



Spatio-temporal prediction of land use and land cover change in Bahi (Manyoni) Catchment, Tanzania, using multilayer perceptron neural network and cellular automata-Markov chain model

Naima A. M. Hersi ·
Deogratias M. M. Mulungu · Joel Nobert

Received: 22 May 2023 / Accepted: 30 November 2023 / Published online: 8 December 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract Evaluation of land use and land cover (LULC) change is among vital tools used for tracking environmental health and proper resource management. Remote sensing data was used to determine LULC change in Bahi (Manyoni) Catchment (BMC) in central Tanzania. Landsat satellite images from Landsat 5 TM and Landsat 8 OLI/TIRS were used, and support vector machine (SVM) algorithm was applied to classify the features of BMC. The obtained kappa values were 0.74, 0.83 and 0.84 for LULC maps of 1985, 2005 and 2021, respectively, which indicates the degree of accuracy from produced being substantial to almost perfect. Classified maps along with geospatial, socio-economic and climatic drivers with sufficient explanatory power were incorporated into MLP-NN to produce transition potential maps. Transition maps were subsequently used in cellular automata (CA)-Markov chain model to predict future LULC for BMC in immediate-future (2035), mid-future (2055) and far-future (2085). The findings indicate BMC is expected to experience significant

expansion of agricultural lands and built land from 31.89 to 50.16% and 1.48 to 9.1% from 2021 to 2085 at the expense of open woodland, shrubland and savanna grassland. Low-yield crop production, water scarcity and population growth were major driving forces for rapid expansion of agricultural lands and overall LULC in BMC. The findings are essential for understanding the impact of LULC on hydrological processes and offer insights for the internal drainage basin (IDB) board to make necessary measures to lessen the expected dramatic changes in LULC in the future while sustaining harmonious balance with livelihood activities.

Keywords CA-Markov chain model · LULC change detection · MLP neural network (MLP-NN) · Semi-arid region · Support vector machine (SVM) · Tanzania

Introduction

Land use and land cover (LULC) change begun since the dawn of civilization. It was primarily driven by human needs for survival. Human converted landscapes to derive their basic needs including food, water, shelter and medicinal products. Regions with favourable terrains, slope condition, soil type and climate were prone to LULC changes (Tewabe & Fentahun, 2020). Concern of spontaneous human expansion of agricultural and pastoral industry at

N. A. M. Hersi (✉) · D. M. M. Mulungu · J. Nobert
Department of Water Resources Engineering, College of Engineering and Technology, University of Dar Es Salaam, P.O. Box 35131, Dar Es Salaam, Tanzania
e-mail: naima.hersi@gmail.com

N. A. M. Hersi
Department of Environmental Engineering and Management, College of Earth Sciences and Engineering, The University of Dodoma, P.O. Box 11090, Dodoma, Tanzania

the expense of natural land cover was voiced out as early as the 1860s (Marsh, 1864). A gain of 4.8 billion ha of agricultural land (cropland and grazeland) has costed the world 2 billion ha of forest and nearly 2.9 billion ha of wild grassland and shrubs (Ritchie & Roser, 2021). It was estimated by Intergovernmental Panel on Climate Change (IPCC)'s Special Report on land use, land use change and forestry that such loss particularly since 1750 to be responsible for one-third of global warming (Watson et al., 2000). Furthermore, LULC is recognized to be a primary driver of land degradation, hydrological system alterations, decreased groundwater recharge, soil degradation and irreversible species extinction (Bogale, 2020; Mengistu et al., 2022; Olorunfemi et al., 2022). Africa is particularly vulnerable to these impacts of LULC changes. It was estimated nearly 65% of ecosystem service value (ESV) in Sub-Saharan Africa is lost due to the degradation of forests and croplands (Fenta et al., 2020) leaving over 1.3 billion people stranded on deteriorating agricultural land (UNCCD, 2017). This poses a barrier to the continent's efforts to achieve food security and improved lives.

Efforts to conserve the environment has been done through legally binding multinational agreements as early as 1910 (Mitchell, 2003). The Kyoto Protocol of 1997 and the Paris Agreement of 2015 are two of the most well-known international agreements. These agreements along with several environmental conservation-related campaigns and projects have reduced deforestation and increased global growing stock levels (Nabuurs et al., 2022; Ritchie & Roser, 2021). However, still forest losses outstrip the gains (Hansen, 2013; Ritchie & Roser, 2021). With current exponential population growth, closing the gap between the losses and gains and overall rejuvenation of the landscapes could be challenging especially in regions where considerable population growth is anticipated such as growing cities. It is even far more challenging in regions with unfavourable climate conditions such as areas of aridity with limited rainfall amounts and high erosion (Yildiz et al., 2018). Therefore, understanding the future pace of LULC under these areas is of paramount importance.

There are a wide range of approaches for LULC prediction being developed and applied across many regions (Mishra et al., 2020; Wang et al., 2021). However, numerous LULC prediction studies have used the hybrid cellular automata (CA)-Markov

chain model for future prediction of LULC patterns (Sang et al., 2011; Harris et al., 2014; Gidey et al., 2017; Hamad et al., 2018; Chanapathi & Thatikonda, 2020; Aburas et al., 2021; Gemtzi, 2021; Getachew et al., 2021; Wang et al., 2021; Ghalehtemouri et al., 2022). The CA-Markov chain model has recently gained popularity because of its simplicity and ability to predict a complex LULC system (Gidey et al., 2017; Wang et al., 2021; Girma et al., 2022). CA works under the principle that changes in any particular grid cell is explained by its current state and state of the neighbouring cells (Gidey et al., 2017), whereas Markov model is a random stochastic model that operates on the basis that the future cell state depends on its current state (Ghalehtemouri et al., 2022). Therefore, the hybrid CA-Markov chain model performs under the influence of proximity that state of LULC class is influenced by the dynamic changes of the adjacent classes (Ghalehtemouri et al., 2022; Gidey et al., 2017). Nevertheless, dynamic changes in LULC often exhibit nonlinearity (Ghalehtemouri et al., 2022; Lamchin et al., 2022). Therefore, it becomes imperative to account for the influence of drivers such as socioeconomic, geospatial and climatic variables (Alqadhi et al., 2021; Girma et al., 2022; Siddik et al., 2022). Methods like the multi-layer perceptron neural network (MLP-NN) offer an effective approach to enhancing our understanding of the LULC process (Eastman, 2020a; Girma et al., 2022). The MLP-NN utilizes a backpropagation algorithm that generalizes transition probabilities while considering drivers that impact LULC transitions (Alqadhi et al., 2021; Girma et al., 2022; Lamchin et al., 2022). This consideration of applying MLP-NN contributes to improving the predictive capabilities of the CA-Markov model. Accuracy and reliability of integrated MLP-NN and CA-Markov chain model have been endorsed by numerous scholars (Alqadhi et al., 2021; Ghalehtemouri et al., 2022; Girma et al., 2022; Rahaman et al., 2023). This influences studies pertaining to future LULC change detection to employ integrated MLP-NN and CA-Markov owing to its performance.

Therefore, this study aims to predict future LULC in Bahi (Manyoni) Catchment (BMC) using CA-Markov chain model for the immediate-future (2035), mid-future (2055) and far-future (2085). BMC became of significance since it covers 14% of the new Tanzania's capital city (i.e. Dodoma) in the southeast

of BMC and other rapidly expanding areas, such as 89% of Kondo Town Council in the northeast, 53% of Singida District in the northwest, and 16% of Manyoni District in the central of BMC. Unprecedented immigration (URT, 2012, 2022) from other regions of the country including the former capital city (i.e. Dar es Salaam) is pressing BMC for a hasty expansion of agricultural land and settlements (Kisamba & Li, 2022; Mubako et al., 2022; Mussa et al., 2021). Therefore, keeping close tabs on the LULC is vital before irreversible degradation occurs. Since no study has attempted to predict future LULC for the entire BMC, therefore, the findings from this study will aid the basin board and other decision-making bodies in planning and decision-making about land conservation, sustainable LULC development and natural resource management.

Description of study area

The BMC is one of the nine catchments in IDB covering 24,770.2 km². The geographical location of the BMC falls within latitude 4°07'12" and 6°44'10" South and longitude 34°39'43" and 36°05'50" East. The catchment is located in the northeast to central part of Tanzania and covers 11 districts, including the Dodoma capital city, Babati, Bahi, Chemba, Kondo rural, Kondo urban, Singida rural, Ikungi, Manyoni, Hanang and Itigi (Fig. 1). The BMC has mean elevation of 1235 m above sea level (a.s.l) and is characterized by mostly gentle slopes, with some steep slopes in the northern part of the catchment. The catchment receives a unimodal rainfall pattern with an average rainfall amount of 620 mm/year usually from December to May and an average temperature ranging from 16 to 27 °C. The major socio-economic activities in the catchment are rain-fed farming, with major crops including maize, sunflower, sorghum, millet, paddy and livestock keeping, mostly cattle and small stock of goats, sheep and pigs (Ghotbi et al., 2015).

Materials and methods

Methodological framework employed to guide this study is presented in Fig. 2. Detailed information on image acquisition and processing, image classification, selection of driver variables, model

training, validation and future of prediction of LULC change were further elaborated in the following subsections.

Images acquisition and processing

There are several satellite images available online for LULC studies; however, for this study, images from Landsat program were used since it is the longest running program for Earth observation. Landsat satellite images considered in this study had spectral resolution of 30 m. The selected images were for years 1985, 2005 and 2021 obtained from the United States Geological Survey (USGS) website <https://earthexplorer.usgs.gov/>. Selection of the images was based on the availability from the oldest to the latest year (i.e. when analysis of downloaded images was performed). The images were downloaded from Landsat 5 Thematic Mapper and Landsat 8 OLI/TIRS. Specifically, Landsat images from collection 2 level 2 were used because of improved usage of ground control points (GCPs) and radiometric calibration. Detailed description of Landsat collections and their products are well summarized in <https://www.usgs.gov/media/files/landsat-collection-1-vs-collection-2-summary>. Bands 1 to 5 of Landsat 5 and bands 2 to 6 of Landsat 8 were used to distinguish closely related vegetation based on vegetation health and density, and various land use categories, including roads, built-up areas, water bodies and bare soils. Dry seasons, i.e. from July to October, were prioritized in order to differentiate closely related land use types such as agriculture vs grassland. During this timeframe, the predominant crops in the region, including maize, sunflower, paddy and sorghum, have undergone harvesting leaving the fields clear. This provides noticeable contrast, making it easier to distinguish agricultural land from grassland. Acquired imageries were processed by removing cloud cover using quality assessment (QA) and surface reflectance sub-datasets. To minimize the effects of gap filling in the multi-temporal characteristics, images from adjacent month/year in the same season (July to October) were used to patch gaps created by the cloud removal process. Later, cloud-free imagery tiles were merged and the study area was extracted.

