



Issue 26(3), 2026

ISSN: 1567-7141

DOI: [10.59490/ejtir.2026.26.3.7816](https://doi.org/10.59490/ejtir.2026.26.3.7816)

TU Delft

OPEN Publishing

Research Article

Revealing the unequal availability and socioeconomic factors of health facilities based on taxi trajectory data and geospatial analysis

Yunying Liu^{1,2}, Huayang Liu², Zhaoqi Xu^{2,3}, Yiming Song², Yixin Wan², Xinyue Gu^{4,*}

¹ School of Architecture and Urban Planning, Nanjing University, China

² School of Geography and Planning, Sun Yat-sen University, China

³ Shanghai International College of Design and Innovation, Tongji University, China

⁴ Guangzhou Institute of Geography, Guangdong Academy of Sciences, China

* corresponding author xinyue.gu@gdas.ac.cn

Keywords

healthcare resource accessibility
two-step floating catchment area
random forest geographically
weighted regression
distance decay
urban inequality

Publishing history

Submitted: 10 September 2024

Revised: 17 October 2025

Accepted: 01 December 2025

Published: 30 September 2026

Cite as

Liu, Y., Liu, H., Xu, Z., Song, Y., Wan, Y., & Gu, X. (2026).

Revealing the unequal availability and socioeconomic factors of health facilities based on taxi trajectory data and geospatial analysis. *European Journal of Transport and Infrastructure Research*, 26(3).

<https://doi.org/10.59490/ejtir.2026.26.3.7816>

Abstract

As China's economy grows, healthcare equity concerns intensify. Historically, healthcare resource allocation often mismatches cities' economic status, and their spatial distribution remains uneven. This study examines healthcare disparities at a grid level using a two-step floating catchment area (2SFCA) that accounts for distance decay. It utilizes big data and API technology to develop a more accurate distance decay function based on travel data, obtaining results that reflect the healthcare fairness. Focusing on Shenzhen, China's largest city with a high number of immigrants and mobile population, it highlights a significant disparity between healthcare and other essential services despite rapid economic development. The approach initially fits a distance decay function with taxi GPS data, correlating driving behavior with hospital proximity. Subsequently, the 2SFCA is enhanced to assess healthcare facility accessibility accurately. Then, random forest geographically weighted regression models relationships among economic, physical environment, social dimensions, and healthcare accessibility in Shenzhen. Results indicate that aging and road network density exhibit the most extensive spatial influence on healthcare accessibility, both positively. These findings deepen understanding of healthcare resource inequality in metropolitan areas and support improved healthcare resource management and policy.

©2026 Authors. Published by TU Delft OPEN Publishing on behalf of the authors. This work is licensed under a Creative Commons Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)) license.

1 Introduction

With the in-depth promotion of the Healthy China initiative and the emphasis on people-oriented urban development, ensuring the well-being of individuals has been placed as a strategic priority for development. Healthcare resources are essential for ensuring the well-being of individuals and the functioning of society, and are directly related to residents' health status and cities' health service capacity, significantly affecting people's overall welfare. Large cities are often the richest in healthcare resources, but the acceleration of urbanization, population growth, and land expansion has made the supply-and-demand imbalance in healthcare resources increasingly prominent, exacerbating structural issues in the healthcare system (Zhang et al., 2022; Zhang et al., 2017). Hence, the distribution of healthcare resources has emerged as a crucial concern in urban planning and development. The study of healthcare resource accessibility is crucial to enhancing people's well-being and ensuring sustainable, superior urban development.

Currently, accessibility research has been conducted across several domains and has been a prominent area of research among Western scholars since the 1960s. Hansen, an American scholar, first introduced the concept of accessibility in 1959, arguing that it primarily refers to the ease of travel between two points (Hansen, 1959). Healthcare resource accessibility refers to the ease with which residents can access healthcare facilities with transportation support (Song et al., 2010; Liao & Jiang, 2018). The study of healthcare resource accessibility primarily focuses on two key areas, one of which is analyzing the accessibility of various demographic groups from the perspective of social equity, with particular attention to vulnerable groups such as the elderly (Jian et al., 2021; Zhang et al., 2017; Shen & Li, 2021; Tao & Shen, 2018). The second is conducting an overall accessibility analysis of specific cities or regions, using spatial accessibility methodologies to effectively identify areas lacking relevant services and optimize the distribution of facilities (Xing & Ng, 2022). Studies have indicated that cities generally show a decreasing trend from the central city to the suburbs, following a center-periphery pattern (Zeng et al., 2017; Gao et al., 2023). Additionally, there is an issue of unequal distribution of spatial accessibility, which poses a challenge to achieving fairness (Cheng & Lian, 2018; Gao et al., 2022). In terms of research scale, the majority of current studies take districts, counties, or subdistricts as the basic unit of analysis (Tong & Chen, 2017; Lv et al., 2022; Hu et al., 2021), and only a few studies examine residential land parcels (Zeng et al., 2017) and community units (Hu et al., 2021), which often fail to fully reflect the differences in healthcare resource accessibility and accurately identify areas with insufficient healthcare services (Lv et al., 2022; Hu et al., 2021). The selection of research units needs further refinement. Therefore, we analyzed accessibility using a 1000×1000 m grid as the minimum unit, compensating for the current research scale, which is limited to districts, counties, and subdistricts.

In recent years, various modeling methods have been applied to assess the accessibility of healthcare resources, and the most commonly used ones include the gravity model, the potential model, and the two-step floating catchment area method (Guagliardo, 2004; Song et al., 2009; Zenget al., 2017). Among them, the reliability of the evaluation results of the gravity model has been questioned (Li et al., 2018), the potential model ignores the difference in travel impedance to different levels of hospitals (Tao & Shen, 2018), and the two-step floating catchment area method incorporates the interactions between supply, demand, and minimum cost into the accessibility measure (Li et al., 2018; Tian et al., 2019). However, the traditional two-step floating catchment area method must account for distance attenuation (Tong & Chen, 2017; Liu et al., 2017). Currently, the rapid development of technology offers new possibilities for precise assessments of accessibility, especially the use of urban traffic data (B. Y. Chen et al., 2020). Research has begun investigating the use of Internet maps and related API services, as well as cell phone signaling data, to improve the two-step floating catchment area method (Tong et al., 2021; Ren & Wang, 2021; Su et al., 2023). Leveraging the rise of urban big data, we introduce a key improvement to the Two-Step Floating Catchment Area (2SFCA) method by deriving a more realistic distance decay function from real-

world cab GPS data. This approach moves beyond the theoretical impedance functions of traditional 2SFCA, thereby achieving a more precise assessment of healthcare accessibility that better reflects actual urban traffic conditions. Then, we use the improved two-step floating catchment area method combined with healthcare facility data and population data to analyze the accessibility features of healthcare resources and the shortage of healthcare in various regions.

