This is the Accepted Manuscript version of an article published by Elsevier in the journal *Building and Environment* in 2026, which is available at: https://doi.org/10.1016/j.buildenv.2026.115095
Cite as: Zhao M, Yang G, Zhang Z, Biljecki F, Zu H, Wu C (2026): Street space as health infrastructure: diagnosing uneven street support for exercise through the spatial triad. *Building and Environment*, 304: 115095.

Street Space as Health Infrastructure: Diagnosing Uneven Street Support for Exercise through the Spatial Triad

Minwei Zhao[a], Guosheng Yang[a], Zhuoni Zhang[a], Filip Biljecki[bc], Hanzhi Zu[a] and Cai Wu[a]*

a Thrust of Urban Governance and Design, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China.

b Department of Architecture, National University of Singapore, Singapore

c Department of Real Estate, National University of Singapore, Singapore

***Corresponding author:**

Dr. Cai Wu

Assistant Professor

Thrust of Urban Governance and Design

The Hong Kong University of Science and Technology (Guangzhou)

caiwu@hkust-gz.edu.cn

Other authors' email addresses:

Minwei Zhao: mzhao886@connect.hkust-gz.edu.cn

Guosheng Yang: gyang559@connect.hkust-gz.edu.cn

Zhuoni Zhang: znzhang@hkust-gz.edu.cn

Filip Biljecki: filip@nus.edu.sg

Hanzhi Zu: hzu765@connect.hkust-gz.edu.cn

Funding: This study is supported by the Guangdong Basic and Applied Basic Research Foundation (grant no. 2025A1515110224), the AI Research and Learning Base of Urban Culture (project 2023WZJD008), and the Guangzhou Municipal Science and Technology Project (Guangzhou Basic and Applied Basic Research Scheme, grant nos. 2024A04J6375 and 2025A03J3490; principal investigator: Zhuoni Zhang).

Street Space as Health Infrastructure: Diagnosing Uneven Street Support for Exercise through the Spatial Triad

Abstract

Urban streets constitute essential everyday health infrastructure within the outdoor built environment, yet their potential to support physical activity remains unevenly realized. Current assessments prioritize formal sports facilities or rely on coarse, area-level metrics, leaving street-level differences in exercise supportiveness less visible. To address this gap, this study develops a theory-informed analytical framework for examining modelled exercise-supportiveness deficits across street networks. We employ Henri Lefebvre's spatial triad as an interpretive lens, conceptualizing urban space as a contradictory and co-productive unity of three interconnected dimensions: Conceived Space (representations of space), Perceived Space (spatial practice), and Lived Space (spaces of representation). Drawing on the research agenda of Critical Quantitative Geography, these dimensions provide a vocabulary for organizing evidence on planned form, visible street conditions, and everyday activity traces. SHAP is then used to examine how the resulting indicator groups are associated with exercise intensity. Using Shenzhen as a case study, the analysis shows that conceived-space indicators account for the largest share of grouped SHAP attribution, while perceived and lived indicators reveal important local low-supportiveness patterns across development contexts. Grouped SHAP attributions are then used to construct a seven-mode diagnostic typology that identifies which C/P/L readings are associated with lower modelled street-level exercise supportiveness. Furthermore, a bivariate spatial association analysis identifies clusters where high population density coincides with low recorded street-level exercise intensity, highlighting locations for closer planning assessment. The framework combines machine-learning attribution with spatial-theoretical interpretation to examine health-supportive everyday movement environments at the street-segment scale.

Keywords: Street-based physical activity; Exercise supportiveness; Explainable machine learning; Social sensing; Spatial mismatch; Outdoor built environment.

1 Introduction

Physical inactivity is globally recognized as a major risk factor for noncommunicable diseases, obesity, mental health issues, and premature mortality (Sallis et al., 2016). As cities become denser and more complex, ensuring equitable access to opportunities for physical activity has become a pressing concern for healthy built-environment research and public health (Kohl et al., 2012; World Health Organization, 2018). Traditionally, research on active living has prioritized public parks, greenways, and designated sports facilities as primary sites for promoting physical activity (Kaczynski & Henderson, 2008). Such studies often rely on coarse measures of accessibility, typically treating opportunities as discrete, point-based destinations (Schipperijn et al., 2017). However, recent scholarship increasingly identifies urban streets as crucial yet underutilized spaces for everyday activity (Thompson, 2013; Kostrzewska, 2017). Unlike formal facilities, streets are ubiquitous and low-barrier, offering continuous environments for walking, jogging, or cycling, particularly in dense and underserved areas with limited designated infrastructure (You, 2016; Altomonte et al., 2020; Gong et al., 2018).

However, current approaches to assessing street-level supportiveness remain methodologically limited. Conventional indicators often overlook the continuity of streets and their capacity to support diverse and spontaneous forms of physical activity. Moreover, mainstream assessments often rely on aggregated walkability indices that are theoretically grounded and useful for describing structural accessibility (Frank et al., 2010; Maghelal & Capp, 2011; Venerandi et al., 2024). Their limitation for the present question is that they often privilege structural dimensions of walkability while giving less attention to eye-level perception and lived social use (Alattar et al., 2021; Biljecki & Ito, 2021; Samuelsson et al., 2019). For instance, a street may be connected in the network yet remain unattractive for everyday exercise when its eye-level environment suggests discomfort, exposure to traffic, or weak pedestrian provision. Understanding streets not merely as conduits for movement but as spaces that shape everyday behaviour therefore requires attention to their inherent complexity. The empirical task is therefore to identify where street-level exercise supportiveness is constrained by structural form, visible street conditions, lived activity traces, or their combined misalignment.

Critical geographer Henri Lefebvre confronted precisely this elusive yet powerful nature of space. The epistemological foundation of his theory lies in the spatial triad, which articulates three interrelated dimensions of space in order to illuminate the complexities of everyday life (Watkins, 2005). Lefebvre argues that space is fundamental to lived experience and that each experience encompasses three mutually constitutive aspects: representations of space (conceived space), spatial practices (perceived space), and spaces of representation (lived space) (Lefebvre, 1991).

Although this triadic framework has been widely employed in studies of environmental inequality and lived urban experience and has become a cornerstone of critical geography, its application in empirical research has long been characterized by an ontological–epistemological–methodological tension. Lefebvre's spatial ontology resists fragmenting space into discrete and measurable variables, whereas quantitative urban analysis relies on operationalized indicators to decompose spatial mechanisms.

Developments in Critical Quantitative Geography have sought to move beyond this binary opposition between dialectics and quantification (Kwan & Schwanen, 2009a) by advocating a post-cultural turn perspective on quantitative approaches (Kwan & Schwanen, 2009b). At the same time, advances in quantitative methodologies have provided a practical methodological "toolbox" for fostering interaction between these traditions.

To address these gaps, this study proposes a theoretically grounded and interpretable analytical framework for examining uneven street-level exercise supportiveness. We employ Lefebvre's spatial triad—conceived space (C), perceived space (P), and lived space (L)—as an interpretive lens to examine the complex interplay among environmental structure, sensory environments, and patterns of use. Within this framework, conceived space captures the rationalities embedded in street hierarchies, connectivity structures, and land-use planning; perceived space encompasses the visual and tactile cues that shape how space is interpreted and navigated; and lived space reflects the situated, symbolic, and affective engagements manifested through presence and activity. By integrating this theoretically informed perspective with explainable machine learning, specifically SHapley Additive exPlanations (SHAP), we aim to reveal association patterns that are both analytically legible for urban research and meaningful for policy reflection, while avoiding the reduction of urban spatial complexity to overly simplistic indicators.

Building on this problem, this study develops a diagnostic workflow with four connected contributions:

(1) We introduce a triad-informed framework for indicator construction, using the conceived, perceived, and lived dimensions as analytical readings that connect critical-geographical theory with measurable traces of street structure, visual environment, and everyday activity. This approach moves beyond ad hoc feature stacking while keeping the proxy nature and interpretive limits of the C/P/L groupings explicit.

(2) We link these spatial factors to actual street-level exercise behaviours using explainable machine learning. This allows structural, visible/material, and lived-space indicators to be compared within one diagnostic setting, extending structurally oriented walkability assessment toward the perceived and lived dimensions emphasized in Lefebvre's triad.

(3) We develop a seven-mode diagnostic typology of model-attributed supportiveness deficits. The typology distinguishes whether lower modelled exercise supportiveness is mainly associated with conceived, perceived, lived, or combined C/P/L readings, offering a segment-scale lens for targeted interpretation.

(4) Finally, our framework enables the identification and contextualization of spatial mismatches through the lens of our triad-informed diagnostic typology. By

cross-referencing mismatch clusters with dominant diagnostic profiles, we examine how the dominant diagnostic profiles may inform context-sensitive improvements in everyday movement support.

Together, these contributions help reconcile data-driven urban analytics with spatial theory by drawing on its structural insights to inform how we organize and interpret multi-source indicators. The triadic orientation provides a flexible way to reason about environmental supportiveness across conceived, perceived, and lived dimensions, while our proxy-based reading links theory to observable traces without treating C/P/L as discrete measurable objects. The resulting synthesis enables a more legible and context-aware diagnosis of street-level exercise supportiveness, with implications for targeted urban interventions across diverse settings.

2 Literature review

2.1 Urban Physical Activity and Exercise Deprivation

A substantial body of research on urban physical activity has focused on identifying how formal designated sports amenities, such as parks, greenways, and purpose-built sports facilities, provide focal nodes for exercise (Kaczynski & Henderson, 2008; Duan et al., 2018; R. Zhang et al., 2024). Within this paradigm, physical activity opportunities are often operationalized through accessibility measures and proximity-based indices that treat destinations as discrete point resources (Schipperijn et al., 2017), while concerns about physical inactivity have motivated public health scholarship emphasizing the uneven distribution of these facilities across populations (Sallis et al., 2016; Kohl et al., 2012; World Health Organization, 2018). Frameworks such as “activity deserts” (Coombes et al., 2010) and related accessibility indices (Pate et al., 2021) thus primarily conceptualize inequality through the availability and spatial allocation of designated infrastructures. Promoting equity in physical activity is therefore examined largely through the lens of facility provision rather than through the more granular mechanics of everyday urban environments. Yet, emerging scholarship increasingly identifies urban streets as crucial yet underutilized spaces for everyday activity (Thompson, 2013; Kostrzewska, 2017). Recent studies emphasize that streets function as everyday, low-threshold public environments where diverse forms of movement unfold (Chen et al., 2024; Mehta, 2008; You, 2016). Their importance becomes especially evident in dense or underserved urban neighborhoods where formal recreational infrastructures are scarce. In such contexts, streets often become the default realm for everyday exercise among populations with limited access to designated amenities, a pattern that is particularly pronounced in high-density cities such as Shenzhen (Sallis et al., 2020; Chen et al., 2016).

Recent evidence reinforces the health and policy relevance of this concern. A large meta-analysis links higher daily step counts with lower all-cause and cardiovascular mortality risk (Banach et al., 2023), while recent reviews show that objectively measured built-environment attributes, walkability, and urban health impact assessment remain central to physical-activity research (Pontin et al., 2022; Westenhöfer et al., 2023; Venerandi et al., 2024). Studies in Chinese cities further suggest that perceived built environment can mediate the relationship between objective built form and leisure-time physical activity (S. Zhang et al., 2024). These studies provide an important foundation, but they also indicate that street-based activity cannot be understood only through facility access or aggregate walkability; visible streetscape quality and socially experienced street use also matter.

This shift in attention from formal spaces to street environments calls for a parallel rethinking of how deprivation is conceptualized. The notion of "exercise deprivation" moves beyond infrastructure absence to encompass relational, multi-dimensional barriers embedded in street-level conditions: ranging from missing sidewalks to visual discomfort, safety-related environmental cues, or limited patterns of use and engagement (Ewing & Handy, 2009; Gehl, 2013; Alfonzo, 2005). However, prevailing approaches often rely on grid-based metrics or node-based proximity, which overlook the continuous character, perceptual attributes, and localised usage patterns of streets (Gehl, 2013). From a micro-scale perspective, a street may be "deprived" not only because of material deficits but also due to perceived barriers and limited social engagement or utilization (Alfonzo, 2005; Handy et al., 2002). These barriers often manifest unevenly, varying significantly from one street segment to the next. Understanding deprivation through this lens foregrounds the interplay of physical form, perception, and lived experience in shaping spatial inequality.

In sum, existing conceptual and analytical frameworks are not yet sufficient to characterize the complexity of street-based exercise deprivation. While shifting the focus to streets is crucial, current metrics often lack the structural depth to capture the "relational barriers" described above. There remains a critical need for a comprehensive framework that moves beyond ad hoc feature selection to structure multi-source data into coherent dimensions—planning, perception, and practice. This gap necessitates the triad-informed indicator construction developed in this study, which seeks to capture the layered nature of street space.

2.2 Measuring Street-Space Supportiveness: Structural, Perceived, and Lived Evidence

Recent research has seen rapid advancements in data-driven analytical methods. Geographic information science, computer vision, mobile sensing, and big urban data have enabled unprecedented granularity in modelling urban representations (Wu et al., 2023; Wu et al., 2024; Zhao et al., 2026). Researchers routinely integrate diverse spatial data to construct comprehensive representations of urban environments, an approach that has become central to contemporary urban health and human-environment interaction studies (Goodchild, 2007; Batty, 2013; Liu et al., 2026).

However, the rapid growth of data-driven urban analytics has not always been matched by theoretical frameworks linking spatial form to lived urban experience (Kim, 1999; Connaway et al., 2011). Heterogeneous indicators, from greenery indices to pedestrian reviews, also risk multicollinearity and scaling inconsistencies (Zhang et al., 2019; Zhong et al., 2020). While composite walkability indices summarize structural accessibility through density, connectivity, land-use mix, and pedestrian infrastructure, they may obscure whether weak street-level supportiveness stems from network configuration, eye-level environmental quality, or lived/social activation (Frank et al., 2010; Maghelal & Capp, 2011). POI density similarly captures functional richness but is rarely combined with perceptual and socio-behavioural evidence (Yue et al., 2017; Talen, 2003). Recent studies have advanced these dimensions separately through space syntax and pedestrian-flow modelling (Stavroulaki et al., 2024), street-view-based measures of walkability, bikeability, exposure, and streetscape diversity (Biljecki & Ito, 2021; Ito et al., 2024; Yang et al., 2026; Zhou et al., 2024), street greenery and active-travel research (Yu et al., 2024), and participatory, VGI, and experiential approaches to movement and eye-level perception (Alattar et al., 2021; Samuelsson et al., 2019). The remaining gap is an integrated street-level diagnostic framework that connects structural, perceived, and lived dimensions of exercise supportiveness.

To mitigate this interpretability challenge, explainable machine learning (XAI) methods, notably SHAP (SHapley Additive exPlanations), have emerged to quantify feature contributions and capture localized spatial variations (Lundberg et al., 2018; Molnar et al., 2020; Guidotti et al., 2019). However, a critical gap remains: existing applications of SHAP in urban research are largely descriptive, serving as post-hoc tools to rank feature importance without integration into broader theoretical frameworks (Samek et al., 2021). Without a guiding theory, SHAP values explain the mathematical weight of a feature, but not necessarily its spatial or social significance.

This theoretical deficit also compromises the spatial resolution of analysis. Lacking a framework grounded in experiential space, researchers often rely on coarse, grid-based units, which obscure micro-scale inequalities and local discontinuities, particularly in fragmented urban typologies (Gavrilidis et al., 2022). Although recent efforts attempt to employ finer units such as streetscape segments (Fang et al., 2025), these often remain statistical constructs rather than theoretically grounded units of lived experience (Gehl, 2013; Lynch, 1964). Consequently, there is a pressing need for a framework that not only utilizes granular units and explainable algorithms but anchors them within a robust spatial theory.

Taken together, these studies offer strong foundations, but few workflows integrate structural, perceptual, and experiential evidence at the street-segment scale to diagnose the sources of low exercise supportiveness. This gap motivates our C/P/L readings and SHAP-based attribution framework.

2.3 Triadic Readings of Urban Space

Henri Lefebvre's spatial triad remains one of the most influential frameworks for interpreting the layered production of urban space (Simonsen, 2005). In The Production of Space (1974), Lefebvre conceptualized space through three interrelated dimensions: conceived space (formal representations and planning rationalities), perceived space (sensory cues and everyday practices), and lived space (symbolic meanings and embodied experiences). This perspective has been widely mobilized across geography, sociology, planning, and architecture to examine how inequalities are produced not only through infrastructure but also through perception and practice (Soja, 2010; Harvey, 2006). In urban health research, it is particularly valuable because it shows how exclusion may stem as much from perceptual and experiential conditions as from material deficits (McLees, 2013; Pierce & Lawhon, 2015).

Applied to street-based exercise, the triad provides an interpretive orientation for understanding why environments that appear similar in form may diverge in their capacity to support movement. Two streets with comparable width or connectivity, for instance, may yield quite different levels of activity depending on safety-related cues, social engagement, and perceived comfort (Handy et al., 2002). Such divergence reflects not only material affordances but also relations of visibility, power, and normativity that shape those who feel entitled to linger, pass, or exercise (Wacquant, 2022). In this sense, exercise deprivation is relational and multidimensional: it arises from the alignment, or misalignment, of planning logics, perceptual cues, and lived practices.