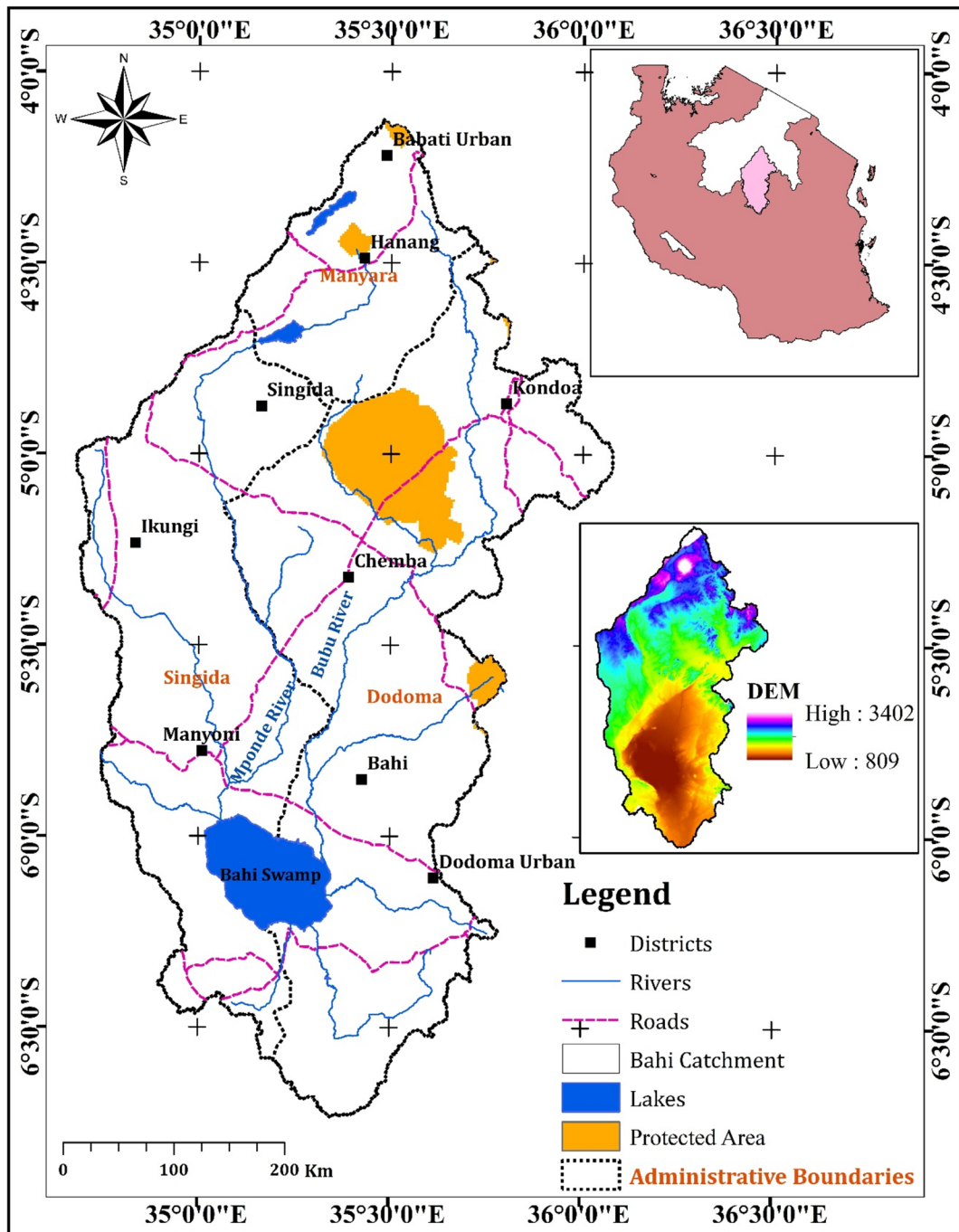


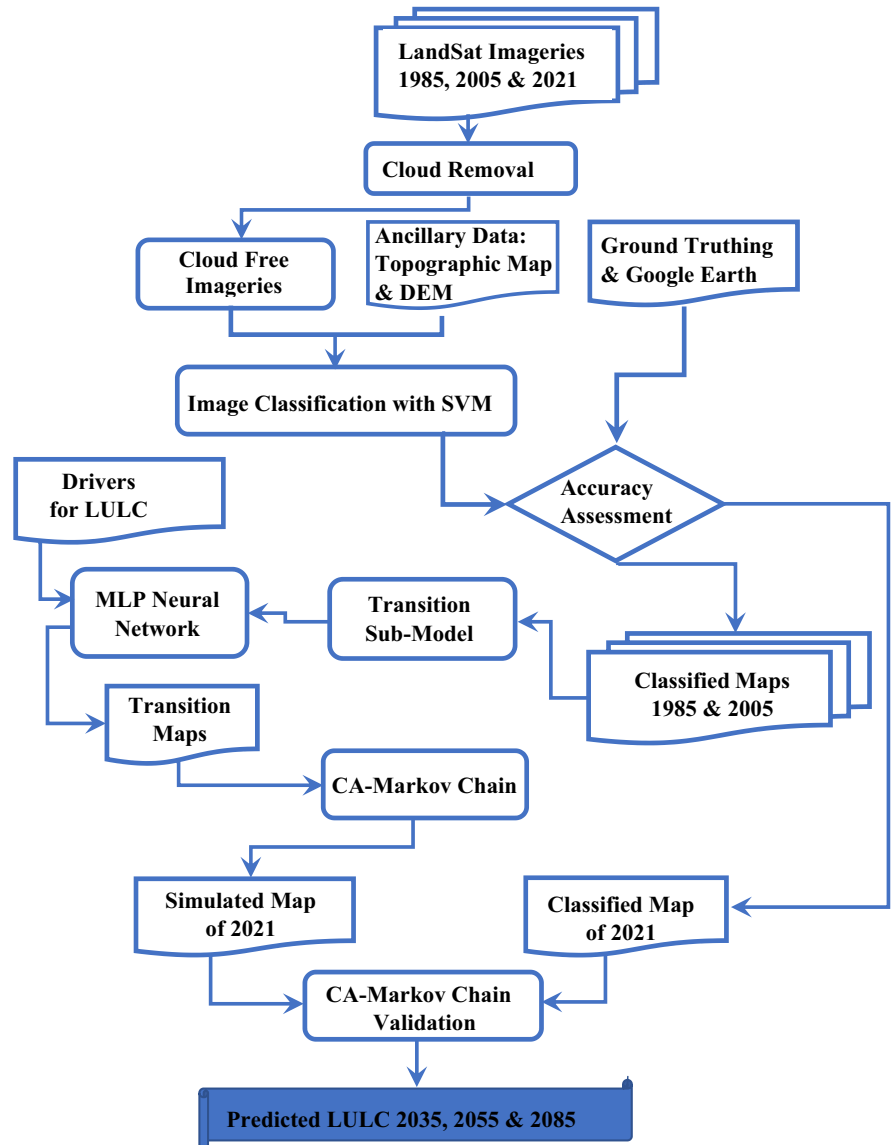
Fig. 1 Location of the BMC

Image classifications

Pixel-based supervised classification was done using support vector machine (SVM) algorithm. SVM

algorithm has been widely used in multiple fields of application (Cervantes et al., 2020). SVM is less susceptible to noise and unbalanced training samples and sizes (Boateng et al., 2020; Noi & Kappas,

Fig. 2 Methodological framework for LULC change prediction



2017). LULC schema was based on type of LULC described in Mainland Tanzania (Mauya et al., 2019). Ten LULC types were used including water, forest, open woodland, shrubland, grassland, agricultural land, wetland, built land, bare land and cloud cover. Description of these LULCs is presented in Table 1. Since the study was conducted in a semi-arid region, the type of grassland referred to here is savanna grassland.

After several training iterations, accuracy assessment was conducted with more than 500 stratified random points. Accuracy assessment for the classified map of 1985 in particular was performed using

toposheet maps, while classified map of 2005 used Google Earth Engine platform per se, and classified map of 2021 used a combination of Google Earth Engine Platform and ground-truthing points. The existing toposheet maps for the study area were prepared from year 1958 to 1984 with fewer details. Being able to use toposheets, classes like open woodland, shrubland, grassland and bare land were merged and classified as shrubland for the purpose of accuracy testing only. Additionally, most of the toposheet maps used did not have built land class; therefore, the class was not evaluated for its accuracy. The sample sizes for the stratified random points per class were

Table 1 LULC types classification in BMC

Class code	LULC class	Primary category *	Description
C_1	Water	Inland water	**Areas with predominant water throughout the year; with little to no sparse vegetation, no rock outcrop and no built-up features
C_2	Forest	Forest	**Any significant clustering of tall (~ 15 m or higher) dense vegetation, typically with a closed or dense canopy, i.e. > 40%
C_3	Open woodland	Forest	*Clustering of tall with a canopy coverage between 10 and 40%
C_4	Shrubland	Forest	**Mix of small clusters of vegetation dispersed on a landscape that shows exposed soil or rock with emergent trees
C_5	Grassland	Non-forest	**Open areas covered in homogeneous grasses with little to no taller vegetation with no obvious human plotting
C_6	Agricultural land	Non-forest	**Human-planted cereals, grasses and crops not at tree height
C_7	Wetland	Swamp	*Area often partially or intermittently covered with water
C_8	Built land	Non-forest	**Human-made structures; major roads, rail networks and large impervious surfaces such as parking plots, office and residential buildings
C_9	Bare land	Non-forest	**Areas of rock or soil with very sparse to no vegetation for the entire year
C_10	Cloud cover	Clouds	Cloud cover and cloud shades

*Mauya et al., (2019) and ** ESRI website: <https://www.esri.com/en-us/arcgis-marketplace/listing/products/43c1d2612d134cc19b7f95deef5469db>

determined by the proportionality of each class on the map, with the most dominating class receiving a larger sample size. Classified LULC maps were then

validated using kappa coefficient in Eqs. 1 and 2 to determine whether there is a statistical significance similarity between classified and observed LULC.

$$\text{Kappa coefficient} = \frac{\text{Overall Agreement} - \text{Probability of Chance Agreement}}{1 - \text{Probability of Chance Agreement}} \tag{1}$$

$$\text{Kappa coefficient} = \frac{\sum_{i=1}^k n_i - \sum_{i=1}^k n_i(C_u C_p)}{n^2 - \sum_{i=1}^k n_i(C_u C_p)} \tag{2}$$

where i is the class number, n is the total number of classified pixels that are being compared to actual class, n_i is the number of pixels belonging to the actual class i that were classified with a class i , C_u is the total number of classified user pixels belonging to class i and C_p is the total number of producer pixels belonging to class i . Kappa coefficient indicates strength of agreement: i.e. < 0.01 considered as no agreement, 0.01–0.20 as slight agreement, 0.20–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement and 0.81–1.00 as almost perfect agreement (Cohen, 1960). Therefore, in this study, LULC maps with kappa coefficient above 0.61 were considered; otherwise, classification was repeated. Post-classification, effect of salt and pepper was reduced by applying majority filter of 3×3 to produce clean classified image.