The city, as a complex mega-system, involves a series of factors that shape the accessibility of healthcare resources. As research progresses, some studies have focused on the factors influencing access to healthcare resources. Some scholars have initially analyzed the influence of socioeconomic factors, among which population density, GDP, and transportation conditions play significant roles (Hansen, 1959; Lv *et al.*, 2022; Cheng & Lian, 2018; Gao *et al.*, 2022). Natural factors such as topography exert greater influence in rural areas (Tong *et al.*, 2021) and cities with diverse terrain (Liu *et al.*, 2023). Currently, geo-weighted regression (Liu *et al.*, 2019), geo-detector (Liu *et al.*, 2017), and decision tree models (Lv *et al.*, 2022) are primarily used to explore the impact factors of accessibility (Liu *et al.*, 2019). However, geo-detector specifically examines variations in influencing factors globally (Wang *et al.*, 2017). At the same time, geo-weighted regression focuses on the spatial heterogeneity of influencing factors at a local scale, and the decision tree model has limitations in accurately capturing geospatial data, thereby affecting the precision of the analysis (Geng *et al.*, 2020). Geographically weighted random forest regression (GWRF) incorporates spatial location information into the standard random forest model. Building a local model that accounts for spatial autocorrelation offers certain advantages in addressing spatial challenges (Y. Luo *et al.*, 2022; Santos *et al.*, 2019). While previous studies have utilized methods like geo-detector and geo-weighted regression, which have limitations in capturing complex non-linear relationships or spatial heterogeneity, our study innovatively adopts the Geographically Weighted Random Forest (GWRF) model. The GWRF model excels at modeling complex non-linear relationships and spatial autocorrelation simultaneously, allowing us to more accurately quantify the spatially varying impacts of economic, physical, and social factors on healthcare accessibility.

Population inversion is a phenomenon in which the number of foreign populations exceeds the number of households registered in the population. As China's largest migrant city and a city with an inverted floating population, Shenzhen's healthcare level and other major livelihood projects are not commensurate with its rapid economic growth (Tong & Chen, 2017; Hu *et al.*, 2024). As of 2022, the number of hospital beds per 1,000 population in Shenzhen is 3.72, significantly lower than the national average of 6.92. Compared with other large cities, Shenzhen's healthcare resources could be more adequate, but most high-quality resources are concentrated in the original Special Economic Zone, leading to a supply-and-demand imbalance. Therefore, it is crucial to investigate the accessibility of healthcare resources and the factors influencing their availability in Shenzhen to assess healthcare service capacity, improve the distribution of healthcare resources, and promote the city's overall health (Shen & Li, 2021).

Therefore, this study presents a comprehensive case study of Shenzhen by making three salient contributions to research on healthcare resource accessibility. Firstly, we address the scale limitation of previous studies by conducting our analysis at an unprecedented 1000m×1000m grid level, enabling a highly detailed identification of healthcare shortages. Then we enhance the precision of the 2SFCA method by incorporating a distance-decay function calibrated using real-time cab GPS data, thereby providing a more reliable measure of travel impedance. Also, we apply the GWRF model innovatively to dissect the influencing factors, offering superior insights into the spatial heterogeneity of the mechanisms underlying accessibility patterns. This multifaceted approach not only accurately assesses the current state of healthcare accessibility in Shenzhen but also establishes a robust framework applicable to other rapidly growing metropolises.

2 Study materials

2.1 Study area

Situated in southern China, Shenzhen is a coastal city enveloped by mountains on three sides and the sea on the remaining side. It is elongated from east to west and narrow from north to south. Shenzhen comprises 10 districts, covers around 1,997 square kilometers, and has a resident population of 17.66 million. Among them, the central area of Shenzhen comprises Futian District, Luohu District, and Nanshan District; Xin'an Street and Xixiang Street in Bao'an District; Minzhi Street and Longhua Street in Longhua District; and Bantian Street, Buji Street, Jihua Street, and Nanwan Street in Longgang District. It is renowned as one of China's most economically advanced cities, with a limited, densely populated land area that necessitates significant healthcare services. Since 2010, the overall quantity of healthcare resources in Shenzhen has consistently increased. However, some research has found significant disparities in the accessibility of healthcare services across locations, particularly in spatial accessibility. Hence, Shenzhen is an ideal testbed for this study.

This study employs a dual approach to analyze healthcare accessibility. Initially, a grid-based approach divides Shenzhen into 2,677 spatial units using a 1,000×1,000m grid. This grid-based analysis enables standardized spatial comparisons and a detailed understanding of micro-level variations in accessibility that might be overlooked by broader administrative boundaries. The grid approach is particularly useful for identifying specific areas within districts that may be underserved, thus offering insights that can inform administrative-level decisions. Furthermore, to reflect the real-world context of healthcare facility planning and decision-making, we integrate administrative boundaries into our analysis. We apply geographically weighted random forest regression at the administrative boundary level to explore spatial heterogeneity in the factors influencing healthcare accessibility. This dual approach ensures that our analysis remains grounded in the administrative context while also benefiting from the finer spatial resolution provided by the grid-based method. By combining grid-based insights with administrative boundary considerations, we provide a comprehensive understanding of healthcare accessibility in Shenzhen. This approach not only identifies potential areas for policy intervention but also ensures that administrative decisions are informed by detailed, location-specific data, ultimately contributing to the more equitable distribution of healthcare resources across the city.

2.2 Data

Healthcare facilities and population data

Due to the hierarchical nature of the medical facility system, the "Shenzhen Medical Institutions Management Measures" categorizes hospitals into three tiers based on their functions and tasks. Tier 1 hospitals serve residents within a single community; Tier 2 hospitals are regional medical centers providing comprehensive healthcare services to multiple communities; and Tier 3 hospitals offer advanced specialist medical services across several areas.

