land

MDPI

Article

A Multi-Scale Workflow for Analysing the Urban Morphological Spectrum: A Comparative Analysis of Three Mid-Sized Cities

Fethi Ahmet Canpolat

Department of Geography, Fırat University, Elazığ 23119, Turkey; facanpolat@firat.edu.tr

Abstract

Traditional urban morphological analyses are structurally limited in terms of both systemic diversity and empirical scale due to reliance on manual methods. In contrast, decoding the complex fabric of rapidly growing cities necessitates data-driven, scalable approaches. To address this gap, this study proposes a multi-scale pipeline that classifies urban form systematically and reproducibly from open spatial data, and applies it comparatively to three Anatolian cities of contrasting typo-morphological character: Elazığ, Erzincan and Mardin. From street networks and building geometries, an integrated morphometric matrix was assembled by computing network topology and orientation metrics, space syntax configurational accessibility, morphological tessellation, coverage area ratio, building form–volume indicators, neighbourhood and adjacency measures, and Local indicators of spatial association (LISA) spatial autocorrelation. After transformation and standardisation, urban typologies were derived through Principal Component Analysis and grouped with spatially weighted k-means (Geo-KMeans). Three findings stand out. First, the cities trace a distinct morphological spectrum that runs from Mardin's organic historical fabric to Erzincan's relatively planned grid structure, with the radial polarised Elazığ occupying an intermediate, transitional position between the two. Second, accessibility and built density prove only weakly related (r = 0.05–0.25). Third, six, five and four morphological typologies emerged, triangulated against LISA hot spot clusters and space syntax maps. Overall, this reproducible framework offers planners a systematic, data-driven basis for exploratory morphological assessment rather than a definitive, universal typology.

Keywords: computational urban morphology; machine learning; space syntax; PCA; cluster analysis; Anatolian cities

Check for updates

Academic Editors: Tao Liu, Yingzhi Qiu and Jiajie Liu

Received: 19 June 2026
Revised: 13 July 2026
Accepted: 15 July 2026
Published: 18 July 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Urban morphology is an interdisciplinary field that examines the physical forms of human settlements, their spatial organisation and their processes of transformation over time [1–3]. The field has traditionally rested on two theoretical strands: the historico-geographical approach developed under the leadership of Conzen and the typo-morphology shaped by Muratori's Italian school [4–6]. Both traditions read the urban fabric qualitatively, through plan units, building types and hierarchies of open space, and thereby pioneered the historical–typological analysis of urban form [7,8]. At the turn of the 21st century, however, a profound shift in geographic data science, the rapid growth in computational capacity and the maturation of global open spatial data infrastructures such as OpenStreetMap (OSM) began to test the limits of this paradigm [9,10].

This shift gave rise to a new research strand, conceptualised as "computational urban morphology" or "urban morphometrics" [11–13]. The approach measures the physical

fabric of cities through hundreds of indicators, analyses them systematically and statistically, and partitions them into typological classes by means of machine learning [14]. Driven by the search for an analytical foundation that renders the interpretive depth of qualitative reading measurable and reproducible, the strand has matured its toolset one instrument at a time [11,15,16]. Fleischmann's momepy toolkit, for instance, offers more than 300 characters spanning dimension, shape, spatial distribution, intensity, connectivity and diversity at the building scale, which makes it the most comprehensive platform for the building-level urban morphometrics tradition [12,15]. Boeing's OSMnx toolkit, in turn, models street networks as directed multigraphs and supports the systematic computation of network metrics such as orientation entropy, node degree and circuity [17,18].

For urban street networks specifically, the Multiple Centrality Assessment (MCA) proposed by Porta and colleagues moves beyond topological abstraction: its "primal" graph approach defines intersections as nodes and street segments as edges, grounding centrality in metric and geographic reality [19]. Within this framework, the closeness, betweenness, straightness and information centrality indices yield evidence-based readings of the critical axes that form a city's skeleton. Such studies share the lineage of the vector-based morphological analyses that this research also pursues.

Alongside these vector-based approaches, a parallel family of indices works with pixel-based raster data. The FRAGSTATS-based landscape metrics developed by McGarigal and Marks quantify the geometric complexity and fragmentation of urban form through indicators such as the perimeter–area fractal dimension and the shape index [20]. Bosch's PyLandStats library carries the same landscape ecology logic into urban morphology metrics such as fragmentation, diversity and density [21]. Moving beyond the comparison of two static states, the DynamicPATCH module of Zhang and colleagues characterises how urban or land use patches evolve through coalescence, splitting, expansion, contraction and disappearance [22].

Processing the high-dimensional datasets that arise from these vector-based indices calls for dimensionality reduction and statistical refinement techniques [23–25]. The Variance Inflation Factor (VIF) serves as a common preliminary diagnostic for controlling multicollinearity [26]. Linear dependence among variables can also be resolved by projecting them onto orthogonal components through Principal Component Analysis (PCA) rather than by discarding variables outright [27]. Morphological indicators are thereby reduced to meaningful dimensions, yielding a sound input set for unsupervised machine learning algorithms. For the clustering itself, this study adopts spatially weighted k-means (Geo-KMeans), which jointly weighs morphological similarity and geographic proximity, drawing on the ClustGeo approach. Rather than fixing the spatial weight parameter (alpha) at a heuristic value, we set it through a data-driven criterion: the value that balances the inertia explained in feature space against that explained in geographic space [28].

Building on these foundations, this research develops a multi-scale workflow that tests these vector-based indices together on three cities representing Türkiye's diversity of urban fabric across both a historical and a functional spectrum. Mardin, Erzincan and Elazığ serve as typological markers not merely through their geographic locations but through their capacity to anchor the measurable outputs of urban form (entropy, grid regularity and centrality) at contrasting extremes. This positioning allows the discriminating power of the proposed computational pipeline to be tested across distinct morphological contexts.

The workflow is structured across three hierarchical scales. At the micro-scale, morphometric characters are identified at building and cell level using Fleischmann's momepy tools [14]. At the meso-scale, the topological and metric centralities of urban networks are measured using the methods of Boeing and Porta [17,19]. At the macro-scale, morphological typology maps are produced by integrating Principal Component Analysis (PCA) with

machine learning algorithms [27,29]. Together, these processes aim to provide a systematic record of the current state of urban form and establish a robust, comparable analytical infrastructure for urban planning.

The study goes beyond a purely empirical spatial analysis by offering an integrated computational framework that combines the precise measurement techniques of computational urban morphology with the classification power of machine learning. This framework addresses the ongoing debate surrounding the role of quantitative and qualitative analysis in morphology [16,24]. Through data standardisation, cross-scale flexibility, and an adjustable parameter structure, the framework proposes a systematic, comparable, and reproducible workflow for characterising urban form. Treating the urban fabric as a mathematical and algorithmic dataset enables spatial analysis to be conducted in a data-driven and reproducible manner, synthesising existing methods. The contribution lies in integration rather than in any new component: none of the individual elements (OSMnx network metrics, momepy morphometrics, space syntax integration, LISA, Geo-KMeans) is new here. What this study adds is their integration into a single pipeline and its comparative application to three Anatolian cities that have not previously been analysed this way, not a new algorithm or measurement technique.

2. Materials and Methods

2.1. Study Area

This study covers Elazığ, Erzincan and Mardin, three cities that embody contrasting geographic, historical and morphological contexts of Anatolia. The decisive selection criterion was to gather, within a single sample, the Anatolian counterparts of the fundamental fabric archetypes defined in the urban morphology literature: organic historical, planned grid and mixed topographic (Figure 1) [2,5,7,30]. These characterisations are then tested quantitatively in Section 3. For each city, the analytical extent was delineated as a single study area polygon, which was manually digitised to follow the contiguous built-up urban fabric rather than administrative or municipal boundaries. These polygons (302.2 km^2 in Elazığ, 230.7 km^2 in Erzincan, and 70.8 km^2 in Mardin) served as both the network query extent and the area denominator for density metrics. Because manual delineation carries an inherent degree of subjectivity, small differences at the study area edge could marginally affect area-normalised metrics, and this should be kept in mind when comparing absolute densities across the three cities.

Elazığ, with a population of approximately 440,000 as of 2025, is a mid-sized city in the southwest of the Eastern Anatolia Region, within the Upper Euphrates section, on a mountain-to-plain transitional topography. As the most populous settlement and the most extensive built environment among the three provinces, it rests on a multi-axial network geometry shaped under topographic constraints. Its movement pattern organises around a radial spine that converges towards the centre across this geometry. This hybrid structure, in which the traditional organic fabric and modern development plans interpenetrate [31], makes Elazığ the representative of the mixed topographic archetype and places it in an intermediate, calibrating position between the morphological extremes of the other two cities.

Erzincan, with a population of about 150,000, lies on an alluvial plain in the Upper Euphrates section of the Eastern Anatolia Region, and it suffered several destructive earthquakes (1939, 1992) during the 20th century [32,33]. The successive reconstruction efforts that followed largely recast the city's physical fabric into an orthogonal grid morphology [34]. Erzincan therefore stands as a clear example of the planned grid archetype, a fabric rebuilt through seismic memory, and provides suitable ground for testing the model's thresholds of "order" and "low entropy".

Mardin, with a population of approximately 173,000, is a historic settlement on a limestone slope in southeastern Anatolia, inscribed on UNESCO's Tentative World Heritage List [32,35]. Its physical morphology was shaped jointly by organic growth continuing from the Middle Ages to the present, a steep hillside topography and traditional stone architecture [36]. The dense, compact historic core on the slope and the planned new neighbourhoods extending towards the plain establish a patterned contrast within the city, which makes Mardin the study's organic historical archetype. Although Mardin falls in the same population class as Erzincan, it contains markedly fewer structures in the study's building inventory. This difference points to a high-density fabric woven from few but compact, contiguous and multi-household structures, a reading confirmed quantitatively in Section 3 [37–39].

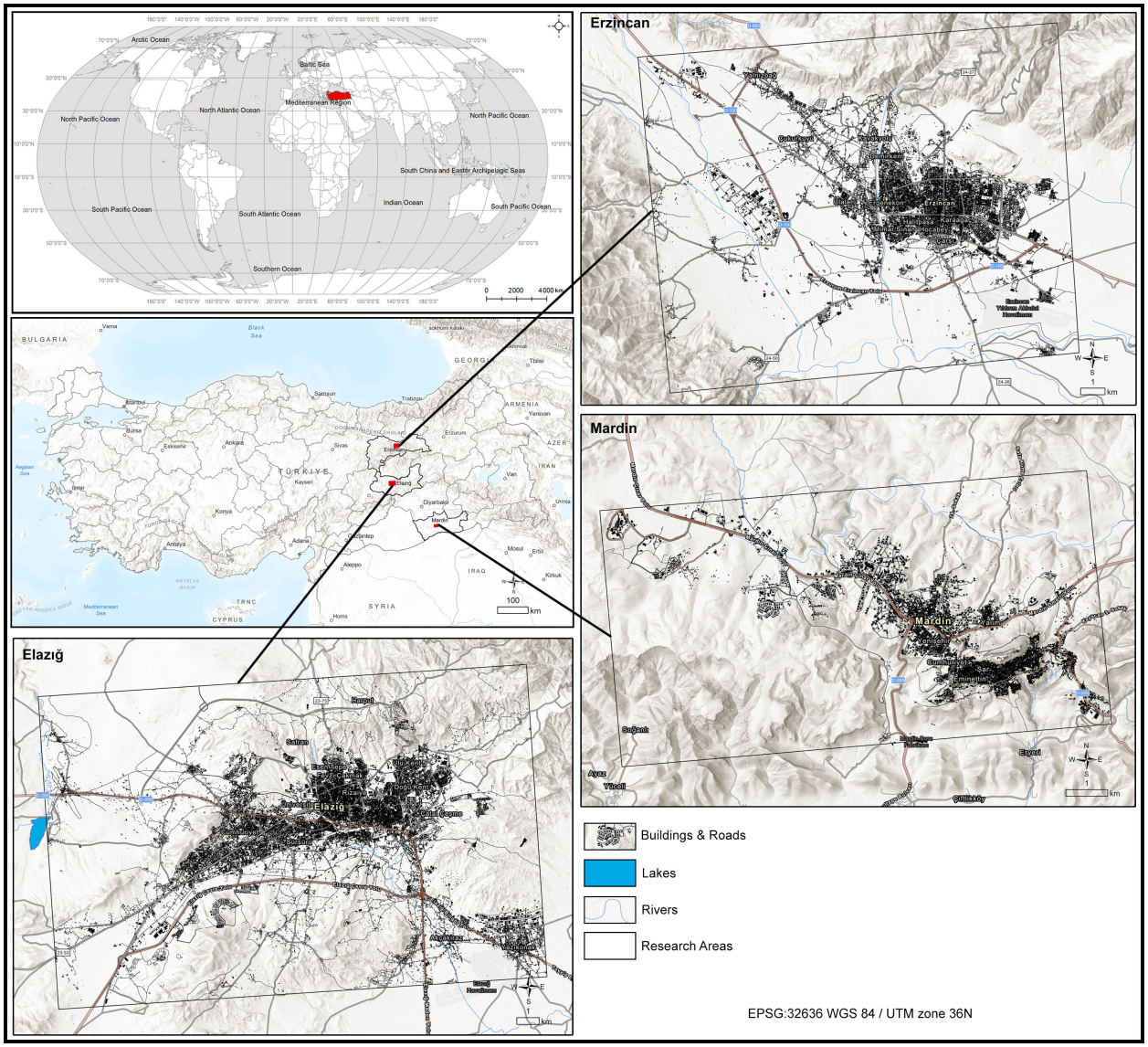


Figure 1. Location map of the study area. Place and road names follow the source basemap and are retained in Turkish.

2.2. Data Sources

The spatial data comprise two core vector datasets for each city: building footprints and roads (Table 1).

Table 1. Summary of the analysis dataset.

City	Building Count	Road Segments	Total Road (km)	Projection
Elazığ	32,209	13,272	1634.3	EPSG:32636
Erzincan	26,255	7727	968.0	EPSG:32636
Mardin	9584	2632	290.6	EPSG:32636

2.2.1. Building Footprint Data

The building footprints were compiled from the open-access GlobalBuildingAtlas (GBA) dataset produced by Zhu et al. [40]. GBA extracts building geometries from PlanetScope satellite imagery through deep learning-based segmentation and integrates them with open-source polygon collections such as OSM and Microsoft. It thereby provides building height estimates and LoD1-level three-dimensional model information at a global scale. Across continents, the height estimation has been reported with root-mean-square-error (RMSE) values ranging from 1.5 to 8.9 m.

Three considerations motivated the choice of GBA over local cadastral data. First, the dataset includes building height and so directly supplies the input required for three-dimensional analyses of urban form based on volume and floor count estimation. Second, it offers a single, consistent source across all study cities, minimising the systematic errors that data heterogeneity introduces into inter-city comparison. Third, its open licence and versioned structure satisfy the standards of academic reproducibility.

During pre-processing, MultiPolygon geometries were first decomposed into single polygons, invalid geometries were repaired with the buffer(0) method, and empty records were removed. Structures with a footprint below 25 m^2 (barracks, sheds and similar temporary elements) were excluded. Overlapping building polygons were then detected automatically. Where the overlap ratio exceeded 2%, the larger building was retained. A smaller building was deleted entirely when its overlap covered more than 80% of its own area, whereas partial overlaps were resolved through clipping. To prevent the topological errors that coincident centroids can cause in the tessellation step, a random offset of ± 0.1 m was applied to each overlapping pair. Finally, the data were transformed from the source EPSG:3857 system to UTM Zone 36N (EPSG:32636), the common metric projection for all three cities, and the pre-processed building layer was checked visually against current satellite imagery.

2.2.2. Street Network Data

The street network data for the three cities were obtained from the open geospatial infrastructure of the Overture Maps Foundation (OMF). Within it, the segments layer under the transportation theme provides the primary geometric input for urban network analysis. Although contributions from the OpenStreetMap community supply the main input to the harmonisation process, the platform also incorporates geometry corrections from proprietary navigation databases and automated topological integration steps [41]. Operated through a partnership of Amazon Web Services, Meta, Microsoft and TomTom, this multi-source platform undergoes regular quality control.