Selection of driver variables

Geospatial, socio-economic and climatic driver variables were identified and tested for their predictive ability. The geospatial variables included were altitude (Elevation) and slope (Slope); socio-economic variables were distance from previous agricultural activities (Dist_Pre_Farming), distance from previous built area (Dist_Pre_Builtland), distance from roads (Dist_Roads), distance from streams (Dist_Stream) and distance from urban centres (Dist_Urban); and climatic variables were annual rainfall, average annual minimum temperature (Tmin) and maximum temperature (Tmax). The last included variable was evidence likelihood of land cover (EC_Likelihood_LC) which is the relative frequency of different land use categories in the area of a given transition period (Eastman, 2020a). Protected areas in BMC, like forests and game reserves, were identified, and constraints were imposed to help the model to maintain

the current reserved LULC. Detailed description of the driver variables and constraints is listed in Table 2.

Drivers’ variable explanatory power were tested using Cramer’s V and their variable’s worth with *p* value. High Cramer’s V indicate high explanatory power despite that it does not guarantee the strong performance (Eastman, 2020a). A cut-off point ≥ 0.15 was considered for this study; the rest of the variables below cut-off were discarded. Higher *p* values indicate low variable’s worth suggesting its rejection. Therefore, the driver variables with *p* values < 0.05 were selected. Qualified variables were used to train the multilayer perceptron neural network (MLP-NN) model.

Model training and transition potential

Transition sub-models were considered for training based on LULC classes that experienced significant changes from year 1985 to year 2005. Transitioned and persisted pixels from the sub-models were trained with the selected driver variables using MLP-NN. Training parameters including start and end learning rates, momentum factor, sigmoid constant and hidden layer nodes were altered until accuracy rate of $\geq 75\%$ and root mean square (RMS) error of < 0.5 was attained (Eastman, 2020b; Moriasi et al., 2007). However, during training, stopping criteria were not confined to the aforementioned criteria to give the model a chance to be performed

better. Therefore, the stopping criteria accuracy rate was set as high as 100%, and RMS was set as low as 0.01. Once a desirable performance was achieved, transition potential maps for each sub-model were then generated.

Validation and future prediction

CA-Markov chain model validation was performed by comparing simulated and classified maps for the year 2021. During simulation of LULC map of 2021, Markov chain analysis uses transition potential maps to simulate the probability of a pixel transitioning (i.e. transition probability matrix) from one land use class to another in different time periods. Transition probability matrix was manually edited until the desired simulated map was obtained. The produced validation map indicates pixels that model managed to capture transition as hits (revealed by green colour), pixels that model predicted transition but persisted as misses (revealed by red colour) and pixels model predicted persistence but transitioned as false alarms (revealed by yellow colour). Once green colour dominated the validation map, accuracy assessment was performed using kappa coefficient (Eqs. 1 and 2) to determine whether there is a statistically significant similarity between simulated and classified LULC. The verified model with kappa coefficient ≥ 0.61 was then applied to predict the LULC for BMC in immediate-future (2035), mid-future (2055) and far future (2085).

Table 2 Description of driver variables data used in prediction of LULC

	Data	Details	Description	Source
Geospatial	STRM-DEM	30-m resolution	Terrain elevation	USGS website
	Slope	Percentage rise	Terrain slope	Generated DEM
Socio-economic	Dist-Pre_Farming	Previous agricultural activities	Euclidean-Distance	Classified map—1985
	Dist-Pre_Builtland	Previous built land		
	Dist-Roads	Main, secondary, primary and tracks	IDBWB	
	Dist-Stream	Ephemeral streams		
	Dist-Urban	Centres		
Climatic	Rainfall, Tmax and Tmin	1985–2005	Observed data	TMA
			GCMs	CMIP6 websites
Other Components	Constraints	Forests and game reserves	Protected areas	IDBWB

CMIP6 Coupled Model Intercomparison Project Phase 6; IDBWB internal drainage basin water board; TMA Tanzania Meteorological Agency; USGS United States Geological Survey

Results

Accuracy of classified LULC maps

The results of the classified LULC maps for the years 1985, 2005 and 2021 are presented in Fig. 3, whereas their corresponding accuracies are presented in confusion matrices in Tables 3, 4, and 5. The obtained accuracy for the 1985 classified map was considered to have substantial agreement with the toposheet maps. Assessment point distribution method used was stratified random, where classes with larger coverage received more points and smaller coverage received fewer points. As shown in Table 3, the assessment points located

in C_2, for instance, only accounted for 9 points (i.e. < 1.8% of the total assessment points), while those in C_3 accounted for 414 points (i.e. > 81.6% of the total assessment points). Misclassification of four assessment points in C_2, the user accuracy became 56%, whereas misclassification of nine assessment points in C_3, the user accuracy is as high as 98%. This implies that accuracy assessment of maps with substantial differences in coverage may result to higher sensitivity in the accuracy of classes with fewer assessment points compared to classes with more points when misclassified. Conversely, the producer accuracy of the same classes with lower user accuracy demonstrated values within the same map contributing to an overall the

Fig. 3 Classified LULC maps for years 1985, 2005 and 2021

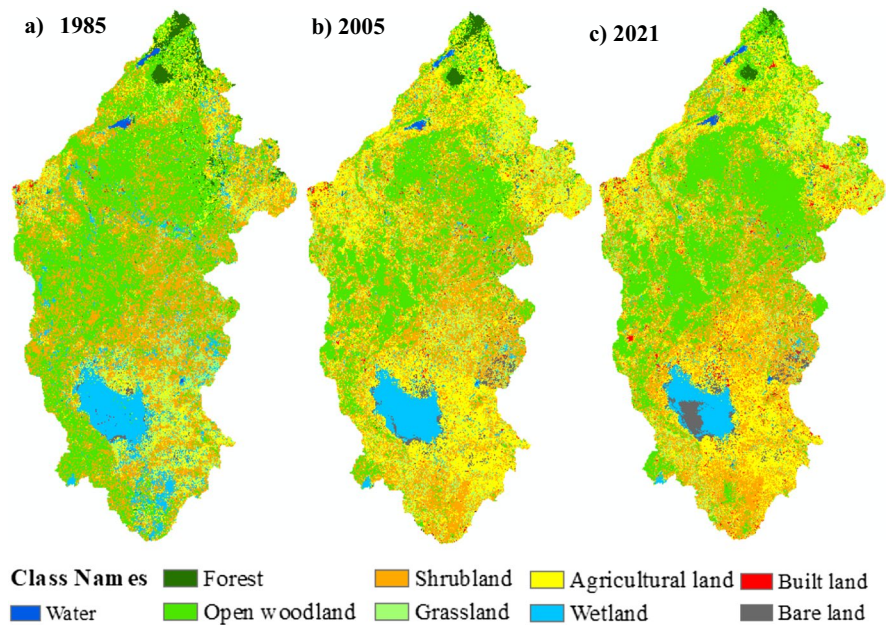


Table 3 Confusion matrix for the year 1985

Class Code	C_1	C_2	C_3	C_6	C_7	Total	User accuracy	Kappa
C_1	7	0	2	0	0	9	0.78	0
C_2	0	5	4	0	0	9	0.56	0
C_3	1	1	405	3	4	414	0.98	0
C_6	0	0	15	36	0	51	0.71	0
C_7	0	0	8	0	16	24	0.67	0
Total	8	6	434	39	20	507	0.00	0
Producer Accuracy	0.88	0.83	0.93	0.92	0.80	0.00	0.93	0
Kappa	0	0	0	0	0	0	0	0.74

Table 4 Confusion matrix for the year 2005

Class code	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	Total	User accuracy	Kappa
C_1	9	0	0	0	0	0	0	0	0	9	1.00	0
C_2	0	9	2	0	0	0	0	0	0	11	0.82	0
C_3	0	1	106	8	0	3	0	0	0	118	0.90	0
C_4	0	0	6	140	3	6	0	0	0	155	0.90	0
C_5	0	0	2	6	41	3	0	0	0	52	0.79	0
C_6	0	0	5	10	3	111	2	0	0	131	0.85	0
C_7	1	0	0	0	0	2	20	0	0	23	0.87	0
C_8	0	0	0	0	0	1	1	9	0	11	0.82	0
C_9	0	0	0	0	1	0	1	1	9	12	0.75	0
Total	10	10	121	164	48	126	24	10	9	522	0.00	0
Producer Accuracy	0.90	0.90	0.88	0.85	0.85	0.88	0.83	0.90	1.00	0	0.87	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0.8332

Table 5 Confusion matrix for the year 2021

Class code	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	Total	User accuracy	Kappa
C_1	9	1	0	0	0	0	0	0	0	10	0.90	0
C_2	0	9	0	0	0	1	0	0	0	10	0.90	0
C_3	0	1	95	4	0	18	0	0	0	118	0.81	0
C_4	0	0	6	147	5	0	0	0	0	158	0.93	0
C_5	0	0	0	3	22	1	0	0	0	26	0.85	0
C_6	0	0	3	8	2	141	1	2	1	158	0.89	0
C_7	0	0	3	0	0	0	15	0	1	19	0.79	0
C_8	0	0	0	0	0	0	0	13	0	13	1.00	0
C_9	0	0	0	1	0	0	1	0	9	11	0.82	0
Total	9	11	107	163	29	161	17	15	11	523	0.00	0
Producer Accuracy	1.00	0.82	0.89	0.90	0.76	0.88	0.88	0.87	0.82	0	0.88	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0.8417

kappa coefficient of 74% (Table 3). According to Cohen (1960), the obtained coefficient signifies substantial agreement between the produced map and observed LULC.

Similarly, for the classified maps of the year 2005 and 2021, both their user and producer accuracies were high leading to their kappa coefficients to reach 83% and 84%, respectively (Tables 4 and 5). This indicates almost perfect agreement with the actual LULC of the area (Cohen, 1960). This concludes that all three classified maps met the expectations of the map producers.

