Examining the Impact of Urban Environment on Healthy Vitality of Outdoor Running based on Street View Imagery and Urban Big Data

Xinyue Gu[1], Lei Zhu[2], and Xintao Liu[1]

1. Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Hong Kong, China
2. School of Architecture, South China University of Technology, Guangzhou, China

Abstract: Urban environments offer a wealth of opportunities for residents to respite from their hectic life. Outdoor running or jogging becomes increasingly popular of an option. Impacts of urban environments on outdoor running, despite some initial studies, remain underexplored. This study aims to establish an analytical framework that can holistically assess the urban environment on the healthy vitality of running. The proposed framework is applied to two modern Chinese cities, i.e., Guangzhou and Shenzhen. We construct three interpretable random forest models to explore the non-linear relationship between environmental variables and running intensity (RI) through analyzing the runners' trajectories and integrating with multi-source urban big data (e.g., street view imagery, remote sensing, and socio-economic data) across the built, natural, and social dimensions, The findings uncover that road density has the greatest impact on RI, and social variables (e.g., population density and housing price) and natural variables (e.g., slope and humidity) all make notable impact on outdoor running. Despite these findings, the impact of environmental variables likely change across different regions due to disparate regional construction and micro-environments, and those specific impacts as well as optimal thresholds also alter. Therefore, construction of healthy cities should take the whole urban environment into account and adapt to local conditions. This study provides a comprehensive evaluation on the influencing variables of healthy vitality and guides sustainable urban planning for creating running-friendly cities.

Keywords: street view imagery, urban pavements, healthy cities, urban vitality, running-friendly cities, running intensity

1. Introduction

Healthy cities are literally defined as urban areas that continuously create and improve a health-promoting environment through providing basic sanitation and health infrastructure to enable residents achieving a better quality of life (Liu et al., 2022; Tang & Long, 2022). Therefore, developing healthy cities has received considerable attentions from many countries across the

Received: 2024-06-18 **Accepted:** 2024-12-16

Corresponding author: Xintao Liu, Associate Professor, specialized in urban informatics. Email: xintao.liu@polyu.edu.hk; Tel.: (852) 2766-5962.

globe, since it is recognized as an effective means to embrace the changing circumstances and promote human health and well-beings (WHO, 2020). Specifically, an emphasis on healthy vitality, which means being vigorous and active to maintain physical and mental health to live, grow, and develop, becomes as of importance. Cities encompassing healthy vitality can reduce health risks associated with non-communicable diseases, such as obesity and cardiovascular disease (Dong et al., 2023; Yang et al., 2024). By fostering environments that promote healthy vitality, cities thus encourage individuals having an active and fulfilling life, thereby enhancing the overall well-being of population (Dong et al., 2023; Qiu & Chang, 2021). Sidewalks, urban parks, and green spaces offer a refreshing break to engage in physical outdoor activities, such as walking, cycling, and jogging (Higuera-Mendieta et al., 2021). These cardiovascular exercises can consolidate physical fitness, facilitate one's mental recovery, and embrace healthy aging, all contributing to an improved quality of life.

Adopting the running / jogging data as a proxy can indeed shed light on the healthy vitality of urban environment, since these data are specifically collected to track the runs and monitor the progress over time, while other modes of travel may be categorized as commuting only. Recent studies have proven that healthy vitality of jogging is closely associated with the built environment (Yang et al., 2023; A. Zhang et al., 2022), especially with public facilities, including road facilities, recreation parks, etc. These greatly raise concerns toward the road accessibility, pedestrians' convenience, and commuters' safety (Qiu & Chang, 2021; Salazar Miranda et al., 2021). Besides, other variables of the built environment that are spatially connected to the commuting routes, such as building density (BD) and sky view index (SVI), also significantly affect the frequency and duration of outdoor sports (Dong et al., 2023; Yang et al., 2024).

Aside from the built environment, researchers also found that natural and social environment around urban footpaths also greatly impacts health vitality (Jiang et al., 2022; Liu et al., 2023). For example, Liu et al. (2022) found that green space in cities commonly attracts a higher jogging flows, indicating its positive effects toward exercising experience. Yang et al. (2024) also pointed out that social economic factors, such as population density, housing, and rental prices, exhibit certain effects on outdoor jogging. All these studies scientifically reveal the impacts of the urban environment on healthy vitality and further promote healthier architectural design and urban planning to a certain extent.

In terms of the modeling the healthy vitality, or specifically the running intensity (RI), most studies introduced multi-source urban big data and machine learning techniques to investigate the non-linear interplay among the concerning factors (Liu et al., 2023; Yang et al., 2024). For example, Yang et al. (2024) employed the Random Forest model to infer the non-linear associations between RI and built environment, visual landscape as well as socio-economic variables. Huang et al. (2023) applied multilevel regression models to assess how the natural environment, built environment, and traffic-related variables affect jogging behavior. Shashank et al. (2022) focused on variables of the built environment that both facilitate and hinder running, and adopted an affordance-based framework to analyze the feature/variable importance.

Despite these studies, the majority of them examine the effects of the urban environment from a specific perspective or at one dimension (Huang et al., 2023), where they do not take the social, natural, and built environments into account as a whole. Therefore, it remains an understudied topic that requires further efforts to fully examine the impact of the urban environment on health vitality.

Indeed, the urban environment as a whole comprises several dimensions; these include the built environment, natural environment, and social environment. These environments all contribute to the healthy vitality of the running experience, and it would be noncomprehensive to simply analyze the impact of only one of these dimensions without considering the other. Also, variables of the natural environment, such as terrain slope (Shashank et al., 2022), temperature (Bernard et al., 2021), humidity (Bernard et al., 2021; Rech et al., 2023), and wind speed (Ferguson et al., 2023; Turrisi et al., 2021), all greatly influence general public's motivation toward outdoor exercise (Huang et al., 2023). Despite that, existing studies neither considered all these variables nor proposed a comprehensive framework to reveal the relationship between jogging behavior and the urban environment. Not to mention most of these previous work focused on the effects of urban block-level environments on jogging activities (Huang et al., 2023; Y. Liu et al., 2023; Yang et al., 2024), failing to model the relationship into a finer urban unit.

This study attempts to construct a multi-environmental analysis framework to explore the urban environmental impacts on healthy vitality at a fine-resolution grid level. Two modern metropolises, Guangzhou and Shenzhen, are chosen in this study because they are regarded as representative healthy cities in mainland China. Nevertheless, their respective urban landscape, cultural heritage, and living habits differ greatly (Gu et al., 2024; Lin et al., 2020; Meng et al., 2020; Xu et al., 2021). Therefore, applying the proposed framework to these two distinct testbeds can reflect, to some extent, the universality of our method. Based on the running trajectories, the study integrates multi-source urban big data, including socio-economic data, street view imagery data, and remote sensing data, to explore the non-linear interplay between the urban environment and running patterns.

The innovative contributions of this study can be summarized as follows: (1) it provides a fine-scale estimation of the healthy vitality of urban pavements, constructing its non-linear relationships with respect to environmental variables; (2) it delivers a comprehensive evaluation of the variables by considering the social, natural, and built dimensions of the urban environment; and (3) it proposes planning and design guidelines that aim at constructing running-friendly cities. By contributing to these aspects, this study advances understanding of urban health dynamics through statistical and geographical analyses. Our findings highlight the importance of integrating various environmental considerations into urban design, which can lead to the formation of comfortable and healthier public spaces. Providing further practical guidelines certainly lead to enhancing the well-being of urban residents by promoting physical activities and reducing exposure to environmental stressors, ultimately fostering the development of healthy urban environment.

2. Materials and Methodology

2.1 Study Area

Two first-tier cities, Guangzhou and Shenzhen, located in the Guangdong-Hong Kong-Macao Greater Bay Area in southern China, are selected as the testbeds, as shown in Figure 1. Although Guangzhou and Shenzhen are in close geographic locations with similar economic status (Lin & Li, 2019), they have very distinct urban morphology and historic progress (Lin et al., 2020). Guangzhou is a thousand-year-old city with rich historical significance, while Shenzhen is the youngest city having rapid development during the past few decades China. This thus implies Guangzhou tends to be a more aging population hub, while Shenzhen attracts xennials and millennials to reside and commute (MacLachlan & Gong, 2022). Besides, studies have pointed out Shenzhen's roadways are commonly designed wider than Guangzhou's, and its infrastructure and buildings are built taller and way modern (Lin et al., 2020; Yang & Zhao, 2022), yielding a notable difference in terms of the built, natural, and social environments. Indeed, both Guangzhou and Shenzhen aim to position themselves as an international first-class healthy city and have continued incorporating healthy streets and roadways in their urban master planning. This makes them suitable testbeds for conducting the comprehensive analyses to examine the impact of urban environment on healthy vitality.