From a geographical perspective, the triad also sharpens long-standing debates about how planned orders meet everyday tactics and negotiations where conceived intentions are appropriated, subverted, or re-made in practice (de Certeau, 1984). Perceived space involves more than visual legibility; it comprises atmospheres, affordances, and embodied judgements that mediate movement, comfort, and safety-related perceptions (Anderson, 2009). Lived space is relational and plural, constituted by intersecting trajectories rather than a single, coherent narrative (Massey, 2005). These insights matter for street-based exercise: the same formal configuration can host distinct rhythms of use, confidence, and co-presence as conditions shift across groups and times of day (Edensor, 2012).

In bridging these theoretical insights with empirical analysis, we adopt the triad not as a rigid taxonomy, but as a flexible reading. In this framework, Conceived space foregrounds institutional form and hierarchy; Perceived space attends to sensory and visual conditions; and Lived space traces rhythms of presence and co-activity. Crucially, we recognize that boundaries between these dimensions remain porous. Sidewalks, road hierarchy, POI diversity, and social-media traces can each carry structural, perceptual, and lived meanings at the same time. We therefore assign each indicator to its most salient reading for the exercise-supportiveness question, while treating overlap as part of the relational character of street space.

Taken together, these dimensions offer a structured yet adaptable vocabulary for examining how movement takes shape along streets. The key analytical challenge is to connect this vocabulary to street-scale evidence without reducing exercise supportiveness to a single accessibility or walkability score. This orientation is consistent with critical quantitative geography, which treats quantitative evidence as a way to open theoretically informed spatial interpretation rather than close it through purely technical measurement (Kwan & Schwanen, 2009a, 2009b). In this study, the C/P/L readings organize the evidence around three diagnostic questions: whether planned street structure provides a usable baseline, whether visible street conditions support comfortable movement, and whether activity traces indicate everyday social use. The subsequent SHAP typology and mismatch analysis then examine how these dimensions combine across Shenzhen's street network.

3 Methodology

To diagnose street-level exercise supportiveness deficits in a spatially structured and interpretable manner, we developed a four-stage analytical framework. The workflow connects street-level exercise intensity with C/P/L indicators, SHAP-based attribution, typology construction, and the bivariate analysis of population density and Keep-derived exercise intensity.

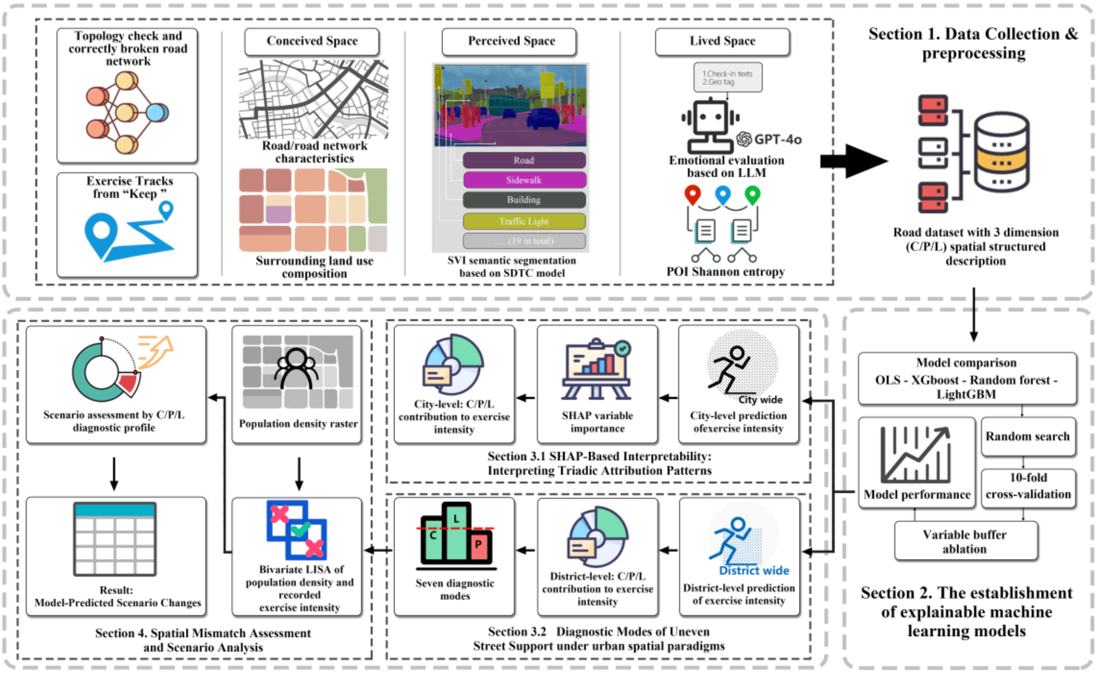


Figure 1 Analytical Workflow: A Theory-Informed, Explainable Pipeline. (1) Data Integration: Aligning multi-source data to street segments to construct C/P/L descriptors. (2) Predictive Modelling: Training gradient-boosted trees (XGBoost) with cross-validation to capture non-linear spatial associations. (3) SHAP-Based Interpretation: Decomposing predictions into triadic attribution patterns at citywide and district scales and synthesizing these contributions into a seven-mode diagnostic typology of supportiveness deficits. (4) Spatial Mismatch Assessment: Comparing population density with recorded Keep-derived exercise intensity and linking the resulting clusters to the diagnostic typology for follow-up planning assessment.

As summarized in Figure 1, the study follows a diagnostic sequence. The workflow begins by aligning multi-source urban data to street segments, then constructs C/P/L proxy indicators for planned structure, visible street conditions, and lived activity traces. Keep trajectories are converted into segment-level exercise-intensity outcomes under alternative matching bands. The resulting street-level dataset is used for model comparison, SHAP-based attribution, seven-mode diagnostic classification, and the bivariate analysis of population density and recorded Keep-derived exercise intensity.

In this sequence, exercise intensity refers to the observed Keep-derived street outcome. Exercise-supportiveness deficits refer to grouped SHAP attributions associated with lower modelled intensity, and mismatch denotes the final comparison between population density and Keep-derived exercise intensity.

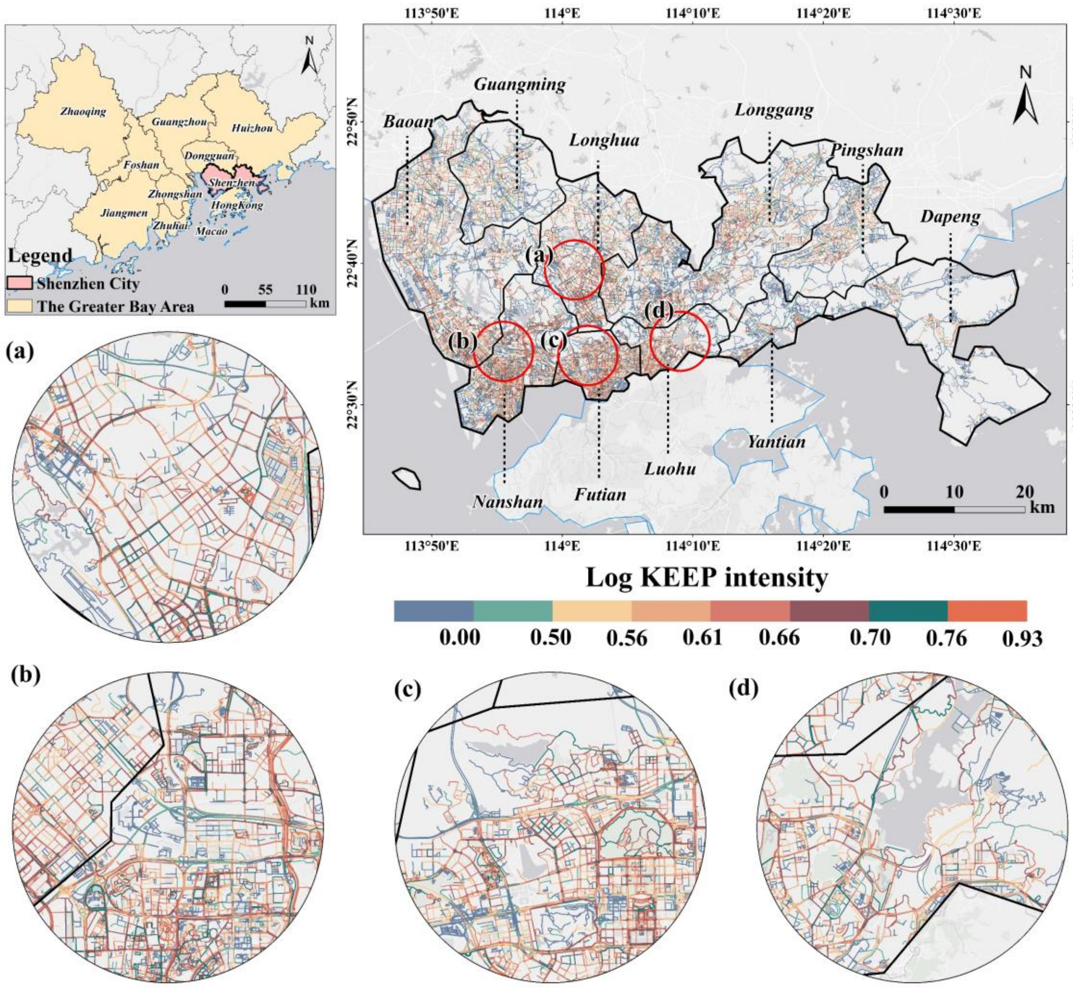


Figure 2 Spatial distribution of street-segment exercise intensity in Shenzhen. The inset map shows the location of Shenzhen within the Greater Bay Area. The main panel presents the log-transformed KEEP intensity at the street-segment level across Shenzhen.

We situate our analysis in Shenzhen, a rapidly growing megacity in southern China (Figure 2). As a frontier of varying urban experiments, Shenzhen provides a compelling laboratory to examine the spatial production of exercise opportunities. Its juxtaposition of rapid, top-down master planning (Conceived space) with vibrant, often informal bottom-up activity (Lived space) creates a complex landscape ideal for diagnosing street-level exercise supportiveness. Uniquely, the city possesses high-resolution, multi-source urban data, enabling a fine-grained empirical translation of the conceived, perceived, and lived representations central to our framework. To ensure analytical consistency, all datasets were aligned to the road segment level and underwent rigorous topological preprocessing (Table 1).

In this study, street space denotes each street-network segment and its surrounding built, visual, and activity context. The main sample therefore retains some automobile-oriented links, such as expressways, trunk roads, ramps, and high-capacity arterials, because they also shape nearby exercise environments and may represent car-oriented constraints on everyday street life. We further test the sensitivity of the results with an accessible-street subset (see *Appendix* E).

Table 1 Data sources and description table

	Data Source	**Description**	**Variable**
Exercise Trajectory	KEEP 2019–2020 exercise trajectory data (KEEP Inc., 2025)	Anonymized Keep GPS trajectories record jogging, walking, and cycling activity among platform users.	'log_d10_norm', 'log_d20_norm', 'log_d30_norm'
Conceived Space	*2024 Amap detailed road network* (Amap, 2025)	High-resolution road network dataset containing road type, hierarchy, connectivity, and functional attributes.	C_free_speed； C_capacity … C_ deg 800m; C_betw_800m C_clo_800m; C_depth_800m
	1m resolution land use remote sensing (Li et al., 2023)	Landuse classification raster providing fine-scale spatial composition around roads, supporting objective environmental context analysis.	C_D_transport; C_D_tree … C_D_water; C_D_building
Perceived Space	Baidu Street View 2024 segmentation results based on STDC model (Fan et al., 2021)	Street-level semantic features extracted from 2024 Baidu Street View using the state-of-the-art STDC segmentation model to represent visual perception.	P_road; P_sidewalk … P_vegetation; P_sky
Lived Space	*Weibo check-in data 2021-2025* (Sina Weibo, 2025)	Spatiotemporal social media data capturing public expression and spatial behaviour.	L_total_weibo_count L_sport_mean; L_urban_mean L_positive_mean L_context_mean
	Amap 2021 POI (Amap, 2025)	Point-of-interest data offering categorical diversity and entropy, representing functional heterogeneity around streets.	L_poi_entropy300

Mismatch analysis	*2022 90m resolution population grid data* (Lebakula et al., 2025)	Gridded population data with 90m resolution reflecting spatial distribution of urban residents, used for mismatch diagnosis with exercise intensity.	Grid population

Table 1 reports the year or period of each data layer. The analysis should be read as an integrated multi-source cross-sectional diagnosis. Some layers, including road hierarchy, network centrality, and broad capacity, are relatively stable over short periods, while recorded exercise activity, Weibo expression, POI composition, and local streetscape scenes may change more quickly, especially in recently redeveloped areas.

As temporal consistency checks, the PRD/GBA 2019-2020 Keep grid comparison and annual Weibo-window analysis show that broad activity and lived-space patterns remain recognizable across available annual files; the full results are reported in *Appendix* H.

The analysis centers on street-level physical activity intensity as the dependent variable, representing the observed outcome of spatial interaction. This metric was derived from anonymized GPS trajectories recorded by the Keep fitness platform (Keep Inc., 2025), capturing jogging, walking, and cycling behaviours. Because the records span 2019-2020, they may also reflect COVID-period changes in outdoor activity and recording behaviour; *Appendix* I provides a brief contextual note and a 2019/2020 year-split check. For each road segment, intersecting trajectory points were counted and normalized by segment length, then log-transformed and z-score standardized to produce robust exercise density metrics under three distance bands (10 m, 20 m, and 30 m). *Appendix L* reports the raw Keep trajectory volume, sampling procedure, projection, and segment coverage under the three matching bands. The resulting spatial pattern of exercise intensity across Shenzhen is shown in Figure 2. The three distance bands represent increasingly permissive trajectory-to-street matching tolerances. We use the 30 m band as the main outcome because it retains the broadest street coverage, while the 10 m and 20 m bands provide stricter construction checks in *Appendix L*. The log transformation reduces the influence of highly skewed activity counts, and z-score standardization keeps continuous variables comparable across the modelling workflow.

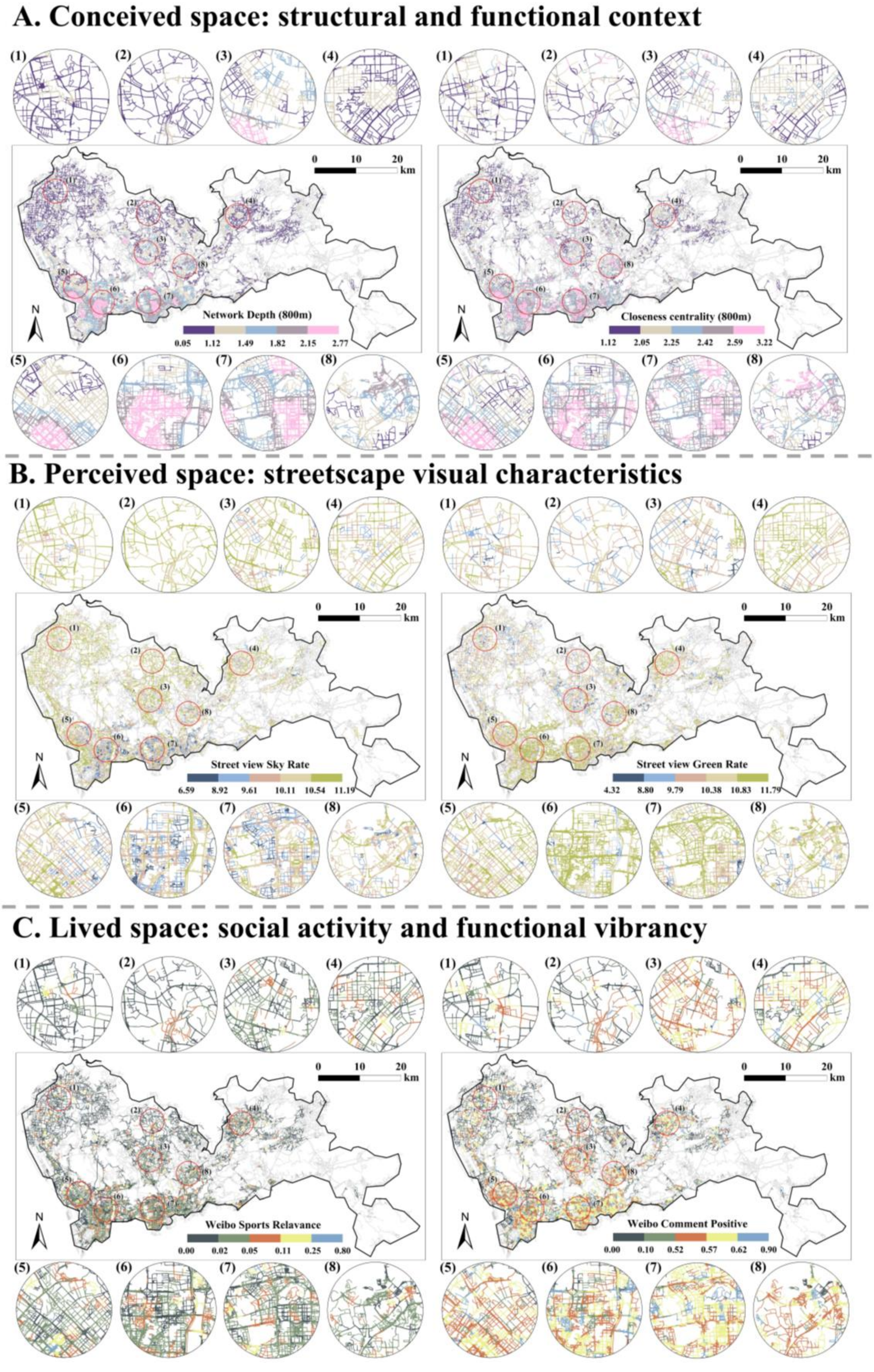


Figure 3 Representative spatial distributions of street-level indicators across the conceived, perceived, and lived dimensions.

