Urban Land Use Planning

Urbanization is persistent globally and has increasingly significant spatial and environmental consequences. It is especially challenging in developing countries due to the increasing pressure on the limited resources, and damage to the bio-physical environment.

1. Introduction

The degree to which human actions have an effect on changes in the environment has been a key subject of study for land management researchers. One can measure these changes, which often have a spatial dimension, both qualitatively and quantitatively. Nevertheless, mapping of temporal and spatial changes of urban and rural land still remains a challenging task because the technological tools and instruments have so far not been adequate to support the daily practice and the spatio-temporal needs of planners and decision makers. The traditional methods used for land use planning such as field surveys and participatory mapping are time consuming, costly and labor intensive. Advancements in the data acquisition technologies and the availability of improved computational power has made it possible to make practical use of the algorithms which earlier could only be called theoretical solutions but could not be practiced [1]. The methods of imaging land use fall in the domain of remote sensing (RS) and processes by which they are utilized for their analysis is through Geographic Information Systems (GIS) tools. RS datasets are a vital source for assessing the land-use and land-cover processes. RS datasets provide large scale coverages right from the regional to the global scale. One of the ways to understand the status and changes in both natural and built environment is through interpretation of RS datasets. In recent decades, RS sensors and techniques have become increasingly sophisticated. They can provide a large volume of data with superior quality and high spatial resolution [2]. The availability of high-resolution data like LIDAR, RADAR, MSS, Hyperspectral, UAV borne data and other commercially available satellite data and data from other airborne platforms has improved the capabilities and understanding of land use planning. While classical methods compartmentalize graphic and non-graphic data for analysis, a combined use of advanced RS and GIS applications can integrate both spatial and socio-economic factors which may support a better understanding of the dynamics of the society needed to improve land use planning.

The spatial analysis underlying land use planning encompasses land use classification, growth, zoning, restrictions, allocation and change. There are various ML algorithms that make it possible to model aspects of land use planning. However, the algorithms differ in their performances, in terms of spatial accuracies and output possibilities. According to Hagenauer et al. (2019) [3] "ML comprises a set of inductive models that recognize patterns and/or minimize the prediction error of complex regression functions, by means of a repeated learning strategy from training data, linking an output such as land-use change to several underlying drivers. Once learned, the model can then be used to estimate previously unseen cases and predict future land-use change. There are many simulation models to model land use change and growth using ML techniques". In other words, the benefits of ML models are numerous, and have the ability of dealing with large amounts of data and a large number of variables, assigning relative importance to the variables, alongside the ability to model complex nonlinear relationships as well as interactions between drivers, while not being grounded in restrictive distributional assumptions of the input data that are hard to achieve in practice [3]. Despite these benefits, there is still limited literature which systematically compares the currently used ML algorithms in light of the specific needs and requirements of land use planning, classification, change, transition and growth.

2. Theoretical Perspective

Land use planning is a process that ensures the judicious use of land by group of people who benefit from it. In order to make the best use of the limited available land and its resources, land use planners and policy makers need to intervene by incorporating guidelines/regulations on the use of land and to sustain the natural resources. The governments and

administrative bodies involved in land use planning impose regulations on the use of land which include but are not limited to zoning, land use control, land use restrictions and allocation [4]. Rapid urbanization leads to substantial unplanned growth which is not healthy for the environment. It leads to degradation, air pollution and contamination of water resources. A major problem faced by urban planner and decision makers is to channelize this growth. In order to ensure the growth is planned and systematic spatial, non-spatial and temporal data has to be made available. To monitor and solve the problems of unplanned growth, EO and GIS data along with socio-economic data provide vital source of information for updating of land use maps.

