictions and their impacts on water resources and subsequently the water budget.

Land is the main element of the natural environmental substrate and is an important carrier of socio-economic development [3]. Land use refers to the human-based utilization of the Earth's surface, while land cover comprises the natural biological, physical and terrestrial surfaces [4]. Changes in land use and land cover (LULC) are mainly attributed to urbanization, infrastructural developments, industrialization, intensification of agriculture and irrigation and overexploitation of grassland and forest covers [5]. Different studies have shown that LULC changes are critical for the assessment of environmental impacts and in determining the relationships with hydrological and ecological issues [6]. Recently, more attention has been given to LULC change due to its extreme impacts on the environment at the basin or watershed levels, as well as its contributions to climate change [7].

Rationalized use and management of land are some of the crucial ways of resolving socio-economic conflicts and contribute to the harmonization of environment, society and economy [8]. However, with increasing population and rapid urbanization, regional and global LULC has undergone enormous changes in the past decades. The modifications of the biophysical systems and the alterations of the land surface energy processes through LULC changes result in fluctuations in surface energy balance, leading to variabilities in climate, which have in turn become responsible for several adverse impacts on different environmental phenomena [9,10]. The changes in LULC are thus critical driving factors for several environmental problems, including enhanced carbon emissions [11,12], biodiversity loss [13], ecosystem productivity decrease [14], soil and land degradation [15] and ecosystem service decline [16]. This implies that LULC change detection and studies play a significant role in monitoring land as a critical resource, how it is transformed naturally and/or due to human activities and its impacts on other resources.

Consequently, changes in LULC are linked to several of the United Nations' (UN) Sustainable Development Goals (SDGs). Specifically, SDG 15 on Life on Land, which demands to "protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation" [17]. Further, the realization of SDG 11 on Sustainable Cities and Communities is based on informed development that relies on accurate land-use information [18,19]. Geospatial information on LULC empowers decision makers to monitor and address environmental challenges, promote sustainable development, ensure inclusive, safe, resilient and sustainable cities as envisioned by SDG 11 [20], implement evidence-based policies and ensure the resilience and sustainability of cities and societies [21]. For sustainable environmental planning, monitoring, identification of possible threats and formulation of mitigation and adaptation strategies, past, present and future LULC information is necessary [8,9].

For LULC mapping and LULC change prediction, a significant amount of research has been based on geospatial modeling using remote sensing (RS) data and GIS techniques [22,23]. The other commonly used approaches for LULC change predictions include models based on: statistics [24]; evolutionary approaches [25]; cellular automata [26]; Markov chain [27]; hybrid models [28]; expert systems [29] and multiagent models [30]. Compared to the RS- and GIS-based methods, the cellular automata (CA)–Markov chain (MC) model is the most widely used approach for LULC change prediction [31,32].

While MC analysis has been used in several case studies for the simulation and prediction of LULC changes, it is best suited for short-term projections [33] due to the fact that it is not spatially explicit, as it works under the physics assumption that the future state depends only on the current state while ignoring the spatial information within LULC classes [34]. Furthermore, its analysis does not offer a geographical dispersal of the verification, which is essential in comprehending the possible influence of predicted variations. Hence, MC can determine the right magnitude of LULC change without the correct change direction. This drawback has been mitigated by creating a superior hybrid

MC-CA model that combines the MC with the empirical CA [28]. A bottom-up dynamic model, CA incorporates a spatial dimension, thus adding direction modeling in LULC change modeling. The MC-CA model has been widely used to understand urban expansion and landscape dynamics in several studies [35–39].

Despite the potential and wide use of the MC-CA model, its adoption presents a hurdle due to the assumption of stationary transitions, making the approach more suited for short-term predictions than long-term projections [40]. Furthermore, MC-CA simulations present uncertainty in the output patterns resulting from the sensitivity of the model to the cell size of the input grid data and to the shapes of the different neighborhood types [41]. Additionally, the self-adaptive ability of MC-CA for modeling nonlinear relationships between the dynamic LULC classes and the change drivers is still not completely reliable [42]. These drawbacks in the MC-CA model can be solved by integrating the model with other dynamic and empirical models [43]. Some of the approaches which have been integrated with MC-CA include multicriteria analysis (MCA), analytical hierarchy process (MCA-AHP) and frequency ratio (FR) [44]; multivariable statistical methods (MSMs) such as linear regression (LiR) [45]; logistic regression (LR) [46]; machine learning methods [47] and artificial neural networks (ANNs) [48]. Nonetheless, due to the complexity of studying land-use change, most of the proposed hybridized models have limitations that include insufficient knowledge about the area of interest and subjectivity in weighting the variables and they cannot model nonlinear phenomena, resulting in unreliable prediction results [49].

In applying the different LULC change prediction models, ref [50] utilized hybrid LR-MC-CA and obtained an accuracy of 89% between simulated and actual land use for case study of Tehran. For urban growth modeling, ref [51] employed RF-CA and SVM-CA models which exhibited a high certainty for modeling complex urban growth in the Wallonia region (Belgium). For forecasting LULC change in Attica (Greece), ref [52] applied RF-CA with notably high overall accuracy of 88.4%. [53] employed the ANN-CA-Markov model for the simulation of urban growth in China with a prediction suitability score of 0.864 and kappa coefficient of 0.78. Similar accuracy of 84% was obtained by [54] using ANN-CA for spatiotemporal analysis and simulations of biophysical indicators under urbanization and climate change scenarios. In another study, ref [55] predicted urban land-use change in Bogota up to 2034 using the MC-ACC with an average validation value of 0.85. In comparing different machine learning models, ref [49] compared different machine learning models, ANN, SVR, RF, decision tree (DT), LR and multivariate adaptive regression splines (MARS), for the prediction of urban growth with the following results. The prediction accuracy results ranged between 68% (LR) and 75% (ANN). A detailed review on the application of ML models for spatiotemporal simulation and prediction of land-use change can be found in Aburas et al. [44]. In addition to the inherent drawbacks with the LULC prediction models, most of the models are developed and tested for well-structured urban landscapes with proper development plans [51]. They may therefore be inapplicable for catchment landscapes with complex and unsystematic evolutions in the spatial–temporal dynamics of land-use change [56].

In the recent past, machine learning (ML) models such as random forest (RF), ANN and support vector machine (SVM) have been proposed for the identification of transitions in land-use classes and for the prediction of LULC changes [51]. ML, with its capabilities such as high learning rates and big data analytics, is able to enhance the data-fitting accuracies, thus producing more intelligent and robust simulations and predictions [57]. Compared to the traditional empirical, statistical and dynamic approaches, machine learning models are more flexible with the ability for efficient learning and processing, do not consider problems related to multivariate covariances and can handle the nonlinearity in LULC data [58]. As such, ML models have the potential to improve the simulation of unsystematic and complex spatial–temporal dynamics in LULC and the prediction of LULC change type and intensity within catchment landscapes. Despite their advantages, the application of ML models for the simulation and prediction of LULC change is still very limited.

For multiple predicted factors, the uncertainty of the inspection variable is influenced by the LULC type and its intensity. This poses a nonlinearity connectivity condition between the LULC variables. To improve the LULC change detection and prediction, this study proposes the simulation and prediction of LULC change using the random forest regression (RFR) ML approach for the prediction of LULC changes in the Gaborone dam catchment (GDC) in Botswana. Catchment management is important not only as a hydrological unit but also as part of the socio-ecological environment that is responsible in part for economic activities and food production and as social security for the residents [59]. However, LULC changes within catchments through, for example, urbanization and deforestation contribute to negative impacts on water quality and water availability and indirectly influence the nature of a watershed ecosystem. This means that, for catchments, the mapping and analysis of the spatiotemporal variations occurring within the watersheds and their interactions with the catchment hydrological components will enhance not only the formulations of water conservation strategies but also overall sustainable planning and management of the catchment. RFR is a nonlinear model suitable for the regression of datasets that do not exhibit linearity. RFR is considered to be able to capture the existing and potential interactions between the type and intensity of LULC classes and the interactions with the influencing or driving factors [47]. The proposed RFR model is compared with two hybrid models, logistic regression-CA (LR-CA) and ANN-CA. LR-CA is proposed due to its ability to take the dynamic process of LULC changes into consideration [60] and can connect the categorical variables and the continuous variables and build the potential relationships between them [61]. Ref. [62] established an LR-CA framework for the prediction of urban changes in the Istanbul metropolitan area, while [63] explored the availability of an integrated LR-CA in predicting the future LULC in the Hamilton area of Ohio in the USA. On the other hand, the ANN has the advantage of being nonparametric and requires minor or no prior knowledge of the distribution of input data in the estimation of nonlinear relationships. In adopting the ANN-CA, the self-organizing and adaptive approach of the ANN can make its integration with CA more self-adaptive than the traditional models [64]. Because of their self-learning ability, the unknown relationships among variables can then be addressed with the automatic approximation of nonlinear functions using the ANN which is considered to be superior to linear-based regression models like LR and AHP [65]. Several researchers have also reported success in applying ANNs for LULC modeling in different regions [66–68].

By simulating the spatiotemporal LULC change within the GDC in Botswana for three decades from 1986 to 2019 at 5-year intervals and incorporating the LULC change drivers comprising physiographical variables (elevation (DEM), slope and aspect) and neighborhood proximity factors (distances to water bodies, roads and urban areas), the objectives of the study are to: (1) derive the spatiotemporal LULC change trends within the Gaborone dam catchment from 1986 to 2019 using machine learning and multiple features; (2) develop an RFR model that integrates transition mapping with driving factors for LULC change prediction within the GDC study area; (3) compare the performance of the proposed RFR-LULC change prediction model with the parametric LR-CA and nonparametric ANN-CA hybrid models for the simulation and prediction of short-term LULC change in GDC for 2025 and 2030 and (4) analyze the relationship between land-use change and dam water variability within the Gaborone dam catchment.

In addition to improving the LULC mapping using machine learning with multiple classification features comprising multispectral bands, vegetation and soil indices and GLCM texture cues, the proposed RFR for LULC change simulation is a novel approach aimed at overcoming the drawbacks in the parametric LR-CA and nonparametric ANN-CA hybrid models especially for complex catchment landscapes. By testing the efficacy of machine learning in simulating and predicting LULC change within dam catchment environments, and for a case study where the proposed approach has not been applied before, this study offers a benchmark for applying machine learning in such environments, with the goal of providing accurate future LULC developments in support of land-use

planning for sustainability [69]. In addition, the current research aligns with the aim and scope of SDG 11 and SDG 15 as it highlights the significance of interdisciplinary studies, innovative approaches and technological advancements in promoting integrated resource monitoring and management for sustainable development at different spatial scales.

2. Materials and Methods

2.1. Study Area

Gaborone dam catchment (GDC) is a sub-catchment in Botswana's Limpopo River Basin (BLRB). BLRB is part of the larger Limpopo River Basin (LRB), located in southern Africa, and is a transboundary basin encompassing portions of Botswana, Mozambique, South Africa and Zimbabwe. It is formed by several sub-catchments, with GDC being one of them (Figure 1). GDC is a sub-catchment within Botswana's Limpopo River Basin (BLRB) in the southeastern part of Botswana and lies between latitudes $25^{\circ}16'$ E to $26^{\circ}30'$ E and longitudes $24^{\circ}42'$ S to $25^{\circ}34'$ S. The catchment overlaps the northwest side of South Africa (Figure 1) and has a total area of approximately 4344 km². It is one of the upstream sub-catchments within the Limpopo River Basin, with Notwane River as the main drainage running from South Africa through the Gaborone dam to the Limpopo River. The catchment has an average altitude of approximately 1292 m AMSL and average temperatures of between 19.7 °C and 32.7 °C and annual average rainfall of approximately 500 mm. The dam catchment is one of the main water catchments and also one of the most densely populated catchments in the arid and semiarid Botswana. The catchment is thus considered as environmentally sensitive resulting from the characteristics of the fragile dry desert environment and its susceptibility to rapid successive changes due to human activities and climate change. Catchment dynamics as influenced by human activities and natural phenomena like climate changes results in LULC changes that make LULC prediction an inevitable step in resource planning and management.

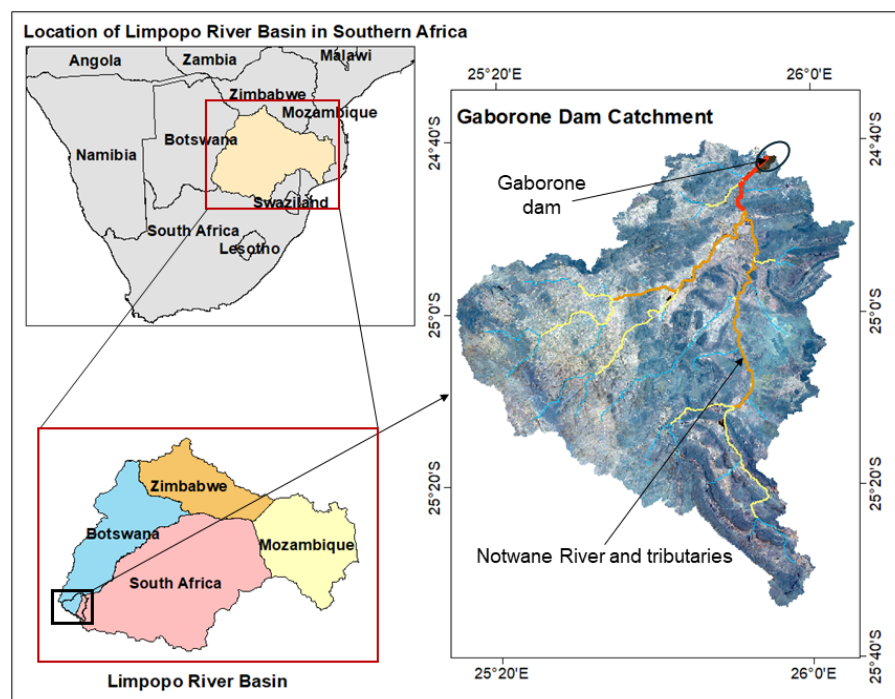


Figure 1. Location map of the transboundary Limpopo River Basin (LRB), Botswana's LRB (BLRB) and the Gaborone dam catchment image from Landsat 8 data overlaid with Notwane River.

2.2. Data

For LULC mapping, Landsat 5 TM was used for 1986, 1989, 1994, 1999, 2004 and 2009, and Landsat 8 OLI was utilized for 2014 and 2019. The Landsat data were downloaded

from the United States Geological Survey Earth Explorer (www.earthexplorer.usgs.gov). To minimize the seasonality effects, the datasets were acquired during the same time of year in April and May. Cloudless images for the study years were filtered for the month using Google Earth Engine (GEE). The pre-processing of the data comprised geometric correction, mosaicking of the different paths and rows and atmospheric and radiometric correction. Geometric correction involved orienting the imagery to the local coordinate projection system, and to minimize radiometric errors, the radiometric tool in ERDAS Imagine was used to calibrate the satellite images. The radiometric correction process involves the conversion of the digital number (DN) as raw data from sensors to top-of-atmosphere reflectance as actual ground surface reflectance. The multitemporal images were all calibrated for top-of-atmosphere (TOA) reflectance. Atmospheric effects restrict the dynamic range with the image having haziness and low contrast. The atmospheric effects on the satellite images were corrected using a haze reduction tool in ERDAS Imagine, based on dark object subtraction (DOS). The multitemporal images were radiometrically calibrated by converting raw digital numbers (DNs) to sensor spectral radiance (L_λ).

The LULC change drivers are usually characterized by proximity-neighborhood factors (distance to major towns and cities, roads, rain networks, water bodies, streams); topography (elevation, slope, aspect) and demographic variables (population and population density). The driver variables are spatially and temporally multifaceted. The LULC change drivers used in the LULC simulation modeling included natural physiographical variables (elevation (DEM), slope and aspect), and neighborhood-proximity factors (distances to water bodies, roads and urban areas) (Table 1). The datasets and the data sources are presented in Table 1. The datasets were recategorized and resampled to 30 m spatial resolution as presented in Figure 2.