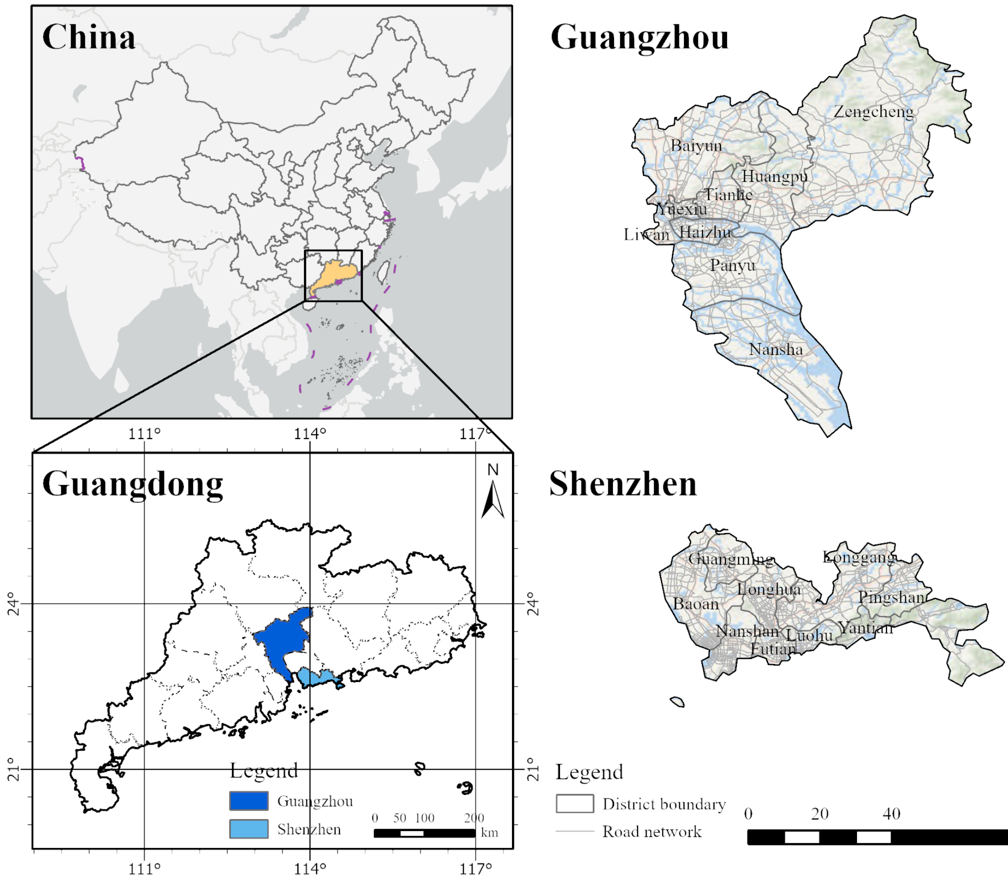

Figure 1. Geographic location of the research areas: Guangzhou-Shenzhen. Note: This study intentionally exclude those low-urbanized districts in Guangzhou, Huadu, and Conghua.

2.2 Data Description

The data adopted in this study include trajectories of running data representing the healthy vitality, where the street view imagery data and urban big data are adopted to quantify the built environment, natural environment, and social environment.

2.2.1 Running Trajectories

Most local residents' habits, especially after the COVID-19 pandemic, tend to become sedentary and inactive (Yang et al., 2024); therefore, cardiovascular exercises, such as running and cycling, have become crucial to help stay in shape and consolidate fitness. Unlike other forms of exercise, jogging simply requires no specialized skills, equipment, or peers (Jiang et al., 2022). This allows anyone to go jogging according to their preference and availability, making it one of most popular forms of exercise in town (Y. Liu et al., 2022). As a result, many studies adopted trajectories of running data as an agent to represent the healthy vitality (Huang et al., 2023; Liu et al., 2022; Yang et al., 2023), which can be attained from sports-tracking platforms, such as Strava (Huang et al., 2023), MapMyRun (Fletcher, 2022), and Keep. Among these sports platforms, Keep has more than 300 million registered users in China, and the cumulative downloads on the App Store platform in mainland China from January to June 2023 exceeded 9.6 million, making it the most popular sports APP in China. Therefore, we crawled all recorded data of running trajectories of users in Guangzhou and Shenzhen in 2020 based on the global running map produced by Keep without containing any personal privacy information (Yang et al., 2024). Then, we carried out pre-processing on the trajectories by removing redundant data, such as abnormal trips, to improve the data quality. The resulting number of valid trajectories was found to be 95,296 in total.

Based on these trajectories, the study areas were further divided into 250 × 250 m grids to facilitate counting of the total running length in each of the spatial grids, which was subsequently used to represent the RI. It can serve as an indicator to represent the healthy vitality of the roads/footpath in the corresponding area. Thus, RI can be formulated as:

$$RI = \frac{\sum S_{n1}}{\sum S_{n2}} \tag{1}$$

where $\sum S_{n1}$ is the total length of the running trajectories in the grid, and $\sum S_{n2}$ is the sum of road/footpath lengths in the grid.

2.2.2 Street View Imagery

Apart from running trajectories, the street view images were collected to semantically infer the environment and quality of running pathways. Street view images have been proven as an effective proxy to record and derive the spatial quality of urban pavements (Wang, 2024), and thus the study employed street view images to analyze the urban micro-environment. Details of data collection and processing are as follows. All the roads in Guangzhou and Shenzhen were first retrieved from the OpenStreetMap (OSM) platform. Then, specific road segments, such as overpasses that were inaccessible by pedestrians, were removed, leaving the remaining roads being divided by points at an interval of 50 m. Then, street view images were crawled with respect to these extracted points, mainly from the Baidu Map API, as it covers a relatively comprehensive range of road street views in China. The street view images crawled were mainly obtained from Year 2019 to 2021, where the searching range expanded to Year 2016 to 2023 in case there is a missing record.

After collecting all the required street view images, the study used the running trajectories to build a 10-m buffer zone, filtering out points around the running route. Ultimately, a total of 36,101 street view images were used for semantic segmentation with deep learning techniques. Specifically, the semantic segmentation of images was implemented based on the pre-trained model in GluonCV. Such deep learning model has been trained based on the Cityscapes dataset, an important urban street imagery segmentation dataset, which has been proven suitable for semantic segmentation of street view images (Cordts et al., 2016; Zhang et al., 2022). After the segmentation, each image was segmented and calculated the SVI, green viewing index (GVI), and other relevant indices (Liu et al., 2024).

2.2.3 Urban Big Data

The study constructs a comprehensive indexing system based on open-source urban big data to reflect the complex urban environment, which can be described into three dimensions: built environment, natural environment, and social environment. Table 1 describes each variable's dimension, selected reason, and the corresponding source.

Table 1. A list of built, natural, and social-related variables of the urban environment.

Dimensions	Variables	Description	References	Sources
Built	Building density (BD)	Total floor area divided by total area	(Liu et al., 2023; Yang et al., 2024)	OpenStreetMap
	Building spacing (BS)	Average distance to adjacent buildings	(Fathi et al., 2020; Lv et al., 2021)	
	Road density (RD)	Length of road divided by total area	(Liu et al., 2023; Yang et al., 2024)	
	Road facilities (RF)	$RF = P_{sidewalk} + P_{trafficlight} + P_{trafficsign} + P_{pole}$	(Lv et al., 2021)	Street view imagery
	Land-use mix (LUM)	Total number of POI points divided by total area	(Huang et al., 2023; Yang et al., 2024)	Gaode (Amap) API platform
	Sky view index (SVI)	$SVI = P_{sky}$	(Yang et al., 2024)	Street view imagery
	Visual motorization index (VMI)	$VMI = P_{road} + P_{trafficlight} + P_{trafficsign} + P_{car} + P_{bus} + P_{train} + P_{motorcycle} + P_{train}$	(Yang et al., 2024)	
Natural	Terrain slope (TS)	Average slope, spatial resolution of 30 m × 30 m	(Shashank et al., 2022)	CGIAR-CSI SRTM dataset

	Wind speed (WS)	Average wind speed, spatial resolution of 250 m × 250 m	(Turrisi et al., 2021)	ERA5 from ECMWF
	Humidity (HUMID)	Average humidity, spatial resolution of 250 m × 250 m	(Bernard et al., 2021; Rech et al., 2023)(Turrisi et al., 2021)	Landsat 8 data based on the GEE platform
	Normalized difference vegetation index (NDVI)	Average NDVI, spatial resolution of 250 m × 250 m	(Huang et al., 2023; Y. Liu et al., 2023)	
	Greening view index (GVI)	$GVI = P_{vegetation}$	(Liu et al., 2023; Yang et al., 2024)	Street view imagery
Social	Population density (PD)	Total population divided by grid area, spatial resolution of 100 m × 100 m	(Liu et al., 2023; Yang et al., 2024)	World pop
	Housing price (HP)	Average housing price	(Yang et al., 2024)	House for sale websites
	Degree of aging (AD)	Total population aged 65 and over, spatial resolution of 100 m × 100 m	(Dong et al., 2023)	World pop

Note: P refers to the percentage of the element, obtained by calculating the pixel points of the element in each street view image divided by the total pixel points of the whole image.

2.3 Analytical Framework and Methods

The comprehensive analytical framework for building a running-friendly city can be depicted in Figure 2. After attaining various variables from three environmental aspects, the process resulted in a total of 25,471 grids, with 13,465 grids formed in Guangzhou and 12,006 grids mapped in Shenzhen, to build regression models. The dependent variable of regression model in this study is the RI (see Eq. 1), and the independent variables are the 16 environmental variables (see Table 1) across the built, natural and social dimensions. Table 2 gives descriptive statistics for the dependent and independent variables.

To validate the variables' noncollinearity, the study first carried out the Ordinary Least Square (OLS) optimization in ArcGIS Pro to detect the relationship (Dong et al., 2023). OLS is the most commonly used regression analysis method, based on the minimizing the prediction error value to calculate the regression coefficients between two sets of parameters (Fox, 1961). It also provides an effective means to examine whether the relationship between healthy vitality and environmental variables is linear or non-linear. If the regression fitting is satisfied enough in the OLS model, it reflects as a linear relationship. Conversely, it indicates that the linear OLS model is not capable of well optimizing the variables, thus implying a non-linear relationship (Gu et al., 2024).

