Automated Spatial Data Processing Pipeline and GIS Decision-Support System for Large-Scale Urban Morphology Analysis

Oydin Shoumarova[1], Anna Vetlugina[1] and Sabina Alieva[2]
[1]*Department of Design, Tashkent University of Architecture and Civil Engineering, Yangi Shahar Str. 9, 100000 Tashkent, Uzbekistan*
[2]*Department of Foundations of Architecture, Azerbaijan University of Architecture and Construction, AZ1073 Baku, Azerbaijan*
ayupovaoydin@gmail.com, anna_vet2015@mail.ru, sabinaaliyeva888@hotmail.com

Keywords: Spatial Decision-Support Systems, Automated Data Pipeline, Smart City Infrastructure, GIS Algorithms, Open Geospatial Data, Spatial Data Science.

Abstract: Rapid urban growth requires advanced computational tools and automated spatial information systems for processing large-scale geospatial data. This paper presents an automated GIS-based data processing pipeline and architectural framework for a Spatial Decision-Support System (SDSS) designed to compute, classify, and visualize complex spatial morphology indicators. Utilizing open-source geospatial data from OpenStreetMap, the developed system automates the extraction, geometric correction, and processing of vector data. It implements an algorithmic workflow based on the Spacematrix methodology to calculate multi-dimensional spatial indicators, including Floor Space Index (FSI), Ground Space Index (GSI), average building height (L), and Open Space Ratio (OSR). The core contribution of this study is a reproducible, data-driven ETL (Extract, Transform, Load) and spatial classification pipeline that handles geometric anomalies and executes data-driven zoning. The system was validated using a dataset of 464 spatial analysis units, automatically classifying them into four distinct morphological typologies. The results demonstrate the efficiency of the developed GIS pipeline as a scalable, morphology-sensitive component for Smart City digital twins and urban information environments.

1 INTRODUCTION

Rapid urban growth requires improving urban environmental quality while increasing residential density through planning tools. In many fast-growing cities, densification of existing areas and peripheral expansion occur simultaneously, often reducing average density and leading to inefficient land use [1]-[3]. High building density can reduce open spaces and worsen infrastructure accessibility [4].

Similar processes characterize Tashkent, where densification of existing residential areas is combined with peripheral expansion, producing a heterogeneous spatial structure of density and open spaces. Urban land increased substantially between 1990 and 2019, mainly due to the conversion of agricultural and other open/green lands [5].

For the quantitative assessment of building characteristics and open-space structure, this study applies the Spacematrix methodology within a GIS framework based on the analysis of density indicators (FSI, GSI), average building height (L), and Open Space Ratio (OSR) [6].

When integrated with GIS, the Spacematrix methodology enables automated calculation and spatial classification of density indicators.

Tashkent is an appropriate case study because it combines multi-storey residential areas with traditional low-rise mahallas, enabling comparison of different residential morphologies and their implications for urban environmental quality.

This study addresses these challenges by developing an automated, reproducible GIS-based framework and data processing pipeline designed for city-scale spatial analysis. Integrated within a modern Spatial Decision-Support System (SDSS) environment, this framework automates data extraction, preprocessing, indicator computation, and spatial zoning. The main technical contributions of this study are as follows: (1) it implements a scalable geospatial data pipeline for the automated extraction and harmonizing of heterogeneous OpenStreetMap data; (2) it develops an algorithmic processing workflow in QGIS using advanced field calculation and geometric correction scripts to compute interconnected spatial metrics (FSI, GSI, L, OSR) across hundreds of analysis units simultaneously; and (3) it introduces a data-driven spatial classification and digital mapping model that generates ready-to-use geospatial layers for Smart City information systems and digital twins.

2 METHODOLOGY

Classical urban density indicators, such as population density and Floor Space Index (FSI), also known as Floor Area Ratio (FAR), are widely used to assess development intensity; however, they have limited explanatory capacity. FSI represents the ratio of total floor area to the block area but does not account for building height distribution, spatial heterogeneity of development, or open-space quality [7], [8]. Therefore, recent studies propose integrated approaches that consider urban density as a system of interrelated indicators of development, population, and open spaces, taking spatial context and scales of analysis into account [7], [9].