Table 1. Description of LULC change driving factors in the Gaborone dam catchment.

Factor Category	Data	Data Sources/Spatial Resolution	Data Description	Units
Natural topographic factors	Elevation	ALOS-PALSAR DEM	DEM	m
	Slope	https://search.asf.alaska.edu/#/ (accessed on 24 November 2022)	Range from 0 to 90	Degrees (°)
	Aspect	12.5 m × 12.5 m	Range from 0 to 360	Degrees (°)
Proximity-neighborhood factors	Distance to water bodies		Euclidean distance to water bodies	km
	Distance to roads	Digitized from Google Earth image https://earth.google.com/web/ (accessed on 12 July 2023)	Euclidean distance to roads and railway lines	km
	Distance to urban areas		Euclidean distance to significant residential points	km
Land use and land cover	Land use land cover (LULC) from Landsat TM and ETM+	Landsat data https://earthexplorer.usgs.gov/ (accessed on 24 November 2022) 30 m × 30 m	Land use land cover for respective years	km ² ; %

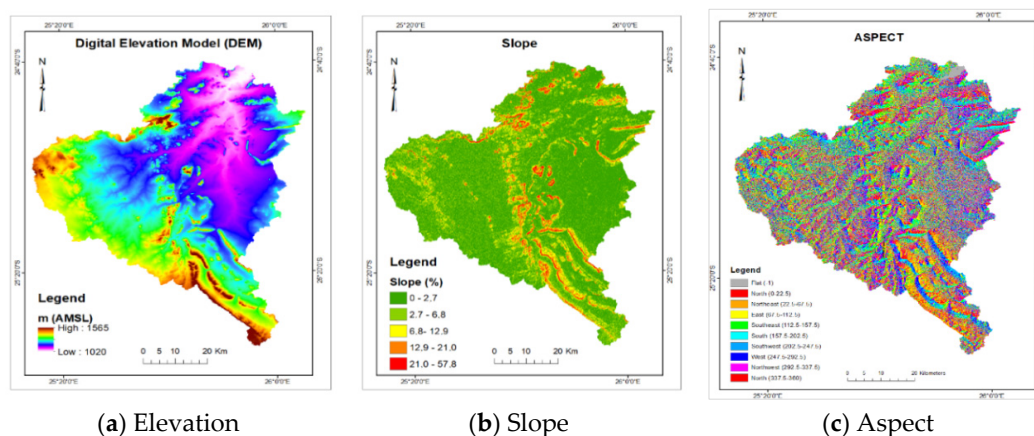


Figure 2. Cont.

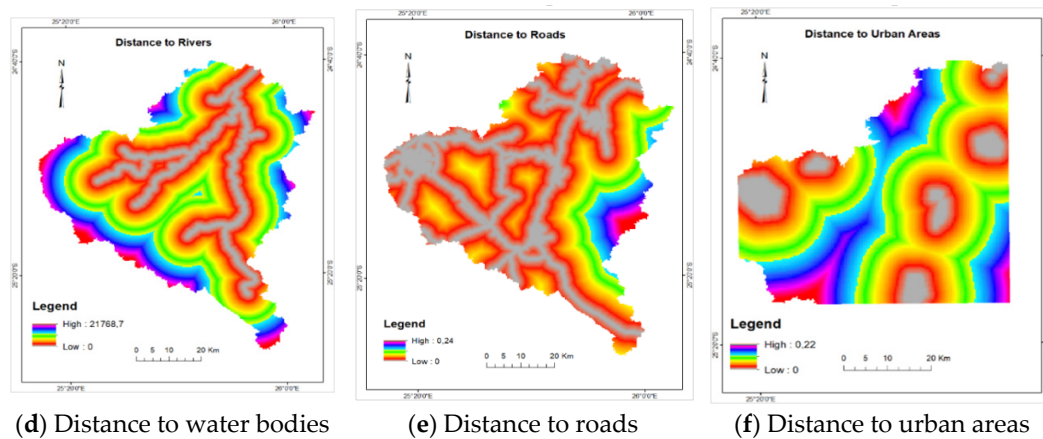


Figure 2. LULC change driving factors for Gaborone dam catchment.

2.3. Methods

The summary flow diagram for the implementation of the LULC simulation, prediction and analysis for the land–water analysis nexus is presented in Figure 3. The implementation is in four phases: (i) LULC mapping and derivation of LULC change-driving factors, (ii) generation of transition probability and transition maps and enforcing the neighborhood influence, (iii) prediction for future LULC scenarios for 2025 and 2030 and (iv) investigating the relationship between LULC change and dam water variability within the catchment.

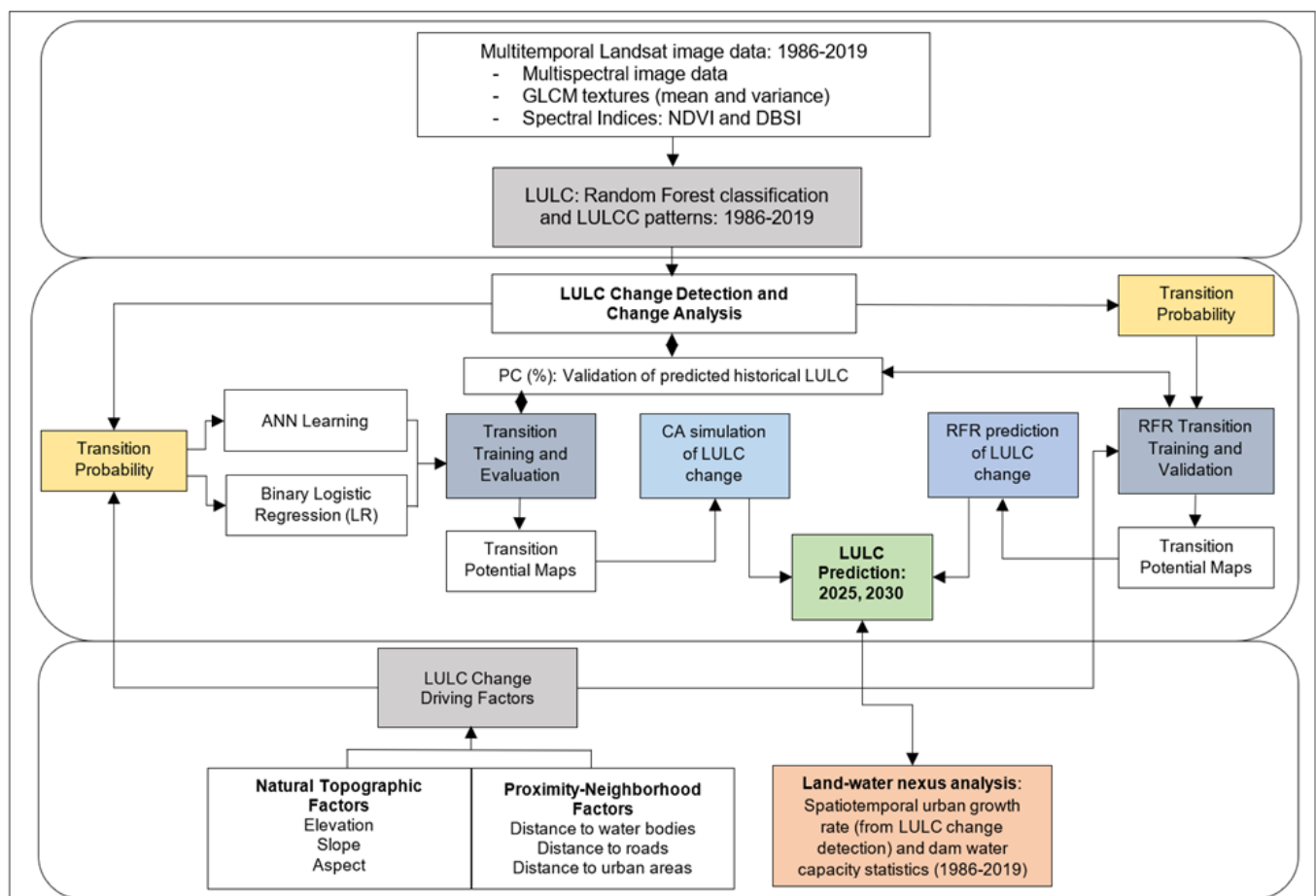


Figure 3. Schematic approach for LULC mapping and prediction using LR-CA, ANN-CA and RFR and land–water nexus analysis.

The LULC classification, simulation and prediction were carried out within the Google Earth Engine (GEE) using the Java programming language and supported by Modules for Land-use Change Simulations (MOLUSCE) in QGIS. The LULC change area transition matrix and the corresponding transition possibility matrix which reveals the likelihood of the LULC transitions in Figure 3 were generated using LR, ANN and RFR. The transition potentials were evaluated for the historical (1999–2019) years and integrated in the prediction of the future LULC cell status using CA (for LR-CA and ANN-CA) and RFR-LULC transition modeling.

2.3.1. LULC Mapping Using Machine Learning with Multiple Input Features

The multitemporal LULC mapping was carried out within the GEE platform using random forest classifier (RFC). RFC uses “parallel ensembling” which fits several decision tree classifiers in parallel on different data sub-samples and uses majority voting or averages for the outcome class as depicted in Figure 4. To build a series of decision trees with controlled variations, RFC combines bootstrap aggregation (bagging) and random feature selection. RFC minimizes the classification overfitting problem and increases the class prediction accuracy and control. As such, the RFC learning model with multiple decision trees is typically more accurate than a single decision-tree-based model, especially in detecting different LULC classes [70,71]. The overall advantage of RFC is that it can produce stable and accurate results even with minimal tuning of the hyperparameters. The algorithm is easy to parameterize, insensitive to overfitting and deals with outliers in training data, reporting the classification error and variable significance [72].

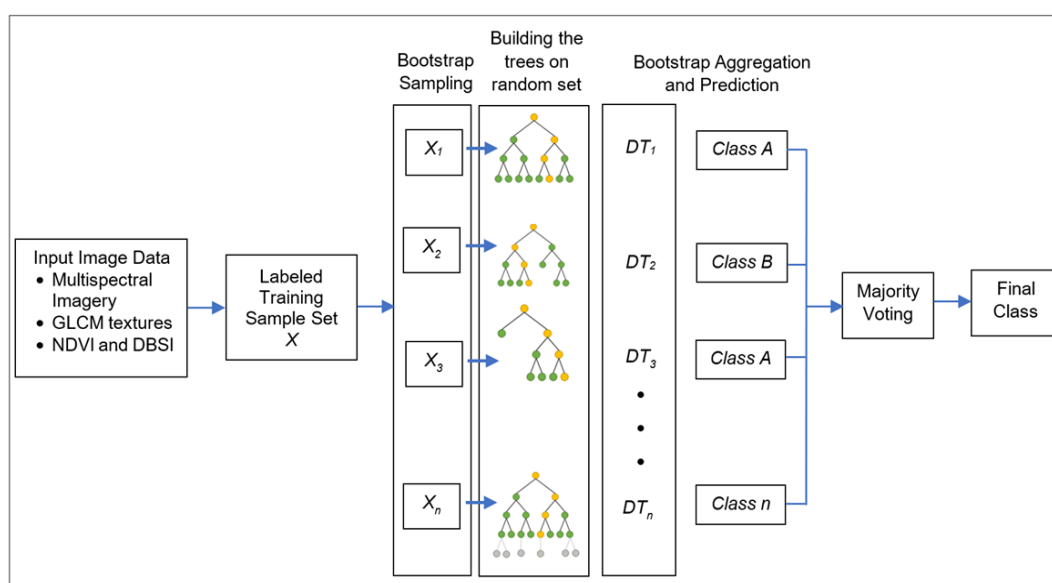


Figure 4. RFC schematic model for LULC classification with multiple input features. The RFC model is trained for each year with different LULC training data.

For each year, the training samples were visually collected in polygons with each polygon comprising 200 pixels. For all the LULC classes except water, 117 polygons were used for training and 20 polygons for validation of the results. For the water class, and due to its being smaller in areal extent, the training and validation comprised 70 and 30 polygons, respectively. The optimization and hyperparameterization of the RFC were implemented as detailed [70]. To improve the classification accuracy, multiple input features comprising mean and variance gray-level co-occurrence matrix (GLCM) textures from the first principal component of the multispectral image data were found to be useful in capturing the structural heterogeneity of classes. In addition, the normalized difference vegetation index (NDVI) (Equation (1)) and dry bare soil index (DBSI) (Equation (2)) were computed and integrated in the classification database. The indices and textural features

are included to enhance different land-cover classes as vegetation and bare soil and to improve the overall classification accuracy. Adopting the EU Copernicus Global Land Cover classification scheme (<https://lcviewer.vito.be/2019> (accessed on 10 August 2023)), the classified classes included tree cover, shrubland (shrubs and savanna), grassland, cropland, water (surface water; dams, rivers and ponds), built-up areas (buildings, roads and airports) and bare soil open land.

The LULC classification accuracy assessment was carried out using the overall accuracy (OA) and the kappa coefficient (K) (Equations (3) and (4)). For n classes with i categories ($i = 1, 2, 3, \dots, n$), the TP, TN, FP and FN representing true positives, true negatives, false positives and false negatives, respectively, are calculated and the OA and K derived.

$$NDVI = (NIR - R) / (NIR + R) \quad (1)$$

$$DBSI = \frac{SWIR1 - G}{SWIR1 + G} - NDVI \quad (2)$$

$$OA = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FN_i)} \times 100\% \quad (3)$$

$$K = \frac{OA - p_e}{1 - p_e} \quad (4)$$

where R is the red wavelength; G is the green wavelength; NIR is the near-infrared; $SWIR1$ is the shortwave infrared band 5 (Landsat 5/7 TM/ETM+) and band 6 (Landsat 8 OLI); $p_e = \sum_{i=1}^N (TP_i + FP_i)(TP_i + FN_i) / P_{Total, i}^2$, and $P_{Total, i} = TP_i + FP_i + TN_i + FN_i$.

2.3.2. Logistic Regression

Logistic regression (LR) is a statistical model in the generalized linear model class. It allows for the formation of multivariate regression relationships between a dependent variable and multivariate independent variables. In LULC prediction, LR uses the binary logistic regression model to analyze and evaluate the driving factors for land-use class transitions. The goal of LR in LULC prediction is to determine the best-fitting model to map the probability of transition of an LULC class to another class based on a set of driving factors (independent variables). The LR model prepares the LULC change probability maps for determining the adaptable model that conveys the correct relationships between the probabilities of the dependent and independent variables. The resultant output of the adaptable model includes the probability surface maps of the dependent variable based on the coefficients of independent variables.

As expressed in Equation (5), LR can be used to determine the transition probability P of various LULC types Y_i (Equation (6)) in a specific spatial location [46].

$$\log it(P_{i,j}) = \ln\left(\frac{P_{i,j}}{1 - P_{i,j}}\right) = \alpha + \sum_{i=1}^n \beta_i X_i \quad (5)$$

$$Prob(Y = 1|X_I) = \frac{Exp(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_i X_i)}{1 - Exp(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_i X_i)} \quad (6)$$

where $P_{i,j}$ is the probability of a class transitioning to another; $P_{i,j} / (1 - P_{i,j})$ is the “odds ratio” of an event which represents the probability that an outcome will occur given a particular condition compared to the odds of the outcome occurring in the absence of that condition; X_i is the independent variable representing the driving factor and $\beta_i = b$ is the estimated regression coefficient of each selected variable X_i .