Analysis of 1985, 2005 and 2021 LULC maps

Findings of LULC classification from the year 1985 to 2021 are presented on Table 6. In 1985, open woodland was the most dominant LULC covering from the northern part to west and southern parts of the catchment whereas shrubland was the second dominant available nearly throughout the catchment with high concentration on the eastern part (Fig. 3a). Considering BMC is a semi-arid region, the weather favours open woodland and shrubland covers making them dominant. The two classes covered > 73% of

Table 6 Coverage of land covers in 1985, 2005 and 2021 for BMC

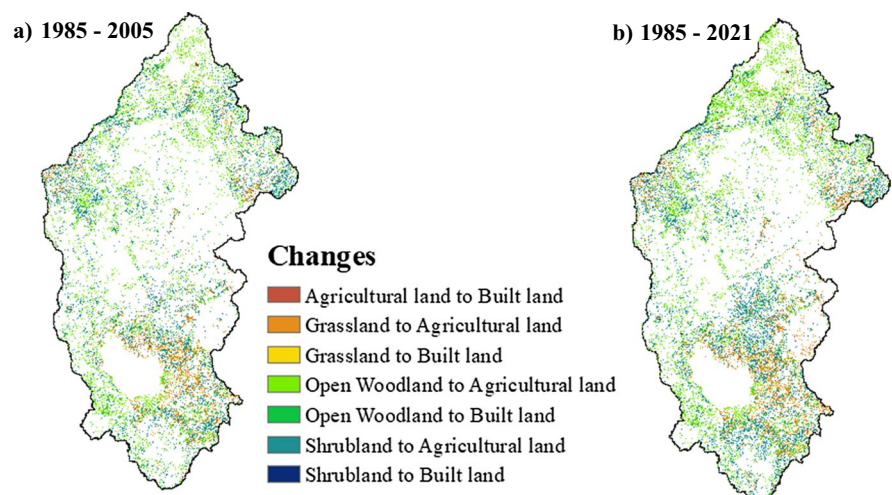
Class code	Class name	Area in km ²			Area in %		
		1985	2005	2021	1985	2005	2021
C_1	Water	164.4	60.5	67.4	0.66	0.24	0.27
C_2	Forest	493.4	325.9	157.6	1.99	1.32	0.64
C_3	Open Woodland	10,051.3	6234.0	6387.9	40.58	25.17	25.79
C_4	Shrubland	8205.5	7652.1	7420.0	33.13	30.89	29.96
C_5	Grassland	1766.4	2398.6	872.7	7.13	9.68	3.52
C_6	Agricultural land	2319.8	6368.6	7898.7	9.37	25.71	31.89
C_7	Wetland	1619.0	1225.5	1104.6	6.54	4.95	4.46
C_8	Built land	36.4	149.8	365.8	0.15	0.60	1.48
C_9	Bare land	105.3	355.3	495.5	0.42	1.43	2.00
C_10	Cloud cover	8.7	0.0	0.0	0.03	0.00	0.00

the LULC in 1985. Grassland on the other hand with a coverage of only around 7% was predominantly located on relatively lower elevations where runoff and sediments from the steep altitudes are deposited (Fig. 1). Agricultural land was the third dominant land cover with an area about 9.37%, found mostly on the northern part and southeast of the catchment.

Figure 4 presents all LULC transitions caused by anthropogenic activities from 1985 to 2005 and from 1985 to 2021. The anthropogenic disturbances transformed natural vegetative covers including open woodland, shrubland and grassland to agricultural lands and built lands. By the year 2021, agricultural land has gained a total of 22.52% becoming the most dominant LULC followed by shrubland and open woodland whereas built land only gained

about 1.33%. The expansion of agricultural land was detected throughout the catchment except in water bodies and some parts at the north, east and west of the catchment (Fig. 3c).

During the period from 1985 to 2021, significant transitions were observed from open woodland, shrubland and grassland to agricultural land, in contrast with the gradual transitions to built land shown in Fig. 5. Figure 6 presents the losses, gains and persistence of these highly transitioned LULC classes. After the identification of the classes that experienced significant changes as a result of anthropogenic disturbances, the transitions were therefore grouped under a single sub-model. The sub-model along with suitable drivers were then used to train the MLP-NN model.

Fig. 4 LULC transitions caused by anthropogenic activities **a** from 1985 to 2005 and **b** from 1985 to 2021

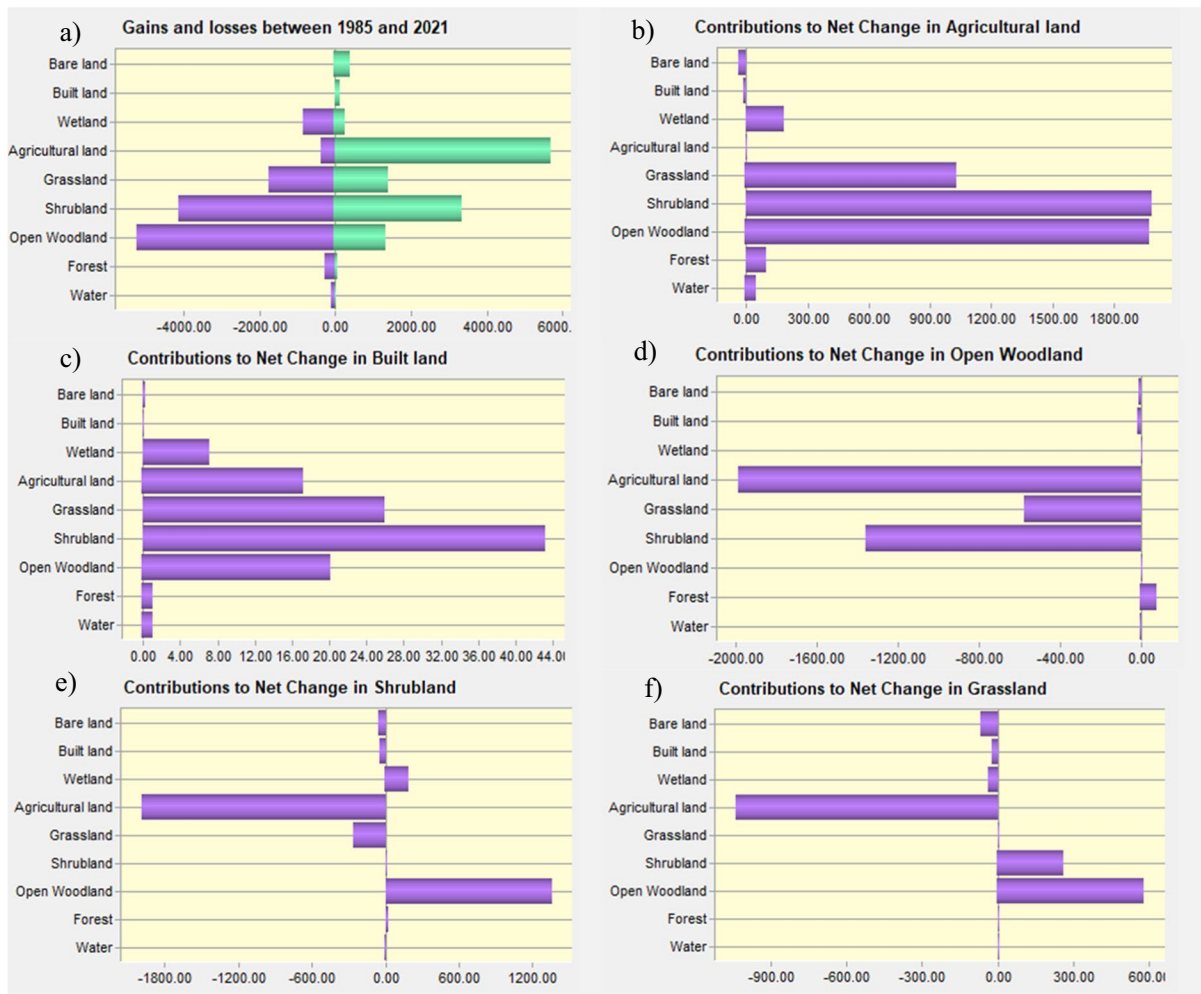


Fig. 5 Overall gain and losses from 1985 to 2021 in km² **a** in all LULC classes, **b** in agricultural land, **c** in built land, **d** in open woodland, **e** in shrubland, and **f** in grassland

Suitable drivers and model training for future prediction

The MPL-NN model was trained using driver variables that has adequate explanatory power. Table 7 presents explanatory power of all tested drivers. Based on the set criteria for the selection of suitable drivers on the “Selection of driver variables” section, only six driver variables had adequate explanatory power, which were Elevation, Rainfall, Dist-Pre_Builtland, Dist-Pre_Farming, Dist_Roads and EC_Likelihood_LC.

Figure 7a, b, and c presents the influence of the driver variables in the MLP-NN model performance.

Figure 7a indicates model performance when one variable was kept constant. Variable 6 (EC_Likelihood_LC) had the most influence on the model performance, followed by variables 4 and 5 (distance from previous anthropogenic disturbances, i.e. farming and built land), whereas variable 2 (Rainfall) had the least influence. This implies that variables 4, 5 and 6 have higher ability to anticipate the transition and/or persistence of most modelled classes far more than the remaining variables. Figure 7b demonstrates variation of model performance when all variables are constant except one. It implies that, independently, each variable performance is quite low even for variable 6 despite of its displayed influence on Fig. 7a. Figure 7c

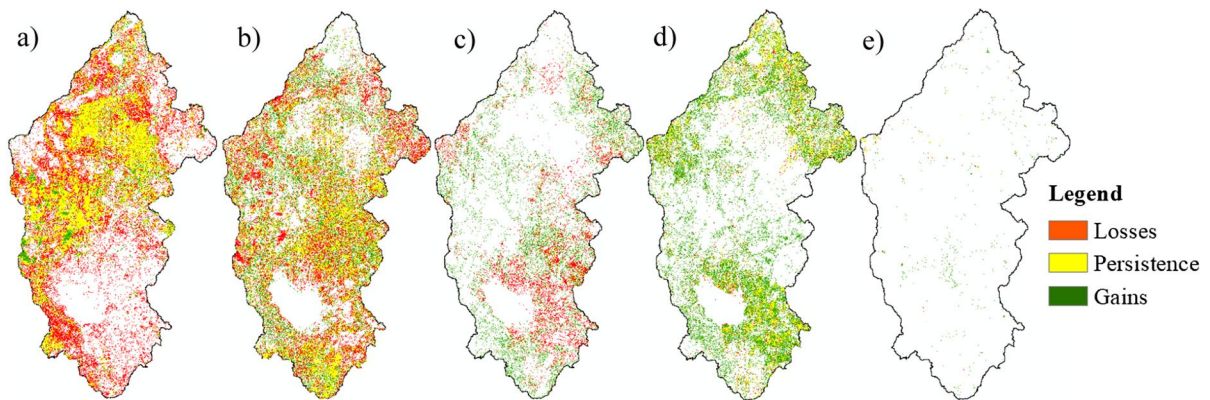


Fig. 6 Losses, persistence and gains from 1985 to 2021 in **a** open woodland, **b** shrubland, **c** grassland, **d** agricultural land, and **e** built land

