A MACHINE LEARNING MODEL TO PREDICT URBAN SPRAWL USING OFFICIAL LAND-USE DATA

Mohamed NOBY
Teaching assistant, B.Sc., Aswan University, Faculty of Engineering, Architectural Department, Egypt, e-mail: mohamednoby@eng.aswu.edu.eg

Mohamed E. ELATTAR
Professor, PhD, The British University in Egypt, Faculty of Engineering, Architectural Department, Egypt, e-mail: mohamed.elattar@bue.edu.eg

Omar HAMDY
Assistant Professor, PhD, Aswan University, Faculty of Engineering, Architectural Department, Egypt, e-mail: omar.hamdy@aswu.edu.eg

Abstract. The rate of global urbanization is constantly increasing. As a result of the massive population growth, there is an increasing demand for further urban development, especially in developing regions such as Aswan city. This paper aims to examine the usage official land-use data in predicting future urban growth until 2046, moreover, to define urban driving forces in case study area. This was done using Similarity weighted model, a machine learning based model to simulate future urban growth. The results show that official land-use data produce a slightly better results' accuracy than remote sensing sources within small to medium scales. The results although reveal that for study region, urban area is expected to expand to cover an area of almost 4460 Feddan by year 2046. The outcome of this research assesses decision makers to accurately predict future urban sprawl areas using available official land-use data.

Key words: urban growth, machine learning, similarity weight, Aswan.

1. Introduction

The global urbanization rate is rapidly rising, associated with a huge population expansion, therefore, there is a rising need for more urban development (Rahman *et al.*, 2010; Ujoh and Ifatimehin, 2010). However, it is believed that in the last century, urban areas increase with a much greater rate than population growth worldwide, especially in less developed countries where growth is frequently erratic and unplanned (Brockerhoff, 1999; Cohen, 2004; Wu *et al.*, 2016). In Egypt, because of the rapid population increase, urban areas tend to spread in all directions. This is especially true in heavily populated regions like Aswan governorate (Hamdy *et al.*, 2017; Mostafa *et al.*, 2021).

Due to the current rapid dynamic expansion, megacities' policymakers face unprecedented challenges in governance, urban planning, and land-use management (Osman *et al.*, 2016). As a result, understanding past, current, and future growth plays a significant part in decision-making (Hamdy *et al.*, 2016). Therefore, spatial modelling has become a crucial process in urban planning globally (Noby *et al.*, 2022). Spatial modelling can be utilized to highlight urban expansion and land changes, which can then be used to reveal information about urban ecosystem management and to track, analyze, and predict potential urban expansion (Nugroho and Al-Sanjary, 2018).

Geographic Information Systems (GIS) play a significant role in urban development prediction and spatial data management (Hamdy *et al.*, 2020). Various GIS based models and approaches are often used to simulate urban development and land-use change. With such models, researchers may easily define and monitor urban expansion as well as determining urban driving factors. There are several techniques used to simulate urban expansion, such as the Cellular Automata - Markov chain model (Arsanjani *et al.*, 2011; Mitsova *et al.*, 2011; Sang *et al.*, 2011), the Logistic Regression model (LR) (Hu and Lo, 2007; Mostafa *et al.*, 2021; Nong and Du, 2011), Similarity Weighted model (Bununu, 2017; Mirakhorlo and Rahimzadegan, 2018; Zubair *et al.*, 2017), the artificial neural networks (ANNs) models (Chettry and Surawar, 2021; Mansour *et al.*, 2023), and the SLEUTH model (Caglioni *et al.*, 2006; Rafiee *et al.*, 2009). Similarity Weighted model is based on machine learning procedures that have been shown to be the optimum choice for non-complex simulations (Nugroho and Al-Sanjary, 2018; Sangermano *et al.*, 2010).

Many works analyzed the historical urban growth in different cities worldwide, while others focused on predicting the future urban expansion. In Egypt, number of researchers studied urban growth within Egyptian governorates (Belal and Moghanm, 2011; Hamdy and Zhao, 2016; Hegazy and Kaloop, 2015; Mahmoud *et al.*, 2019; Mohamed, 2012), however, only few papers focus on Aswan governorate (Hamdy *et al.*, 2014, 2016). Moreover, the lack of updated governmental official planning data forces many studies to use remotely sensed data for land-use mapping. However, for small to medium scale, the mapping process requires a high level of accuracy which is almost impossible to be obtained using remotely sensed data (Fuller *et al.*, 2003). Such a problem can be solved using official governmental land-use maps which present a detailed historical urban data.

The main objective of this paper is to examine the usage of available official land-use governmental data to analyze the historical urban expansion and simulate the predicted future urban sprawl in Aswan city. The used methodology depends on machine learning based similarity weighted model, which is used to predict future urban growth in case study area for years 2026 and 2046, moreover, defining urban driving forces of urban growth with the aid of GIS software to process the available official land-uses data.