As motorization in the city continues to rise, over 80% of residents opt to use motor cars as their primary mode of transportation when traveling to hospitals (Haynes et al., 2006). This study focuses on analyzing hospital accessibility by car and therefore does not consider unclassified and level 1 hospitals, which have more limited catchment areas. Then, the study obtained data on 22 second-class hospitals and 61 third-class hospitals in Shenzhen from the official website of the Shenzhen Municipal Commission of Health. The number of beds and the number of specialized doctors are significant indicators of the magnitude and quality of hospitals and are frequently employed to evaluate the extent of healthcare facility provision (Y. Chen et al., 2022; Zhong et al.,

2016); the study compiles data from the official website of the Shenzhen government and each hospital, gathering information on each hospital individually.

The population data of Shenzhen is sourced from the LandScan population dataset, developed by the US Department of Energy's Oak Ridge National Laboratory (ORNL) and provided by East View Cartographic. This dataset has a resolution of 1,000m x 1,000m and incorporates various data sources, including high-resolution remotely sensed data, geographic information systems (GIS), and census data (He et al., 2024). Statistical and spatial modeling techniques are employed to calculate and predict population distribution, age, sex, and other demographic information for each global cell (pixel). The resulting data is presented in a grid format, currently considered the most accurate, reliable, and geographically based data for global vital statistics analysis (He et al., 2024). It is widely utilized in various demographic studies (Kyaw et al., 2023; Li & Derudder, 2022; Li & Liu, 2018; Zhang et al., 2023). LandScan outperforms other widely used demographic datasets (GPW, GHS-POP, WorldPop) and statistical population sources (censuses and yearbooks) in terms of spatial granularity and estimation error. It is considered the top performer among these datasets (Yin et al., 2021).

Healthcare travel time

Utilizing OD data that includes the origin and destination points of residents' journeys to hospitals is optimal for travel research. However, this data has privacy concerns and is not readily accessible. Taxis also significantly affect hospital travel due to their flexibility and quickness. Cab OD datasets are commonly used to analyze residents' travel behavior patterns, mostly because of the convenient access to cab data that provides exact information on origin-destination (Jiao et al., 2022; Wang et al., 2020; Yu et al., 2018).

The cab dataset is sourced from Shenzhen cab enterprises and comprises trajectory data of 747 cabs on October 1, 2018. It consists of 4.275 million raw data points, encompassing vehicle speed, color, time, license plate number, latitude and longitude coordinates, vehicle status (vacant or occupied), direction, and mileage traveled. This study determines the beginning and end of a trip based on the vehicle status field. When the vehicle status changes from vacant to occupied, it signifies the start of the trip. At this point, the latitude/longitude coordinates and time are recorded as the starting point.

Conversely, when the vehicle status changes from occupied to vacant, it indicates the end of the trip. The latitude/longitude coordinates and time at this point are recorded as the trip's ending point. Trips with distances less than 500 meters or speeds greater than 120 km/h are removed after extracting the trip origin-destination (OD) pairs. A total of 464,700 OD trip data were retrieved.

Socioeconomic data

The availability of healthcare resources is often highly associated with socioeconomic conditions and physical environment (Lv et al., 2022; Shen & Li, 2021; Hu et al., 2021). While the grid-based analysis provides a high-resolution view of accessibility patterns, understanding these patterns at the subdistrict (street) level is crucial for policy-making and urban planning. Administrative units are the primary entities responsible for implementing public health policies and resource allocation. Therefore, to bridge our high-resolution findings with actionable insights for urban governance, this study employs the subdistrict as the fundamental unit of analysis for the Geographically Weighted Random Forest Regression (GWRf) model.

Consequently, the healthcare accessibility values, originally calculated for each 1km grid cell, were aggregated at the subdistrict scale by taking their mean. This approach provides a standardized measure of accessibility for each administrative unit. Six indicators were selected to represent the three dimensions of each subdistrict: Economy, Physical Environment, and Society. The specific variables and their units of measurement are as follows: the Economy dimension includes the

Average Monthly Rental Cost (yuan) and the Total Value of Family Cars (ten thousand yuan); the Physical Environment dimension encompasses the Percentage of Housing Built After 2000 (%) and Road Density (m/km²); and the Society dimension includes the Percentage of the Population with a College Degree or Above (%) and the Percentage of the Population Aged 60 and Above (%). The data for these variables, except the road network data, were gathered from the Seventh National Population Census of Shenzhen at the subdistrict level. The road network data used to calculate density was sourced from the OpenStreetMap platform. It is worth noting that population density was initially considered as a candidate explanatory variable, given its potential confounding effect on healthcare accessibility. However, it was ultimately excluded from the GWRF model for two primary reasons. First, population density exhibited significant multicollinearity with other variables (VIF > 10), undermining model robustness. Second, and more fundamentally, healthcare accessibility itself is derived from population-weighted calculations in the E2SFCA method (Section 2.3.1); thus, reintroducing population density as an independent variable would create conceptual overlap and endogeneity, potentially leading to biased interpretations. A geographically weighted random forest (GWRF) was then conducted to analyze the spatial heterogeneity of the influencing factors on healthcare accessibility at the administrative level. Table 1 shows the descriptive statistics of the explanatory variables in this study.

Table 1. Descriptive statistics of variables. (The unit of analysis for all variables is the subdistrict (street).)

Variable (Unit)	Mean	Std	Min	Max
Average Monthly Rental Cost (yuan)	2097.116	1350.687	570.403	5690.259
Percentage of Housing Built After 2000 (%)	0.632	0.191	0.192	0.951
Percentage of Population with a College Degree or Above (%)	0.337	0.145	0.118	0.660
Percentage of Aged 60 and Above (%)	0.061	0.026	0.021	0.118
Road Density (m/km)	11863.426	7128.486	2280.248	31472.618
Total Value of Family Cars (ten thousand yuan)	10.518	5.267	4.267	28.429

2.3 Research methodology

Accessibility calculations: An improved two-step mobile search approach

The conventional 2SFCA approach comprises two distinct stages. The first stage involves calculating the supply-demand ratio R_i for each supply point i . This is done by dividing the capacity S_i of each supply point i by the total population of each demand point within a given region d_0 , as stated in Equation (1).

$$R_i = \frac{S_i}{\sum_{k \in \{d_k \leq d_0\}} P_k} \quad (1)$$

where P_k is the population at demand point k , and d_{ki} is the distance between demand point k , and supply point i .