To construct a triad-informed empirical reading of street conditions, we organized multi-source indicators into conceived, perceived, and lived proxy groups at the street-segment level. Figure 3 presents six representative indicators selected from these three proxy groups, highlighting both their citywide spatial distributions and their neighborhood-scale heterogeneity through enlarged local views.

Conceived space, reflecting the "representations of space" and planning rationalities, was characterized through the structural skeleton of the street network and its functional context (Figure 3a). We constructed this description using two primary sources. First, a 2024 high-resolution road network dataset from Amap (Amap, 2025) provided intrinsic metadata including free-flow speed, design capacity, and functional classification. To capture topological hierarchy, network geometries were corrected to ensure graph validity, allowing the calculation of global and local centrality measures (degree, closeness, betweenness) within 800-meter ego-graphs, which provide a pedestrian-scale local network context for comparing street accessibility. Complementing this network logic, we incorporated 1-meter resolution land use classification data (Li et al., 2023) to compute the surrounding functional composition within 100-meter buffers, delineating the planned context through proportions of residential, commercial, transportation, vegetation, water, and building classes.

Perceived space corresponds to the sensory deciphering of the environment, representing the visual interface encountered by pedestrians. We characterized this dimension by applying semantic segmentation to 2024 Baidu Street View imagery using the state-of-the-art STDC model (Fan et al., 2021). *Appendix* N reports the model benchmark information and a local output plausibility audit for this streetscape-segmentation layer. Frontal and lateral views were sampled along street networks and parsed into visual categories such as sky, vegetation, sidewalk, façade, and vehicle (Figure 3b). These aggregated visual proportions were normalized by road length to form a perceptual profile for each segment, reflecting key experiential qualities such as visual enclosure, greenery, and the legibility of urban form.

Lived space embodies the "representational spaces"—the complex, symbolic, and affective reality overlaid on the physical form. Since this dimension reflects dynamic rhythms of practice rather than static objects, we approximated it through two digital proxies. First, over 1.5 million Weibo check-in posts (2021–2025) were scored with GPT-4o using a structured JSON prompt across eight experiential dimensions, including urban relevance, affective tone, sport-relatedness, safety/walkability cues, and immediate geographical context (*Appendix* A). *Appendix* A reports three complementary validation exercises. A preliminary 50-post human–machine consistency check for the positive indicator showed high agreement, with human–human accuracy of 0.88 and Cohen's κ of 0.81, and coder–GPT-4o accuracy of 0.85–

0.86 with κ values of 0.77–0.78. In the cross-model check, the GPT-4o and Kimi scores were strongly correlated across the core urban and positive indicators (Pearson r=0.859; Spearman ρ=0.776; MAE =0.089). In the larger 300-post human-coder validation, the average agreement between the three coders and GPT-4o was moderate (mean Pearson r=0.707; mean Spearman ρ=0.601; mean MAE =0.169). The Weibo layer provides a complementary trace of digitally mediated social expression as one observable aspect of lived street activity. Second, we calculated POI Shannon entropy from 2021 Amap POI data (Amap, 2025) to represent the diversity of surrounding amenities. Together, these indicators capture the symbolic, affective, and social dimensions of urban space that go beyond physical form.

While our primary predictive modelling operates at the street segment level, the final comparison between population density and recorded exercise intensity requires demographic data available only in raster format. We therefore rasterized the study area into a uniform 200 × 200m grid to bridge these spatial units. The 200 m cell size provides a neighborhood-scale overlay that is fine enough to retain local variation in recorded exercise intensity while large enough to contain sufficient street length and population counts for stable aggregation; *Appendix* O reports grid-size sensitivity checks. Population estimates from 2022 90-meter raster data (Lebakula et al., 2025) were aggregated into this new grid using area-weighted interpolation. Simultaneously, street-level C/P/L features and exercise variables were aggregated using length-weighted means within each grid cell. Finally, all continuous variables across the workflow (except categorical descriptors) were log-transformed and z-score normalized to ensure statistical comparability and suitability for gradient-based machine learning.

3.1 Quantifying Spatial Drivers: Predictive Modelling and SHAP Attribution

With the triad-informed dataset established, we evaluated how well the grouped C/P/L variables explain street-level exercise intensity. Specifically, we used the C/P/L predictors for each street segment to regress the corresponding Keep-derived exercise-intensity outcome, comparing four models with increasing flexibility: Ordinary Least Squares, Random Forest, LightGBM, and XGBoost. Because all models used the same street-segment predictors and outcome, the comparison identifies a suitable predictive basis for subsequent SHAP attribution. Hyperparameter tuning was performed via randomized search with 10-fold cross-validation to compare model performance under random folds. The parameter space and final selected hyperparameters for each model are detailed in *Appendix* C.

3.2 From Attribution to Diagnosis: SHAP-Based Interpretation and Supportiveness-Deficit Typology

Building on the best-performing model, we applied SHapley Additive exPlanations (SHAP) to decompose predictions into additive contributions of input features. Feature-level SHAP values were summed within the C, P, and L indicator groups for each segment. Signed grouped values identify whether a reading raises or lowers the predicted exercise intensity, while absolute grouped values are used to summarize contribution shares. The citywide and district-level SHAP share diagrams are normalized within their respective samples, so the values should be read as compositional attribution profiles. This interpretation was conducted at two scales: a citywide reading that summarizes contributions into the three canonical dimensions, and a district-level refinement where the conceived dimension was further disaggregated into structural and functional components to capture local heterogeneity in planning legacies. To check predictor redundancy before interpreting SHAP attributions, we conducted Pearson/Spearman correlation diagnostics, VIF analysis, and a correlation-pruned sensitivity model; the full results are reported in *Appendix* M.

We then translated grouped SHAP attributions into a diagnostic typology of model-attributed supportiveness deficits. For each segment, feature-level SHAP values were summed within the C, P, and L indicator groups. A C/P/L reading was flagged when its grouped contribution was negative and fell within the selected low-supportiveness tail, indicating that the corresponding indicator group lowered predicted exercise intensity relative to the learned model baseline. The low-supportiveness tail is defined citywide using the strongest 20% of negative grouped contributions. Combining the three C/P/L flags yields seven diagnostic modes: C-only, P-only, L-only, CP, CL, PL, and CPL (shown in Table 2). To make the SHAP attribution results comparable across streets, we translate the signed grouped C/P/L SHAP contributions into diagnostic supportiveness-deficit configurations. Formally, a segment s is classified using the signed grouped SHAP contributions $S_C(s)$, $S_P(s)$, and $S_L(s)$:

$$D_k(s) = \mathbb{1}\big(S_k(s) < 0 \text{ and } |S_k(s)| \geq Q_{0.80}(\{|S_k(u)|: S_k(u) < 0,\ u \in \mathcal{S}\})\big), \quad k \in \{C, P, L\}.$$

$$Type(s) = \{k \in \{C, P, L\}: D_k(s) = 1\}.$$

where s denotes a street segment, $k \in \{C, P, L\}$ denotes the conceived, perceived, or lived reading, $\mathcal{F}_k$ is the set of features assigned to dimension k, and $\phi_j(s)$ is the SHAP value of feature j for segment s. $S_k(s) = \sum_{j \in \mathcal{F}_k} \phi_j(s)$ is the signed grouped SHAP contribution of dimension k. $Q_{0.80}(\cdot)$ denotes the 80th percentile of the absolute negative grouped SHAP contributions. $\mathbb{1}(\cdot)$ is an indicator function, $D_k(s)$ is the binary deficit flag for dimension k, and $Type(s)$ records the flagged C/P/L dimensions for segment s.

Table 2 Definition of SHAP-based diagnostic modes.

Mode	Judgment criterion	Diagnostic meaning
Not deprived	D_C=0, D_P=0, D_L=0	No C/P/L grouped contribution is in the low-supportiveness tail.
C-only	D_C=1, D_P=0, D_L=0	Conceived indicators alone lower modelled exercise supportiveness.
P-only	D_C=0, D_P=1, D_L=0	Perceived streetscape indicators alone lower modelled supportiveness.
L-only	D_C=0, D_P=0, D_L=1	Lived/platform-mediated indicators alone are weak or negative.
CP	D_C=1, D_P=1, D_L=0	Conceived and perceived supportiveness deficits co-occur.
CL	D_C=1, D_P=0, D_L=1	Conceived and lived supportiveness deficits co-occur.
PL	D_C=0, D_P=1, D_L=1	Perceived and lived supportiveness deficits co-occur.
CPL	D_C=1, D_P=1, D_L=1	All three C/P/L readings lower modelled supportiveness.

We also examined the stability and internal consistency of the C/P/L attribution. *Appendix* K assesses classification under alternative citywide and district-specific thresholds. *Appendix* F shows that observed Keep intensity declines as more C/P/L readings are flagged. Because the SHAP contributions were fitted to the same outcome, this ordering indicates internal consistency; validation against an independent behavioural or field-based measure remains necessary. Table 2 further explains the structural, streetscape, and activity-context conditions represented by each diagnostic mode. *Appendix* G shows that the population-density and GDP check captures broad exposure and socioeconomic context.

3.3 From Diagnosis to Spatial Mismatch Assessment

Finally, we conducted a Bivariate Mismatch Analysis to compare gridded population density with recorded Keep-derived exercise intensity. Using the 200-meter grid integrated in the data construction phase, we applied Local Indicators of Spatial Association (LISA) to identify localized clusters where high population density coincides with low recorded exercise intensity. The grid layer is used only in this final overlay step, where population density and street-level exercise intensity are aggregated within each cell. These High-Population/Low-Exercise-Intensity clusters were then cross-referenced with the segment-level diagnostic typology. Population density indicates where residents are concentrated; exercise demand may also vary with demographic composition, health, preferences, socioeconomic conditions, and access to alternative facilities. This synthesis links the observed spatial pattern to the model attribution: the grid analysis identifies areas where high population density coincides with low recorded exercise intensity, while the typology indicates which C/P/L dimension is most relevant for local diagnosis.

This framework connects prediction, attribution, typology, and mismatch screening to support later interpretation of where modelled street supportiveness is low and which C/P/L readings are most relevant for follow-up assessment.

4 Analyses and Discussion

4.1 Model Performance and Predictive Basis for Attribution

Following the model-comparison procedure described in the Methods section, Table 3 reports the predictive performance of OLS, Random Forest, LightGBM, and XGBoost for the same street-level outcome and C/P/L predictor set.

The comparative results in Table 3 show that the flexible ensemble models fit the observed exercise-intensity pattern better than the linear baseline under random cross-validation. XGBoost yields the highest mean R^2 (0.7023 ± 0.0043) and the lowest RMSE (0.0449 ± 0.0003), and is used for the subsequent SHAP analysis. The gap from the linear model is consistent with possible interactions, thresholds, and local heterogeneity in the observed associations.

Performance declines as validation groups become more spatially separated. Random five-fold cross-validation yields R^2 = 0.690, compared with 0.461 for 500 m grid-grouped validation and 0.330 for subdistrict-grouped validation; alternative street-office boundaries produce a similar R^2 of 0.327. This decline indicates more limited transfer across neighborhoods. *Appendix* J reports the full multi-scale results.

Table 3 Cross-validated model performance comparison across regression algorithms (mean ± standard deviation over 10 folds).

Model	CV R² (Mean ± Std)	CV RMSE (Mean ± Std)
XGBoost	0.7023 ± 0.0043	0.0449 ± 0.0003
RandomForest	0.3995 ± 0.0043	0.0638 ± 0.0005
LightGBM	0.5661 ± 0.0047	0.0542 ± 0.0004
LinearRegression	0.2800 ± 0.0137	0.0699 ± 0.0010

4.2 Interpreting Triadic Attribution Patterns: Multi-Scalar SHAP Interpretation

4.2.1 Global Attribution Hierarchy and Non-linear Associations

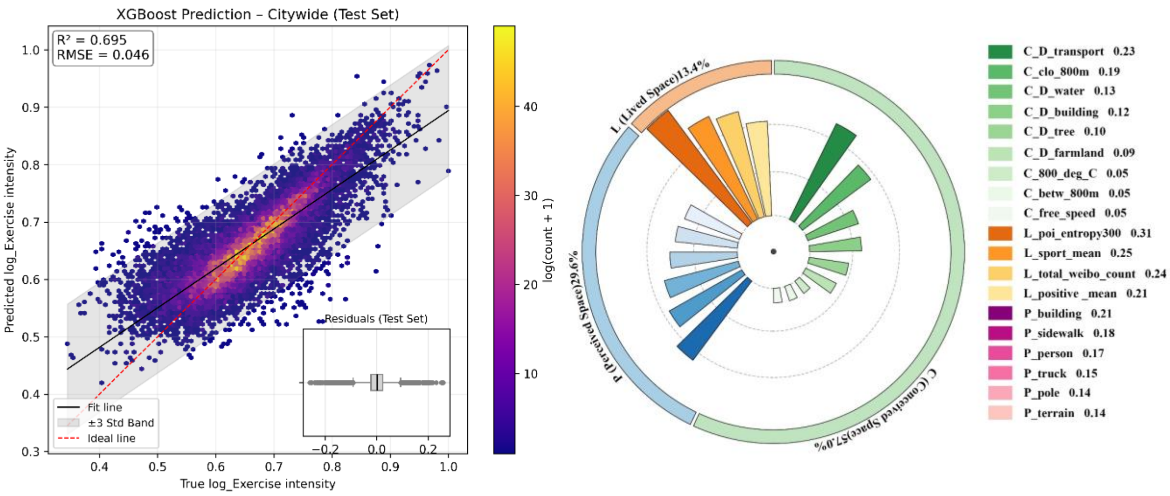


Figure 4 Citywide model fit and triadic SHAP contribution overview (single random hold-out test set).

To unpack the spatial drivers of exercise intensity, we first examined model performance and feature attribution at the citywide scale. In the random hold-out sample shown in Figure 4, predicted and observed exercise intensity yield an R^2 of 0.695. Decomposing the predictions by SHAP group indicates that conceived-space indicators account for the largest share of model attribution (57.0%), followed by perceived-space indicators (29.6%) and lived-space indicators (13.4%). This pattern suggests that planned street structure, transport infrastructure, and land-use context provide the strongest statistical baseline in the fitted model. Perceived and lived indicators remain important within the fitted model, although their attribution shares do not establish causal mechanisms.

The SHAP summaries in Figure 5 further disaggregate these dimensions, showing how specific features are associated with predicted movement patterns. Within the dominant Conceived dimension, variables such as transport density ($C\_D_{\text{transport}}$), closeness centrality ($C_{\text{clo\_800m}}$), and proximity to water bodies (C_D_{water}) display consistent, positive decision-path trajectories. These SHAP trajectories suggest that higher functional connectivity and spatial accessibility are associated with higher predicted exercise intensity. Consequently, the urban network structure appears as an important baseline condition for street-level movement in the fitted model.

