The Egyptian Journal of Remote Sensing and Space Sciences 25 (2022) 541–550

HOSTED BY

Contents lists available at ScienceDirect

The Egyptian Journal of Remote Sensing and Space Sciences

journal homepage: www.sciencedirect.com

Research Paper

Machine learning-based monitoring and modeling for spatio-temporal urban growth of Islamabad

Adeer Khan, Mehran Sudheer *

Department of Civil Engineering, Capital University of Science and Technology, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 8 November 2021
Revised 8 February 2022
Accepted 24 March 2022
Available online 31 March 2022

Keywords:
Machine learning
Urban growth
Artificial neural network
Cellular automata
LULC prediction
Islamabad

ABSTRACT

LULC maps are important thematic maps that provide a baseline for monitoring, assessing, and planning activities. This study incorporates spatio-temporal land use/ land cover (LULC) monitoring (1991–2021) and urban growth modeling (2021–2041) of Islamabad, Pakistan to deduce the changes in various LULC classes in the past and the future by incorporating realistic influential thematic layers and Artificial Neural Network-Cellular Automata (ANN-CA) machine learning algorithms. Three decades of Landsat satellite imagery were used to classify LULC maps using a random forest algorithm with high Kappa indexes ranging from 0.93 to 0.97. Simulations for 2011 and 2021 were done for well-calibration of the model with Kappa (>0.85) and spatial similarity (>75%) using the MOLUSCE plugin in QGIS software. Future predictions were done for the years 2031 and 2041 to analyze and study the future urban growth patterns. The satellite-based LULC maps during 1991–2021 exhibited a 142.4 km^2 increase in net urban growth. This had detrimental effects on other classes: net decrease of forests by 38.4 km^2 and waterbodies by 2.9 km^2. The projected increase of urban areas in 2021–2041 will be 58.2 km^2. Visual urban sprawl assessment on LULC maps was done to highlight the type of sprawls. Overall, it was sensed that the city's urbanization has been unplanned and erratic; leading to dire consequences on the environmental and urban systems. Therefore, the study necessitates better monitoring and better planning of urbanization by enforcing policies and necessary measures.

1. Introduction

Urban sprawl is the negative growth of urbanization that leads to the uncontrolled rapid development of cities. In recent decades, across the world, urban processes have seen vast transformations, and have transformed many geographies from their previous stagnant states (Coq-Huelva and Asián-Chaves, 2019; Liu et al., 2018). America (Bueno-Suárez and Coq-Huelva, 2020) and Europe (Aurambout et al., 2018) have recently witnessed a large increase in urbanization. Due to the economic development and population growth in developing countries; Asia particularly is witnessing worrisome urbanization, where China's sudden growth has led to urban sprawl menacing its eco-environmental quality and socioeconomic sustainability (Li and Li, 2019). South Asian countries face shared difficulties; therefore, their urban sprawl problems are becoming a grave concern for their low-income growth and the unsettling population trends (Ranagalage et al., 2021; Hatab et al., 2019). Pakistan though a young nation, has seen tremendous urban growth (Rana and Bhatti, 2018; Zafar and Zaidi, 2019). Islamabad is a new city and is the only planned city of Pakistan. It has seen enormous urbanization during the post-90s. The increase of built-up areas will consequently be a poised challenge for socio-economical and environmental factors for a developing country like Pakistan (Hatab et al., 2019).

Machine learning (ML) has become an important tool for researchers to analyze their data and predict reliable and accurate outcomes (Rogan et al., 2008; Patil et al., 2017). Urban growth patterns along with their driving factors have been geospatially and statistically modeled by researchers using ML algorithms; those hybrid models are Cellular Automata (CA) (Tripathy and Kumar, 2019), CA-Markov (CA-M) (Baqa et al., 2021; Otuoze et al., 2020), CA Markov Chain (CA-MC) (Vinayak et al., 2021) and CA-logistic regression (LR) (Mustafa et al., 2018a). Amongst these models the most used and result promising one is CA (Tripathy and Kumar, 2019). The modeling algorithm is done using thematic layers that influence the change dynamics. Researchers have used elevation, hillshade, aspect, slope, land use/land cover (LULC) maps, proxi-

Peer review under responsibility of National Authority for Remote Sensing and Space Sciences.

* Corresponding author at: Capital University of Science and Technology, Islamabad Expressway, Kahuta Road, Zone-V Sihala, Islamabad, Pakistan.
E-mail address: mehransudheer8@gmail.com (M. Sudheer).

https://doi.org/10.1016/j.ejrs.2022.03.012
1110-9823/

mate maps, population, GDP, roughness and richness indexes, employment rate, zone density, the distance between regions, etc (Mustafa et al., 2018a; Abbas et al., 2021; Patil et al., 2017). Urban growth predictions have been applied across the world: America (Hu and Lo, 2007), Europe (Rienow and Goetzke, 2015), China (Zhang et al., 2011), India (Vinayak et al., 2021), and Pakistan (Baqa et al., 2021).

The purpose of this research was to identify and predict the urban growth of Islamabad and further provide key insight into the hitherto and upcoming situation of urbanization to the policy-makers and the administrative authorities. Such studies influence better urban planning and contribute to the prevention of further urban sprawls. A well-calibrated ANN-CA approach to simulate and predict future urbanization on three decades of satellite imagery's classified LULC maps was done to assess urban growth. Model validation by Kappa statistics, confusion (error) matrix, spatial similarity, and RMSE error were done to approve the accuracy of the model. To the best of the authors' knowledge, no research has been done on spatio-temporal modeling and visual urban sprawl assessment of Islamabad LULC maps to highlight the type of sprawls and their movement's pattern across the decades. By doing so, it is then expected that the city's overall growth behavior will be deduced and the effects it will have on urban systems can then be contemplated by the urban planners. It is anticipated that the study's mitigation propositions will be based upon the findings in the research.

2. Study area

Islamabad is Pakistan's capital city and is designated as Islamabad Capital Territory, it covers an area of 906.5 km^2 and is located at 33°49′ N latitude and 72°24′ E longitude coordinates (Fig. 1). Its altitude ranges between 457 and 610 m. As a planned city, it has been divided into 5 major zones. Islamabad has a humid subtropical climate due to the nearby Margalla Hills with hot and humid summers, and cool winters. It receives an average yearly rainfall of 1143 mm. According to (The Government of Pakistan, 2017), Islamabad's urban population is 1,014,825 whereas rural is 991,747; and the average annual growth rate from 1998 to 2017 was 4.91% – the highest in Pakistan. And, the current population density is 2207 persons/ km^2. The city registered the highest population growth rate of 5.2 % from 1981 to 1998, and 4.91 % from 1998 to 2017 (Department of Finance, Government of Pakistan, 2017). Urbanization then was sure to take effect followed by creeping urban sprawl.

3. Methods and data used

3.1. Satellite datasets acquisition

Multi-temporal LANDSAT satellite images were acquired from the USGS GLOVIS site (https://glovis.usgs.gov). A brief description of satellite data and other variables is provided in Table 1.