From a computational perspective, the Spacematrix methodology is modeled as an algorithmic framework embedded within a Spatial Decision-Support System. In this study, the spatial parameters are treated as multi-dimensional data vectors processed through a sequence of mathematical expressions within the GIS environment. The indicator calculation algorithms are structured as follows:

- Floor Space Index (FSI) - ratio of total floor area to block area;
- Ground Space Index (GSI) - proportion of built-up area within the block;
- L - average building height (storeys);
- Open Space Ratio (OSR) - ratio of open space to total floor area within the unit of analysis; the manuscript does not state explicitly whether A_open is limited to unbuilt land within each residential unit's own boundary or also incorporates adjoining streets and public spaces; consistent with common Spacematrix practice, A_open would typically refer to unbuilt land within the unit boundary only, but this scope should be confirmed and stated explicitly by the authors for full reproducibility.
- Built form typology - classification of urban morphological types.

Indicator formulas are defined as follows:

$$FSI = A_{floor}/A_{unit} \ , \quad (1)$$

$$GSI = A_{footprint}/A_{unit} \ , \quad (2)$$

$$OSR = A_{open}/A_{floor} \ , \quad (3)$$

$$L = A_{floor}/A_{footprint} \ . \quad (4)$$

where:

- A_{floor}- total floor area of buildings;
- $A_{footprint}$- total building footprint area;
- A_{unit}- area of the residential block (unit of analysis);
- A_{open}- open space area within the block.

These indicators were computed in GIS using expressions (1)-(4), ensuring consistent analysis of large spatial datasets.

2.1 Study Area and Unit of Analysis

The study covers the entire urbanized territory of Tashkent, including historic, Soviet-era, and contemporary residential areas. The units of analysis are microdistricts, mahallas, and modern residential complexes, enabling comparison of urban morphology across different development types.

The analysis units were derived from the OpenStreetMap land-use layer by selecting polygons with the attribute *landuse = 'residential'*. These polygons represent residential urban fabric and were used as spatial units. In total, 464 residential units were identified. Building morphology indicators (FSI, GSI, OSR, and L) were calculated in GIS and used to position each unit within the Spacematrix diagram, enabling comparison of residential density types across the city.

2.2. Data Sources and GIS Processing

The analysis is based on open spatial data from OpenStreetMap (OSM), including building footprints, land-use information, and street networks. Within GIS, residential buildings were filtered, geometric errors corrected, attributes standardized, and datasets projected into a unified coordinate reference system to form a geospatial database for automated spatial analysis [10], [11]. Although OSM contains some spatial inconsistencies and incomplete attributes, its coverage is generally sufficient for city-scale analysis in Tashkent [12].

All spatial calculations and classification procedures were conducted in GIS using contemporary geospatial processing techniques. The workflow follows a sequential pipeline including data preprocessing, indicator computation, spatial classification, and mapping.

OSM data were extracted in November 2025. Prior to analysis, geometric inconsistencies in the building layer were corrected using the Fix Geometries tool in QGIS.

Building height information was obtained from the OSM attribute building:levels, which was available for approximately 10% of buildings. For buildings without this attribute, a default value of one storey was assigned. This assumption mainly affects traditional mahalla-type areas, where low-rise detached and semi-detached housing predominates, and therefore reflects the dominant morphological condition of many missing-data cases. Since the analysis is conducted at the level of residential units rather than individual buildings, the assumption is not expected to substantially distort the resulting typology. Its main effect is a possible shift in the absolute values of FSI, L, and OSR for some units, while the broader contrast between low-rise, moderate-density, and compact forms remains preserved.