2.3.3. Multilayer Perceptron (MLP) Artificial Neural Networks

The backpropagation MLP neural network can perform nonparametric regression analysis. With an input layer, an output layer and one or more hidden layers between the input and output layers, each hidden and output layer neuron processes its inputs by multiplying each input I_i^n by a weight W_i^j , summing the product and then processing the sum (if that exceeds the neuron threshold, then the neuron is activated) using a nonlinear activation function to derive I_j^{n+1} . Equation (7) defines how a neuron in the receiver layer receives values from the neurons from the sender layer, where I_i^n is the input value from the i th neuron in the sender layer and I_j^{n+1} (Equation (8)) is the output generated by the j th neuron in the receiver layer. W_i^j denotes the weights of the input values and b_j is a bias value added to the summation of all inputs. f is the function which determines how each cell in the grid will change based on the neighboring LULC classes. The determination of f is dependent on the characteristics of the data in the training and the suitability of the ANN model structure in determining the transition probability of LULC change using multiple output neurons for simulating the LULC changes within the ANN-CA structure.

$$\text{net} = \sum_{i=0}^n w_{ij} I_i^n \quad (7)$$

$$I_j^{n+1} = f(\text{net}) = f\left(\sum_i w_{ij} I_i^n + b_j\right) \quad (8)$$

The training data are a set of points (pixels) that are randomly selected from the maps. Each pixel is represented by a vector that is composed of the values of the driving factors. In addition, each pixel is assigned to a land-use type of the corresponding LULC map for a given year. The points are used to train the corresponding ANNs, with an ANN for each land use and its complements. The resulting probability maps are used as potential transition maps to improve the CA simulation. By combining the transition rules, the transition matrix and the potential maps that are produced by the ANN, the ANN-CA considers both spatial and temporal dynamics of LULC change and adequately incorporates the influence of the driving forces.

The transition potential for this study was trained with a momentum of 0.05 and learning rate of 0.1 for the stabilization of the learning graph. The number of iterations was set to 150 to minimize model overfitting. These thresholds are obtained from the ANN learning process and are derived when the highest transition probability is greater than the threshold value of 0.9 as proposed in previous experiments [67,73]. Below a probability of 0.9, the cells remain unchanged as the threshold of 0.9 is used to keep the LULC changes stable in each iteration, thus obtaining fine simulation patterns. In the ANN-CA simulation, the state of the new cell is determined by the existing state of a current cell and changes in the neighborhood cells in the CA. The ANN-CA simulation selects raster data, such as classes of LULC, raster of spatial parameters and transition potential model, based on the ANN algorithm [74]. The potential changes are determined for each class, and the simulation creates a raster of the most likely transitions. The simulation examines a fixed number of pixels, with the greatest certainty for each transition corresponding to the most likely transitions, and then it adjusts the class of the pixel. Further detailed implementation steps for the ANN-CA model are outlined [67].

2.3.4. Cellular Automata (CA) LULC Change Prediction Model

CA has the ability to model the proximity influences, which are considered as an essential spatial element that reflects the dynamics of land-use changes. The basic principle of CA is that the past LULC patterns affects future development through the local interactions which collectively constitute the global growth patterns within a region [75]. The CA simulation assumes that an LULC class has a higher tendency to transition to another category if the neighboring regions belong to or are influenced by that class category [76]. Discrete in space, time and state, CA is able to carry out complex time-space simulations [77].

CA consists of grid cells, cell space, neighbor, rule and time. Each cell has an internal state with a value which belongs to a set and updating these states is carried out simultaneously with the transition rule of each neighboring cell. Every cell is equivalent to a pixel on the area map. Each land-use category is represented by a cellular state and the cell data state S_{t+1} are decided by the cell and its neighboring cells in the S^t state. The transition rules then determine the change in the cell state. The results are derived from the suitability maps that represent the potential of a cell to change from one state to another. In general, the CA model shows a cell's interaction and its state of change which affects the spatiotemporal pattern of neighboring cells. The cell neighborhood is determined by a filter and the closer the distance between the nuclear cell and the neighbor, the higher the influence weight. In the CA implementation, a contiguity filter of 5×5 pixels with $30 \text{ m} \times 30 \text{ m}$ spatial resolution is adopted as suggested by [55]. By combining the weight with the transition probability, the next potential state of the adjacent cells is derived. The transition model is expressed as in Equation (9):

$$S_{i,j}^{t+1} = f(S_{i,j}^t, \Omega_{i,j}^t, V) \quad (9)$$

where t and $t + 1$ are the beginning and end of simulations; $S_{i,j}^{t+1}$ and $S_{i,j}^t$ are the state of the cell in row i and column j at time t and $t + 1$; $\Omega_{i,j}^t$ is the state of neighbors of the cell in row i and column j at t ; V is the set of suitability factors and f is the transition law or function considered either as the sum [78] or as the product of all the terms [79,80].

2.3.5. Random Forest Regression for LULC Prediction

For multiple predicted factors, the uncertainty of the inspection variable changes with the LULC type and its intensity. This implies that the interactions between the LULC classes and the driving factors tend to be nonlinear. RFR is a nonlinear model that is suitable for the regression of datasets that do not exhibit linearity. RFR can capture the existing and potential interactions between the type and intensity of an LULC class and the influencing (driving) factors [81]. In its implementation, the proposed RFR-LULC change simulation and prediction model first extracts the multitemporal class area changes, from which the transition probability matrices for the successive years are derived. This is followed by the generation of transition potential maps which are combined with the transition probabilities to simulate the future LULC future scenarios.

The RFR machine learning technique for modeling LULC can be regarded as a classification problem [82] and can formally be represented such that if $C = \{c_1, c_2, \dots, c_k\}$ is a set of k LULC classes represented in grids, then each grid cell at time point t can be represented as a row vector $x^t = \{x_1^t, \dots, x_i^t, \dots, x_n^t\}$, where x_i^t represents the i th spatial attribute value assigned to the cell. Let y^t from C be an LULC class of x^t at time t , then a prediction function $f_p: x^t \rightarrow y^{t+1}$ can be applied over every x^t on the study area grid on the condition that $f_p(x^t) = y^{t+1}$ holds whenever the cell for x^t changes its land use to the class y^{t+1} . The model constraint is that each cell can belong to only one class at any time t . The transition function f_p maps the grid at time t to the LULC classes at the next time $t + 1$, and the ML learns the function f_p' which approximates the unknown f_p using the training set in which all of the attribute values at time t and land-use classes at time $t + 1$ are known a priori.

For each time interval $(t, t + 1)$, the initial dataset $I_{t, t+1}$ is created and contains the driver attributes x^t with y^{t+1} as the LULC class at time $t + 1$ for each grid cell. To learn the predictive function f_p' , a training set $\{(x^t, y^{t+1})_i\}$, $i = 1, 2, \dots, n$, where n is the training set sample, is constructed from the initial dataset $(I_{t, t+1})$, with x_i^t representing the grid cell at time t and y_i^{t+1} being the corresponding LULC class at time $t + 1$. The training model is tested using an independent test set comprising $\{(x^{t+1}, y^{t+2})_i\}$, $i = 1, 2, \dots, m$, where m is the test dataset samples, and the test dataset is derived from the corresponding initial set $I_{t+1, t+2}$. This implies that three time points are used for training $(t, t + 1)$ and for testing $(t + 1, t + 2)$ the proposed RFR prediction model.

The RFR training process generally involves each tree using concentrated training data and learning from samples of the random sets. The predicted variables from the random set increase exponentially without trimming of the trees and the second step is iterated until the number of trees increases. In the final iteration, the average predicted LULC types and intensity are derived [83]. In predicting the areal extents of LULC classes using RFR, the class area (CA) for each classification year is computed according to Equation (10):

$$CA = (C * X * Y) / 100,000 \quad (10)$$

where CA = area of the LULC class type (in square kilometers); C = class cell count; X = height of class cell and Y = width of class cell.

To evaluate the predictive power of the approximation, it is necessary to test the model that is built in the previous step with inputs x^{t+1} and to compare $f_p'(x^{t+1})$ with known y^{t+2} . Hence, RFR-LULC change prediction model building and validation assume the availability of data on spatial attributes and land-use classes from three different time epochs.

2.3.6. Validation of LULC Prediction

For the training of the LR-CA, ANN-CA and RFR models, two consecutive years ($t, t + 1$) are used, and the prediction performance is measured by comparing the predicted with the classified LULC at ($t + 2$). From the derived difference statistics, the LULC prediction results are validated using the percentage of correctness (PC) in Equation (11). PC determines the agreement and disagreement between the simulated and the reference LULC maps.

$$PC = \frac{p_a - p_e}{p_i - p_e} \times 100\% \quad (11)$$

where p_a is the proportion of observed agreements (actual accuracy); p_e is the proportion of agreements expected by chance and p_i is the ideal accuracy (100%). $p_a = \sum_{i=1}^c p_{ij}$ and $p_e = \sum_{i=1}^c p_i T_p T_j$, where p_{ij} is the i th and j th cell of the contingency table, $p_i T_p$ is the sum of all cells in the i th row, $p T_j$ is the sum of all cells in the j th column and c is the count of the raster category.

3. Results

3.1. LULC Change Patterns in the Gaborone Dam Catchment

Figure 5 presents the classified LULC maps for the eight study years using the RF classifier. As already stated above, the mapping adopted the EU Copernicus Global Land Cover classification scheme (<https://lcviewer.vito.be/2019> (accessed on 10 August 2023)). The summary accuracy of the LULC classification results is presented in Table 2. The overall accuracy (OA) is observed to be above 81% for all the years, with the highest OA at 89.6% for the year 2004, and the lowest OA at 81.3% for 2009. The corresponding kappa coefficients ranged from 0.74–0.84 (Table 2). The LULC classification accuracy for 2019 was dynamically compared with the EU Copernicus Global Land Cover of the same year, with high a degree of visual agreements for the classes.

Table 2. Overall Accuracy (OA) and Kappa coefficients for LULC classification.

Year	1986	1989	1994	1999	2004	2009	2014	2019
OA (%)	88.9	89.1	83.4	85.2	89.6	81.3	85.2	84.8
Kappa Index	0.82	0.84	0.76	0.78	0.84	0.74	0.78	0.80

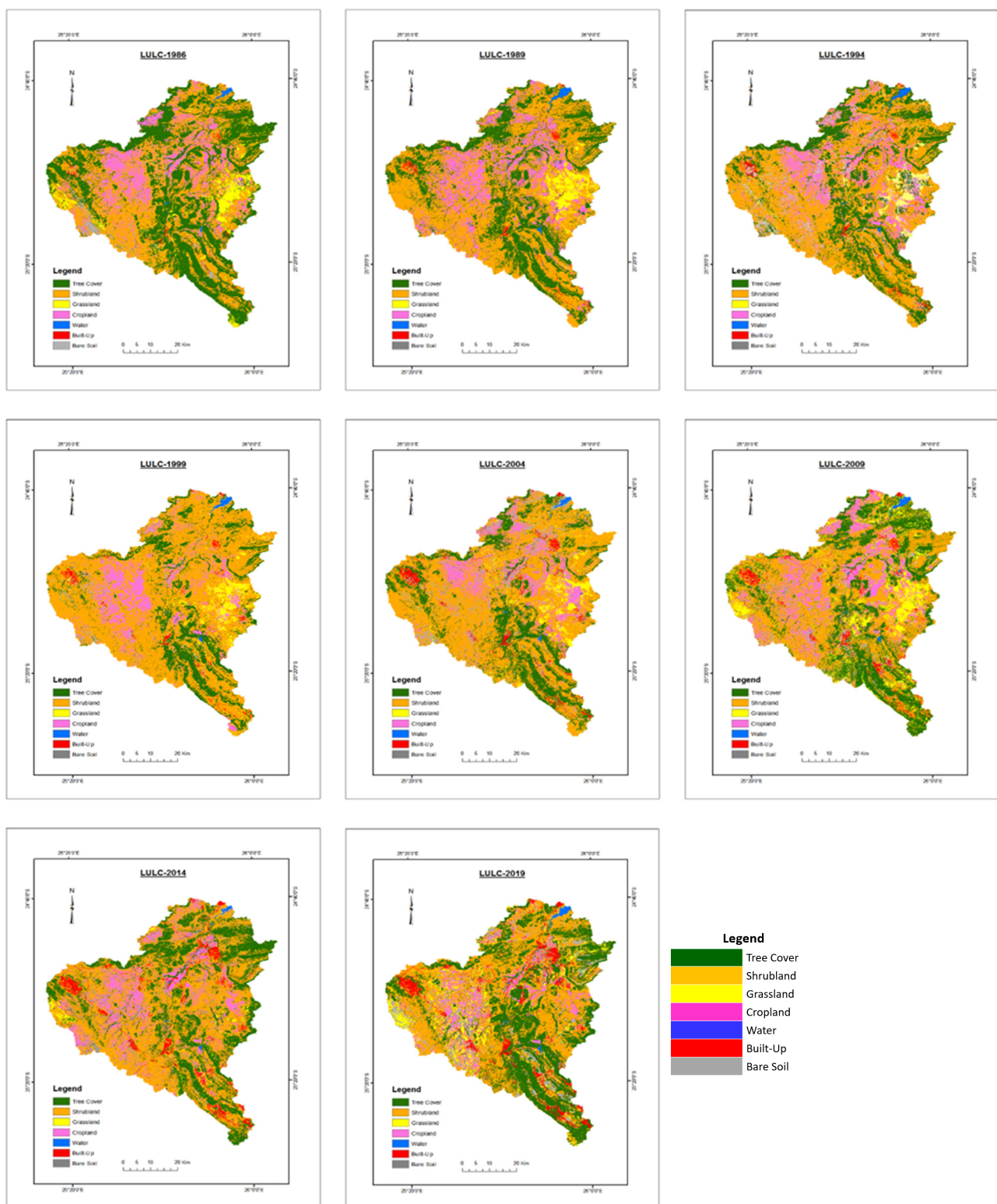
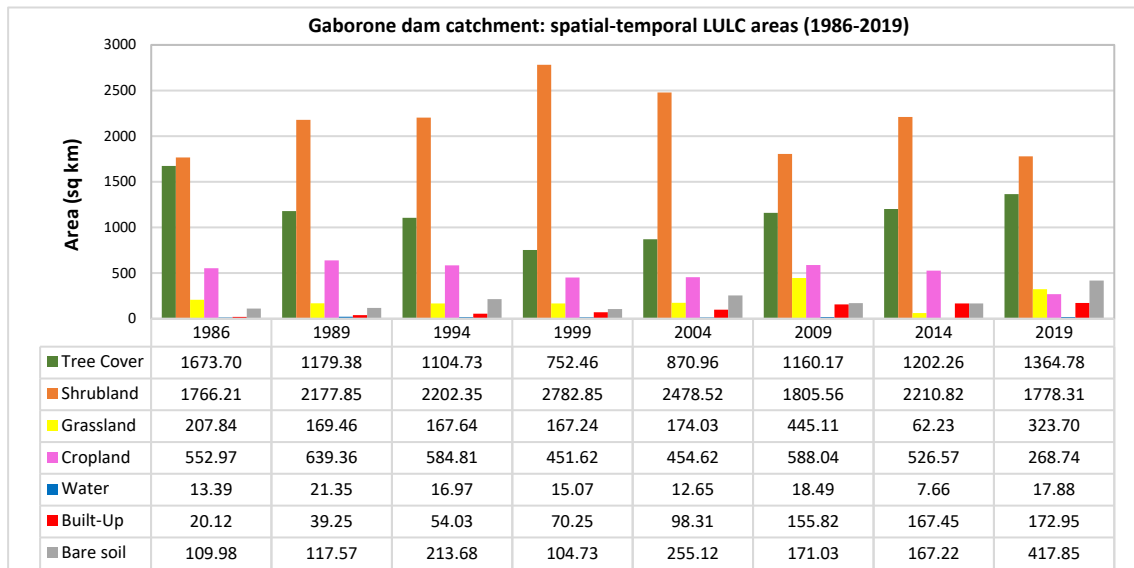


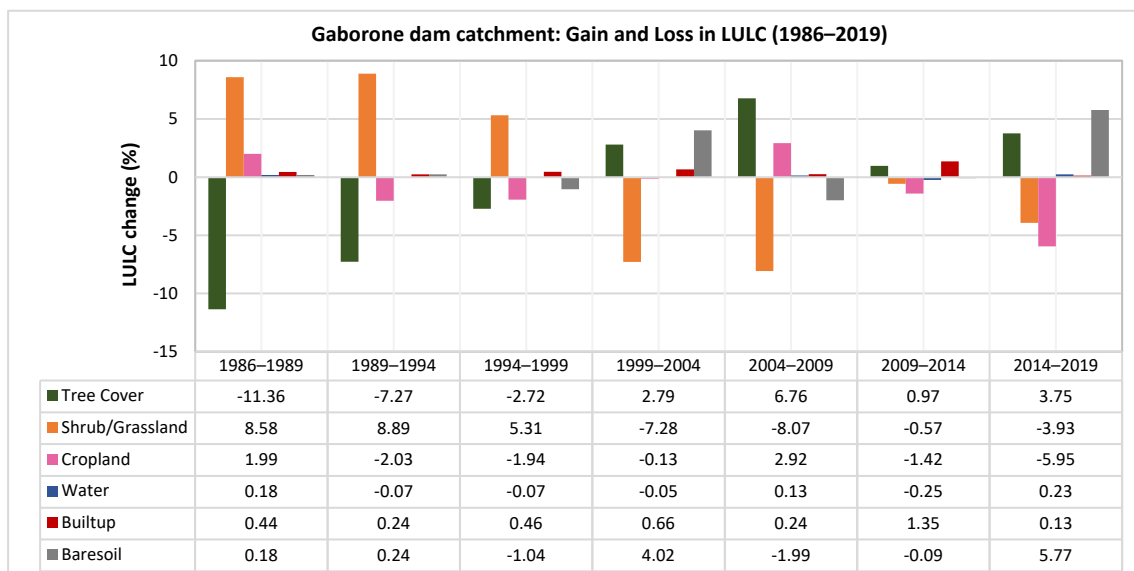
Figure 5. LULC classified maps of the Gaborone dam catchment from 1986 to 2019.

