An Automated Approach based on Machine Learning for Tracking Urban Expansion: Case of Study in Gharbia Governorate, Egypt

Wael Ahmed[1, 2*], Islam Atef [3], Moustafa Baraka [4], Adel ElShazly [1],Ahmad M. Senousi [1, 2],Walid Darwish[1, 2]

[1] Public Works Department, Faculty of Engineering, Cairo University,
Giza, Egypt, 12613.(walid.darwish & wael.mohamed) @cu.edu.eg
[2] NAMAA for Engineering Consultations, Dokki , Giza, Egypt
[3] Department of Civil Engineering, King Fahd University of Petroleum
Minerals, Dhahran 31261, Saudi Arabia
[4] Civil Engineering Program, German University in Cairo 11835, Egypt

Keywords: Random Forest, LULC, Gharbia, Change Detection, Spectral Indices.

Abstract

Addressing the United Nations Sustainable Development Goals, particularly sustainable cities and communities (SDG 11), and the protection of terrestrial ecosystems (SDG 15), is closely linked to understanding patterns of urbanization. Rapid urban growth significantly influences ecosystem functions, including transportation, housing, and economic development. Monitoring this growth and analyzing performance patterns are essential for supporting decision-making and guiding urban planning and management. This study presents an automatic approach for monitoring urban expansion by applying the Random Forest machine learning classifier from 2015 to 2025 using Google Earth Engine. The method exploits spectral indices not only for unsupervised classification but also for training the Random Forest classifier, thereby ensuring a fully automated workflow. The proposed approach is applied to Gharbia Governorate, a region which lacks surrounding desert margins and is instead entirely composed of fertile agricultural land, to monitor urban expansion in three-year intervals. The proposed study, which achieved a kappa coefficient exceeding 0.96 across all study periods, revealed a gradual decline in agricultural land from 75.5% in 2015 to 72.7% in 2025. These outcomes offer valuable insights to support evidence-based planning and promote sustainable land use management.

1. Introduction

Urban extent has expanded more than forty-fold since 1870, highlighting two centuries of urban sprawl and its long-term environmental implications (Li et al., 2021). Urban growth exerts adverse effects on systems such as transportation, housing, employment, and privacy, with its impacts being most severe in developing countries where expansion is often rapid and unplanned (Mostafa et al., 2021). The study of land use and land cover (LULC) change is crucial for understanding current land utilization and providing a foundation for effective planning and management. LULC has emerged as a key element in contemporary approaches to monitoring environmental change (Atef et al., 2024) and managing natural resources (Teck et al., 2023), particularly as human activity and expansion have altered or modified nearly 75% of natural landscapes worldwide (Hussain et al., 2024; Brown and Anand, 2024).

Consequently, multi-use or mixed landscapes that integrate natural, urban, and semi-urban land-use types are becoming increasingly widespread, reflecting the complex dynamics of human–environment interactions (Roy et al., 2025). Green spaces, particularly forests and agricultural lands, constitute essential assets for sustaining environmental balance, conserving biodiversity, and supporting economic development; therefore, the systematic monitoring of their spatiotemporal dynamics through Remote Sensing and Geographic Information Systems (GIS) is of critical importance (Atef et al., 2024; Valjarević et al., 2025).

Remote sensing combined with GIS provides an effective framework for detecting and quantifying land use changes, delivering timely, accurate, and cost-efficient insights into spatial and temporal dynamics (Wang et al., 2024). Variety of remote sensing techniques have been applied to generate accurate LULC maps which are essential for tracking and analyzing transitions. Those techniques are classified into traditional statistical approaches such as the Maximum Likelihood Classifier (MLC) and advanced machine learning methods like neural networks, decision trees, random forests (RF), and support vector machines (SVM) (Atef et al., 2023).

Random Forest (RF) and Classification and Regression Trees (CART) were applied and evaluated in the western Oregon region (USA), which comprised five land cover classes. The results indicated that RF outperformed CART in producing more accurate (LULC) maps (Gülci et al., 2025). Support Vector Machines (SVM), Random Forest (RF), and Classification and Regression Trees (CART) were analyzed and their performance evaluated through accuracy assessments. The findings revealed that RF classifiers outperformed both SVM and CART classifiers in terms of classification accuracy (Loukika et al., 2021).