To evaluate the influence of incomplete height information, a sensitivity check was conducted using a sample of 26 residential units representing different density conditions across the study area. Two scenarios were compared for buildings lacking the *building:levels* attribute: a baseline scenario assigning one storey and an alternative scenario assigning two storeys. The comparison showed that most sampled units retained the same morphological classification under both scenarios, while only three units (11.5% of the sample) shifted across threshold boundaries in the alternative scenario, all located close to class limits. This sensitivity check covers 26 of the 464 residential units (5.6% of the dataset), and the sampling procedure used to select these 26 units (e.g., random vs. stratified across density zones) is not specified in the present description; both aspects limit the extent to which this check can be generalized as evidence that the citywide impact of missing height data is small, particularly in high-rise districts where building height has a stronger influence on FSI. A larger, explicitly stratified sensitivity sample - oversampling higher-density zones - is recommended for a future revision to strengthen this validation.

These results indicate that uncertainty related to missing height data has a limited effect on the overall typological structure and mainly influences borderline cases rather than the dominant spatial pattern of residential morphology. Since the typological classification is based on relative differences between residential units rather than precise building-level estimates, the adopted assumption does not substantially affect the identification of major morphological types and is therefore acceptable for planning-oriented comparative analysis at the city scale.

2.2.1 Data Pipeline Architecture and Workflow Implementation

Spatial analysis and indicator computation were implemented in QGIS 3.40.6. The workflow included data preprocessing, delineation of residential units, computation of FSI, GSI, L, and OSR, classification of density indicators, and cartographic visualization.

Workflow:

1) Data acquisition - OSM building footprints and *landuse = 'residential'* polygons (November 2025 extract).
2) Preprocessing - Fix Geometries was used to correct invalid geometries; Delete Duplicate Geometries removed duplicate features; Reproject Layer harmonized the coordinate reference system; residential buildings were selected using Select by Expression.
3) Unit delineation - residential land-use polygons from OSM were dissolved and subdivided by major roads to produce 464 residential units.
4) Indicator computation - building footprint area and floor area were calculated using the Field Calculator; indicators FSI, GSI, L and OSR were then derived.
5) Classification - FSI and GSI were classified using empirical thresholds.
6) Mapping outputs - density indicator maps and morphological zoning layers were produced.

The structural architecture of the developed spatial data infrastructure and the sequential logic of the ETL (Extract, Transform, Load) processing stages are visually conceptualized in the pipeline diagram below (Fig. 1).

2.3 Classification of FSI and GSI Indicators

To facilitate morphological interpretation, the FSI and GSI values were classified into four classes based on their empirical distribution and morphological differences across residential areas of Tashkent, following previously applied block-level density mapping approaches [13].

Class thresholds were defined using empirical and quantile-based distributions [14] and refined through visual inspection of mapped results to ensure consistency with dominant morphological patterns [15]. Within the Spacematrix framework, the threshold values were locally adapted to the urban context of Tashkent. The specific quantile statistic used (e.g., quartiles vs. another percentile split), and the precise criteria applied during the subsequent visual-inspection refinement (as opposed to a formally optimized method such as Jenks natural breaks), are not fully documented in the present description; reporting these procedural details explicitly is recommended in a future revision to allow the classification to be reproduced independently.