In the second phase, the accessibility A_j of demand point j is determined by summing the supply-demand ratios R_i within the same region d_0 , as obtained in the first step, as stated in Equation (2).

$$A_j = \sum_{i \in \{d_i \leq d_0\}} R_i \quad (2)$$

In light of the current circumstances where the desire to travel diminishes as the distance increases, scholars have proposed several attenuation functions $f(d_{ki})$, as stated in Equation (3)(4), such as the Gaussian function, kernel function, and gravity function, to quantify the impact of distance attenuation (W. Luo & Qi, 2009; Polzin et al., 2014; Tao et al., 2014). Using GIS network analysis to calculate the distance between supply and demand is prone to biased estimates due to inaccuracies

in road network data. Internet map developer platforms, such as Google Maps and AutoNavi Maps, offer navigation path-planning services that provide real-time journey-time data. These platforms also enable users to calculate more precise and realistic travel costs, which can be useful for accessibility studies (Lyu et al., 2020; Pengfei et al., 2024).

$$R_i = \frac{S_i}{\sum_{k \in \{d_{ki} \leq d_0\}} P_k f(d_{ki})} \quad (3)$$

$$A_j = \sum_{i \in \{d_{ij} \leq d_0\}} R_i f(d_{ij}) = \sum_{i \in \{d_{ij} \leq d_0\}} \frac{s_i f(d_{ij})}{\sum_{k \in \{d_{ki} \leq d_0\}} P_k f(d_{ki})} \quad (4)$$

This research utilizes cab trajectories to extract the trip patterns of individuals who use a car or drive to the hospital. We then apply a decay function to these patterns, enhancing the scientific and logical aspect of the analysis while maintaining its in situ characteristics. Regarding supply, previous studies have used a 1:2 ratio to estimate hospitals' overall service capacity by combining the number of beds and doctors (J. Luo et al., 2018). Figure 1 displays the spatial distribution of 83 hospitals in the study area, together with the depiction of their comprehensive score indicators. The population of the Shenzhen grid, measuring 1000m x 1000m, was utilized as the scale for demand assessment.

Hence, this study introduces an enhanced 2SFCA technique that relies on the fitted decay function. The cab trajectories to tertiary and secondary hospitals are defined as drop-off locations within 100 m of the hospitals' centers of mass. A total of 46,000 hospital visits are recorded, with an average of 554 visits per hospital.

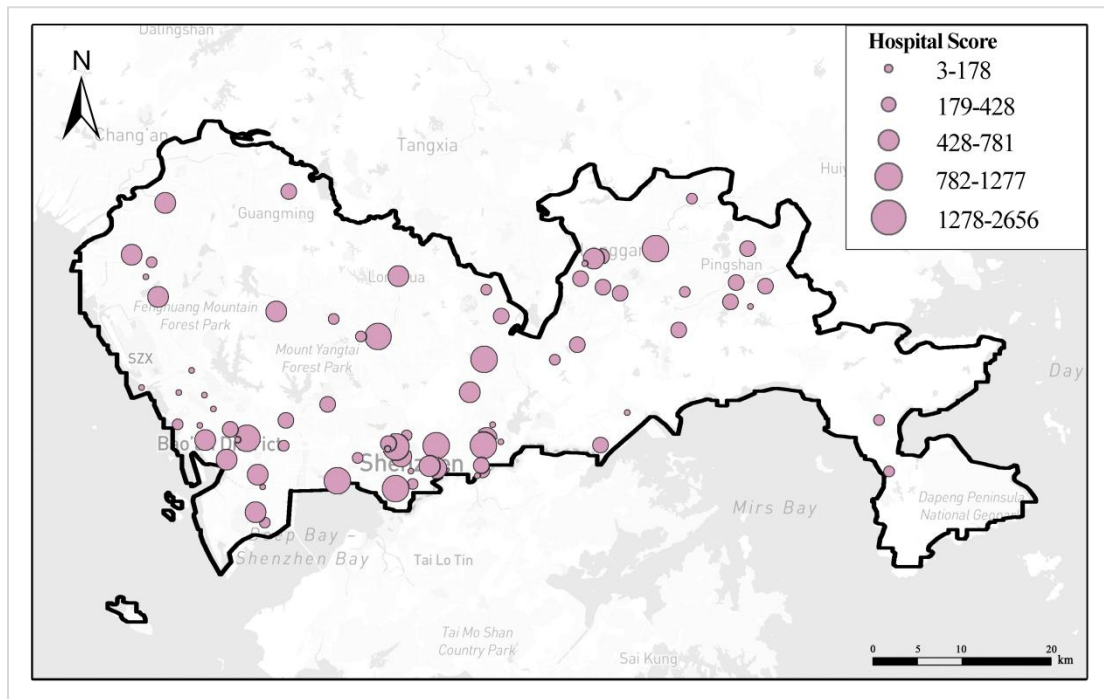


Figure 1. Spatial distribution of hospitals in Shenzhen and their composite score.

Using the frequency distribution of trips at different intervals (e.g., Figure 2a), the data before the peak was approximated, assuming that the hospital's appeal would remain constant as the distance before the peak rose. This is due to the hospital's proximity, which encourages more pedestrian traffic and reduces the need for car transportation. Subsequently, the usage frequency was standardized to reflect the decay weights, as illustrated in Figure 2b.

Ultimately, a Gaussian function is applied to the scatter plot following the peak to establish the decay function representing the decrease in the number of Shenzhen residents going to the hospital.

Figure 2c displays both the fitted function and the actual scenario. The color orange represents the fitted value of the function. The explicit equation for the attenuation function $f(t)$ is presented below, with t denoting the time taken for travel:

$$f(t) = \begin{cases} 1 & t \leq 11 \\ 2.609 * e^{-0.105t} + 0.001 & t > 11 \end{cases} \quad (5)$$

To calculate travel times, we used the route-planning API from Gaode Maps, Alibaba's web-based mapping service. We selected the morning peak hour (7 a.m. to 9 a.m.) and the off-peak hour (0 a.m. to 2 a.m.) of a typical weekday in June 2023 as our time frames. Additionally, we used map API to retrieve traffic times from the following link: <https://lbs.amap.com/api/uri-api/guide/travel/route>. We collected journey times for a total of 222,191 routes from 2,677 grids to 83 hospitals for both periods. This data will be used in the decay function $f(t)$ to calculate the healthcare accessibility for each grid cell. It is important to note that if the starting point is far from the road, the navigation system will initially direct the user towards the road. As a result, accessibility in larger enclosed areas, such as parks, will be considerably reduced compared to that in surrounding areas.