In recent years, the Google Earth Engine (GEE) platform has been increasingly employed for spatial mapping and analysis across a wide range of research domains, including the monitoring of urban expansion. This issue is particularly critical in Egypt's Gharbia Governorate, which lacks surrounding desert margins and is instead entirely composed of fertile agricultural land. The challenge of urban sprawl in Egypt is further exacerbated by the fact that approximately 96% of the national territory consists of uninhabited desert, thereby placing disproportionate pressure on the limited arable areas (Mostafa et al., 2021). This research employs Landsat imagery spanning the period 2015–2025 to examine the spatiotemporal dynamics of land use and land cover (LULC) in Gharbia Governorate, a rep-

* Corresponding author

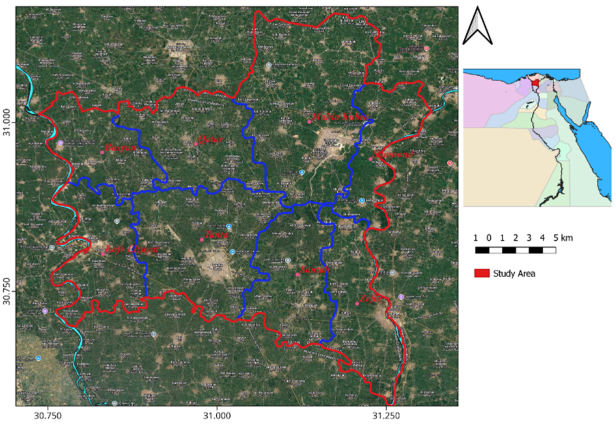


Figure 1. Gharbia Governorate Study Area

resentative region of the Nile Delta. Land cover classification was performed using the Random Forest algorithm within a GIS framework, targeting three dominant categories: agricultural land, urban areas, and water bodies. The results aim to support evidence-based land management strategies and highlight the pressing need for sustainable urban planning in agriculturally productive areas of the Egyptian Delta.

2. Study Area and Material

Gharbia Governorate is situated at the core of Egypt’s Nile Delta, positioned at approximately 30.87° N latitude and 31.03° E longitude. It is bounded by Kafr El-Sheikh Governorate to the north, Menoufia Governorate to the south, and flanked to the east and west by the Damietta and Rosetta branches of the Nile, respectively. Administratively, Gharbia comprises eight districts, covering a total area of about 1,943.27 km² (equivalent to 462,684 feddans), as illustrated in Figure 1. According to the Central Agency for Public Mobilization and Statistics (CAPMAS) (Gharbia-Governorate, 2025), the population of Gharbia reached 5,379,417 in July 2022. In comparison, earlier records indicate populations of 5,066,000 in 2018, 4,011,320 in 2006, and 3,790,670 in 2001, reflecting a steady demographic growth over the past two decades (Mostafa et al., 2023). Landsat 8 imagery from the years 2015, 2018, 2021, and 2025 were selected for analysis. To ensure consistency and enhance the accuracy of change detection, images acquired in May—characterized by minimal cloud cover—were used for all study years. GEE code editor and its provider services were used as a calculation platform.

3. Proposed Methodology

The monitoring of LULC changes in Gharbia Governorate was implemented on GEE platform following the workflow presented in Figure 2. The methodology consists of three main stages: (I) automatic selection of training samples, (II) land cover classification using the Random Forest (RF) algorithm, and (III) Mapping & Change Detection. The procedures corresponding to each stage are described in detail in the following sections.