2. Case study

The study location is in the southern Egyptian city of Aswan. It is located between longitudes 32° 51'E to 32° 55'E and the latitudes 24° 1'N to 24° 7'N. The research region area is about 17.5 thousand Feddan, located on the Nile River and is limited to the east by mountainous area, and from the west with agricultural land. Most urban areas are found on the east bank of the Nile River as illustrated in Fig. 1. The study period of the urban growth history begins from 1986 and extends to 2006 with time interval of 10 years due to the availability of trusted official land-use data for these years.

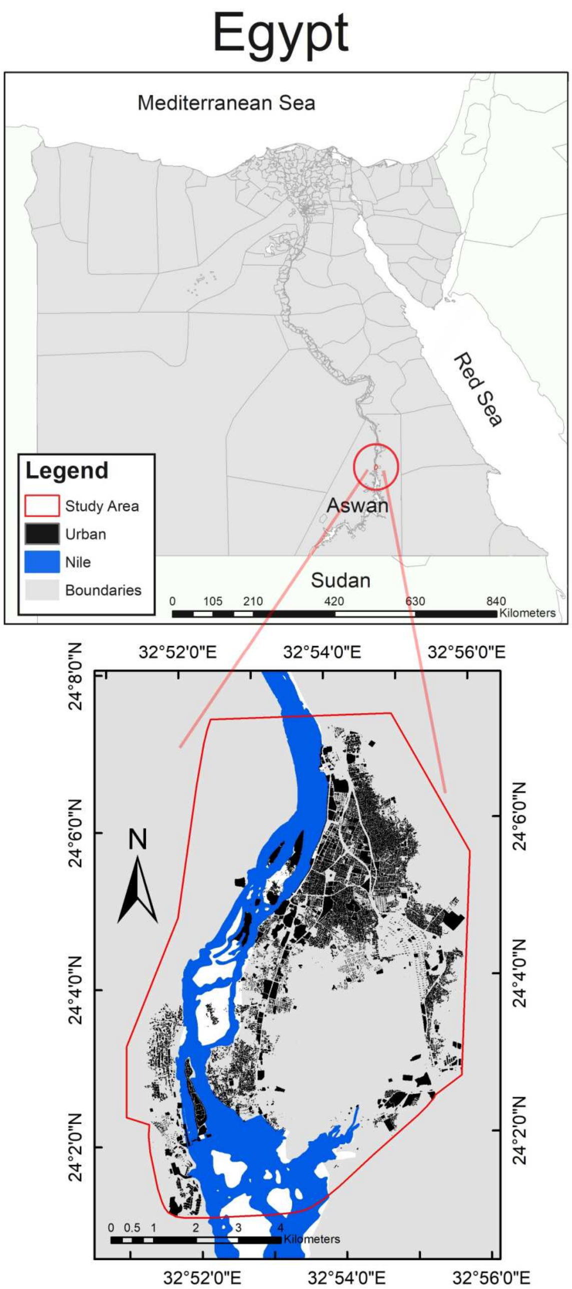


Fig. 1. The case study area of Aswan city.

3. Methodology

The research methodology in this paper applied Land Change Modeler (LCM) to predict future urban growth, which offers an applicable machine learning based techniques for urban growth prediction compared with traditional techniques. Accordingly, historical urban data were obtained from the official land-use data for study years, created by General Organization of Physical Planning (GOPP). Land-uses were classified into 2 classes: - i) urban; and ii) non-urban, which are the main classes needed to detect historical urban expansion. The obtained vector data were converted to raster using GIS software, then clipped to study area boundaries.

It is necessary to monitor urban areas throughout time in order to discover growth tendencies. Similarity weighted model integrated in LCM was used to analyze changes happened between years 1986 and 1996, then transition probability matrix was computed to define the probability of change (Hamdy *et al*., 2016; Losiri *et al*., 2016). The second phase was to identify the urban driving forces based on an assessment of relevant studies (Achmad *et al*., 2015; Hamdy *et al*., 2017; Hu and Lo, 2007). Four factors were used, which are (i) distances from roads; (ii) distance from services; (iii) Distance from Nile; and (iv) Digital Elevation Model (DEM), these factors were checked to be of an accepted Cramer's V value of more than 0.15. The model examined the weight of each driving factor to finally predict the urban area of year 2006. The following step was to validate the accuracy of the predicted map using the validate module as well as the Relevant Operational Characteristic (ROC). Finally, the model was used to predict further urban growth maps for years 2026 and 2046. The overall methodology is shown in Fig. 2.

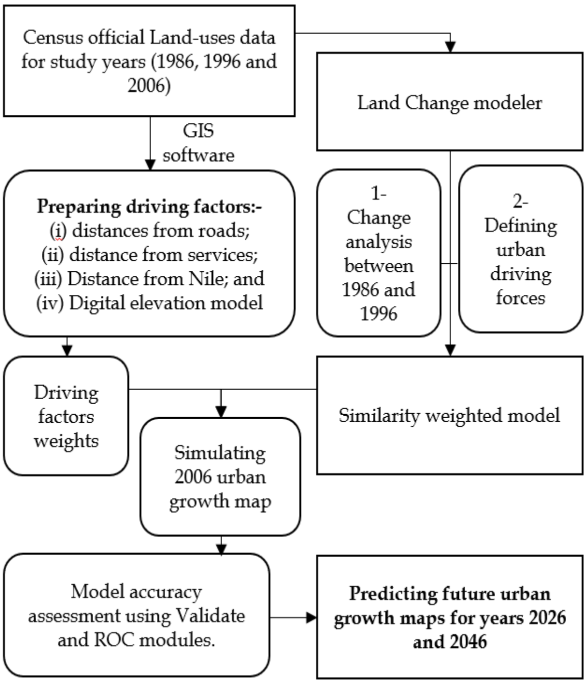


Fig. 2. The overall methodological framework.