3.2. Thematic layers

ML prediction algorithms require variables that are responsible for a particular change. For LULC changes the most influential factors that link anthropogenic activities were considered which are the same for urban sprawl across the world (Li and Li, 2019). Like other variables, proximate variables help analyze the driving causes for land-use change (Serneels and Lambin, 2001; Nahuelhual et al., 2012). The socioeconomic factors are necessary to understand the change dynamics. HDI (Human Development Index) and population are precursors of development; leading to the increase of built-up areas (Yumashev et al., 2020). Therefore, population and HDI data are assessed to generate a potential reason for the change in classes.

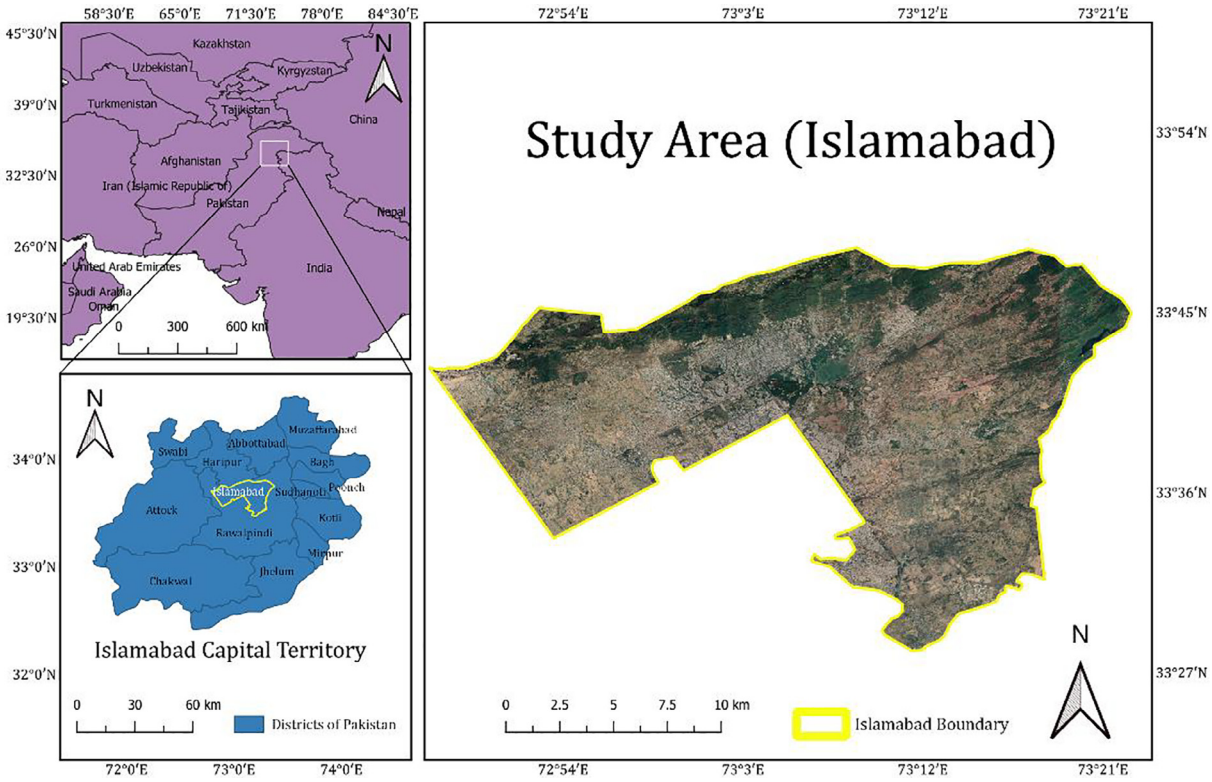

Fig. 1. Locational map of the study area (Islamabad).

Table 1
Datasets used in the study.

Data type	Details	Source	Date of acquisition
LANDSAT 5 TM	Path: 150; Row: 37	USGS GLOVIS	September 18, 1991
LANDSAT 5 TM	Path: 150; Row: 37	USGS GLOVIS	September 10, 2000
LANDSAT 5 TM	Path: 150; Row: 37	USGS GLOVIS	September 25, 2011
LANDSAT 8 OLI	Path: 150; Row: 37	USGS GLOVIS	August 19, 2021
Thematic layers maps	Digital elevation model (DEM)	NASA EARTHDATA	2021
	Slope	Calculated from DEM	–
	Aspect	Calculated from DEM	–
	Hillshade	Calculated from DEM	–
	Distance from main roads	OpenStreetMap	2021
	Distance from waterways	OpenStreetMap	2021
	Distance from the city center	OpenStreetMap	2021
Socioeconomic factors	Population (1991 to 2031)	UN country profile	2021
	HDI (1991 to 2019)	UN country profile	2021

3.3. LULC classification

Training data was prepared for supervised ML classification of satellite imagery according to different Landsat bands. Random Forest (RF) – an ML algorithm, was used for classification using Orfeo Toolbox applied through QGIS (Quantum GIS) as it gave higher overall and class-specific accuracy in this study case. According to studies, it is established that the RF algorithm has the highest accuracy for LULC classifications and remote sensing due to its ability to handle high-dimensional data sets (Talukdar et al., 2020; Ramezan et al., 2021; Rumora et al., 2020; Abdi, 2019).

For modeling ease to assess urban sprawl from built-up area changes, four classes were subsequently selected. Islamabad's open areas have wild vegetation, so barren soil is very unlikely in the monsoon seasons. The others class training data was at first supervised for different non-required spectral intensities. Therefore, many sub-classes were generated for them. Since the study is built-up area oriented the many sub-classes pixels were merged to the Others class. A brief description is provided in Table 2. The assessment of LULC classification maps of years (1991, 2001, 2011 & 2021) was validated by kappa coefficients with high ranges from 0.927 to 0.977. The finalized maps were then further analyzed for LULC change dynamics to assess the urban growth for 1991–2021.

3.4. LULC change analysis

To calculate the spatiotemporal change data from the LULC classification maps, the area analysis tool of the Modules for Land-Use Change Simulation (MOULSCE), a free and most widely used plugin for urban modeling and future scenario simulations through QGIS (Abbas et al., 2021; El-Tantawi et al., 2019; Kafy et al., 2021; Reddy et al., 2019; Gantumur et al., 2020) was used. The data was calculated for the study intervals of (1991 to 2001, 2001 to 2011, and 2011 to 2021) to generate change maps on which further ANN modeling was done.

Table 2
LULC classes characteristics.

Class Feature	Description
Built-up Area	Impervious man-made surfaces that include residential, commercial, and other structures.
Forest	Lush green forests and collections of high-density trees.
Waterbodies	Rivers, lakes, and man-made ponds and dams.
Others	Low Vegetative matter and soil, fallow agricultural land, rock outcrops, etc.