3.1 Automatic Selection of Training Samples

To ensure full automation of the classification process and consider that the selection of training samples is a critical factor in Random Forest (RF) classification, three spectral indices were

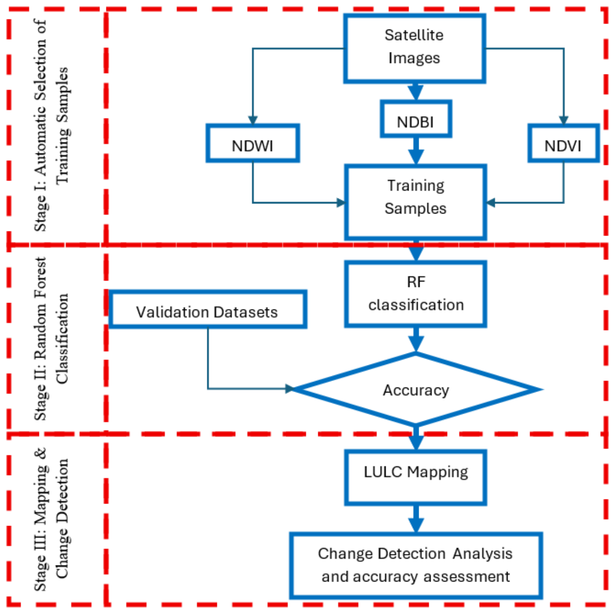


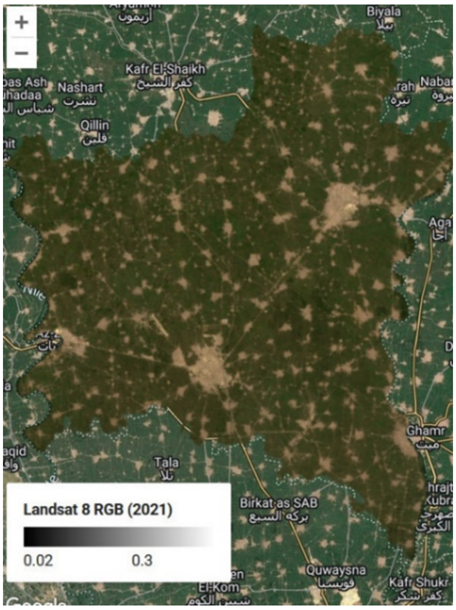
Figure 2. Proposed Flowchart for LULC classification on the GEE platform

employed to generate predefined training locations for each land cover class within the study area. These indices (Guha and Govil, 2022) – results shown in Figure 3, derived from Landsat 8 spectral bands, are presented in Equations 1,2,and 3. Each index has been proven effective in discriminating specific land cover category. The Normalized Difference Water Index (NDWI), which enhances the contrast between the green and near-infrared bands, was applied to extract initial training samples for water bodies as NDWI value increases towards +1 in areas with water bodies. The Normalized Difference Built-up Index (NDBI), highlighting the difference between near-infrared and shortwave infrared bands, was utilized to identify built-up areas as values for built up regions increased towards +1. Similarly, the Normalized Difference Vegetation Index (NDVI), which emphasizes the contrast between near-infrared and red bands, was adopted to generate initial vegetation samples. NDVI values approach +1 if health and dense vegetation existed. As illustrated in Figure 3 (c & f), a clear distinction emerges between NDBI and NDWI in identifying water and built-up areas. This difference is particularly significant when applying threshold-based classification. In the proposed methodology, an automated thresholding strategy is adopted to ensure accurate delineation of dominant classes. Specifically, NDWI values greater than 0.1 are classified as water, NDBI values exceeding 0.4 are assigned to built-up areas, and NDVI values above 0.5 are recommended for vegetation. These threshold values were selected based on empirical testing and visual inspection, and they were found to provide an optimal balance between omission and commission errors. Sensitivity analysis indicated that lowering the thresholds increased false positives, while higher thresholds led to the exclusion of valid class samples.

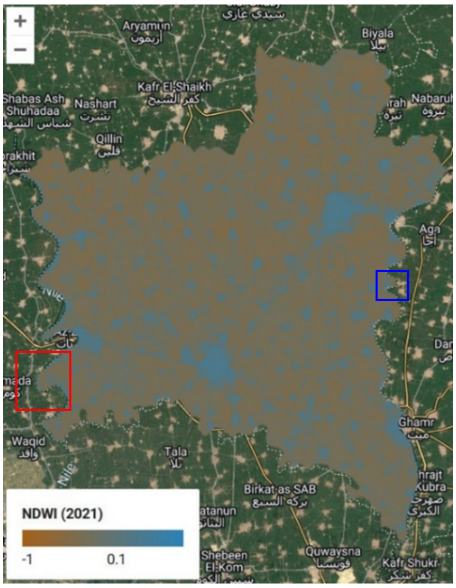
$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{1}$$