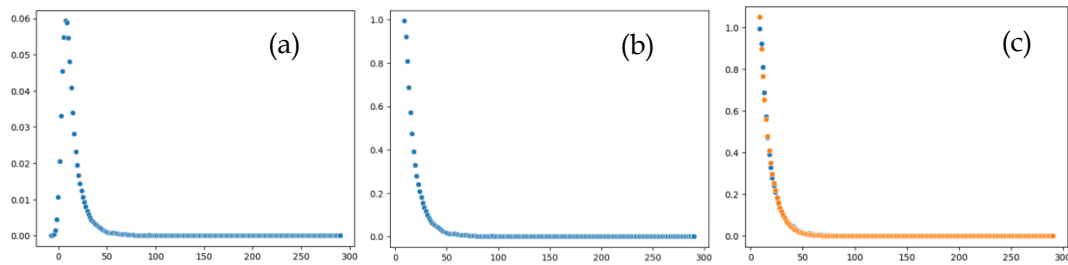


Figure 2. The process of defining the decay function: (a) frequency distribution of trips of different durations; (b) removal of the pre-peak portion and transformation of the data; (c) curve fitting to obtain the parameters of the decay function.

Accessibility analysis: Getis-Ord Gi and geographically weighted random forest

To assess disparities in the availability of healthcare resources in Shenzhen, we employed the Getis-Ord Gi statistic to find significant clusters of high and low accessibility. Furthermore, to investigate the factors that affect the level of healthcare services and their connection to socioeconomic progress, we perform zonal statistics on each grid cell to obtain the mean non-peak healthcare accessibility across different streets in Shenzhen. Subsequently, we examine the correlation between accessibility and socioeconomic characteristics through geographically weighted random forest regression (GWRF).

Geographically weighted random forest regression (GWRF) enhances the modeling of spatially correlated data by integrating spatial autocorrelation and the nonlinear association between features and target variables (Fayaz & Galasso, 2024). It incorporates spatial location information into the standard random forest model and constructs a separate model for each sample location. Building a local model that accounts for spatial autocorrelation offers certain advantages for addressing spatial challenges (Y. Luo et al., 2022; Santos et al., 2019). We selected GWRF for two critical reasons tied to Shenzhen's urban complexity. First, healthcare accessibility is driven by socioeconomic factors that likely exhibit nonlinear dynamics—for instance, road network density may show threshold effects (e.g., accessibility improves markedly only after a minimum density is reached) or synergistic interactions with aging populations. GWRF captures such intricate, non-additive relationships, whereas GWR assumes linearity and would obscure these patterns. Second, spatial heterogeneity is a core focus of this study, as resources vary drastically between central and peripheral districts; GWRF directly addresses this by embedding local spatial weights into each tree, mitigating residual autocorrelation that global RF ignores. This hybrid approach optimally balances spatial granularity and predictive power for Shenzhen's heterogeneous urban fabric (D.

Wu et al., 2024). Crucially, GWRF outputs variable importance metrics (e.g., the Gini index) rather than fixed coefficients, thereby revealing the spatial intensity of drivers such as road density or aging. This allows us to quantify where factors exert the strongest influence, highlighting spatially divergent hotspots without oversimplifying to a single directional effect. GWRF has the same fundamental principles as GWR, using the most basic form of a regression equation as an illustrative example (Equation (6)).

$$Y_i = ax_i + e \quad (6)$$

Where Y_i is the value of the dependent variable at the i th observation, x_i is the i th sample point, a is the nonlinear predicted value of x_i based on the RF, and e is the model error term. This equation is usually constructed from all the data and does not account for the spatial distribution of the samples. However, in GWRF, equation (6) is extended as:

$$Y_i = a(u_i, v_i)x_i + e \quad (7)$$

Where $a(u_i, v_i)x_i$ is the local RF prediction corrected for x_i by spatial weighting at location i and (u, v) are the coordinates. A sub-model is built at each sample point location, and only nearby observations are considered (Georganos et al., 2021). The region where the submodel operates is called the kernel, and the radius of the kernel is called the bandwidth. The kernel usually consists of two types: "adaptive," defined by the n nearest sample points, and "fixed," defined by a circle with a radius of the bandwidth. The adaptive kernel is used for the computation due to the uneven distribution of the sample points. For the implementation of GWRF, we used the R package SpatialML (Kalogirou & Georganos, 2022), which builds on the geographical random forest framework of Georganos et al. (2021). The dataset contains 73 observations and six independent variables, with 21 nearest-neighbor observations considered for each local model. The model was a regression, and a total of 500 decision trees were generated, with a randomly selected number of variables (m try) of 2 per split, a target node size of 5, a variable importance measure of Gini impurity, and a splitting rule of variance.