In remotely sensed data, spatial resolution and temporal frequencies are the two important factors to study land use change. Proprietary and open-source data provided by various international and national agencies are custodians of EO and GIS data. The land administrators and land use planners need information on built-up area, non-built-up area, water bodies, green area, forests (urban), land use patterns, road network, drainage systems, etc. Most of this data can be obtained and/or extracted from images. Some types of data, such as open street map, USGS earth explorer, Sentinel hub, Copernicus open access hub, GHSL, etc., are repositories of open-source data and are utilized for spatial data requirements. Apart from open-source data, commercially available satellite images, aerial photographs, LIDAR data are used for various land use planning application such as cadastral boundary extraction, 3D feature extraction, 3D modeling, etc. The first step in the various stages of land use planning is to study the past data which is available in the form of master plan, census data and other statistical data available with various government and non-government agencies, the second stage is to study how far the plan has been fulfilled by studying with the help of EO based data against the master plan projections and the third stage will be projecting the land use plan for future, depending on the projection of the statistical data being generated as future projection, i.e., it will be geared towards fulfillment of those projections which will also be dependent on the government policies. A comprehensive geospatial database has to be developed to assess the existing land use and model the future changes. The type of data requirement depends on the type of land use planning problem in hand which needs to be addressed. Depending on the purpose, imagery could be of high resolution and in multi band or it may suffice to have imagery of low resolution as can be seen in **Table 1** depicting the data requirements and the applications to measure the indicators.

Table 1. Land use planning indicators with measurements, data required and applications.

Indicators	Measurements	Data	Application
Urban expansion	**Built-up density, settlement patterns, population distribution**	**EO based data, i.e., classified images, building footprints, urban heat islands**	**Classification and simulation (CA, spatial logistics regression, SVM, random forest, CNN)**
Land restrictions	**Land use/land cover change, built-up and non-built-up spaces,**	**Master plan, building by-laws, land use regulations**	**Classification, extraction of EO products like DEM, vegetation cover**
Land allocations	**Govt. policies, population growth, population distribution**	**Census data, socio-economic data**	**Spatial logistic regression, cellular automata**
Zoning	**Govt. policies and by-laws**	**Master plan, classified images**	**Planned development**
Land use change	**Settlement patterns, urban growth processes, (aggregated, compact, dispersed) population growth**	**Spatio-temporal EO based data**	**Spatial metrics, cellular automata, spatial logistic regression, agent-based modeling**

2.1. Machine Learning Based Algorithms

Several ML algorithms have been tested for their performance on different kinds of datasets for land use classification and simulation of land use planning processes. The more popular algorithms are support vector machine, neural network, Markov random field, GANS and random forest. These algorithms are experimented on different data sets individually and in combination. The article is a review of the functionalities of these algorithms and their application in land use planning.

There is ongoing research for new methods of ML to take land use mapping to a higher plane. Support vector machines (SVMs) have been applied in a number of research papers and have been compared for their performance in land use classification with other ML algorithms such as random forest (RF), neural network. SVMs is a group of non-parametric ML algorithms. The core operation of SVMs is to construct a separating hyperplane (i.e., a decision boundary) on the basis of the properties of the training samples, specifically their distribution in feature space. In many instances, classification in high dimension feature spaces results in over-fitting in the input space, however, in SVMs over-fitting is controlled through the principle of structural risk minimization. The empirical risk of misclassification is minimized by

maximizing the margin between the data points and the decision boundary [5]. In case of computational requirements SVMs work well with small data sets with fewer outliers [6]. Among the decision tree algorithms like CART (classification and regression tree), ID3 (iterative dichotomizer 3) more commonly used algorithm for land use classification is RF. One of the benefits of RF algorithm is that it can be used for both classification and regression. RF works as an ensemble learning algorithm based on decision tree classifiers, bagging, and bootstrapping. Each tree is trained by bootstrapping, using different samples from the training data. Additionally, each tree is trained using a random subset of the predicting variables. RF may use thousands of decision trees, where each tree casts a vote and the prediction of the class is decided by the majority vote. A big committee of randomly created decision tree determines the classification, hence the name, random forest [7]. RF can handle large number of variables without need of deleting any and bringing out the relative importance of each of the variables. As compared to SVMs, RF do not have hyperparameters to tune like choosing the right kernel, regularization, penalty, the slack variable, however, the complexity and computational cost increase with the increase in the number of trees in the forest. A list of some of the most commonly used variables is given in **Table 2**.

Table 2. List of image features.