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \tag{2}$$

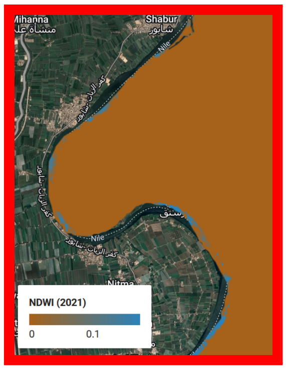
$$NDVI = \frac{NIR - Red}{NIR - Red} \tag{3}$$



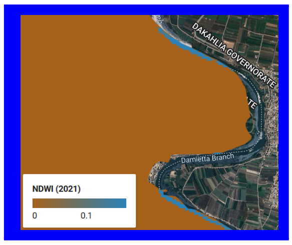
(a) Visible bands composites



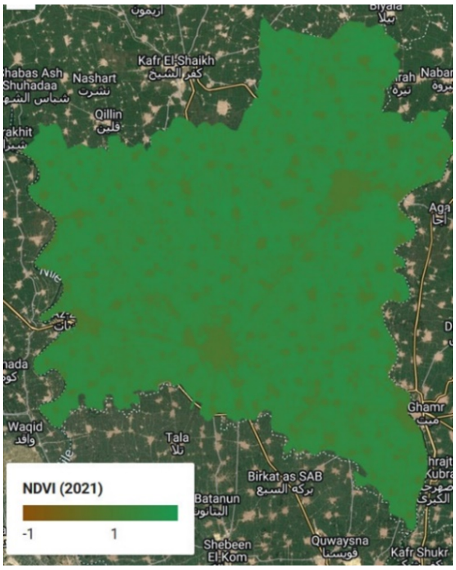
(b) NDWI



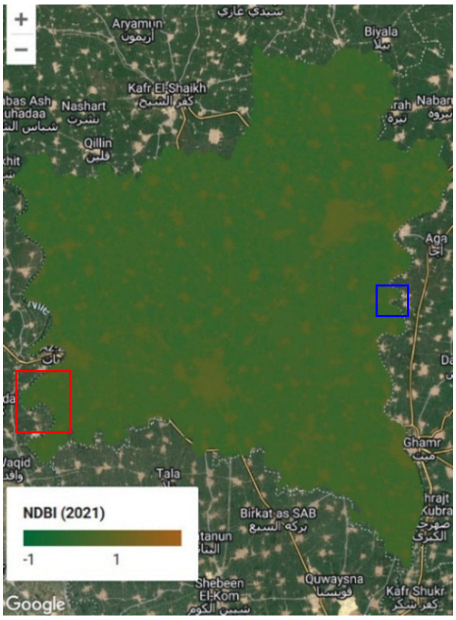
(c) NDWI zoomed for red Square in (b)



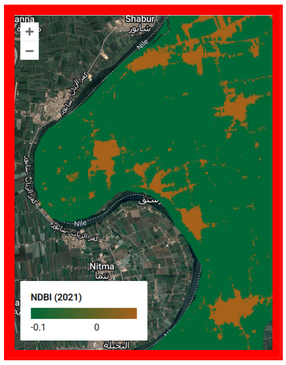
(d) NDWI zoomed for blue Square in (b)



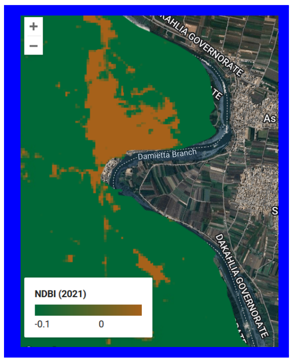
(e) NDVI



(f) NDBI



(g) NDBI zoomed for blue Square in (f)



(h) NDBI zoomed for red Square in (f)

Figure 3. Training Samples: Visible bands composites, NDWI, NDVI, and NDBI for the study area.