3 Results

3.1 Accessibility analysis

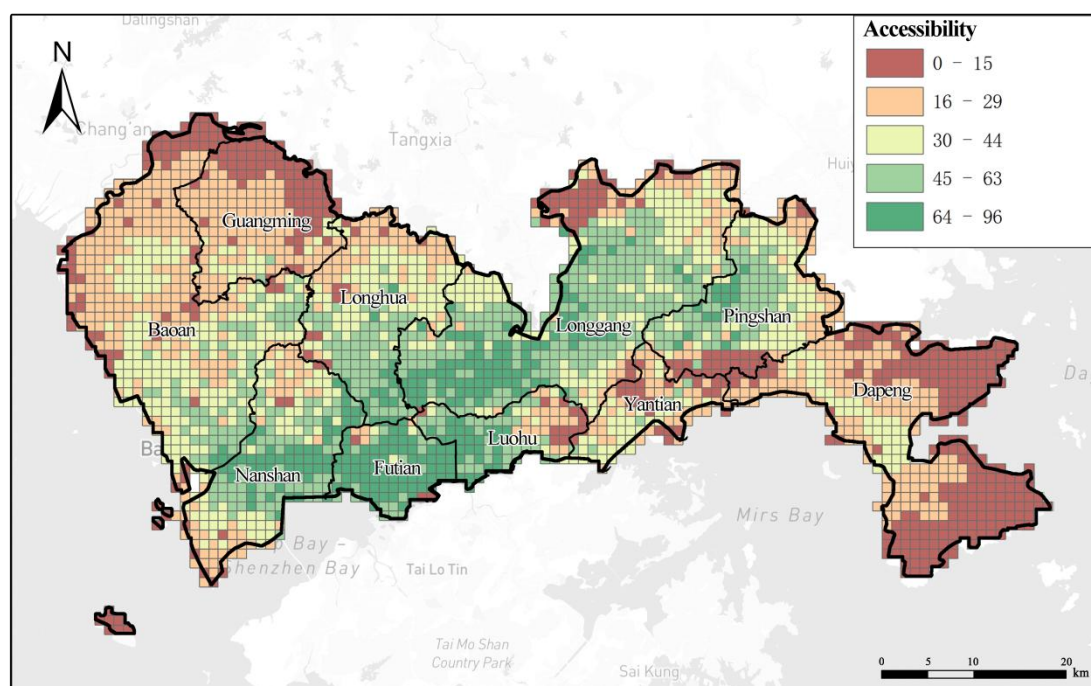
The enhanced mobility-two-step movement search approach was employed to compute the accessibility of healthcare resources in each grid in Shenzhen using the driving journey mode. The regional distribution of healthcare resource accessibility exhibits clear spatial variation, as depicted in Figures 3a and 3b. The accessibility values of Nanshan, Futian, Luohu, and Longgang in the southwest and central parts of the city are consistently higher than those of other locations, regardless of peak period. Conversely, the areas with lower accessibility values are primarily concentrated in the southeast and northwest. Areas with high value, such as Futian, Luohu, Pingshan, and Longgang, have excellent access to healthcare resources since a significant number of high-quality hospitals surround them. Low-value locations are farther from areas where high-quality hospitals are concentrated.

Additionally, the infrastructure, such as roads and public service facilities, could be of a better standard in these areas. As a result, residents in these areas have greater restrictions in accessing healthcare treatment due to the distance they have to travel. Consequently, the level of healthcare accessibility in these areas is the poorest.

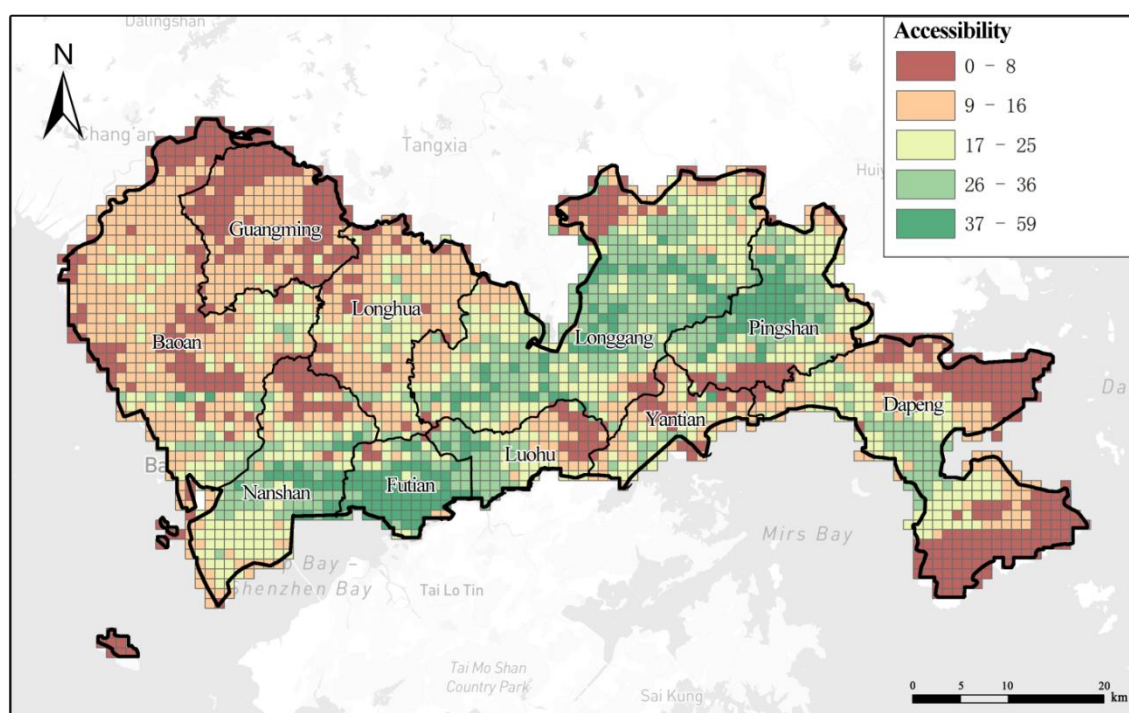
We examined the peak and off-peak periods individually. Based on Shenzhen's traffic management regulations, this study defines peak hours as 7:00 to 9:00 in the morning and 5:30 to 7:30 p.m., which align with the city's traffic restrictions. Healthcare resources are less accessible during peak periods than during off-peak hours.

During non-peak hours (as shown in Figure 3a), when the impact of big green spaces is excluded, Shenzhen exhibits superior accessibility to healthcare resources. When considering the size of the subdistricts, the streets Xiangmihu, Lianhua, and Huaifu in Futian and Luohu districts have greater accessibility. The streets Longcheng, Baolong, Pingshan, Billing, and Nanwan in Longgang district also have higher accessibility. In the Nanshan district, only the town center has higher healthcare accessibility. In the southern part of Longhua district, subdistricts such as Bantian and Jihua have higher accessibility due to a higher concentration of healthcare resources.

During the peak periods (as shown in Figure 3b), the area with limited accessibility grows from the outer edges towards the center due to traffic congestion. This results in forming a band of places with high accessibility running from the southwest to the northeast in the center. When comparing off-peak hours to other times, getting healthcare resources in low-value communities in the southeast and northwest is more challenging. Pingshan, Longgang, and Dapeng New Area in the eastern region have superior accessibility compared to other locations, primarily due to their excellent transportation infrastructure. Shiyan, Hangcheng, and Xixiang subdistricts in Baoan District, as well as Minzhi and Bantian subdistricts in Longhua District and Gongming and Martian subdistricts in Guangming District, have considerably reduced accessibility values during off-peak hours compared to their normal subdistrict scale.