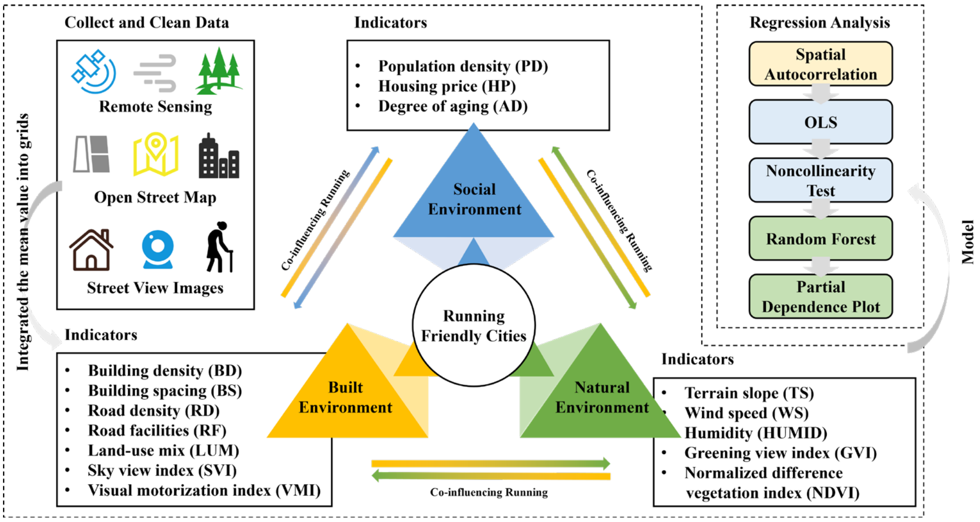

Figure 2. The overall analytical framework of the study. (Note: Icons on the chart are from https://www.iconfont.cn/)

Previous studies have shown that different urban environment variables have non-linear impacts on healthy vitality (Dong et al., 2023; Kim & Lee, 2023). Therefore, the study thus explores a suitable non-linear regression model to reveal the relationship. After comparing different models' performances, we intentionally select the random forest model to carry out the regression analysis. The random forest model is a traditional method that combines random node optimization and bagging to construct forests of uncorrelated trees using a CART-like process (Ho, 1995). It has been proven to be an effective way to model the non-linear impacts of the urban environment (Peng et al., 2023; Wu et al., 2021). To further improve the performance of random forest model, we utilized the GridSearchCV under scikit-learn in Python and plotted validation curves to visualize and understand the parameter optimization in order to determine the best model parameters (Belete & Huchaiah, 2022).

Table 2. Descriptive statistics of all environmental variables.

Variable	Max.	Min.	Mean	S.D.
Dependent variable				
Running intensity (RI)	34736.16	0	36.50	388.45
Built environmental variable				
Building density (BD)	0.87	0	0.09	0.001
Building spacing (BS)	86.97	0	5.65	11.70
Road density (RD)	0.10	0	0.01	0.01
Road facilities (RF)	0.71	0	0.02	0.03
Land-use mix (LUM)	1828.00	0	46.08	73.75
Sky view index (SVI)	0.65	0	0.12	0.15
Visual motorization index (VMI)	0.54	0	0.13	0.14
Natural environmental variable				
Terrain slope (TS)	31.21	0	4.93	3.27

Wind speed (WS)	2.16	0	1.32	0.27
Humidity (HUMID)	0.70	0	0.61	0.02
Normalized difference vegetation index (NDVI)	0.84	-0.54	0.34	0.16
Greening view index (GVI)	0.89	0	0.12	0.16
Social environmental variable				
Population density (PD)	358.97	0	80.12	85.34
Housing price (HP)	627852.00	0	9053.77	22263.92
Degree of aging (AD)	40.56	0	4.11	6.00

3. Results

3.1 Spatial Pattern of Healthy Vitality

We first demonstrate the spatial distribution of the healthy vitality in Guangzhou and Shenzhen, as shown in Figure 3. The histograms presented quantify the differences of RI between different districts in each of these two cities. From the statistical perspective, one can observe that the overall RI in Guangzhou is nearly 1.5 times higher than that in Shenzhen, which implies that the healthy vitality of Guangzhou is significantly better. In addition, the healthy vitality of Tianhe, Haizhu, and Yuexiu districts (which are rather old, well-developed districts in the city center) in Guangzhou plays a dominant role in promoting outdoor jogging activities, with an average RI of over 30.

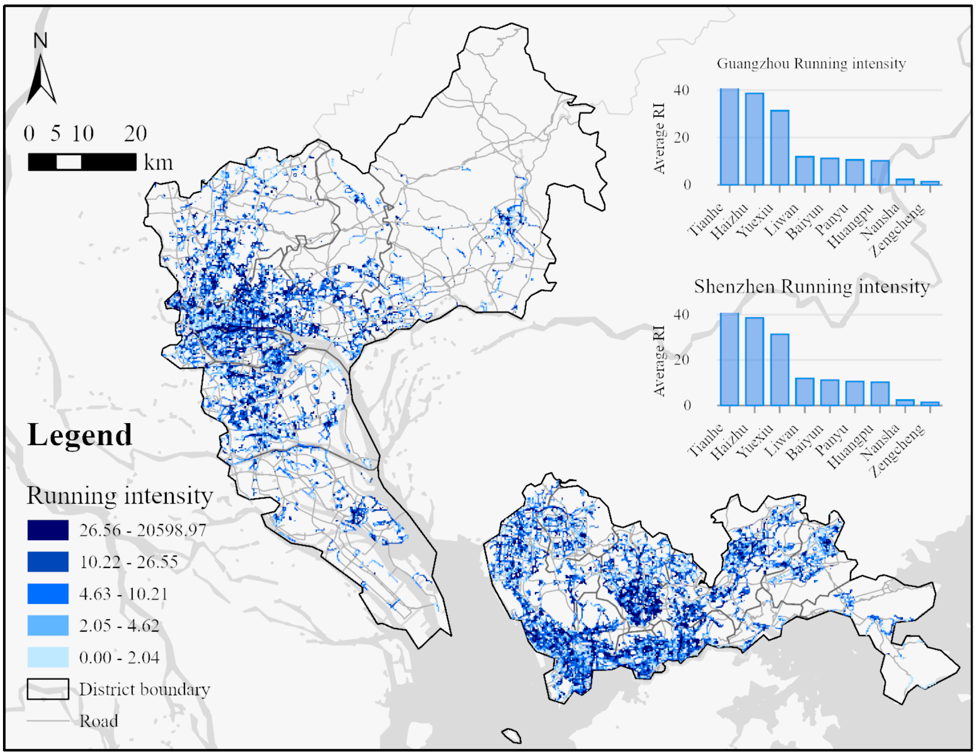

Figure 3. A comparison of RI distribution in Guangzhou and Shenzhen.

From the geospatial perspective, the distribution of RI in Guangzhou is more aggregated, showing a monocentric spatial structure. In comparison, the RI in Shenzhen tends to be decentralized and exhibits a polycentric structure. As expected, outdoor jogging activities in the cities follow the spatial structure of the respective cities. In both cities, the highest RI is found in the central districts, i.e., Tianhe, Haizhu, and Yuexiu districts in Guangzhou and Futian, Luohu, and Nanshan districts in Shenzhen. It can be concluded that healthy vitality seems closely related to a city's center, as these areas have higher population and more facilities. On the contrary, healthy vitality is relatively lower in those suburban areas of the cities, such as the Nansha and Zengcheng districts in Guangzhou and the Yantian district in Shenzhen.

3.2 Importance Ranking of Various Environments on Healthy Vitality

Our analysis subsequently proceeds to model the relationship between healthy vitality and multiple variables of the urban environment. As mentioned in the Section 2.3, we first employed the OLS model in ArcGIS Pro to validate the Variance Inflation Factor (VIF) and the non-linear effects of independent variables. The results of the OLS modeling are summarized in Table 3, which reveals that all the variables are non-collinear, since the VIFs are all less than five. Furthermore, the overall performance of the model is unsatisfactory, with R-squared yielded 0.004928 and Akaike information criterion (AIC) equivalent to 375,900. This implies that the urban environment has a non-linear effect on RI, which needs to be further modeled using the random forest model.

Table 3. The summary results of the OLS model are based on the variables in the study.

Variable	Coefficient	StdError	Probability	VIF
Intercept	-87.2224	79.2278	0.2709	--------
Building density (BD)	2.5185	1.6015	0.1158	1.0135
Building spacing (BS)	0.1781	0.2348	0.4480	1.2800
Road density (RD)	-3493.8064	358.5119	0.0000*	1.3104
Road facilities (RF)	-82.2313	96.4007	0.3936	1.7119
Land-use mix (LUM)	-0.0604	0.0428	0.1583	1.6894
Sky view index (SVI)	-20.2166	27.8070	0.4672	3.0127
Visual motorization index (VMI)	28.7144	37.7333	0.4467	4.8981
Terrain slope (TS)	1.4817	0.8386	0.0773	1.2782
Wind speed (WS)	0.9298	11.0745	0.9331	1.5470
Humidity (HUMID)	264.2025	130.2725	0.0426	1.0963
Normalized difference vegetation index (NDVI)	-25.5657	18.1278	0.1585	1.4858
Greening view index (GVI)	16.7672	21.5942	0.4375	2.0617
Population density (PD)	-0.0375	0.0513	0.4646	3.2494
Housing price (HP)	0.0005	0.0001	0.0000*	1.2244
Degree of aging (AD)	-0.0550	0.7155	0.9387	3.1239

Note: Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$)

As the relationship between RI and various variables are deemed to be non-linear, we then proceed to access the relationship based on random forest model. Prior to constructing the model, our experiments conduct a hyperparameter tuning: scales the data into a training / test set with a ratio of 8:2 and sets the random state as 105. Meanwhile, the following best fitted modeling parameters are input for the random forest model, i.e., max depth = 11, min samples leaf = 1, min samples split = 3, and *n* estimators = 100. With reference to these model settings, the random forest model of Guangzhou-Shenzhen is constructed, and the model performance is evaluated, as shown in Table 4. It can be found that all the metrics perform well, and no over- or under-fitting issues appear. This thus implies the non-linear relationship constructed by the random forest model is reliable.

Table 4. Results of random forest model constructed for Guangzhou-Shenzhen.