3.2 Land Cover Classification using the Random Forest (RF) Algorithm

The Random Forest (RF) classifier, implemented through the smileRandomForest function in GEE, was applied using 50 decision trees, which provided a balance between computational efficiency and classification accuracy. For each land cover class, 500 sample points were automatically generated from objects generated from indices for each class, ensuring adequate representation of the study area, as shown in Figure 4. The dataset was subsequently partitioned into training and validation subsets, with 70% of the samples allocated for training and the

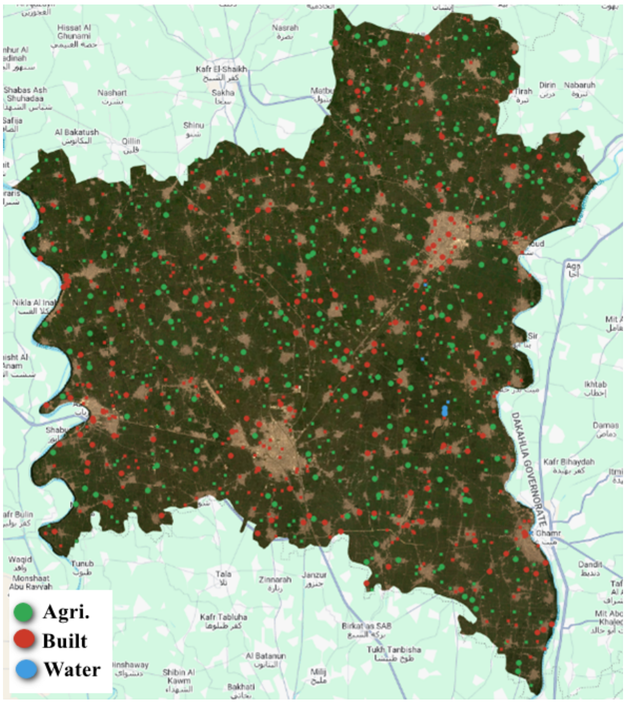


Figure 4. The automatic selection of training and testing points [green: Vegetation, Red: Built, and navy blue: water]

remaining 30% reserved for accuracy assessment and classification results shown in Figure 6-c. To facilitate visualization and distinction, the points were symbolized by both color and shape: agricultural areas were represented in green, built-up areas in red, and water bodies in blue.

3.3 Accuracy assessment of the classified results

The accuracy of the classification was assessed using the 2021 Landsat image. Evaluation metrics included overall accuracy (OA) and the Kappa coefficient, which are widely adopted in remote sensing studies to validate classification reliability. The overall accuracy was calculated as the proportion of correctly classified pixels in relation to the total number of validation samples. To provide a more detailed assessment of class-level performance, a confusion matrix was generated, from which user's accuracy, producer's accuracy, and the Kappa statistic were derived. For validation, 30% of the ground reference points were randomly selected and reserved as independent testing data, while the remaining 70% were used for training the classifier.

4. Analysis and Discussion

The classification was comparatively reported for 2021 as shown in Figure 5 and overall accuracy and kappa coefficients achieved 98.3% and 0.967 respectively. So, RF was applied to four different time slots: 2015, 2018, and 2025 to get LULC mapping as shown in Figure 6.

Based on the LULC classification, statistical analysis, and spatial visualization of changes (Table 1), agricultural land remains the dominant class, accounting for approximately 75% of the study area, compared to 25% for built-up regions. However, the proportion of agricultural land has gradually declined from 75.5% in 2015 to 72.7% in 2025, reflecting the loss of cultivated land primarily due to urban expansion. Correspondingly,

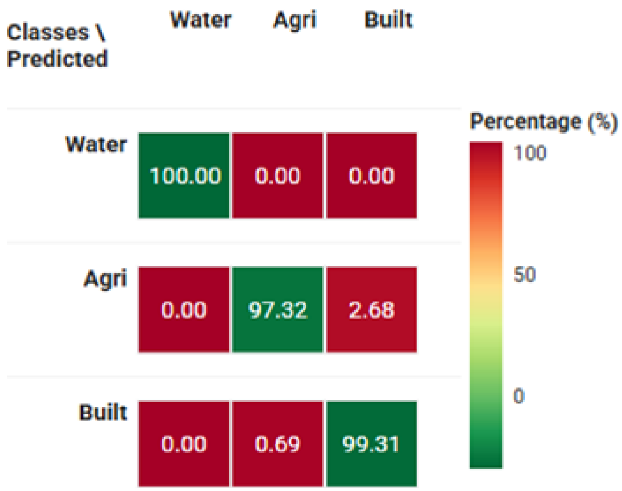


Figure 5. fig:The heatmap presents the area (%) information of LULC classes for 1500 training and test points.