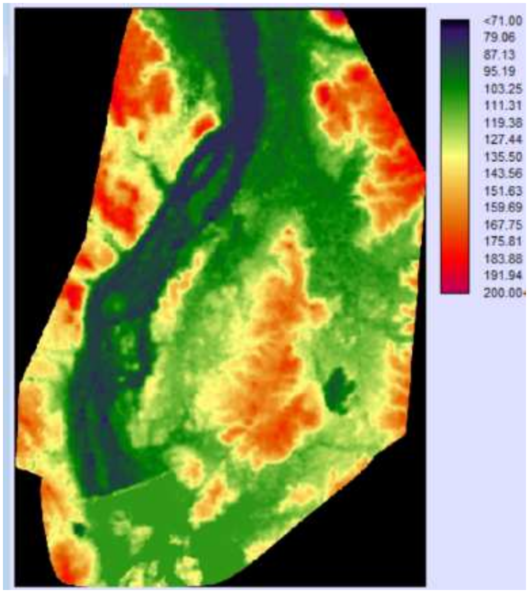
4. Results

The analysis of changes in urban area for the period 1986 to 1996 shows that urban area in study area has increased by about 468 Feddan during the period 1986 to 1996. Four urban driving factors were used for future urban prediction as shown in Fig. 3, where each variable was evaluated using Cramer's V value to detect the most effective urban driving factor. Distance to services variable was found to be the most effective factor in urban growth with Cramer's V value of 0.93, while DEM had the least effect as illustrated in Table 1.

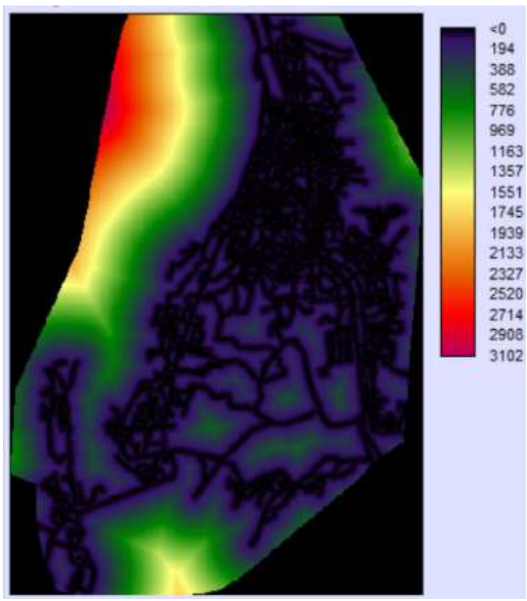
Table 1. Test and Selection of Site and Driver Variables.

Driver Variables	Cramer's V
DEM	0.1377
Distance to Nile	0.1505
Distance to Roads	0.3302
Distance to Services	0.3953

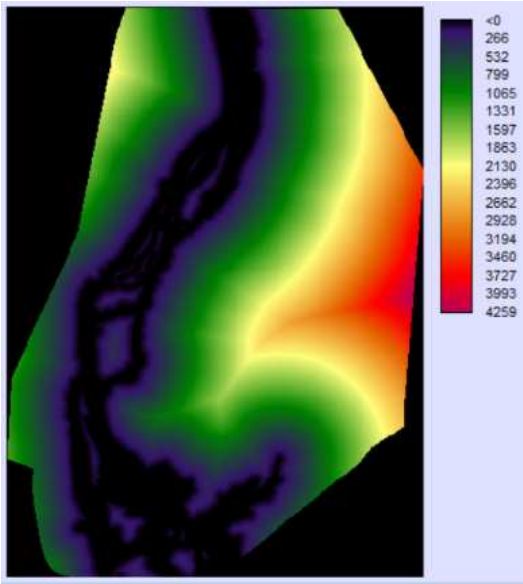
Over the 20-year historical urban expansion period, the urban area in our study zone grew gradually. The urban area in 1986 was 1859 Feddan, which expanded to 2770 Feddan in 2006 as shown in Fig. 4. For the next 20-year simulation period, urban area is predicted to increase to 3690 Feddan in 2026, and finally to 4459 Feddan in 2046, as illustrated in Fig. 4. Table 2 shows the percentage of urbanization for the study period, the highest growth rate is predicted to happen before 2026 with urban increasing percentage of about 33%, however, the growth rate becomes 21% for the period 2026 to 2046 as illustrated in Fig. 5.