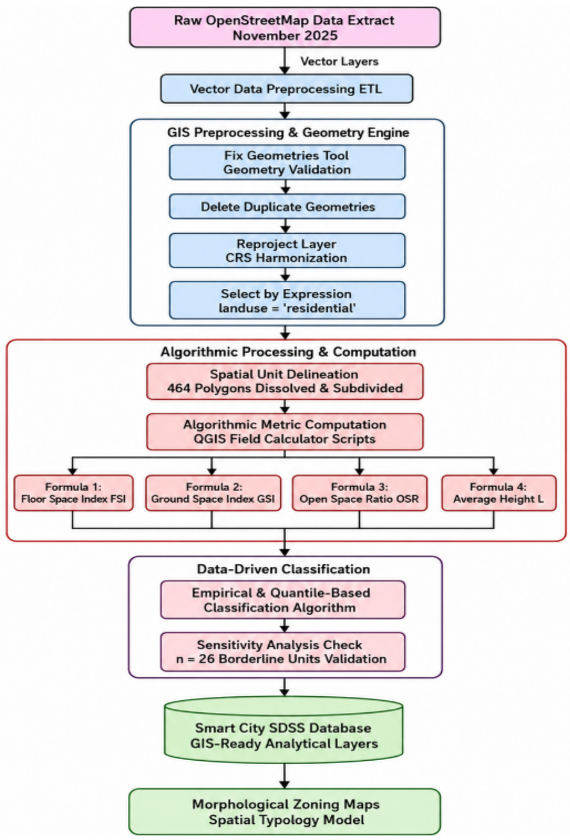


Figure 1: Architectural scheme and data flow of the automated spatial data processing pipeline.

The classification of residential density types was based on threshold ranges of the FSI and GSI indicators derived from the empirical distribution of density values in the dataset. Four classes were defined using the following thresholds: FSI < 0.80, 0.80-1.50, 1.50-2.50, and ≥ 2.50, and GSI < 0.10, 0.10-0.25, 0.25-0.40, and ≥ 0.40.

To assess the robustness of the classification, the threshold values were slightly adjusted and the resulting spatial patterns were compared. The overall morphological structure remained broadly consistent under minor threshold adjustments and moderate uncertainty in input data.

OSR and L were calculated for all residential units and used in the subsequent typological analysis, but were not classified into classes.

2.4 Urban Morphological Typology Framework

From a computational perspective, the typology construction is based on the analysis of GIS-derived indicators and represents an empirically informed classification model of residential morphology. The analysis included 464 residential zones representing dominant residential morphologies of Tashkent.

Building on the previously defined FSI and GSI classes, residential areas were further analyzed within the Spacematrix framework. Indicators were calculated in GIS and visualized through a Spacematrix diagram (Fig. 2) to explore density relationships within the residential fabric. The resulting typological groupings can be reused in urban information systems and planning tools.

The identified morphological differentiation is consistent with earlier typological findings for Tashkent obtained using the Spacematrix approach [16], while the present study expands this analysis through GIS-based mapping.

The resulting typological grouping of residential areas is presented in the Results section (Fig. 3).

3 RESULTS

To identify spatial patterns of built-up density across Tashkent, the indicators FSI, GSI, OSR, and L were calculated and mapped at the level of urban blocks (Fig. 4). The resulting maps represent GIS-generated analytical layers.

3.1 Spatial Distribution of FSI

The FSI map (Fig. 4a) shows clear spatial variability in development intensity across Tashkent. Higher FSI values concentrate in central areas and along major transport corridors, while traditional mahalla neighborhoods show lower values. Most blocks fall within intermediate ranges.

3.2 Spatial Distributionof GSI

The GSI map (Fig. 4b) reflects density in terms of the proportion of land covered by buildings rather than building height. Unlike FSI, high GSI values (class 4) occur not only in central districts but also in mid-rise residential blocks and compact mahalla areas, where buildings occupy a substantial share of the land surface. Moderate GSI values (classes 2-3) dominate across most of the city and correspond to relatively balanced morphologies with preserved courtyard and semi-open spaces.

Comparison of the FSI and GSI maps shows that high FSI does not necessarily coincide with very high GSI: some areas combine tall buildings with limited ground coverage, while others exhibit extensive ground coverage associated with predominantly low-rise development.

3.3 Spatial Distribution of L

The L map (Fig. 4c) indicates that a large portion of the city is characterized by low-rise development (1-2.5 storeys), typical of peripheral areas and traditional mahalla morphologies dominated by one- and two-storey buildings. Areas with low- and mid-rise development (2.5-4.5 and 4.5-7 storeys) occur in a fragmented pattern without clearly defined spatial zones. High-rise development (>7 storeys) shows a point-based distribution and appears as isolated insertions rather than continuous zones.