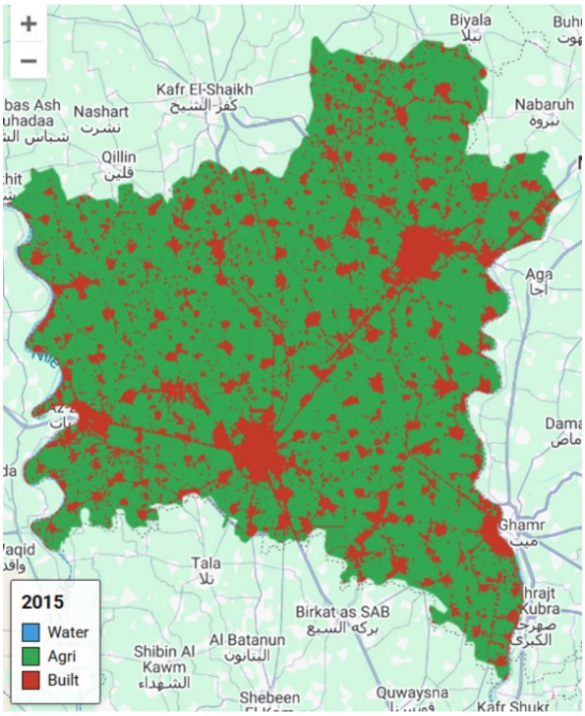
Table 1. Details of the LULC area (km2) in Gharbia governorate for 2015, 2018, 2021, and 2025

Year	Water (km^2)	Agri (km^2)	Built (km^2)
2015	9.894	1498.297	475.764
2018	9.227	1469.064	505.664
2021	10.437	1389.837	583.681
2025	9.967	1443.024	530.964

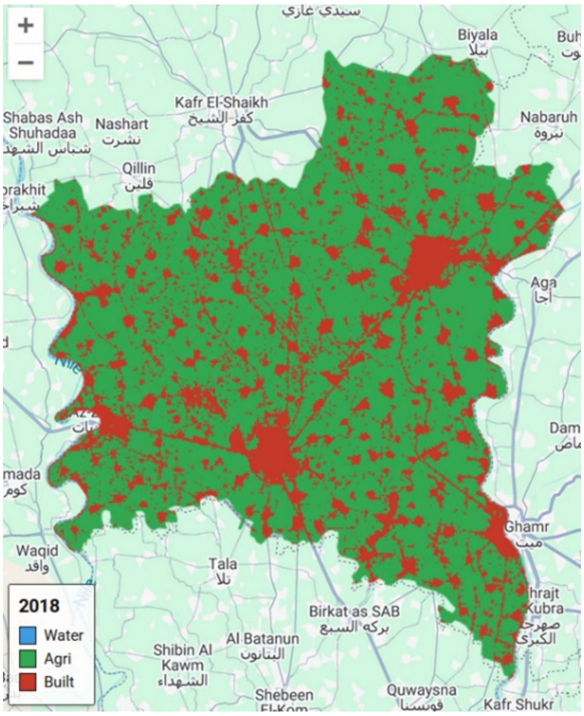
the built-up class increased from 23.98% in 2015 to 26.76% in 2025. The water class, which mainly represents irrigation canals, exhibited only marginal changes. These minor variations may be attributed either to the actual stability of water bodies or to the challenges in precisely delineating canal boundaries, which are occasionally misclassified as vegetation. Spatial patterns of urban expansion were observed predominantly around the centers of the governorate and along major infrastructure, such as roads and railway stations. This trend reflects the absence of desert or unutilized land available for expansion, forcing urban growth to occur primarily at the expense of agricultural land. Similar patterns of agricultural land loss and urban sprawl have been reported across other Nile Delta governorates, including Dakahlia and Sharqia, where rapid population growth and infrastructure development continue to exert pressure on fertile farmland, reinforcing the regional challenge of sustainable land management.