3.5. CA-ANN modeling and validation measures

The ANN model is a reliable tool that has been used in numerous research studies for future LULC predictions (Rahman and Esha, 2020; Saputra and Lee, 2019). The purpose of the model is to generate a transitional potential map using different computational intelligence aspects. The strategy of the model is to handle enormous amounts of uncertain data. ANN incorporates fuzzy logic, by describing the terrain on a continuous range from 0 to 1. The alteration of the weigh connections between geographically linked neurons is an important element of ANN (Bhattacharya et al., 2021). They are dependent on the computation power offered (MohanRajan et al., 2020). Nevertheless, it is highly suitable for urban growth modeling as it can connect well between the complex relationship of huge data fed and extracted.

LULC maps of 1991 and 2001 and the spatial variables were used to predict the map of 2011. MOLUSCE plugin also offers validation of the simulated and actual map to approve the accuracy of the model using % of correctness and Kappa validation coefficients. In the transition potential modeling module of MOLUSCE, based on satisfactory results it was concluded that the ANN model for the study data training was better with a neighborhood value of 1x1 pixels, the learning rate of 0.001, maximum iterations set to 100, with 12 hidden layers, and the momentum value of 0.001 (Perović et al., 2018).

Then in the cellular automata simulation module, using an iteration value of 1, the LULC map of 2011 was simulated. Afterward, in the validation module, the simulated LULC map of 2011 was validated with the reference satellite-based LULC map of 2011. The performance of the algorithms was also assessed by confusion matrix derivates such as classification accuracy and Kappa statistics. In addition, the accuracy of the prediction was assessed by Root Mean Square Error (RMSE) (Achu et al., 2021). After obtaining satisfactory results, the process was repeated for simulation of the LULC map of 2021 from the LULC maps 2001 and 2011 along with the thematic layers. And, then for the 2031 forecast, LULC maps of 2011 and 2021 were used; as for the 2041 forecast, LULC maps of 2001 and 2021 maps were used. The procedure for the entire study is mentioned in Fig. 2.

3.6. Urban sprawl assessment

Urban sprawl assessments have been done through many strategies that involve computations and equations such as diversity, density, and spatial compactness measures (Steurer and Bayr, 2020; Mehriar et al., 2020; Magidi and Ahmed, 2019). But here the approach was a simple one and was done to provide a visual representation to the planners by describing the characteristics of

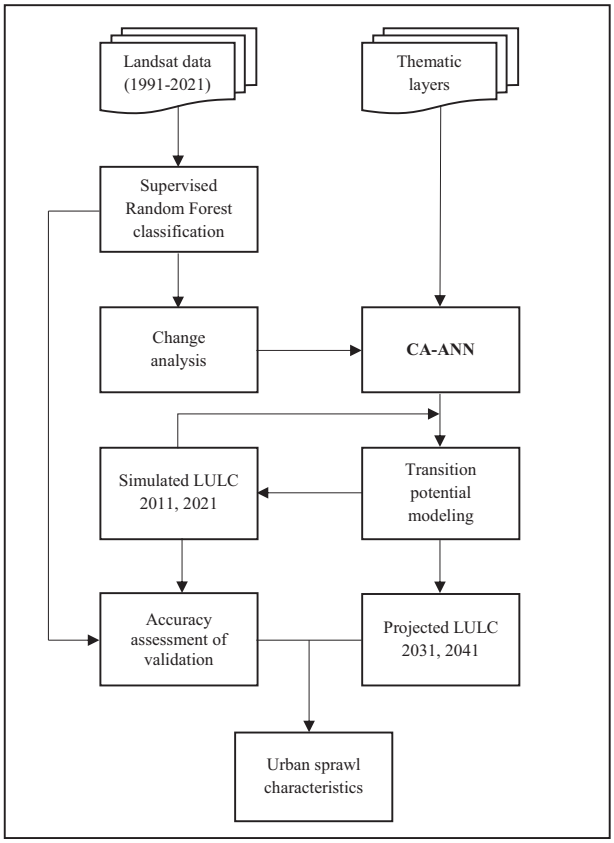

Fig. 2. Methodology chart adopted for the study.

urban sprawl. Urban sprawls often have certain characteristics that can be easily deduced from LULC maps (Ewing, 2008).

In the LULC maps, the built-up areas that fit the characteristics were chosen, and vector lines were drawn across them. Consequently, for the next decade, the changing and the well-matched built-up areas were considered. The considered sprawl types are described briefly in Table 3.

4. Results and discussion

The urban growth patterns were deduced from the LULC maps for the years 1991–2021. Later on, the urban growth influencing thematic layers were prepared to calibrate the model and then simulate for the years 2031 and 2041. The model's accuracy was verified with simulations of 2011 and 2021, as its validation was needed for best results.

4.1. Spatio-temporal LULC mapping and urban growth

It was found that for the study period (1991–2021), the built-up area has continuously increased from 50.7 km^2 to 193.1 km^2 with an annual rate of change of 4.7 km^2 and 0.47%. The built-up area increased from 50.7 km^2 (5%) to 119.5 km^2 (11.9%) (Fig. 3 and Table 4); then it had a low growth of 0.8% occupying 128.3 km^2 (12.7%) land in 2011. It again spiked with the growth of 6.4%, increasing to 193.1 km^2 (19.1 %) in 2021. Table 4 and 5 defines the Spatio-temporal area and percentage changes for the LULC classes according to the decades.

Table 3
Brief descriptions of sprawl types used for assessment.

Sprawl Type	Description
Strip development	This development follows the direction of the highways, as the commercials and homes are aligned on them.
Scattered development	It happens when societies start consuming open lands for economical reasons. This leads to large developments in far lands near the urban areas.
Leapfrog development	It is the dispersed development of parcels situated next to urbanized areas.

For 1991–2001 built-up area increase is justified by the high population growth rate of 5.7% to 3.52% and HDI of 0.508 to 0.577 in Table 6. It can be deduced that 2011′s high HDI would in turn then have more influence on the next decade, and it is confirmed as 2021′s built-up area showed a much higher increase than the previous decades. So, it can be logically deduced that HDI had been increasing up to 2021, though that data has not been calculated so far. Noticeably, population and HDI influence urban growth, and in recent years it is noted that sudden growths cause urban sprawl problems. Both of the parameters have been high in Islamabad due to the chief reason that it is a new capital city, and the high migration rates into the city are due to the high HDI and GDP of the city.

4.2. Urban growth transition potential modeling

Using CA-ANN, urban growth modeling was done to simulate LULC patterns for 2011 and 2021; which were validated with satellite-based LULC maps. And, then the prediction of the future LULC patterns was done for the years 2031 and 2041 using the previous 3 decades LULC maps, and the contributing thematic layers: DEM, slope, aspect, hill shade, distance from waterways and roads, distance from the city center.