3.4 Spatial Distribution of OSR

The OSR map (Fig. 4d) reveals spatial variability in the ratio of open space to total floor area across residential blocks in Tashkent. Low OSR values (0-0.30) characterize densely built-up areas and indicate high land coverage by buildings. Moderate OSR values (0.30-0.60) correspond to the most prevalent urban fabric with a relative balance between built-up and open spaces. High OSR values (≥0.60) are mainly associated with peripheral and low-density areas and reflect a higher amount of open space relative to build floor area.

3.5 Morphological Zoning of Residential Development

Based on the integrated analysis of density indicators, a GIS-based morphological zoning of residential development was performed. The zoning represents a classification model combining multiple indicators into a unified typological structure. The following section presents the characteristics of the four identified morphological types of residential development (Zones A-D). Their statistical characteristics and the distribution of residential units (n = 464) are summarized in Table 1.

3.5.1 Zone A - Dispersed Low-Rise Structure

Zone A is characterized by low FSI and GSI values, low building heights, and high OSR.

Spatial distribution: Zone A is located in peripheral residential areas and isolated inner-city enclaves. It includes traditional mahalla-type areas with dispersed low-rise development and other low-intensity residential fabrics. These areas show low land-use intensity, limited infrastructure pressure, and lower accessibility and service provision than denser urban zones.

3.5.2 Zone B - Moderate-Density Structure

Zone B is characterized by moderate density, medium building heights, and relatively high OSR.

Spatial distribution: Zone B represents the dominant residential fabric of Tashkent, with a balanced relationship between built-up and open spaces, stable environmental conditions, moderate infrastructure pressure, and potential for selective densification and public-space improvement.

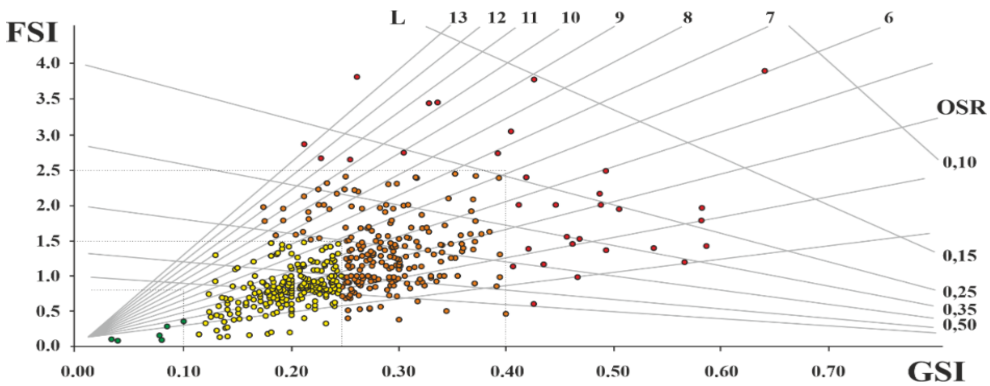


Figure 2: Spacematrix diagram of residential blocks based on FSI and GSI with reference lines of L and OSR.