5. Conclusion

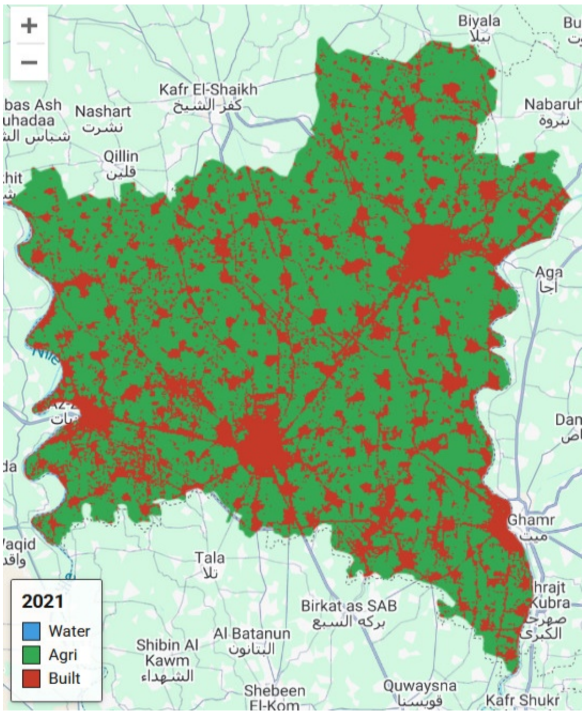
Urban growth has a significant impact on ecosystem functions such as transportation, housing, utilities, employment, and economic development. Land use and land cover (LULC) classification plays a pivotal role in managing urban expansion and provides clear insights for decision-makers, particularly in developing countries. Freely available satellite imagery (e.g., Landsat 8) can be used for large-scale urban growth management and prediction. In this study, we introduce a fully automated approach to classify and assess urban expansion in building areas of Gharbia Governorate. The approach utilizes GEE and automatically extracts training samples using commonly applied indices (e.g., NDVI, NDWI, NDBI) to train Random Forest (RF) classifiers. The proposed method achieved an overall accuracy



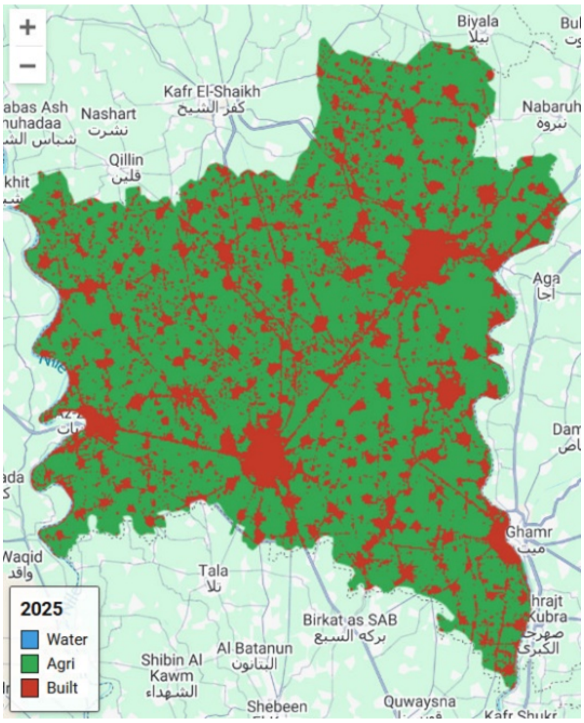
(a) 2015



(b) 2018



(c) 2021



(d) 2025

Figure 6. LULC Mapping of Gharbia: (a) 2015, (b) 2018, (c) 2021, and (d) 2025.