Table 7 Results on the explanatory power of the tested drivers

Driver variables	Cramer's value	<i>P</i> value	Suitability	Input on MPL-NN
Elevation	0.2772	<0.001	Suitable	Variable 1
Rainfall	0.1876	<0.001	Suitable	Variable 2
Dist_Roads	0.1936	<0.001	Suitable	Variable 3
Dist-Pre_Farming	0.2400	<0.001	Suitable	Variable 4
Dist-Pre_Builtland	0.2301	<0.001	Suitable	Variable 5
EC_Likelihood_LC	0.3801	<0.001	Suitable	Variable 6
Dist_Urban	0.1244	<0.001	Not suitable	Not used
Slope	0.1224	<0.001	Not suitable	Not used
Tmin	0.1222	<0.001	Not suitable	Not used
Tmax	0.1047	<0.001	Not suitable	Not used

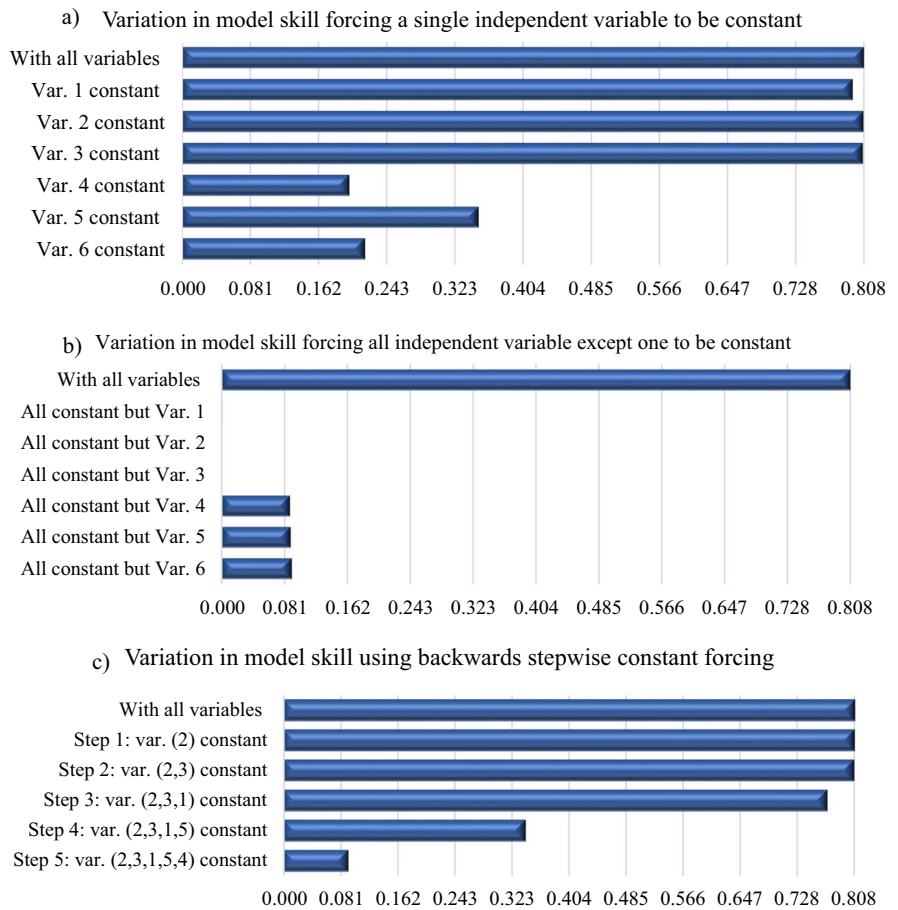
is backward stepwise forcing, where one by one variable is held constant until only one variable is left. The figure demonstrated the impact on model performance becomes significant when variable 5 and then variable 4 were held constant along with other least influencing variables. This implies that the overall analysis revealed that these variables depend on one another to achieve optimum model performance.

The parameters and performance of the trained model are presented in Table 8. The performance in Table 8 indicates the successful model training since the RMS (root mean square) error is <0.50 and accuracy rate >75% (Eastman, 2020b; Moriasi et al., 2007).

After successful model training, results of the transition and persistence of the selected LULC are presented in Table 9. It was observed from the

Table 9 that transitions performed better than persistence. This is because of the driver variables 4 to 6 were based on land use categories of the area of 1985 inclusive of previous anthropogenic activities which was farming and built land. Persistence of grassland was very high compared to the other LULC types, likely due to its smaller coverage making its training more efficient. However, the persistence of shrubland and open woodland on the other hand was the lowest. This suggests that variables indicated a higher likelihood of some pixels from open woodland and shrubland transitioning to either agricultural land and built land, while they persisted in the 2005 LULC map. The model was aided by the adjusting transition probability grid matrix for correction. Otherwise, the model is considered to be generally successful.

Fig. 7 Sensitivity of driver variables on MLP-NN Model



Prediction of future LULC maps

Prior to prediction, the performed accuracy assessment between the generated and classified 2021

LULC map produced a kappa coefficient of >0.89 (Table 10). Moreover, their error map with nearly 74% hits (captured transitions), 23% misses (predicted transition but persisted), 3% false alarms (predicted persistence but transitioned) and persistence (areas that persisted from 1985 to 2021). Based on Table 10 and Fig. 8, these results demonstrate nearly

Table 8 Parameters and performance

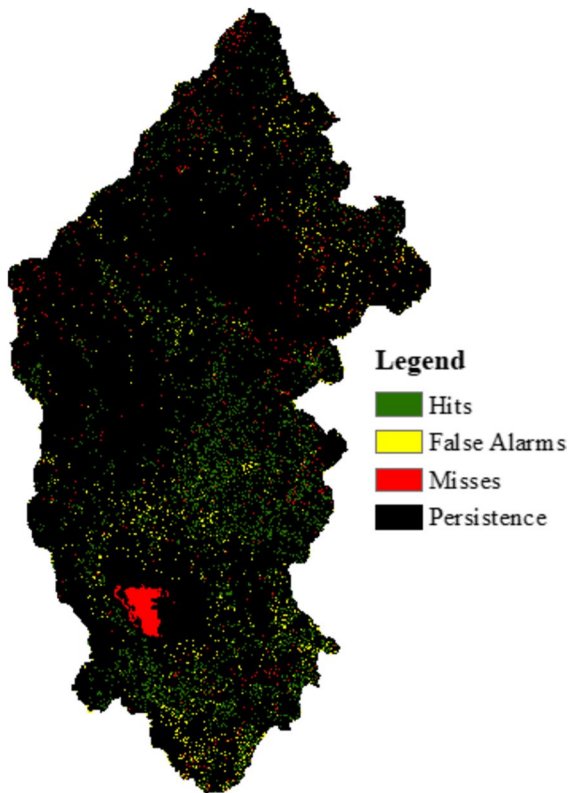
Parameters	
Input layer neurons	6
Hidden layer neurons	9
Output layer neurons	11
Requested samples per class	10,000
Final leaning rate	0.0000
Momentum factor	0.5
Sigmoid constant	1
Iterations	10,000
Performance	
Training RMS	0.1703
Testing RMS	0.1707
Accuracy rate	82.58%
Skill measure	0.8084

Table 9 Model skill breakdown by transition and persistence

Class	Skill measure
Transition: open woodland to agricultural land	0.8823
Transition: open woodland to built land	0.8505
Transition: shrubland to agricultural land	0.8811
Transition: shrubland to built land	0.7549
Transition: grassland to agricultural land	0.7628
Transition: grassland to built land	0.8131
Persistence: open woodland	0.6513
Persistence: shrubland	0.6638
Persistence: grassland	0.8627
Persistence: agricultural land	0.9303

Table 10 Confusion matrix between classified and generated LULC map of year 2021

Class code	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	Total	User accuracy	Kappa
C_1	6	2	0	0	0	0	1	0	0	9	0.67	0
C_2	0	8	2	0	0	0	0	0	0	10	0.80	0
C_3	0	0	122	2	2	0	0	1	0	127	0.96	0
C_4	0	0	6	143	1	2	0	0	0	152	0.94	0
C_5	0	0	0	0	21	5	0	0	0	26	0.81	0
C_6	0	2	2	1	0	148	0	0	1	154	0.96	0
C_7	1	1	0	0	0	1	20	0	0	23	0.87	0
C_8	0	0	0	0	1	2	0	7	0	10	0.70	0
C_9	0	0	0	0	0	0	5	0	5	10	0.50	0
Total	7	13	132	146	25	158	26	8	6	521	0.00	0
Producer accuracy	0.86	0.62	0.92	0.98	0.84	0.94	0.77	0.88	0.83	0.00	0.92	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0.8966

**Fig. 8** Error map between classified and generated LULC map of year 2021

perfect agreement between the generated and classified LULC maps of 2021 (Cohen, 1960), indicating the CA-Markov chain model was ready to generate future LULC maps for the year 2035, 2055 and 2085.

Coverage of predicted future LULC maps by CA-Markov chain for the years 2035, 2055 and 2085 are presented in Table 11. It is evident that the main change was predicted to be from open woodland, shrubland and grassland to agricultural land and built land. The results demonstrated agricultural land and built land are expected to increase from 2021 to 2085 agricultural land becoming the most dominant LULC whereas built land becoming the fourth dominant LULC after shrubland and open woodland (Tables 6 and 11).

Discussion

This study has classified and analysed spatio-temporal changes in LULC maps for the years 1985, 2005 and 2021 and subsequently employed a cellular automata Markov chain model to project LULC maps for the years 2035 (immediate-future), 2055 (mid-future) and 2085 (far-future) within the Bahi (Man-yoni) Catchment (BMC), Tanzania. The accuracy results of the classified LULC maps of the year 2005 and 2021 (Tables 4 and 5) indicated almost perfect agreement with the observed LULC in the study area (Cohen, 1960). However, the accuracy of the classified LULC map of year 1985 ranked slightly lower than the remaining LULC maps (Table 3). This could be attributed by lack of available reference ground-truth data, paper quality of a toposheet map which brought difficulties during georeferencing as well as use of outdated toposheet maps (Daba & You, 2022;

Table 11 Coverage of the predicted future land covers in 2035, 2055 and 2085 in BMC

Class code	Class name	Area in km ²			Area in %		
		2035	2055	2085	2035	2055	2085
C_1	Water	60.3	60.1	59.6	0.24	0.24	0.24
C_2	Forest	324.2	324.1	317.1	1.31	1.31	1.28
C_3	Open woodland	4606.1	3893.8	3335.4	18.60	15.72	13.47
C_4	Shrubland	5288.7	4308.1	3730.8	21.35	17.39	15.06
C_5	Grassland	1233.6	1172.8	1082.0	4.98	4.73	4.37
C_6	Agricultural land	11,046.6	12,044.8	12,424.1	44.60	48.63	50.16
C_7	Wetland	1229.2	1227.7	1226.2	4.96	4.96	4.95
C_8	Built land	633.2	1390.5	2257.0	2.56	5.61	9.11
C_9	Bare land	348.2	348.3	338.1	1.41	1.41	1.37

Nath et al., 2023; Piškinaite and Veteikis 2023). It is worth noting that some of these toposheets date back to 1958, which is more than 20 years before the classification year of 1985. Based on Cohen (1960) criteria of agreements concludes that all three produced LULC maps represent the spatio-temporal situation of BMC and that the classified maps could be used for further analysis and predictions.