Two considerations made Overture Maps preferable to OSM. The first is the segment gaps observed in OSM. The second is that the topological consistency steps in OMF's automated workflow (node–edge integration and the removal of duplicate geometries) more reliably secure the data homogeneity essential to comparative inter-city analysis.

Before the graph was built, the street geometries were prepared for space syntax analysis with the roadcleaner tool of the QGIS (version 3.40) Space Syntax Toolkit (version 0.3.10) [42] and then passed to the Python-based pipeline. There, only LineString geometries were retained, and coordinate precision was rounded to a 0.1 m grid to resolve

topological disconnections. The cleaned geometries were converted into an intersection-based primal graph with the momepy library (version 0.9.1). To ensure compatibility with the betweenness centrality and space syntax computations, the graph was then cast as a directed multigraph (MultiDiGraph). Missing edge lengths were filled by deriving them from the geometry length and, failing that, from the Euclidean distance between the terminal node coordinates. As a final step, the largest weakly connected component was selected and disconnected fragments were dropped from the analysis.

2.3. Methodological Framework

Designed as a six-stage, modular and reproducible computational pipeline, the analytical process runs from data acquisition through spatial typology classification to visualisation (Figure 2). The first stage acquires street network data [41] and building footprints [40] for the three cities. These are then pre-processed in the second stage, covering geometric validity checks and coordinate system transformations. The third stage, which forms the analytical core, comprises five branches executed concurrently: (i) street network metrics, (ii) space syntax analysis, (iii) morphological tessellation and coverage area ratio (CAR), (iv) building morphology metrics, and (v) neighbourhood and adjacency indicators. The remaining stages are detailed in the subsections below.

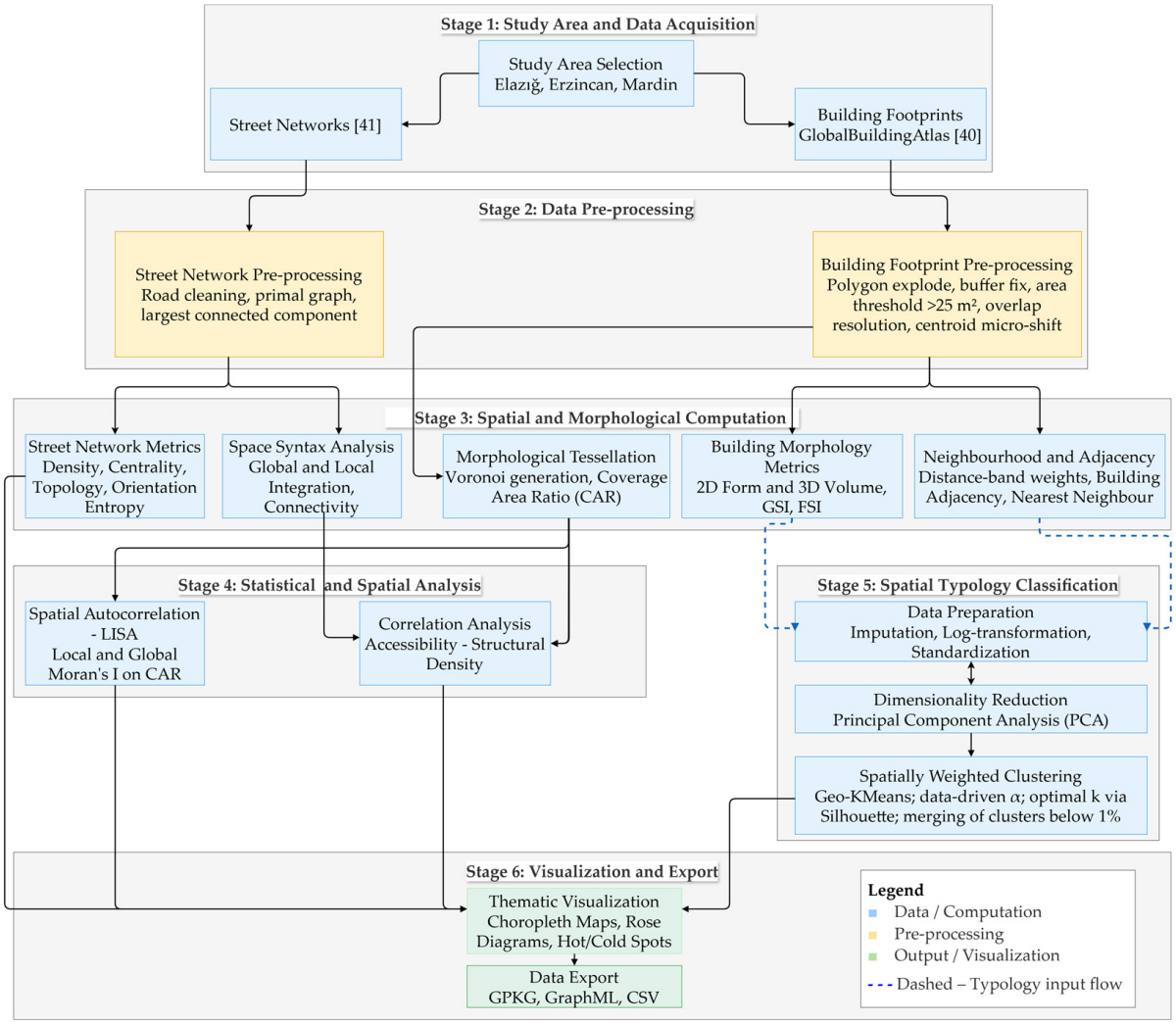


Figure 2. The six-stage urban morphology analysis workflow. Input data: building footprints [40] and street network data [41].

2.3.1. Street Network Topology and Space Syntax

The topological and geometric properties of the street network were analysed with OSMnx (version 2.0.7) for graph construction and pre-processing [18] and with NetworkX (version 3.4.2) for metric computation [43]. Basic connectivity was assessed through intersection count and areal density (nodes/km^2), street segment count, mean segment length, total street length density (km/km^2) and mean node degree (k_avg). To capture circulation patterns, circuity, the ratio of actual road length to the Euclidean distance between two endpoints, was determined together with the proportions of dead ends and four-way intersections. Orientational diversity was then measured by summarising each segment's bearing angle, length-weighted, in a 36-bin (5°) polar histogram. From this distribution, Shannon entropy (H) and the grid order index derived from it ($O = 1 - H/\log n$) were computed [10,30].

To represent potential movement flows and the distribution of centrality across the network, betweenness centrality was computed following Porta et al. [19]. To keep the computational load manageable in large networks, the sample size was set to $k = \min(\max(200, n \div 20), 500)$. The results were normalised, and a fixed random seed (seed = 42) ensured reproducibility. The distributional inequality of betweenness across the network was then summarised with the Gini coefficient.

Circuit structure and topological complexity were assessed with the three indices defined by Garrison and Marble: alpha, the circuit ratio ($\alpha = \mu/(2n - 5)$); beta, the edge–node relationship ($\beta = e/n$); and gamma, the ratio of observed edges to the theoretical maximum ($\gamma = e/3(n - 2)$) [44]. Here e denotes edges, n nodes and μ the number of independent circuits.

Spatial accessibility and integration were assessed within the space syntax theory, whose foundations were laid by Hillier and Hanson [45]. The graph was rendered undirected and the analysis confined to the largest connected component. To gauge the macroscale movement potential, the mean topological depth (mean depth, MD) of each node was computed. MD was first converted into Relative Asymmetry ($RA = 2(MD - 1)/(n - 2)$) and then divided by the reference graph coefficient (D-value), which standardises independently of system size, to yield Real Relative Asymmetry (RRA = RA/D). Because global integration is defined as the inverse of RRA (1/RRA), high integration corresponds to a topologically shallower, more accessible network [45,46].

For local scale accessibility, Penn's approach was adopted [47]. A Breadth-First Search (BFS) defined the accessible sub-network of each node within topological step radii of R = 3 and R = 5. For each sub-network, mean topological depth was obtained by dividing the summed depths to the node by ($n_r - 1$), and local integration was then derived using the D-value specific to that sub-network's size. Edge-level values were assigned as the arithmetic mean of an edge's two terminal node values. Connectivity was computed as normalised degree centrality, each intersection's neighbour count divided by ($n - 1$).

2.3.2. Morphological Character at the Building and Fabric Scale

Building size was quantified through floor area and the longest axis length (LAL), both derived directly from the geometry with momepy. Formal complexity and regularity were then captured by four complementary measures. Circular compactness expresses the circularity of the form as the ratio of the building's area to that of a circle with the same perimeter. A high value signals proximity to a circle, a low value an irregular or elongated shape [48]. The Equivalent Rectangular Index (ERI) reflects how closely the building's area–perimeter relationship matches that of a theoretical rectangle [11]. Finally, the aspect ratio of the minimum rotated rectangle records the building's mode of elongation, while elongation itself measures the extent to which the form extends along its main axis [14].

To quantify the vertical dimension, the height and floor count of each building were determined. Where height was not available directly in metres, it was derived from the floor count using a fixed coefficient (3.0 m/floor), and volume followed from the product of floor area and height. In line with the density framework of Berghauser Pont and Haupt [49], two related measures were derived: the Ground Space Index (GSI), the ratio of floor area to tessellation cell area, and the Floor Space Index (FSI), the product of GSI and floor count. Floor count itself is also reported separately as a vertical density proxy, since no parcel-level floor-area-to-plot-area ratio was computed.

To measure the spatial relationship between building footprints and the surrounding open space without relying on parcel data, the morphological tessellation method of Fleischmann et al. was applied [50]. The method generates, from the footprints, continuous Voronoi polygons that represent each building's catchment area. Geometric stability is secured through a 100 m buffer boundary, a 0.4 m shrink distance and a 0.1 m segment tolerance. For each cell, the coverage area ratio (CAR = floor area/cell area) was computed. Identical to GSI at the cell scale, this ratio captures the horizontal density of the built environment at high resolution.

To characterise the spatial relationships buildings establish with their immediate surroundings, a distance band weight matrix (DistanceBand) with a 100 m threshold was constructed using libpysal (version 4.13.0) [51]. For each building, adjacency was computed with momepy's BuildingAdjacency as the ratio of separate (non-contiguous) structures to the total number of buildings within the 100 m neighbourhood context; a value near 1 marks a fully detached fabric, while lower values mark attached, contiguous clusters. Neighbour distance, in turn, gives the mean spacing of buildings within the same context and complementarily describes the clustering pattern.

2.3.3. Spatial Autocorrelation of Urban Density (LISA)

To establish whether the urban density fabric is homogeneous, spatially clustered or random, the Local Indicators of Spatial Association (LISA) developed by Anselin were applied [52]. By quantifying the local spatial dependence of each observation unit separately, LISA identifies statistically significant clusters and outlier regions.

The input variable is the coverage area ratio (CAR) assigned to the morphological tessellation cells. CAR is a dimensionless ratio of building footprint area to cell area that expresses horizontal settlement density and takes a value between 0 and 1 for each cell. The cells were generated by the method of Fleischmann et al. with a 100 m buffer zone, a 0.4 m shrink distance and a 0.1 m segment tolerance, and each cell served as an observation unit [50].

To define the spatial interaction, a Queen contiguity weight matrix was constructed with libpysal and row-standardised [51]. The Queen matrix is a binary structure linking every pair of cells that share an edge or a corner. The Local Moran's I statistic was then computed over the CAR values with the esda library (version 2.7.0). Significant cells fall into four types: High–High (HH) hot spots, where the dense fabric is continuous, Low–Low (LL) cold spots, where the sparse fabric forms integrated areas and High–Low (HL) and Low–High (LH) spatial outliers, which depart markedly from their neighbours. Cells below the significance threshold were classified as non-significant (NS).

To anchor the local findings within an integrated frame, the Global Moran's I statistic was also computed. It summarises the spatial autocorrelation among all cell pairs in a single scalar: +1 denotes full clustering, 0 a random distribution and -1 full dispersion. Global Moran's I thereby provides a city-wide reference against which the local LISA results can be read.

2.3.4. The Relationship Between Accessibility and Built Density

To quantify the relationship between the topological accessibility of the street network and the physical density of the built environment, a correlation analysis was conducted. Because the network and building layers have independent scales and geometric bases, they were matched through a spatial join. Each building's centroid was assigned to the nearest street segment by Euclidean distance with the sjoin_nearest function of GeoPandas (version 1.1.3), transferring that segment's accessibility value to the building record. No distance upper bound was imposed, and where a match was ambiguous the first segment found was retained.

The study's independent variable is global integration, which represents the accessibility of the urban network. It rests on the shortest topological distances among all node pairs in the street network. The raw values are normalised to remove the scale effect of the total number of streets, yielding a standardised, city-wide Real Relative Asymmetry (RRA) index. A low RRA places a node at a shorter topological distance from the rest of the network, that is, at a high level of integration.

The dependent variables are two complementary indicators of urban density, horizontal and vertical. Horizontal density is captured by the coverage area ratio (CAR), the building footprint area divided by its own tessellation cell area. Vertical capacity is captured by the Floor Space Index (FSI), the total construction area (floor area $\times$ floor count) divided by that same cell area.

Two complementary coefficients were computed in parallel: the Pearson coefficient (r) gauges the linear relationship, whereas the Spearman rank coefficient (ρ) gauges the monotonic relationship free of distributional assumptions. Together they expose both the linear and the non-linear components of the accessibility–density pattern. Significance was tested at the two-tailed $p < 0.01$ level.

2.3.5. Multivariate Morphological Typology Inference

Eight building-level variables were included in the cluster analysis: floor area (floor_area), circular compactness (compact), coverage area ratio (car), building adjacency (adj), neighbour distance (neighbor_dist), fractal dimension (fractal_dim), building height (height_m) and floor count (floors). Missing values were completed with median imputation. A log(1 + x) transformation was applied to right-skewed variables such as floor area, CAR, neighbour distance and building height, and all variables were then standardised by Z-score (StandardScaler; scikit-learn, version 1.7.2). To remove multicollinearity and reduce dimensionality, Principal Component Analysis (PCA) was applied, with the number of components fixed dynamically at the point where the cumulative explained variance ratio first exceeded the 90% threshold.

To identify urban typologies without detaching them from their geographic context, a spatially weighted k-means (Geo-KMeans) model drawing on the ClustGeo approach was built [28]. Because classical clustering weighs only morphological similarity, it gathers similar buildings into the same class wherever they lie in the city. Geo-KMeans adds geographic location to that similarity, so that typologies form spatially coherent regions rather than scattered patches. In the model, each building's standardised X and Y coordinates were appended to its PCA component values, scaled by a spatial weight coefficient; as this coefficient grows, location carries more weight and the typology regions become more contiguous. The coefficient was not assigned heuristically: at a fixed cluster number (k = 6), a sweep that jointly weighed morphological homogeneity and spatial continuity identified the value maximising the balance of the two criteria as optimal.

The optimal number of clusters (k) was kept between 4 and 6 to preserve interpretability and was selected within this range by the Silhouette score, subject to a parsimony rule:

whenever increasing k by one improved the Silhouette score by less than 0.005, the smaller k was retained, since so marginal a gain in separation did not justify the added typological complexity. This score compares how close each building is to its own cluster against how far it is from the nearest neighbouring cluster, on a scale from −1 to +1; a high value marks well-separated clusters. The Davies–Bouldin and Calinski–Harabasz indices served as auxiliary measures when candidate solutions were compared [53,54]. As a final step, clusters comprising fewer than 1% of the total buildings were merged into the nearest neighbouring cluster on the basis of the Euclidean distance between cluster centroids.

2.3.6. Visualisation

To convey the spatial distributions and statistical patterns emerging from the quantitative analyses, more than 15 thematic outputs were produced with Matplotlib (version 3.10.9). These include a choropleth network map rendering the street hierarchy from betweenness centrality, a length-weighted polar rose diagram summarising street orientations, and four-panel maps of the building-level formal indicators. The four panels display, respectively, circular compactness, fractal dimension, ERI and convexity (φ).