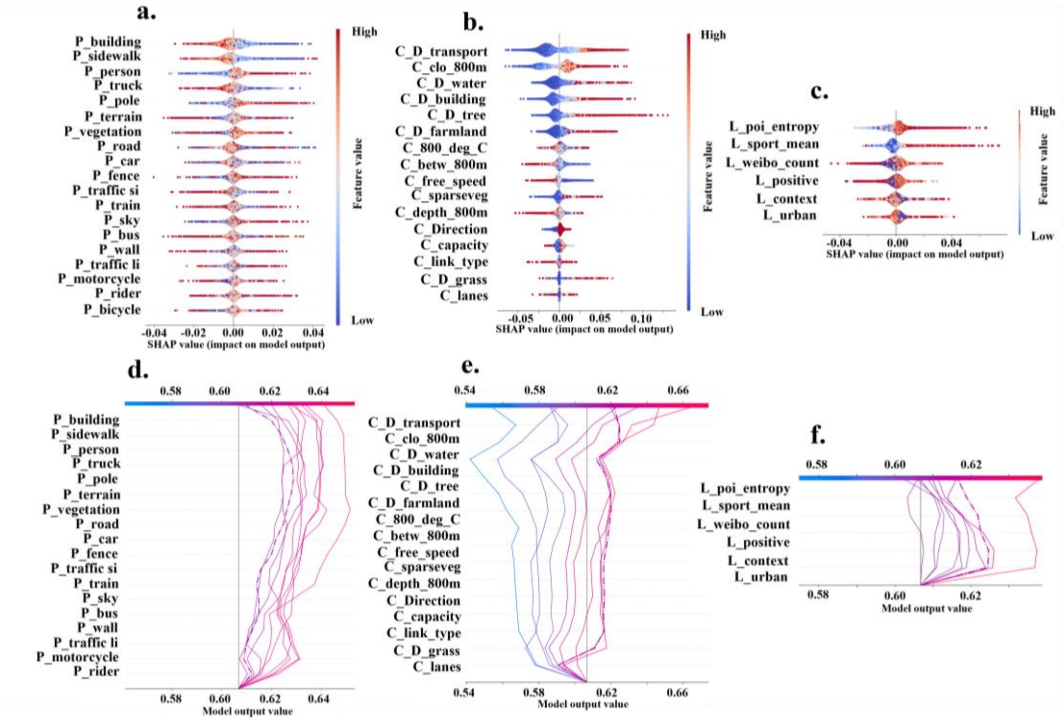


Figure 5 SHAP summary for (a.) Perceived Space variables (P); (b.) Conceived Space variables (C); (c.) Lived Space variables (L), and decision path for (d.) Perceived Space variables (P); (e.) Conceived Space variables (C); (f.) Lived Space variables (L)

In contrast, perceived-space features derived from street-level imagery show more complex, nonlinear or threshold-like associations in the fitted model. Variables such as building enclosure (P_{building}) and sidewalk proportion (P_{sidewalk}) exhibit both positive and negative contributions depending on their intensity. The decision paths reveal specific thresholds beyond which additional built form or infrastructure density begins to yield diminishing returns, lowering predicted exercise intensity rather than raising it. Similar saturation effects appear for terrain features (P_{terrain}). These patterns are consistent with the idea that sensory environments may support movement within certain ranges of enclosure and visual complexity, a reading that aligns with research on legibility and human-scale comfort.

Finally, lived-space features contribute more modestly to the overall attribution profile but become more influential at high quantiles. Indicators such as total Weibo activity, positive sentiment, and sport-related discourse show sharper contributions once social activity or emotional positivity is already high. This pattern is interpreted as a threshold-like association within the fitted model rather than direct evidence that social vibrancy causes further movement. Taken together, these results suggest a complementary attribution pattern across the triad-informed indicator groups: structural indicators describe baseline access conditions, visible streetscape indicators point to sensory-comfort thresholds, and lived-space proxies mark contexts where recorded social activity is already stronger.

4.2.2 District-Level Heterogeneity and Spatial Signatures

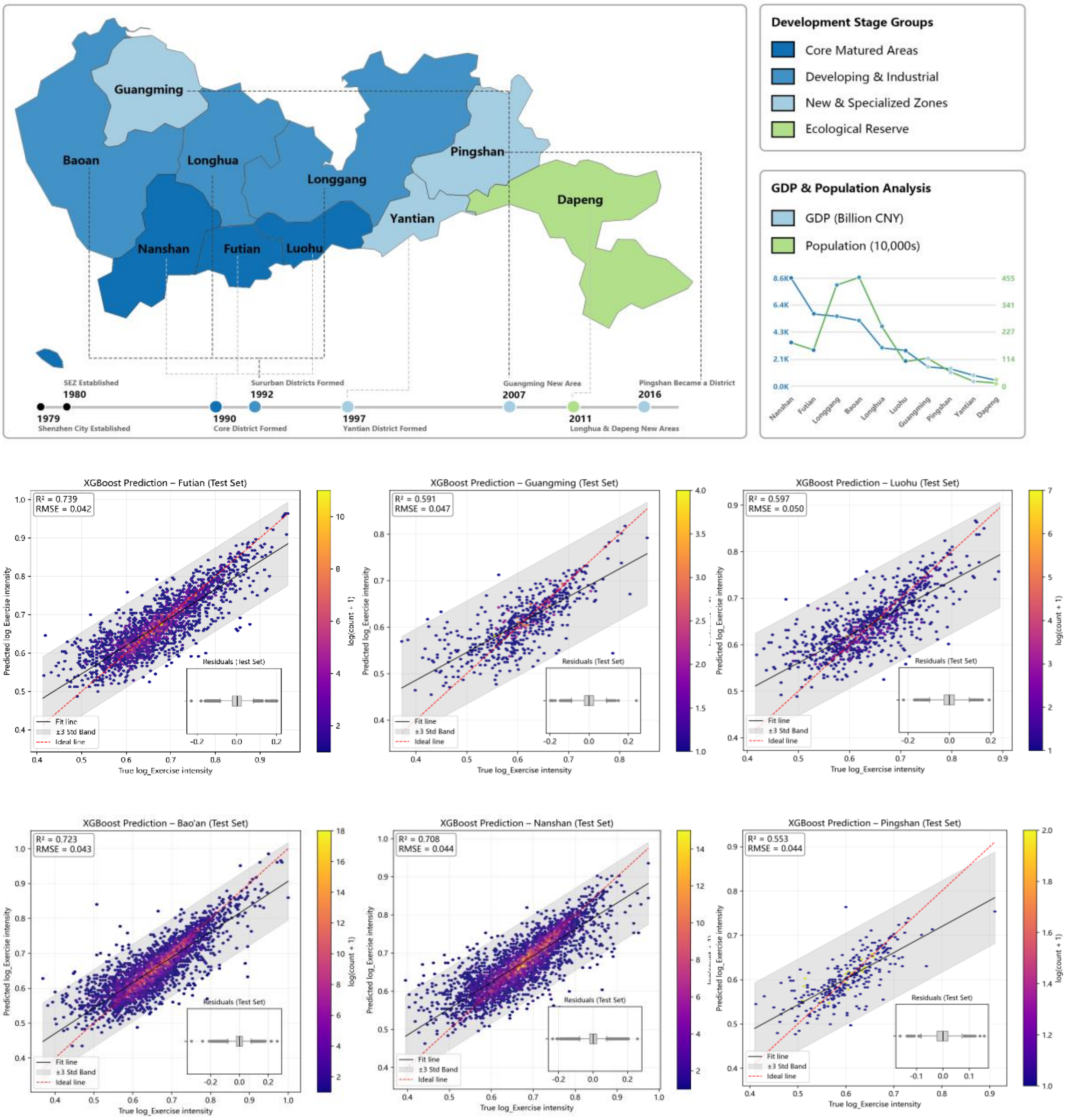


Figure 6 District-specific XGBoost model performance, development stage context, and demographic–economic profiles across six Shenzhen districts. The upper panel shows Shenzhen's administrative districts grouped by development stage (core matured, developing/industrial, new and specialized zones, ecological reserve) with a timeline of key district formations (1979–2016). The line graph depicts 2022–2023 GDP (billion CNY) and population (10,000s) for each district, illustrating marked differences in economic scale and demographic scale. The lower scatter plots display predicted versus observed street-level exercise intensity from district-specific XGBoost models, with 95 % confidence bands and residual histograms.

While the citywide analysis reveals broad patterns, it risks masking local variations rooted in urban development history. To examine this spatial heterogeneity, we conducted a subregional decomposition across six representative districts spanning a developmental gradient: Futian and Luohu (mature institutional cores), Bao'an and

Nanshan (evolving innovation-industrial zones), and Guangming and Pingshan (emerging peripheral new towns). By synthesizing district-specific XGBoost scatterplots with development timelines and socioeconomic profiles (Figure 6), we contextualized how varying governance regimes and urban form maturity are associated with localized supportiveness patterns.

The district-specific models show that within-district fit varies across development contexts. R^2 reaches 0.739 in Futian and is approximately 0.55 in Guangming and Pingshan. In the mature cores, recorded Keep activity is more consistently associated with measured street and land-use indicators. Lower fit in peripheral new towns may reflect weaker spatial legibility, fragmented public-realm conditions, and unmeasured behavioural or development-stage factors (Lynch, 1964; Batty, 2013).

To interpret these local attribution patterns, we refined our analytical resolution by disaggregating the Conceived dimension into (road-intrinsic connectivity) and (surrounding functional context). The resulting SHAP composition rose diagrams (Figure 7) visualize distinct spatial signatures that evolve with the district's development stage. In the mature cores of Futian and Luohu, the profiles exhibit a balanced triadic synergy, with a mix of structure, land use, and perceived quality complemented by a visible contribution from Lived space. In Luohu, lived space accounts for nearly 15% of the grouped SHAP attribution, aligning with its long-established urban fabric where sustained economic density has produced a vibrant social presence.

The innovation-industrial zones display a diverging trajectory. Nanshan shows high Perceived-space contributions alongside a robust Structural base, reflecting a "co-evolution" where infrastructure planning and streetscape beautification have advanced in tandem. In contrast, Bao’an displays a mismatch where Land-use and Perceived factors dominate but lived contribution is weaker, reflecting its status as a maturing mix where large-scale planning has established the skeleton but human-scale social activation lags behind. Finally, in the peripheral new towns of Guangming and Pingshan, the spatial logic is overwhelmingly dominated by macro-planning ($C_{landuse}$), suggesting that activity patterns remain heavily structured by coarse zoning rather than fine-grained human-scale affordances. These observed asymmetries suggest that low supportiveness is also shaped by development timing, motivating the diagnostic typology that follows.

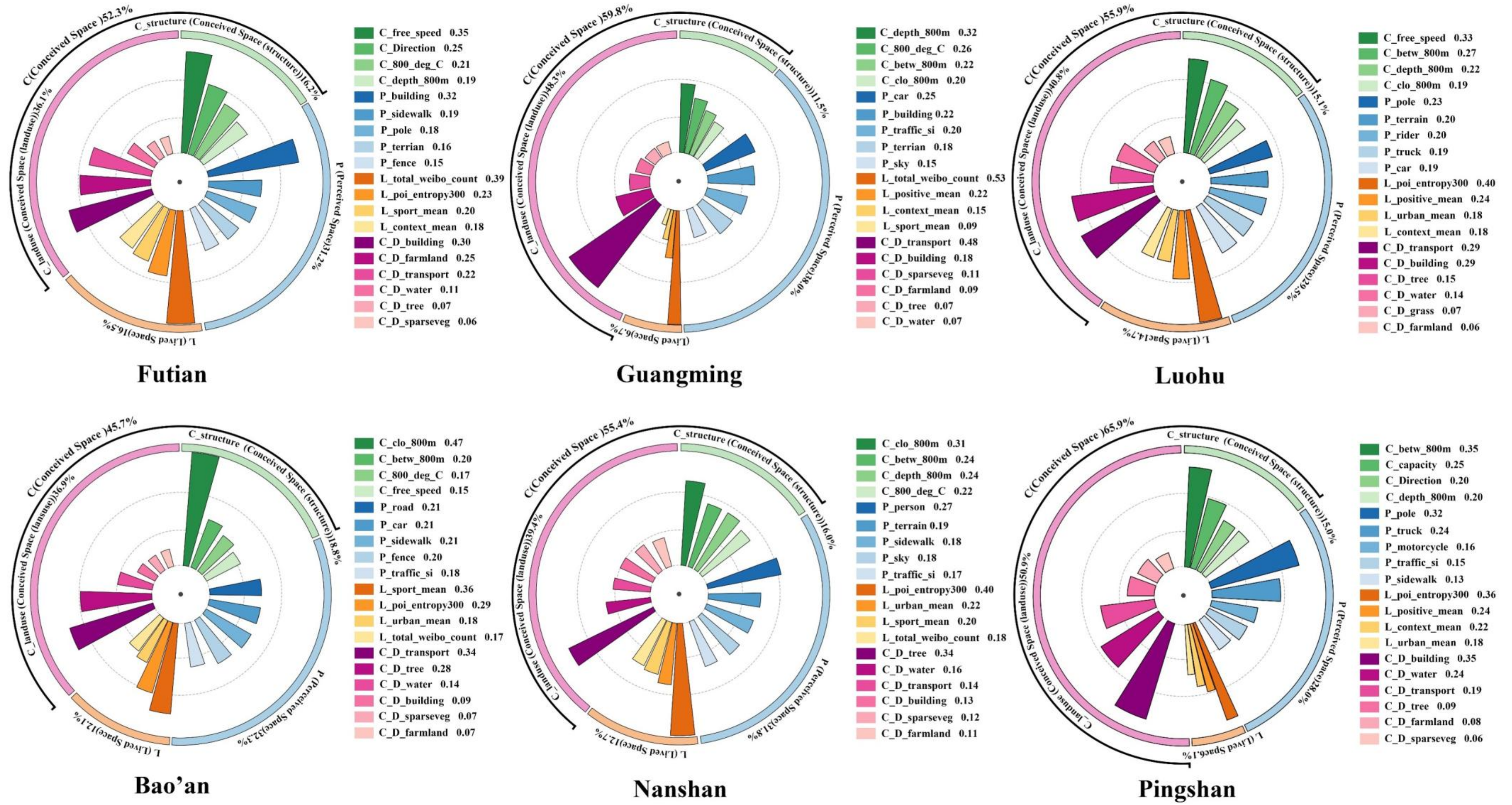


Figure 7 District-level SHAP-based triadic contribution rose diagrams. The diagrams compare C/P/L attribution profiles across six representative districts. The outer structure keeps the conceived, perceived, and lived ordering used in the citywide SHAP summary, while the conceived-space contribution is further shown as a nested split between road-intrinsic network structure and surrounding land-use context.

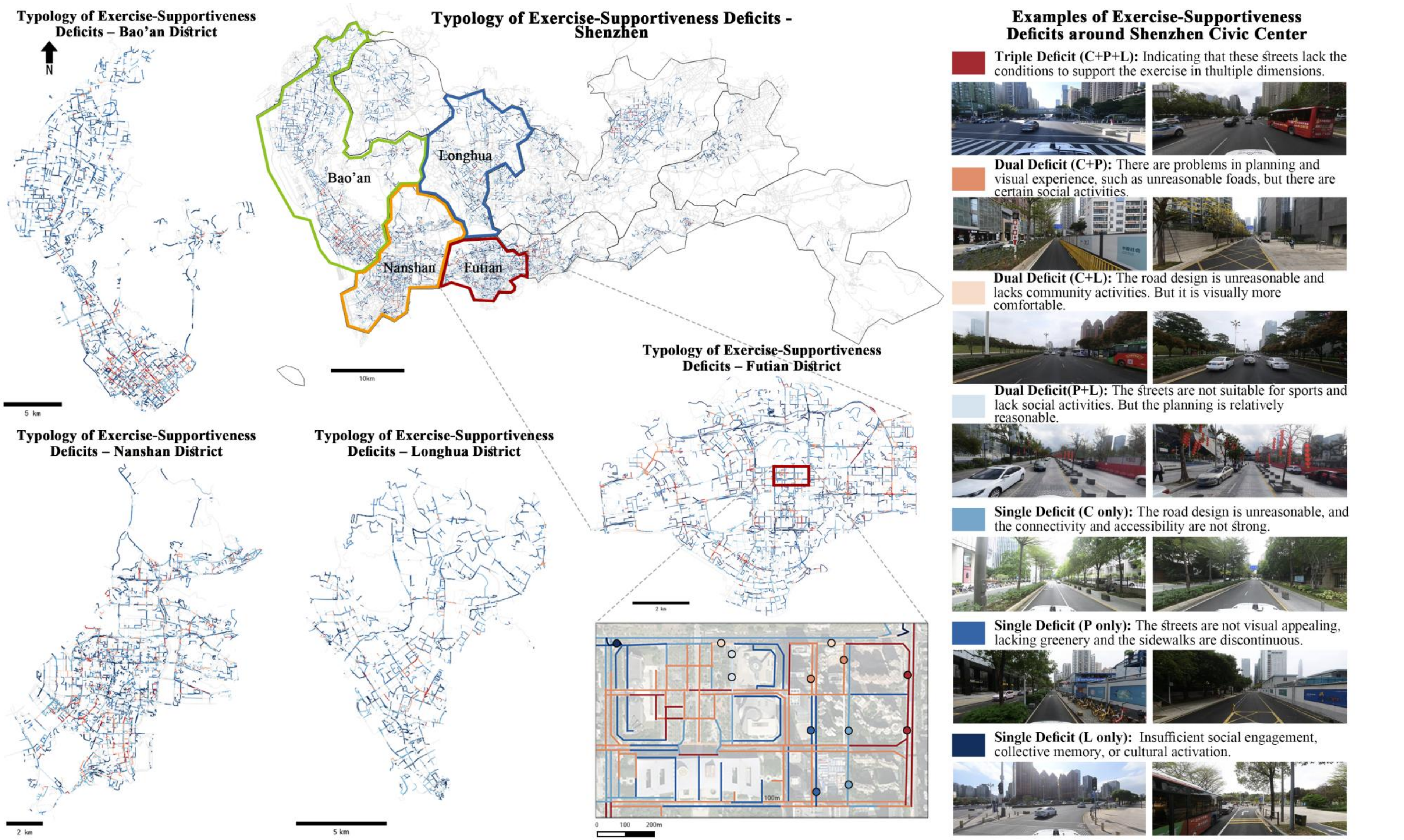


Figure 8 SHAP-based diagnostic typology of street-level exercise supportiveness deficits across Shenzhen and representative districts. The left and center panels map the seven diagnostic modes (C-only, P-only, L-only, CP, CL, PL, CPL) at the street-segment level for Shenzhen as a whole and for four representative districts (Futian, Nanshan, Bao'an, and Longhua). The inset zooms into a cluster near the Shenzhen Civic Center to illustrate fine-grained patterns. The right-hand column shows ground photographs of typical streets corresponding to each diagnostic mode, from CPL to single-domain modes, illustrating how C/P/L readings associated with lower modelled supportiveness appear in selected street examples

4.3 A Diagnostic Typology of Uneven Street Support for Exercise

To translate these triadic insights into a street-level diagnosis, we synthesized the SHAP-derived contributions into a spatially explicit typology. By applying a relative threshold to the Conceived, Perceived, and Lived dimensions, each street segment was classified into one of seven diagnostic modes: CPL, CP, CL, PL, C-only, P-only, or L-only. Figure 8 maps these resulting typologies across the city and within four representative districts-Futian, Nanshan, Bao'an, and Longhua-alongside ground-level photographs illustrating each mode. When interpreted in conjunction with the development timelines and socioeconomic profiles presented in Figure 6, this typology provides a multidimensional context for understanding how specific low-supportiveness patterns emerge under distinct development contexts.