4.2.1. Influence of thematic layers on urban growth

The simulated study results show a diverse result from the thematic layers, as they were influenced by the waterways, roads, and geomorphic variables. The result was the scatter of built-up area along roads and rivers; and the origination of remote pockets of development near the urban areas, rather than the densification of the city's core areas. The variables' contribution in the study is further explained in detail in relevance with the validation simulations of 2011 and 2021 in Fig. 3.

4.2.1.1. Digital elevation model (DEM). The city's terrain has sharp changing elevations (Fig. 4 (a)) formed by the waterways (Fig. 4 (c)) inside the urbanized areas, and when the model trained on LULC maps of 1991 and 2001, it learned that the built-up areas were built on almost stable and leveled regions, as at that time, the city had been newly built on plain areas. Due to this reason, the simulations model focused on plain areas (Fig. 3 (2011 & 2021)), such as near roads, waterways, and open fields.

4.2.1.2. Slope. Plausibly due to the elevation changes, slopes would be generated that are known to exert great control over the growth of built-up areas. The previous built-up areas were built on slopes of <4° (dark purple regions) (Fig. 4 (e)); which is also the majority of the city's central area. Urbanization had occurred around these slopes due to the ease and economy of construction. Consequently, the LULC simulations focused on the non-undulatory areas by avoiding slopes >8° and resulting in the prediction of built-up areas near roads, waterways, and open fields where pockets of built-up areas were generated.

4.2.1.3. Aspect. Aspect describes the direction of the slope, and its inclusion was done to simulate a realistic scenario. It acted in coordination with DEM and slope variables by accurately predicting the simulations. The eccentricity of the aspect (Fig. 4 (d)) highlights that the city has an irregular terrain.

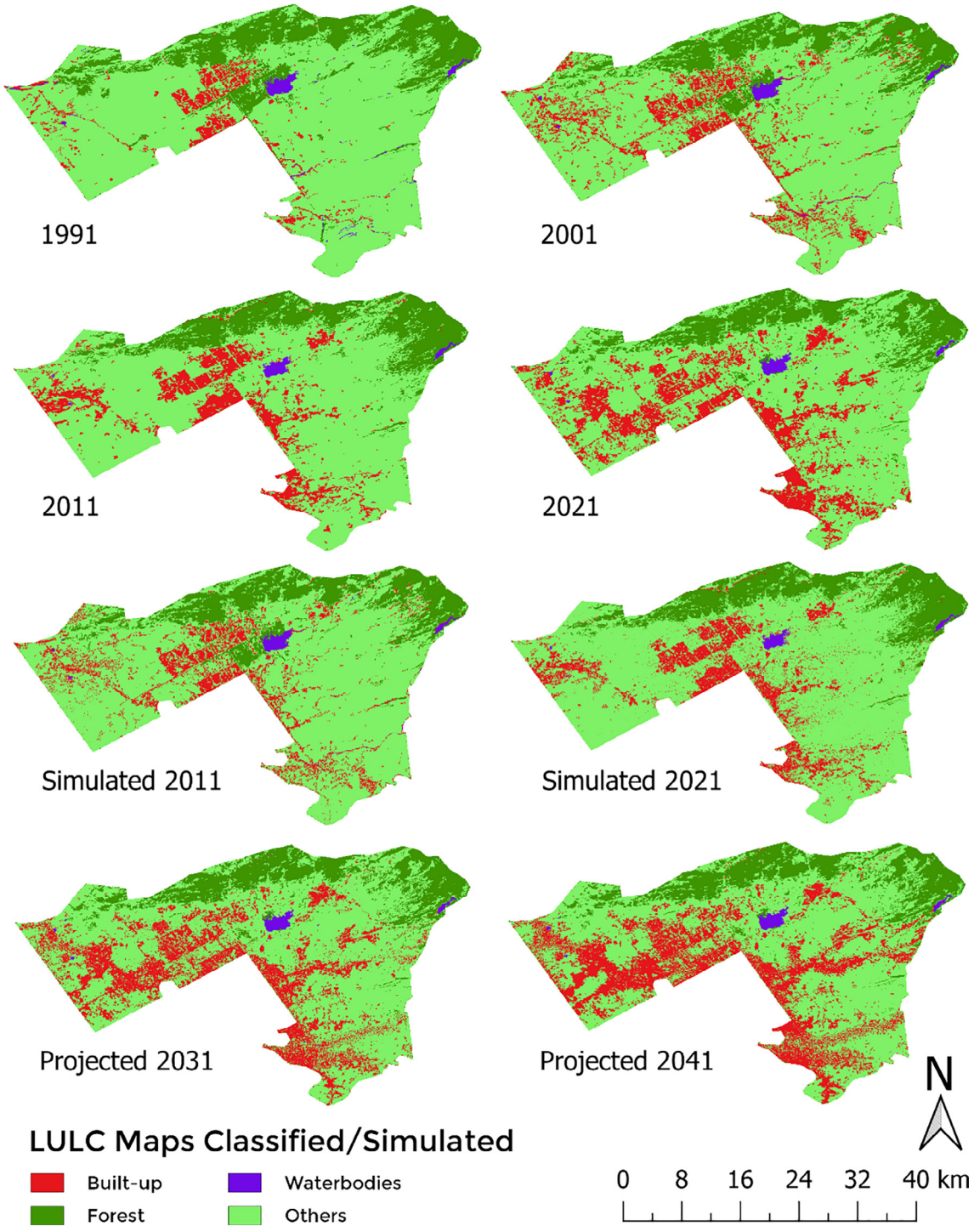

Fig. 3. Satellite based LULC maps for years 1991 to 2021 with simulations of 2011 and 2021 and projections for 2031 and 2041.

Table 4
LULC area from 1991 to 2021 (km^2) and the annual rate of change (ARC).

LULC Class	1991		2001		2011		2021		Annual Rate of Change	
	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
Built-up land	50.7	5.0	119.5	11.9	128.3	12.7	193.1	19.1	4.7	0.47
Forest	193.3	19.2	161.7	16.0	169.1	16.8	154.9	15.4	−1.3	−0.13
Waterbodies	10.9	1.1	9.7	1.0	7.1	0.7	8.1	0.8	−0.1	−0.01
Others	753.6	74.7	717.7	71.1	704.2	69.8	652.5	64.7	−3.4	−0.33

Table 5
The increase or decrease in the area (km^2) and % (from the total area) for the classes from 1991 to 2021.

LULC Class	1991–2001		2001–2011		2011–2021	
	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
Built-up land	68.8	6.8	8.7	0.9	64.8	6.4
Forest	−31.6	−3.1	7.4	0.7	−14.2	−1.4
Waterbodies	−1.2	−0.1	−2.6	−0.3	1.0	0.1
Others	−36.0	−3.6	−13.5	−1.3	−51.7	−5.1

Table 6
Population and HDI per capita for Islamabad according to the study.