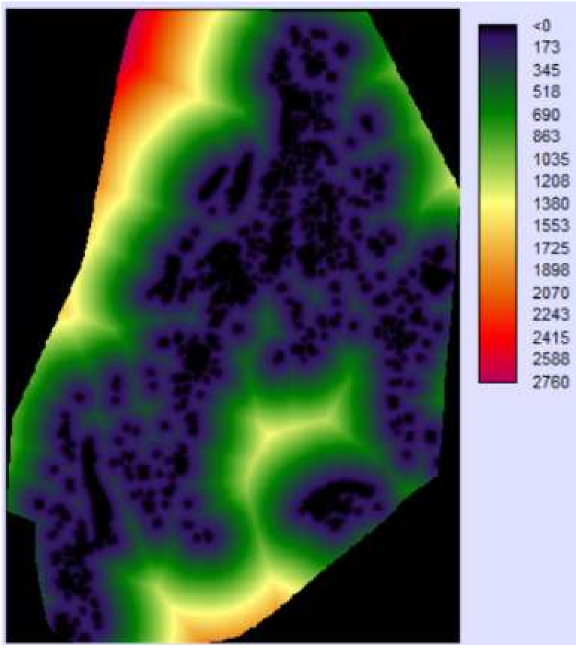
a. DEM



b. Distance from road



c. Distance from Nile

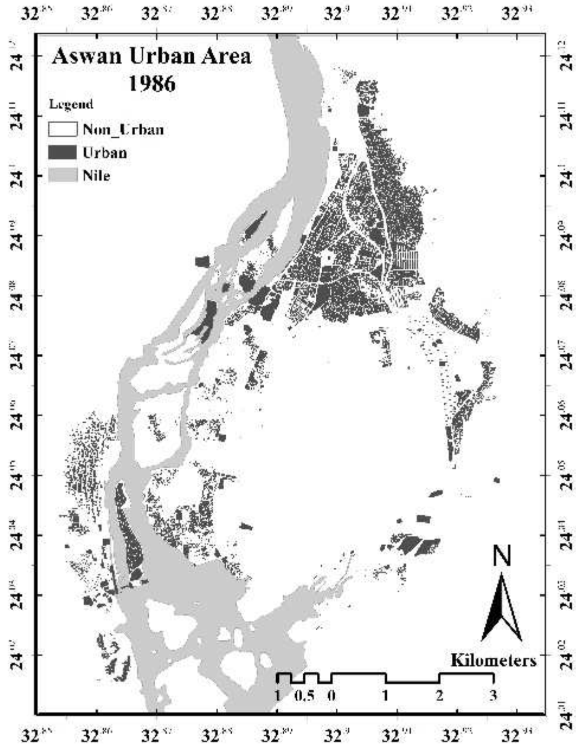


d. Distance from services

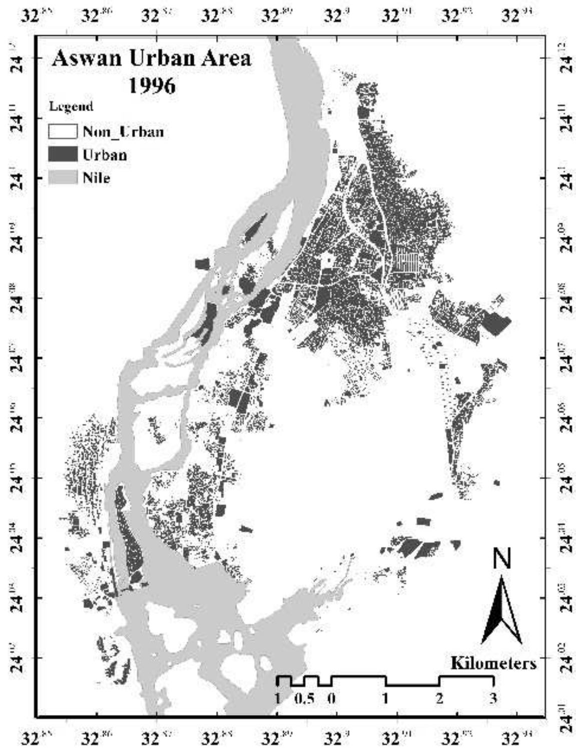
Fig. 3. Urban driving factors used in urban growth prediction model.

Table 2. Urban growth for the study period.

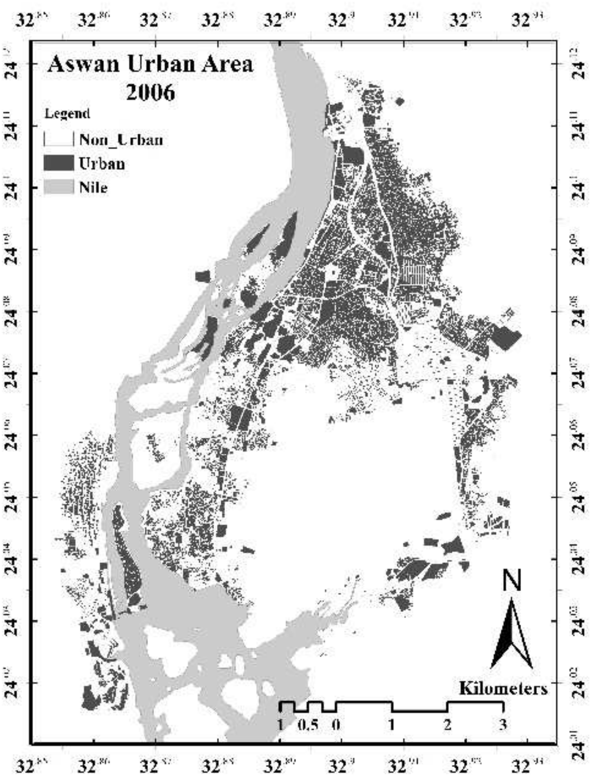
Year	Area (Feddan)	Urban Increase (Feddan)	Urban Increase Percentage (%)
2046	4459	769	20.85%
2026	3690	920	33.20%
2006	2770	443	19.02%
1996	2327	468	25.19%
1986	1859	0	0
Total	**15105**	——	——