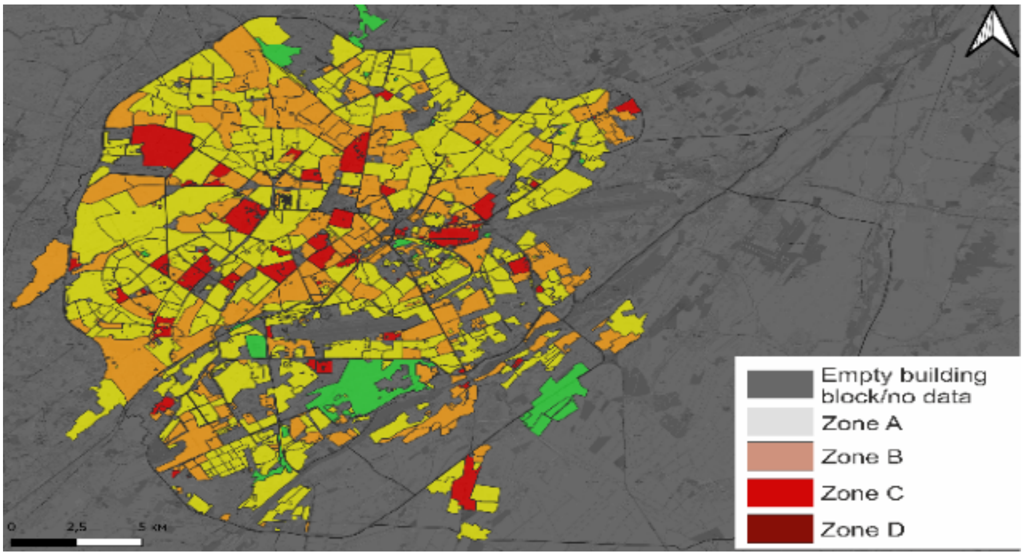
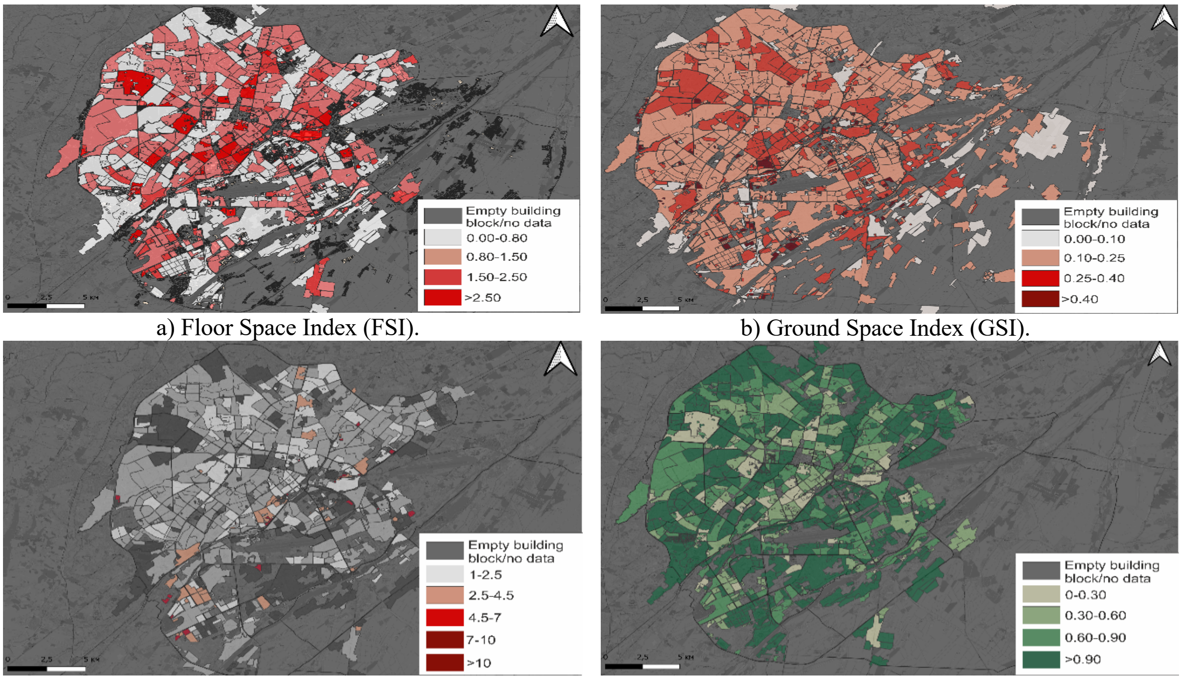


Figure 3: GIS-generated morphological zoning model of residential development in Tashkent (Zones A-D).



a) Floor Space Index (FSI).

b) Ground Space Index (GSI).

c) Average Building Height (L).

d) Open Space Ratio (OSR).

Figure 4: Spatial distribution of density indicators in Tashkent.

Table 1: Statistical characteristics and indicative value ranges of residential morphological zones.