The findings of the classified LULC maps from the years 1985 to 2021 revealed that in 1985, BMC was predominantly covered by natural vegetation, which subsequently transitioned to agricultural lands across most of the catchment (Figs. 3 and 4).

The transition was primarily driven by population growth and low crop yield due to water scarcity and/or poor soil quality which brings strong incentive to expand agriculture to increase production and meet food demand (Dolan et al., 2021; Rosa et al., 2020). Furthermore, clearing of natural vegetation could also be attributed by logging for timber and rotational grazing for livestock keeping. It is a common practice in Tanzania for livestock keepers to burn shrubland to stimulate new out-sprout of vegetations for grazing their cattle herds (Tanzania Forest Services 2013). Studies conducted by Valone et al. (2002) and Mora et al. (2022) on the effects of fire and grazing on an arid regions found out fire has a temporary effect on the vegetation species that could be reversed. However, continuous large and high-severity burning could cause permanent transition from shrubland to grassland (Orr et al., 2022). Since this was beyond the scope of this study to analyse the tolerance of vegetation species subjected to fire, this calls for further studies. The segments of the catchment that remained unaltered were mostly protected areas. For instance,

in the northern part of the catchment has Hanang forest reserve and *Swagaswaga* game reserve as well as on the east there is *Mbuga ya Goima* reserve and southeast is *Kigongkwe/Chigongkwe* reserve. There are also other forest reserves at the edges of the catchment, which extends beyond the northern part including *Nou*, *Salanga* and *Isabe* and on the east there is *Chenene* west forest reserves. These protected areas are under either the state management (Tanzania Forest Services Agency) or Joint Forest Management (JFM) constituting of miombo woodland, Itigithickets, savanna-grassland and shrublands except for *Mbuga ya Goima* and *Kigongkwe/Chigongkwe*, which constitutes of mostly savanna-grassland and shrubland (USAID, 2019).

Interestingly, the western part of the catchment (Makuru and Saranda wards in Manyoni district and Sanzawa ward in Chemba district) and northwest (Mughunga ward in Singida rural district) were not subjected to anthropogenic activities despite not being under protection. These areas are distant from the town centres such as Manyoni and Singida urban centres, respectively. This could be further explained by Von Thunen’s model of agricultural land use established in 1826, which relates the use of suitable agricultural lands close to the commercial centres being triggered by the lower transportation costs (Peet et al., 1967). Hence, the more distant the agricultural areas are from the centres, the less likely for them to be utilized. As for built land, it is sparsely located throughout the catchment (Figs. 4 and 6e). Table 6 indicates continuous growth of built land from 1985 to 2021, with relatively higher growth from 2005 to 2021 compared to 1985 to 2005. The higher growth from 2005 to 2021 was associated with population

growth and/or active relocation of government offices to capital city from 2015 (Mnyampanda & Chindeng-wike, 2021), which attracted movement from various parts of the country seeking opportunities. However, the overall expansion of built land from 1985 to 2021 is still small, as indicated in Table 6. This could be attributed to some of the town centres being located a few kilometres outside the catchment area, such as Dodoma city centre and Singida town centre.

The CA-Markov chain model successfully generated a LULC map for the year 2021 with no significant differences compared to the classified LULC map of 2021 (Table 10). However, despite of CA-Markov model success, the model failed to mimic the state of Bahi swamp of the year 2021 (Fig. 8). The dried-out part of the swamp, which exposed the bottom soils, was detected and classified as bare land on the classified LULC map of year 2021. However, since the CA-Markov chain model was trained using previously classified LULC maps from 1985 and 2005, which had full coverage of the Bahi swamp, it was unable to detect the dried part of the swamp in 2021, resulting in errors. This means that the CA-Markov model predicted the persistence of water in the swamp instead of bare land. This is one of the limitations of using the CA-Markov chain model, which could not be resolved in this study, even with the inclusion of climatic drivers such as rainfall. Aguejidad (2021) and Gemitzi, (2021) were also concerned with ability of CA-Markov chain model performance accuracy when subjected to abrupt changes as a result from natural disasters, climate variabilities and/or geopolitical uncertainties. This is a cross-cutting drawback for most LULC change prediction models. Major limitation of LULC prediction models is not being able to predict non-linear transformations, which is usually the case for areas undergoing rapid changes (Hao et al., 2018; Aburas et al., 2019; Aguejidad, 2021; Antwi-Agyakwa et al., 2023).

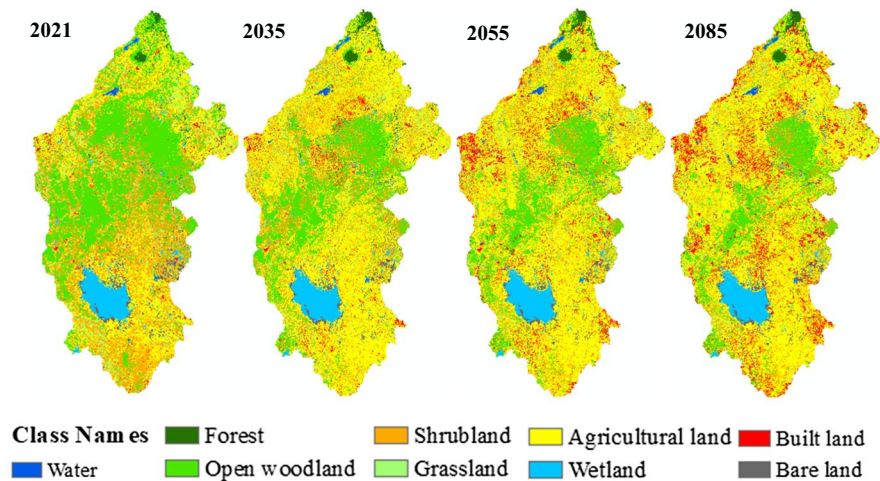
Normally, the drying out of water bodies is associated with reduced rainfall amounts and higher temperatures in the surrounding areas (Mir et al., 2021). However, it was not the case for Bahi swamp. Based on the yearly statement on the status of Tanzania climate of 2021, it was recorded to have an increased rainfall amounts of 276.6 mm and with the temperature increase between 1 and 2 °C compared to a long-term average (TMA, 2022). The Clausius-Clapeyron relation, a 1 °C increase in temperature can increase

the rate of evaporation approximately about 7% (Kim et al., 2022). However, the actual evaporation depends on multiple factors including initial temperature, relative humidity, wind speed, availability of water and surface area of water bodies, and plant cover can affect the total amount of water available for evaporation (Penman, 1956). Since this study did not establish the rate of evaporation, therefore the influence of future changes in rainfall and temperature on water bodies within the catchment remains unknown.

The findings highlight a significant shift in the LULC composition towards increased agricultural and built land at the expense of natural vegetation (Tables 6 and 11). It seems that the historical pattern of growth of agricultural land and built land at the expense of natural vegetative land covers since 1985 will prevail. The pattern's perseverance from historical to future predicted LULC was based on model training that relied on sub-model used, which dealt with only major changes caused by anthropogenic disturbances (see the "Analysis of 1985, 2005 and 2021 LULC maps" section). It is acknowledged that there are other LULC caused by human activities during the period between 1985 and 2005. However, because the sub-model used in the study accounted for over 80.5% of all anthropogenic changes during that time, it is reasonable to assume that any remaining changes would have minimal impact on future predictions.

It was also noted that despite a predicted substantial increase of built land from 2021 to 2035, the growth of agricultural land gradually slowed from 2035 to 2085. Several studies predicted similar results that the growth of agricultural land will either slow down or decrease in the future (Belay & Mengistu, 2021; Lu et al., 2019; Mahdian et al., 2023; Zhu et al., 2022). This slowing growth is linked to the limited availability of land suitable for agricultural expansion. The majority of the remaining unconverted lands with potential for agriculture are protected areas, especially on the north and northeast (Fig. 9). If the forest management becomes ineffective, there is a high possibility of encroachment as some farms are already positioned quite close to the borders of these protected areas since 2021. With the fast-growing population, promoting a revolution from extensive to intensive farming practices is necessary to minimize the conversion of native lands to agricultural lands. This will increase productivity in small areas

Fig. 9 Generated maps by CA-Markov chain for the year 2021, 2035, 2055 and 2085



and may lower the urge to encroach on the protected areas when management practices become ineffective. Predictions indicate that by 2085 that some parts of Makuru and Saranda wards in Manyoni district, as well as Sanzawa ward in Chemba district, are the only remaining areas of the catchment that have remained as non-agricultural land. This could be because of the farming extended from the northern and southeast towards the west of the catchment. Since the remaining area is not under protection, it is anticipated that it is at risk of being converted to agriculture in the future. The findings obtained here are consistent with prior researches conducted in this and other water basin in Tanzania, which have identified farming as the fastest-growing LULC in the country (Berakhi et al., 2015; Msofe et al., 2019; Mussa et al., 2021; Näschen et al., 2019; Twisa & Buchroithner, 2019). Prolonged farming can lead to a reduction in infiltration and an increase in runoff due to agricultural activities, such as ploughing, tilling and harvesting, that disturb the soil structure and increase soil compaction (Nunes et al., 2020; Voorhees et al., 1978). Consequently, the soil’s ability to absorb water decreases, and more water runs off the surface, causing erosion. This raises concerns about the alteration of hydrological processes which may affect groundwater recharge. Moreover, prolonged farming accompanied with excessive use of agrochemicals such as fertilizers and insecticides can lead to decreased soil fertility and eventually contributing to food insecurity (Popp et al., 2013). Therefore, with the expected growing demand for agricultural land, this study provides necessary information for the basin board

regarding the areas that are more susceptible to changes for proper LULC planning.