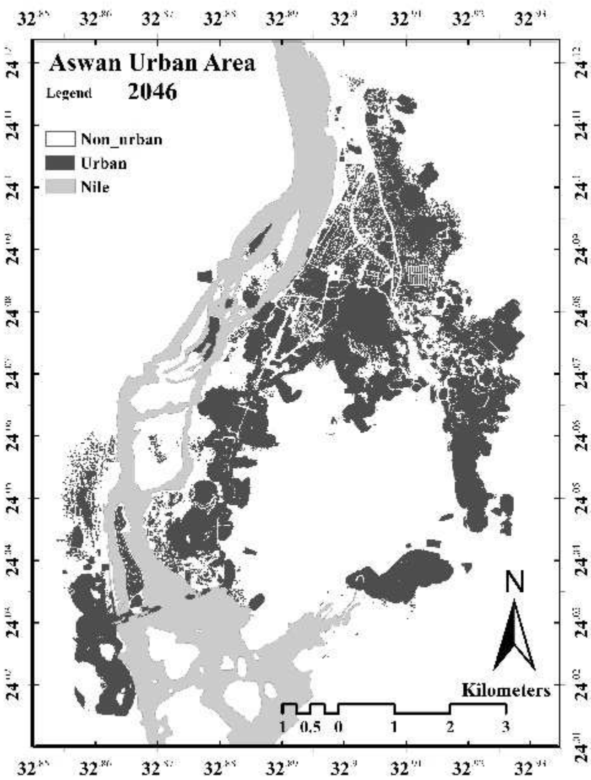
a. Urban areas for 1986



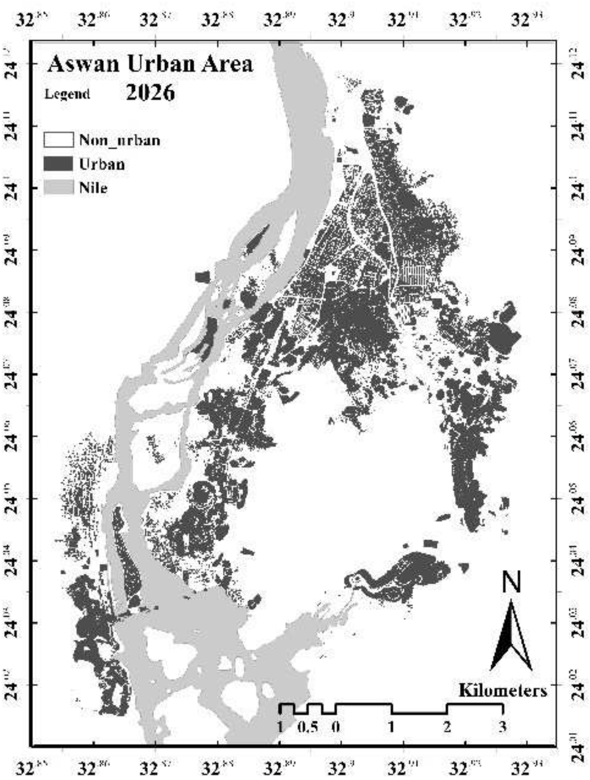
B. Urban areas for 1996



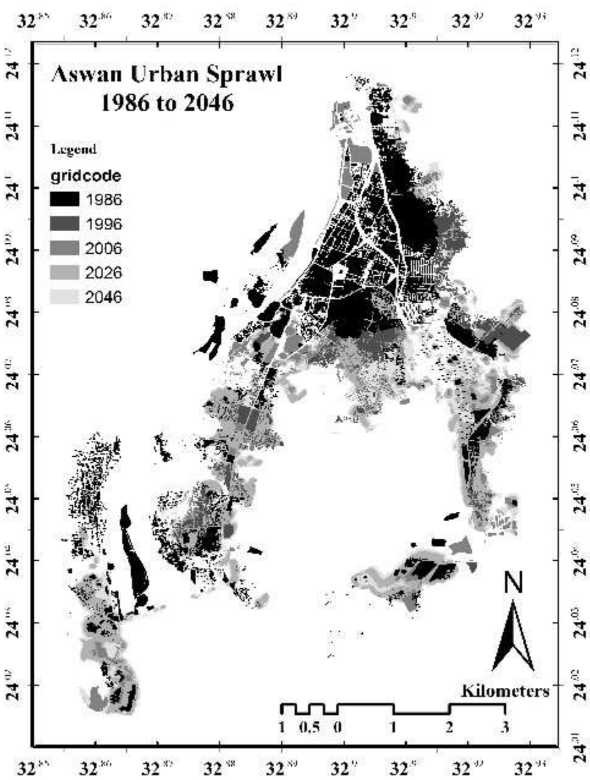
C. Urban areas for 2006



e. Future urban growth prediction for year 2046

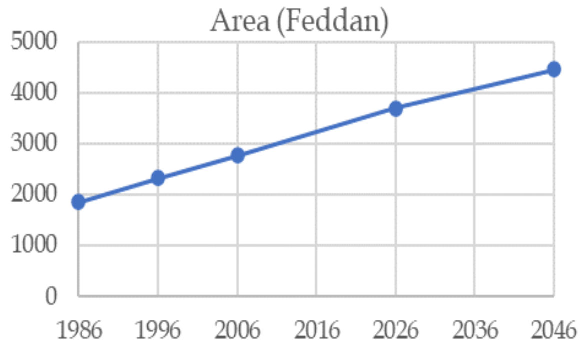


d. Future urban growth prediction for year 2026

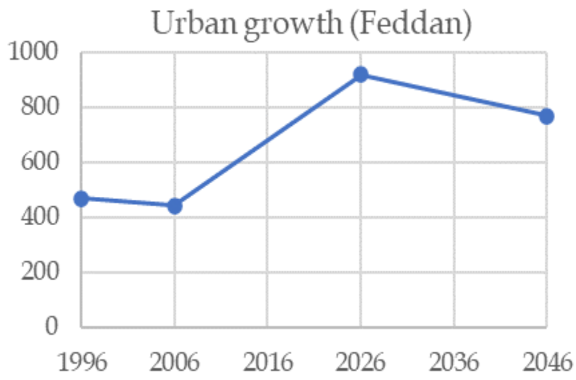


f. Overall urban sprawl map for the study period 1986 to 2046

Fig. 4. Urban growth maps for study period.



a. Urban areas for the period 1986 to 2046.



b. Urban growth for each interval.

Fig. 5. Urban growth trend in Aswan.

Based on the Kappa spatial correlation statistic, both the relative operating characteristic (ROC) and the validate module were used to examine the amount of agreement between the observed 2006 land-use map and the simulated one. All statistics for the validate module were substantially above 0.90 as shown in Table 3, while the ROC value stands at 0.85. These values indicate that the two datasets agreed extremely well; hence, the transition probability matrix may be utilized to forecast the distribution pattern of land-use datasets in the research region.

Table 3. The validation results for the simulated map

K. No	0.9504
K. location	0.9364
K. Standard	0.9358
K. location Strata	0.9364
ROC	0.8575

5. Discussions

The results reveal that urban expansion in Aswan city is nearly linear and is limited to few directions due to the city topography, setting urban areas is surrounded from the east with high mountainous area, while by the west is the Nile. Moreover, urban areas have increased by around 900 Feddan in 20-year period, expanding from 1859 Feddan in 1986 to around 2700 Feddan in 2006. However, urban areas are expected to further expand to cover an area of almost 4460 Feddan by year 2046. This urban expansion is mostly influenced by the distance to facilities, particularly health and educational services which attract urban areas in case study region.