From the classification results, the LULC change trends in the catchment are presented in Figure 6. Shrubland, tree cover and cropland are observed to be the dominant LULC in terms of area coverages (Figure 6a). Tree cover is noted to decline from 1673.70 km² to 752.46 km² during the 1989 to 1999 period. The decline in tree cover could be attributed to the conversion of the forest cover into shrubland, which is mainly driven by urban development and agricultural activities. As a consequence, an increase in shrubland,

bare soil and built-up areas is observed. Grassland decreased from 207.84 km² in 1986 to 167.24 km² in 1999 followed by a surge in the 2004–2009 period. Cropland had an increase from 552.97 km² in 1986 to 639.36 km² in 1989 but reduced between 1994 and 1999 from 584.81 km² to 451.62 km². From 2004 to 2019, tree cover had an increase in area coverage following a previous decline. Shrubland and grassland reduced in area followed by a rise in bare soil and built-up areas. Water bodies had a marginal decline in area followed by an increase from 7.66 km² in 2014 to 17.88 km² in 2019.



(a)



(b)

Figure 6. Spatiotemporal variability of LULC in the Gaborone dam catchment: (a) class area coverage and (b) LULC gain and loss.

A gradual but steady growth is observed for the built-up areas throughout the 30-year period from approximately 20 km² (1986) to 173 km² (2019). Particularly rapid growth was observed in the total area covered by the built-up class from 1999–2009 as the area increased from 70.25 km² to 172.95 km². This increase was, however, at the expense of areas covered by shrubland which decreased considerably. The area covered by surface

water increased from 13.39 km² in 1986 to 21.35 km² in 1989, and thereafter a steady decline was observed, which is mainly attributed to climate variations. The corresponding class areal gains and losses are presented in Figure 6b and shows net 30-year losses for tree cover (−10.7%), cropland (−0.6%) and water bodies (−0.3%), while gains are observed for shrubland/grassland (+6.7%), bare soil (+1.3%) and built-up areas (+3.4%). Some of the observed changes, especially in water and natural vegetation cover, are not only attributed to anthropogenic activities but also to climatic conditions. Due to the high correlations between grassland and shrubland covers, the two classes are combined into one class as shrubland/grassland or just shrubland for further analysis.

3.2. LULC Class Transition Analysis

The transitional probabilities for the different LULC classes between 1986 and 2019 are presented in Table 3. During the early years with minimal built-up area, it was observed that the built-up area for the period of 1986–1989 had the likelihood of being converted to shrubland or grassland with a transitional probability of >50%. This is also perceived for 1989–1994 and 1994–1999 periods as the prospect of the built-up class remaining the same is shown to be unlikely with <35% transitional probability. From 1999 onwards, however, the built-up class is observed to be resistant to change to other land-use classes with >40% probability of remaining unchanged.

Table 3. Transition probability matrix for Gaborone dam catchment from 1986 to 2019.

Class	Transition Years	Tree Cover	Shrub/Grassland	Cropland	Water	Built-Up	Bare Soil
Tree Cover	1986–1989	0.5778	0.3707	0.0337	0.0001	0.0050	0.0128
	1989–1994	0.6325	0.3181	0.0321	0.0004	0.0051	0.0119
	1994–1999	0.4730	0.5014	0.0152	0.0005	0.0038	0.0060
	1999–2004	0.6582	0.2967	0.0275	0.0005	0.0036	0.0135
	2004–2009	0.6393	0.3118	0.0283	0.0007	0.0089	0.0110
	2009–2014	0.6140	0.3443	0.0190	0.0000	0.0102	0.0125
	2014–2019	0.7236	0.2295	0.0096	0.0003	0.0034	0.0337
Shrub/Grassland	1986–1989	0.0923	0.7340	0.1253	0.0047	0.0116	0.0321
	1989–1994	0.1380	0.7087	0.1028	0.0001	0.0086	0.0418
	1994–1999	0.0827	0.8294	0.0681	0.0000	0.0111	0.0086
	1999–2004	0.1192	0.7185	0.0944	0.0001	0.0192	0.0486
	2004–2009	0.2101	0.6390	0.1113	0.0002	0.0119	0.0275
	2009–2014	0.1978	0.6534	0.0980	0.0000	0.0235	0.0272
	2014–2019	0.1974	0.6336	0.0463	0.0017	0.0209	0.1001
Cropland	1986–1989	0.0322	0.3578	0.5700	0.0000	0.0039	0.0361
	1989–1994	0.0301	0.4002	0.4400	0.0001	0.0081	0.1215
	1994–1999	0.0411	0.5430	0.3732	0.0003	0.0204	0.0220
	1999–2004	0.0243	0.5294	0.3115	0.0002	0.0151	0.1194
	2004–2009	0.0545	0.4629	0.4363	0.0001	0.0092	0.0369
	2009–2014	0.0458	0.4496	0.4253	0.0000	0.0266	0.0527
	2014–2019	0.0515	0.5369	0.2693	0.0005	0.0165	0.1253
Water	1986–1989	0.0691	0.0208	0.0048	0.8798	0.0063	0.0192
	1989–1994	0.1453	0.0173	0.0300	0.7524	0.0264	0.0287
	1994–1999	0.0275	0.0248	0.0071	0.8268	0.0212	0.0926
	1999–2004	0.0139	0.0122	0.0012	0.7526	0.0010	0.2191
	2004–2009	0.0012	0.0088	0.0000	0.9666	0.0003	0.0230
	2009–2014	0.0355	0.2053	0.0162	0.3918	0.0035	0.3478
	2014–2019	0.0216	0.0024	0.0000	0.9317	0.0098	0.0346

Table 3. Cont.

Class	Transition Years	Tree Cover	Shrub/Grassland	Cropland	Water	Built-Up	Bare Soil
Built-Up	1986–1989	0.1020	0.5229	0.0652	0.0004	0.2097	0.0999
	1989–1994	0.0666	0.3962	0.0712	0.0005	0.3788	0.0867
	1994–1999	0.0242	0.5414	0.0418	0.0010	0.3441	0.0475
	1999–2004	0.0661	0.3315	0.0793	0.0024	0.3893	0.1315
	2004–2009	0.0496	0.3207	0.0460	0.0005	0.4885	0.0947
	2009–2014	0.0322	0.2954	0.0375	0.0001	0.6021	0.0327
	2014–2019	0.0634	0.2100	0.0145	0.0002	0.6237	0.0882
Bare Soil	1986–1989	0.0870	0.6300	0.1738	0.0009	0.0128	0.0955
	1989–1994	0.0844	0.5060	0.1777	0.0023	0.0599	0.1698
	1994–1999	0.0253	0.4538	0.3042	0.0005	0.0377	0.1785
	1999–2004	0.0759	0.3237	0.1213	0.0033	0.0528	0.4230
	2004–2009	0.0638	0.3489	0.2565	0.0196	0.0675	0.2437
	2009–2014	0.0258	0.4204	0.1446	0.0016	0.1195	0.2882
	2014–2019	0.0496	0.4067	0.0463	0.0375	0.0485	0.4115

The water class showed a constant trend of remaining unchanged throughout the entire study period. This is especially evident for the periods of 1986–1989, 1994–1999, 2004–2009 and 2014–2019 exhibiting very minimal transitional probabilities of 87%, 82%, 96% and 93%, respectively. Similarly, the shrub/grassland class also showed a steady trend of remaining unchanged throughout the study period with a probability of >63% except for bare soil with 32–63% conversion probability. The cropland class showed a high possibility of being converted to another land-use class from the year 1994 onwards, having a probability of <43% to remain as cropland, implying that it may be converted to shrub/grassland with a probability of >44%. Tree cover showed moderate resistance to conversion to other land-use classes. This is especially the case during the periods of 1986–1989 and 1989–1994 with corresponding probabilities of 57% and 63%. Conversely, from 1994 to 1999, tree cover had a likelihood of being converted to shrub/grassland exhibiting a transitional probability of 50%. Nonetheless, the tree cover class remained unchanged from 1999 onwards with a transitional probability of >60%.

3.3. Calibration of LULC Transition Potential

The LULC calibrations were carried out for 1994, 1999, 2004, 2009, 2014 and 2019 to evaluate the performance of the prediction algorithms. The average percentage of correctness was computed between the classified and predicted LULC transition maps. The LR prediction results (Figure 7) show a fair correlation with the reference classified LULC in terms of areas covered by the vegetation classes, with shrubland being the most underestimated vegetation cover in 1999. LR, however, underestimated the built-up area except for 1994 and 2019. The most accurate built-up area prediction using LR was recorded in 2019 with 167 km² compared to the classified area of 172.95 km², indicating an error of 3.2% (5.51 km²). LR predicted an increase of 82 km² in cropland from 1994–1999, with a steady decrease between the next subsequent years of −37 km² from 2004–2009 and −35 km² from 2014 to 2019. Bare soils had mixed prediction results with overestimation in 1994, 2004, 2014 and 2019, while underestimation was observed in 1999 and 2009.

Using ANN prediction, the results in Figure 8 present the areas covered by the LULC classes from 1994 to 2014. For the years 1999, 2009 and 2019, there is a considerable difference between the classified and predicted results for shrubland. This could be attributed to the fluctuations of the areas covered by the vegetation classes which affected the learning of the prediction model. The built-up area shows a consistent increase throughout the years for both classified and predicted output, however, with under- and overestimations for different years. The water class is the most accurately predicted class as the differences between the classified and the predicted areas are observed to be minimal.

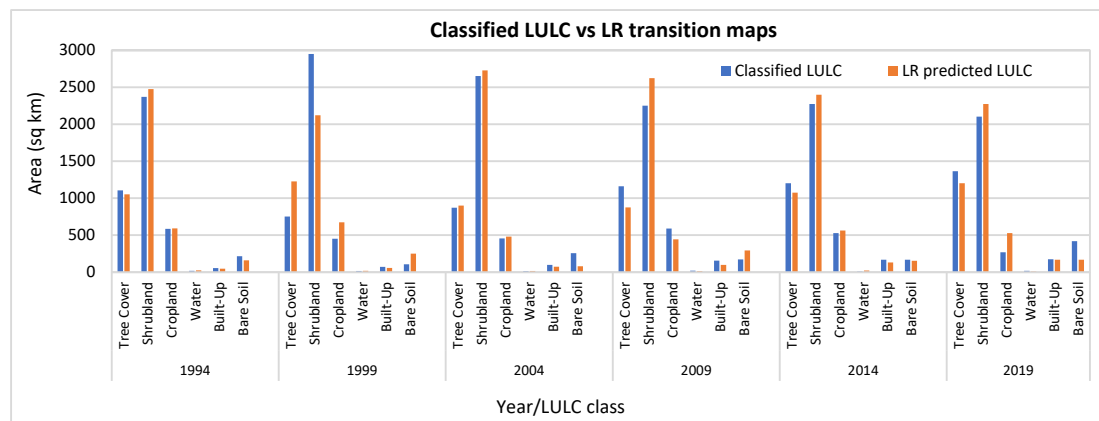


Figure 7. Comparison of LR-predicted LULC and classified LULC.

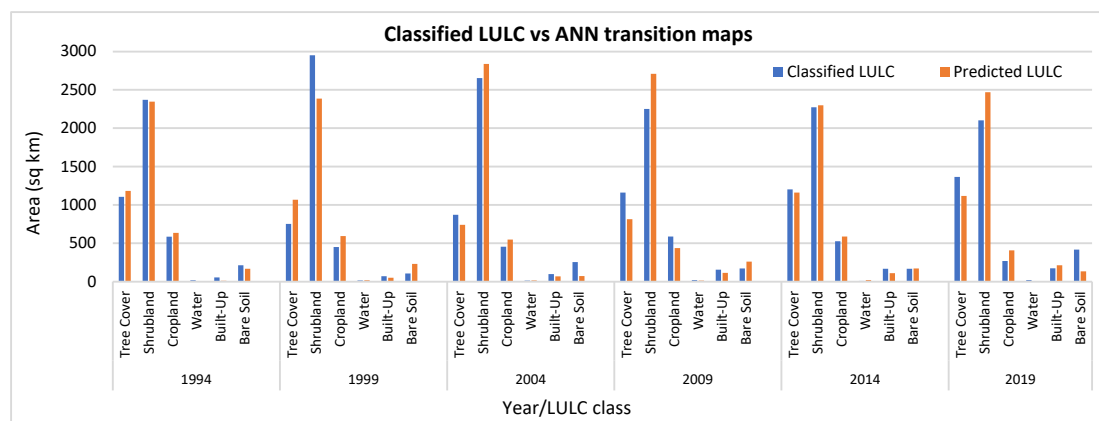


Figure 8. Comparison of ANN-predicted LULC with classified LULC.

As compared to the LR and ANN LULC class prediction validation results, the results in Figure 9 show that RFR predictions have a higher degree of closeness to the classified LULC. Except for shrubland which is overestimated by RFR in all the years except in 2019, the rest of the predicted classes have marginal differences from the classified LULC classes. The average LULC validation of the prediction accuracies per year using LR, ANN and RFR models is presented in Figure 10 with marginal difference between LR (PC = 60.7%) and ANN (PC = 62.1%), while the average percentage correctness for RFR is observed at 84.9%.