The macro-scale morphological outputs comprise a two-panel map presenting tessellation cell area and CAR together; five line maps for the space syntax configurational-accessibility metrics (global integration, local integration R3 and R5, connectivity and mean depth); the LISA hot-/cold-spot map; and the Geo-KMeans typology classification map. The building height and floor count distributions, which represent three-dimensional urban density, were presented as a further four-panel group comprising a choropleth map, a histogram, a kernel density estimate (KDE) and a GSI–FSI scatter plot. All maps were saved at 300 DPI with a north arrow, a dynamic scale bar and descriptive legends. The colour palettes were drawn from colour-blind-friendly, perceptually distinct options: turbo for betweenness centrality, inferno for space syntax integration and viridis for connectivity.

To ensure transparency and reproducibility, the analyses were implemented as a modular software pipeline in a Python (version 3.10.11) environment. The pipeline comprises six modules: computation of network and building metrics, topological accessibility analysis, spatial autocorrelation (libpysal, esda), dimensionality reduction and machine learning-based clustering, statistical correlation and academic visualisation (Matplotlib). All spatial outputs were saved in the lossless, interoperable GeoPackage (GPKG) format, topological structures in GraphML and summary statistics as CSV tables. Two stochastic steps, the multiple k-means initialisations (n_init = 20) and the LISA permutation tests (999 permutations), were fixed with a single seed (random_state = 42) so that the results reproduce exactly.

3. Results

3.1. Street Network Topology and Space Syntax: Patterns of Connectivity, Centrality and Accessibility

Assessed through Boeing's topological and geometric indicators, the three street networks display a clear gradation in connectivity and circulation (Figure 3, Table 2) [10,30]. Elazığ has the densest network: an intersection density of 28.82 nodes/km^2 and a street network density of 5.41 km/km^2 are the highest among the three cities, while a relatively low dead-end proportion (8.8%) reflects the lightly branching, cyclical character of its circulation. Erzincan, though numerically sparser, is topologically the most reticular. Its highest mean node degree (3.09), lowest dead-end proportion (7.5%) and highest four-way intersection proportion (23.2%) expose the circuit structure that the regular right-angled grid core lends to the network. Mardin presents the opposite picture: the mean node degree falls to 2.85, the dead-end proportion rises to 16.5%, and the four-way intersection

proportion remains lowest at 15.8%. In combination, these values point to a dead-end-oriented, tree-like hierarchy rather than a looped one.

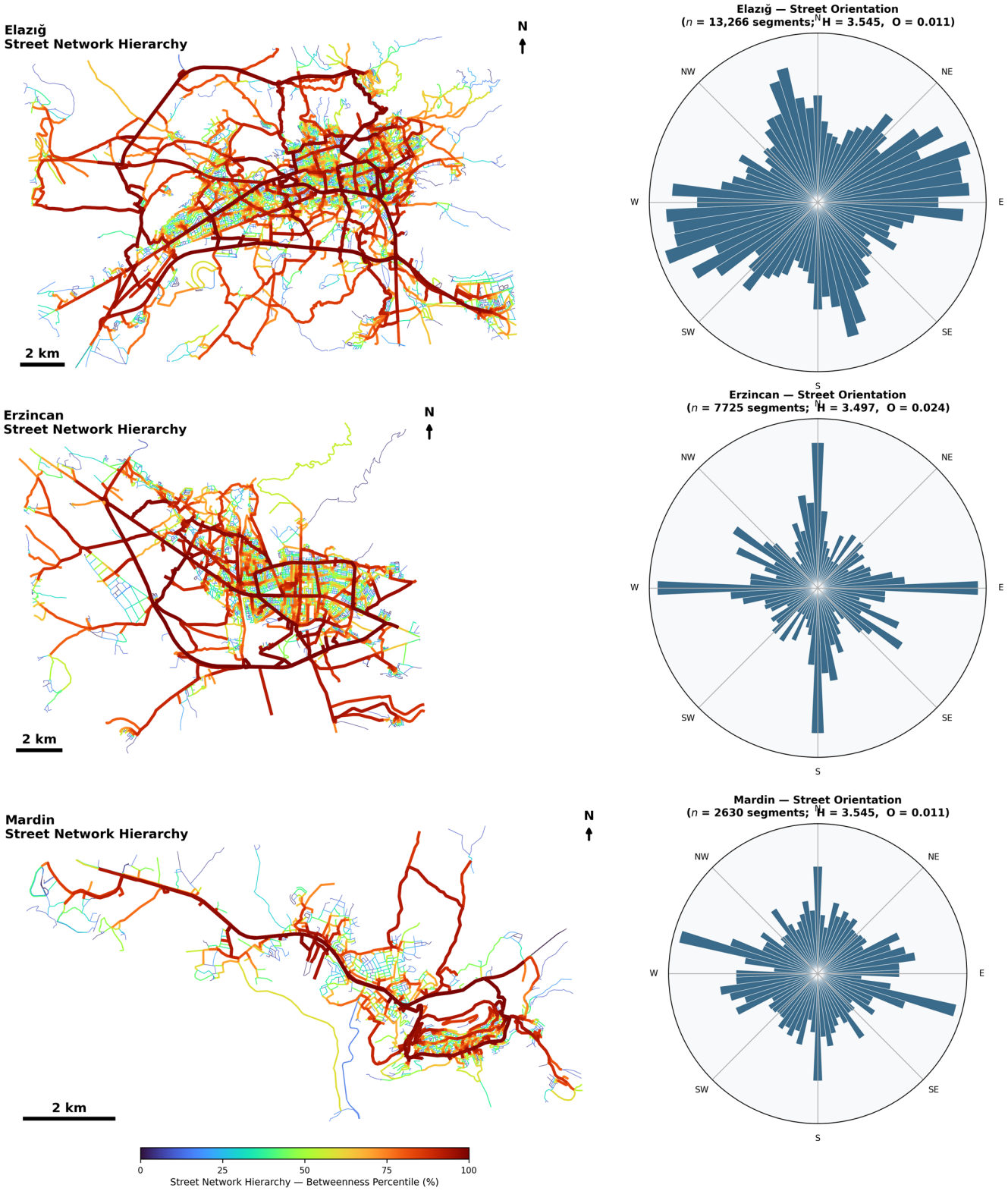


Figure 3. Comparative street network analysis of Elazığ, Erzincan and Mardin. The left column maps the street network hierarchy, with each edge coloured and weighted by its betweenness centrality percentile (0–100%; turbo colormap) to highlight the primary movement structure of each city. The right column shows the corresponding length-weighted orientation rose diagrams, reporting orientation entropy (H) and order (O) for each network.

Table 2. Street network topology and space syntax metrics for the three study cities (Elazığ, Erzincan and Mardin).

Metric	Elazığ	Erzincan	Mardin
Network size and geometry			
Intersection count	8708	5000	1844
Street segment count	13,272	7727	2632
Total street length (km)	1634.3	968.0	290.6
Median segment length (m)	72.4	76.0	57.3
Mean segment length (m)	123.1	125.3	110.4
Intersection density (per km^2)	28.82	21.67	26.04
Street density (km per km^2)	5.41	4.20	4.10
Connectivity and circulation			
Average node degree, k_avg	3.05	3.09	2.85
Dead-end proportion	0.088	0.075	0.165
Four-way intersection proportion	0.204	0.232	0.158
Circuity, average	1.031	1.033	1.076
Orientation			
Orientation entropy, H_o (nats)	3.545	3.497	3.545
Orientation order, O	0.011	0.024	0.011
Topological indices (Garrison and Marble)			
Alpha (α)	0.262	0.273	0.214
Beta (β)	1.524	1.545	1.427
Gamma (γ)	0.508	0.515	0.476
Centrality [19]			
Mean betweenness centrality	0.0044	0.0062	0.0119
Betweenness Gini	0.819	0.795	0.796
Mean closeness centrality	0.0260	0.0316	0.0445
Space syntax (Hillier and Hanson; Penn)			
Global integration (RRA-based)	0.280	0.317	0.386
Mean depth (topological steps)	39.17	32.03	23.26
Local integration, R = 3	1.432	1.448	1.351
Local integration, R = 5	1.132	1.140	1.084
Connectivity (degree centrality)	0.000349	0.000615	0.001540

Indicators of geometric regularity reinforce this distinction. Circuity, the ratio of route length to Euclidean distance, peaks in Mardin at 1.076, whereas Elazığ (1.031) and Erzincan (1.033) offer far more linear connections. Orientation entropy, which measures the diversity of street bearings, runs in all three cities close to the theoretical upper bound of the 36-bin polar histogram (≈3.58 nats) (Elazığ and Mardin 3.545; Erzincan 3.497), so no network carries a dominant grid orientation. The grid order index, however, is roughly twice as high in Erzincan (0.024) as in Elazığ and Mardin (both ≈0.011), which confirms that the most regular orientational structure belongs to Erzincan. Garrison and Marble's topology indices preserve the same ordering: alpha (Erzincan 0.273; Elazığ 0.262; Mardin 0.214), beta (1.545; 1.524; 1.427) and gamma (0.515; 0.508; 0.476) are all highest in Erzincan and lowest in Mardin [44]. In sum, Elazığ presents a densely connected, Erzincan a reticular grid, and Mardin a sparse, tree-like topological organisation.

Examined through betweenness and closeness centrality after Porta et al., the raw means interact strongly with network size (Figure 3, Table 2) [19]. Mardin records the highest mean betweenness (0.0119) and closeness centrality (0.0445) of the three cities. Much of this advantage is scalar rather than structural: both measures rise as a network shrinks and its mean topological distance falls, so Mardin's favourable position reflects its smaller size far more than any genuine centrality advantage. In Elazığ, the most extensive

network, mean betweenness (0.0044) and closeness centrality (0.0260) remain the lowest, as expected, while Erzincan (0.0062; 0.0316) falls between the two extremes.

By contrast, the Gini coefficient, which captures the inequality in the distribution of movement potential, offers a more discriminating view [30]. On this measure the ordering reverses: Elazığ rises to the highest value at 0.819, whereas Mardin (0.796) and Erzincan (0.795) stay close together at relatively low levels. Through movement therefore concentrates on a few dominant axes in Elazığ, while the potential spreads more evenly across Erzincan and Mardin. Considered together, the two readings underline that the network size effect and genuine structural concentration must be kept apart when centrality is interpreted.

At the global scale, within Hillier and Hanson's space syntax theory, the expected inverse relationship between global integration and mean topological depth holds in all three cities (Figure 4, Table 2) [45]. Mardin, with the lowest mean depth (23.26 steps), attains the highest global integration (0.386), followed by Erzincan (32.03; 0.317); Elazığ, with the most extensive network (39.17; 0.280), records the lowest. This ordering matches the closeness centrality pattern exactly (Mardin > Erzincan > Elazığ), which mathematically expresses the inverse of mean depth. The two independent indicators converge on the same ordering, Mardin highest and Elazığ lowest, although this apparent advantage is partly a network size effect: because global integration is itself sensitive to network size, Mardin's high value reflects its compact extent as much as any structural difference.

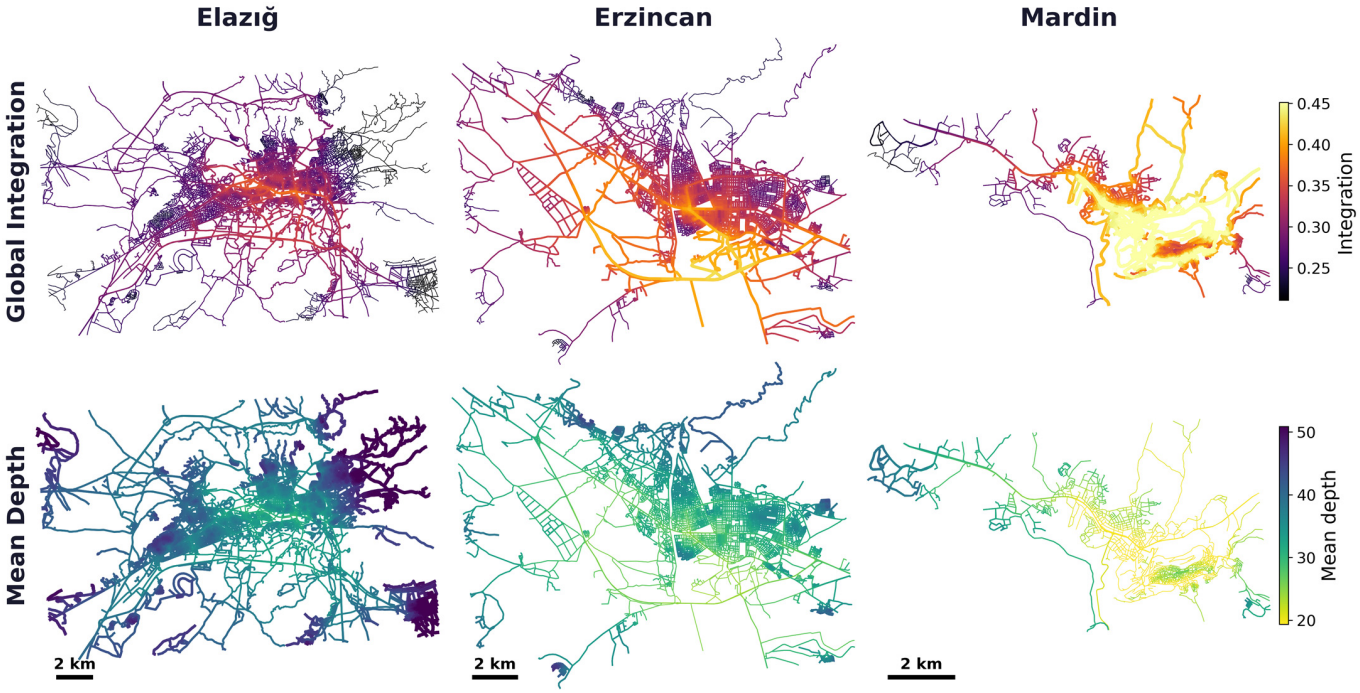


Figure 4. Comparative space syntax maps of global integration and mean depth across Mardin, Erzincan and Elazığ. **Upper row**: global integration (edge-level, inferno scale); **lower row**: mean depth (viridis_r scale). Colour ramps are globally calibrated to the 2nd–98th percentile of pooled values across all three cities to ensure cross-city comparability. Line width varies proportionally with the metric value. Computed via a node/segment topological approach on the primary street network [19]; absolute values are not directly comparable with axial-map-based studies.

At the local scale the picture changes markedly. For local integration computed at the R = 3 and R = 5 topological radii after Penn, the highest scores belong to Erzincan (R3 = 1.448; R5 = 1.140) [47]. Elazığ runs close behind (1.432; 1.132), whereas Mardin records the lowest local integration (1.351; 1.084). Mardin, the most advantaged city globally, thus fails

to hold that advantage locally: Erzincan's reticular grid fabric sustains the formation of local sub-centres, while Mardin's compact structure limits local accessibility. The R3 > R5 > global integration hierarchy holds in all three cities (for Mardin, 1.351 > 1.084 > 0.386).

Connectivity, based on degree centrality, is the highest in Mardin (0.00154) and the lowest in Elazığ (0.000349). The raw mean node degree is the lowest in Mardin (2.85), yet normalised connectivity lifts the same city to the top. This is not a contradiction but a scaling effect: because degree centrality is inversely sensitive to network size, Mardin's high value again reflects its compact extent. Across all three cities, the high-integration corridors coincide with the urban main axes. The spatial logic of that coincidence nonetheless varies: in Elazığ integration concentrates on a radial spine, in Erzincan it forms a more diffuse and balanced network of local centrality, and in Mardin the compact structure underwrites global accessibility while constraining local organisation. Set side by side, these space syntax patterns (Figure 5) show how decisive the choice of scale is for morphological interpretation.