(OA) of 98.3% and a kappa coefficient of 0.967. In conclusion, the proposed approach demonstrates both efficiency and reliability, benefiting from the automatic extraction of training samples and the computational power of GEE. The findings reveal that water bodies in Gharbia Governorate remain relatively stable, whereas agricultural lands are experiencing considerable decline due to continuous urban expansion. Urban growth exerts adverse effects on systems such as transportation, housing, employment, and privacy, with its impacts being most severe in developing countries where expansion is often rapid and unplanned.

References

Atef, I., Ahmed, W., Abdel-Maguid, R. H., 2023. Modelling of land use land cover changes using machine learning and GIS techniques: a case study in El-Fayoum Governorate, Egypt. *Environmental Monitoring and Assessment*, 195.

Atef, I., Ahmed, W., Abdel-Maguid, R. H., 2024. Future land use land cover changes in El-Fayoum governorate: a simulation study using satellite data and CA-Markov model. *Stochastic Environmental Research and Risk Assessment*, 38(2), 651–664.

Brown, L. M., Anand, M., 2024. Impacts of urban land-cover on plant community structure and biodiversity in a multi-use landscape. *Landscape Ecology*, 39, 197. https://link.springer.com/10.1007/s10980-024-01988-9.

Guha, S., Govil, H., 2022. Annual assessment on the relationship between land surface temperature and six remote sensing indices using landsat data from 1988 to 2019. *Geocarto International*, 37, 4292-4311.

Gülci, S., Wing, M., Akay, A. E., 2025. Land Use and Land Cover (LULC) Mapping Accuracy Using Single-Date Sentinel-2 MSI Imagery with Random Forest and Classification and Regression Tree Classifiers. *Geomatics*, 5, 29. https://www.mdpi.com/2673-7418/5/3/29.

Hussain, S., Mubeen, M., Nasim, W., Karuppannan, S., Ahmad, A., Amjad, M., Fahad, S., Tariq, A., Akram, W., 2024. Assessing the impact of land use land cover changes on soil moisture and vegetation cover in Southern Punjab, Pakistan using multi-temporal satellite data. *Geology, Ecology, and Landscapes*.

Li, X., Zhou, Y., Hejazi, M., Wise, M., Vernon, C., Iyer, G., Chen, W., 2021. Global urban growth between 1870 and 2100 from integrated high resolution mapped data and urban dynamic modeling. *Communications Earth & Environment*, 2, 201. https://www.nature.com/articles/s43247-021-00273-w.

Loukika, K. N., Keesara, V. R., Sridhar, V., 2021. Analysis of land use and land cover using machine learning algorithms on google earth engine for Munneru river basin, India. *Sustainability (Switzerland)*, 13.

Mostafa, E., Li, X., Sadek, M., 2023. Urbanization Trends Analysis Using Hybrid Modeling of Fuzzy Analytical Hierarchical Process-Cellular Automata-Markov Chain and Investigating Its Impact on Land Surface Temperature over Gharbia City, Egypt. *Remote Sensing*, 15(3). https://www.mdpi.com/2072-4292/15/3/843.

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.

Roy, S., Ray, A., Halder, B., 2025. Impacts of landform dynamics on peri-urban livelihood in the southern outskirt of Kolkata municipal corporation. *Letters in Spatial and Resource Sciences*, 18(1), 1–27.

Teck, V., Poortinga, A., Riano, C., Dahal, K., Legaspi, R. M. B., Ann, V., Chea, R., 2023. Land use and land cover change implications on agriculture and natural resource management of Koah Nheaek, Mondulkiri province, Cambodia. *Remote Sensing Applications: Society and Environment*, 29, 100895.

Valjarević, A., Morar, C., Brasanac-Bosanac, L., Cirkovic-Mitrovic, T., Djekic, T., Mihajlović, M., Milevski, I., Culafic, G., Luković, M., Niemets, L., Sehida, K., Kaplan, G., 2025. Sustainable land use in Moldova: GIS & remote sensing of forests and crops. *Land Use Policy*, 152, 107515. https://linkinghub.elsevier.com/retrieve/pii/S0264837725000481.

Wang, J., Wang, Y., Qi, Z., 2024. Remote sensing data assimilation in crop growth modeling from an agricultural perspective: New insights on challenges and prospects.