Variables/Image Feature	Description	Examples	Description	Application in Land Use Planning
Spectral features	Provide information regarding the spectral response of objects, which differ for land coverage types, states of vegetation, soil composition, building materials [8]	NDVI (normalized difference vegetation index)—to measure/identify biomass	NDVI = (NIR−RED)/(NIR+RED)	Distinguishing built-up areas from non-built-up, green vegetation from barren land
		SAVI (soil adjusted vegetation index)	SAVI = 1.5 × (NIR-R)/(NIR+R+0.5)	Differentiate between vegetation and built-up
		BAI (built-up area index)	BAI = (B-NIR)/(B+NIR)	Built-up areas index has good performance in detecting asphalt and concrete surfaces
		NDWI (normalized difference water index)	NDWI = (G-NIR)/(G+NIR)	Enhances water features and helps in distinguishing water features from other ground objects
Texture features	Characterize the spatial distribution of intensity values of an image and data on contrast, uniformity, rugosity, etc. [8]	GLCM (grey level co-occurrence matrix)—specifically relevant when measuring, qualifying	GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image	Measuring spatial patterns which are repetitive on the image like crop land and built-up
Structural features	Help in identifying the spatial arrangement of elements in terms of the randomness or regularity of their distribution [8]	Edge detection filter specifically relevant when measuring, qualifying	Edge detection is a technique used to find the boundaries of features in an image. This uses an algorithm that searches for discontinuities in pixel brightness in an image that is converted to grayscale. ("Applying Edge Detection To Feature Extraction And Pixel Integrity," n.d [9])	For shape recognition, edge enhancement

Apart from SVM and RF, another ML algorithm which has been widely applied for land use classification is deep learning methods. In 2006, deep learning was introduced by Hinton et al. (2015) [10]. Deep learning methods are representation learning methods composed of multiple layers and each layer computes a new data representation from the representation in the previous layers of artificial neurons creating a hierarchy of data abstractions [10]. Among the group of deep learning methods is convolutional neural network (CNN) composed of convolution and pooling that are concluded by a fully connected neural network layer and a proper activation function, i.e., in models that directly reconstruct an output image prediction, such as U-Net and generative models, the fully connected network and activation function is not needed [11]. In deep learning, artificial neural network (ANN) has been gaining importance in land use planning studies. ANN is a computational ML model based on multilayer perception composed of processing elements forming three kinds of layer

(input, hidden, output) which are called perceptrons [12]. Deep learning algorithms work well with relatively large datasets with supporting infrastructure to train them in reasonable time. There has been increasing interest in the Markov random field-based methods for land use classification and land use change as it helps in generating a smooth classification pattern. Markov random field (MRF) is a statistical model based on probability theory which efficiently represents dependency between pixels in a spatial domain. MRF is useful for characterizing spatial-contextual information and has been commonly used for image segmentation, texture analysis, edge detection and image restoration. MRF has been used for linear feature detection with satisfactory results. MRF for modeling spatial context relies on its relationship to Gibbs random field which is a useful way to apply MRF to deal with context [13]. In the paper "Identifying Urban Poverty Using High-Resolution Satellite Imagery and Machine Learning Approaches: Implications for Housing Inequality" [14], six types of image features perimeter, line segment detector (LSD), Hough transform, gray-level co-occurrence matrix GLCM, HoG, and local binary patterns (LBP) were extracted to identify urban poverty in Wuhan, China. The paper utilizes four machine learning regression approaches random forest (RF), Gaussian process regression (GPR), support vector regression (SVR), and neural network (NN) to study whether the features derived are helpful in differentiating urban poverty. It was concluded in the paper that textural features are important in identifying urban poverty in the study area.

In addition to the above-mentioned ML algorithms, there are several simulation models for the purpose of mapping and growth projections of land use. There are mainly two groups according to the key mechanisms to simulate the process of land use change rule-based/process-based models and empirical-statistic models [15].

2.2. Urban Land Use Models

Cellular Automata (CA)have been defined as discrete spatio-temporal dynamic systems based on local rules. In cellular models, geographic space is represented in the form of a geographic grid, such as the cells in a raster Geographic Information System. They are preferred when model states and the probabilities of transitions among those states are known and stable. They are most suitable for measuring, detecting and predicting change processes such as land use change and urban growth [16].

Cellular automata have capacity to handle temporal dynamics. Cellular automata have the following basic features:

- States: each cell can take an integer value that corresponds to the current state of that cell. There is a finite set of states.

- Neighborhood: is a collection of cells that interact with the current one. To perform simulations on a satellite image we normally take the eight surrounding pixels as neighborhood.

- Transition function (f): takes as input arguments the cell and neighborhood values and returns the new state of the current cell.