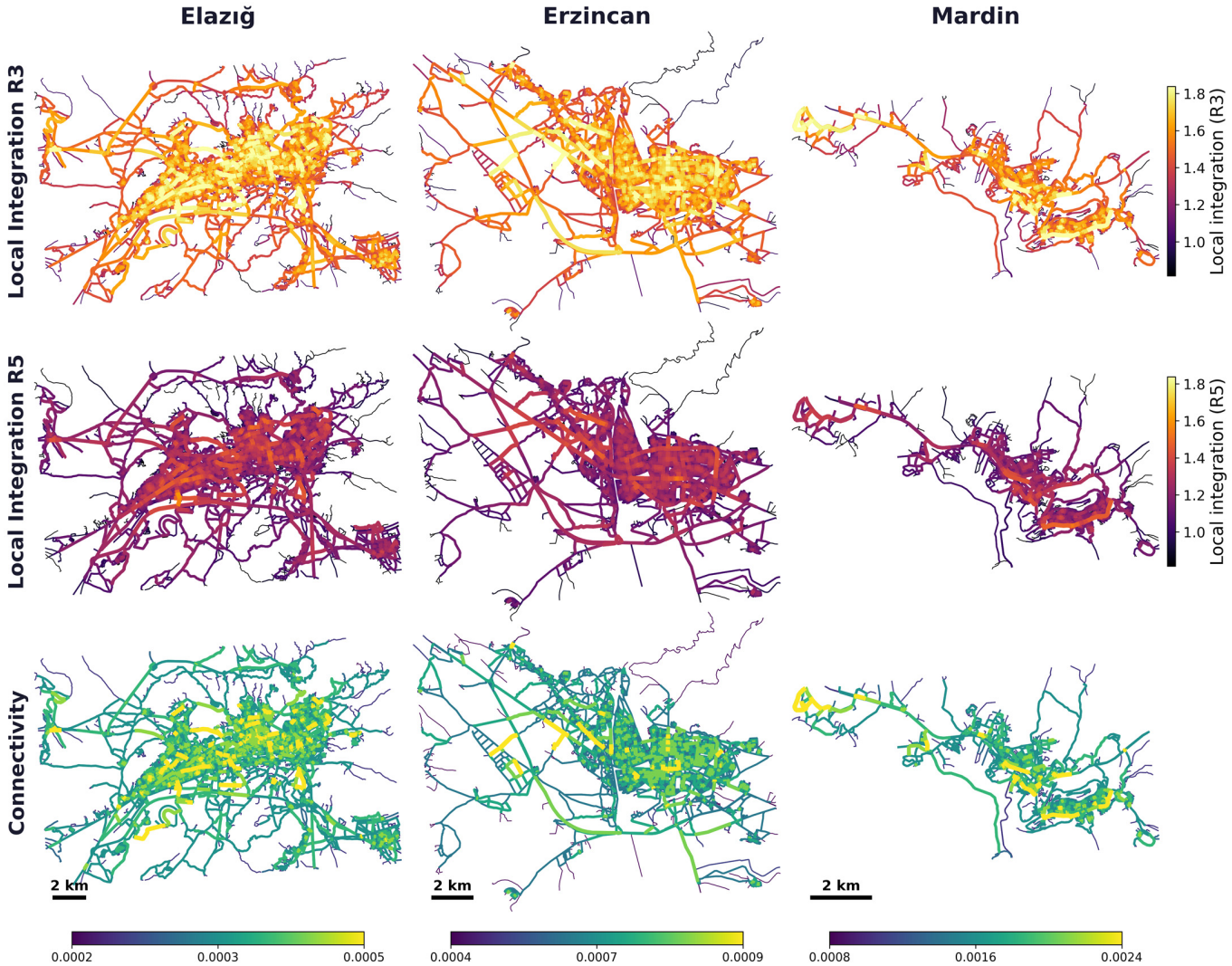


Figure 5. Comparative space syntax maps of local integration (radii R3 and R5) and connectivity across Mardin, Erzincan and Elazığ. Rows from top to bottom: local integration R3, local integration R5, and connectivity (normalised degree centrality). Colour ramps are globally calibrated to the 2nd–98th percentile of pooled values; line width varies with value for the integration rows and is uniform for connectivity. All metrics were computed on the primary street network using a topological segment approach [19,47].

3.2. Morphological Character at the Building and Fabric Scale: Form–Size Metrics, Morphological Tessellation and Neighbourhood Relations

Compared through the two-dimensional form–size indicators computed with Fleischmann's momepy toolkit, the building scale reveals a two-tier pattern: inter-city differentiation is strong along the size axis but only moderate along the shape axis (Figure 6, Table 3) [14]. Size diverges most sharply. The median floor area is the highest in Elazığ (189.4 m^2), intermediate in Erzincan (160.9 m^2) and the lowest in Mardin (114.3 m^2), and the longest axis length follows the same order (20.8 m, 19.1 m and 16.4 m, respectively).

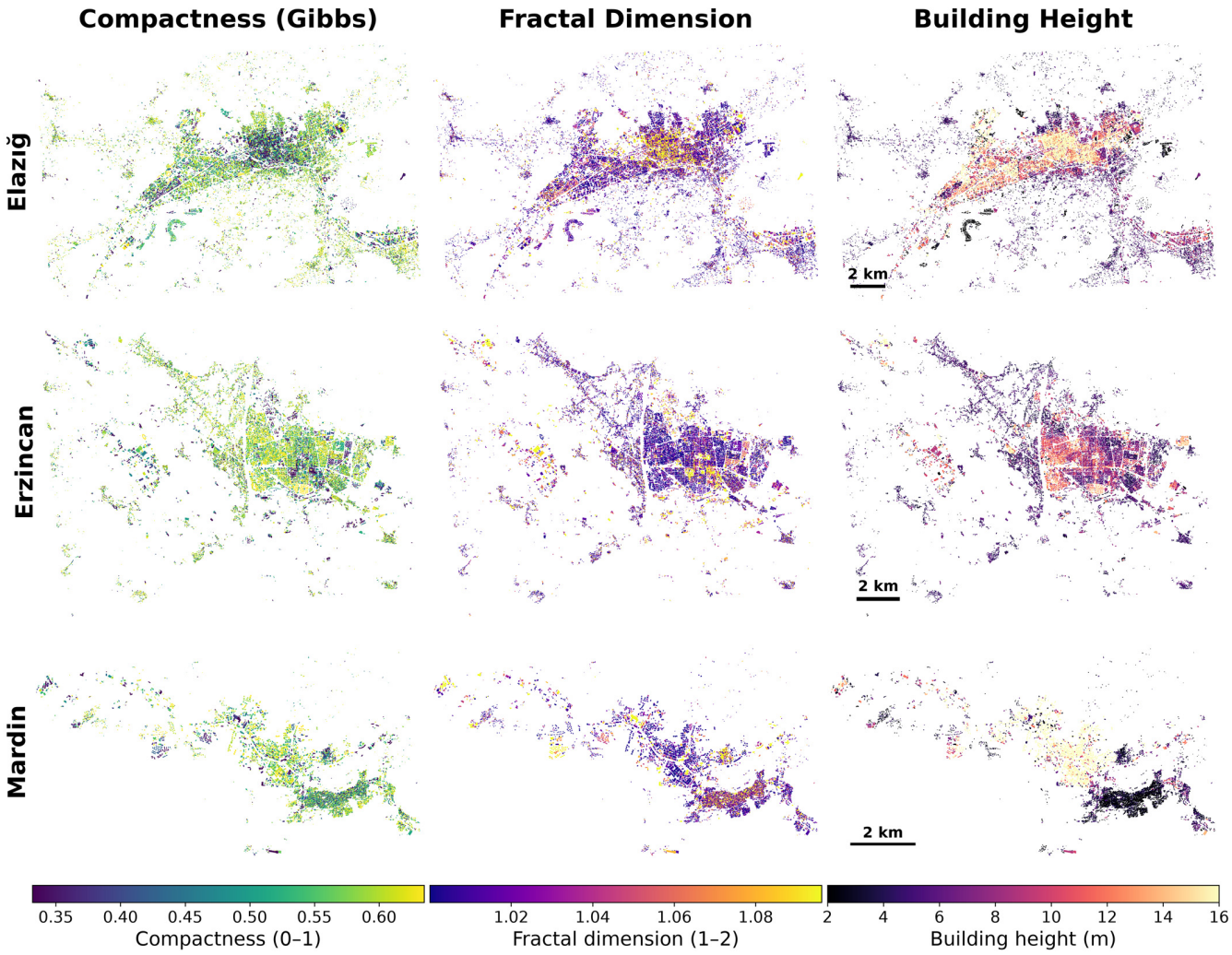


Figure 6. Building-scale morphological atlas of Elazığ, Erzincan and Mardin. Rows correspond to the three cities and columns to three discriminating building-level metrics: circular compactness (Gibbs), fractal dimension and building height. Each footprint is shaded by its metric value, and the colour scales are shared column-wise (a common 2nd–98th-percentile range across the three cities) to enable direct inter-city comparison. Metrics were computed with momepy [14,50].

Formal regularity, by contrast, diverges far less. The medians of the Equivalent Rectangular Index (ERI), which gauges the footprint's area–perimeter agreement with a theoretical rectangle, and of convexity (PHI) reach saturated values (≈1.000) in all three cities, and the median corner count is fixed at 4 throughout. The ERI-MRR, the short/long side ratio of the minimum rotated rectangle, is likewise almost identical across the cities, within the 0.706–0.713 range [11].

Table 3. Building- and fabric-scale morphological metrics (city-wide medians) for Elazığ, Erzincan and Mardin.

Metric	Elazığ	Erzincan	Mardin
Building count	32,209	26,255	9584
2D shape–size (median)			
Floor area (m^2)	189.4	160.9	114.3
Longest axis length (m)	20.8	19.1	16.4
Circular compactness (Gibbs)	0.573	0.581	0.557
Elongation ratio	0.706	0.713	0.710
Equivalent Rectangular Index, ERI	1.000	1.000	1.000
MRR aspect ratio, ERI-MRR	0.706	0.713	0.710
Convexity, PHI	1.000	1.000	1.000
Fractal dimension	1.012	1.010	1.016
Corner count	4	4	4
3D vertical/volumetric (median)			
Building height (m)	5.0	5.0	3.0
Floor count	1.7	1.7	1.0
Building volume (m^3)	1044.4	825.4	310.7
Ground Space Index, GSI	0.141	0.165	0.284
Floor Space Index, FSI	0.255	0.286	0.333
Height coeff. of variation (CV)	0.685	0.486	1.404
Height Shannon entropy (nats)	1.128	0.864	0.833
Morphological tessellation (median)			
Cell area (m^2)	1972.0	1208.0	471.5
Coverage area ratio, CAR	0.141	0.165	0.284
Neighbourhood/adjacency (median)			
Building adjacency	1.000	1.000	0.859

At the median, then, the footprints are predominantly quadrilateral and regular. Circular compactness and fractal dimension, however, trace a consistent if subtle pattern: compactness is the lowest (0.557) and fractal dimension the highest (1.016) in Mardin, whereas the most compact (0.581) and least fractal (1.010) structures occur in Erzincan [20,48]. Because "low compactness + high fractal dimension" denotes an irregular boundary, Mardin's geometries have the most irregular contour and Erzincan's the most regular. These formal differences remain numerically small (≈0.02 in compactness, ≈0.006 in fractal dimension), so the variable that truly distinguishes the cities at the building scale is size far more than formal complexity.

The three-dimensional character shows that the cities pursue fundamentally different routes to density (Table 3). In vertical development, Mardin separates clearly from the other two: the median building height is 5.0 m (≈1.7 floors) in Elazığ and Erzincan but falls to 3.0 m (1.0 floor) in Mardin, and the median volume is 1044.4 m^3, 825.4 m^3 and only 310.7 m^3, respectively. The density indicators of Berghauser Pont and Haupt expose the mechanism behind this divergence: the balance between horizontal occupancy and vertical density [49]. Among these, the Ground Space Index (GSI), the share of the tessellation cell occupied by the building, is the highest in Mardin (0.284) and the lowest in Elazığ (0.141), and the Floor Space Index (FSI), the product of GSI and floor count, is again the highest in Mardin (0.333; Erzincan 0.286; Elazığ 0.255). Floor count itself reverses the order exactly, at 1.7 in Elazığ and Erzincan and 1.0 in Mardin. Together these values reveal that Mardin reaches its relatively high structural density through horizontal occupancy rather than height, whereas Elazığ and Erzincan offset low ground coverage with additional floors.

Mardin, in short, follows a "compact but low" regime of densification, and Elazığ and Erzincan a "sparse but tall" one.

For vertical diversity, the height coefficient of variation (CV; standard deviation over mean) and the Shannon entropy computed over 5 m height classes appear contradictory at first, yet read together they complement one another. CV is by far the highest in Mardin (1.404; Elazığ 0.685; Erzincan 0.486), whereas height entropy peaks in Elazığ (1.128 nats; Erzincan 0.864; Mardin 0.833). The two measures follow different logics: CV inflates the relative dispersion that a few comparatively tall structures create in low mean Mardin, while entropy rises in Elazığ, where the height classes are more evenly populated. Mardin's distribution is therefore "low mean–high relative deviation", Elazığ's a "multi-class balanced distribution", with Erzincan intermediate on both axes.

The morphological tessellation approach, which assigns each building a catchment area without requiring parcel data (using a 100 m buffer, 0.4 m shrink, and 0.1 m segment tolerance), quantifies this density divergence at the grain of the urban fabric (Figure 7) [50]. The median cell area is the largest in Elazığ (1972.0 m^2), intermediate in Erzincan (1208.0 m^2), and the smallest in Mardin (471.5 m^2). Because small cells mark a dense, closely set fabric, this ordering confirms that Mardin is the densest structure and Elazığ the sparsest. The coverage area ratio (CAR), calculated as the building footprint area over the cell area and identical to the Ground Space Index (GSI) at the cell scale reinforces this pattern, standing at 0.284 in Mardin, 0.165 in Erzincan, and 0.141 in Elazığ. The horizontal-occupancy findings of the three-dimensional analysis are thereby confirmed on a parcel-independent basis: Mardin's "small cell + high occupancy" dynamic yields a compact fabric, whereas Elazığ's "large cell + low occupancy" results in a loose one.

The neighbourhood metrics, which directly gauge the detached or attached character of the fabric, were computed over a 100 m threshold neighbourhood with momepy and libpysal's DistanceBand method [14]. Building adjacency expresses the proportion of structures standing as independent units within that 100 m context: values near 1 mark a largely detached fabric, low values attached clusters. The median reaches 1.000 in Elazığ and Erzincan, and the very low standard deviations (0.022 and 0.010, respectively) confirm the internal consistency of this detached character. In Mardin the median falls to 0.859 while the standard deviation rises sharply to 0.317, as a substantial share of buildings form attached clusters and the internal diversity of the fabric reaches its maximum. Neighbour distance, the mean spacing of buildings within the 100 m band, varies within a narrow range: 44.7 m in Elazığ, 47.6 m in Erzincan and 50.1 m in Mardin. That Mardin combines the most attached fabric with the highest neighbour distance may seem contradictory, but the two indicators act at different scales: adjacency on direct contiguity and neighbour distance on an average across the 100 m band. Mardin's building stock, few in number (9584) and fragmented on sloping terrain, widens the band average despite tight local clusters. These direct measurements complete numerically the gradation anticipated from the tessellation: Mardin compact attached, Erzincan intermediate and Elazığ sparse detached.

3.3. Spatial Autocorrelation of Urban Density (LISA): Patterns of Concentration and Dispersion

The spatial clustering of the cell-level coverage area ratio (CAR) was tested with Global Moran's I on a Queen contiguity weight matrix (Figure 8, Table 4) [52]. In none of the three cities is density distributed at random. The values are strong and positive (Mardin 0.709, Elazığ 0.697, Erzincan 0.620), and all reach the $p = 0.001$ level. Mardin's highest value arises because the compact historic core forms a single, sharp density focus: low- and high-density areas separate cleanly, and similar cells gather into tidy district clusters. Elazığ clusters almost as strongly, since the density built up around the radial street spine produces a comparable pattern. Erzincan, by contrast, lags markedly: its grid-based, relatively regular

fabric spreads density across several small foci rather than one dominant centre, and its lower Moran's I marks a less focal, polycentric and gently graded organisation.