Year	Population	Population Growth Rate	HDI
1991	362,411	5.70%	0.508
2001	588,725	3.52%	0.577
2011	832,715	3.53%	0.696
2021	1,163,584	3.05%	–
2031	1,514,316	2.52%	–

4.2.1.4. Hill shade. Hill shade (Fig. 4 (f)) is a form of 3d representation of surface that accounts for the relative position of the sun. Much like aspect, its inclusion would allow for a better simulation; as the orientation of a planned city and a structure are influenced by the path of the sun for sustainable design, ventilation, lighting, etc. Accordingly, it was less likely that built-up areas were found under the shades of hills.

4.2.1.5. Distance from main roads. The variable is a proximity map of the main roads. And, it displays a very high density of main roads in the city central parts and lower density in the suburban areas (Fig. 4 (b)). Urbanizations occur mostly near the main roads; as they allow high accessibility. This consequently promotes strip developments along the roads as the nearer ones are given high priority for growth and vice versa. This realistic event is also emulated in the simulations, as densification of built-up areas occurred along the main roads (Fig. 4 (Simulated 2011 & 2021)).

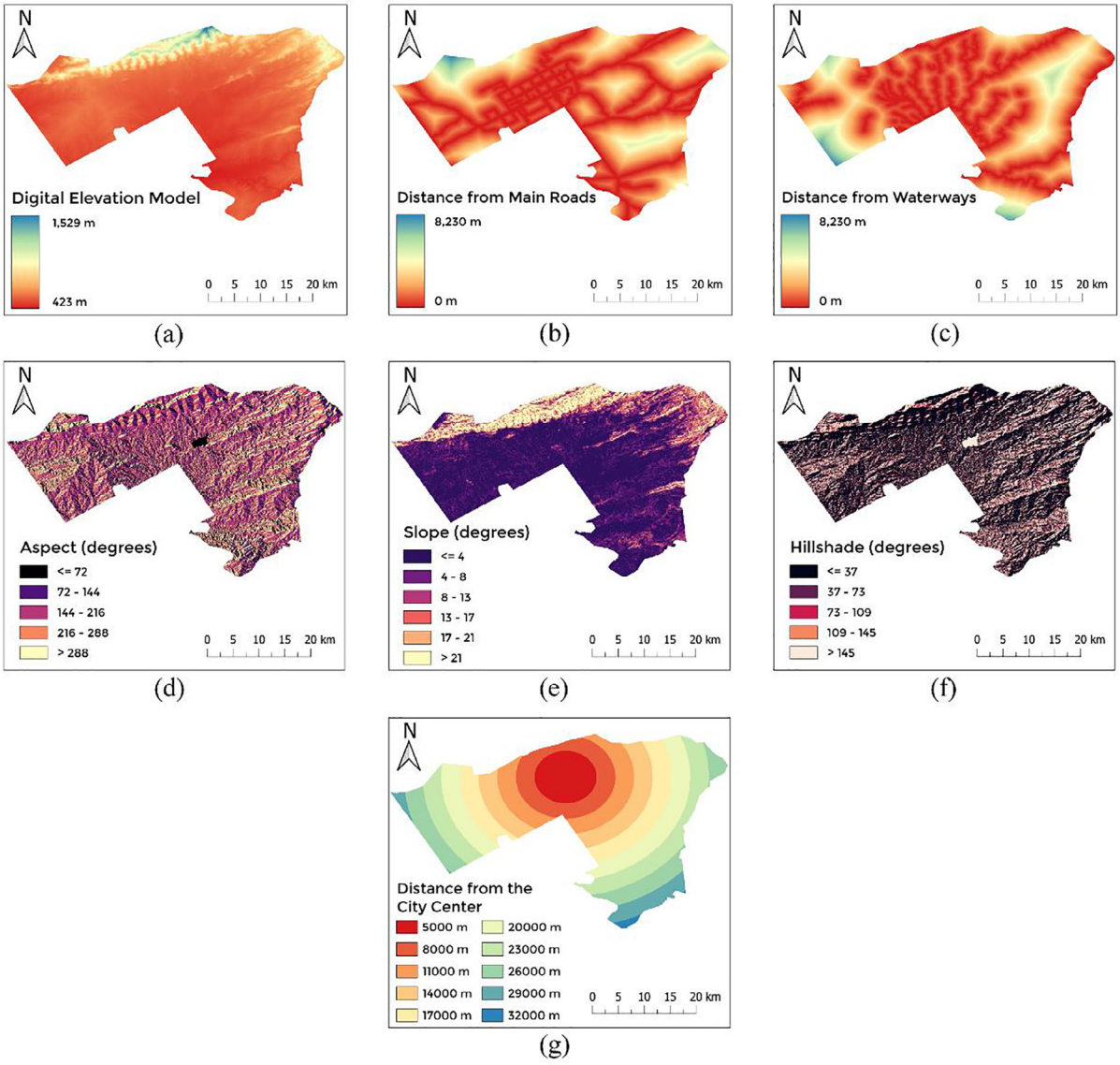

Fig. 4. Thematic layers: (a) DEM, (b) distance from main roads, (c) distance from waterways, (d) aspect, (e) slope, (f) hillshade, (g) distance from the city center.

Table 7
Simulated and satellite-based LULC for 2021.

LULC Class	Satellite-based		Simulated		Accuracy	Kappa Value	
	Area (km^2)	%	Area (km^2)	%		ANN	Validation
Built-up land	193.1	19.1	174.8	17.3	75.9%	0.71	0.86
Forest	154.9	15.4	161.2	16.0			
Waterbodies	8.1	0.8	6.3	0.6			
Others	652.5	64.7	666.5	66.1			

Table 8
Simulated and satellite-based LULC for 2011.

LULC Class	Satellite-based		Simulated		Accuracy	Kappa Value	
	Area (km^2)	%	Area (km^2)	%		ANN	Validation
Built-up land	128.3	12.7	111.2	11.0	80.2%	0.83	0.94
Forest	169.1	16.8	160.5	15.9			
Waterbodies	7.1	0.7	9.1	0.9			
Others	704.2	69.8	727.9	72.2			

Table 9
Confusion matrix of built-up pixels for satellite-based and simulated 2021 LULC map.

		Predicted		Percentage Correct	Overall Accuracy
		Non-built-up	Built-up		
Actual	Non-built-up	843,740	62,368	93.12%	82.09%
	Built-up	138,318	76,247	64.46%	

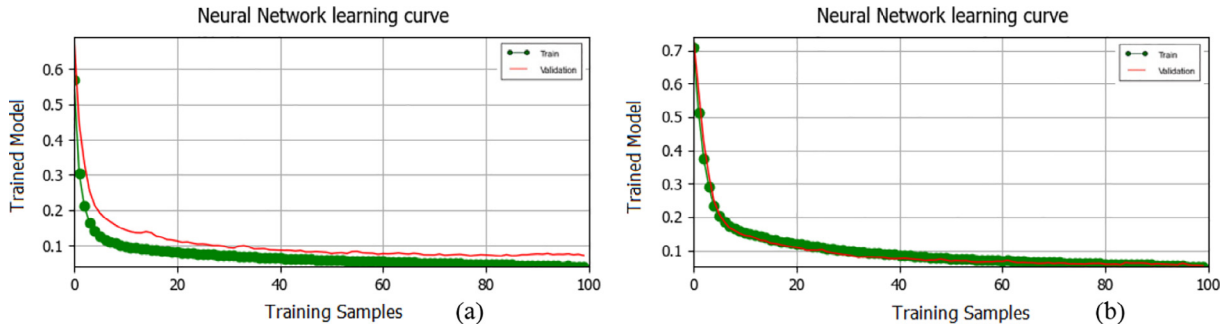

Fig. 5. ANN plot for the 2011 (a) and 2021 (b) simulations.