	R-square	MAE	RMSE	MAPE
Trained model	0.8103	31.3675	178.2063	83.6362
Tested model	0.8305	32.9757	120.1579	101.4985

Table 5 shows the importance ranking of the variables based on the random forest model of Guangzhou-Shenzhen. As expected, RD is the most important variable influencing RI and determine whether a running event occurs. This is followed by HP and LUM in second and third place, respectively, implying that urbanization significantly affects outdoor running in the city. Right after, several important natural variables ranked from 4^{th} to 8^{th} (except the 7^{th}) show that NDVI, TS, HUMID, and WS in some way influence the RI. PD and AD respectively occupy the 7^{th} and 9^{th} ranking, which shows that the influence of the social environment on RI also made an impact that cannot be neglected. Those variables related to the built environment, including VMI, SVI, RF, and BS, also contribute mildly to RI as they are listed at the bottom of the ranking.

Overall, the impact of both natural and social variables is rather mild, while the impact of built elements, except for the road density, is generally limited. Although the built environment is the basis for outdoor joggings taking place in public spaces, the variability of RI is not largely restricted by other built environments, but is rather influenced by the natural and social environment in the city. In particular, the three natural dimensions, NDVI, TS, and HUMID, which are relevant to influencing the experience of outdoor running, cannot be neglected with respect to their impact on the healthy vitality of the city.

Table 5. Feature importance ranking of random forest model for Guangzhou-Shenzhen.

Dimensions	Variables	Importance	Rank
Built	Road density (RD)	0.6408	1
Social	Housing price (HP)	0.1323	2
Built	Land-use mix (LUM)	0.0502	3
Natural	Normalized difference vegetation index (NDVI)	0.0482	4
Natural	Terrain slope (TS)	0.0330	5
Natural	Humidity (HUMID)	0.0284	6
Social	Population density (PD)	0.0160	7
Natural	Wind speed (WS)	0.0140	8

Built	Building density (BD)	0.0118	9
Social	Degree of aging (AD)	0.0118	10
Built	Visual motorization index (VMI)	0.0047	11
Built	Sky view index (SVI)	0.0038	12
Natural	Greening view index (GVI)	0.0035	13
Built	Road facilities (RF)	0.0010	14
Built	Building spacing (BS)	0.0007	15

3.3 Relationship between Healthy Vitality and Urban Environment

After modeling the importance ranking of all urban environment variables, the study further reveals several key relationships between RI and the predictor variables by using the partial dependence plots, which can help reveal the specific influences of each variable and guide further analysis or decision-making.

As shown in Figure 4, BD shows a sharp peak in partial dependence at around 0.2, indicating that RI increases significantly with building density up to this point, after which the effect stabilizes. Similarly, BS exhibits a steep jump in partial dependence up to approximately 10, suggesting that RI increases with building spacing up to this threshold, beyond which the effect becomes stabilized. RD displays certain fluctuations but generally indicates an increasing trend as road density increases, implying a positive relationship with RI.

In contrast, RF shows a gradual decrease in partial dependence as their presence increases, implying that more road facilities might be associated with lower RI. The LUM plot is relatively flat, indicating that the land-use mix has little to no effect on RI. The SVI demonstrates a sharp increase at around 0.2, suggesting that RI increases significantly with the sky view index up to this point, after which the effect stabilizes. VMI decreases in partial dependence as it increases, implying that higher visual motorization is associated with lower RI.

Similar to BS and LUM, TS exhibits an increase in the PDP, indicating a positive relationship with RI. WS shows a notable increase at approximately 1.0, suggesting that RI escalates with wind speed up to this point, beyond which the effect stabilizes. HUMID displays a certain degree of fluctuations, but in general indicates a mild decreasing trend in partial dependence as humidity increases, suggesting a negative effects. The NDVI, on the other hand, shows an increase in partial dependence, indicating a positive relationship that more vegetation cover associates with a higher RI. The GVI also echoes and displays a similar pattern of NDVI in which a higher greening view index is associated with a higher RI. Regarding the social environment, PD and HP exhibit relatively flat plots, indicating that these variables have little to no effect on the RI. Finally, AD shows a notable growth in partial dependence, resulting in a positive relationship with respect to RI.

Overall, the PDP analysis reveals those variables namely BD, BS, SVI, TS, WS, NDVI, GVI, and AD have a positive correlation with respect to RI. In contrast, RF, VMI, and HUMID demonstrate a negative relationship with RI. Variables, such as LUM, PD, and HP, appear to have little to no effect on the RI.

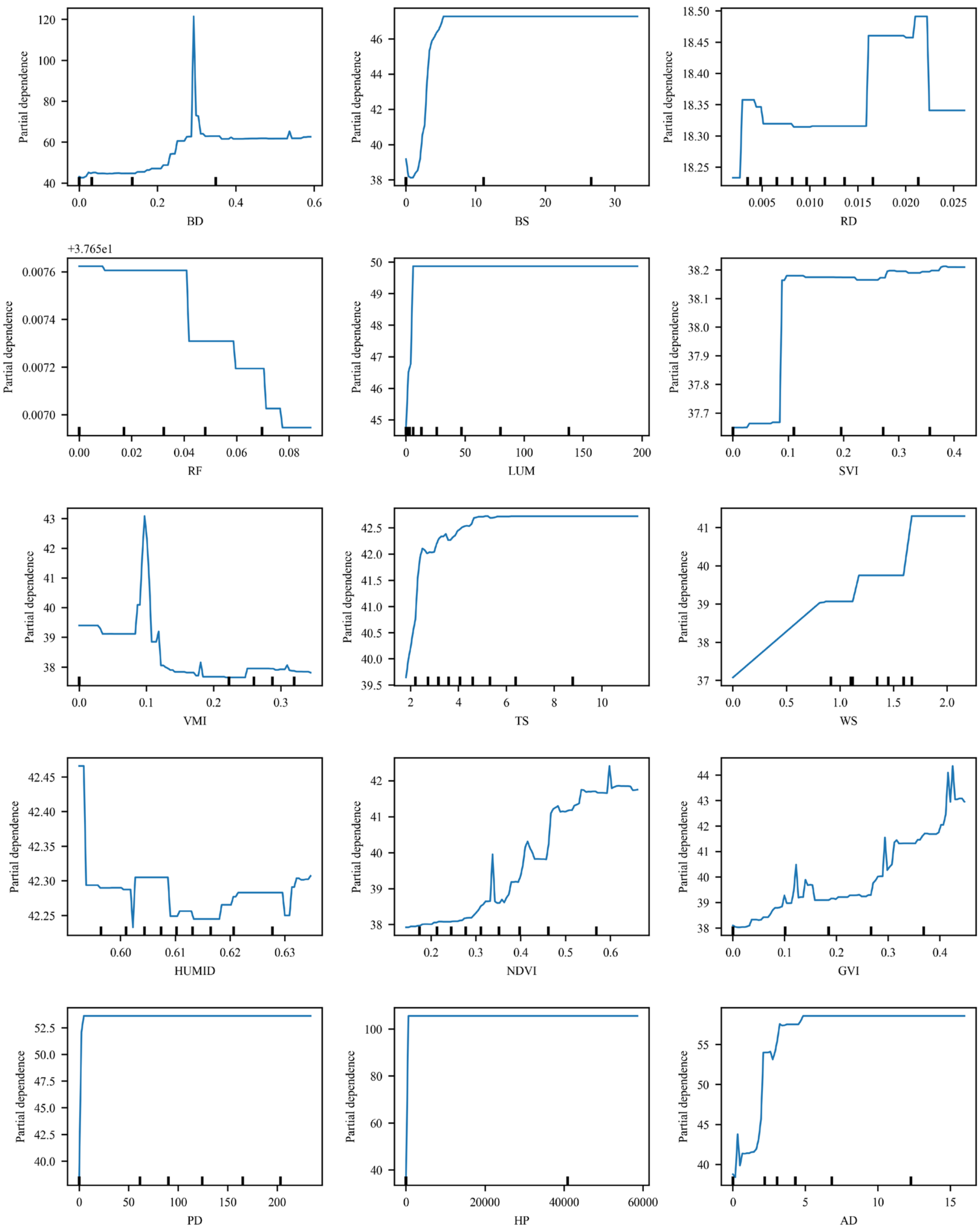

Figure 4. Partial dependence plots (PDPs) of each environmental variable based on the random forest model of Guangzhou-Shenzhen.

3.4 Comparing the RI Relationships Different Cities

A comparative analysis of RI across multiple regions / study areas can aid in providing an in-depth comprehensive understanding of the impact of each independent variable(s) on the dependent variable(s) (Gu et al., 2024; Wu et al., 2022). The study thus conducts two random forest models based on the dataset of Guangzhou and Shenzhen, respectively, to compare the results with each other and Guangzhou-Shenzhen region at large.

Similarly, the study first performs hyperparameter tuning by scaling the data to the training / test set with a ratio of 5.5: 4.5, random state as 42, and inputting the following modeling parameters into the model: max depth = 11, min samples leaf = 1, min samples split = 5, and n estimators = 100. Table 6 shows the model performance of the random forest model in Guangzhou. Although the overall model performs slightly worse than the model of Guangzhou-Shenzhen, most metrics are still well aligned and deemed to be acceptable. This implies that the non-linear relationship established by the random forest model is well grounded.

Table 6. Results of random forest model constructed for Guangzhou.

	R-square	MAE	RMSE	MAPE
Trained model	0.6990	29.4682	208.2436	81.2317
Tested model	0.4910	44.9839	372.0665	103.7392

As for the random forest model in Shenzhen, the study scales the data into a training / test set with a ratio of 6:4, random state as 42 and inputs the following modeling parameters into the model after hyperparameter tuning: max depth = 12, min samples leaf = 1, min samples split = 15 and *n* estimators = 50. Table 7 shows the model performance of the random forest model in Shenzhen. Although the model performs slightly worse in terms of R-squared, it is still comparable to the resulting Guangzhou-Shenzhen and Guangzhou models in terms of three commonly evaluated metrics. In addition, the model does not seem to suffer from overfitting, which is also considered reliable.