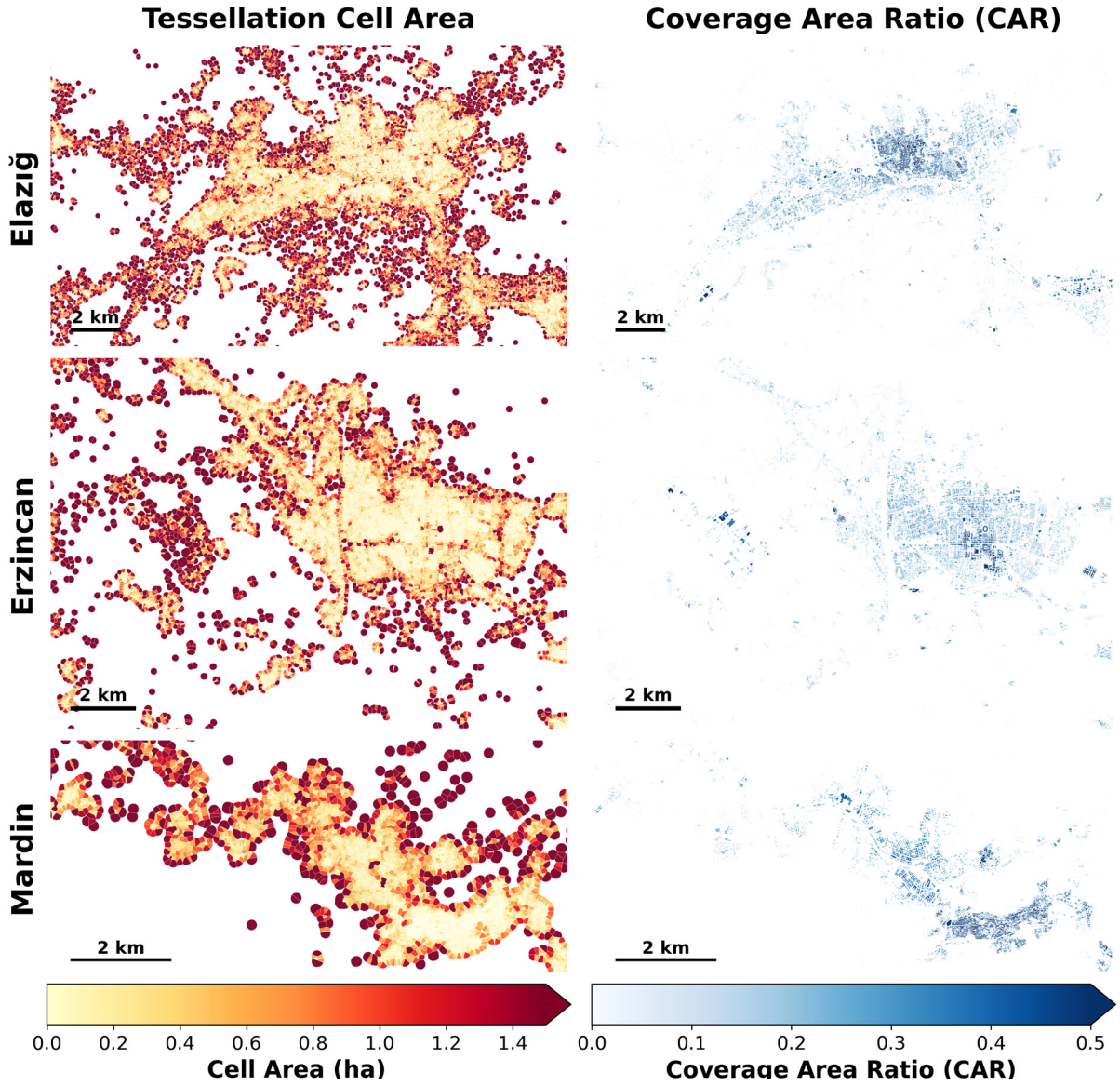


Figure 7. Morphological tessellation of Elazığ, Erzincan and Mardin (top to bottom). For each city, the left panel maps the tessellation cell area and the right panel the cell-level coverage area ratio (CAR), defined as building footprint area over cell area. Cells were generated from the building footprints with momepy (100 m buffer limit, 0.4 m inward shrink, 0.1 m segment tolerance). The colour scales are shared across the three cities and rendered as a single colorbar beneath each column, so that cell area and CAR are directly comparable between the cities [14,50].

Table 4. Global and local spatial autocorrelation of the coverage area ratio (CAR) in Elazığ, Erzincan and Mardin: Global Moran's I (with pseudo *p*-value) and the distribution of significant LISA cluster types (High–High (hot spots) (hot spot), Low–Low (cold spots) (cold spot), High–Low and Low–High (spatial outliers) (outlier) (outlier)) alongside non-significant cells.

City	Moran's I (CAR)	*p*	HH	LL	HL	LH	NS
Elazığ	0.697	0.001	18.0%	26.6%	0.6%	1.0%	53.8%
Erzincan	0.620	0.001	19.2%	22.0%	0.8%	1.2%	56.9%
Mardin	0.709	0.001	20.2%	25.6%	0.4%	0.8%	53.0%

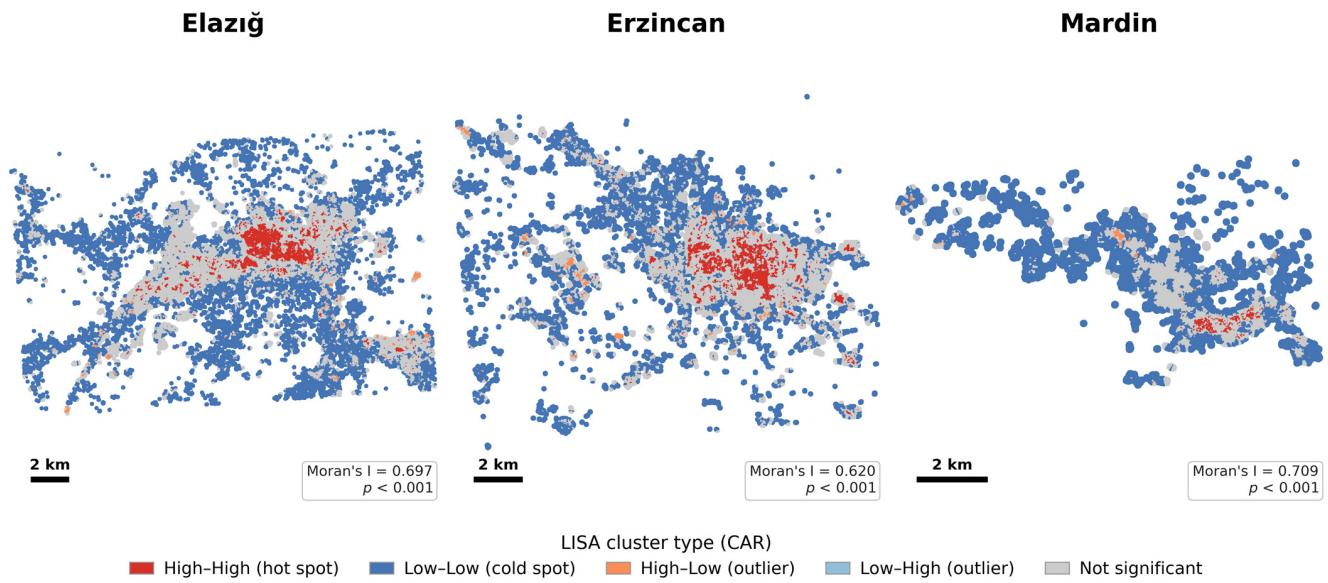


Figure 8. Local spatial clustering of the coverage area ratio (CAR): LISA hot/cold-spot maps for Elazığ, Erzincan and Mardin (top to bottom). Tessellation cells are classified as High–High (hot spots), Low–Low (cold spots), High–Low and Low–High (spatial outliers), or non-significant (NS), using Queen contiguity weights at $p < 0.05$ [52]. Each panel reports the Global Moran's I and its pseudo p-value, and the legends give the per-class cell counts.

Analysed through the LISA class distribution, the local patterns give the global gradation a spatial logic (Figure 8, Table 4). The share of high-density hot spots (High–High) is the highest in Mardin (20.2%), intermediate in Erzincan (19.2%) and the lowest in Elazığ (18.0%); on the maps these clusters gather in the urban core and coincide with the high-coverage, fine-grained centre fabric seen in Mardin's tessellation. The share of low-density cold spots (Low–Low), by contrast, peaks in Elazığ (26.6%), followed by Mardin (25.6%) and Erzincan (22.0%). This broad peripheral cold-spot cover matches Elazığ's largest cell area and sparsest edge fabric, confirming the city's sharp centre–periphery contrast quantitatively.

Spatial outliers (High–Low and Low–High cells) remain extremely rare in all three cities (Elazığ 1.5%; Erzincan 2.0%; Mardin 1.2% in total), so transitions between density clusters are sharp rather than gradual and anomalous density islands are few. Cells with no significant clustering (not significant) are the most frequent in Erzincan (56.9%) and the least frequent in Mardin (53.0%). Mardin thus commits the largest share of its density fabric to significant clusters, and Erzincan the smallest, an outcome consistent with Erzincan's lowest Global Moran's I and its polycentric, grid-based fabric.

3.4. The Relationship Between Accessibility and Built Density: Correlation Analysis

To relate the accessibility potential of the street network to built environment density, the two were matched at the building scale: each building's centroid was assigned to the nearest street segment (sjoin_nearest), carrying that axis's global integration value to the structure. No maximum distance threshold was applied: every building is matched to its nearest segment regardless of distance, since the street network covers the full extent of the built-up study area and unmatched buildings would otherwise have to be dropped. Integration was then correlated with two density indicators, CAR for horizontal occupancy and FSI for floor area density, using Pearson and Spearman coefficients (Table 5). A common pattern emerges. The accessibility–density relationship is positive in all three cities but structurally weak. For horizontal density (CAR), the Pearson coefficient is 0.170 in Elazığ, 0.118 in Mardin and only 0.050 in Erzincan. Even in Elazığ, the strongest case,

integration explains about 3% of the variance in CAR, and in Erzincan the share falls below 1%. Network centrality alone, therefore, does not determine horizontal building density.

Table 5. Correlation between street network accessibility (global integration) and built-up density (horizontal CAR and vertical FSI) in Elazığ, Erzincan and Mardin. Pearson r and Spearman ρ coefficients are reported with sample sizes (*N*); each building was linked to its nearest street segment by centroid (sjoin_nearest). All coefficients are significant at $p < 0.01$, although the effect sizes are weak.

City	*N*	Int ↔ CAR Pearson r	Spearman ρ	Int ↔ FSI Pearson r	Spearman ρ	*p*
Elazığ	32,209	0.170	0.148	0.247	0.199	<0.01
Erzincan	26,255	0.050	0.063	0.112	0.108	<0.01
Mardin	9584	0.118	0.109	0.129	0.124	<0.01

Accessibility's relationship with vertical density (FSI) follows the same city ordering and strengthens somewhat over CAR in all three cities, with Pearson coefficients of 0.247 in Elazığ, 0.129 in Mardin and 0.112 in Erzincan. Accessibility thus aligns with floor area density more than with horizontal occupancy: around more integrated axes, buildings tend to grow upward rather than outward. Even so, the tendency stays weak; in the strongest case (Elazığ, FSI) the explained variance reaches only about 6%. Overall, the two indicators give a consistent order: the strongest in Elazığ, the weakest in Erzincan on both density axes, with Mardin being intermediate. Elazığ's higher coefficients accord with density, and especially vertical density, gathering towards the main axes of its radial converging spine, and with the high betweenness Gini noted earlier. Erzincan's near-zero CAR correlation reflects a diffuse, polycentric grid in which density spreads across the base rather than binding to particular accessible axes.

Comparing the linear (Pearson) and monotonic (Spearman) coefficients exposes subtler structure within the relationship. The two families largely coincide, and the small inter-city gaps trace distributional character. In Elazığ, Pearson exceeds Spearman (0.170 versus 0.148 for CAR; 0.247 versus 0.199 for FSI), which suggests that extreme values carry part of the relationship, namely a handful of observations on the most integrated spine axes, where high accessibility meets high density. In Erzincan, by contrast, Spearman is very close to, and even slightly above, Pearson, indicating a weak tendency spread across the base rather than driven by extremes, in keeping with the city's diffuse grid. These deviations remain small and leave the overall picture of weakness intact.

The significance values, finally, call for caution given the sample size. Every coefficient is significant at $p < 0.01$, yet this rests on tens of thousands of observations (N: Elazığ 32,209; Erzincan 26,255; Mardin 9584), and in samples this large even very small coefficients reach significance. A further caveat compounds this: the buildings are spatially autocorrelated rather than independent observations, which violates the assumption behind the conventional *p*-value and inflates apparent significance beyond what sample size alone would produce. We therefore treat these correlations as descriptive and exploratory, not as a formal test of association and do not lean on their *p*-values as evidence of anything beyond the direction of the relationship. The result therefore attests to the existence and direction of the relationship, not its strength; effect size and statistical significance must be judged separately. Seen in this light, the picture is clear: all three cities show a detectable but practically weak and positive relationship. Highly integrated axes draw density to a degree, but its principal determinants lie largely off the network, and space syntax configurational accessibility explains only a small portion of the variance. A full account of the density pattern would require multivariate models incorporating off-network variables such as terrain slope, the zoning regime, the legacy of the historic core and distance to the centre.

3.5. Multivariate Morphological Typology Inference: Data Pre-Processing, Dimensionality Reduction (PCA) and Spatially Weighted Clustering

The building-level typologies were identified through a multi-stage flow of pre-processing, dimensionality reduction and spatial clustering (Table 6). Pre-processing is standard: missing and invalid values are filled by median imputation, right-skewed size and distance variables are symmetrised with a log(1 + x) transformation, and all variables are then standardised using Z-score. The analysis rests on eight variables: floor area, circular compactness, coverage area ratio (CAR), building adjacency, neighbour distance, fractal dimension, building height and floor count [14].

Table 6. Dimensionality reduction and clustering results for the building-level morphological typology of Elazığ, Erzincan and Mardin.

City	PCA (≥90%)	PC1/PC2 var.%	α* (Data-Driven)	Optimal k	Final Typology	Silhouette
Elazığ	5 components	37.1/23.0	1.75	6	6	0.245
Erzincan	5 components	34.8/26.0	1.75	5	5	0.240
Mardin	5 components	40.4/23.4	1.75	4	4	0.265

α*: the data-driven optimal spatial weight parameter selected for Geo-KMeans; the asterisk denotes the optimal value.

The standardised variables enter a Principal Component Analysis (PCA) that selects, dynamically, the number of components needed to explain at least 90% of the total variance [27]. Five components meet this threshold in every city (cumulative explained variance 92.1% in Elazığ, 93.2% in Erzincan, 94.2% in Mardin), and the pairwise correlations among the input variables (Figure 9) and the resulting loadings (Figure 10) both share a consistent structure. The first component (37.1% in Elazığ, 34.8% in Erzincan, 40.4% in Mardin) reads as a "size and vertical density axis", driven mainly by building height, floor count, floor area and coverage ratio. The second (23.0–26.0%) forms a "formal regularity axis" on which compactness and fractal dimension load in opposition. In Mardin specifically, building adjacency also loads strongly on the first component, which then polarises between a "large/tall and detached" pole and a "small/low, attached and high-coverage" one. This is the direct counterpart, in the reduced space, of Mardin's low-rise, high-coverage attached fabric noted in earlier sections.