In the mature administrative core of Futian, which is characterized by high GDP and population density, the analysis reveals a prevalence of L-only and PL diagnostic modes concentrated around civic facilities and arterial roads. This indicates that while the structural and visual environmental conditions are adequate, social activation and everyday use often lag behind. This phenomenon echoes the "saturation effect" noted in Section 4.2, suggesting that in highly developed cores, the primary barrier is no longer structural access but the fostering of vibrant social practice.

Conversely, Nanshan, an innovation district developed earlier in Shenzhen's history, displays a different spatial logic. Here, P-only and CP patterns predominate, reflecting a condition where sound infrastructure exists but is compromised by perceptual barriers such as visual clutter, weak shading, or pedestrian discontinuities. This pattern suggests that even in high-growth, mixed-use zones, sensory refinement may remain an important constraint on comfortable and sustained active use.

Moving to the transitional districts, Bao'an exhibits a complex spatial mosaic. Its older industrial corridors are characterized by C-only and CP diagnostic modes, pointing to lingering deficiencies in both network structure and visual quality. In contrast, its newer housing clusters display sporadic L-only diagnostic patterns. This latter pattern mirrors the situation in Longhua, a rapidly developing "new town," where physical connectivity has outpaced social and experiential activation. Longhua's strong prevalence of C-only and L-only diagnostic modes underlines a significant temporal lag: the physical city has been built, but the "lived city" of social memory and cultural activation has yet to fully emerge.

These subregional variations are consistent with the spatial-triad reading that supportiveness deficits often reflect imbalance rather than simple absence. Streets may underperform not only because they lack design or activity, but because planned structure, visible street conditions, and lived activity traces do not align. The SHAP-based typology translates model attributions into a diagnostic language for interpreting which proxy dimensions are associated with lower street-level exercise supportiveness.

Crucially, this typology serves not only an explanatory function but also a planning-diagnostic one. Each diagnostic mode implies a distinct intervention emphasis: structural redesign for C-heavy deficits, environmental articulation for P-heavy deficits, or social activation for L-heavy deficits. This use of model attribution as planning-screening evidence is consistent with critical quantitative geography's effort to make quantitative analysis more interpretive and socially consequential (Kwan & Schwanen, 2009a), while the typology itself should be read as a diagnostic guide rather than a direct measure of spatial justice (Soja, 2010).

4.4 From Diagnosis to Intervention: Spatial Mismatch Analysis

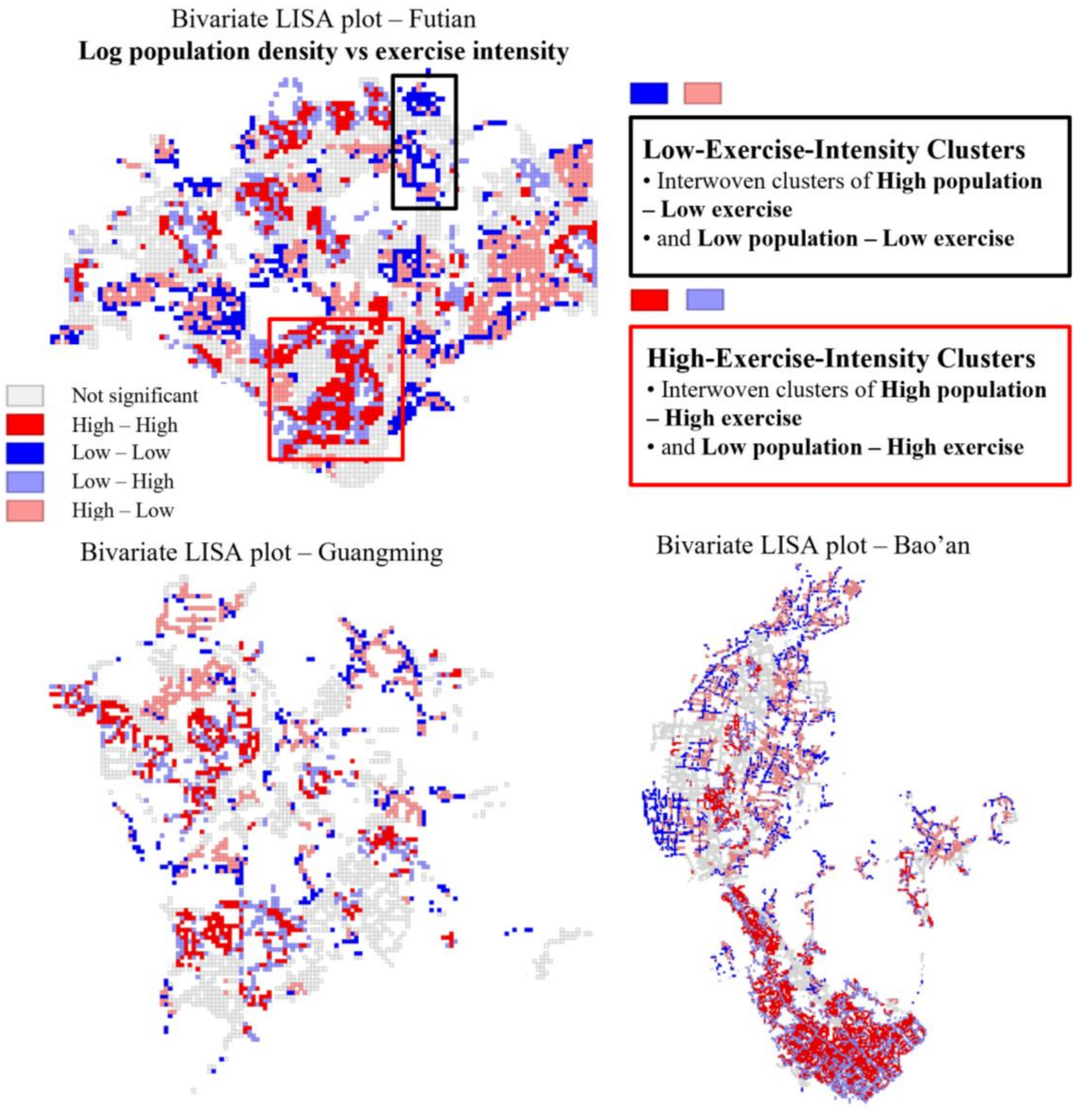


Figure 9 Bivariate LISA analysis of population density and recorded street-level exercise intensity. Maps show local spatial associations across Shenzhen and representative districts. Red, dark blue, light blue, and pink denote High–High, Low–Low, Low–High, and High–Low clusters, respectively. High-Population/Low-Exercise-Intensity clusters identify locations where high population density coincides with low recorded exercise intensity and may warrant closer planning assessment.

In complex urban systems, patterns of recorded exercise intensity can be examined alongside population concentration (Xie et al., 2023). We therefore used grid-level bivariate LISA to compare log population density with recorded Keep-derived exercise intensity. This analysis describes how recorded street-level activity varies across more and less densely populated areas.

As visualized in Figure 9, four local association classes emerge from the bivariate analysis. High-High and Low-Low indicate spatial correspondence between population density and recorded exercise intensity; High-Low identifies locations where high population density coincides with low recorded exercise intensity, while Low-High shows the converse pattern. Other cells have no significant local association.

This spatial reading reveals potential “High-Population/Low-Exercise-Intensity (HL)” clusters. In Futian, for example, mismatch zones cluster around older arterial corridors where density outstrips aging infrastructure; in Bao'an and Guangming, they appear along industrial–residential edges, reflecting a lag in public-realm investment relative to population influx. These patterns suggest that the planning challenge involves not only infrastructure provision but also its distributional alignment with everyday life.

Examining these mismatch clusters also links the diagnostic results to intervention scoping. While the LISA analysis identifies where population–exercise-intensity mismatch is concentrated, the SHAP-derived typology helps indicate which C/P/L reading may be most relevant for follow-up diagnosis. By cross-referencing these HL clusters with the diagnostic modes identified in Section 4.3, we can outline possible intervention emphases. For instance, C-heavy HL grids may call for closer review of connectivity and road structure; P-heavy HL grids may point to streetscape and pedestrian-comfort improvements; and L-heavy HL grids may require community activation strategies or additional field validation. This integration of grid-level targeting and segment-level diagnosis translates model outputs into planning hypotheses for further assessment. Detailed intervention simulations and quantitative assessments of these scenarios are provided in the *Appendix* D.

4.5 Limitations and Scope of Inference

This study provides a diagnostic reading of street-based exercise supportiveness, and its findings should be interpreted within the proxy-based nature of the framework. Lefebvre’s conceived, perceived, and lived spaces are fluid and overlapping categories; translating them into grouped indicators necessarily prioritizes empirical operability over philosophical completeness. The C/P/L variables should therefore be understood as measurable traces of broader spatial processes rather than comprehensive representations of urban experience.

The underlying data layers also carry platform-specific and temporal limitations. Keep trajectories reflect recorded activity among digitally connected fitness users and may underrepresent elderly residents, children, low-income groups, casual walkers, and people who exercise without fitness apps. Activity patterns in 2020 may also be affected by COVID-period disruptions. Similarly, Weibo check-ins capture digitally mediated social expression, while street-view imagery represents visual composition but not the full embodied sensory experience of streets. Safety is also represented indirectly. Future validation could incorporate pedestrian and cycling counts, field observations, resident surveys, health surveys, safety audits, or other activity datasets where available.

Methodologically, SHAP values describe statistical associations and do not establish causal mechanisms, particularly given correlations among urban features. The resulting typology functions as a screening and interpretive tool for recorded exercise supportiveness and does not provide definitive causal explanations or direct evidence of spatial injustice. Because the findings are shaped by Shenzhen's morphological, institutional, and redevelopment context, applications to other cities would require local calibration and, where possible, synchronized annual datasets, longitudinal designs, participatory validation, and field-based evidence to support stronger claims about unmet need or behavioural change.

5 Conclusion

This study extends structurally oriented walkability assessment by bringing visible streetscape conditions and platform-mediated lived-space traces into a triad-informed diagnosis of street-level exercise supportiveness. Using Henri Lefebvre's spatial triad as a lens, the study links environmental form, sensory quality, and social practice to modelled street-level supportiveness. The Shenzhen results show that low supportiveness is often associated with misalignment among these dimensions. Conceived attributes provide an important structural baseline, but local attribution patterns vary, ranging from behavioural saturation in mature cores to structural coarseness in new towns. In this reading, streets may fail to support activity when planned structure, visible street conditions, and lived activity traces fall out of alignment. Practically, the resulting diagnostic typology can support more differentiated planning assessment. Diagnostic profiles can indicate where structural redesign, environmental improvement, or community activation may warrant closer local evaluation. The triad-informed approach provides a structured basis for examining streets as everyday health-supportive infrastructure and for identifying locally relevant planning questions.

Acknowledgements

This study is supported by the Guangdong Basic and Applied Basic Research Foundation (grant no. 2025A1515110224), the AI Research and Learning Base of Urban Culture (project 2023WZJD008), and the Guangzhou Municipal Science and Technology Project (Guangzhou Basic and Applied Basic Research Scheme, grant nos. 2024A04J6375 and 2025A03J3490; principal investigator: Zhuoni Zhang).

References

Alattar, M. A., Cottrill, C., & Beecroft, M. (2021). Public participation geographic information system (PPGIS) as a method for active travel data acquisition. *Journal of Transport Geography*, *96*, 103180.

Alfonzo, M. A. (2005). To walk or not to walk? The hierarchy of walking needs. *Environment and behaviour*, *37*(6), 808-836.

Altomonte, S., Allen, J., Bluyssen, P. M., Brager, G., Heschong, L., Loder, A., ... & Wargocki, P. (2020). Ten questions concerning well-being in the built environment. *Building and Environment*, *180*, 106949.

Amap. (2025). *Point of Interest (POI) data for urban spatial analysis* [Data set]. AutoNavi Software Co., Ltd. https://www.amap.com

Anderson, B. (2009). Emotion, space and society. *Affective Atmospheres*, *2*(2), 77-81.

Banach, M., Lewek, J., Surma, S., Penson, P. E., Sahebkar, A., Martin, S. S., ... & Bytyçi, I. (2023). The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis. *European journal of preventive cardiology*, *30*(18), 1975-1985.

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in human geography*, *3*(3), 274-279.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and urban planning*, *215*, 104217.

Chen, J., Tao, Z., Wu, W., Wang, L., & Chen, D. (2024). Influence of urban park pathway features on the density and intensity of walking and running activities: a case study of Shanghai city. *Land*, *13*(2), 156.

Chen, T., Hui, E. C. M., Lang, W., & Tao, L. (2016). People, recreational facility and physical activity: New-type urbanization planning for the healthy communities in China. *Habitat International*, *58*, 12-22.

Connaway, L. S., Dickey, T. J., & Radford, M. L. (2011). “If it is too inconvenient I'm not going after it:” Convenience as a critical factor in information-seeking behaviours. *Library & Information science research*, *33*(3), 179-190.

Coombes, E., Jones, A. P., & Hillsdon, M. (2010). The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Social science & medicine*, *70*(6), 816-822.

De Certeau, M. (1984). The Practice of Everyday Life. 1984. *Trans. Steven Rendall. Berkeley: U of California P*.

Duan, Y., Wagner, P., Zhang, R., Wulff, H., & Brehm, W. (2018). Physical activity areas in urban parks and their use by the elderly from two cities in China and Germany. *Landscape and Urban Planning*, *178*, 261-269.

Edensor, T. (Ed.). (2012). *Geographies of rhythm: nature, place, mobilities and bodies*. Ashgate Publishing, Ltd..

Ewing, R., & Handy, S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban design*, *14*(1), 65-84.

Fan, M., Lai, S., Huang, J., Wei, X., Chai, Z., Luo, J., & Wei, X. (2021). Rethinking bisenet for

real-time semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9716-9725).
Fang, Y. N., Namaiti, A., Zhang, S., & Feng, T. (2025). Multimodal Data-Driven Visual Sensitivity Assessment and Planning Response Strategies for Streetscapes in Historic Districts: A Case Study of Anshandao, Tianjin. *Land*, *14*(5), 1036.
Frank, L. D., Sallis, J. F., Saelens, B. E., Leary, L., Cain, K., Conway, T. L., & Hess, P. M. (2010). The development of a walkability index: application to the Neighborhood Quality of Life Study. *British journal of sports medicine*, *44*(13), 924-933.
Gavrilidis, A. A., Popa, A. M., Onose, D. A., & Gradinaru, S. R. (2022). Planning small for winning big: Small urban green space distribution patterns in an expanding city. *Urban Forestry & Urban Greening*, *78*, 127787.
Gehl, J. (2013). *Cities for people*. Island press.
Gong, F. Y., Zeng, Z. C., Zhang, F., Li, X., Ng, E., & Norford, L. K. (2018). Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Building and Environment*, *134*, 155-167.
Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211-221.
Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., & Turini, F. (2019). Factual and counterfactual explanations for black box decision making. *IEEE Intelligent Systems*, *34*(6), 14-23.
Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: views from urban planning. *American journal of preventive medicine*, *23*(2), 64-73.
Harvey, D. (2006). *Spaces of global capitalism*. Verso.
Ito, K., Quintana, M., Han, X., Zimmermann, R., & Biljecki, F. (2024). Translating street view imagery to correct perspectives to enhance bikeability and walkability studies. *International Journal of Geographical Information Science*, *38*(12), 2514-2544.
Kaczynski, A. T., & Henderson, K. A. (2008). Parks and recreation settings and active living: a review of associations with physical activity function and intensity. *Journal of physical activity and health*, *5*(4), 619-632.
Keep Inc. (2025). *Keep user activity data for exercise behaviour analysis* [Data set]. https://www.gotokeep.com
Kim, Y. O. (1999). *Spatial Configuration, Spatial Cognition and Spatial Behaviour: the role of architectural intelligibility in shaping spatial experience*. University of London, University College London (United Kingdom).
Kohl, H. W., Craig, C. L., Lambert, E. V., Inoue, S., Alkandari, J. R., Leetongin, G., & Kahlmeier, S. (2012). The pandemic of physical inactivity: global action for public health. *The lancet*, *380*(9838), 294-305.
Kostrzewska, M. (2017, October). Activating public space: how to promote physical activity in urban environment. In *IOP conference series: materials science and engineering* (Vol. 245, No. 5, p. 052074). IOP Publishing.
Kwan, M. P., & Schwanen, T. (2009a). Critical quantitative geographies. *Environment and Planning A*, *41*(2), 261-264.