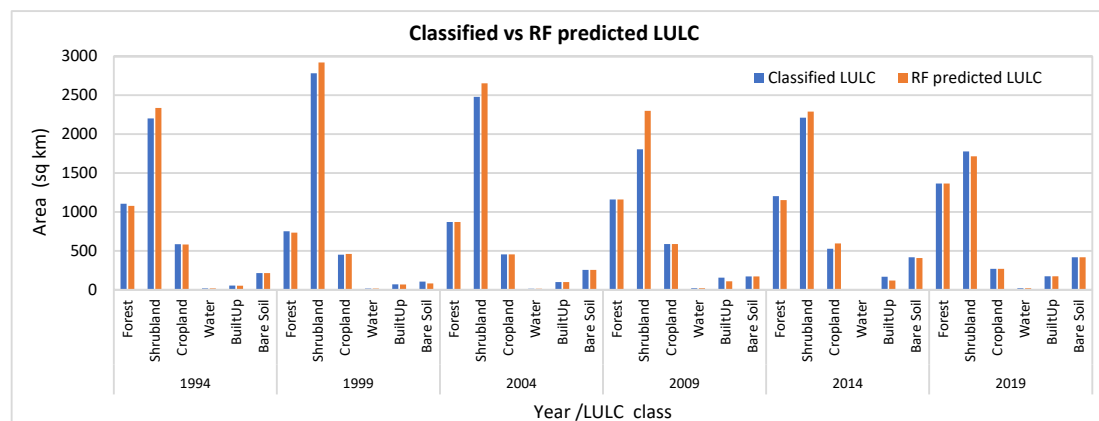


Figure 9. Validated classified LULC and predicted LULC using RF.

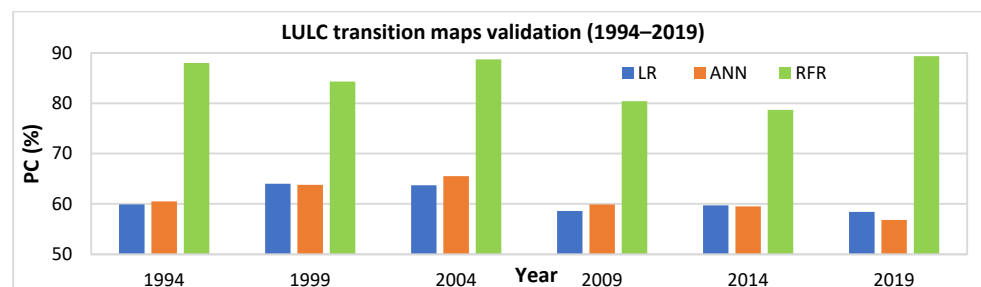
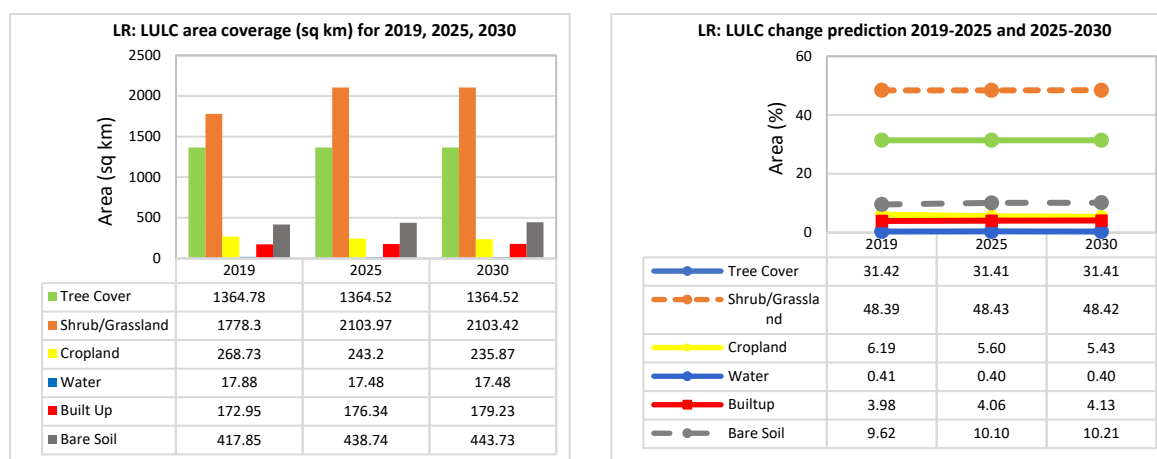


Figure 10. Average LULC prediction percent correctness using LR, ANN and RFR for Gaborone dam catchment.

3.4. LULC Prediction for 2025 and 2030

Using LR-CA, ANN-CA and RFR models, the LULC prediction summary results for each class are presented in Figure 11 and Table 4 for 2019–2030. From the results, water bodies are predicted to remain constant in area for all three prediction models, with LR-CA and ANN-CA predicting minor decrease of -0.01% for 2019–2025 and no changes for 2025–2030. RFR showed a total increase of $+0.05\%$ for the entire duration of 2019–2030. LR-CA and ANN-CA predicted the built-up area to change by nearly the same magnitude of $+0.14\%$ between 2019 and 2030, thus occupying an area of 179.23 km^2 increased from 172.95 km^2 . RFR, however, showed a net increase of $+2.6\%$ within the same period with a higher growth rate of $+1.63\%$ predicted from 2025–2030, thus predicting an increase in urban built-up area from 3.98% (2019) to 6.56% (2030). Tree cover prediction largely remained nearly constant at 31.41% from LR-CA and ANN-CA predictions for the two time periods. RFR however, showed a net decrease of -0.68% for the two periods which can be explained by the observed increase in built-up area taking up more space within the catchment. For shrubland/grassland, only LR-CA predicted a net increase of $+0.03\%$ from 2019–2030, however, for the same period ANN-CA and RFR showed respective decreases of -0.19% and -8.91% . The decrease in tree cover, shrubland/grassland and cropland may not only be explained by the increase in built-up area but also by the increase in bare soil cover of 8.79% , with most of the increase (8.65%) observed from 2025–2030.



(a)

Figure 11. Cont.

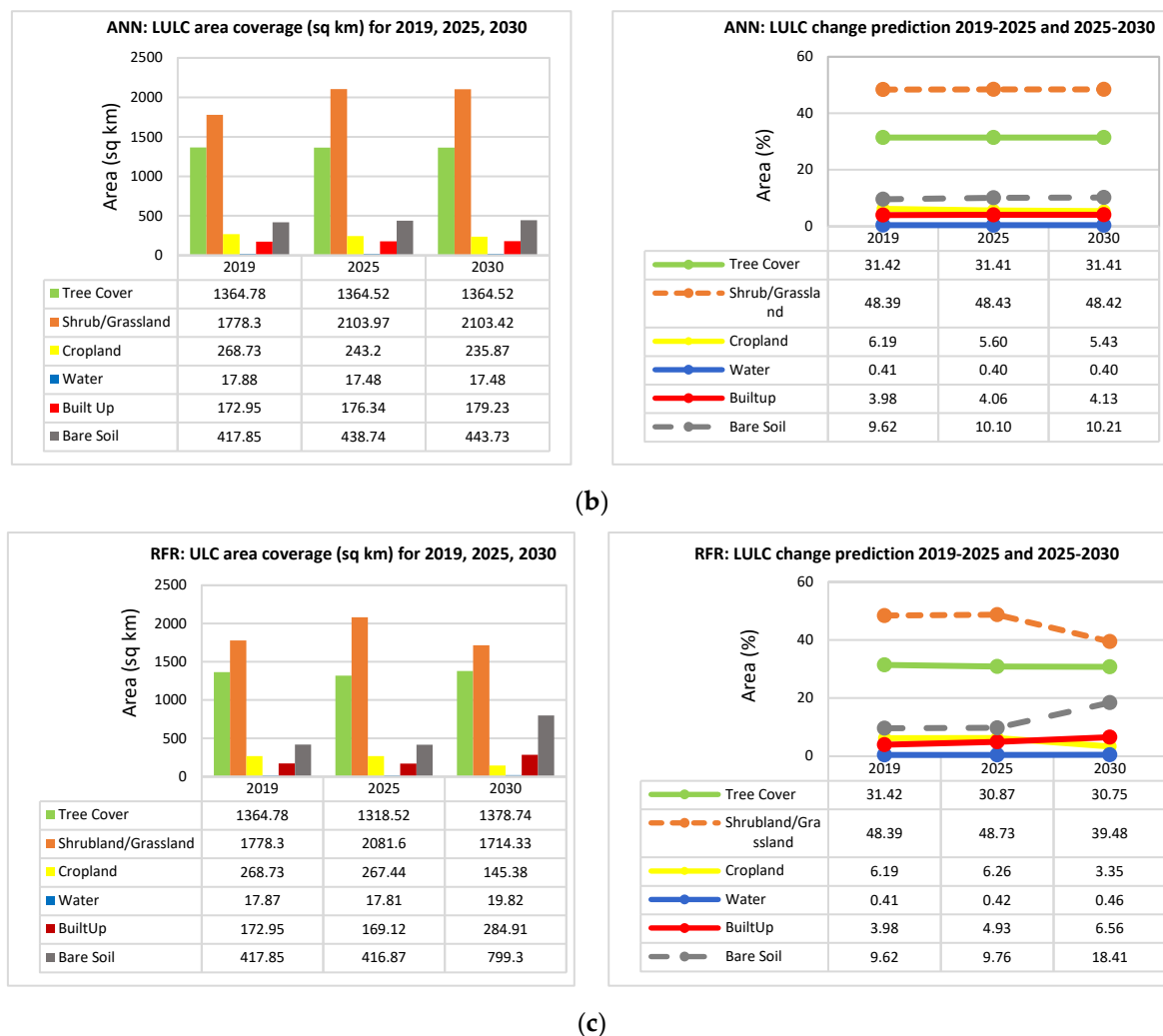


Figure 11. LULC prediction results for Gaborone dam catchment from 2019 to 2030 using: (a) LR-CA, (b) ANN-CA and (c) RFR prediction results.

Table 4. Predicted LULC change in the Gaborone dam catchment (2019–2030).

5-Year LULC Predictions (%)						
LULC Class	LR-CA		ANN-CA		RFR	
	2019–2025	2025–2030	2019–2025	2025–2030	2019–2025	2025–2030
Tree Cover	−0.01	0.00	−0.02	0.00	−0.55	−0.12
Shrubland/Grassland	0.04	−0.01	−0.02	−0.17	0.34	−9.25
Cropland	−0.59	−0.17	−0.60	−0.17	0.07	−2.91
Water	−0.01	0.00	−0.01	0.00	0.01	0.04
Built-Up	0.08	0.07	0.07	0.07	0.95	1.63
Bare Soil	0.48	0.11	0.47	0.12	0.14	8.65

The results in Figure 12 show the RFR-predicted LULC for 2025 and 2030, as RFR presented the best validation and prediction results as compared to LR-CA and ANN-CA (Figure 10). From visual comparison of the 2025 and 2030 predicted LULC, most of the increase in built-up area is detected in the southeast parts of the catchment. These are also the regions with close proximity to forest tree cover and hence are more suitable for

settlement given the semiarid and arid climatic conditions within the catchment. The forest cover is observed to be largely conserved, and so is the dam water body.

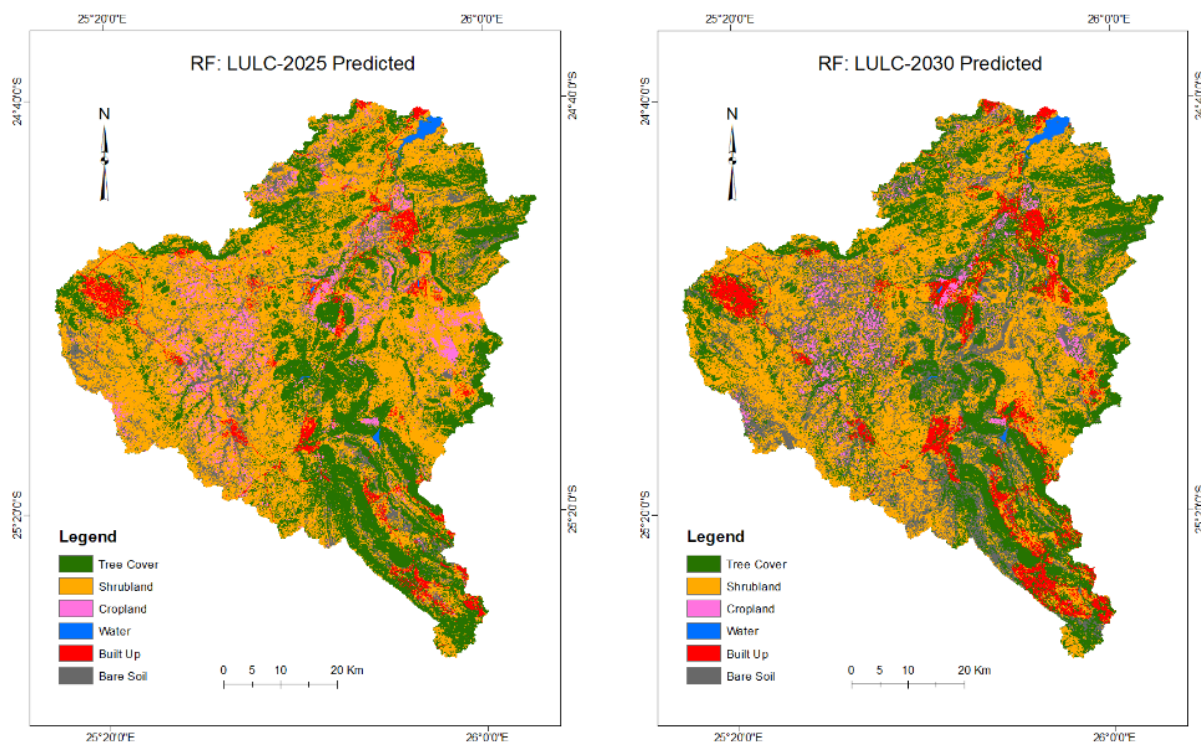


Figure 12. RFR-predicted LULC for Gaborone dam catchment for 2025 and 2030.

4. Discussions

4.1. Comparison of the LULC Prediction Models

The prediction of the future LULC changes is important for, among others, urban and regional planners, hydrologists, water policymakers and environmentalists in making appropriate decision policies for future developments and sustainability. As such, spatiotemporal LULC changes in catchments are modeled to detect the consequences of long-term interactions between humans and the environment [84], hence contributing to sustainable development [2]. For this purpose, location-based models are commonly utilized [85] to simulate the LULC dynamics.

Cellular automata are the most widely used location-based models especially for simulating the spatiotemporal evolution of LULC and urban expansion [86]. The simplicity and explicit representation of LULC changes [87] make the CA model a standard model for simulating such dynamics. While the CA model assumption is based on the effects of past changes on future transitions [75], this background convention limits the model's ability to realistically simulate the complex nature of LULC over time. In addition, the transition rule which extracts the state of a cell over time [86] and is an essential function in the CA model varies in terms of geographical regions and neighborhood interactions. Integration of CA with other statistical and geospatial models can enhance the CA model's predictability [3,75]. Commonly used reinforcement methods are the LR model [88,89], Markov chain analysis (MCA) [90,91], agent-based model [92,93] and ANNs [2,86].

This study successfully compared hybrid statistical LR-CA and nonparametric ANN-CA models with a machine learning RFR approach for the simulation and prediction of LULC in a dam catchment. The models were validated using the degree of correctness and ANN performed better LR with a percentage of correctness of 62.1% which was slightly higher than that of LR by 1.4%. The results of the validation of the performance prediction models show that the proposed RFR outperformed the hybrid models by about 23% in percent accuracy. While LR lacks representation of the effects of LULC drivers [86] and

spatial dependency [94], the ANN algorithm is more able to address the spatial probability of changes and can be trained to estimate the probability of occurrence from nonlinear functions through training by weight change and calibration to simulate a more realistic projection of LULC changes [68,95]. The combination of LR and ANN is envisaged to increase the LULC change prediction accuracy, however, the main drawback in CA is the inability to include trends from previous states and driver variables responsible for LULC changes in the basin area. This could contribute to lower performance of the LR-CA and ANN-CA hybrid models.