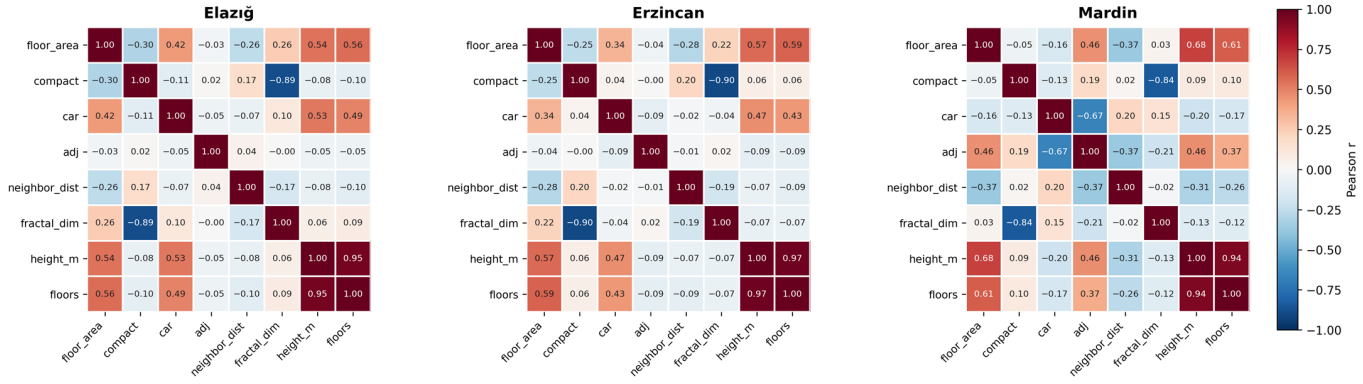


Figure 9. Pearson correlation matrix among the eight building-level clustering variables, by city (three-panel heat map; annotated Pearson r, diverging colour scale centred on zero). Floor area, CAR, neighbour distance and height are log1p-transformed prior to correlation, matching the transformation applied before clustering (Section 2.3.5).

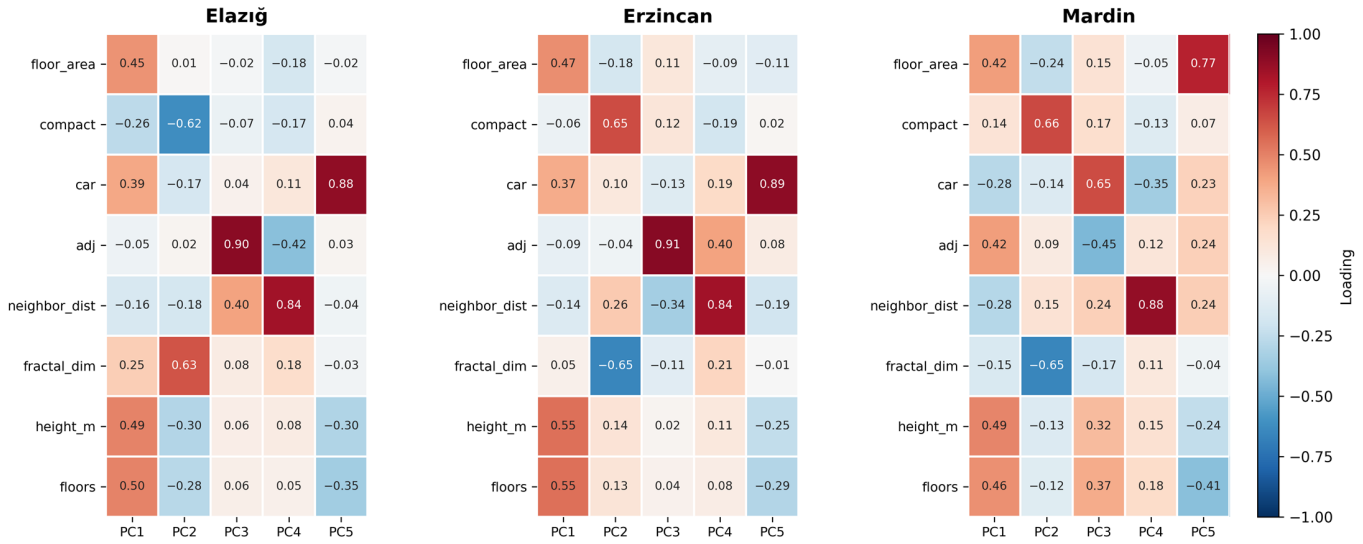


Figure 10. Principal component loadings for the eight building-level clustering variables, by city (five components retained per the 90% explained variance threshold; Section 2.3.5). Cell colour and annotated value both encode the loading; the diverging scale is centred on zero.

Clustering proceeds with the spatially weighted k-means (Geo-KMeans) model adapted from the ClustGeo logic, which appends the buildings' standardised coordinates to the PCA space through a spatial weight (alpha) [28]. This weight is set not by heuristic but by a data-driven criterion: at a fixed cluster number (k = 6), alpha is swept while the inertia the clustering explains in feature (PCA) space (Q0) and in coordinate space (Q1) is computed, and the value maximising the normalised Q0 + Q1 balance is adopted. It converges to $\alpha^* = 1.75$ in all three cities, and the near-flat balance curve over the 1.5–2.0 range shows the result to be robust to this parameter. The number of clusters is then chosen within the 4–6 range by the Silhouette criterion under the parsimony rule defined in Section 2.3.5, with the Davies–Bouldin and Calinski–Harabasz indices reported only for reference. Interpretability considerations cap the search at six, and clusters below 1% of the total are merged into the nearest centroid. In the final solution, the typologies number six in Elazığ, five in Erzincan and four in Mardin. The highest Silhouette value (0.265) belongs to Mardin, while Elazığ (0.245) and Erzincan (0.240) sit just below. In Erzincan, k = 6 attains a marginally higher Silhouette score (0.2451) than the adopted k = 5 (0.2403). The 0.0048 gain falls below the 0.005 parsimony threshold, so the five-cluster solution was retained as the more parsimonious representation. Elazığ and Mardin, by contrast, already reach their maximum Silhouette score within the search range at the adopted k. Mardin's smaller count signals that its fabric gathers into fewer, more sharply separated groups, the cluster separability reflection of the polarised morphology evident throughout (attached dense core ↔ detached sparse periphery).

Figures 9 and 10 make explicit the variable redundancy within the eight-variable input set. Height and floor count correlate at r = 0.94–0.97 across all three cities, which is expected given that floor count is itself derived from height (or vice versa) through the fixed 3.0 m/floor conversion whenever direct height data are unavailable (Section 2.3.2). The two variables load almost identically on the first principal component in every city, so they act as a single "size" dimension rather than two independent ones. Compactness and fractal dimension are similarly redundant, correlating at r = −0.84 to −0.90 and loading with opposite signs on the second component, consistent with both indices capturing the same underlying formal regularity axis from complementary directions. Floor area, CAR and adjacency show weaker, more city-specific correlations ($|r| \leq 0.68$). Adjacency, in particular, loads predominantly on its own component (PC3) in Elazığ and Erzincan

(loadings of 0.90 and 0.91), whereas in Mardin it splits almost evenly between PC1 (0.42) and PC3 (−0.45), reflecting the attached/detached polarisation already noted above. In every city, adjacency captures information that the other seven variables do not. The redundant variables were retained in the clustering, since PCA projects them onto orthogonal components before the Geo-KMeans step, so the redundancy inflates neither the distance metric nor the cluster count directly. Even so, height, floor count, compactness and fractal dimension are best read as two effective dimensions rather than four when interpreting the typology (Section 4.5).

The typology profiles (Table 7, Figure 11) present common typological groups and context-specific distinctions side by side. Each city features an 'ordinary settlement' typology consisting of low-rise, regular, detached structures, which make up about one-third of the buildings in Elazığ (typology 3) and the largest proportion in Erzincan (typology 1). Alongside this is a core typology with a higher floor count and a higher coverage ratio, concentrated in the centre of the maps (Elazığ typology 1 and Erzincan typology 4). Low-coverage, sparse periphery typologies also recur across the three cities. The irregularly shaped typology, characterised by low compactness and a high fractal dimension, constitutes a distinct category in Elazığ (typology 4) and Erzincan (typology 0). However, in Mardin's four-typology solution, it loses its independence and disperses among the other types. Mardin's dominant typology (typology 0) illustrates this: its mean floor count of 0.9 is not an error but a direct consequence of its height profile, since 56% of its buildings carry a GlobalBuildingAtlas-derived height below 3 m, the fixed 3.0 m/floor conversion threshold for a single storey (Section 2.3.2). This is consistent with the many low, single-storey and ancillary attached structures that make up this dense historic core fabric, and with the height estimation uncertainty already flagged for Mardin's micro-scale stock (Section 4.6).

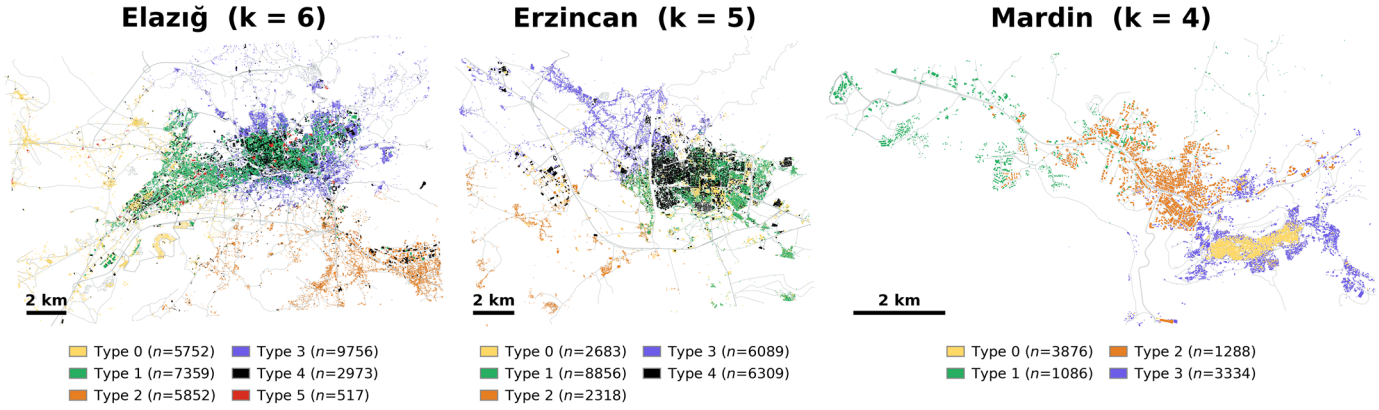


Figure 11. Urban fabric typologies of Elazığ, Erzincan and Mardin (top to bottom), derived from spatially weighted clustering (Geo-KMeans) of eight building-level morphometric variables (floor area, circular compactness, coverage area ratio, building adjacency, neighbour distance, fractal dimension, building height and floor count). Each footprint is coloured according to its assigned typology class, and the legends report the per-class building counts. The number of typologies (six for Elazığ, five for Erzincan and four for Mardin) corresponds to the optimal k selected by the Silhouette criterion within a search range of k = 4–6, after clusters comprising fewer than 1% of the building stock were merged into the nearest cluster centre.

Building adjacency proves most decisive, as expected, in Mardin. The city's dominant typology (typology 0, 40.4% of buildings) is a physically continuous historic core fabric of single-storey, very-high-coverage, markedly attached structures; with CAR reaching 0.477 and adjacency falling to 0.369, it is the densest and most continuous fabric observed across the three cities. Set against it are a wholly detached high-rise typology averag-

ing 8.3 floors (typology 2) and a sparse detached, low-rise periphery (typology 1), with the fourth typology (typology 3) bridging the extremes as a partly attached, medium-coverage transitional fabric. As a whole, the analysis gathers into one classification the morphological signatures traced separately in earlier sections. Elazığ and Erzincan present multi-typology structures resting on a largely detached stock and a graded centre–periphery gradient. Mardin, by contrast, sets out a less numerous but more polarised pattern, in which the attached dense historic core stands sharply apart from the detached high-rise and sparse periphery typologies. Synthesised at the building scale, the typology map thus confirms the three-city distinction that recurs across street network, centrality, density and adjacency: Elazığ radial polarised, Erzincan diffuse polycentric and Mardin compact attached polarised.

Table 7. Per-typology morphometric profiles for the building-level Geo-KMeans solution of each city. Values are cluster means, and building counts and shares are computed over the classified building stock (Elazığ n = 32,209; Erzincan n = 26,255; Mardin n = 9584). CAR is reported on the original (back-transformed) scale, while floor count, adjacency, compactness and fractal dimension are untransformed cluster means. Mardin's typology 0 records a mean floor count below 1 (0.9), a height-driven effect explained in the main text rather than a computational error.

City	Typology	*n* (%)	Floors	CAR	Adjacency	Compactness	Fractal Dim.	Dominant Character
Elazığ (k = 6)	0	5752 (17.9)	1.5	0.092	0.998	0.55	1.02	Low-rise, sparse detached
	1	7359 (22.8)	4.7	0.283	0.997	0.55	1.02	Dense high-rise core
	2	5852 (18.2)	1.6	0.088	0.999	0.56	1.02	Low-rise, sparse detached
	3	9756 (30.3)	1.7	0.147	0.999	0.57	1.01	Ordinary settlement (dominant)
	4	2973 (9.2)	3.1	0.236	0.998	0.33	1.10	Irregularly shaped
	5	517 (1.6)	3.1	0.196	0.855	0.53	1.03	Partly attached small group
Erzincan (k = 5)	0	2683 (10.2)	1.9	0.173	0.999	0.34	1.09	Irregularly shaped
	1	8856 (33.7)	1.6	0.181	0.999	0.58	1.01	Ordinary settlement (dominant)
	2	2318 (8.8)	1.4	0.085	1.000	0.55	1.02	Sparse detached periphery
	3	6089 (23.2)	1.5	0.094	1.000	0.56	1.02	Low-coverage settlement
	4	6309 (24.0)	3.3	0.273	0.996	0.57	1.01	Dense high-rise core
Mardin (k = 4)	0	3876 (40.4)	0.9	0.477	0.369	0.50	1.05	Attached historic core (dominant)
	1	1086 (11.3)	1.4	0.066	0.994	0.54	1.03	Sparse detached periphery
	2	1288 (13.4)	8.3	0.218	0.990	0.55	1.02	Detached high-rise
	3	3334 (34.8)	1.2	0.201	0.875	0.57	1.02	Partly attached transitional fabric

4. Discussion

Taken together, the findings demonstrate how much can be gained by joining computational urban morphology to unsupervised machine learning in the analysis of morphological diversity, and they contribute empirically to the theoretical debate on the form–process relationship in cities. The discussion is organised along six axes.

4.1. Orientation Entropy and Order: Beyond the Organic–Grid Dichotomy

In a comparison spanning more than a hundred cities, Boeing showed that urban form occupies an uninterrupted continuum between ordered and disordered poles [30]. The orientation and regularity results here confirm that continuum in the Anatolian case. Furthermore, they reveal that apparently identical indicators can issue from quite different morphogenetic processes.

The practically equal orientation entropy of Elazığ and Mardin is a case in point. In Mardin, high entropy, read with the highest circuity (1.076), the highest dead-end proportion (16.5%) and the lowest four-way intersection proportion (15.8%), points to the organic accumulation of Conzen's plan units in accommodation to topographic constraints [5,39]. The disorder is the natural product of a historic fabric terracing on a slope. In Elazığ, by contrast, a similar entropy reflects the mixed network geometry shaped on the multi-axial topography described by Canpolat [31], where disorder arises not from organic accumulation but from the superposition of plan fragments at different orientations. Orientation

entropy alone, therefore, does not distinguish fabric type. Read together with the four-way intersection proportion, the grid order index and the topological indices, it yields meaningful typo-morphological distinctions.

The most regular network belongs to Erzincan, whose four-way intersection proportion (23.2%) and Garrison and Marble topology indices ($\alpha = 0.273$; $\beta = 1.545$; $\gamma = 0.515$) are the highest of the three cities [44]. These values nonetheless fall well below those of the strong grid cities in Boeing's dataset [30]: although Erzincan is the most orthogonal of the three provinces, in absolute terms it holds a relatively orthogonal intermediate position rather than a true planned grid. That position fits the systematic reconstruction after the 1939 and 1992 earthquakes and aligns with Conzen's concepts of morphological persistence and the resilience of the plan unit [5]. The traces of reconstruction surface not only in the street network but also at the building scale: high compactness (Gibbs index 0.581), low fractal dimension (1.010) and relatively high topological circuitry ($\alpha = 0.273$). This mutual reinforcement of network and building geometry shows that the morphological signature of planned intervention has stayed legible for more than half a century.