Table 10
Predicted area statistics for years 2031 and 2041.

LULC Class	2021		2031		2041	
	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
Built-up land	193.1	19.1	205.6	20.4	251.3	24.9
Forest	154.9	15.4	148.4	14.7	146.2	14.5
Waterbodies	8.1	0.80	7.9	0.78	7.7	0.76
Others	652.5	64.7	646.8	64.1	603.4	59.8

4.2.1.6. Distance from waterways. The distance from waterways like main roads is a proximity map that would promote higher growth to areas nearer and vice versa. In the simulations, its effect was highlighted by the strip development of built-up areas along the rivers (Fig. 3 (Simulated 2011)). The city's high vegetation is due to the branching system of waterways originating from the lakes and hills (Fig. 4 (c)).

4.2.1.7. Distance from city center. The distance from the city center is a proximity buffer (Fig. 4 (g)) based on a monocentric city model, created to simulate a directional urban growth pattern. The main city is mostly situated around the central areas; as in the early 80s, the new city was mostly run by elite government employees and workers; who could afford living expenses. But most probably later on, as the population increased, the influx settled in remote areas due to cheap land rates. However, due to the increase of GDP and HDI of the city; the densification of central areas has started, this was well detected in the LULC maps and the simulations (Fig. 3).

4.3. Validation

The validation module in the MOLUSCE plugin does an accuracy assessment of the simulated and reference satellite-based LULC

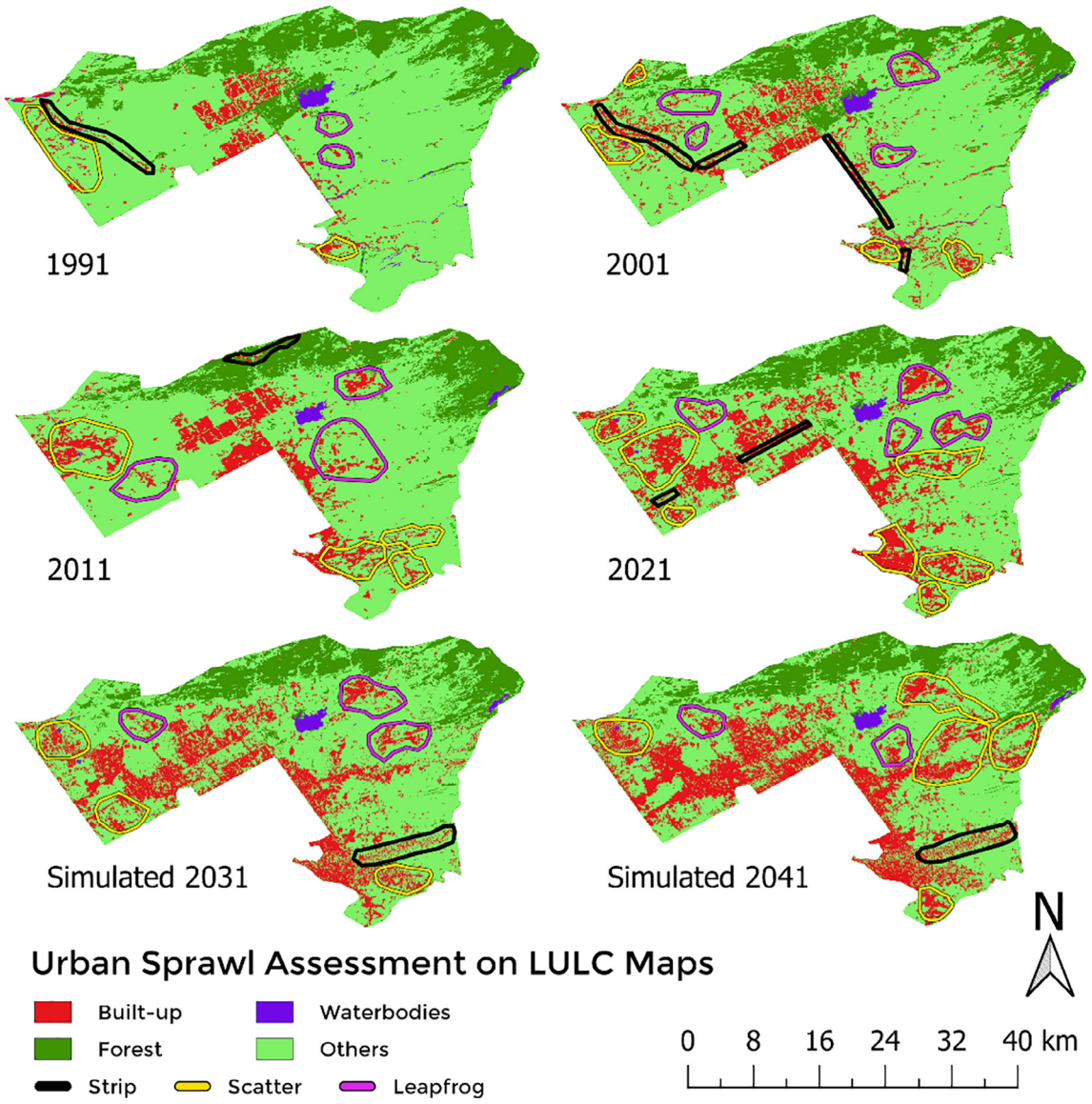

Fig. 6. Urban sprawl assessment of the satellite-based and predicted LULC maps.

map. Two accuracy assessments for the years 2011 and 2021 were done so that the model can be well validated before any future predictions. The statistical validation was achieved by a high Kappa value (>0.85) in Tables 7 & 8; while the spatial similarity was good enough (>75%) in Fig. 3. The high Kappa values in ANN modeling and validation indicate an excellent accuracy (Foody, 2020) of the modeling; which can then further be used for predictions. Tables 7 & 9, show that the built-up areas for satellite-based and simulated LULC maps were analogous.

The overall accuracy for the LULC simulation and actual 2021 was 82.09% in the confusion matrix in Table 9. While the RMSE error for actual and simulated LULC's of 2011 and 2021 were 0.46859 and 0.50260. Fig. 5 highlights the ANN plot of model learning with training and validation data. Huge spikes did not emerge, and the curve followed a smooth pattern.