The transition function is applied to each cell of the grid across several iterations. Therefore, cellular automata have an evolution process because some cells are changing their states across the different iterations [17]. The most commonly used cellular automata model is the Slope, Land cover, Exclusion, Urban growth, Transport and Hill shade (SLEUTH) model which has been in application for a long time. SLEUTH model has been widely used for simulating urban growth and land use change. SLEUTH is open source and has been developed in C programming language. As described by Berberoğlu et al. (2016) [18].

> "The program involves as a series of nested loops: the outer control loop repeatedly executes each growth "history", retaining cumulative statistical data, while the inner loop executes the growth rules for a single iteration, assumed to be a "-year-." The rules apply to one cell at a time and the whole grid is updated as the iterations complete".

Statistical modeling methods are widely used for modeling, assessing, qualifying, quantifying and predicting the degree/extent/direction of land use change and growth. An example of a statistical model is the logistic regression model. Logistic regression is a predictive statistical modeling technique which applies multivariate regression to predict future land use based on historical land use changes, their spatial (change) characteristics and other potential drivers [15]. It is easier to model land use change using statistical modeling methods as their calibration is not so computationally intensive compared to rule-based models like cellular automata [19]. In logistic regression, social and economic factors like population density, accessibility to services, distance to commercial and industrial area, mean incomes, etc., can be incorporated in the model. Logistic regression analysis has been one of the most widely used approaches in the past two decades for predictive land use modeling by means of variation of inductive modeling [19].

Agent-based modeling (ABM), which is a forward-looking simulation technique which calculates "agents," each of which represents an actor and how they interact with their "environment" or the total system. The models represent real and imagined scenarios, which allow for the discovery of potentially emergent issues or phenomena. Such models have increasingly been used to analyze complex issues like land use change. Agents are independent entities which have set goals to achieve. The agents can be countries, landowners, land tenants, citizens, etc. ABM is used to simulate human behavior in cities, for example whereby policy makers, planners or citizens are entities (agents) which interact with the city environment and are capable of making urban planning decisions [16]. In a rule-based approach, their behavior is fixed, meaning that decision-making functions and algorithms remain unchanged (i.e., they always react in the same way when confronted with a particular situation). While agents react to changes in their spatial and social environment, they neither adapt their rules in response nor intelligently learn from previous experiences. ABM can be useful to study the changes in the land use and to evaluate the projections.

The various development environment used for agent-based modeling are Anylogic, Cormas, Cougaar (via OpenMap) Framsticks, Janus (using JaSIM), MASON, Repast, SeSAm, VisualBots, and NetLogo. These environments provide tools to develop an agent-based model and provide a platform to represent model components, control model function, and evaluate and visualize model output [20].

Hybrid approaches have been developed by integrating different ML methods which has resulted in better performance and assessment. An example of a hybrid method is the integration of logistic regression, Markov chain and cellular automata [21] to model urban expansion in the metropolitan area of Tehran, Iran. The results of the simulation were compared with the actual land use map and the result matched 89% between the simulated and the actual. Another example is integration of cellular automata Markov chain (CA-MC) with artificial neural network [22] to enhance the simulation capacity in predicting the changes in land use. The study integrates ANN and CA-MC to incorporate several driving forces (economic, spatial and environmental variables) that impact land use change. The integration and the influence of the driving forces improved the model prediction. Among the hybrid approaches Kamusoko and Gamba (2015) [23] tested random forest–cellular automata to study urban land change in Harare metropolitan province, Zimbabwe. Cellular automata was used to calculate multiple-step transition rates from land use/land cover maps (1984, 2002 and 2008). RF model was used to compute transition potential maps. The study then compared this model with SVM-CA and logistic regression (LR) and CA. The result showed that RF-CA outperformed SVM-CA and LR-CA models. Hybrid Urban Expansion Model (HUEM) was used by Mustafa and Cools (2018) [24] that integrates LR, CA, AB to simulate future urban development in Wallonia, Belgium. The urban expansion is simulated between 1990 and 2000. The calibration results are analyzed by comparing the projection for the 2000 simulated map with the actual 2000 land use map. The HEUM model uses three agent sets, developer agent, farmer agent and planning permission authority agent. The performance of HEUM is compared with other spatial expansion models, i.e., Logit model, CA model, CA-Logit model. The comparison shows that the performance of HEUM model is better than other models in terms of allocation ability.

References

1. Samardžić, M.; Kovačević, M.; Bajat, B.; Dragićević, S. Machine learning techniques for modelling short term land-use change. ISPRS Int. J. Geo-Inf. 2017, 6, 387.