4.2. Space Syntax: Network Size, Topological Depth and Local–Global Decoupling

The space syntax findings largely confirm the basic principles of Hillier and Hanson, and of Hillier, while also showing context-specific deviations [45,55]. These deviations reveal how the theory's universal expectations are reshaped within the topographic and historical context of Anatolian cities.

The core pattern confirming the theory is the inverse relationship between global integration and network size (Mardin 0.386 > Erzincan 0.317 > Elazığ 0.280). This order is internally consistent with both mean depth (Elazığ 39.2 > Erzincan 32.0 > Mardin 23.3) and closeness centrality (Mardin 0.0445 > Erzincan 0.0316 > Elazığ 0.0260), and it matches the expectation, intrinsic to Relative Asymmetry normalisation, that integration falls in large networks. Elazığ, the most extended and deepest network and the one with the highest betweenness Gini (0.819), thus presents a radial polarised structure in which through movement concentrates on a few spine axes.

A theoretically important result is that Mardin couples high physical density (smallest cell area, highest CAR) with the lowest local integration (R3 = 1.351; R5 = 1.084), which shows that physical compactness and topological accessibility need not go together. Although its compact historic core is dense, Mardin's narrow, dead-end-heavy, slope-conforming organic network depresses local accessibility. Erzincan, conversely, records the highest local integration (R3 = 1.448; R5 = 1.140), the local scale advantage of its relatively reticular fabric. Yet this advantage does not carry to the global scale, where Erzincan remains mid-ranked. Such local–global decoupling accords with Penn's discussion of how topological information is preserved across scale transitions, and it shows that local organisation capacity and global accessibility can act as independent dimensions [47].

One methodological caveat applies here: integration was computed with a segment-/node-based topological approach on the primary network rather than the classical axial map [19], so the absolute values are not directly comparable with axial-based studies. Because the same method was applied identically to the three provinces, however, the relative inter-city comparisons remain valid. On balance, the findings show that the basic principles of space syntax theory hold in mid-sized Anatolian cities, while two kinds of decoupling mark the limits of that validity. First, the most integrated axes do not always coincide with the densest fabrics. Second, an axis's pedestrian-scale accessibility does not directly reflect its position within the city as a whole. The framework, therefore, must be read together with each city's cadastral, historical and topographic context.

4.3. The Accessibility–Density Relationship: An Anatolian Test of Hillier's Proposition

A central finding is that Hillier's proposition that "accessibility and density co-evolve" receives strong empirical support in none of the three provinces [55]. One of the theory's central assumptions, then, holds only limited validity in mid-sized Anatolian cities.

The empirical picture is clear. The global integration–FSI correlations (Pearson r = 0.11–0.25) and the integration–CAR correlations (r = 0.05–0.17) are statistically significant ($p < 0.01$), yet the explained variance remains around 6%, even at its strongest. That significance is a sample effect of tens of thousands of observations rather than a sign of a strong relationship. The link is relatively the clearest in Elazığ (0.247 with FSI) and the weakest in Erzincan (0.050 with CAR), a contrast that follows the cities' development regimes. State-led post-earthquake reconstruction appears to have largely decoupled density decisions from accessibility configuration in Erzincan, whereas at Elazığ's larger scale market dynamics have begun a partial alignment with street access that has not yet hardened into structural integration. The weak accessibility–density link is thus not a universal theoretical failure but the imprint of a context in which density is shaped chiefly by off-network factors: the planning regime, topography and historical heritage.

Horizontal and vertical density relate to one another quite differently. The CAR–FSI correlation is very high in Elazığ (0.823) and Erzincan (0.840) but markedly low in Mardin (0.237). This pattern is partly definitional: because GSI and CAR are identical in the dataset and FSI is derived as FSI = GSI × floor count, in Elazığ and Erzincan, where floor count is low and homogeneous, FSI becomes almost a linear multiple of CAR, inflating the correlation. Mardin's low value, by contrast, is a genuine morphological signal, since the floor count heterogeneity introduced by a separate high-rise typology averaging ~8.3 floors statistically detaches FSI from horizontal coverage. Paradoxically, the city with the lowest correlation has the most diverse vertical structure, while the two with the highest have the most homogeneous.

This distinction carries a critical caution for theoretical interpretation. The seemingly simultaneous horizontal and vertical growth in Elazığ and Erzincan is not the counterpart of a genuine form–process pattern of the kind defined for European cities within the density framework of Berghauser Pont and Haupt [49]. It reflects, rather, the low vertical diversity and the partly definitional character of the measure: the apparently strong horizontal–vertical coupling is a statistical by-product of a low, homogeneous building stock, not a morphogenetic co-evolution. Correlations among density indicators must therefore be read with the measures' definitional dependencies and the level of vertical diversity in mind. Otherwise, tautological relationships risk being mistaken for morphological findings.

4.4. Morphological Tessellation and Spatial Clustering Patterns

The LISA findings quantify the spatial structure of density across morphological tessellation cells. Therefore, the tessellation serves not only as a measure of grain size, but also as an analytical basis for the spatial organisation of density [50,52]. Strong, positive and significant spatial autocorrelation in all three provinces (Global Moran's I = 0.620–0.709; $p < 0.001$) confirms that the CAR distribution is random in no city. The ordering of the values (Mardin 0.709 > Elazığ 0.697 > Erzincan 0.620), together with the proportions of non-significant cells (the highest in Erzincan at 56.9%, the lowest in Mardin at 53.0%), reveals that Mardin holds the most centralised and Erzincan the most diffuse density structure.

Beneath this global ordering, the local clustering patterns correspond to three distinct density models. Mardin, with the highest hot spot (High–High) share (20.2%) and the lowest non-significant proportion, shows a single-core, sharply bounded concentration gathered in the historic fabric, consistent with its compact, attached, low-rise high-coverage signature. Elazığ, with the highest cold-spot (Low–Low) share (26.6%), draws the sharpest

centre–periphery contrast between a clear core and a broad sparse periphery, a polarisation that coincides with its radial spine and high betweenness Gini. Erzincan, by contrast, combines the highest spatial outlier share (HL + LH ≈ 2.0%) with the highest non-significant proportion, pointing to a diffuse, transitional and multi-focal distribution produced by systematic post-earthquake reconstruction. That planned intervention spread density across the base rather than into one sharp gradient also fits the city's relatively regular grid and high local integration scores.

The contrasts among the three provinces (single-core in Mardin, polarised radial in Elazığ and diffuse multi-focal in Erzincan) suggest that urban density is best assessed not through a static centre–periphery model but alongside the dynamics of polycentric development [56]. Erzincan's diffuse pattern shows, in particular, that planned reconstruction can generate a multi-focal density structure. LISA thus serves as a diagnostic that distinguishes not only how strongly density clusters but also by what spatial logic, historical accumulation or planned intervention, it is organised.

4.5. Typology Inference: The Competence and Context Sensitivity of Unsupervised Classification

In the typology analysis, the number of clusters was selected within the four-to-six range by the Silhouette criterion under the parsimony rule defined in Section 2.3.5 (Davies–Bouldin and Calinski–Harabasz served only for reference), the search was capped at six for interpretability, and clusters below 1% of the total were merged into the nearest centre. The final typologies number six in Elazığ, five in Erzincan and four in Mardin. The highest Silhouette value (0.265), in Mardin, shows that its typologies separate most cleanly. Including the building adjacency and neighbour distance variables proved decisive: by introducing a distinct neighbourhood/spacing dimension, they split the fabric, especially in Mardin, into a physically attached dominant typology (≈40.4% of the stock) and detached high-rise and sparse periphery typologies.

The typology families recurring across all three provinces (a dominant regular low-rise fabric, a dense centre core, an irregularly shaped fabric and a sparse periphery) show that certain basic fabric types persist in Anatolian cities too. That the families' morphological content and proportions shift from city to city, in turn, reveals that unsupervised clustering yields context-sensitive typologies rather than universal categories [2,7,29]. By producing meaningful, spatially coherent clusters in every province (separating the attached historic core from the detached/high-rise fabric in Mardin, foregrounding the planned fabric in Erzincan, and grading centre to periphery in Elazığ), the method confirms its capacity for exploratory typology. The proposed workflow's contribution is to place that capacity on a reproducible, objective footing, offering a measurable alternative to traditional typology studies grounded in expert judgement [12,57].

4.6. Methodological Contribution and Limitations

The study's methodological contribution lies in uniting OSMnx-based street network analysis [10,17], morphological tessellation and building metrics from momepy [14,50], the space syntax configurational accessibility framework [45] and LISA spatial autocorrelation [52] within a single, open and reproducible workflow. This integration lets the findings be triangulated against one another, though not statistically cross-validated in the strict sense: the coverage area ratio (CAR) is both the LISA input variable and one of the eight Geo-KMeans clustering features, so the overlap between LISA hot spots and typology boundaries partly reflects this shared input rather than fully independent corroboration. The LISA hot spot clusters, the space syntax integration patterns and the typology boundaries overlap, which shows that three distinct analytical perspectives converge and that the three-city morphological typology recurring throughout the study is robust across meth-

ods. All parameters were reported explicitly and held identical across the three provinces, securing the basic precondition for inter-city comparison.

The framework's sensitivity to parameter choices was tested systematically. Typologies obtained by varying the spatial weight (alpha) and the PCA variance threshold (80%; 90%) on a grid were compared through the Adjusted Rand Index (ARI). The PCA threshold proved negligible (ARI = 0.92–0.96), and the number of typologies stayed relatively robust under every spatially weighted setting. The decisive factor is alpha: even within the reasonable range (1.5 to 2.0), the exact class assignments shifted moderately (pairwise ARI = 0.66–0.99 across the three cities). Because of this sensitivity, alpha was set not heuristically but by a data-driven criterion, grounded in the ClustGeo logic, that balances the inertia the clustering explains in feature space (Q0) against that in spatial space (Q1) [28]. The procedure converged to $\alpha^* = 1.75$ in all three provinces, and the flat balance curve over the 1.5–2.0 range confirmed robustness across that interval. This tuning was first carried out at a fixed reference k = 6 for all three cities. Since the finally adopted k differs from that reference in Mardin (k = 4) and Erzincan (k = 5), alpha was then re-estimated independently at each city's own adopted k as an explicit check. In all three cities the criterion converged to the same $\alpha^* = 1.75$ regardless of whether it was tuned at the k = 6 reference or at the city's own adopted k, and the reported Silhouette scores were unchanged to four decimal places, confirming that the alpha choice is not an artefact of the reference k used to derive it. In Erzincan, re-fitting with the re-estimated alpha reassigned a small number of buildings near cluster boundaries (under 1% of the stock) between two typologies of similar character, consistent with the k-means initialisation sensitivity already noted below. The optimal k, the typology count and every other city's result remained unchanged. Even so, both this criterion and the literature-based fixed parameters, such as the tessellation and PCA thresholds, warrant re-testing before transfer to cities of different scale and data density.

A further robustness check asks what the Geo-KMeans typology would look like if built from smaller, conceptually coherent subsets of the eight building-level variables rather than from all eight together. Table 8 re-runs the full pipeline (log1p transform where applicable, standardisation, PCA, data-driven alpha, and Silhouette-based k selection under the same parsimony rule) on a building-form-only set (floor area, compactness, fractal dimension), a density-only set (CAR, height, floor count) and a neighbourhood-only set (adjacency, neighbour distance), and compares each against the paper's actual typology via the Adjusted Rand Index. Two patterns hold in all three cities. First, every reduced subset returns a higher Silhouette score than the full solution (0.29–0.45 versus 0.24–0.27), a pattern that reflects the Silhouette index's well-known bias toward lower-dimensional, more easily separable spaces rather than a genuinely "better" typology. Silhouette alone, for this reason, cannot justify a reduced variable set. Second, no subset reproduces the full solution closely: ARI ranges from 0.34 to 0.74, well short of the near-1 values that would indicate redundancy. The density-only subset comes closest to the full solution in Elazığ and Erzincan (ARI 0.74 and 0.66), consistent with the strong loadings these variables carry on the first component (Figure 10), while building form and neighbourhood variables diverge further, particularly in Mardin (ARI as low as 0.34), where adjacency is already known to drive the typology's structure (Section 4.5). Each indicator group therefore contributes information the others do not, which supports retaining the full eight-variable input rather than a reduced one. One caveat applies: with only two or three input variables, PCA reduces each subset to at most two components (Table 8), so these are near-no-reduction clusterings rather than a like-for-like PCA comparison with the eight-variable case.

Table 8. Indicator group ablation: Geo-KMeans re-run on reduced variable subsets for each city, compared against the full eight-variable solution reported in Table 7 via the Adjusted Rand Index (ARI).

City	Variable Group	*n* PCs	*α**	Optimal k	Silhouette	ARI vs. Full
Elazığ	Full (baseline, 8 vars.)	5	1.75	6	0.245	1.000
Elazığ	Building-form-only (3)	2	1.25	4	0.290	0.536
Elazığ	Density-only (3)	2	1.25	4	0.323	0.736
Elazığ	Neighbourhood-only (2)	2	1.00	6	0.365	0.442
Erzincan	Full (baseline, 8 vars.)	5	1.75	5	0.240	1.000
Erzincan	Building-form-only (3)	2	1.25	4	0.310	0.489
Erzincan	Density-only (3)	2	1.25	4	0.294	0.663
Erzincan	Neighbourhood-only (2)	2	1.00	5	0.371	0.373
Mardin	Full (baseline, 8 vars.)	5	1.75	4	0.265	1.000
Mardin	Building-form-only (3)	2	1.25	4	0.359	0.344
Mardin	Density-only (3)	2	1.25	4	0.359	0.537
Mardin	Neighbourhood-only (2)	2	1.00	4	0.446	0.601

α*: α denotes the spatial weight parameter of the Geo-KMeans model, re-estimated independently for each variable subset via the data-driven ClustGeo Q0–Q1 inertia-balance criterion; the * denotes the optimal value.

The study's remaining limitations fall under six headings. The first, and methodologically the weightiest, is that space syntax was run with a topological segment/node approach on the primary network rather than the classical axial/angular line analysis [19]. Because axial and angular segment analyses foreground long sight–movement lines and route continuity, they may order the integration hierarchy differently [45,47], so the absolute integration/depth values here cannot be matched one-to-one with the axial-based literature. The relative inter-city orderings remain consistent, since the identical pipeline was applied throughout, and re-analysing the same networks with axial/angular segment analysis would usefully test the results' sensitivity to the choice of representation [42]. Second, building heights were derived from an open global raster dataset [40], whose resolution limit may introduce uncertainty, especially in Mardin's micro-scale stock. Third, k-means guarantees no global optimum and is sensitive to initial centres; n_init = 20 repetitions, the spatially weighted (Geo-KMeans) set-up and the data-driven alpha selection mitigate this but cannot remove it entirely, and the partly definitional link between horizontal and vertical density should likewise be kept in view (see Section 4.3). Fourth, the completeness of Overture Maps data may vary across cities, and the narrow streets of Mardin's historic fabric may be under-represented [41]. Fifth, the three cities were selected from the outset to instantiate the organic historical, planned grid and mixed topographic archetypes that the results subsequently confirm. The workflow's capacity to recover this pattern reflects, at least in part, a purposive rather than a fully independent sample, and its generalisation to cities outside this a priori framework remains to be tested. Finally, the analysis is cross-sectional and does not track the temporal evolution of morphological transformation, whereas a longitudinal perspective would deepen understanding of the form–process relationship.

Overall, this section clarifies the study's dual position. On the one hand, by uniting four morphometric traditions in a single, transparent and parameter-sensitivity-documented pipeline, it enables findings to be triangulated across methods, within the caveat noted above regarding CAR's dual role. On the other, it reports explicitly its dependencies on the choice of representation (topological versus axial), the data sources (Overture, a global height raster) and the parameter decisions (alpha, the PCA threshold), thereby marking the limits within which the results generalise. The study positions itself, therefore, not as a definitive and universal typology claim but as a framework that priori-

tises reproducibility and methodological transparency and that offers a basis for future scale-extension and representation comparison work.