4.4. Future urban growth

The predictions were done by the CA-based urban growth model, which was validated with satisfactory results. Table 9 for the year 2031 illustrates that the built-up area is projected to increase by 12.5 km^2 and the urban growth of 1.3%. And, for the year 2041 (Table 10); the built-up area is projected to increase by 45.7 km^2, and the urban growth of 4.5%. The study exhibits that high urban growth is expected to occur in the year 2041.

4.5. Built-up area sprawl assessment

The urban sprawl assessment was done to highlight the type of sprawl occurring in this study. Three types of urban sprawl criteria were selected as their visual assessment was possible through LULC maps. It was detected that scattered development was the main contributor to urban sprawl. The setup of societies in remote areas of the cities promoted scattered development, which led to strip development across the main roads (Fig. 6). Consequently, leapfrog developments started in the open regions between the main city and the scattered developments.

5. Conclusions and recommendations

This study incorporates spatio-temporal LULC monitoring (1991–2021) and urban growth modeling (2021–2041) of Islam-

abad using the ANN-CA model. Seven thematic layers along with three decades of satellite-based LULC maps were used to simulate LULC maps for 2011 and 2021 to validate the used model using the MOLUSCE plugin and other accuracy assessment measures. Then, the validated model was used to predict future LULC maps of 2031 and 2041. Visual urban sprawl assessment was done on satellite-based and predicted LULC maps to highlight the type of urban sprawl.

1. In the period 1991–2021, the satellite-based LULC showed a net gain of 142.4 km^2 (14.1%) in built-up areas; this rapid urban growth had crucially altered other features by reducing: 38.4 km^2 (-3.8%) of forest land, 2.9 km^2 (-0.3%) of waterbodies land, and 101.1 km^2 (-10.2%) others land.
2. The CA-based predicted LULC of 2031 and 2041 show that 58.2 km^2 (5.7%) urban growth will occur in 2021–2041.
3. Visual urban sprawl assessment of LULC maps exhibits that scattered and leapfrog developments were the influencing urban sprawl types. It was further deduced from the LULC maps that the urbanization of the city has been uneven.
4. The increase of built-up areas will lead to chaotic urban sprawls that would have drastic effects on urban systems – infrastructures, environments, and networks. The predicted maps show the strip urbanizations around the rivers; which will ultimately cause the lowering of aquifers, and lead to a water crisis in the city.

It is recommended that the city should promote sustainable urbanization by inducing society planning, environmental, and controlled development policies. Revision of the city's master plan is urgently needed to incorporate the predicted urbanization, and prevent the imminent chaotic situation.

The research is intended to be useful for policymakers, urban planners, and administrative authorities; so that they can grab the bigger picture of the current and upcoming urban sprawl. Those, who want to further pursue the research; should consider satellite data at shorter intervals and should employ more variables that would in turn allow them to explore new results.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The careful review and constructive suggestions by the anonymous reviewers are gratefully acknowledged.

References

Abbas, Z., Yang, G., Zhong, Y., Zhao, Y., 2021. Spatiotemporal change analysis and future scenario of LULC Using the CA-ANN approach: a case study of the Greater Bay Area, China. *Land* 10 (6), 584. https://doi.org/10.3390/land10060584.

Abdi, A.M., 2019. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GISci. Remote Sens. 57 (1), 1–20. https://doi.org/10.1080/15481603.2019.1650447.

Achu, A., Thomas, J., Aju, C., Gopinath, G., Kumar, S., Reghunath, R., 2021. Machine-learning modelling of fire susceptibility in a forest-agriculture mosaic landscape of southern India. Ecol. Inf. 64,. https://doi.org/10.1016/j.ecoinf.2021.101348 101348.

Aurambout, J.-P., Barranco, R., Lavalle, C., 2018. Towards a simpler characterization of urban sprawl across Urban Areas in Europe. Land, MDPI 7 (1), 33. https://doi.org/10.3390/land7010033.

Baqa, M.F., Chen, F., Lu, L., Qureshi, S., Tariq, A., Wang, S., Jing, L., Hamza, S., Li, Q., 2021. Monitoring and modeling the patterns and trends of urban growth using urban sprawl matrix and CA-Markov Model: a case study of Karachi, Pakistan. *Land* 10 (7), 700.

Bhattacharya, R.K., Chatterjee, N.D., Das, K., 2021. Land use and Land Cover change and its resultant erosion susceptible level: an appraisal using RUSLE and Logistic Regression in a tropical plateau basin of West Bengal, India. Environ. Dev. Sustain. 23, 1411–1446. https://doi.org/10.1007/s10668-020-00628-x.

Bueno-Suárez, C., Coq-Huelva, D., 2020. Sustaining what is unsustainable: a review of urban sprawl and urban socio-environmental policies in North America and Western Europe. Sustainability 12 (11), 4445. https://doi.org/10.3390/su12114445.

Coq-Huelva, D., Asián-Chaves, R., 2019. Urban sprawl and sustainable urban policies. A review of the Cases of Lima, Mexico City and Santiago de Chile. *Sustainability* 11 (20), 5835. https://doi.org/10.3390/su12114445.

Department of Finance, Government of Pakistan, 2017. Population, Labour Force & Employment. In: Pakistan Economic Survey 2017–18. Finance Department, Islamabad, pp. 180–188.

El-Tantawi, A.M., Bao, A., Chang, C., Liu, Y., 2019. Monitoring and predicting land use/cover changes in the Aksu-Tarim River Basin, Xinjiang-China (1990–2030). Environ. Monit. Assess. 191, 480. https://doi.org/10.1007/s10661-019-7478-0.

Ewing, R.H., 2008. Characteristics, causes, and effects of sprawl: a literature review. In: Marzluff, J.M., Shulenberger, E., Endlicher, W., Alberti, M., Bradley, G., Ryan, C., Simon, U., ZumBrunnen, C. (Eds.), Urban Ecology. Springer US, Boston, MA, pp. 519–535.

Foody, G.M., 2020. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. Remote Sens. Environ. 239,. https://doi.org/10.1016/j.rse.2019.111630 111630.

Zhao, Y. (2020). Spatiotemporal dynamics of urban expansion and its simulation using CA-ANN model in Ulaanbaatar, Mongolia. *Geocarto Int.*. 10.1080/10106049.2020.1723714.

Hatab, A.A., RigoCavinato, M.E., Lindemer, A., Lagerkvist, C.-J., 2019. Urban sprawl, food security and agricultural systems in developing countries: a systematic review of the literature. Cities 94, 129–142. https://doi.org/10.1016/j.cities.2019.06.001.

Hu, Z., Lo, C., 2007. Modeling urban growth in Atlanta using logistic regression. Comput. Environ. Urban Syst. 31 (6), 667–688. https://doi.org/10.1016/j.compenvurbsys.2006.11.001.

Kafy, A. -A., Abdullah-Al-Faisal, Rahman, M. S., Islam, M., Rakib, A. A., Islam, M. A., Sattar, G. S. (2021). Prediction of seasonal urban thermal field variance index using machine learning algorithms in Cumilla, Bangladesh. *Sustainable Cities Soc., 64*, 102542. 10.1016/j.scs.2020.102542.