Conclusion

This study aimed to predict future land use and land cover change in Bahi (Manyoni) catchment, Tanzania for the immediate-future (2035), mid-future (2055) and far-future (2085) using integrated MLP-NN and CA-Markov chain model. The classified LULC maps of the years 1985, 2005 and 2021 were used to train the model. The accuracy assessment revealed remarkable agreement with observed conditions, reaffirming reliability of the classified LULC maps for further analysis and predictive applications. It should be noted that the land use for 2021 included abrupt changes associated drying of Bahi swamp. Despite limitations encountered by the integrated MLP-NN-CA-Markov chain when predicting abrupt changes, such as the drying out of the Bahi swamp, nevertheless, the overall trained model managed to simulate almost perfectly the LULC map of 2021, thus indicating its high predictive capability. The findings of LULC in BMC has indicated significant shifts in the landscape dynamics of the catchment. The transition from natural vegetation to agricultural lands emerged as a dominant pattern from 1985 to 2021, driven by factors such as water scarcity, low crop yields, population growth and anthropogenic activities such as logging and grazing. The study demonstrated that most of these anthropogenic changes are expected to spread from the north and southeast to the west part

covering nearly the entire catchment except for the protected areas and water bodies. It was noted that the patterns are likely to continue into the future, though the expansion of agricultural lands might gradually slow due to constraints of land availability. To address the potential challenges arising from these changes, the study suggested promoting a transition from extensive to intensive farming practices. This approach could reduce conversion rate of native lands to agriculture. Furthermore, as agriculture continues to expand, concerns about altered hydrological processes including groundwater recharge arise, underlining the need for strategic LULC planning and management. This research provides a valuable contribution to understanding the complex interplay of anthropogenic activities and environmental changes within the Bahi (Manyoni) Catchment. The findings offer insights for sustainable land management strategies, aiding policymakers and stakeholders in making informed decisions to balance development with conservation efforts in this vital catchment. As Tanzania and other regions globally face similar challenges of land use transformation, this study's insights resonate beyond the Bahi (Manyoni) catchment, contributing to the broader discourse on effective land use planning and management.

Recommendations

This study offers valuable insights and directions for future research in the field of land use and land cover (LULC) dynamics. The following are some areas that may benefit from further investigation:

- Since the study revealed challenges in predicting non-linear transformations, particularly instance of abrupt changes like the drying out of the Bahi swamp, future research could investigate deeper into understanding and addressing this limitation to improve the accuracy of LULC prediction model in abrupt changing environment.
- This study did not incorporate impact of future climate variables to the projected land use. Since it is known that climate affect LULC, hence future study may consider incorporating future climate variables into the LULC prediction model.
- The study projected LULC changes up to 2085, yet uncertainties increase with longer projection periods. Hence, future research could investigate the reliability and robustness of LULC predictions over extended timeframes, addressing potential uncertainties that arise in long-term modelling.

Acknowledgements The authors would like to acknowledge The University of Dodoma for the financial support to the first author's Ph.D studies.

Author contributions NA.M.H: Conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft. DM.M.M: Validation, writing (review and editing), supervision. JN: Writing (review and editing), supervision.

Funding The research leading to these results received funding from The University of Dodoma.

Data Availability Landsat imagery data are available at USGS website with the following URL: <https://earthexplorer.usgs.gov/>

Declarations All authors have read, have understood and have complied as applicable with the statement on “Ethical responsibilities of Authors” as found in the Instructions for Authors.

Competing interests The authors declare no competing interests.

References

- Aburas, M. M., Ahamad, M. S. S., & Omar, N. Q. (2019). Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: A review. *Environmental Monitoring and Assessment*, *191*(4). <https://doi.org/10.1007/s10661-019-7330-6>
- Aburas, M. M., Ho, Y. M., Pradhan, B., Salleh, A. H., & Alazaiza, M. Y. D. (2021). Spatio-temporal simulation of future urban growth trends using an integrated CA-Markov model.
- Aguejdad, R. (2021). The influence of the calibration interval on simulating non-stationary urban growth dynamic using CA-markov model. *Remote Sensing*, *13*(3), 1–20. <https://doi.org/10.3390/rs13030468>
- Alqadhi, S., Mallick, J., Balha, A., Bindajam, A., Singh, C. K., & Hoa, P. V. (2021). Spatial and decadal prediction of land use/land cover using multi-layer perceptron-neural network (MLP-NN) algorithm for a semi-arid region of Asir. *Saudi Arabia. Earth Science Informatics*, *14*(3), 1547–1562. <https://doi.org/10.1007/s12145-021-00633-2>
- Antwi-Agyakwa, K. T., Afenyo, M. K., & Angnuureng, D. B. (2023). Know to predict, forecast to warn: A review of

- flood risk prediction tools. *Water*, 15(3), 427. <https://doi.org/10.3390/w15030427>
- Belay, T., & Mengistu, D. A. (2021). Impacts of land use/land cover and climate changes on soil erosion in Muga watershed, Upper Blue Nile basin (Abay), Ethiopia. *Ecological Processes*, 10(1). <https://doi.org/10.1186/s13717-021-00339-9>
- Berakhi, R. O., Oyana, T. J., & Adu-Prah, S. (2015). Land use and land cover change and its implications in Kagera river basin. *East Africa. African Geographical Review*, 34(3), 209–231. <https://doi.org/10.1080/19376812.2014.912140>
- Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms k-nearest-neighbor, support vector machine, random forest and neural network: A review. *Journal of Data Analysis and Information Processing*, 08(04), 341–357. <https://doi.org/10.4236/jdaip.2020.84020>
- Bogale, A. (2020). Review, impact of land use/cover change on soil erosion in the Lake Tana Basin, Upper Blue Nile. *Ethiopia. Applied Water Science*, 10(12), 1–6. <https://doi.org/10.1007/s13201-020-01325-w>
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189–215. <https://doi.org/10.1016/j.neucom.2019.10.118>
- Chanapathi, T., & Thatikonda, S. (2020). Investigating the impact of climate and land-use land cover changes on hydrological predictions over the Krishna river basin under present and future scenarios. *Science of the Total Environment*, 721, 137736. <https://doi.org/10.1016/j.scitotenv.2020.137736>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Daba, M. H., & You, S. (2022). Quantitatively assessing the future land-use / land-cover changes and their driving factors in the upper stream of the awash river based on the CA – Markov model and their implications for water resources management.
- Dolan, F., Lamontagne, J., Link, R., Hejazi, M., Reed, P., & Edmonds, J. (2021). Evaluating the economic impact of water scarcity in a changing world. *Nature Communications*, 1–10. <https://doi.org/10.1038/s41467-021-22194-0>
- Eastman, R. J. (2020a). TerrSet2020 Geospatial Monitoring and Modeling System Manual. Clark Labs.
- Eastman, R. J. (2020b). TerrSet2020 Geospatial Monitoring and Modeling System Tutorial. *Clark Labs*.
- Fenta, A. A., Tsunekawa, A., Haregeweyn, N., Tsubo, M., Yasuda, H., Shimizu, K., et al. (2020). Cropland expansion outweighs the monetary effect of declining natural vegetation on ecosystem services in sub-Saharan Africa. *Ecosystem Services*, 45(August), 101154. <https://doi.org/10.1016/j.ecoser.2020.101154>
- Gemitzi, A. (2021). Predicting land cover changes using a CA Markov model under different shared socioeconomic pathways in Greece. *Giscience and Remote Sensing*, 58(3), 425–441. <https://doi.org/10.1080/15481603.2021.1885235>
- Getachew, B., Manjunatha, B. R., & Bhat, H. G. (2021). Modeling projected impacts of climate and land use/land cover changes on hydrological responses in the Lake Tana Basin, upper Blue Nile River Basin. *Ethiopia. Journal of Hydrology*, 595(January), 125974. <https://doi.org/10.1016/j.jhydrol.2021.125974>
- Ghalehtemouri, K. J., Shamsoddini, A., Mousavi, M. N., Ros, BintiChe, & F., & Khedmatzadeh, A. (2022). Predicting spatial and decadal of land use and land cover change using integrated cellular automata Markov chain model based scenarios (2019–2049) Zarrin-Rüd River Basin in Iran. *Environmental Challenges*, 6(November 2021), 100399. <https://doi.org/10.1016/j.envc.2021.100399>
- Ghotbi, A., Shafieifar, M., & Daneshvar, S. (2015). *Integrated water resources management and development plan (IWRMDP) for internal drainage basin (IDB). Volume 1: Main Report*.
- Gidey, E., Dikinya, O., Sebego, R., Segosebe, E., & Zenebe, A. (2017). Cellular automata and Markov Chain (CA_Markov) model-based predictions of future land use and land cover scenarios (2015–2033) in Raya, northern Ethiopia. *Modeling Earth Systems and Environment*, 3(4), 1245–1262. <https://doi.org/10.1007/s40808-017-0397-6>
- Girma, R., Fürst, C., & Moges, A. (2022). Land use land cover change modeling by integrating artificial neural network with cellular Automata-Markov chain model in Gidabo river basin, main Ethiopian rift. *Environmental Challenges*, 6(November 2021), 100419. <https://doi.org/10.1016/j.envc.2021.100419>
- Hamad, R., Balzter, H., & Kolo, K. (2018). Predicting land use/land cover changes using a CA-Markov model under two different scenarios. *Sustainability (switzerland)*, 10(10), 1–23. <https://doi.org/10.3390/su10103421>
- Hansen, M. C. (2013). *High-Resolution Global Maps of, 850(November)*, 850–854. <https://doi.org/10.1126/science.1244693>
- Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, 56(1), 108–141. <https://doi.org/10.1002/2016RG000549>
- Harris, R. M. B., Grose, M. R., Lee, G., Bindoff, N. L., Porfiri, L. L., & Fox-Hughes, P. (2014). Climate projections for ecologists. *Wiley Interdisciplinary Reviews: Climate Change*, 5(5), 621–637. <https://doi.org/10.1002/wcc.291>
- Kim, S., Sharma, A., Wasko, C., & Nathan, R. (2022). Linking total precipitable water to precipitation extremes globally. *Earth's Future*, 10(2). <https://doi.org/10.1029/2021EF002473>
- Kisamba, F. C., & Li, F. (2022). Analysis and modeling urban growth of Dodoma urban district in Tanzania using an integrated CA–Markov model. *GeoJournal*, (0123456789). <https://doi.org/10.1007/s10708-022-10617-4>
- Lamchin, M., Lee, W. K., & Wang, S. W. (2022). Multi-temporal analysis of past and future land-cover changes of the third pole. *Land*, 11(12). <https://doi.org/10.3390/land11122227>
- Lu, Y., Wu, P., Ma, X., & Li, X. (2019). Detection and prediction of land use/land cover change using spatiotemporal data fusion and the Cellular Automata–Markov model. *Environmental Monitoring and Assessment*, 191(2). <https://doi.org/10.1007/s10661-019-7200-2>