The accuracy validation of urban sprawl prediction results shows a slightly higher values for kappa coefficients compared to other studies applied on an area of small to medium scale and depending on remotely sensed data (Azizi *et al.*, 2022; Baqa *et al.*, 2021; Shrestha *et al.*, 2023). This may be because the land-use classification technique greatly relies on both user and procedure accuracies which in most cases lower the overall accuracy of the classified land-use maps compared to the official maps. However, the data availability of consecutive land-use maps crucial for urban expansion monitoring may limit most studies to use remotely sensed data.

Urban sprawl is usually uncontrolled expansion that is driven by a variety of causes rather than planning regulations (Hamma and Petrişor, 2018). Population growth, damaged infrastructure, lack of open spaces, social problems, lack of affordable housing, and internal migration are all thought to be the major drivers of urban sprawl globally (Gouda *et al.*, 2016; Habibi and Asadi, 2011;

Nguyen *et al.*, 2022; van Vliet, 2019). However, in Egypt, political aspects and construction laws have a significant influence on urban sprawl. Moreover, the rapid population growth along with relatively low average per capita income are the primary causes that encourage individuals to dwell in informal settlements in Egypt (Gouda *et al.*, 2016; Shalaby *et al.*, 2012). Integrating the causative drivers of urban sprawl into urban planning process would be a suitable choice for managing urban growth and minimizing the uncontrolled urban sprawl.

6. Conclusions

This paper aim was to test the usage of official land-uses data in a machine learning based urban growth modeler, also to identify the influence of urban driving factors in case study area. Overall results show that although most relevant researches relies on remotely sensed data for urban sprawl monitoring (Baqa *et al.*, 2021; Hamdy *et al.*, 2016; Khawaldah *et al.*, 2020; Kisamba and Li, 2022), our study results confirm that for small to medium scale regions, it is more accurate to use official land-use data when available for a higher results accuracy, while remote sensing provide a suitable accuracy for urban sprawl prediction for regional scale.