Li, G., Li, F., 2019. Urban sprawl in China: differences and socioeconomic drivers. Sci. Total Environ. 673, 367–377. https://doi.org/10.1016/j.scitotenv.2019.04.080.

Liu, Y., Fan, P., Yue, W., Song, Y., 2018. Impacts of land finance on urban sprawl in China: the case of Chongqing. Land Use Policy 72, 420–432. https://doi.org/10.1016/j.landusepol.2018.01.004.

Magidi, J., Ahmed, F., 2019. Assessing urban sprawl using remote sensing and landscape metrics: A case study of City of Tshwane, South Africa (1984–2015). Egypt. J. Remote Sens. Space Sci. 22 (3), 335–346. https://doi.org/10.1016/j.ejrs.2018.07.003.

Mehriar, M., Masoumi, H., Mohino, I., 2020. Urban sprawl, socioeconomic features, and travel patterns in middle east countries: a case study in Iran. Sustainability 12 (22), 9620. https://doi.org/10.3390/su12229620.

MohanRajan, S.N., Loganathan, A., Manoharan, P., 2020. Survey on Land Use/Land Cover (LU/LC) change analysis in remote sensing and GIS environment: techniques and challenges. Environ. Sci. Pollut. Res. 27, 29900–29926. https://doi.org/10.1007/s11356-020-09091-7.

Mustafa, A., Heppenstall, A., Omrani, H., Saadi, I., Cools, M., Teller, J., 2018. Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. Comput. Environ. Urban Syst. 67, 147–156.

Nahuelhual, L., Carmona, A., Lara, A., Echeverría, C., González, E., 2012. Land-cover change to forest plantations: proximate causes and implications for the landscape in south-central Chile. Landscape Urban Plann. 107 (1), 12–20. https://doi.org/10.1016/j.landurbplan.2012.04.006.

Otuoze, S.H., Hunt, D.V., Jefferson, I., 2020. Predictive modeling of transport infrastructure space for urban growth phenomena in developing countries' cities: a case study of Kano — Nigeria. Sustainability 13 (1), 308. https://doi.org/10.3390/su13010308.

Patil, S.D., Gu, Y., Dias, F.S., Stieglitz, M., Turk, G., 2017. Predicting the spectral information of future land cover using machine learning. Int. J. Remote Sens. 38 (20), 5592–5607. https://doi.org/10.1080/01431161.2017.1343512.

Perović, V., Jakšić, D., Jaramaz, D., Koković, N., Čakmak, D., Mitrović, M., Pavlović, P., 2018. Spatio-temporal analysis of land use/land cover change and its effects on soil erosion (Case study in the Oplenac wine-producing area, Serbia). Environ. Monit. Assess. 190 (675). https://doi.org/10.1007/s10661-018-7025-4.

Rahman, M.T., Esha, E.J., 2020. Prediction of land cover change based on CA-ANN model to assess its local impacts on Bagerhat, southwestern coastal Bangladesh. Geocarto Int. https://doi.org/10.1080/10106049.2020.1831621.

Ramezan, C.A., Warner, T.A., Maxwell, A.E., Price, B.S., 2021. Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. Remote Sensing 13 (3), 368. https://doi.org/10.3390/rs13030368.

Rana, I. A., Bhatti, S. S., 2018. Lahore, Pakistan – Urbanization challenges and opportunities. *Cities, 72*(Part B), 348–355. 10.1016/j.cities.2017.09.014.

Ranagalage, M., Morimoto, T., Simwanda, M., Murayama, Y., 2021. Spatial analysis of urbanization patterns in four rapidly growing south Asian cities using sentinel-2 Data. Remote Sensing 13 (8), 1531. https://doi.org/10.3390/rs13081531.

Reddy, C.S., Pasha, S.V., Satish, K.V., Unnikrishnan, A., Chavan, S.B., Jha, C.S., Diwakar, P.G., Dadhwal, V.K., 2019. Quantifying and predicting multi-decadal forest cover changes in Myanmar: a biodiversity hotspot under threat. Biodivers. Conserv. 28 (5), 1129–1149.

Rienow, A., Goetzke, R., 2015. Supporting SLEUTH – Enhancing a cellular automaton with support vector machines for urban growth modeling. Comput. Environ. Urban Syst. 49, 66–81. https://doi.org/10.1016/j.compenvurbsys.2014.05.001.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., Roberts, D., 2008. Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. Remote Sens. Environ. 112 (5), 2272–2283. https://doi.org/10.1016/j.rse.2007.10.004.

Rumora, L., Miler, M., Medak, D., 2020. Impact of various atmospheric corrections on sentinel-2 land cover classification accuracy using machine learning classifiers. ISPRS Int. J. Geo-Inf. 9 (4), 277. https://doi.org/10.3390/ijgi9040277.

Saputra, M.H., Lee, H.S., 2019. Prediction of land use and land cover changes for North Sumatra, Indonesia, using an artificial-neural-network-based cellular automaton. Sustainability 11 (11), 3024. https://doi.org/10.3390/su11113024.

Serneels, S., Lambin, E.F., 2001. Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. Agric. Ecosyst. Environ. 85 (1–3), 65–81. https://doi.org/10.1016/S0167-8809(01)00188-8.

Steurer, M., Bayr, C., 2020. Measuring urban sprawl using land use data. Land Use Policy 97,. https://doi.org/10.1016/j.landusepol.2020.104799 104799.

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—a review. Remote Sensing 12 (7), 1135.

The Government of Pakistan, 2017. Provisional Summary Results of 6th Population and Housing Census-2017. Pakistan Bureau of Statistics, Islamabad.

Tripathy, P., Kumar, A., 2019. Monitoring and modelling spatio-temporal urban growth of Delhi using Cellular Automata and geoinformatics. Cities 90, 52–63.

Vinayak, B., Lee, H.S., Gedem, S., 2021. Prediction of land use and land cover changes in mumbai city, india, using remote sensing data and a multilayer perceptron neural network-based Markov Chain model. Sustainability 13 (2), 471. https://doi.org/10.3390/su13020471.

Yumashev, A., Ślusarczyk, B., Kondrashev, S., Mikhaylov, A., 2020. Global indicators of sustainable development: evaluation of the influence of the human development index on consumption and quality of energy. Energies 13 (11), 2768. https://doi.org/10.3390/en13112768.

Zafar, S., Zaidi, A., 2019. Impact of urbanization on basin hydrology: a case study of the Malir Basin, Karachi, Pakistan. Reg. Environ. Change 19, 1815–1827. https://doi.org/10.1007/s10113-019-01512-9.

Zhang, Q., Ban, Y., Liu, J., Hu, Y., 2011. Simulation and analysis of urban growth scenarios for the Greater Shanghai Area, China. Comput. Environ. Urban Syst. 35 (2), 126–139. https://doi.org/10.1016/j.compenvurbsys.2010.12.002.