Kwan, M. P., & Schwanen, T. (2009b). Quantitative revolution 2: The critical (re) turn. *The Professional Geographer*, *61*(3), 283-291.

Lebakula, V., Sims, K., Reith, A., Rose, A., McKee, J., Coleman, P., ... & Bright, E. (2025). LandScan Global 30 arcsecond annual global gridded population datasets from 2000 to 2022. *Scientific Data*, *12*(1), 495.

Lefebvre, H. (1991). *The Production of Space*. Oxford: Blackwell.

Li, Z., He, W., Cheng, M., Hu, J., Yang, G., & Zhang, H. (2023). SinoLC-1: The first 1-meter resolution national-scale land-cover map of China created with the deep learning framework and open-access data. *Earth System Science Data Discussions*, *2023*, 1-38.

Liu, Q., Dai, S., Yang, G., & Wu, C. (2026). Understanding multi-perspective urban green space patterns under urban expansion: Evidence from Guangzhou. *Sustainable Cities and Society, 147*, 107534.

Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.

Lynch, K. (1964). *The image of the city*. MIT press.

Maghelal, P. K., & Capp, C. J. (2011). Walkability: A Review of Existing Pedestrian Indices. *Journal of the Urban & Regional Information Systems Association*, *23*(2).

Massey, D. B. (2005). For space.

McLees, L. (2013). A postcolonial approach to urban studies: Interviews, mental maps, and photo voices on the urban farms of Dar es Salaam, Tanzania. *The Professional Geographer*, *65*(2), 283-295.

Mehta, V. (2008). Walkable streets: pedestrian behaviour, perceptions and attitudes. *Journal of urbanism*, *1*(3), 217-245.

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., ... & Bischl, B. (2020, July). General pitfalls of model-agnostic interpretation methods for machine learning models. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers* (pp. 39-68). Cham: Springer International Publishing.

Pate, R. R., Dowda, M., Saunders, R. P., Colabianchi, N., Clennin, M. N., Cordan, K. L., ... & Shirley, W. L. (2021). Operationalizing and testing the concept of a physical activity desert. *Journal of Physical Activity and Health*, *18*(5), 533-540.

Pierce, J., & Lawhon, M. (2015). Walking as method: Toward methodological forthrightness and comparability in urban geographical research. *The professional geographer*, *67*(4), 655-662.

Pontin, F. L., Jenneson, V. L., Morris, M. A., Clarke, G. P., & Lomax, N. M. (2022). Objectively measuring the association between the built environment and physical activity: a systematic review and reporting framework. *International Journal of Behavioural Nutrition and Physical Activity*, *19*(1), 119.

Sallis, J. F., Cerin, E., Kerr, J., Adams, M. A., Sugiyama, T., Christiansen, L. B., ... & Owen, N. (2020). Built environment, physical activity, and obesity: findings from the international physical activity and environment network (IPEN) adult study. *Annual review of public health*, *41*(1), 119-139.

Sallis, J. F., Cerin, E., Conway, T. L., Adams, M. A., Frank, L. D., Pratt, M., ... & Owen, N. (2016). Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. *The lancet*, *387*(10034), 2207-2217.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, *109*(3), 247-278.

Samuelsson, K., Colding, J., & Barthel, S. (2019). Urban resilience at eye level: Spatial analysis of empirically defined experiential landscapes. *Landscape and Urban Planning*, *187*, 70-80.

Schipperijn, J., Cerin, E., Adams, M. A., Reis, R., Smith, G., Cain, K., ... & Sallis, J. F. (2017). Access to parks and physical activity: An eight country comparison. *Urban forestry & urban greening*, *27*, 253-263.

Simonsen, K. (2005). Bodies, sensations, space and time: The contribution from Henri Lefebvre. *Geografiska Annaler: Series B, Human Geography*, *87*(1), 1-14.

Sina Weibo. (2025). *Check-in data from the Chinese social media platform Weibo* [Data set]. Sina Corporation. https://weibo.com

Soja, E. (2010). Spatializing the urban, Part I. *City*, *14*(6), 629-635.

Stavroulaki, I., Ivarsson, O., Pont, M. B., & Verendel, V. (2024). From explanations to predictions. Developing a predictive model of pedestrian flows on existing and planned streets. In *Proceedings of the 14th International Space Syntax Symposium, SSS 2024* (pp. 1705-1734).

Talen, E. (2003). Neighborhoods as service providers: a methodology for evaluating pedestrian access. *Environment and planning B: Planning and Design*, *30*(2), 181-200.

Thompson, C. W. (2013). Activity, exercise and the planning and design of outdoor spaces. Journal of Environmental Psychology, 34, 79-96.

Venerandi, A., Mellen, H., Romice, O., & Porta, S. (2024). Walkability Indices—The state of the Art and future directions: A systematic review. *Sustainability*, *16*(16), 6730.

Wacquant, L. (2022). The Invention of the'Underclass': A Study in the Politics of Knowledge. John Wiley & Sons.

Watkins, C. (2005). Representations of space, spatial practices and spaces of representation: An application of Lefebvre's spatial triad. *Culture and organization*, *11*(3), 209-220.

Westenhoefer, J., Nouri, E., Reschke, M. L., Seebach, F., & Buchcik, J. (2023). Walkability and urban built environments—a systematic review of health impact assessments (HIA). *BMC public health*, *23*(1), 518.

World Health Organization. (2018). *Global action plan on physical activity 2018-2030: more active people for a healthier world*. World Health Organization.

Wu, C., Zhao, M., & Ye, Y. (2023). Measuring urban nighttime vitality and its relationship with urban spatial structure: A data-driven approach. *Environment and Planning B: Urban Analytics and City Science*, *50*(1), 130-145.

Wu, C., Wang, J., Wang, M., & Kraak, M. J. (2024). Machine learning-based characterisation of urban morphology with the street pattern. *Computers, Environment and Urban Systems*, *109*, 102078.

Xie, H., Wang, X., Wang, Z., Shi, Z., Hu, X., Lin, H., ... & Liu, X. (2023). Mismatch between infrastructure supply and demand within a 15-minute living circle evaluation in Fuzhou, China. *Heliyon*, *9*(9).

Yang, G., Su, Y., Zhao, M., Li, C., & Wu, C. (2026). Beyond single snapshots: Quantifying multi-scale heterogeneity from street-view imagery and what it reveals about perception. *Cities, 175*, 107227.

You, H. (2016). Characterizing the inequalities in urban public green space provision in Shenzhen, China. *Habitat international*, *56*, 176-180.

Yu, J., Zhang, H., Dong, X., & Shen, J. (2024). The impact of street greenery on active travel: A narrative systematic review. *Frontiers in public health*, *12*, 1337804.

Yue, Y., Zhuang, Y., Yeh, A. G., Xie, J. Y., Ma, C. L., & Li, Q. Q. (2017). Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*, *31*(4), 658-675.

Zhang, R., Cao, L., Wang, L., Wang, L., Wang, J., Xu, N., & Luo, J. (2024). Assessing the relationship between urban park spatial features and physical activity levels in Residents: A spatial analysis Utilizing drone remote sensing. *Ecological Indicators*, *166*, 112520.

Zhang, L., Ye, Y., Zeng, W., & Chiaradia, A. (2019). A systematic measurement of street quality through multi-sourced urban data: A human-oriented analysis. *International journal of environmental research and public health*, *16*(10), 1782.

Zhang, S., Ran, L., Fan, X., Zhang, Y., & Guo, H. (2024). Perceived built environment as a mediator linking objective built environment and leisure-time physical activity in Chinese cities. *Scientific Reports*, *14*(1), 17091.

Zhao, M., Šćepanović, S., Law, S., Obadić, I., Wu, C., & Quercia, D. (2026). UST-GNN: A unified spatial–topological graph neural network framework for urban analytics demonstrated through a case study on urban health prediction. *Computers, Environment and Urban Systems*, *129*, 102466.

Zhong, T., Lü, G., Zhong, X., Tang, H., & Ye, Y. (2020). Measuring human-scale living convenience through multi-sourced urban data and a geodesign approach: Buildings as analytical units. *Sustainability*, *12*(11), 4712.

Zhou, H., Wang, J., Widener, M., & Wilson, K. (2024). Examining the relationship between active transport and exposure to streetscape diversity during travel: A study using GPS data and street view imagery. *Computers, Environment and Urban Systems*, *110*, 102105.

Appendix A: Prompt Design for LLM-assisted Urban Perception Scoring

To analyze street-scale social media texts in terms of urban perception and experiential context, we developed an LLM-assisted semantic scoring approach using a structured prompt. The core prompt used in the analysis is presented below.

System Prompt:

You are an expert in urban space perception analysis. Please evaluate each text along the following eight dimensions, using a score from 0 to 1 with two decimal places, and return the results in standard JSON format.

User Prompt Template:

Please score the following microblog texts across multiple affective and contextual dimensions, using a scale from 0 to 1 with two decimal places. The dimensions are defined as follows:
1. Urban: the extent to which the text involves evaluation or description of urban space, streets, landscape, environment, urban renewal, or related urban elements.
2. positive: the extent to which the text expresses positive or energetic emotions (e.g., happiness, satisfaction, vitality).
3. relax: the extent to which the text conveys relaxation, calmness, or comfort (e.g., slow-paced life, quietness, ease).
4. explore: the extent to which the text expresses curiosity, interest, or a desire to explore or visit.
5. social: the extent to which the text reflects social interaction, gathering, companionship, or collective activity.
6. sport: the extent to which the text is related to sport or physical activity.
7. safe_walk: the extent to which the text refers to safety or walkability, such as road conditions, lighting, accessibility, or walking experience. Use 0.50 if this aspect is not mentioned.
8. context: the extent to which the text reflects the immediate geographical context or present-moment experience, that is, the consistency between the description and the real location.
The microblog texts are:
{text_block}

Please return the results in the following format only, without any explanation. Each text should correspond to one JSON object:

```
{
   "1": {"urban": 0.85, "positive": 0.76, "relax": 0.42, "explore": 0.67, "social": 0.31, "sport": 0.05, "safe_walk": 0.50, "context": 0.92},
   "2": {...},
   ...
```

"10": {...}
}"""

The semantic scoring used in the analysis was generated with GPT-4o. Each cleaned Weibo post was evaluated under the structured JSON prompt above across eight dimensions: urban, positive, relax, explore, social, sport, safe_walk, and context. The street-level lived-space model used four aggregated semantic indicators from this output: urban, positive, sport, and context.

Validation and reproducibility checks for the LLM-assisted indicators

We conducted three complementary checks of the LLM-derived semantic indicators. First, a preliminary human–machine consistency check focused on the positive indicator. Two trained human coders with urban-planning backgrounds independently evaluated the positive affect expressed in 50 sampled Weibo posts using a five-level ordinal scale. For comparison, the continuous GPT-4o positive scores were converted to the same five-level scale using equal-width intervals: Level 1 for $0.00 \le x < 0.20$, Level 2 for $0.20 \le x < 0.40$, Level 3 for $0.40 \le x < 0.60$, Level 4 for $0.60 \le x < 0.80$, and Level 5 for $0.80 \le x \le 1.00$. Human–human agreement reached an accuracy of 0.88 and a Cohen's κ of 0.81. Agreement between the two coders and the GPT-4o ratings was similarly high, with accuracies of 0.85 and 0.86 and κ values of 0.77 and 0.78, respectively (Table A.1). Minor disagreements occurred mainly for posts assigned to Level 3, where the affective tone was neutral, ambiguous, or mixed.

Second, we conducted a cross-model consistency check for the core urban and positive indicators. GPT-4o and Kimi independently evaluated 200 sampled posts using the same Chinese-language scoring prompt and JSON output structure. A total of 195 posts produced valid outputs from both models. The GPT-4o–Kimi comparison yielded a Pearson correlation of 0.859, a Spearman correlation of 0.776, and a mean absolute error (MAE) of 0.089 (Table A.2).

Third, three trained human coders independently evaluated 300 sampled posts for the urban and positive dimensions. Their scores were compared with the original GPT-4o outputs and with one another. Across the three coder–GPT-4o comparisons, the mean Pearson correlation was 0.707, the mean Spearman correlation was 0.601, and the mean MAE was 0.169. Inter-coder consistency was generally higher, with Pearson correlations ranging from 0.764 to 0.887 and Spearman correlations ranging from 0.672 to 0.852 (Table A.3).

The main reliability analyses focus on urban and positive because these dimensions provide relatively direct and auditable indications of place-related expression and affective tone. The sport score is retained as a task-specific lived-space indicator, but explicit sport-related language is sparse in general Weibo posts and is therefore interpreted as auxiliary evidence. Taken together, the three checks indicate moderate consistency between the LLM-derived and human-coded urban and positive indicators, while also revealing some coder-specific variation. These results support their use as approximate, platform-mediated proxies in the lived-space analysis rather than as definitive measures of urban experience.

Table A.1 Preliminary 50-post human–machine consistency check for the positive-affect indicator.

Comparison	N	Metric	Result
Coder A vs Coder B	50	Acc / kappa	0.88 / 0.81

Coder A vs GPT-4o	50	Acc / kappa	0.85 / 0.77
Coder B vs GPT-4o	50	Acc / kappa	0.86 / 0.78

Table A.2 Cross-model consistency between GPT-4o and Kimi for the core urban and positive indicators.

Validation comparison	N	Pearson	Spearman	MAE
GPT-4o vs Kimi	195	0.859	0.776	0.089

Table A.3 Three-human-coder validation of the original GPT-4o urban and positive scores.

Validation comparison	N	Pearson	Spearman	MAE
Coder A vs original GPT-4o	300	0.725	0.578	0.134
Coder B vs original GPT-4o	300	0.668	0.602	0.236
Coder C vs original GPT-4o	300	0.728	0.624	0.138
Coder A vs Coder B	300	0.764	0.672	0.182
Coder A vs Coder C	300	0.859	0.749	0.092
Coder B vs Coder C	300	0.887	0.852	0.136