Comparatively, in a study carried out for Addis Ababa (Ethiopia) using MC-CA, for example, an average validation accuracy of 87% was achieved for three test years of 2005, 2011 and 2015 [96]. However, in the current research, the employment of ANN with the CA model resulted in an overall accuracy of 62.1%. This is because, as opposed the previous studies that were based on well-structured urban areas, the current study considered an entire watershed in which LULC and driving factors may be more heterogeneous and complex. Aburas et al. [44] utilized the MC-CA model with AHP and frequency ratio (FR) to simulate urban growth in Seremban (Malaysia). The models performed well in determining important factors for urban growth with similar accuracies of 88.1% and 88.2%. However, their incorporation of subjective weighting of variables renders the results difficult to replicate when using other experts to weigh the driving forces [97]. Other studies applied physical and proximity drivers for LULC change simulation based on integrated models, including the ANN-CA-MCA [98], ANN-CA [60,68,86] and CA-MCA [39,88,99].

Despite the extensive efforts to improve LULC predictions, there appear to be shortcomings resulting from subjective methods of weighting the variables such as the approach used in AHP. LR shows limitations where there is insufficient knowledge about the area of interest or failure in covering all aspects and variables affecting land-use change. Ref. [100], however, implemented an LR model and achieved an accuracy of 81%. Compared to LR, ANN can be considered as an unbiased tool that is appropriate to assign weights that are derived with minimum prediction errors. As a result, it is fair to say that the ANN approach reduces inaccuracy as well as the possibility of expert bias. Ref. [68], for example, integrated an ANN with CA-MC which improved the accuracy from 86.3% using CA-MC to 90%. This implied that the integration of CA-MC with the ANN allows the model to capture the different variables and dynamics behind land transformations, which significantly improves the CA-MC model's prediction capability.

Influenced by multiple factors, the traditionally used empirical–statistical models (e.g., MC and regression models) and dynamics models (e.g., CA model, agent-based model and system dynamic models) are capable of predicting land-use change in the future for LULC, however, they fail to provide precise explanations on the impacts of the LULC-change-driving factors or variables and tend to either overestimate or underestimate the prediction of LULC changes [47]. Hybrid or integrated models (e.g., MC-CA, MC-CA-ANN, LR-CA model and the conversion of land-use and its effects (CLUE) model) serve to improve LULC change prediction by combining elements of different modeling techniques have thus been suggested to improve the LULC change prediction. Nonetheless, due to the complexity of studying land-use change, especially with catchments, most of the proposed hybrid models have limitations that include insufficient knowledge about the area of interest and subjectivity in weighting the variables and they cannot model nonlinear phenomena, resulting in unreliable the prediction results [49].

RFR can capture the nonlinear relationships between factors and deal with complex patterns and changes in land use with great efficiency. This is based on its provision of nonlinearities and its ability to deal with missing or fuzzy data as well [44]. Thus, RFR can detect potential interdependencies through implied driving forces. Moreover, the significance of using the RFR model is that the model illustrates the effects of each driving factor used in the simulation operation and specifies which factors affect the land change more to give a clearer understanding of the land change process. RFR acts independently regardless of the statistical data distribution or the lack of statistics for

specific variables [101]. The outcome from this study demonstrates the ability of RFR to train for prediction even with limited inputs and driving forces, thus allowing for the detection of potential interdependencies [49].

4.2. Case Study Assessment

From the LULC prediction results for the case study, the highest net change is observed in the decrease in natural vegetation cover, with forest cover predicted to decrease by -0.02% (LR and ANN) and by -0.68% (RFR). Shrubland/grassland are predicted to decrease by -0.19% using the ANN and -8.91% using RFR. An equivalent gain is detected in bare soil having the highest increase ranging from $+0.59\%$ (from LR and ANN) to $+8.79\%$ using RFR. Cropland area is also predicted to decrease during the 10-year period by -2.8% using RFR and marginally by -0.77% according to LR and ANN predictions. From the RFR predictions of built-up area from 2019–2030, a net increase of 2.49% is determined. Compared to the magnitude of decrease in vegetation cover and cropland, it can be inferred that most of the cropland areas are likely to be converted to built-up areas, while the vegetated surface is mainly converted to bare soil. The water body is predicted to decrease marginally by -0.01% using LR and ANN models and to increase by $+0.05\%$ from RFR simulations. Conversely, it is observed that LR- and ANN-based predictions only resulted in a $+0.15\%$ increase in built-up area and a -0.77% decrease in cropland. Comparatively, the magnitudes and directions of the LULC predictions using LR and ANN are observed to be nearly equal, however, they tended to underestimate future LULC scenarios. It can also be argued that while the RFR predictions represented the likely scenarios in the future, the large magnitude of changes in 10 years, especially for vegetation cover and bare soil, could signal minor overestimation, especially given that the driving factors did not include climate and economic variables.

Judging from the LULC prediction results, it is observed that areas adjacent to the main road networks were more prone to change to urban areas. This observation is also linked to distance to commercial land use since high-intensity built-up areas are often located within the neighborhoods of the main road networks [102]. The fact is that LULC change modeling requires a deep understanding of the driving factors [86], historical changes [103] and environmental predictors [104]. Extensive research has modeled the effects of physical factors, including elevation, slope and aspect [86,98,105], on LULC changes. Several attempts have explored the importance of proximity factors in LULC changes. Among others, distance to roads [94,106], distance to water bodies and rivers and distance to cities [2,68] were the most important factors. Thus, understanding the changes and dynamics of LULC is challenging and requires modeling techniques utilized with spatiotemporal data [99]. Likewise, the selection of the most suitable predictors which detect the nature and structure of LULC and determine the pattern of changes is essential in LULC change modeling.

To improve the current results, a more integrated model is still needed to overcome the limitations in simulating human behavior as well as policies in LULC change prediction. Thus, future studies will consider multiple-scenario simulations and predictions to appropriately address the uncertainty in land development problems relating to land-use policies [50]. Furthermore, socio-economic factors that affect the urban expansion process need to be included since human behaviors and their interactions with natural and social changes also play critical roles in the dynamic process. Similarly, the inclusion of climate variables should be considered as a boundary condition for improving the prediction of natural land cover such as vegetation and water classes. Although adding more variables is preferable as it is expected to reduce errors, some of the data may not be accessible all the time. In addition, there are model limitations in which nonlinear and qualitative variables cannot be incorporated. Lastly, for a given case study focal and local transition rules should be considered, including the global transition rule for CA models, in which different areas may be subjected to different urban expansion policies and dynamics.

4.3. Insights into LULC Change and Land–Water Sustainability

In general, human activities characterized by urban development, agricultural activities and deforestation systematically leads to LULC changes, which then results in environmental changes that impacts on the earth–atmosphere interactions and sustainable development [107]. The increasing population and related human activities coupled with climate change have progressively led to the decline in per capita water availability [108]. The decline is estimated to have a 6-fold increase over the past century and is predicted to rise annually at a rate of 1% [109]. The increase in land-use development thus has direct effects on multiple environmental aspects, which are interwoven with sustainable development [110,111]. In terms of water resources, LULC changes can impact different hydrological processes and responses [112]. This means that the accurate quantification of the various hydrological factors under varied catchment conditions is critical for continuous and sustainable management of water resources.

Figure 13 shows the land-cover trends and dynamics within the Gaborone dam catchment following from the LULC classification (1986–2019) and LULC prediction (2019–2030). The increase in population within the catchment, as characterized by an increase in built-up area, can be correlated to the decreases in shrub/grassland and cropland. The unprecedented increase in population results in numerous environmental and human problems, such as disturbance in economic development, poor livelihood and law and order situations, limiting the available water supply and deteriorating of its quality and contributing to climate change [113]. The competing demands on land use as observed by the gains and losses in Figure 13 can also lead to conflicts in terms of demand for water supply, for example, between companies and the resident population, in the form of conversion of agricultural land to residential and/or the conversions of natural habitats for agricultural production. These conflicts resulting from different land-use demands and options have been observed to increase in most parts of the world in recent years [111,114]. This means that land use can be understood as the functional dimension of land for different human demands and goals, and its continuous monitoring is necessary for sustainable planning and development.

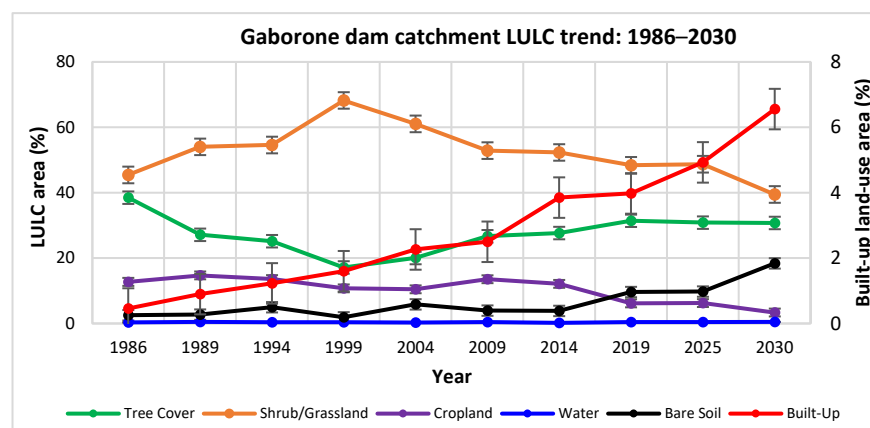


Figure 13. Trends in LULC change in Gaborone dam catchment from 1986 to 2030.

LULC changes through urbanization and deforestation have direct influence on hydrological patterns such as infiltration, evaporation and runoff [115,116]. Figure 14 presents the relationship between the urban growth rate and the dam water capacity from 1986 to 2020. Assuming the dam relies on the catchment for its water supply, the trends shows that the urban population, which has the highest demand for water, continues to grow while the dam water capacity is observed to have a general decline over time. The decreasing pattern in dam water surface area illustrates a trend towards stress on the dam water availability, which can also be attributed to decreasing precipitation combined with increasing consumption demands. While the analysis of impacts of land use on water resources is

not the goal of this study, the identification of the general trends of LULC change and their correlations with water resources will assist the local management authorities to enable a sustainable use of the available water resources. This implies that land, water and environmental managers should consistently quantify and analyze the LULC changes (Figure 13) within the catchment and how the changes are related to or influence the dam water availability (Figure 14).

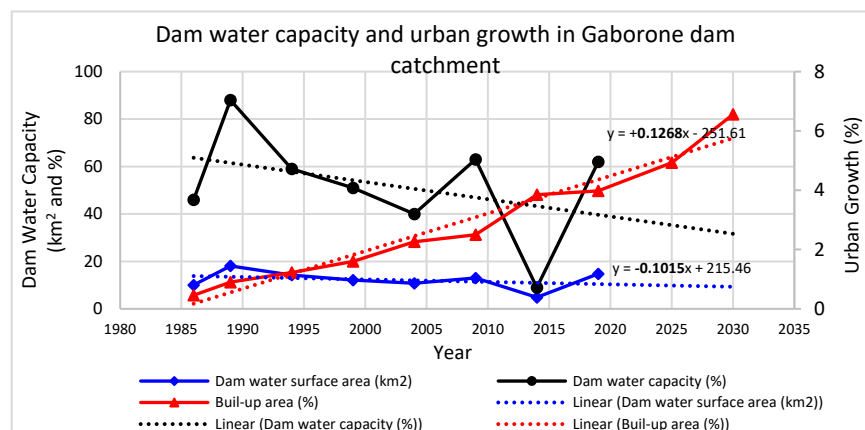


Figure 14. Relationship between urban land-use development and water supply in the Gaborone dam catchment.

In recent decades, catchments have witnessed significant transformations stemming from the interplay of socio-economic processes, demographic dynamics and climate/environmental shifts, which have reshaped the layout of sustainable development. There is therefore the need to understand the social–ecological system so as to strengthen strategies that ensure sustainable socio-economic benefits to local people, while minimizing ecosystem degradation to allow for the sustainable utilization and protection of the resource base. As such, there is need for targeted policies and instruments for addressing vulnerable regions, such as catchments. Information on land use and land management at basin scales is not only important in managing the water resource but also for water allocation for an array of interrelated sectors (e.g., water supply, agriculture, energy production, ecosystems, forest management, energy production, etc.), as well as basin environmental management and protection, the design and operation of water infrastructure and broader issues related to water resource development, planning policy and governance. Therefore, accurate information on land use and its prediction are not only important for planning and management of land-use practice but also in determining how its evolution impacts on water availability and supply.

The integrated understanding of the temporal dynamics of the two resources is important for sustainable land management (SLM) and sustainable water management (SWM). To meet the UN goals for sustainability of SLM and SWM, countries need to set specific land and water policies and targets [117]. Such policies will ensure sustainable environmental planning, monitoring, identification of possible threats and the formulation of mitigation and adaptation strategies [8,9]. In addition to the UN’s established international goals for land use, countries need to set specific land-use policies and targets within the scope of their relevant domestic SDG strategies [117].

By comprehensively analyzing the existing interactions between land and water, different scenarios for a new land–water nexus can be developed for the advancement of relevant guiding targets as set by UNEP, FAO, OECD and the World Energy Outlook. The nexus can enable the mapping of the current and future hot spots of available resources as well as their productivity across different sectors. From such “nexus maps”, planners and managers can formulate means and ways of managing and improving resource productivity [1]. Above all, suitable models are required for simulation of future patterns, to predict growth rates

and for accurate understanding of the cause–effect relationships of the forces driving LULC change within catchments.

5. Conclusions

Prediction of future LULC scenarios within hydrological units such as water catchments is important for effective land use planning, sustainable management of land and water resources, and for planning towards climate adaptation and mitigation for overall sustainable management. This paper presents results of the simulation and prediction of LULC change within the Gaborone dam catchment (Botswana) by comparing two hybrid models, LR–CA and ANN–CA, and the proposed random forest regression (RFR) machine learning model for catchment land-use change modeling. For LULC change detection within the catchment, land-use classification of Landsat TM and ETM+ was carried out from 1986–2019 using the random forest classifier. In the generation of the LULC change transition area matrix and the corresponding change transition possibility matrix, the RFR outperformed LR and the ANN, obtaining a higher percentage of accuracy in the validation of simulated LULC for the previous years with respective accuracies of 84.9%, 62.1% and 60.7%. For the prediction of the LULC change scenarios in the catchment for the 2019–2025 and 2025–2030 periods, physiographic factors (elevation, slope and aspect) and proximity-neighborhood variables (distances to water bodies, roads and urban areas) were integrated as land-use change drivers in the models. From the prediction results, RFR detected a net increase of 2.58% in built-up area over 1 years for 2019–2025 (0.95%) and 2025–2030 (1.63%), while for the same period, LR-CA and ANN-CA predicted a net increase in built-up area by nearly the same rates of 0.15% and 0.14%, respectively. RFR predictions showed bare soil cover increasing to 8.9%, which was nearly equivalent to the combined predicted decrease in tree cover and grassland at 9.5%. ANN-CA and LR-CA detected equal marginal decreases in vegetation cover (−0.02%) and increases in bare soil cover at 0.55%, while cropland was predicted to have a net decrease in percent cover by LR-CA (−0.76%), ANN-CA (−0.77%) and RFR (−2.84%). The study results shows that the proposed RFR is suitable for the prediction of LULC patterns in complex catchment environments as it can capture the potential interactions between the type and intensity of LULC classes and their interactions with the driving parameters. Given that improving land-use management and efficiency is a necessary step towards achieving specific SDGs (e.g., SDG 11 and SDG 15), the approach and results in this study hold significant potential for informed policy formulation and decision making for sustainable catchment monitoring and development, as it provides accurate spatial, quantitative and qualitative LULC change scenarios and their relationships with dam water availability at the catchment scale. The demonstrated relationship between built-up land use and dam water is important in addressing the cross-sectoral links between resources for increased and efficient monitoring and exploitation of resources at local scales. Overall, the findings of this study can aid environmental and urban planners and policymakers to better understand the nature and potential patterns of land-use change within the catchment and the critical factors that drive the LULC change processes. This can aid in the formulation of more appropriate and applicable policies and strategies for mitigating any impacts from the LULC transformations. In addition to the land–water nexus, it is recommended that research should be carried out on the impact assessment of LULC changes on water quality and quantity and the integration of micro-climate changes within the catchment.