Table 7. Results of random forest model constructed for Shenzhen.

	R-square	MAE	RMSE	MAPE
Trained model	0.2147	30.6020	249.8227	95.4698
Tested model	0.1507	39.3604	313.3357	112.9087

Table 8 compares the importance ranking of the environmental variables in Guangzhou-Shenzhen, Guangzhou and Shenzhen. Also, the last column of Table 8 presents the average rankings across the three different models to reflect a more generalized impact. As expected, there exist notable differences in the importance ranking between different testbeds. RD always ranks first on average, while RF ranks last in all the three rankings, making it the least influential on average. This suggests that constructing suitable dense roads could be effective for outdoor jogging activities to promote healthy vitality.

In terms of other variables, HP always ranks as top three regarding its impact on RI in Guangzhou-Shenzhen and Guangzhou, but it drops to the tenth place in Shenzhen. This can be attributed to the high HP in Shenzhen with limited variation, making its impact insignificant. TS ranks as third on average, which is in the middle of the pack among all three regression models, with little variation in impact. It is closely followed by HUMID, which ranks fourth overall and second in Shenzhen, implying that outdoor jogging activities are more sensitive to humidity in Shenzhen than Guangzhou. The fifth variable on average is LUM, which ranks lower in Guangzhou, possibly due to the lack of mixed-use areas in this city. VMI ranks sixth on average, also receiving a higher rank in Guangzhou (fourth) and Shenzhen (third) but only being listed on the 11th place in Guangzhou-Shenzhen, as overtaken by other natural variables. Although BD and PD tie at 7th on average, PD shows a significantly different impact in different regions: PD ranks second in Guangzhou but 12th in Shenzhen, which can be explained by the distribution of population density.

Table 8. Comparison of the importance ranking of the variables based on three random forest models in different research areas: Guangzhou-Shenzhen (GS), Guangzhou (GZ) and Shenzhen (SZ).

Variables	GS	GZ	SZ	Avg.
Road density (RD)	1	1	1	1
Housing price (HP)	2	3	10	2
Land-use mix (LUM)	3	9	5	5
Normalized difference vegetation index (NDVI)	4	12	9	9
Terrain slope (TS)	5	7	4	3
Humidity (HUMID)	6	8	2	4
Population density (PD)	7	2	12	7
Wind speed (WS)	8	11	15	13
Building density (BD)	9	5	7	7
Degree of aging (AD)	10	10	13	11
Visual motorization index (VMI)	11	4	3	6
Sky view index (SVI)	12	15	6	11
Greening view index (GVI)	13	6	11	10
Road facilities (RF)	14	14	14	15
Building spacing (BS)	15	13	8	14

In contrast to VMI, NDVI ranks relatively lower in Guangzhou (12th) and Shenzhen (ninth) but fourth in Guangzhou-Shenzhen, resulting in an average ranking of ninth. GVI takes the tenth place on average, which ranks relatively high in Guangzhou (sixth), but its rankings in Guangzhou-Shenzhen and Shenzhen are relatively low, suggesting that the greening environment of urban pavements may have a greater effect on health in cities with a similar structure to Guangzhou. SVI and AD both rank 11th on average, with AD being stabilized in all three models, while SVI has a greater impact in Shenzhen. BS, which is similar to SVI, reflects the distance between buildings, resulting in a greater impact in Shenzhen but ranks lower in Guangzhou-Shenzhen. The similar effects of SVI and BS suggest that the spacing of high-rise buildings plays an important role in influencing healthy vitality toward young and modern cities, such as Shenzhen.

To further reveal the distinct patterns on how various variables impact RI, the study displays and compares the partial dependence plots for the respective three models of Guangzhou-Shenzhen, Guangzhou, and Shenzhen, as shown in Figure 5 to Figure 7, respectively. It can be seen that the three models show consistent patterns for most of the variables, indicating a steady relationship between RI and the urban environment.

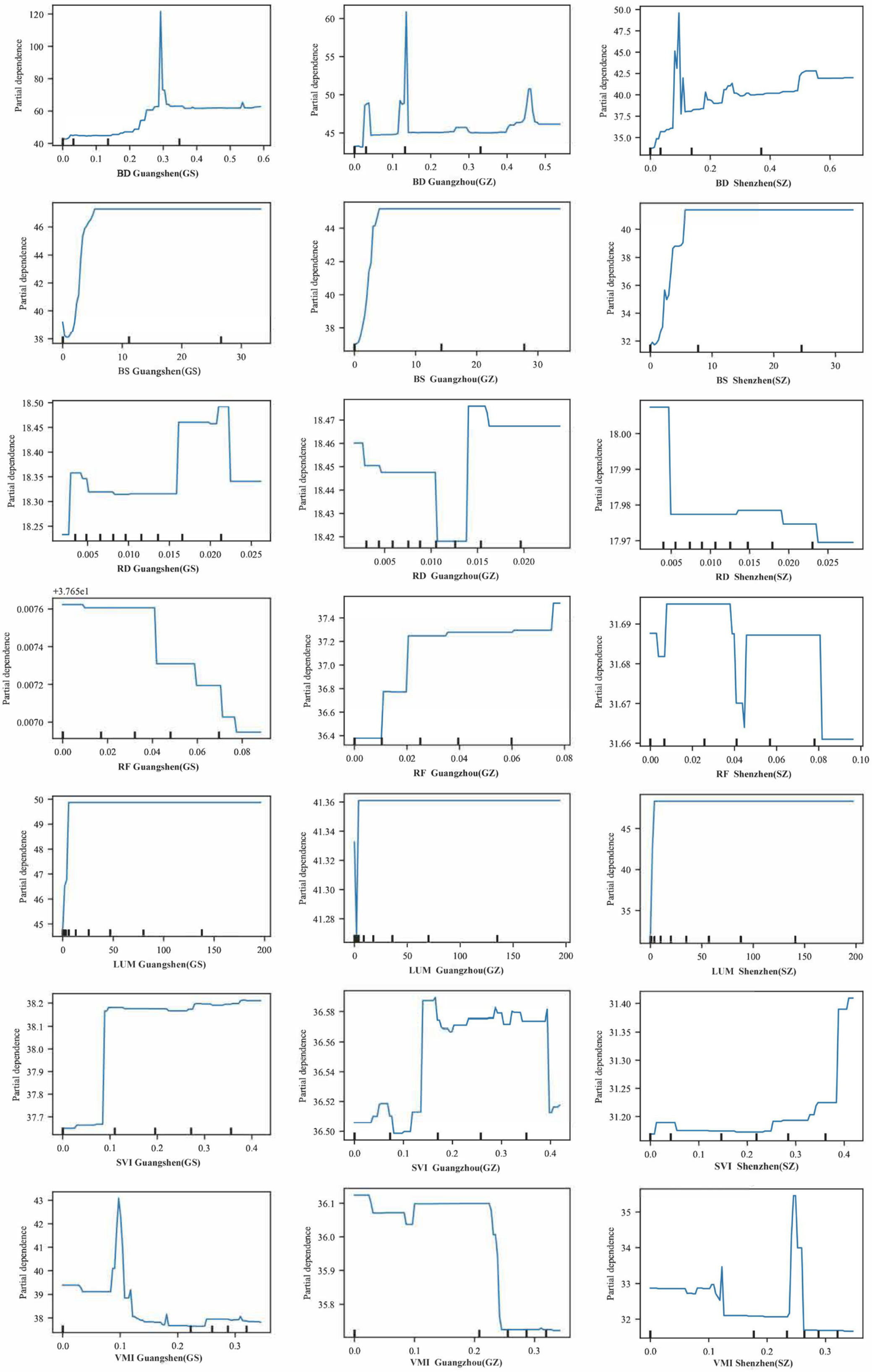

Figure 5. Comparison of the partial dependence plots (PDPs) of built environmental variables in Guangzhou-Shenzhen (GS), Guangzhou (GZ), and Shenzhen (SZ).

Regarding the impact of the built environmental variables (Figure 5), there exist certain similarities among BD, BS, LUM, SVI, and VMI across all models. In particular, as the value of each variable increases, the impact of BD shows a sharp increase at approximately 0.1 and subsequently stabilizes. Furthermore, the impact of BS shows a steep increase up to approximately 10, after which it stabilizes in all models. LUM remains constant, and SVI shows a sharp jump at around 0.1 and 0.3 in all models, while the effect of VMI rises significantly before dropping. Thus, when the value of BD reaches between 0.1 and 0.4, BS, LUM, SVI, and VIM are respectively above 8, 20, 0.1, and 0.1 to 0.3. As a result, their impacts on RI are notably significant.