Appendix B: Citywide SHAP Analysis - Variable Dependency Diagram

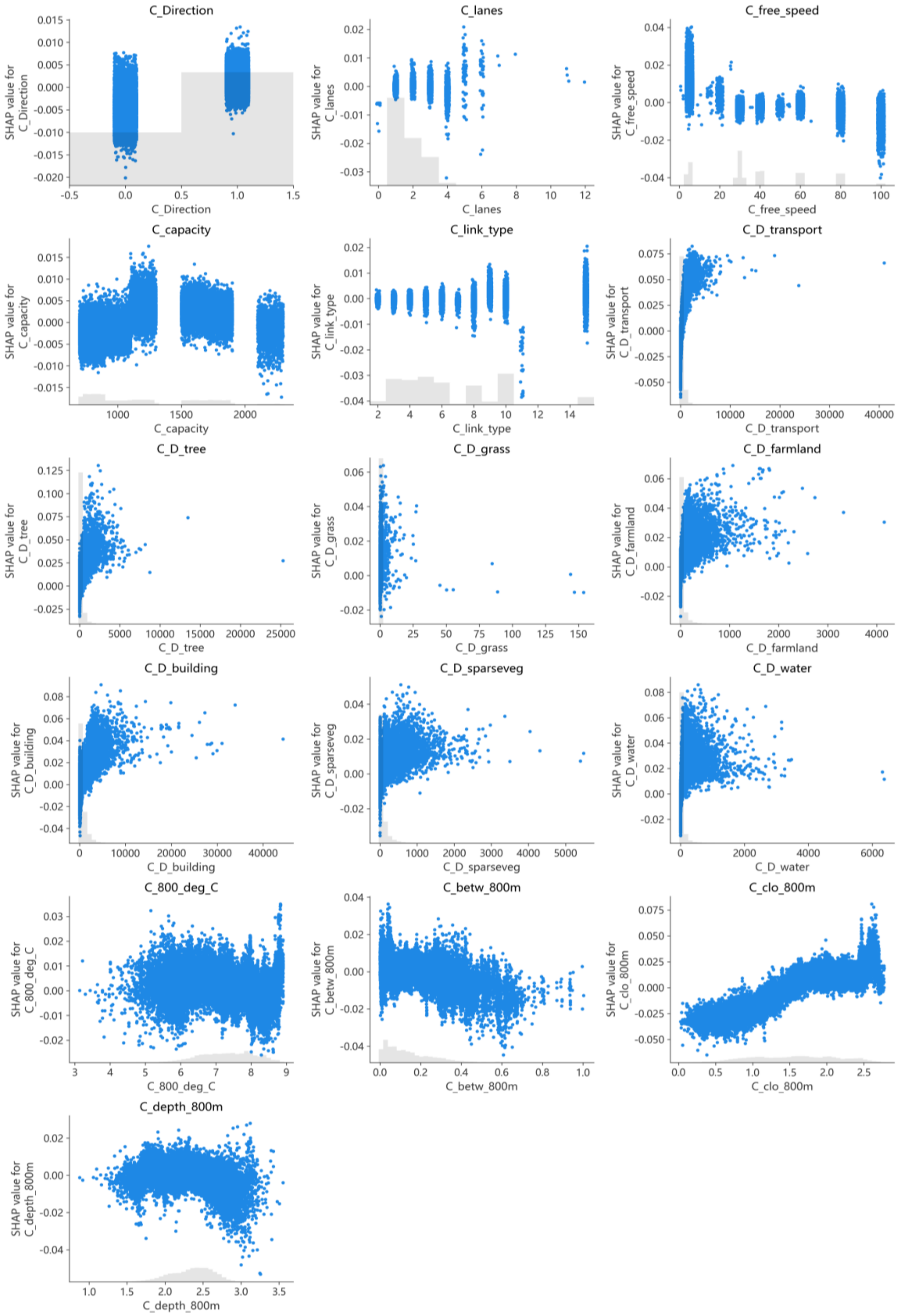


Figure B.1 SHAP dependency graph of C-variables

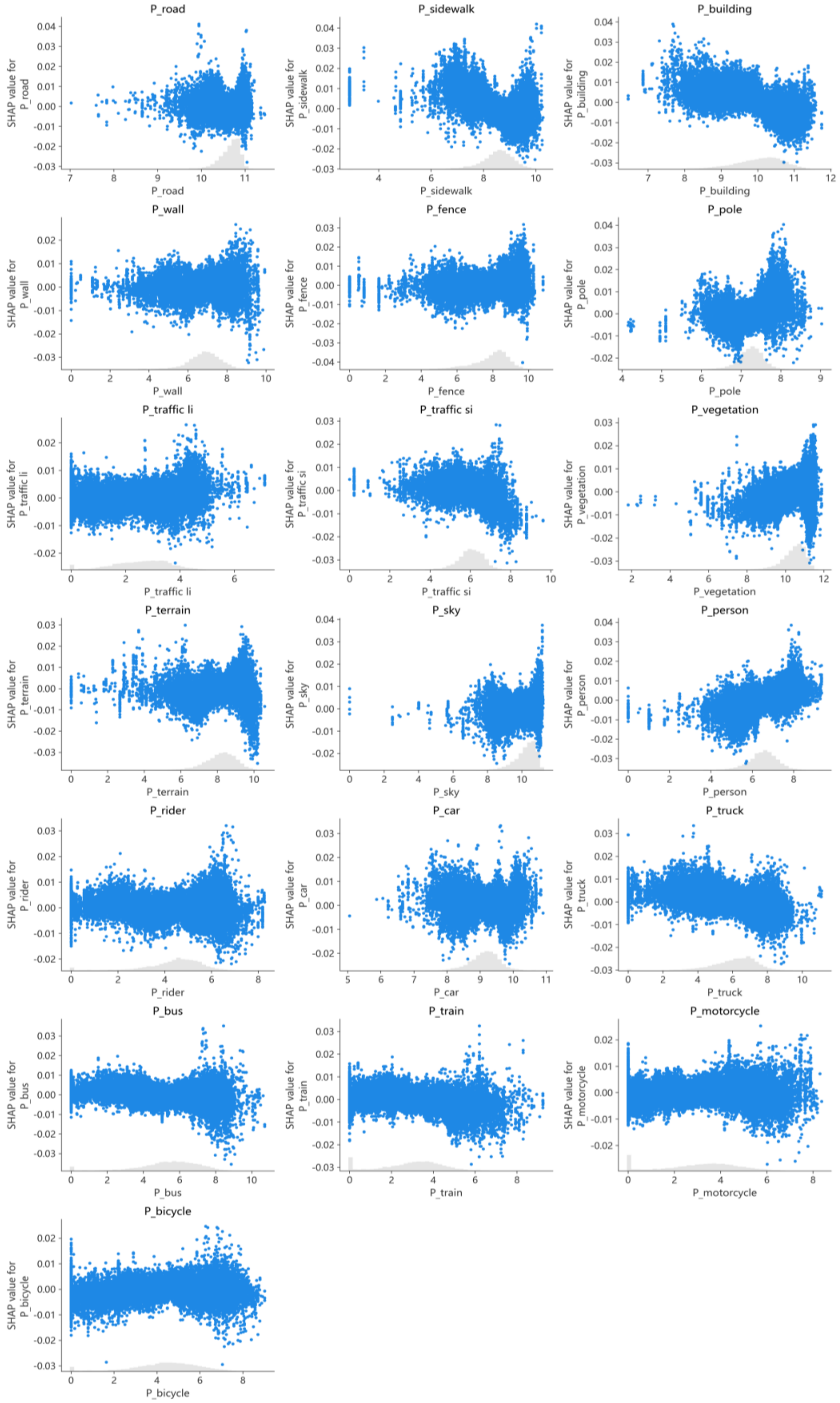


Figure B.2 SHAP dependency graph of P-variables

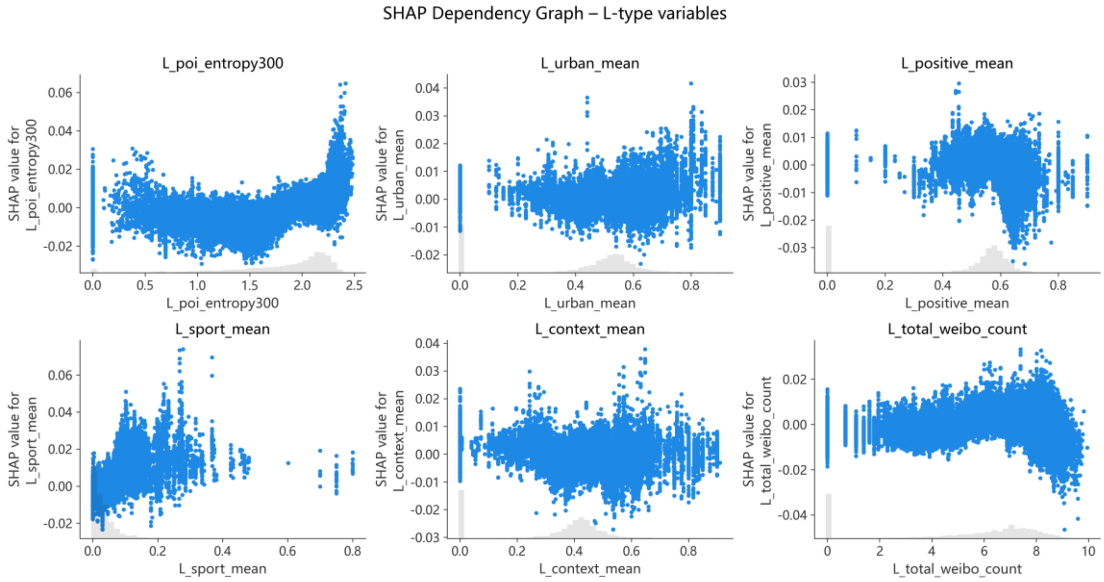


Figure B.3 SHAP dependency graph of L-variables

Appendix C: Hyperparameter Search Space and Final Configuration

To ensure robust model performance and avoid overfitting, we conducted a randomized search across a carefully selected parameter space for each regression algorithm. The search employed 10-fold cross-validation and was performed separately for each model using the training set.

Table C.1 Random Search Parameter Space

Model	Parameter	Search Range
XGBoost	max_depth	[3, 4, 5, 6, 8]
	learning_rate	[0.01, 0.05, 0.1, 0.2]
	subsample	[0.6, 0.8, 1.0]
	colsample_bytree	[0.6, 0.8, 1.0]
	gamma	[0, 0.1, 0.5, 1]
RandomForest	max_depth	[4, 6, 8]
	n_estimators	[100, 150, 200, 300]
	max_features	['auto', 'sqrt', 'log2']
LightGBM	max_depth	[4, 6, 8]
	learning_rate	[0.01, 0.05, 0.1]
	num_leaves	[31, 63, 127]
	feature_fraction	[0.6, 0.8, 1.0]

Table C.2 Optimal Hyperparameters for XGBoost Model

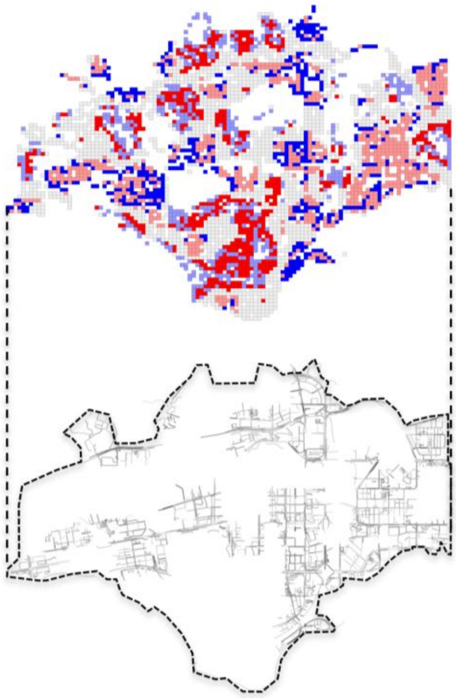
Parameter	Value	Description
objective	'reg:squarederror'	Loss function for regression tasks
n_estimators	200	Number of boosting rounds
max_depth	8	Maximum depth of each decision tree
learning_rate	0.05	Step size shrinkage
subsample	0.8	Fraction of samples used per tree
colsample_bytree	0.8	Fraction of features used per tree
gamma	0.1	Minimum loss reduction to make a split
random_state	42	Seed for reproducibility

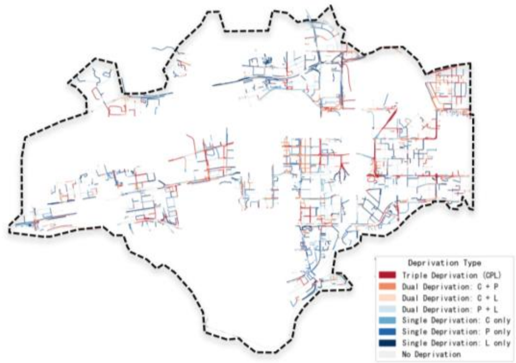
Appendix D: SHAP-Informed Intervention Simulation

Table D.1 Summary of intervention simulations across Futian, Bao'an, and Guangming, including intervention type, variable count, and model-predicted changes.

	Type of intervention	Intensity of intervention	Number of variables	Predicted change
Futian	*C*	20%	5	3.71%
	P	20%	5	2.54%
	L	20%	5	3.21%
	C+P	20%	5	6.32%
	C+L	20%	5	6.80%
	L+P	20%	5	3.75%
	C+P+L	20%	5	7.00%
	C+P+L	20%	10	8.05%
	C+P+L	30%	10	8.17%
	C+P+L	30%	15	10.02%
Top 10 variables: 'L_total_weibo_count', 'P_sky', 'C_free_speed', 'L_positive_mean', 'P_building', 'C_depth_800m', 'P_fence', 'C_D_tree', 'P_car', 'L_urban_mean'				
Bao'an	*C*	20%	5	1.85%
	P	20%	5	5.67%
	L	20%	5	0.77%
	C+P	20%	5	7.25%
	C+L	20%	5	2.65%
	L+P	20%	5	5.67%
	C+P+L	20%	5	7.25%
	C+P+L	20%	10	9.38%
	C+P+L	30%	10	9.38%
	C+P+L	30%	15	11.53%
Top 10 variables: 'P_car', 'P_truck', 'C_free_speed', 'P_wall', 'P_road', 'P_sidewalk', 'P_building', 'C_D_sparseveg', 'L_poi_entropy300', 'P_bus'				
Guangming	*C*	20%	5	1.99%
	P	20%	5	7.04%
	L	20%	5	0.18%
	C+P	20%	5	6.79%
	C+L	20%	5	1.66%
	L+P	20%	5	7.04%
	C+P+L	20%	5	6.79%
	C+P+L	20%	10	8.24%
	C+P+L	30%	10	8.31%
	C+P+L	30%	15	7.93%
Top 10 variables: 'P_car', 'P_building', 'C_800_deg_C', 'P_traffic si', 'P_terrain', 'C_D_tree', 'P_fence', 'P_truck', 'P_road', 'C_clo_800m'				

These perturbation scenarios in Table D.1 are model-based sensitivity exercises rather than causal estimates of real-world intervention effects.





	Type of intervention	Intensity of intervention	Number if variables	Exercise improves
Futian	C	20%	5	3.71%
	P	20%	5	2.54%
	L	20%	5	3.21%
	C+P	20%	5	6.32%
	C+L	20%	5	6.80%
	L+P	20%	5	3.75%
	C+P+L	20%	5	7.00%
	C+P+L	20%	10	8.05%
	C+P+L	30%	10	8.17%
	C+P+L	30%	15	10.02%

Figure D.1 SHAP-informed intervention simulation workflow for High-Population/Low-Exercise-Intensity areas in Futian District. The left panel shows grid-level clusters where high population density coincides with low recorded exercise intensity; the right panel maps the seven diagnostic modes for the same area, overlaid with the simulated intervention enhancements. The accompanying table reports model-predicted changes in exercise supportiveness under different intervention types (C, P, L, and their combinations) and intensities.

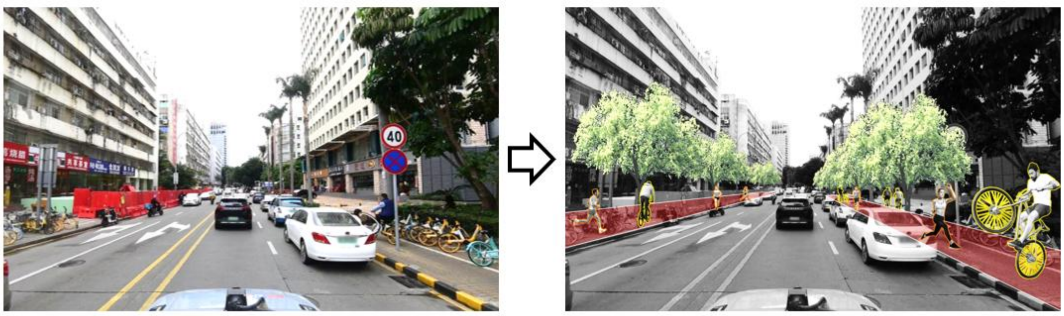


Figure D.2 Street-level intervention scenarios illustrating how conceived (C), perceived (P), and lived (L) spaces can be improved. The left image shows the current condition, while the right visualizes hypothetical enhancements

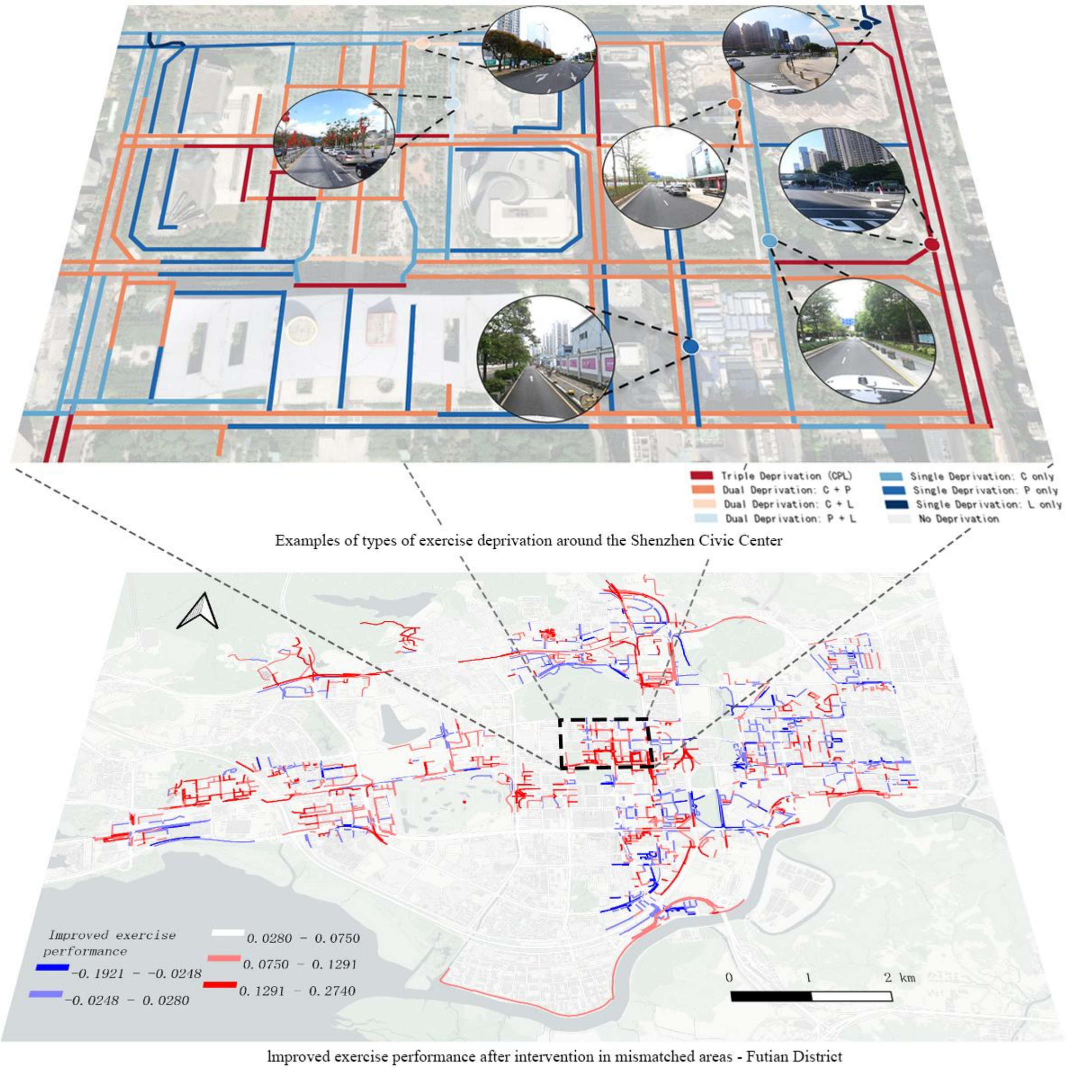


Figure D.3 Examples of diagnostic supportiveness-deficit modes around the Shenzhen Civic Center (top) and model-predicted changes in exercise supportiveness under hypothetical perturbation scenarios (bottom) in Futian District.