Identifying the variables that influence urban sprawl is a highly effective approach to understand urban expansion and assisting decision-makers and planners in predicting how urban is expected to expand in the future. Some variables have influenced urban expansion in Aswan city. However, the most effective variable was the distance to services followed by the distance from roads. These results reveal that modifying such factors could help controlling urban sprawl direction. The outcome of this research may assess decision makers to accurately predict future urban sprawl areas using available official land-use data. However, more work may be needed to assess urban expanding towards natural hazard zones, particularly flush flood and landslide hazardous zones.

REFERENCES

Achmad A., Hasyim S., Dahlan B., Aulia D. N. (2015), *Modeling of urban growth in tsunami-prone city using logistic regression: Analysis of Banda Aceh, Indonesia,* Applied Geography **62**: 237–246.

Arsanjani J. J., Kainz W., Mousivand A. J. (2011), *Tracking dynamic land-use change using spatially explicit Markov Chain based on cellular automata: the case of Tehran,* International Journal of Image and Data Fusion **2(4):** 329–345.

Azizi P., Soltani A., Bagheri F., Sharifi S., Mikaeili M. (2022), *An Integrated Modelling Approach to Urban Growth and Land Use/Cover Change,* Land **11:** 1715.

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.

Belal A. A., Moghanm F. S. (2011), *Detecting urban growth using remote sensing and GIS techniques in Al Gharbiya governorate, Egypt,* The Egyptian Journal of Remote Sensing and Space Science **14(2):** 73–79.

Brockerhoff M. (1999), *Urban growth in developing countries: a review of projections and predictions,* Population and Development Review **25(4):** 757–778.

Bununu Y. A. (2017), *Integration of Markov chain analysis and similarity-weighted instance-based machine learning algorithm (SimWeight) to simulate urban expansion,* International Journal of Urban Sciences **21(2):** 217–237.

Caglioni M., Pelizzoni M., Rabino G. A. (2006), *Urban sprawl: A case study for project gigalopolis using SLEUTH model,* in: El Yacoubi S., Chopard B., Bandini S. (Eds.), Proceedings of the 7th International Conference on Cellular Automata for Research and Industry, 20-23 September

2006, Perpignan, France, Computer Science, Berlin, Germany, pp. 436–445.

Chettry V., Surawar M. (2021), *Delineating urban growth boundary using remote sensing, ANN-MLP and CA model: a case study of Thiruvananthapuram urban agglomeration, India,* Journal of the Indian Society of Remote Sensing **49(10):** 2437–2450.

Cohen B. (2004), *Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts,* World Development **32(1):** 23–51.

Fuller R. M., Smith G. M., Devereux B. J. (2003), *The characterisation and measurement of land cover change through remote sensing: problems in operational applications,* International Journal of Applied Earth Observation and Geoinformation **4(3):** 243–253.

Gouda A.A., Hosseini M., Masoumi H. E. (2016), *The Status of Urban and Suburban Sprawl in Egypt and Iran,* GeoScape **10:** 1–15.

Habibi S., Asadi N. (2011), *Causes, Results and Methods of Controlling Urban Sprawl,* Procedia Eng **21:** 133–141.

Hamdy O., Zhao S. (2016), *A Study on Urban Growth in Torrent Risk Areas in Aswan, Egypt,* Journal of Architecture and Planning **81:** 1733–1741.

Hamdy O., Zhao S., El-atty H. A., Ragab A., Salem M. (2020), *Urban Areas Management in Developing Countries : Analysis the Urban Areas Crossed with Risk of Storm Water Drains, Aswan-Egypt,* International Journal of Urban and Civil Engineering **14(3):** 96–102.

Hamdy O., Zhao S., Osman T., Salheen M. A., Eid Y. Y. (2016), *Applying a hybrid model of Markov chain and logistic regression to identify future urban sprawl in Abouelreesh, Aswan: A case study,* Geosciences **6(4):** 43.

Hamdy O., Zhao S., Salheen M. A., Eid Y. Y. (2017), *Analyses the driving forces for urban growth by using IDRISI® Selva models Abouelreesh-Aswan as a case study,* International Journal of Engineering and Technology **9(3):** 226.

Hamdy O., Zhao S., Salheen M., Eid Y. (2014), *Using Arc GIS to analyse urban growth towards torrent risk areas (Aswan city as a case study),* Earth and Environmental Science **20:** 12009.

Hamma W., Petrişor A.-I. (2018), *Urbanization and risks: case of Bejaia city in Algeria,* Human Geographies **12:** 97–114.

Hegazy I. R., Kaloop M. R. (2015), *Monitoring urban growth and land use change detection with GIS and remote sensing techniques in Daqahlia governorate Egypt,* International Journal of Sustainable Built Environment **4(1):** 117–124.

Hu Z., Lo C. P. (2007), *Modeling urban growth in Atlanta using logistic regression,* Computers, Environment and Urban Systems **31(6):** 667–688.