(a)



(b)
Figure 3. Accessibility of healthcare resources by grid in Shenzhen (a. Off-peak; b. peak).

3.2 Hot spot analysis

During the off-peak period (0:00-2:00), the availability of healthcare resources in Shenzhen shows distinct spatial distribution patterns. The primary areas of high activity are focused on the central part of the city, specifically Nanshan District, Futian District, and Luohu District. Additionally, these hotspots extend to Longhua District and Longgang District in the northeastern part of the city. During this period, certain subdistricts like Xiangmihu, Futian, Shahe, Shatou, Zhaogang, and Yuanling gained significant attention as popular areas. In contrast, subdistricts like Minzhi, Bantian, Jihua, and Nanwan were considered less prominent peripheral areas (Figure 4a).

It is important to note that there are gaps between hotspots, indicating a discontinuous distribution due to the presence of sizable green spaces such as Tonglangshan Park and Yinhushan Park. Cold spots are primarily located on the outskirts of the southeastern and northwestern areas of the city, specifically in the Gongming, Longgang, Maluan, Dapeng, and Nan'ao subdistricts (Figure 4a).

During the peak period (7:00-9:00), the distribution of hotspots shows noticeable clustering patterns. The hotspots are primarily concentrated in the central areas, specifically in Nanshan District, Futian District, and Luohu District. Additionally, there is a notable concentration of hotspots in Pingshan District and Longgang District. Futian District and Pingshan District are the primary locations for accessibility hotspots. More precisely, Xiangmihu, Shatou, Shahe, Futian, and Yuanling subdistricts in the city center, as well as Pingshan and Longtian subdistricts in Pingshan District, are notable for being easily accessible areas. In contrast, the subdistricts in Longcheng, Maluan, and Yuanshan exhibit minor hotspots, as seen in Figure 4b.

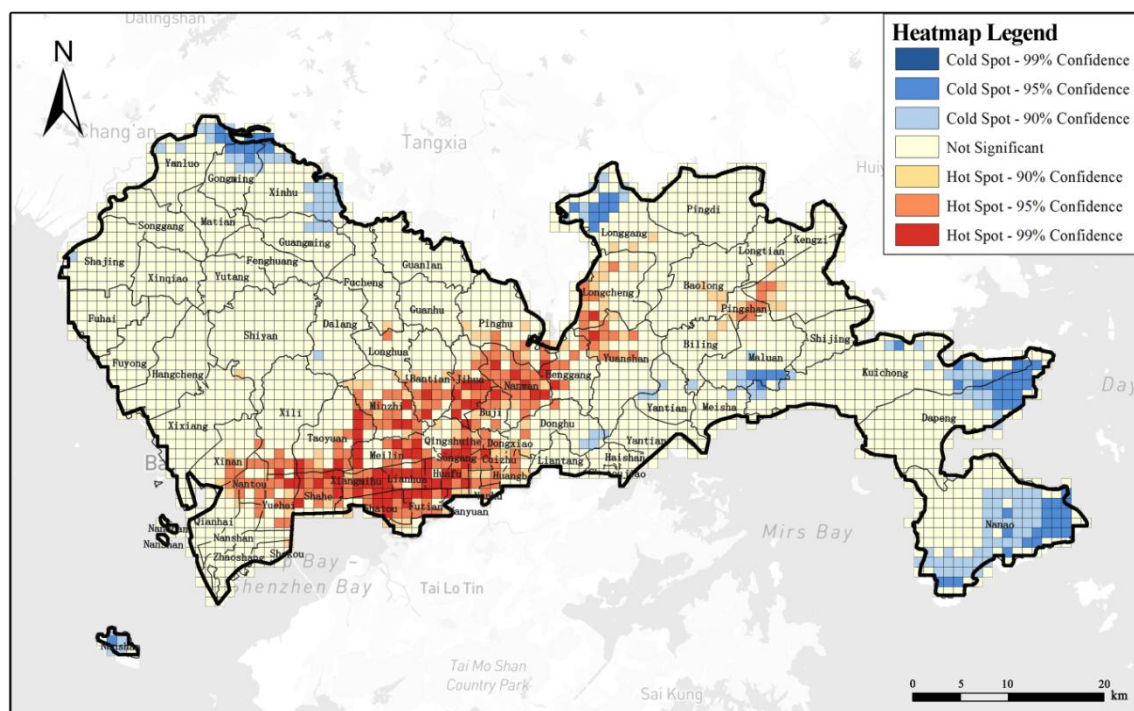
There is a notable shift in the distribution of hotspots in peak periods compared to the off-peak period. During the peak period, the hotspot areas in the city center decreased significantly, while a new hotspot area with improved accessibility emerged in the Pingshan and Longgang districts. There is a hypothesis that the shrinking of the hotspot in the city center is caused by increased road congestion during peak hours, resulting in reduced accessibility under similar conditions. On the

other hand, the emergence of new hotspots in Pingshan and Longgang districts during the busiest time can be explained by their greater distance from the city center and better traffic conditions at that time (Figure 4b). Additionally, these districts have a higher concentration of healthcare resources compared to other districts outside the city center (Figure 3). As a result, they have become a new focal point for accessibility during the peak period.