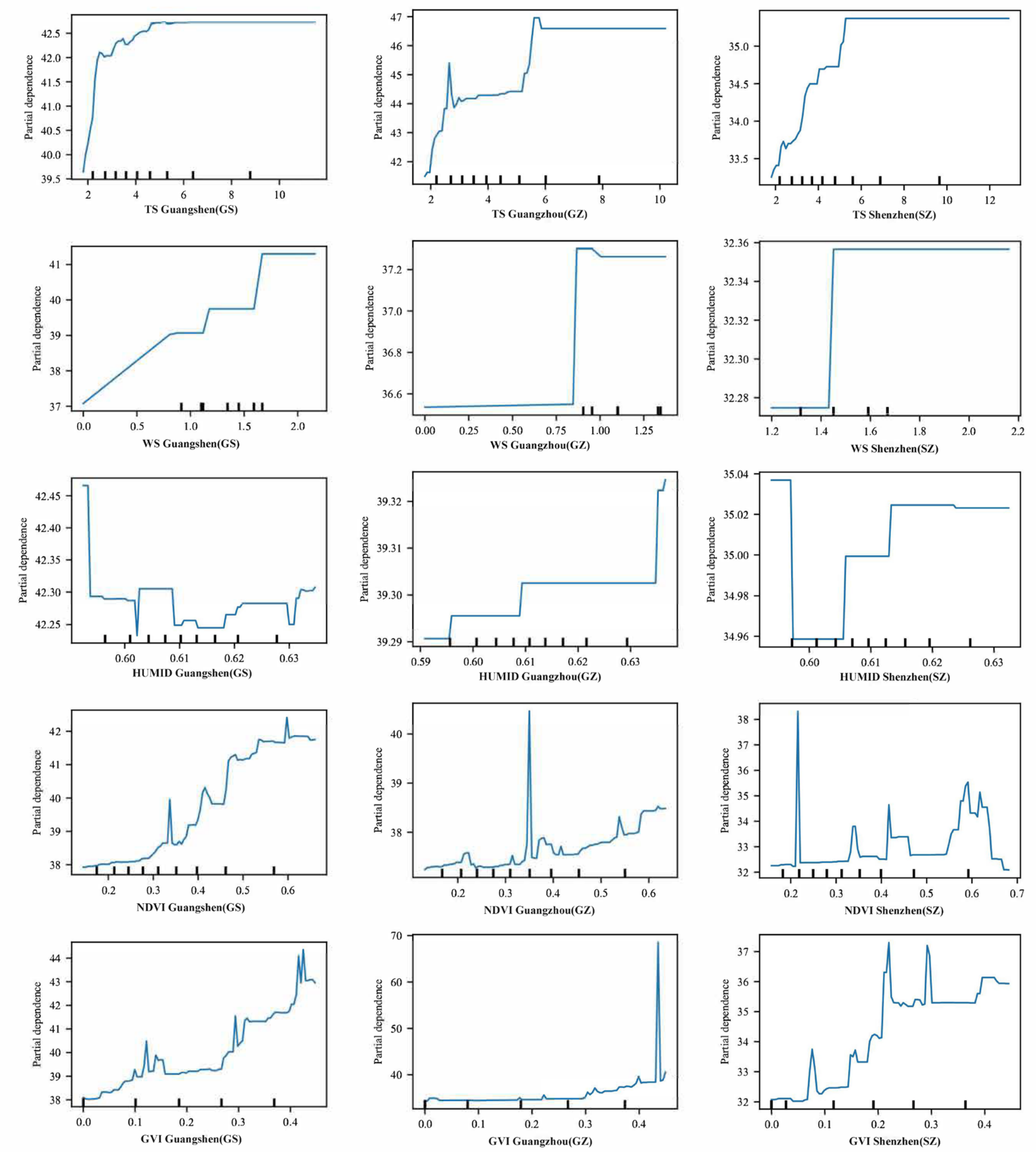

Figure 6. Comparison of the partial dependence plots (PDPs) of natural environmental variables in Guangzhou-Shenzhen (GS), Guangzhou (GZ), and Shenzhen (SZ).

Indeed, there exist certain glaring differences in terms of the impact of RD and RF on RI among all the models. The partial dependence of Shenzhen shows a sharp reduction in the effect of RD at around 0.01 when the value increases. One possible reason can be attributed to the historic development of Guangzhou, where such a mature city's road network is designed to be more pedestrian-friendly, thus facilitating outdoor physical activities. In contrast, Shenzhen is a rapidly expanding, compact city with a denser road network and a larger vehicle ownership (Meng et al., 2020). It is also the region with the highest and most concentrated traffic-related carbon emissions in the Guangdong Province (Deng et al., 2023). In Shenzhen, certain areas and communities with high road density may also suffer from issues, such as traffic congestion, air pollution, and noise, which in turn lead to a sharp decrease in the impact of road traffic on RI. In contrast, the random forest models of Guangzhou-Shenzhen and Guangzhou both show similar mild fluctuations along a generally increasing trend. In addition, RF shows a sharp increase at around 0.05 in Shenzhen, while the corresponding partial dependence of Guangzhou-Shenzhen and Shenzhen gradually decrease as their value increases. This implies that a reasonable increase in RD in the range of 0.1 to 0.2 can contribute to health vitality. A further increase of density does not promote RI and may even cause a negative impact in some cities.

Regarding the impacts of natural environmental variables (see Figure 6), only TS shows similar patterns among all models, where the impact of which gradually increases with higher TS value. On the contrary, WS, HUMID, NDVI, and GVI all demonstrate different patterns across the three random forest models. WS shows a spike at around 1.5 in the PDP of Shenzhen, while in Guangzhou-Shenzhen and Guangzhou, the corresponding PDPs are recorded with a sharp increase of around 1.0. In addition, NDVI and GVI both demonstrate more fluctuations in Shenzhen than the rest of the two with increasing values. With WS greater than 1, a smaller HUMID, and a higher NDVI and GVI, all can aid in promoting healthy vitality.

Notedly, when comparing the PDPs of humidity, one can observe that HUMID shows a sharp increase at around 0.61 in Shenzhen, and records slight fluctuations coupled with a decreasing pattern in the two other PDPs. This thus explains why humidity ranks second in Shenzhen and eighth in Guangzhou. In Guangzhou, the effect of humidity is smaller as the PDP curve changes gently, and when the humidity varies between 0.60 and 0.63, the variation of RI is rather limited, basically remaining in between 39.29 and 39.32. On the other hand, the effect of humidity is larger in Shenzhen, and the curve varies significantly. When the humidity varies between 0.60 and 0.63, the variation in RI is slightly larger, i.e., from 34.96 to 35.04. This indicates that the humidity has a significantly stronger impact on RI in Shenzhen, probably because Shenzhen is located nearby the coast suffer higher humidity, and the range of variation is also larger than that of Guangzhou (Zhou et al., 2022).

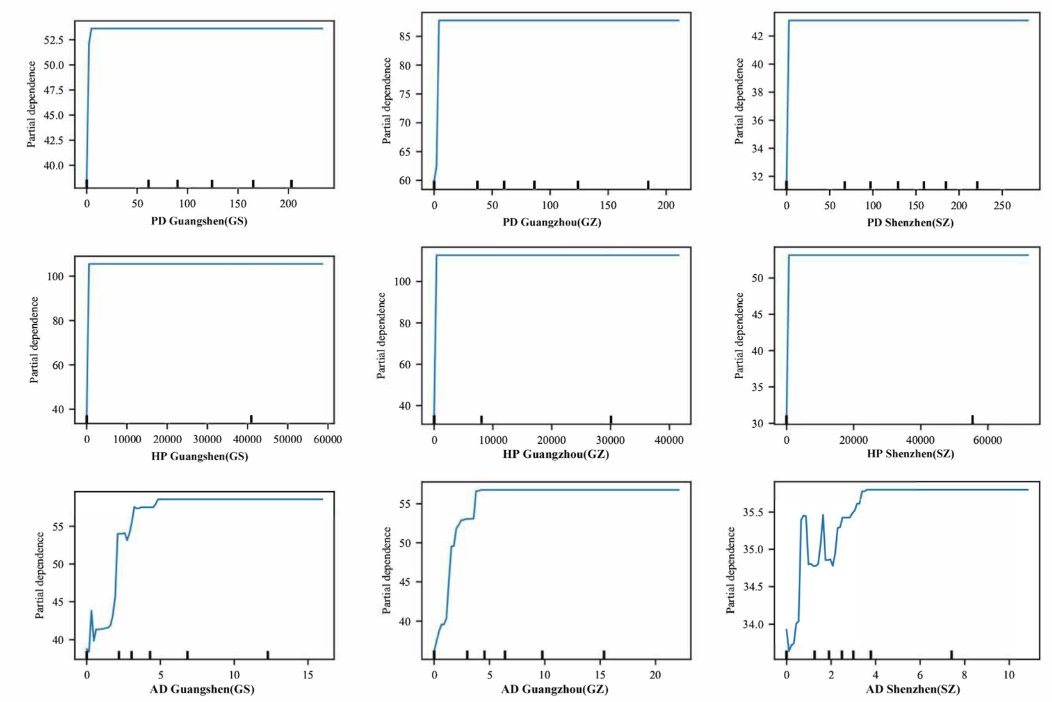

Figure 7. Comparison of the partial dependence plots (PDPs) of social environmental variables in Guangzhou-Shenzhen (GS), Guangzhou (GZ), and Shenzhen (SZ).

Meanwhile, the PDPs of GVI show differences in their influence on RI between Guangzhou (GZ) and Shenzhen (SZ). In Guangzhou, the impact of GVI remains relatively stable at lower values (i.e., 0 to 0.3), only with a significant increase when GVI exceeds 0.3 and obviously at 0.4. This suggests that high levels of green space in Guangzhou can significantly contribute to the RI. Corresponding green places, such as large recreational parks and landscaped areas, are likely the gathering spots frequented by runners, which result in higher importance rankings. Conversely, GVI on RI in Shenzhen shows a gradual and steady increase over the whole range of values (i.e., 0 to 0.4), with no abrupt changes. This suggests that even lower levels of GVI can gradually contribute to higher RI in Shenzhen. However, a lower GVI often means that the area is not vastly covered by green and does not attract higher RI, which is why the GVI in Shenzhen only ranks 11th.

As for the impacts of the social environment (Figure 7), these variables all have very similar impacts among the three random forest models. PD and HP remain constant, while AD shows a sharp increase at around 5 when its value increases. Thus, enhancing these helps promote healthy vitality, having comparative applicability in all cities.

4. Discussion

Based on the findings of this study, it is evident that environmental variables play a significant role in influencing RI. The research identifies key variables that significantly impact RI, which are RD, HP, VMI, and HUMID.