Author Contributions: Conceptualization and Methodology; Y.O.O.; Project Administration, Y.O.O., B.N., P.O., D.B.M., G.A. and B.P.; Funding Acquisition, Y.O.O. and J.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was funded by both the USAID Partnerships for Enhanced Engagement in Research (PEER) under the PEER program cooperative agreement number: AID-OAA-A-11-00012 and the University of Botswana, Office of Research and Development (ORD).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in this study were obtained from the United States Geological Survey (USGS): <https://earthexplorer.usgs.gov/> (accessed on 12 July 2023 and 24 November 2022). The rest of the data are as presented in this paper. The image data classification and analyses were carried out within the Google Earth Engine (GEE).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Hoff, H.; Iceland, C.; Kuylenstierna, J.; Te Velde, D.W. Managing the water-land-energy nexus for sustainable development. *UN Chron.* **2012**, *49*, 4. [\[CrossRef\]](#)
- Abbas, Z.; Yang, G.; Zhong, Y.; Zhao, Y. Spatiotemporal change analysis and future scenario of LULC using the CA-ANN approach: A case study of the greater bay area, China. *Land* **2021**, *10*, 584. [\[CrossRef\]](#)
- Tirumala, R.D.; Tiwari, P. Importance of Land in SDG Policy Instruments: A Study of ASEAN Developing Countries. *Land* **2022**, *11*, 218. [\[CrossRef\]](#)
- Islam, K.; Rahman, M.F.; Jashimuddin, M. Modeling land use change using cellular automata and artificial neural network: The case of Chunati Wildlife Sanctuary, Bangladesh. *Ecol. Indic.* **2018**, *88*, 439–453. [\[CrossRef\]](#)
- Goldewijk, K.K. Estimating global land use change over the past 300 years: The HYDE database. *Glob. Biogeochem. Cycles* **2001**, *15*, 417–433. [\[CrossRef\]](#)
- Yirsaw, E.; Wu, W.; Shi, X.; Temesgen, H.; Bekele, B. Land use/land cover change modeling and the prediction of subsequent changes in ecosystem service values in a coastal area of China, the Su-Xi-Chang Region. *Sustainability* **2017**, *9*, 1204. [\[CrossRef\]](#)
- Abuelaish, B.; Olmedo, M.T.C. Scenario of land use and land cover change in the Gaza Strip using remote sensing and GIS models. *Arab. J. Geosci.* **2016**, *9*, 274. [\[CrossRef\]](#)
- Cao, M.; Chang, L.; Ma, S.; Zhao, Z.; Wu, K.; Hu, X.; Gu, Q.; Lü, G.; Chen, M. Multi-scenario simulation of land use for sustainable development goals. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 2119–2127. [\[CrossRef\]](#)
- Duveiller, G.; Caporaso, L.; Abad-Viñas, R.; Perugini, L.; Grassi, G.; Arnet, A.; Cescatti, A. Local biophysical effects of land use and land cover change: Towards an assessment tool for policy makers. *Land Use Policy* **2020**, *91*, 104382. [\[CrossRef\]](#)
- Davin, E.L.; Rechid, D.; Breil, M.; Cardoso, R.M.; Coppola, E.; Hoffmann, P.; Jach, L.L.; Katragkou, E.; de Noblet-Ducoudré, N.; Radtke, K.; et al. Biogeophysical impacts of forestation in Europe: First results from the LUCAS (Land Use and Climate Across Scales) regional climate model intercomparison. *Earth Syst. Dyn.* **2020**, *11*, 183–200. [\[CrossRef\]](#)
- Cao, J.; Zhang, X.; Deo, R.; Gong, Y.; Feng, Q. Influence of stand type and stand age on soil carbon storage in China's arid and semi-arid regions. *Land Use Policy* **2018**, *78*, 258–265. [\[CrossRef\]](#)
- Zhu, E.; Deng, J.; Zhou, M.; Gan, M.; Jiang, R.; Wang, K.; Shahtahmassebi, A. Carbon emissions induced by land-use and land-cover change from 1970 to 2010 in Zhejiang, China. *Sci. Total Environ.* **2019**, *646*, 930–939. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Koning, G.H.J.; Benítez, P.C.; Muñoz, F.; Olschewski, R. Modelling the impacts of payments for biodiversity conservation on regional land-use patterns. *Landsc. Urban Plan.* **2007**, *83*, 255–267. [\[CrossRef\]](#)
- Xinyang, Y.; Qiang, Z.; Xiaomin, Y.; Sheng, W.; Xueyuan, R.; Funian, Z. An overview of distribution characteristics and formation mechanisms in global arid areas. *Adv. Earth Sci.* **2019**, *34*, 826.
- Meaza, H.; Tsegaye, D.; Nyssen, J. Allocation of degraded hillsides to landless farmers and improved livelihoods in Tigray, Ethiopia. *Nor. Geogr. Tidsskr. Nor. J. Geogr.* **2016**, *70*, 1–12. [\[CrossRef\]](#)
- Hoyer, R.; Chang, H. Assessment of freshwater ecosystem services in the Tualatin and Yamhill basins under climate change and urbanization. *Appl. Geogr.* **2014**, *53*, 402–416. [\[CrossRef\]](#)
- United Nations. *The Sustainable Development Goals Report 2016*; UN Department of Economic and Social Affairs: New York, NY, USA, 2016.
- Li, F.; Yigitcanlar, T.; Nepal, M.; Nguyen, K.; Dur, F. Machine Learning and Remote Sensing Integration for Leveraging Urban Sustainability: A Review and Framework. *Sustain. Cities Soc.* **2023**, *96*, 104653. [\[CrossRef\]](#)
- Milojevic-Dupont, N.; Creutzig, F. Machine learning for geographically differentiated climate change mitigation in urban areas. *Sustain. Cities Soc.* **2021**, *64*, 102526. [\[CrossRef\]](#)
- Mokhtari, Z.; Amani-Beni, M.; Asgarian, A.; Russo, A.; Qureshi, S.; Karami, A. Spatial prediction of the urban inter-annual land surface temperature variability: An integrated modeling approach in a rapidly urbanizing semi-arid region. *Sustain. Cities Soc.* **2023**, *93*, 104523. [\[CrossRef\]](#)
- Shorabeh, S.N.; Kakroodi, A.A.; Firozjaei, M.K.; Minaei, F.; Homaei, M. Impact assessment modeling of climatic conditions on spatial-temporal changes in surface biophysical properties driven by urban physical expansion using satellite images. *Sustain. Cities Soc.* **2022**, *80*, 103757. [\[CrossRef\]](#)
- Liping, C.; Yujun, S.; Saeed, S. Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangxi, China. *PLoS ONE* **2018**, *13*, e0200493. [\[CrossRef\]](#)
- Wang, S.W.; Gebru, B.M.; Lamchin, M.; Kayastha, R.B.; Lee, W.K. Land use and land cover change detection and prediction in the Kathmandu district of Nepal using remote sensing and GIS. *Sustainability* **2020**, *12*, 3925. [\[CrossRef\]](#)

24. Hyandye, C.; Mandara, C.G.; Safari, J. GIS and logit regression model applications in land use/land cover change and distribution in Usangu catchment. *Am. J. Remote Sens.* **2015**, *3*, 6–16. [\[CrossRef\]](#)
25. Aitkenhead, M.J.; Aalders, I.H. Predicting land cover using GIS, Bayesian and evolutionary algorithm methods. *J. Environ. Manag.* **2009**, *90*, 236–250. [\[CrossRef\]](#)
26. Lu, Y.; Wu, P.; Ma, X.; Li, X. Detection and prediction of land use/land cover change using spatiotemporal data fusion and the Cellular Automata–Markov model. *Environ. Monit. Assess.* **2019**, *191*, 68. [\[CrossRef\]](#)
27. Yang, X.; Zheng, X.Q.; Lv, L.N. A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. *Ecol. Model.* **2012**, *233*, 11–19. [\[CrossRef\]](#)
28. Subedi, P.; Subedi, K.; Thapa, B. Application of a hybrid cellularautomaton–Markov (CA–Markov) model in land-use change prediction: A case study of Saddle Creek Drainage Basin, Florida. *Appl. Ecol. Environ. Sci.* **2013**, *1*, 126–132.
29. Stefanov, W.L.; Ramsey, M.S.; Christensen, P.R. Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers. *Remote Sens. Environ.* **2001**, *77*, 173–185. [\[CrossRef\]](#)
30. Ralha, C.G.; Abreu, C.G.; Coelho, C.G.; Zaghetto, A.; Macchiavello, B.; Machado, R.B. A multi-agent model system for land-use change simulation. *Environ. Model. Softw.* **2013**, *42*, 30–46. [\[CrossRef\]](#)
31. Palmate, S.S.; Wagner, P.D.; Fohrer, N.; Pandey, A. Assessment of uncertainties in modelling land use change with an integrated cellular automata–Markov chain model. *Environ. Model. Assess.* **2022**, *27*, 275–293. [\[CrossRef\]](#)
32. Jana, A.; Jat, M.K.; Saxena, A.; Choudhary, M. Prediction of land use land cover changes of a river basin using the CA–Markov model. *Geocarto Int.* **2022**, *37*, 14127–14147. [\[CrossRef\]](#)
33. Sinha, P.; Kumar, L. Markov land cover change modeling using pairs of time-series satellite images. *Photogramm. Eng. Remote Sens.* **2013**, *79*, 1037–1051. [\[CrossRef\]](#)
34. Khan, A.M.; Li, Q.; Saqib, Z.; Khan, N.; Habib, T.; Khalid, N.; Majeed, M.; Tariq, A. MaxEnt modelling and impact of climate change on habitat suitability variations of economically important Chilgoza Pine (*Pinus gerardiana* Wall.) in South Asia. *Forests* **2022**, *13*, 715. [\[CrossRef\]](#)
35. Saxena, A.; Jat, M.K.; Kumar, S. Sensitivity analysis and retrieval of optimum SLEUTH model parameters. *Geocarto Int.* **2022**, *37*, 7431–7444. [\[CrossRef\]](#)
36. Puertas, O.L.; Henríquez, C.; Meza, F.J. Assessing spatial dynamics of urban growth using an integrated land use model. Application in Santiago Metropolitan Area, 2010–2045. *Land Use Policy* **2014**, *38*, 415–425.
37. Han, H.; Yang, C.; Song, J. Scenario simulation and the prediction of land use and land cover change in Beijing, China. *Sustainability* **2015**, *7*, 4260–4279. [\[CrossRef\]](#)
38. Keshtkar, H.; Voigt, W.; Alizadeh, E. Land-cover classification and analysis of change using machine-learning classifiers and multi-temporal remote sensing imagery. *Arab. J. Geosci.* **2017**, *10*, 154. [\[CrossRef\]](#)
39. Rimal, B.; Zhang, L.; Keshtkar, H.; Haack, B.N.; Rijal, S.; Zhang, P. Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 154. [\[CrossRef\]](#)
40. Tariq, A.; Mumtaz, F.; Majeed, M.; Zeng, X. Spatio-temporal assessment of land use land cover based on trajectories and cellular automata Markov modelling and its impact on land surface temperature of Lahore district Pakistan. *Environ. Monit. Assess.* **2023**, *195*, 114. [\[CrossRef\]](#)
41. Kocabas, V.; and Dragicevic, S. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Comput. Environ. Urban Syst.* **2006**, *30*, 921–953. [\[CrossRef\]](#)
42. Ghosh, P.; Mukhopadhyay, A.; Chanda, A.; Mondal, P.; Akhand, A.; Mukherjee, S.; Nayak, S.K.; Ghosh, S.; Mitra, D.; Ghosh, T.; et al. Application of Cellular automata and Markov-chain model in geospatial environmental modeling–A review. *Remote Sens. Appl. Soc. Environ.* **2017**, *5*, 64–77. [\[CrossRef\]](#)
43. Felegari, S.; Sharifi, A.; Moravej, K.; Golchin, A.; Tariq, A. Investigation of the relationship between ndvi index, soil moisture, and precipitation data using satellite images. *Sustain. Agric. Syst. Technol.* **2022**, 314–325.
44. Aburas, M.M.; Ho, Y.M.; Ramli, M.F.; Ash’aari, Z.H. Improving the capability of an integrated CA–Markov model to simulate spatio-temporal urban growth trends using an Analytical Hierarchy Process and Frequency Ratio. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *59*, 65–78. [\[CrossRef\]](#)
45. Chaplot, V.; Le Brozec, E.C.; Silvera, N.; Valentin, C. Spatial and temporal assessment of linear erosion in catchments under sloping lands of northern Laos. *Catena* **2005**, *63*, 167–184. [\[CrossRef\]](#)
46. Shahbazian, Z.; Faramarzi, M.; Rostami, N.; Mahdizadeh, H. Integrating logistic regression and cellular automata–Markov models with the experts’ perceptions for detecting and simulating land use changes and their driving forces. *Environ. Monit. Assess.* **2019**, *191*, 422. [\[CrossRef\]](#)
47. Gu, G.; Wu, B.; Zhang, W.; Lu, R.; Feng, X.; Liao, W.; Pang, C.; Lu, S. Comparing machine learning methods for predicting land development intensity. *PLoS ONE* **2023**, *18*, e0282476. [\[CrossRef\]](#)
48. Ozturk, D. Urban growth simulation of Atakum (Samsun, Turkey) using cellular automata–Markov chain and multi-layer perceptron–Markov chain models. *Remote Sens.* **2015**, *7*, 5918–5950. [\[CrossRef\]](#)
49. Shafizadeh-Moghadam, H.; Asghari, A.; Tayyebi, A.; Taleai, M. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Comput. Environ. Urban Syst.* **2017**, *64*, 297–308. [\[CrossRef\]](#)
50. Arsanjani, J.J.; Helbich, M.; Kainz, W.; Boloorani, A.D. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 265–275. [\[CrossRef\]](#)