5. Conclusions

This study developed a multi-scale computational urban morphology workflow that integrates OSMnx, momepy, space syntax, and LISA into a single, reproducible pipeline, and applied it comparatively to Elazığ, Erzincan, and Mardin (Figure 2). The findings converge on five main conclusions.

First, the workflow succeeds in mapping, in measurable form, the morphological spectrum of Anatolian cities of differing typo-morphological character. The four-way intersection proportion (0.158–0.232), orientation entropy (3.497–3.545 nats), grid order index (0.011–0.024), building compactness (0.557–0.581) and fractal dimension (1.010–1.016) produce a systematic differentiation along a continuous axis running from Mardin's organic historical fabric to Erzincan's relatively planned grid structure. Mardin is at one end of this axis, with the lowest proportion of four-way intersections, the lowest compactness and the highest fractal dimension. Erzincan is at the other end, with the highest proportion of four-way intersections, the highest grid order and the highest topological circuit complexity. Elazığ occupies an intermediate, transitional position between the two.

Second, Hillier's proposition that "accessibility and density co-evolve" does not receive strong empirical support in any of the three cities [55]. The correlations between global integration and density remain weak (r = 0.05–0.25), with even the strongest explaining only about six percent of the variance. This suggests that densification in Anatolian cities is shaped far more by cadastral structure, zoning legislation, and political–economic processes than by street network configuration.

Third, the correlation between the coverage area ratio (CAR) and the Floor Space Index (FSI) runs very high in Erzincan and Elazığ ($r > 0.82$), implying that horizontal and vertical growth proceed concurrently in these two cities. This strong relationship is partly definitional. Because FSI is calculated by multiplying CAR with the floor count, the low and homogeneous floor profiles in both cities structurally tighten the correlation. In contrast, Mardin's low CAR–FSI correlation (r = 0.24) reflects a stark vertical heterogeneity. This indicates the presence of a distinct high-rise typology that completely decouples urban density from horizontal ground coverage.

Fourth, spatially weighted unsupervised clustering (Geo-KMeans), with the spatial weight parameter chosen using a data-driven criterion, produces six meaningful and spatially coherent morphological types in Elazığ, five in Erzincan, and four in Mardin. The reduction to four types in Mardin demonstrates that these typologies are not context-independent universal categories but classes highly sensitive to local morphological structure. Including the building adjacency variable separates, especially in Mardin, the physically continuous (attached) historic core from the detached high-rise fabric. Triangulated against LISA hot spot clusters and space syntax integration maps (with the CAR overlap caveat noted in Section 4.6), these classifications show that unsupervised machine learning can offer a reproducible, data-driven complement to expert judgement in deciphering urban fabric.

Fifth, the systematic differentiation in building compactness and fractal dimension numerically documents the relationship between urban form and historical depth. Earthquake-reconstructed Erzincan marks the extreme of geometric simplicity (highest compactness, lowest fractal dimension), whereas historic Mardin marks the extreme of geometric complexity (lowest compactness, highest fractal dimension). This contrast empirically demonstrates the decisive effect of post-disaster reconstruction and conservation policies on the continuity of morphological heritage.

For planning practice, these findings carry several concrete implications. In Mardin's historic core, the pairing of a high-coverage ratio with the lowest local integration argues that density increases in conservation areas should not proceed without improving local topological accessibility. The decoupling of accessibility from density in Erzincan indicates that reconstruction decisions were taken with little regard to network configuration, arguing for space syntax analysis as a decision support tool in future development plan revisions. In Elazığ, the concentration of movement on a few spine axes within its most extended and topologically deepest network (the highest betweenness Gini, 0.819) implies that transport investment should target the topological decoupling between the centre and the broad, low-density periphery.

Four directions follow for future work. The first is to track the temporal evolution of urban form through longitudinal analysis. The second is to benchmark deep learning classifiers, such as Graph Neural Networks (GNNs), against the k-means results. The third is to extend the sample to a metropolitan-scale city such as Ankara, systematically testing the scale effect and the transferability of the parameters. The fourth is to validate the present findings with high-resolution LiDAR-based building heights. Re-analysing space syntax with axial/angular segment analysis would, in addition, be a valuable extension for testing how sensitive the integration hierarchy is to the choice of representation.

In conclusion, the vector-based hybrid typo-morphological framework produces robust, comparable, and applicable results for Anatolian cities of historical depth. By bringing together open-source tools, standard data formats, and a reproducible, parameter-sensitivity-documented pipeline, the workflow establishes an analytical ground that reaches beyond mere description. At the same time, it should be read not as a "universal typology model" but as an application calibrated to three mid-sized Anatolian cities. The approach carries urban morphology research from traditional descriptive methods toward a data-driven, algorithmic classification practice and offers a standard, reproducible reference framework for studies of urban form in Türkiye and comparable geographies.

Funding: This work was supported by the Scientific Research Projects Coordination Unit of Fırat University (FÜBAP) under Grant İSBF.25.28.

Data Availability Statement: The Python analysis pipeline, the GeoPackage outputs and the component-loading matrices used in this study are openly available at https://github.com/facanpolat-tr/Urban-Morphological-Spectrum (accessed on 14 July 2026). The building footprint data were obtained from Zhu et al. [40] and the street network data from the Overture Maps Foundation [41]; the relevant licence terms are stated in the data repository.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Larkham, P.J. The Study of Urban Form in Great Britain. *Urban Morphol.* **2006**, *10*, 117–141. [CrossRef]
2. Moudon, A.V. Urban Morphology as an Emerging Interdisciplinary Field. *Urban Morphol.* **1997**, *1*, 3–10. [CrossRef]
3. Whitehand, J.W.R. British Urban Morphology: The Conzenian Tradition. *Urban Morphol.* **2001**, *5*, 103–109. [CrossRef]
4. Muratori, S. *Studi per una Operante Storia Urbana di Venezia*; Istituto Poligrafico dello Stato: Rome, Italy, 1959.
5. Conzen, M.R.G. *Alnwick, Northumberland: A Study in Town-Plan Analysis*; George Philip: London, UK, 1960.
6. Maretto, M. Saverio Muratori: Towards a Morphological School of Urban Design. *Urban Morphol.* **2013**, *17*, 93–106. [CrossRef]
7. Caniggia, G.; Maffei, G.L. *Architectural Composition and Building Typology*; Alinea: Copenhagen, Denmark, 2001.
8. Kropf, K. *The Handbook of Urban Morphology*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
9. D'Acci, L. On Urban Morphology and Mathematics. In *The Mathematics of Urban Morphology*; D'Acci, L., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 1–18.
10. Boeing, G. Spatial Information and the Legibility of Urban Form: Big Data in Urban Morphology. *Int. J. Inf. Manag.* **2021**, *56*, 102013. [CrossRef]

11. Dibble, J.; Prelorendjos, A.; Romice, O.; Zanella, M.; Strano, E.; Pagel, M.; Porta, S. On the origin of spaces: Morphometric foundations of urban form evolution. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 707–730. [CrossRef]
12. Fleischmann, M.; Romice, O.; Porta, S. Measuring Urban Form: Overcoming Terminological Inconsistencies for a Quantitative and Comprehensive Morphologic Analysis of Cities. *Environ. Plan. B Urban Anal. City Sci.* **2021**, *48*, 2133–2150. [CrossRef]
13. Ye, Y.; Van Nes, A. Quantitative Tools in Urban Morphology: Combining Space Syntax, Spacematrix and Mixed-Use Index in a GIS Framework. *Urban Morphol.* **2014**, *18*, 97–118. [CrossRef]
14. Fleischmann, M. Momepy: Urban Morphology Measuring Toolkit. *J. Open Source Softw.* **2019**, *4*, 1807. [CrossRef]
15. Fleischmann, M.; Feliciotti, A.; Kerr, W. Evolution of Urban Patterns: Urban Morphology as an Open Reproducible Data Science. *Geogr. Anal.* **2022**, *54*, 536–558. [CrossRef]
16. Oliveira, V.; Porta, S. Quantitative and Qualitative Analysis in Urban Morphology: Systematic Legacy and Latest Developments. *Proc. Inst. Civ. Eng. Urban Des. Plan.* **2025**, *178*, 75–87. [CrossRef]
17. Boeing, G. OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. *Comput. Environ. Urban Syst.* **2017**, *65*, 126–139. [CrossRef]
18. Boeing, G. Modeling and Analyzing Urban Networks and Amenities with OSMnx. *Geogr. Anal.* **2025**, *57*, 567–577. [CrossRef]
19. Porta, S.; Crucitti, P.; Latora, V. The Network Analysis of Urban Streets: A Primal Approach. *Environ. Plan. B Plan. Des.* **2006**, *33*, 705–725. [CrossRef]
20. McGarigal, K.; Marks, B.J. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure*; USDA Forest Service: Washington, DC, USA, 1995.
21. Bosch, M. PyLandStats: An open-source Pythonic library to compute landscape metrics. *PLoS ONE* **2019**, *14*, e0225734. [CrossRef] [PubMed]
22. Zhang, A.; Pontius, R.G., Jr.; Bilintoh, T.M.; Sangermano, F.; Rogan, J. DynamicPATCH: Method and software for spatially explicit dynamic patch transition characterization. *Landsc. Ecol.* **2025**, *40*, 132. [CrossRef]
23. Araldi, A.; Fusco, G. From the Street to the Metropolitan Region: Pedestrian Perspective in Urban Fabric Analysis. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 1243–1263. [CrossRef]
24. Cai, C.; Guo, Z.; Zhang, B.; Wang, X.; Li, B.; Tang, P. Urban Morphological Feature Extraction and Multi-Dimensional Similarity Analysis Based on Deep Learning Approaches. *Sustainability* **2021**, *13*, 6859. [CrossRef]
25. Jochem, W.C.; Leasure, D.R.; Pannell, O.; Chamberlain, H.R.; Jones, P.; Tatem, A.J. Classifying Settlement Types from Multi-Scale Spatial Patterns of Building Footprints. *Environ. Plan. B Urban Anal. City Sci.* **2021**, *48*, 1161–1179. [CrossRef]
26. O'Brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **2007**, *41*, 673–690. [CrossRef]
27. Jolliffe, I.T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef] [PubMed]
28. Chavent, M.; Kuentz-Simonet, V.; Labenne, A.; Saracco, J. ClustGeo: An R package for hierarchical clustering with spatial constraints. *Comput. Stat.* **2018**, *33*, 1799–1822. [CrossRef]
29. Gil, J.; Beirão, J.N.; Montenegro, N.; Duarte, J.P. On the Discovery of Urban Typologies: Data Mining the Many Dimensions of Urban Form. *Urban Morphol.* **2012**, *16*, 27. [CrossRef]
30. Boeing, G. Urban Spatial Order: Street Network Orientation, Configuration, and Entropy. *Appl. Netw. Sci.* **2019**, *4*, 1–19. [CrossRef]
31. Canpolat, F.A. Kentsel Değişimin Coğrafi Boyutları: Elazığ Örneği. *Al Farabi Uluslararası Sos. Bilim. Derg.* **2019**, *3*, 70–105.
32. TURKSTAT. Turkish Statistical Institute. 2025. Available online: https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr (accessed on 17 June 2026).
33. Uysal, A.; Sunkar, M. Erzincan Ovası ve Çevresinin Depremselliği ile Depremlerin Mekânsal ve Yoğunluk Analizler. *Fırat Univ. J. Soc. Sci.* **2024**, *34*, 419–436. [CrossRef]
34. Günay, B. *Urban Design Is a Public Policy*; METU Faculty of Architecture Press: Ankara, Türkiye, 1999.
35. UNESCO World Heritage Centre. Mardin Cultural Landscape. 2000. Available online: https://whc.unesco.org/en/tentativelists/1406/ (accessed on 17 June 2026).
36. Dursun, F. From Quarry to Monument: Considering Mardin Stone (SE, Türkiye) as the Symbol of Architectural and Cultural Heritage. *Geoheritage* **2024**, *16*, 64. [CrossRef]
37. Alioğlu, E.F. Geleneksel Mardin Şehir Dokusu ve Evleri Üzerine Bir Deneme. Ph.D. Thesis, Institute of Science and Technology, Istanbul Technical University, Istanbul, Türkiye, 1989.
38. Ozorhon, G.; Ozorhon, I.F. Learning from Mardin and Cumalıkızık: Turkish Vernacular Architecture in the Context of Sustainability. *Arts* **2014**, *3*, 175–189. [CrossRef]
39. Genç, İ. Tarihi Kentlerde Yürünebilirlik Olgusunun Kentleşme Süreci İçinde Değerlendirilmesi: Mardin Örneği. Ph.D. Thesis, Düzce University, Düzce, Türkiye, 2022.
40. Zhu, X.X.; Chen, S.; Zhang, F.; Shi, Y.; Wang, Y. GlobalBuildingAtlas: An Open Global and Complete Dataset of Building Polygons, Heights and LoD1 3D Models; Technical University of Munich. 2025. Available online: https://mediatum.ub.tum.de/1782307 (accessed on 17 June 2026).

41. Overture Maps Foundation. Overture Maps Data: Transportation Theme. 2024. Available online: https://overturemaps.org/download/ (accessed on 17 June 2026).
42. Gil, J.; Varoudis, T.; Karimi, K.; Penn, A. The Space Syntax Toolkit: Integrating Street Network and Urban Form Analysis in GIS. In *Proceedings of the 10th International Space Syntax Symposium*; University College London: London, UK, 2015.
43. Hagberg, A.; Swart, P.J.; Schult, D.A. *Exploring Network Structure, Dynamics, and Function Using NetworkX*; Los Alamos National Laboratory (LANL): Los Alamos, NM, USA, 2008.
44. Garrison, W.L.; Marble, D.F. *The Structure of Transportation Networks*; U.S. Army Transportation Command: Washington, DC, USA, 1962.
45. Hillier, B.; Hanson, J. *The Social Logic of Space*; Cambridge University Press: Cambridge, UK, 1984.
46. Hillier, B.; Penn, A.; Hanson, J.; Grajewski, T.; Xu, J. Natural Movement: Or, Configuration and Attraction in Urban Pedestrian Movement. *Environ. Plan. B Plan. Des.* **1993**, *20*, 29–66. [CrossRef]
47. Penn, A. Space Syntax and Spatial Cognition: Or Why the Axial Line? *Environ. Behav.* **2003**, *35*, 30–65. [CrossRef]
48. Gibbs, J.P. *Urban Research Methods*; Van Nostrand: Ann Arbor, MI, USA, 1961.
49. Berghauser Pont, M.; Haupt, P. *Spacematrix: Space, Density and Urban Form—Revised Edition*; TU Delft OPEN Publishing: Delft, The Netherlands, 2023. Available online: https://books.open.tudelft.nl/home/catalog/view/38/51/177 (accessed on 14 July 2026).
50. Fleischmann, M.; Feliciotti, A.; Romice, O.; Porta, S. Morphological Tessellation as a Way of Partitioning Space: Improving Consistency in Urban Morphology at the Plot Scale. *Comput. Environ. Urban Syst.* **2020**, *80*, 101441. [CrossRef]
51. Rey, S.J.; Anselin, L. PySAL: A Python Library of Spatial Analytical Methods. In *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*; Fischer, M.M., Getis, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 175–193.
52. Anselin, L. Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [CrossRef]
53. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI–1*, 224–227. [CrossRef]
54. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1–27. [CrossRef]
55. Hillier, B. *Space is the Machine: A Configurational Theory of Architecture*; Cambridge University Press: Cambridge, UK, 1996.
56. Batty, M. *The New Science of Cities*; The MIT Press: Cambridge, MA, USA, 2013.
57. Wolf, L.J.; Knaap, E.; Rey, S. Geosilhouettes: Geographical Measures of Cluster Fit. *Environ. Plan. B Urban Anal. City Sci.* **2021**, *48*, 521–539. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.