We simulated SHAP-informed interventions by perturbing key explanatory variables in the top-ranked SHAP dimensions and recomputing predicted exercise scores within mismatch areas. Formally, for a given segment s, the adjusted prediction under a hypothetical intervention δ on feature x_k is given by:

$$\hat{f}_s' = f(x_1, \ldots, x_k + \delta, \ldots, x_n)$$

Aggregated intervention effects were summarized by typology and region, resulting in a structured intervention enhancement table that specifies (1) which spatial domain (C/P/L) to target, (2) which specific features to improve (e.g., greening, sidewalk continuity, POI richness), and (3) the model-predicted change in movement supportiveness under the hypothetical perturbation.

Appendix E: Accessible-Street Robustness Check

To examine whether the results depend on automobile-oriented road links, we repeated the full modelling and attribution workflow under two restricted street samples. The first sample retained walking- or cycling-permitted segments while excluding motorway and trunk links. The second used a stricter active-street subset, including footways, cycleways, residential streets, living streets, service roads, tertiary roads, unclassified roads, and tracks. For each sample, we re-estimated the XGBoost model, recalculated grouped SHAP contributions, and reconstructed the seven-mode diagnostic typology using the same classification rule as in the main analysis.

The results are stable across the two filters. Excluding motorway and trunk links slightly improves model performance, with R^2 increasing from 0.695 to 0.699, while the grouped C/P/L contribution shares remain almost unchanged. Under the stricter active-street subset, performance decreases moderately, as expected given the smaller and more homogeneous sample, but the C/P/L attribution structure remains highly similar. The seven-mode typology shares also vary only marginally across samples. These results indicate that the diagnostic patterns reported in the main analysis are not driven by high-speed motorized links.

Table E.1 Accessible-street robustness test for model performance and grouped SHAP contribution.

Sample	N	R^2	RMSE	MAE	C share	P share	L share
Main sample	85,903	0.695	0.0458	0.0328	57.0%	29.6%	13.4%
Accessible streets, excluding motorway/trunk	84,032	0.699	0.0454	0.0324	56.9%	29.8%	13.3%
Strict active-street subset	57,923	0.667	0.0498	0.0359	57.0%	29.3%	13.7%

Note. C, P, and L denote the conceived-, perceived-, and lived-space indicator groups, respectively. Grouped SHAP shares are normalized within each sample.

Table E.2 Accessible-street robustness test for seven-mode deprivation typology shares.

Sample	Not deprived	C-only	P-only	L-only	CP	CL	PL	CPL
Main sample	58.3%	10.1%	7.4%	8.8%	4.1%	2.7%	5.4%	3.1%
Accessible streets, excluding motorway/trunk	58.7%	9.9%	7.5%	8.5%	3.9%	2.9%	5.4%	3.2%
Strict active-street subset	58.5%	9.7%	7.5%	9.0%	4.2%	2.8%	5.0%	3.3%

Note. Typology shares indicate the percentage of street segments assigned to each diagnostic mode within each sample. “Not deprived” denotes segments for which no C/P/L grouped SHAP contribution falls into the low-supportiveness tail.

Appendix F: Internal Consistency of SHAP-Based Diagnostic Modes

This appendix describes the internal ordering of the SHAP-based diagnostic modes using the Keep-derived exercise-intensity outcome fitted by the model. Mean and median intensity decline as more C/P/L readings are flagged, with the lowest values for CPL. Because the same outcome underlies the SHAP contributions and the comparison, the pattern indicates internal consistency; validation against an independent behavioural or field-based outcome remains necessary.

Table F.1 Observed exercise intensity by number of SHAP-deprived dimensions.

Number of deprived dimensions	N	Mean	Median	Mean difference vs. 0 dims
0	50,122	0.704	0.699	0.000
1	22,640	0.623	0.619	-0.081
2	10,520	0.572	0.570	-0.132
3	2,621	0.506	0.508	-0.198

Table F.2 Observed exercise intensity by SHAP-based deprivation mode.

Deprivation mode	N	Mean	Median	Mean difference vs. not deprived
Not deprived	50,122	0.704	0.699	0.000
C-only	8,652	0.600	0.600	-0.104
P-only	6,391	0.624	0.621	-0.080
L-only	7,597	0.649	0.645	-0.055
CP	3,557	0.539	0.544	-0.164
CL	2,351	0.564	0.565	-0.140
PL	4,612	0.600	0.598	-0.104
CPL	2,621	0.506	0.508	-0.198

Appendix G: Population-Density and GDP Control Robustness Check

This appendix reports a contextual robustness check that adds population density and gridded GDP to the XGBoost model. The test evaluates whether broad exposure and socioeconomic context absorb the C/P/L attribution structure.

Table G.1 Population-density and GDP control robustness check.

Model	N	R^2	RMSE	MAE	C share	P share	L share	Control share
Baseline	85,903	0.695	0.0458	0.0328	57.0%	29.6%	13.4%	0.0%
+ Population	85,903	0.699	0.0455	0.0326	58.0%	28.8%	13.2%	2.3%
+ Population + GDP	85,903	0.712	0.0445	0.0319	58.1%	29.7%	12.2%	6.4%

Table G.2 Seven-mode typology shares after adding population-density and GDP controls.

Model	Not dep.	C-only	P-only	L-only	CP	CL	PL	CPL
Baseline	58.3%	10.1%	7.4%	8.8%	4.1%	2.7%	5.4%	3.1%
+ Population	58.6%	9.9%	7.6%	8.5%	3.9%	3.0%	5.3%	3.1%
+ Population + GDP	58.6%	9.9%	7.3%	8.7%	4.2%	2.8%	5.4%	3.2%

Appendix H: Temporal Consistency Checks for Keep and Weibo Layers

This appendix reports two temporal consistency checks used to evaluate whether the broad activity and lived-space patterns are dominated by a single annual file. These checks do not remove temporal mismatch, but they help bound its likely influence on the diagnostic interpretation.

Table H.1 Pearl River Delta-wide Keep 2019/2020 grid-pattern consistency.

Subset	N	Pearson	Spearman	Top-decile Jaccard	Active grid cells
All active grid-cell union	13,671	0.428	0.448	0.551	6,924 / 12,497
Grid cells active in both years	5,750	0.675	0.676	--	--

Table H.2 Annual Weibo indicator consistency with pooled 2021–2025 lived-space indicators.

Year	Posts	Active segments	Count Spearman	Mean score Spearman
2021	190,101	74,831	0.899	0.633
2022	100,018	72,178	0.885	0.575
2023	113,478	70,075	0.884	0.580
2024	246,165	70,795	0.903	0.594
2025	56,501	52,774	0.822	0.390

The annual-window analysis indicates that Weibo-derived count patterns remain broadly consistent across the available yearly files. This supports using the pooled 2021-2025 Weibo layer to identify persistent street-level social-expression signals, while interpreting these indicators as platform-mediated activity traces rather than population-representative survey measures.

The full corpus used for semantic scoring contained over 1.5 million Weibo posts. The annual counts reported in Table H.2 refer only to posts retained after street-segment matching and the removal of invalid or advertising-related posts.

Appendix I: COVID-Period Context and Keep Year-Split Robustness

The Keep trajectories used in the study cover 2019-2020 and therefore include both pre-pandemic activity and activity recorded during COVID-period management. Because the trajectories include records from 2020, they may reflect temporary changes in outdoor activity, route choice, and recording behaviour associated with the COVID-19 period.

Table I.1 Year-split robustness check for Keep activity patterns in the PRD/GBA dataset.

Comparison	2019 trajectories	2020 trajectories	Common active cells	Pearson	Spearman
2019 vs. 2020, 1 km grid	13,840	44,967	6,924	0.675	0.676

Top-decile hotspot Jaccard overlap = 0.551. The moderate year-to-year correlations suggest that the broad spatial structure of recorded Keep activity was not fully driven by a single pandemic-year pattern, although the 2020 records should still be interpreted with caution.

Appendix J: Multi-scale Spatial Validation

Random cross-validation may overstate predictive performance when nearby street segments share similar network, streetscape, and activity conditions. We therefore added spatially grouped validation while keeping street segments as the prediction and interpretation unit. Grid cells and administrative boundaries are used only to define held-out folds. The results show the expected decline in predictive performance as hold-out groups become spatially broader, indicating that extrapolation across larger urban areas is harder than random validation suggests.

Table J.1 Multi-scale street-segment spatial validation results.

Validation strategy	Model setting	Folds	R^2	R^2 SD	RMSE	MAE
Random 5-fold CV	Main specification	5	0.690	0.005	0.0459	0.0328
Grid-grouped 100 m 5-fold CV	Tuned	5	0.610	0.014	0.0514	0.0383
Grid-grouped 200 m 5-fold CV	Tuned	5	0.542	0.010	0.0557	0.0419
Grid-grouped 500 m 5-fold CV	Tuned	5	0.461	0.016	0.0604	0.0461
Subdistrict-grouped 5-fold CV	Tuned	5	0.330	0.031	0.0671	0.0522
Alternative street-office boundary CV	Tuned	5	0.327	0.007	0.0673	0.0521

Appendix K: Threshold Sensitivity of the SHAP-based Diagnostic Typology

This appendix tests whether the seven-mode diagnostic typology is sensitive to the selected low-supportiveness threshold. The main analysis uses a citywide 20% cutoff to preserve comparability across Shenzhen's street-network distribution. We compare this baseline with citywide 15%, 25%, and 30% cutoffs and a district-specific 20% threshold. The results show that nearby citywide cutoffs produce high agreement with the baseline, while observed Keep intensity continues to decline as more C/P/L dimensions are flagged.

Table K.1 Sensitivity of the SHAP-based diagnostic typology to alternative thresholds.

Threshold	Any deprived	Exact mode agree.	Binary agree.	0 dim mean	1 dim mean	2 dim mean	3 dim mean
Citywide 15%	33.2%	86.1%	91.5%	0.695	0.610	0.557	0.490
Citywide 20%	41.7%	100.0%	100.0%	0.704	0.623	0.572	0.506
Citywide 25%	49.3%	85.9%	92.3%	0.712	0.635	0.584	0.519
Citywide 30%	56.4%	73.4%	85.3%	0.721	0.645	0.594	0.531
District-specific 20%	41.2%	87.5%	92.8%	0.703	0.624	0.573	0.510

Appendix L: Keep Trajectory Data Volume and Street-Matching Procedure

This appendix reports the raw Keep trajectory volume and the street-matching steps used to construct the dependent variable. It also reports segment coverage under the 10 m, 20 m, and 30 m matching bands.

Table L.1 Raw volume and processing stages of the Keep trajectory data.

Processing stage	Count	Role in dependent-variable construction
Raw Keep trajectories, PRD/GBA source	58,807	Original trajectory lines before Shenzhen clipping
Keep trajectories clipped to Shenzhen	18,209	Trajectory lines retained within the Shenzhen study area
10 m sampled Keep trajectory points	11,585,134	Point samples generated from Shenzhen Keep trajectories for road matching
Street segments in integrated model input	115,365	Road/street segments after integrating C/P/L indicators

Segments with non-zero 30 m Keep intensity	87,467	Segments covered by the main 30 m dependent-variable band
Segments allowed in final analytic sample	112,643	Segments passing the street-use eligibility filter
Final modelling segments	85,903	Segments with non-zero 30 m Keep intensity that also passed the street-use eligibility filter

Table L.2 Road-segment coverage under alternative Keep trajectory matching bands.

Distance band	Covered segments	Share of all segments	Covered after eligibility filter	Share after eligibility filter
10 m	71,811	62.2%	70,547	62.6%
20 m	82,234	71.3%	80,782	71.7%
30 m	87,467	75.8%	85,903	76.3%

Appendix M: Multicollinearity Diagnostics and Correlation-pruned Sensitivity

This appendix reports predictor-redundancy diagnostics used before interpreting SHAP attributions. We assessed the 41 predictors used in the main model using pairwise Pearson correlations, pairwise Spearman correlations, and variance inflation factors (VIFs). Because the model is used primarily for prediction and grouped diagnostic attribution, these diagnostics are interpreted as transparency and robustness checks rather than as a requirement that all predictors be mutually independent.

The strongest pairwise correlations are internal to the same C/P/L reading rather than cross-dimensional. This suggests that redundancy mainly reflects related indicators within road-network, streetscape, or platform-mediated lived-space families, not strong confusion among the three analytical readings.

Table M.1 Pairwise feature-correlation diagnostic summary.

Method and threshold	Strong pairs	Within-group pairs	Cross-group pairs
Pearson $\|r\| >= 0.80$	12	12	0
Pearson $\|r\| >= 0.90$	4	4	0
Spearman $\|rho\| >= 0.80$	9	9	0
Spearman $\|rho\| >= 0.90$	4	4	0

Table M.2 Variance inflation factor diagnostic summary.

N features	Median VIF	90th pct. VIF	Max finite VIF	VIF > 5	VIF > 10
41	1.88	13.71	33.96	7	6

The VIF results indicate moderate-to-high redundancy for a limited subset of predictors. High VIFs are expected for some road-hierarchy, network, and semantic indicators because they describe related aspects of the same urban environment.

Table M.3 Correlation-pruned model sensitivity check.

Model	Features	R2	RMSE	MAE	C share	P share	L share	Mode agreement
Full predictor set	41	0.695	0.0458	0.0328	57.0%	29.6%	13.4%	100.0%
Correlation-pruned predictors	38	0.695	0.0458	0.0327	54.4%	31.6%	13.9%	75.8%

For predictor pairs with Spearman |rho| >= 0.90, we retained the feature with stronger baseline XGBoost importance and removed the redundant counterpart. The pruned model reduced the feature set from 41 to 38 predictors, kept model performance essentially unchanged, and preserved the same C > P > L grouped-attribution ordering. These diagnostics support group-level robustness, while individual-variable SHAP rankings should still be interpreted cautiously under correlated predictors.

Appendix N: STDC Segmentation Model Transparency and Local Output Audit

This appendix provides model-level transparency for the STDC semantic-segmentation layer used to construct perceived-space streetscape indicators. The benchmark values are taken from the original STDC reporting, while the local audit summarizes the aggregate class composition of the Shenzhen street-view outputs. These checks support use of the layer as an aggregated streetscape proxy, but they do not replace a Shenzhen-specific pixel-level validation dataset.

Table N.1 Reported STDC benchmark performance used for model transparency.

Model / setting	Benchmark	mIoU	Speed / hardware
STDC1-Seg50	Cityscapes test	71.9%	250.4 FPS / GTX 1080Ti
STDC2-Seg75	Cityscapes test	76.8%	97.0 FPS / GTX 1080Ti
STDC1-Seg	720 x 960 speed-accuracy comparison	73.0%	197.6 FPS
STDC2-Seg	720 x 960 speed-	73.9%	152.2 FPS

	accuracy comparison

Table N.2 Local Shenzhen street-view segmentation output plausibility audit.

Class / audit item	Mean share or value
Street-view sampling records	198,788
Expected image-equivalent pixels per record	approximately 172,800
Road	24.43%
Vegetation	21.44%
Sky	18.14%
Building	17.40%
Car	6.05%
Sidewalk	3.69%

Appendix O: 200 m Grid Rationale and Grid-size Sensitivity Check

This appendix reports the rationale and sensitivity check for the 200 m grid used in the final bivariate analysis of population density and recorded Keep-derived exercise intensity. The predictive model and SHAP typology remain street-segment based; the grid is introduced to compare gridded population density with length-weighted recorded exercise intensity. The 200 m baseline provides a middle-scale compromise: it is finer than coarse neighborhood units, coarser than the native 90 m population raster to reduce sparse-cell instability, and large enough to contain sufficient street length for aggregation.

Table O.1 Grid-size sensitivity for the bivariate population-density and recorded-exercise-intensity analysis. HP/LEI denotes High-Population/Low-Exercise-Intensity.

Grid	Valid cells	HP/LEI cells	HP/LEI share	Score rho vs 200 m	District score rho vs 200 m
100 m	42,111	1,606	3.81%	0.763	0.950
200 m	17,000	676	3.98%	1.000	1.000
300 m	9,234	369	4.00%	0.708	1.000
500 m	4,040	125	3.09%	0.621	0.983

The High-Population/Low-Exercise-Intensity share remains similar across 100 m, 200 m, and 300 m grids and declines mainly at the coarser 500 m scale. Continuous mismatch scores remain positively associated with the 200 m baseline, and district-level rankings are stable across grid sizes. The grid map is therefore interpreted as a planning screening layer.