51. Rienow, A.; Mustafa, A.; Krelaus, L.; Lindner, C. Modeling urban regions: Comparing random forest and support vector machines for cellular automata. *Trans. GIS* **2021**, *25*, 1625–1645. [\[CrossRef\]](#)
52. Gounaridis, D.; Chorianopoulos, I.; Symeonakis, E.; Koukoulas, S. A Random Forest-Cellular Automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales. *Sci. Total Environ.* **2019**, *646*, 320–335. [\[CrossRef\]](#)
53. Zhang, X.; Zhou, J.; Song, W. Simulating urban sprawl in china based on the artificial neural network-cellular automata-Markov model. *Sustainability* **2020**, *12*, 4341. [\[CrossRef\]](#)
54. Roy, B.; Rahman, M.Z. Spatio-temporal analysis and cellular automata-based simulations of biophysical indicators under the scenario of climate change and urbanization using artificial neural network. *Remote Sens. Appl. Soc. Environ.* **2023**, *31*, 100992. [\[CrossRef\]](#)
55. Cuellar, Y.; Perez, L. Assessing the accuracy of sensitivity analysis: An application for a cellular automata model of Bogota's urban wetland changes. *Geocarto Int.* **2023**, *38*, 2186491. [\[CrossRef\]](#)
56. Zhang, X.; Ren, W.; Peng, H. Urban land use change simulation and spatial responses of ecosystem service value under multiple scenarios: A case study of Wuhan, China. *Ecol. Indic.* **2022**, *144*, 109526. [\[CrossRef\]](#)
57. Ambarwulan, W.; Yulianto, F.; Widiatmaka, W.; Rahadiati, A.; Tarigan, S.D.; Firmansyah, I.; Hasibuan, M.A.S. Modelling land use/land cover projection using different scenarios in the Cisadane Watershed, Indonesia: Implication on deforestation and food security. *Egypt. J. Remote Sens. Space Sci.* **2023**, *26*, 273–283. [\[CrossRef\]](#)
58. Pourmohammadi, P.; Adjero, D.A.; Strager, M.P.; Farid, Y.Z. Predicting developed land expansion using deep convolutional neural networks. *Environ. Model. Softw.* **2020**, *134*, 104751. [\[CrossRef\]](#)
59. Fernald, A.; Tidwell, V.; Rivera, J.; Rodríguez, S.; Guldán, S.; Steele, C.; Ochoa, C.; Hurd, B.; Ortiz, M.; Koykin, K.; et al. Modeling sustainability of water, environment, livelihood, and culture in traditional irrigation communities and their linked watersheds. *Sustainability* **2012**, *4*, 2998–3022. [\[CrossRef\]](#)
60. Hu, X.; Li, X.; Lu, L. Modeling the land use change in an arid oasis constrained by water resources and environmental policy change using cellular automata models. *Sustainability* **2018**, *10*, 2878. [\[CrossRef\]](#)
61. Kamusoko, C.; Gamba, J. Simulating urban growth using a random forest-cellular automata (RF-CA) model. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 447–470. [\[CrossRef\]](#)
62. Cetin, M.; Demirel, H. Modelling and simulation of urban dynamics. *Fresenius Environ. Bull.* **2010**, *9*, 2348–2353.
63. Fu, X.; Wang, X.; Yang, Y.J. Deriving suitability factors for CA-Markov land use simulation model based on local historical data. *J. Environ. Manag.* **2018**, *206*, 10–19. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Berberoğlu, S.; Akin, A.; Clarke, K.C. Cellular automata modeling approaches to forecast urban growth for Adana, Turkey: A comparative approach. *Landsc. Urban Plan.* **2016**, *153*, 11–27. [\[CrossRef\]](#)
65. Pijanowski, B.C.; Tayyebi, A.; Doucette, J.; Pekin, B.K.; Braun, D.; Plourde, J. A big data urban growth simulation at a national scale: Configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment. *Environ. Model. Softw.* **2014**, *51*, 250–268. [\[CrossRef\]](#)
66. Sajan, B.; Mishra, V.N.; Kanga, S.; Meraj, G.; Singh, S.K.; Kumar, P. Cellular automata-based artificial neural network model for assessing past, present, and future land use/land cover dynamics. *Agronomy* **2022**, *12*, 2772. [\[CrossRef\]](#)
67. Saputra, M.H.; Lee, H.S. Prediction of land use and land cover changes for North Sumatra, Indonesia, using an artificial-neural-network-based cellular automaton. *Sustainability* **2019**, *11*, 3024. [\[CrossRef\]](#)
68. Gharaibeh, A.; Shaamala, A.; Obeidat, R.; Al-Kofahi, S. Improving land-use change modeling by integrating ANN with Cellular Automata-Markov Chain model. *Heliyon* **2020**, *6*, e05092. [\[CrossRef\]](#)
69. Ackerschott, A.; Kohlhase, E.; Vollmer, A.; Hörisch, J.; von Wehrden, H. Steering of land use in the context of sustainable development: A systematic review of economic instruments. *Land Use Policy* **2023**, *129*, 106620. [\[CrossRef\]](#)
70. Ouma, Y.O.; Keitsile, A.; Nkwae, B.; Odirile, P.; Moalafhi, D.; Qi, J. Urban land-use classification using machine learning classifiers: Comparative evaluation and post-classification multi-feature fusion approach. *Eur. J. Remote Sens.* **2023**, *56*, 2173659. [\[CrossRef\]](#)
71. Ouma, Y.; Nkwae, B.; Moalafhi, D.; Odirile, P.; Parida, B.; Anderson, G.; Qi, J. Comparison of machine learning classifiers for multitemporal and multisensor mapping of urban LULC features. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *43*, 681–689. [\[CrossRef\]](#)
72. Biau, G.Š.; Scornet, E. A random forest guided tour. *TEST* **2016**, *25*, 197–227. [\[CrossRef\]](#)
73. Li, X.; Yeh, A.G.O. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 323–343. [\[CrossRef\]](#)
74. Jogun, T.; Lukić, A.; Gašparović, M. Simulation model of land cover changes in a post-socialist peripheral rural area: Požega-Slavonia County, Croatia. *Croat. Geogr. Bull.* **2019**, *81*, 31–59. [\[CrossRef\]](#)
75. Santé, I.; García, A.M.; Miranda, D.; Crecente, R. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landsc. Urban Plan.* **2010**, *96*, 108–122. [\[CrossRef\]](#)
76. Memarian, H.; Balasundram, S.K.; Talib, J.B.; Sung, C.T.B.; Sood, A.M.; Abbaspour, K. Validation of CA-Markov for simulation of land use and cover change in the Langat Basin, Malaysia. *J. Geogr. Inf. Syst.* **2012**, *4*, 26322. [\[CrossRef\]](#)
77. He, D.; Zhou, J.; Gao, W.; Guo, H.Y.U.S.; Yu, S.; Liu, Y. An integrated CA-markov model for dynamic simulation of land use change in Lake Dianchi watershed. *Acta Sci. Nat. Univ. Pekin.* **2014**, *50*, 1095–1105.

78. Roodposhti, M.S.; Aryal, J.; Bryan, B.A. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. *Environ. Model. Softw.* **2019**, *112*, 70–81. [\[CrossRef\]](#)
79. Hewitt, R.; and Díaz-Pacheco, J. Stable models for metastable systems? Lessons from sensitivity analysis of a cellular automata urban land use model. *Comput. Environ. Urban Syst.* **2017**, *62*, 113–124. [\[CrossRef\]](#)
80. Wu, X.; Liu, X.; Zhang, D.; Zhang, J.; He, J.; Xu, X. Simulating mixed land-use change under multi-label concept by integrating a convolutional neural network and cellular automata: A case study of Huizhou, China. *GIScience Remote Sens.* **2022**, *59*, 609–632. [\[CrossRef\]](#)
81. Roy, B. A machine learning approach to monitoring and forecasting spatio-temporal dynamics of land cover in Cox's Bazar district, Bangladesh from 2001 to 2019. *Environ. Chall.* **2021**, *5*, 100237. [\[CrossRef\]](#)
82. Gahegan, M. On the application of inductive machine learning tools to geographical analysis. *Geogr. Anal.* **2000**, *32*, 113–139. [\[CrossRef\]](#)
83. Yan, W.; Meimei, X.; Yujia, T.; Huan, G.; Rong, C.; Zhiyi, S.; Xinying, A. Prediction and Early Warning Model for Environmental Data and Circulatory System Disease Death with Machine Learning. *Data Anal. Knowl. Discov.* **2022**, *6*, 79–92.
84. Liu, J.; Shao, Q.; Yan, X.; Fan, J.; Zhan, J.; Deng, X.; Kuang, W.; Huang, L. The climatic impacts of land use and land cover change compared among countries. *J. Geogr. Sci.* **2016**, *26*, 889–903. [\[CrossRef\]](#)
85. Tong, X.; Feng, Y. A review of assessment methods for cellular automata models of land-use change and urban growth. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 866–898. [\[CrossRef\]](#)
86. Xu, Q.; Wang, Q.; Liu, J.; Liang, H. Simulation of land-use changes using the partitioned ANN-CA model and considering the influence of land-use change frequency. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 346. [\[CrossRef\]](#)
87. Grinblat, Y.; Gilichinsky, M.; Benenson, I. Cellular automata modeling of land-use/land-cover dynamics: Questioning the reliability of data sources and classification methods. *Ann. Am. Assoc. Geogr.* **2016**, *106*, 1299–1320. [\[CrossRef\]](#)
88. Guan, D.; Zhao, Z.; Tan, J. Dynamic simulation of land use change based on logistic-CA-Markov and WLC-CA-Markov models: A case study in three gorges reservoir area of Chongqing, China. *Environ. Sci. Pollut. Res.* **2019**, *26*, 20669–20688. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Siddiqui, A.; Siddiqui, A.; Maithani, S.; Jha, A.K.; Kumar, P.; Srivastav, S.K. Urban growth dynamics of an Indian metropolitan using CA Markov and Logistic Regression. *Egypt. J. Remote Sens. Space Sci.* **2018**, *21*, 229–236. [\[CrossRef\]](#)
90. He, J.; Li, X.; Yao, Y.; Hong, Y.; Jinbao, Z. Mining transition rules of cellular automata for simulating urban expansion by using the deep learning techniques. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 2076–2097. [\[CrossRef\]](#)
91. Saadani, S.; Laajaj, R.; Maanan, M.; Rhinane, H.; Aaroud, A. Simulating spatial-temporal urban growth of a Moroccan metropolitan using CA-Markov model. *Spat. Inf. Res.* **2020**, *28*, 609–621. [\[CrossRef\]](#)
92. Liu, Y.; Cao, X.; Li, T. Identifying driving forces of built-up land expansion based on the geographical detector: A case study of Pearl River Delta urban agglomeration. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1759. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Mozaffaree Pour, N.; Oja, T. Prediction power of logistic regression (LR) and Multi-Layer perceptron (MLP) models in exploring driving forces of urban expansion to be sustainable in Estonia. *Sustainability* **2021**, *14*, 160. [\[CrossRef\]](#)
94. Ullah, S.; Ahmad, K.; Sajjad, R.U.; Abbasi, A.M.; Nazeer, A.; Tahir, A.A. Analysis and simulation of land cover changes and their impacts on land surface temperature in lower Himalayan region. *J. Environ. Manag.* **2019**, *245*, 348–357. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Saha, T.K.; Pal, S.; Sarkar, R. Prediction of wetland area and depth using linear regression model and artificial neural network based cellular automata. *Ecol. Inform.* **2021**, *62*, 101272. [\[CrossRef\]](#)
96. Mohamed, A.; Worku, H. Simulating urban land use and cover dynamics using cellular automata and Markov chain approach in Addis Ababa and the surrounding. *Urban Clim.* **2020**, *31*, 100545. [\[CrossRef\]](#)
97. Cannemi, M.; García-Melón, M.; Aragonés-Beltrán, P.; Gómez-Navarro, T. Modeling decision making as a support tool for policy making on renewable energy development. *Energy Policy* **2014**, *67*, 127–137. [\[CrossRef\]](#)
98. Girma, R.; Fürst, C.; Moges, A. Land use land cover change modeling by integrating artificial neural network with cellular Automata-Markov chain model in Gidabo river basin, main Ethiopian rift. *Environ. Chall.* **2022**, *6*, 100419. [\[CrossRef\]](#)
99. Singh, S.K.; Mustak, S.; Srivastava, P.K.; Szabó, S.; Islam, T. Predicting spatial and decadal LULC changes through cellular automata Markov chain models using earth observation datasets and geo-information. *Environ. Process.* **2015**, *2*, 61–78. [\[CrossRef\]](#)
100. Maithani, S. A neural network based urban growth model of an Indian city. *J. Indian Soc. Remote Sens.* **2009**, *37*, 363–376. [\[CrossRef\]](#)
101. Zhou, Y.; Zhang, F.; Du, Z.; Ye, X.; Liu, R. Integrating cellular automata with the deep belief network for simulating urban growth. *Sustainability* **2017**, *9*, 1786. [\[CrossRef\]](#)
102. Simwanda, M.; Murayama, Y.; Ranagalage, M. Modeling the drivers of urban land use changes in Lusaka, Zambia using multi-criteria evaluation: An analytic network process approach. *Land Use Policy* **2020**, *92*, 104441. [\[CrossRef\]](#)
103. Karimi, F.; Sultana, S.; Babakan, A.S.; Suthaharan, S. An enhanced support vector machine model for urban expansion prediction. *Comput. Environ. Urban Syst.* **2019**, *75*, 61–75. [\[CrossRef\]](#)
104. Msofe, N.K.; Sheng, L.; Lyimo, J. Land use change trends and their driving forces in the Kilombero Valley Floodplain, Southeastern Tanzania. *Sustainability* **2019**, *11*, 505. [\[CrossRef\]](#)
105. Bonilla-Bedoya, S.; Mora, A.; Vaca, A.; Estrella, A.; Herrera, M.Á. Modelling the relationship between urban expansion processes and urban forest characteristics: An application to the Metropolitan District of Quito. *Comput. Environ. Urban Syst.* **2020**, *79*, 101420. [\[CrossRef\]](#)

106. Rahnama, M.R. Forecasting land-use changes in Mashhad Metropolitan area using Cellular Automata and Markov chain model for 2016–2030. *Sustain. Cities Soc.* **2021**, *64*, 102548. [[CrossRef](#)]
107. Turner, B.L.; Lambin, E.F.; Reenberg, A. The emergence of land change science for global environmental change and sustainability. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 0666–20671. [[CrossRef](#)] [[PubMed](#)]
108. Shukla, S.; Meshesha, T.W.; Sen, I.S.; Bol, R.; Bogena, H.; Wang, J. Assessing Impacts of Land Use and Land Cover (LULC) Change on Stream Flow and Runoff in Rur Basin, Germany. *Sustainability* **2023**, *15*, 9811. [[CrossRef](#)]
109. UN-Water. *Water and Climate Change. The United Nations World Water Development Report*; UNESCO: Paris, France, 2020.
110. Kong, X.; Fu, M.; Zhao, X.; Wang, J.; Jiang, P. Ecological effects of land-use change on two sides of the Hu Huanyong Line in China. *Land Use Policy* **2022**, *113*, 105895. [[CrossRef](#)]
111. Lafuite, A.-S.; Denise, G.; Loreau, M. Sustainable land-use management under biodiversity lag effects. *Ecol. Econ.* **2018**, *154*, 272–281. [[CrossRef](#)]
112. Welde, K.; Gebremariam, B. Effect of land use land cover dynamics on hydrological response of watershed: Case study of Tekeze Dam watershed, northern Ethiopia. *Int. Soil Water Conserv. Res.* **2017**, *5*, 1–16. [[CrossRef](#)]
113. Murmu, P.; Kumar, M.; Lal, D.; Sonker, I.; Singh, S.K. Delineation of groundwater potential zones using geospatial techniques and analytical hierarchy process in Dumka district, Jharkhand, India. *Groundw. Sustain. Dev.* **2019**, *9*, 100239. [[CrossRef](#)]
114. Zachrisson, A.; Björstig, T.; Thellbro, C.; Neumann, W.; Svensson, J. Participatory comprehensive planning to handle competing land-use priorities in the sparsely populated rural context. *J. Rural Stud.* **2021**, *88*, 1–13. [[CrossRef](#)]
115. Loveland, T.R.; Mahmood, R. A design for a sustained assessment of climate forcing and feedbacks related to land use and land cover change. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 1563–1572. [[CrossRef](#)]
116. Scanlon, B.R.; Reedy, R.C.; Stonestrom, D.A.; Prudic, D.E.; Dennehy, K.F. Impact of land use and land cover change on groundwater recharge and quality in the southwestern US. *Glob. Change Biol.* **2005**, *11*, 1577–1593. [[CrossRef](#)]
117. Pauliuk, S.; Heeren, N. Material efficiency and its contribution to climate change mitigation in Germany: A deep decarbonization scenario analysis until 2060. *J. Ind. Ecol.* **2021**, *25*, 479–493. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.