2. Cao, C.; Dragićević, S.; Li, S. Land-use change detection with convolutional neural network methods. Environments 2019, 6, 25.

3. Hagenauer, J.; Omrani, H.; Helbich, M. Assessing the performance of 38 machine learning models: The case of land consumption rates in Bavaria, Germany. Int. J. Geogr. Inf. Sci. 2019, 33, 1399–1419.

4. Amler, B.; Betke, D.; Eger, H.; Ehrich, C.; Kohler, A.; Kutter, A.; von Lossau, A.; Müller, U.; Seidemann, S.; Steurer, R. Land Use Planning Methods, Strategies and Tools; Deutsche Gesellschaft Für Technische Zusammenarbeit (GTZ) GmbH: Eschborn, Germany, 1999.

5. Gidudu, A.; Hulley, G.; Marwala, T. Classification of images using Support Vector Machines. arXiv 2007, arXiv:0709.3967v1.

6. Pal, M.; Mather, P.M. Support vector machines for classification in remote sensing. Int. J. Remote Sens. 2005, 26, 1007–1011.

7. Ruiz Hernandez, I.E.; Shi, W. A Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analysis. Int. J. Remote Sens. 2018, 39, 1175–1198.

8. Duque, J.C.; Patino, J.E.; Betancourt, A. Exploring the potential of machine learning for automatic slum identification from VHR imagery. Remote Sens. 2017, 9, 895.

9. hd-pro, Applying Edge Detection to Feature Extraction and Pixel Integrity. (n.d.). Retrieved 13 April 2019. Available online: https://medium.com/hd-pro/applying-edge-detection-to-feature-extraction-and-pixel-integrity-2d39d9460842 (accessed on 16 January 2021).

10. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444.

11. Vali, A.; Comai, S.; Matteucci, M. Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. Remote Sens. 2020, 12, 2495.

12. Maithani, S. A neural network based urban growth model of an Indian city. J. Indian Soc. Remote Sens. 2009, 37, 363–376.

13. Tso, B.; Mather, P.M. Classification Methods for Remotely Sensed Data, 2nd ed.; CRC Press, LLC: Boca Raton, FL, USA, 2001; pp. 255–278.

14. Li, G.; Cai, Z.; Qian, Y. Identifying Urban Poverty Using High-Resolution. Satellite. Land 2021, 10, 648.

15. Nong, Y.; Du, Q. Urban growth pattern modeling using logistic regression. Geo-Spat. Inf. Sci. 2011, 14, 62–67.

16. Clarke, K.C. Cellular Automata and Agent-Based Models. In Handbook of Regional Science; Springer: Berlin/Heidelberg, Germany, 2015.

17. Espínola, M.; Ayala, R.; Leguizamón, S.; Menenti, M. Classification of Satellite Images Using the Cellular Automata Approach. In The Open Knowlege Society; A Computer Science and Information Systems Manifesto: Berlin/Heidelberg, Germany, 2008; pp. 521–526.

18. Berberoğlu, S.; Akin, A.; Clarke, K.C. Cellular automata modeling approaches to forecast urban growth for adana, Turkey: A comparative approach. Landsc. Urban Plan. 2016, 153, 11–27.

19. Hu, Z.; Lo, C.P. Modeling urban growth in Atlanta using logistic regression. Comput. Environ. Urban Syst. 2007, 31, 667–688.

20. National research council. Advancing Land Change Modeling; The National Academic Press: Washington, DC, USA, 2014; pp. 93–96.

21. Arsanjani, J.J.; Helbich, M.; Kainz, W.; Boloorani, A.D. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. Int. J. Appl. Earth Obs. Geoinf. 2012, 21, 265–275.

22. Gharaibeh, A.; Shaamala, A.; Obeidat, R.; Al-kofahi, S. Heliyon Improving land-use change modeling by integrating ANN with Cellular Automata-Markov Chain model. Heliyon 2020, 6, e05092.

23. Kamusoko, C.; Gamba, J. Simulating Urban Growth Using a Random Forest-Cellular Automata (RF-CA) Model. ISPRS Int. J. Geo-Inf. 2015, 4, 447–470.

24. Mustafa, A.; Cools, M.; Saadi, I.; Teller, J. Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM). Land Use Policy 2017, 69, 529–540.