Khawaldah H. A., Farhan I., Alzboun N. M. (2020), *Simulation and prediction of land use and land cover change using GIS, remote sensing and CA-Markov model.* Global Journal of Environmental Science and Management **6(2):** 215–232.

Kisamba F. C., Li F. (2022), *Analysis and modelling urban growth of Dodoma urban district in Tanzania using an integrated CA–Markov model.* GeoJournal **88:** 511–532.

Losiri C., Nagai M., Ninsawat S., Shrestha R. P. (2016), *Modeling urban expansion in Bangkok metropolitan region using demographic–economic data through cellular automata-Markov chain and multi-layer perceptron-Markov chain models,* Sustainability **8(7):** 686.

Mahmoud H., Alfons R., M. Reffat R. (2019), *Analysis of The Driving Forces of Urban Expansion in Luxor City by Remote Sensing Monitoring,* International Journal of Integrated Engineering **11(6):** 296–307.

Mansour S., Ghoneim E., El-Kersh A., Said S., Abdelnaby S. (2023), *Spatiotemporal Monitoring of Urban Sprawl in a Coastal City Using GIS-Based Markov Chain and Artificial Neural Network (ANN),* Remote Sensing **15(3):** 601.

Mirakhorlo M. S., Rahimzadegan M. (2018), *Integration of SimWeight and Markov Chain to Predict Land Use of Lavasanat Basin TT,* Kntu-Nmce **2(4):** 1–9.

Mitsova D., Shuster W., Wang X. (2011), *A cellular automata model of land cover change to integrate urban growth with open space conservation,* Landscape and Urban Planning **99(2):** 141–153.

Mohamed E. (2012), *Analysis of urban growth at Cairo, Egypt using remote sensing and GIS,* Natural Science **4:** 355-361

Mostafa E., Li X., Sadek M., Dossou J. F. (2021), *Monitoring and Forecasting of Urban Expansion Using Machine Learning-Based Techniques and Remotely Sensed Data: A Case Study of Gharbia Governorate, Egypt,* Remote Sensing **13(22):** 4498

Nguyen H. D., Dang D.K., Nguyen Q.-H., Bui Q.-T., Petrişor A.-I. (2022), *Evaluating the effects of climate and land use change on the future flood susceptibility in the central region of Vietnam by integrating land change modeler,* *machine learning methods,* Geocarto International **37:** 12810–12845.

Noby M., Michitaka U., Hamdy O. (2022), *Urban Risk Assessments: Framework for Identifying Land-uses Exposure of Coastal Cities to Sea Level Rise, a Case Study of Alexandria,* SVU-International Journal of Engineering Sciences and Applications **3(1):** 78–90.

Nong Y., Du Q. (2011), *Urban growth pattern modeling using logistic regression,* Geo-Spatial Information Science **14(1):** 62–67.

Nugroho F., Al-Sanjary O. (2018), *A Review of Simulation Urban Growth Model,* International Journal of Engineering and Technology (UAE) **7(4.11):** 17-23.

Osman T., Divigalpitiya P., Osman M. M., Kenawy E., Salem M., Hamdy O. (2016), *Quantifying the Relationship between the Built Environment Attributes and Urban Sustainability Potentials for Housing Areas,* Buildings **6(3):** 39

Rafiee R., Mahiny A. S., Khorasani N., Darvishsefat A. A., Danekar A. (2009), *Simulating urban growth in Mashad City, Iran through the SLEUTH model (UGM),* Cities **26(1):** 19–26.

Rahman A., Aggarwal S. P., Netzband M., Fazal S. (2010), *Monitoring urban sprawl using remote sensing and GIS techniques of a fast growing urban centre, India,* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **4(1):** 56–64.

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):** 938–943.

Sangermano F., Eastman J. R., Zhu H. (2010), *Similarity weighted instance-based learning for the generation of transition potentials in land use change modeling,* Transactions in GIS **14(5):** 569–580.

Shalaby A.A., Ali R. R., Gad A. (2012), *Urban sprawl impact assessment on the agricultural land in Egypt using remote sensing and GIS: a case study, Qalubiya Governorate,* Land Use Science **7:** 261–273.

Shrestha M., Mitra C., Rahman M., Marzen L. (2023), *Mapping and Predicting Land Cover Changes of Small and Medium Size Cities in Alabama Using Machine Learning Techniques,* Remote Sensing **15(1):** 106.

Ujoh F., Ifatimehin O. O. (2010), *Understanding urban sprawl in the Federal Capital City, Abuja: Towards sustainable urbanization in Nigeria,* Journal of Geography and Regional Planning **2(5):** 106.

van Vliet J. (2019), *Direct and indirect loss of natural area from urban expansion,* Nature Sustainability **2:** 755–763.

Wu Y., Li S., Yu S. (2016), *Monitoring urban expansion and its effects on land use and land cover changes in Guangzhou city, China,* Environmental Monitoring and Assessment **188(1):** 54.

Zubair O. A., Ji W., Weilert T. E. (2017), *Modeling the impact of urban landscape change on urban wetlands using similarity weighted instance-based machine learning and Markov model,* Sustainability **9(12):** 2223.

Received: 8 March 2023 • **Revised**: 30 March 2023 • **Accepted**: 9 April 2023