Zone	FSI (min-max)	GSI (min-max)	L (mean ± SD)	OSR (mean ± SD)	n	Share (%)
A	0.11 - 0.49	0.011 - 0.10	2.24 ± 0.97	1.40 ± 0.35	30	6.5
B	0.14 - 1.49	0.10 - 0.25	3.69 ± 1.55	0.78 ± 0.12	235	50.6
C	0.28 - 2.45	0.15 - 0.40	4.41 ± 2.08	0.61 ± 0.18	143	30.8
D	1.50 - 4	0.21 - 0.65	5.19 ± 2.56	0.22 ± 0.07	56	12.1

3.5.3 Zone C - Dense Low- and Mid-Rise Structure

Zone C is characterized by elevated GSI combined with moderate FSI, indicating high ground coverage with relatively limited building heights.

Spatial distribution: Zone C is concentrated in central and adjacent areas and is associated with compact residential fabrics with limited courtyards. It includes compact mahalla-type areas and low- to mid-rise microdistricts with limited open spaces, requiring improvements in courtyards, public spaces, and environmental comfort.

3.5.4 Zone D - High-Density Compact Structure

Zone D is characterized by the highest spatial intensity, with elevated GSI, moderate-to-high FSI and building heights, and low OSR.

Spatial distribution: Zone D shows a clustered pattern mainly in central areas and along major transport corridors, embedded within surrounding residential fabrics and reflecting intensive densification. It is associated with high land-use intensity and pressure on transport, engineering, and public space infrastructure.

4 DISCUSSION

Residential density patterns in Tashkent are shaped not only by building height but also by the interaction between GSI and OSR. Comparison of FSI and GSI shows that high spatial intensity may result from multi-storey development or compact low-rise fabrics. Areas with high GSI and low OSR show reduced spatial permeability and higher infrastructure pressure.

Traditional mahalla-type areas exhibit significant morphological diversity, ranging from dispersed low-density patterns to compact low-rise configurations with high ground coverage. Therefore, mahalla development cannot be treated as a single morphological category. GIS-derived indicators capture this variability, enabling differentiation of traditional residential fabrics. In quantitative terms, this diversity is reflected in the zone distribution already reported in Table 1: only n = 30 units (6.5%) fall into the lowest-density Zone A, while the majority of units are distributed across the moderate-density Zones B and C (n = 235, 50.6%, and n = 143, 30.8%, respectively), with a smaller compact/high-rise Zone D (n = 56, 12.1%); a full cross-tabulation linking individual mahalla names or districts to these zones was not compiled in the present study and is identified as a natural extension for future spatially disaggregated analysis.

The classification thresholds used in this study are partly data-driven and partly interpretation-guided. They were derived from the empirical distribution of FSI and GSI values and supported by quantile-based logic but were also refined through visual interpretation of mapped spatial patterns to ensure morphological coherence in the context of Tashkent. For planning-oriented GIS analysis, this approach is acceptable because the aim is not to define universally fixed thresholds, but to identify spatially meaningful and operational categories that can support comparative interpretation, zoning, and decision-making within a specific urban context.

The identified morphological zones represent consistent configurations of density, compactness, and open-space structure corresponding to distinct urban environments with different constraints and potentials. Such typologies support the interpretation of urban density patterns and morphology-sensitive densification strategies.

5 CONCLUSIONS

This study demonstrates the effectiveness of integrating the Spacematrix methodology with GIS-based analysis for large-scale assessment of residential density and open space structure. The combined analysis of FSI, GSI, L, and OSR enabled the identification of four morphological types of residential development in Tashkent. The resulting density indicators and zoning represent geospatial outputs that can be reused in urban planning systems. In practice, these outputs can be integrated into urban planning decision-support systems as GIS layers for identifying areas suitable for infill development, open-space protection, and infrastructure monitoring.

The typological framework provides a basis for context-sensitive urban planning and supports densification regulation, infrastructure prioritization, and improvement of public and open spaces. The proposed typology helps distinguish areas suitable for infill development from those requiring open-space preservation and infrastructure pressure reduction. The GIS-supported approach enables comparative spatial analysis of residential morphology and can be applied to other rapidly growing cities.