The results of the importance ranking indicate that RD consistently ranks first in terms of their impact on RI. This suggests that the density and layout of roads in a neighborhood are crucial for promoting running activities (Liu et al., 2023). Urban planners should consider and design communities formed with considerable road density to facilitate outdoor running. This could involve designing communities with interconnected streets and fewer dead-ends, providing runners with different route options, and ultimately leading to a convenient and enjoyable running journey. The high ranking of HP, a very important socio-environmental variable, well aligns with the findings as reported in Yang et al. (2024), where areas with higher housing prices exhibit higher RI. It is worth noting that natural environment variables, such as NDVI, TS, HUMID, and WS, all are crucial for driving the RI. Previous studies have reported the impact of NDVI (Liu et al., 2022; Zhou et al., 2024); however, limited attention has been paid to broader natural environment variables, including TS, HUMID, and WS. These variables also profoundly influence outdoor running activity and need to be taken into consideration (Rech et al., 2023). Our study extends these findings by constructing a comprehensive framework encompassing built-nature-society environments.

On the other hand, the results show that the larger the RF, VMI, and HUMID values, the less significant in terms of positive contribution toward RI. This suggests that certain environmental variables, such as obstacles on the road (Huang et al., 2023), high-traffic-volume roads (Yang et al., 2024), and high humidity (Bernard et al., 2021), can hinder running activities. Urban planners should place an emphasis on controlling these adverse variables. For instance, those running pathway / sidewalk can be designed with free of obstacles at one point. Also, urban planners should ensure the intensity of motorization nearby should reach an acceptable level. Reducing the surrounding noise and exhausting fumes can thus contribute to a pleasant running environment (Huang et al., 2023).

Interestingly, the study also uncovers that increasing the degree of variables, including BS, LUM, SVI, TS, WS, PD, HP, and AD, can significantly raise their impacts on RI. For example, BS and SVI are positively associated with RI, and urban planners can thus design communities with wider spacing or road width to create a more open and welcoming running environment (Basu & Sevtsuk, 2022). Since green visibility and NDVI have positive benefits toward running activity, landscape designers can also ensure communities planted with enough trees and shrubs to increase the vegetation index (Yang et al., 2024) on one hand, and to facilitate cooling effects with comfortable temperature and wind speed, on the other hand (Ferguson et al., 2023). Also, terrain slope is positively correlated with RI, possibly due to the overrepresentation of running trajectories collected from athletic runners (Huang et al., 2023). HP, PD, and AD are also positively correlated with RI, which aligns with the findings from previous studies (Huang et al., 2023; Y. Liu et al., 2023).

On should note that the impact of these environmental variables reveals strong regional characteristics. Despite being a first-tier city, Guangzhou has its rich historic development (i.e., 2,234 years old), while Shenzhen is rather a newly formed city (i.e., 40 years old). As a result, these two cities have a structural difference in the urban environment and their respective landscape is unlikely similar. Our findings also reveal such a phenomenon where impacts of certain variables, such as HP and HUMID, vary significantly between these two different regions. This thus reminds urban planners to consider the local circumstances when designing running friendly communities.

Finally, the impact of GVI on health is strong in cities, such as Guangzhou, implying that GVI plays a key role in promoting physical activity (Yang et al., 2021; Zang et al., 2020, 2023). Despite its low ranking of importance found in Shenzhen, increasing the GVI can accommodate in boosting RI to a certain extent (Zhong et al., 2024). This suggests that greening construction around urban pavement and sidewalk can contribute to the healthy vitality. Urban planners should pay attention to enhance the greening construction in cities, such as planting trees and reserving green spaces, yielding a more welcoming and healthier running environment.

Taken together, improving the urban environment may be a viable strategy for encouraging physical activity and promoting civic health. The results of this study provide a valuable reference for planning strategies that aim at promoting running activities and improving urban health vitality. The importance ranking of variables can be used as a basis for prioritizing interventions, and the non-linear effects demonstrated by partial dependency plots are critical for developing refined guidelines for environments that promote jogging.

In short, this study sheds light on how environmental variables influence RI. The findings of this study can inform urban planning strategies to promote outdoor exercises and improve public health. However, there are certain limitations in healthy vitality assessments and influencing evaluation. It would be ideal to integrate additional time-series data related to different outdoor sports activities to represent healthy vitality in cities. Besides, the current study mainly focuses on two cities. Future research could reflect to some extent the broad applicability of the framework by selecting cities with different characteristics as research subjects (e.g. differences in population size, level of economic development, cultural background, and climate zones, etc.).

5. Conclusions

This study offers a comprehensive and comparative analysis of the effects of urban environments on the healthy vitality of running in two modern Chinese metropolises, i.e., Guangzhou and Shenzhen. The study has established three interpretable random forest models using running trajectories coupled with multi-source urban big data across three analytical dimensions, i.e., built, natural, and social environments.

Our findings reveal that urban health vitality is influenced by a variety of environmental variables. Among which road density, housing price, and terrain slope have a significant and influential role toward the health vitality of residents. Furthermore, increasing environmental variables, such as building spacing, land-use mix, NDVI, temperature, wind speed, and population density can significantly contribute to the RI. The experimental results also highlight the importance of greening construction in promoting running activities, especially in cities with urban morphology to Guangzhou.

Taken together, this study provides valuable insights into the urban environment affecting the healthy vitality of outdoor running activities and recommends a number of planning guidance to form running-friendly cities. The findings can shed light on urban planning strategies that aim at promoting running activities and improving public health in the era of rapid urbanization.

Acknowledgment

This study would like to acknowledge the following three fundings: the NSFC General Program (42171455), the Hong Kong RGC Research Impact Fund (R5011-23), and the Hong Kong General Research Fund (15204121).

References

Basu, R., & Sevtsuk, A. (2022). How do street attributes affect willingness-to-walk? City-wide pedestrian route choice analysis using big data from Boston and San Francisco. *Transportation Research Part A: Policy and Practice*, *163*, 1–19. https://doi.org/10.1016/j.tra.2022.06.007

Belete, D. M., & Huchaiah, M. D. (2022). Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications*, *44*(9), 875–886. https://doi.org/10.1080/1206212X.2021.1974663

Bernard, P., Chevance, G., Kingsbury, C., Baillot, A., Romain, A.-J., Molinier, V., Gadais, T., & Dancause, K. N. (2021). Climate Change, Physical Activity and Sport: A Systematic Review. *Sports Medicine*, *51*(5), 1041–1059. https://doi.org/10.1007/s40279-021-01439-4

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). *The Cityscapes Dataset for Semantic Urban Scene Understanding*. 3213–3223. https://openaccess.thecvf.com/content_cvpr_2016/html/Cordts_The_Cityscapes_Dataset_CVPR_2016_paper.html

Deng, X., Chen, W., Zhou, Q., Zheng, Y., Li, H., Liao, S., & Biljecki, F. (2023). Exploring spatiotemporal pattern and agglomeration of road CO2 emissions in Guangdong, China. *Science of The Total Environment*, *871*, 162134. https://doi.org/10.1016/j.scitotenv.2023.162134

Dong, L., Jiang, H., Li, W., Qiu, B., Wang, H., & Qiu, W. (2023). Assessing impacts of objective features and subjective perceptions of street environment on running amount: A case study of Boston. *Landscape and Urban Planning*, *235*, 104756. https://doi.org/10.1016/j.landurbplan.2023.104756

Fathi, S., Sajadzadeh, H., Mohammadi Sheshkal, F., Aram, F., Pinter, G., Felde, I., & Mosavi, A. (2020). The Role of Urban Morphology Design on Enhancing Physical Activity and

Public Health. *International Journal of Environmental Research and Public Health*, *17*(7), Article 7. https://doi.org/10.3390/ijerph17072359

Ferguson, T., Curtis, R., Fraysse, F., Olds, T., Dumuid, D., Brown, W., Esterman, A., & Maher, C. (2023). Weather associations with physical activity, sedentary behaviour and sleep patterns of Australian adults: A longitudinal study with implications for climate change. *International Journal of Behavioral Nutrition and Physical Activity*, *20*(1), 30. https://doi.org/10.1186/s12966-023-01414-4

Fletcher, O. (2022). 'Friendly' and 'noisy surveillance' through MapMyRun during the COVID-19 pandemic. *Geoforum*, *133*, 11–19. https://doi.org/10.1016/j.geoforum.2022.05.004

Fox, R. A. (1961). Principles and Procedures of Statistics with Special Reference to the Biological Sciences. *Journal of the Royal Statistical Society Series D: The Statistician*, *11*(3), 170–171. https://doi.org/10.2307/2987461

Gu, X., Tang, X., Chen, T., & Liu, X. (2024). Predicting the network shift of large urban agglomerations in China using the deep-learning gravity model: A perspective of population migration. *Cities*, *145*, 104680. https://doi.org/10.1016/j.cities.2023.104680

Gu, X., Wu, Z., Liu, X., Qiao, R., & Jiang, Q. (2024). Exploring the Nonlinear Interplay between Urban Morphology and Nighttime Thermal Environment. *Sustainable Cities and Society*, *101*, 105176. https://doi.org/10.1016/j.scs.2024.105176

Higuera-Mendieta, D., Uriza, P. A., Cabrales, S. A., Medaglia, A. L., Guzman, L. A., & Sarmiento, O. L. (2021). Is the built-environment at origin, on route, and at destination associated with bicycle commuting? A gender-informed approach. *Journal of Transport Geography*, *94*, 103120. https://doi.org/10.1016/j.jtrangeo.2021.103120

Huang, D., Tian, M., & Yuan, L. (2023). Sustainable design of running friendly streets: Environmental exposures predict runnability by Volunteered Geographic Information and multilevel model approaches. *Sustainable Cities and Society*, *89*, 104336. https://doi.org/10.1016/j.scs.2022.104336

Jiang, H., Dong, L., & Qiu, B. (2022). How Are Macro-Scale and Micro-Scale Built Environments Associated with Running Activity? The Application of Strava Data and Deep Learning in Inner London. *ISPRS International Journal of Geo-Information*, *11*(10), Article 10. https://doi.org/10.3390/ijgi11100504