- Mahdian, M., Hosseinzadeh, M., Siadatmousavi, S. M., Chalipa, Z., Delavar, M., Guo, M., et al. (2023). Modeling impacts of climate change and anthropogenic activities on inflows and sediment loads of wetlands: Case study of the Anzali wetland. *Scientific Reports*, *13*(1), 1–18. <https://doi.org/10.1038/s41598-023-32343-8>
- Marsh, G. P. (1864). *Man and nature , or physical geography as modified by human action by George P. Marsh*. Sampson Low, Son and Marston.
- Mauya, E. W., Mugasha, W. A., Njana, M. A., Zahabu, E., & Malimbwi, R. (2019). Carbon stocks for different land cover types in Mainland Tanzania. *Carbon Balance and Management*, *14*(1), 1–13. <https://doi.org/10.1186/s13021-019-0120-1>
- Mengistu, T. D., Chung, I. M., Kim, M. G., Chang, S. W., & Lee, J. E. (2022). Impacts and implications of land use land cover dynamics on groundwater recharge and surface runoff in East African Watershed. *Water (Switzerland)*, *14*(13). <https://doi.org/10.3390/w14132068>
- Mir, B. H., Kumar, R., Lone, M. A., & Tantray, F. A. (2021). Climate change and water resources of Himalayan region—review of impacts and implication. *Arabian Journal of Geosciences*, *14*(12). <https://doi.org/10.1007/s12517-021-07438-z>
- Mishra, P. K., Rai, A., & Rai, S. C. (2020). The Egyptian Journal of Remote Sensing and Space Sciences Land use and land cover change detection using geospatial techniques in the Sikkim Himalaya, India. *The Egyptian Journal of Remote Sensing and Space Sciences*, *23*(2), 133–143. <https://doi.org/10.1016/j.ejrs.2019.02.001>
- Mitchell, R. B. (2003). International environmental agreements: A survey of their features, formation, and effects. *Annual Review of Environment and Resources*, *28*, 429–461. <https://doi.org/10.1146/annurev.energy.28.050302.105603>
- Mnyampanda, A. M., & Chindengwike, J. (2021). Relationship between asset quality and financial performance of commercial banks before and after shifting capital city located to Dodoma Region. *Tanzania. SSRN Electronic Journal*, *1*, 149–157. <https://doi.org/10.2139/ssrn.3989888>
- Mora, J. L., Badía-Villas, D., & Gómez, D. (2022). Fire does not transform shrublands of *Echinospartum horridum* (Vahl) Rothm. into grasslands in the Pyrenees: Development of community structure and nutritive value after single prescribed burns. *Journal of Environmental Management*, *315*(November 2021). <https://doi.org/10.1016/j.jenvman.2022.115125>
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*, 885–900. American society of agricultural and biological engineers
- Msofe, N. K., Sheng, L., & Lyimo, J. (2019). Land use change trends and their driving forces in the Kilombero Valley Floodplain. *Southeastern Tanzania. Sustainability (switzerland)*, *11*(2), 1–25. <https://doi.org/10.3390/su11020505>
- Mubako, S., Nnko, H. J., Peter, K. H., & Msongaleli, B. (2022). Evaluating historical and predicted long-term land use/land-cover change in Dodoma Urban District, Tanzania: 1992–2029. *Physics and Chemistry of the Earth*, *128*(August). <https://doi.org/10.1016/j.pce.2022.103205>
- Mussa, K. R., Mjemah, I. C., & Machunda, R. L. (2021). Natural groundwater recharge response to climate variability and land cover change perturbations in basins with contrasting climate and geology in Tanzania. *Earth*, *2*(3), 556–585. <https://doi.org/10.3390/earth2030033>
- Nabuurs, G., Mrabet, R., Hatab, A. A., Bustamante, M., Clark, H., Havlík, P., et al. (2022). Chapter 7: Agriculture, Forestry and Other Land Uses (AFOLU). In P. R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, et al. (Eds.), *Climate Change 2022: Mitigation of Climate Change. Working Group III contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Näschen, K., Diekkrüger, B., Evers, M., Höllermann, B., Steinbach, S., & Thonfeld, F. (2019). The impact of land use/land cover change (LULCC) on water resources in a tropical catchment in tanzania under different climate change scenarios. *Sustainability (Switzerland)*, *11*(24). <https://doi.org/10.3390/su11247083>
- Nath, A., Koley, B., Choudhury, T., Saraswati, S., Ray, B. C., Um, J., & Sharma, A. (2023). Assessing coastal land-use and land-cover change dynamics using geospatial techniques, 1–23.
- Noi, P. T., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 Imagery. *Sensors (Basel, Switzerland)*, *18*(1). <https://doi.org/10.3390/s18010018>
- Nunes, M. R., Karlen, D. L., & Moorman, T. B. (2020). Sustainability tillage intensity effects on soil structure indicators — A US meta-analysis.
- Olorunfemi, I. E., Olufayo, A. A., Fasinmirin, J. T., & Komolafe, A. A. (2022). Dynamics of land use land cover and its impact on carbon stocks in Sub-Saharan Africa: An overview. *Environment, Development and Sustainability*, *24*(1), 40–76. <https://doi.org/10.1007/s10668-021-01484-z>
- Orr, D. A., Bates, J. D., & Davies, K. W. (2022). Grazing intensity effects on fire ignition risk and spread in sagebrush steppe. *Rangeland Ecology and Management*, (xxxx), 1–10. <https://doi.org/10.1016/j.rama.2022.08.004>
- Peet, J. R., Horvath, R. J., & Sinclair, R. (1967). The present pertinence of von Thunen theory. *Taylor & Francis, Ltd on behalf of the Association of American Geographers*, *57*(4), 810–815. <http://www.jstor.org/stable/2561774>. Accessed 6 Mar 2023.
- Penman, H. L. (1956). Estimating evaporation. *Eos, Transactions American Geophysical Union*, *37*(1), 43–50. <https://doi.org/10.1029/TR037i001p00043>
- Piškinaitė, E., & Veteikis, D. (2023). The results of digitizing historical maps: Comparison of Lithuanian land-use structure in the 19th and 21st Centuries.
- Popp, J., Pet, K., & Nagy, J. (2013). Pesticide productivity and food security . A review, 243–255. <https://doi.org/10.1007/s13593-012-0105-x>
- Rahaman, H., Singh, R., Masroor, M., & Sajjad, H. (2023). Integrating remote sensing derived indices and machine learning algorithms for precise extraction of small surface

- water bodies in the lower Thoubal river watershed. *India. Journal of Cleaner Production*, 422(July), 138563. <https://doi.org/10.1016/j.jclepro.2023.138563>
- Ritchie, H., & Roser, M. (2021). Forests and deforestation. *Our World In Data*. Our World In Data. <https://ourworldindata.org/forests-and-deforestation>. Accessed 27 December 2022
- Rosa, L., Chiarelli, D. D., Rulli, M. C., Angelo, J. D., & Odorico, P. D. (2020). Global agricultural economic water scarcity, (April), 1–11.
- Sang, L., Zhang, C., Yang, J., Zhu, D., & Yun, W. (2011). Simulation of land use spatial pattern of towns and villages based on CA-Markov model. *Mathematical and Computer Modelling*, 54(3–4), 938–943. <https://doi.org/10.1016/j.mcm.2010.11.019>
- Siddik, M. S., Tulip, S. S., Rahman, A., Islam, M. N., Haghghi, A. T., & Mustafa, S. M. T. (2022). The impact of land use and land cover change on groundwater recharge in north-western Bangladesh. *Journal of Environmental Management*, 315(October 2021), 115130. <https://doi.org/10.1016/j.jenvman.2022.115130>
- Tanzania Forest Services. (2013). A training manual on fire awareness and education: Training at the village level in Tanzania, 1–64.
- Tewabe, D., & Fentahun, T. (2020). Assessing land use and land cover change detection using remote sensing in the Lake Tana Basin, Northwest Ethiopia. *Cogent Environmental Science*, 6(1). <https://doi.org/10.1080/23311843.2020.1778998>
- TMA. (2022). Statement on the Status of Tanzania Climate in 2021. *Tanzania Meteorological Authority (TMA)*.
- Twisa, S., & Buchroithner, M. F. (2019). Land-use and land-cover (LULC) change detection in Wami river basin, Tanzania. *Land*, 8(9). <https://doi.org/10.3390/land8090136>
- UNCCD. (2017). *United Nations Convention to Combat Desertification. 2017. The Global Land Outlook, first edition*.
- URT. (2012). *Basic demographic and socio-economic profile*. United Republic of Tanzania (URT) Dar es Salaam, Tanzania.
- URT. (2022). *Population and Housing Census: Preliminary Results*. United Republic of Tanzania (URT) Dodoma, Tanzania.
- USAID. (2019). Biodiversity and Tropical forestry (FAA 118/119) Assessment. *United States Agency for International Development (USAID)*, (7200).
- Valone, T. J. ., Nordell, S. E. ., & Ernest, S. . K. . M. (2002). Effects of fire and grazing on an arid grassland ecosystem. *Southwestern Association of Naturalists*, 47(4), 557–565. <https://www.jstor.org/stable/3672659>
- Voorhees, W. B., Senst, C. G., & Nelson, W. W. (1978). Compaction and soil structure modification by wheel traffic in the Northern Corn Belt. *Soil Science Society of America Journal*, 42(2), 344–349. <https://doi.org/10.2136/sssaj1978.03615995004200020029x>
- Wang, S. W., Munkhnasan, L., & Lee, W. (2021). Land use and land cover change detection and prediction in Bhutan's high altitude city of Thimphu, using cellular automata and Markov chain. *Environmental Challenges*, 2(December 2020). <https://doi.org/10.1016/j.envc.2020.100017>
- Watson, R. T., Noble, I. R., Bolin, B., Ravindranath, N. H., Verardo, D. J., & Dokken, D. J. (2000). *Intergovernmental panel on climate change special report on land use, land-use change and forestry*. Cambridge University Press. <https://www.ipcc.ch/report/land-use-land-use-change-and-forestry/>. Accessed 30 Dec 2022.
- Yildiz, O., Altundağ, E., Çetin, B., Teoman Güner, Sarginci, M., & Toprak, B. (2018). Experimental arid land afforestation in Central Anatolia, Turkey. *Environmental Monitoring and Assessment*, 190(6). <https://doi.org/10.1007/s10661-018-6724-1>
- Zhu, L., Song, R., Sun, S., Li, Y., & Hu, K. (2022). Land use/land cover change and its impact on ecosystem carbon storage in coastal areas of China from 1980 to 2050. *Ecological Indicators*, 142(July). <https://doi.org/10.1016/j.ecolind.2022.109178>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.