The results show that residential development in Tashkent is dominated by medium-density morphology (Zone B), accounting for 50.6% of the analyzed blocks. Low-density areas (Zone A) represent 6.5% of the sample, while compact mid-rise and high-density forms (Zones C and D) account for 30.8% and 12.1%, respectively. The analysis also indicates a gradual decrease in OSR with increasing density, with average OSR declining from 1.40 in Zone A to 0.22 in Zone D.

REFERENCES

[1] S. Angel, P. Lamson-Hall, A. Blei, S. Shingade, and S. Kumar, "Densify and expand: A global analysis of recent urban growth," Sustainability, vol. 13, no. 7, p. 3835, 2021.

[2] G. Xu, Z. Zhou, L. Jiao, and R. Zhao, "Compact urban form and expansion pattern slow down the decline in urban densities: A global perspective," Land Use Policy, vol. 94, p. 104563, 2020.

[3] H. Lu, Z. Shang, Y. Ruan, and L. Jiang, "Study on urban expansion and population density changes based on the inverse S-shaped function," Sustainability, vol. 15, no. 13, p. 10464, 2023.

[4] A. Chokhachian, K. Perini, S. Giulini, and T. Auer, "Urban performance and density: Generative study on interdependencies of urban form and environmental measures," Sustainable Cities and Society, vol. 53, p. 101952, 2020.

[5] I. Aslanov et al., "Applying remote sensing techniques to monitor green areas in Tashkent, Uzbekistan," in E3S Web of Conferences, vol. 258, p. 04012, 2021.

[6] M. Berghauser Pont and P. Haupt, Spacematrix: Space, Density and Urban Form, Rev. ed. Delft, The Netherlands: TU Delft Open Publishing, 2023.

[7] K. Dovey and E. Pafka, "The urban density assemblage: Modelling multiple measures," Urban Design International, vol. 19, no. 1, pp. 66-76, 2014.

[8] Z. Bao and C. Li, "Progress on the study of urban architecture FAR," Progress in Geography, vol. 29, pp. 396-402, 2010 (in Chinese).

[9] M. Wurm et al., "Inferring floor area ratio thresholds for the delineation of city centers based on cognitive perception," Environment and Planning B: Urban Analytics and City Science, vol. 48, no. 2, pp. 265-279, 2021.

[10] A. Karduni, A. Kermanshah, and S. Derrible, "A protocol to convert spatial polyline data to network formats and applications to world urban road networks," Scientific Data, vol. 3, no. 1, pp. 1-7, 2016.

[11] C. Fang, L. Zhou, X. Gu, X. Liu, and M. Werner, "A data-driven approach to urban area delineation using multi-source geospatial data," Scientific Reports, vol. 15, no. 1, p. 8708, 2025.

[12] F. Mara, C. Anselmi, F. Deri, and V. Cutini, "The divergent geographies of urban amenities: A data comparison between OpenStreetMap and Google Maps," Sustainability, 2025, [Online]. Available: https://doi.org/10.3390/su17209016.

[13] K. Hackenbroch, U. Altrock, and H. Sterly, "Governance of the megacity: Challenges for service provision and planning," in MegaCities MegaChallenge-Informal Dynamics of Global Change: Insights from Dhaka, Bangladesh and the Pearl River Delta, China. Stuttgart, Germany: Schweizerbart, 2019, pp. 66-70.

[14] C. Hennig and C. Viroli, "Quantile-based classifiers," Biometrika, vol. 103, no. 2, pp. 435-446, 2016.

[15] J. Wang, W. Huang, and F. Biljecki, "Learning visual features from figure-ground maps for urban morphology discovery," Computers, Environment and Urban Systems, vol. 109, p. 102076, 2024.

[16] O. A. Shoumarova, A. V. Vetlugina, and M. A. Norboeva, "Typology of urban form and density indicators: Application of the Spacematrix methodology to the districts of Tashkent," Universum: Engineering Sciences, vol. 6, no. 135, pp. 43-47, Mar. 2025, [Online]. Available: https://doi.org/10.32743/UniTech.2025.135.6.20297 (in Russian).