Kim, S., & Lee, S. (2023). Nonlinear relationships and interaction effects of an urban environment on crime incidence: Application of urban big data and an interpretable machine learning method. *Sustainable Cities and Society*, *91*, 104419. https://doi.org/10.1016/j.scs.2023.104419

Lin, J., & Li, X. (2019). Large-scale ecological red line planning in urban agglomerations using a semi-automatic intelligent zoning method. *Sustainable Cities and Society*, *46*, 101410. https://doi.org/10.1016/j.scs.2018.12.038

Lin, J., Wan, H., & Cui, Y. (2020). Analyzing the spatial factors related to the distributions of building heights in urban areas: A comparative case study in Guangzhou and Shenzhen. *Sustainable Cities and Society*, *52*, 101854. https://doi.org/10.1016/j.scs.2019.101854

Liu, C., Fan, C., & Mostafavi, A. (2022). *Graph Attention Networks Unveil Determinants of Intra- and Inter-city Health Disparity* (arXiv:2210.10142). arXiv.

https://doi.org/10.48550/arXiv.2210.10142

Liu, S., Ge, J., Ye, X., Wu, C., & Bai, M. (2023). Urban vitality assessment at the neighborhood scale with geo-data: A review toward implementation. *Journal of Geographical Sciences*, *33*(7), 1482–1504. https://doi.org/10.1007/s11442-023-2139-1

Liu, Y., Gu, X., Wang, Z., & Anderson, A. (2024). Urban greenery distribution and its link to social vulnerability. *Urban Forestry & Urban Greening*, *101*, 128542. https://doi.org/10.1016/j.ufug.2024.128542

Liu, Y., Hu, J., Yang, W., & Luo, C. (2022). Effects of urban park environment on recreational jogging activity based on trajectory data: A case of Chongqing, China. *Urban Forestry & Urban Greening*, *67*, 127443. https://doi.org/10.1016/j.ufug.2021.127443

Liu, Y., Li, Y., Yang, W., & Hu, J. (2023). Exploring nonlinear effects of built environment on jogging behavior using random forest. *Applied Geography*, *156*, 102990. https://doi.org/10.1016/j.apgeog.2023.102990

Lv, M., Wang, N., Yao, S., Wu, J., & Fang, L. (2021). Towards Healthy Aging: Influence of the Built Environment on Elderly Pedestrian Safety at the Micro-Level. *International Journal of Environmental Research and Public Health*, *18*(18), Article 18. https://doi.org/10.3390/ijerph18189534

MacLachlan, I., & Gong, Y. (2022). China's new age floating population: Talent workers and drifting elders. *Cities*, *131*, 103960. https://doi.org/10.1016/j.cities.2022.103960

Meng, L., Sun, Y., & Zhao, S. (2020). Comparing the spatial and temporal dynamics of urban expansion in Guangzhou and Shenzhen from 1975 to 2015: A case study of pioneer cities in China's rapid urbanization. *Land Use Policy*, *97*, 104753. https://doi.org/10.1016/j.landusepol.2020.104753

Peng, Z., Lu, W., Hao, T., Tang, X., Huang, J., & Webster, C. (2023). Cost-aware generative design for urban 'cool spots': A random forest-principal component analysis-augmented combinatorial optimization approach. *Energy and Buildings*, *295*, 113317. https://doi.org/10.1016/j.enbuild.2023.113317

Qiu, W., & Chang, H. (2021). The interplay between dockless bikeshare and bus for small-size cities in the US: A case study of Ithaca. *Journal of Transport Geography*, *96*, 103175. https://doi.org/10.1016/j.jtrangeo.2021.103175

Rech, C. R., Godtsfriedt, C. E. S., Arins, G. C. B., de Zorzi, V. N., Pazin, J., Hino, A. A. F., & Lopes, A. A. dos S. (2023). How do the year's seasons and specific weather indices affect physical activity and the patterns of use of public open spaces in the Brazilian context? *International Journal of Behavioral Nutrition and Physical Activity*, *20*(1), 124. https://doi.org/10.1186/s12966-023-01521-2

Salazar Miranda, A., Fan, Z., Duarte, F., & Ratti, C. (2021). Desirable streets: Using deviations in pedestrian trajectories to measure the value of the built environment. *Computers, Environment and Urban Systems*, *86*, 101563. https://doi.org/10.1016/j.compenvurbsys.2020.101563

Shashank, A., Schuurman, N., Copley, R., & Lear, S. (2022). Creation of a rough runnability index using an affordance-based framework. *Environment and Planning B: Urban Analytics and City Science*, *49*(1), 321–334. https://doi.org/10.1177/23998083211003886

Tang, L., & Long, H. (2022). Simulating the development of resilient human settlement in

Changsha. *Journal of Geographical Sciences*, *32*(8), 1513–1529. https://doi.org/10.1007/s11442-022-2008-3

Tin Kam Ho. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282. https://doi.org/10.1109/ICDAR.1995.598994

Turrisi, T. B., Bittel, K. M., West, A. B., Hojjatinia, S., Hojjatinia, S., Mama, S. K., Lagoa, C. M., & Conroy, D. E. (2021). Seasons, weather, and device-measured movement behaviors: A scoping review from 2006 to 2020. *International Journal of Behavioral Nutrition and Physical Activity*, *18*(1), 24. https://doi.org/10.1186/s12966-021-01091-1

Wang, Y. (2024). Fusing multi-source social media data and street view imagery to inform urban space quality: A study of user perceptions at Kampong Glam and Haji Lane. *Urban Informatics*, *3*(1), 21. https://doi.org/10.1007/s44212-024-00052-w

WHO. (2020). *Healthy Cities Effective Approach to a Changing World*.

Wu, H., Lin, A., Xing, X., Song, D., & Li, Y. (2021). Identifying core driving factors of urban land use change from global land cover products and POI data using the random forest method. *International Journal of Applied Earth Observation and Geoinformation*, *103*, 102475. https://doi.org/10.1016/j.jag.2021.102475

Wu, Z., Qiao, R., Zhao, S., Liu, X., Gao, S., Liu, Z., Ao, X., Zhou, S., Wang, Z., & Jiang, Q. (2022). Nonlinear forces in urban thermal environment using Bayesian optimization-based ensemble learning. *Science of The Total Environment*, *838*, 156348. https://doi.org/10.1016/j.scitotenv.2022.156348

Xu, Q., Yang, R., Zhuang, D., & Lu, Z. (2021). Spatial gradient differences of ecosystem services supply and demand in the Pearl River Delta region. *Journal of Cleaner Production*, *279*, 123849. https://doi.org/10.1016/j.jclepro.2020.123849

Yang, Ao, Y., Ke, J., Lu, Y., & Liang, Y. (2021). To walk or not to walk? Examining non-linear effects of streetscape greenery on walking propensity of older adults. *Journal of Transport Geography*, *94*, 103099. https://doi.org/10.1016/j.jtrangeo.2021.103099

Yang, C., & Zhao, S. (2022). Urban vertical profiles of three most urbanized Chinese cities and the spatial coupling with horizontal urban expansion. *Land Use Policy*, *113*, 105919. https://doi.org/10.1016/j.landusepol.2021.105919

Yang, Hu, J., Liu, Y., & Guo, W. (2023). Examining the influence of neighborhood and street-level built environment on fitness jogging in Chengdu, China: A massive GPS trajectory data analysis. *Journal of Transport Geography*, *108*, 103575. https://doi.org/10.1016/j.jtrangeo.2023.103575

Yang, Li, Y., Liu, Y., Fan, P., & Yue, W. (2024). Environmental factors for outdoor jogging in Beijing: Insights from using explainable spatial machine learning and massive trajectory data. *Landscape and Urban Planning*, *243*, 104969. https://doi.org/10.1016/j.landurbplan.2023.104969

Zang, P., Liu, X., Zhao, Y., Guo, H., Lu, Y., & Xue, C. Q. L. (2020). Eye-Level Street Greenery and Walking Behaviors of Older Adults. *International Journal of Environmental Research and Public Health*, *17*(17), Article 17. https://doi.org/10.3390/ijerph17176130

Zang, P., Qiu, H., Zhang, H., Chen, K., Xian, F., Mi, J., Guo, H., Qiu, Y., & Liao, K. (2023). The built environment's nonlinear effects on the elderly's propensity to walk. *Frontiers in*

Ecology and Evolution, *11*. https://doi.org/10.3389/fevo.2023.1103140

Zhang, A., Song, L., & Zhang, F. (2022). Perception of pleasure in the urban running environment with street view images and running routes. *Journal of Geographical Sciences*, *32*(12), 2624–2640. https://doi.org/10.1007/s11442-022-2064-8

Zhang, Y., Zhang, J., Xu, K., Tang, D., Li, Y., Wang, X., & Zhang, K. (2022). An improved method for urban Walk Score calculation considering perception of the street environment. *Transactions in GIS*, *26*(3), 1399–1420. https://doi.org/10.1111/tgis.12909

Zhong, Q., Li, B., Jiang, B., & Dong, T. (2024). Unleashing the potential of urban jogging: Exploring the synergistic relationship of high-density environments and exercise on Residents' Health. *Journal of Cleaner Production*, *466*, 142882. https://doi.org/10.1016/j.jclepro.2024.142882

Zhou, W., Liang, Z., Fan, Z., & Li, Z. (2024). Spatio–temporal effects of built environment on running activity based on a random forest approach in nanjing, China. *Health & Place*, *85*, 103176. https://doi.org/10.1016/j.healthplace.2024.103176

Zhou, Zhang, D., Cao, Y., Wang, Y., & Zhang, G. (2022). Spatio–temporal evolution and factors of climate comfort for urban human settlements in the Guangdong–Hong Kong–Macau Greater Bay Area. *Frontiers in Environmental Science*, *10*. https://doi.org/10.3389/fenvs.2022.1001064