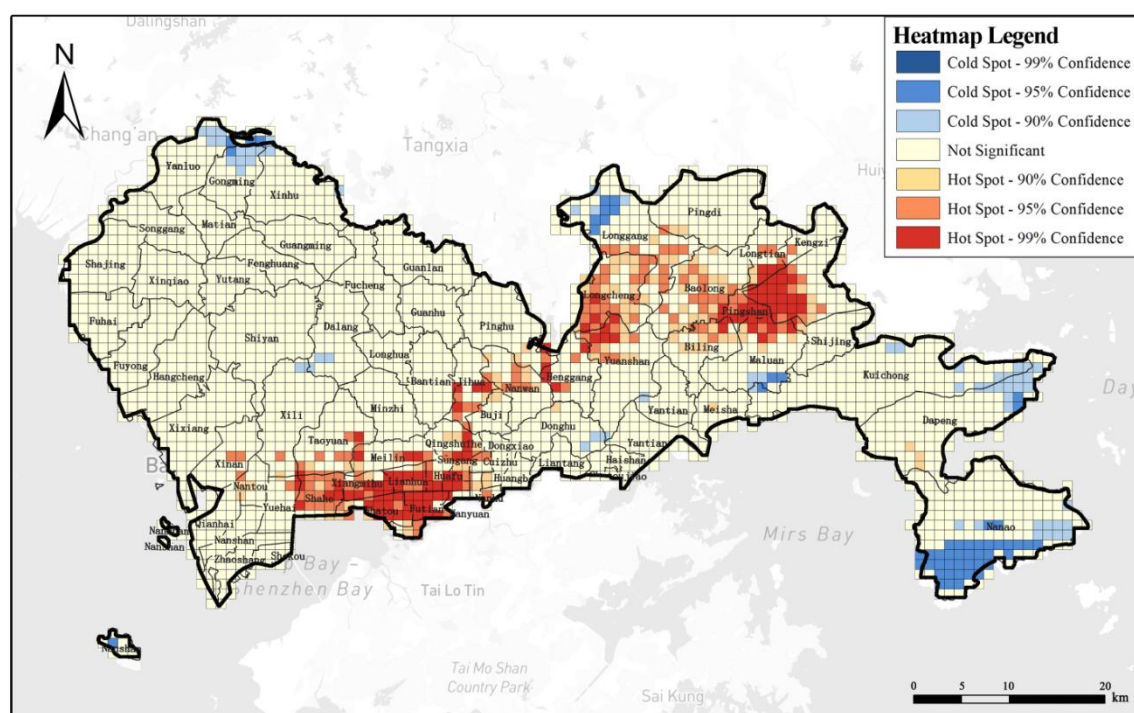
Concerning the distribution of cold spots, there was a general decline in their significance during peak time compared to the off-peak period. However, the relevance of the largest cold spot, located in SA, increased (Figure 4b).

To summarize, the city center is the most prominent area of accessibility hotspots during peak and off-peak periods. However, the range of hotspots is broader during off-peak periods than during peak periods. The hotspot located between Pingshan District and Longgang District is only observable during peak periods.

Regarding the areas with limited accessibility, the distribution of cold spots during both periods is similar but more pronounced during off-peak periods. The cold spots primarily occur in the periphery of the southeastern and northwestern corners of the research region and exhibit a fragmented distribution. It is crucial to acknowledge that these cold spots are typically located in expansive enclosed spaces, such as verdant parks. Due to vehicle prohibitions in these locations, the guidance system typically instructs users to walk to the exit on the nearby road. This approach is time-consuming and greatly hampers the accessibility of healthcare resources in these places, resulting in a much lower degree of access compared to neighboring areas.



(a)



(b)

Figure 4. Spatial distribution of cold hotspots of healthcare resources accessibility during off-peak (a) and peak (b) periods in Shenzhen.

In general, the availability of healthcare resources during the peak period indicates a gradual shift from the city center to the outskirts, moving from areas with high demand to areas with low demand. This reflects the concentration of healthcare resources in the city center of Shenzhen, while the allocation of healthcare resources in the peripheral regions of the city needs to be revised.

3.3 Importance analysis based on geographically weighted random forest regression

In this study, we used geographically weighted random forest regression (GWRf) to explore the spatial heterogeneity in the relationship between various socioeconomic factors and the accessibility of healthcare resources. We initially considered peak and non-peak periods to understand the driving mechanisms affecting healthcare accessibility. However, after careful consideration, we decided to retain the non-peak results as the primary focus of our analysis. This decision was made because non-peak periods account for most of the time when residents seek healthcare services and thus provide a more comprehensive view of accessibility under typical conditions.

The rationale for emphasizing non-peak times is based on the observation that healthcare accessibility during non-peak hours is more stable and reflective of the general situation experienced by residents throughout most of the day. Peak periods, while important, are characterized by higher traffic congestion, which introduces additional variability in travel times and may not accurately represent the underlying spatial distribution of healthcare resources.

To ensure the robustness of our analysis, we first assessed multicollinearity among the predictor variables using the Variance Inflation Factor (VIF). Consider multicollinearity among various influencing factors and remove variables of monthly rental expenses whose variance inflation factor (VIF) exceeds 7.5. After this analysis, five key factors were retained: Percentage of Housing Built After 2000 (VIF = 3.06), Percentage of Population with a College Degree or Above (VIF = 3.57), Percentage of Aged 60 and Above (VIF = 7.27), Road Density (VIF = 3.73), and Total Value of Family

Cars (VIF=4.67). The GWRF model was selected for this study as it effectively captures the complex, potentially non-linear relationships between socioeconomic factors and healthcare accessibility without requiring pre-specified functional forms. After optimizing the parameters, the optimal model $R^2=0.720$ was obtained, $AIC=-1002.12$, $AICc=-1000.847$. To validate the scientific reliability of our results, we reviewed relevant literature and found that the R-squared values for Random Forest and Geographically Weighted Regression (GWR) models typically range between 0.6 and 0.9. For instance, Yiwen Hu et al. (2024) reported R-squared values between 0.719 and 0.804 for Random Forest models, and between 0.413 and 0.681 for GWR models. Similarly, Tsion Mulat Tebeje et al. (2024) obtained an R-squared value of 0.426 in their GWR study (Tebeje et al., 2024). Moreover, Khan et al. (2022) obtained up to 0.849 with GWRF in corn yield prediction, and Hu and Liang (2024) reported adjusted R^2 of 70-80% in analyzing road network - public facility diversity in Changsha. These findings suggest that the R^2 value obtained in this study is reasonable and credible. These findings suggest that the R-squared value we obtained is reasonable and credible. Therefore, our Random Forest Geographically Weighted Regression model demonstrates high predictive accuracy, and the socioeconomic factors we selected can partially explain the accessibility of healthcare resources. The spatial distribution of each factor's importance is mapped in Figures 5a-e. To aid interpretation, the color gradient in each map (from blue to red) represents the spectrum of local variable importance, where red areas indicate subdistricts where the factor is a stronger predictor of healthcare accessibility, and blue areas indicate where it is a weaker predictor. The following sections detail the patterns for each dimension.

Economic factor

The relative importance of the economic factor, represented by Total Household Vehicle Value, is mapped in Figure 5 based on the GWRF model output. The Gini Index values shown represent the local variable importance, which quantifies the relative contribution of this factor to explaining variations in healthcare accessibility across subdistricts. A higher index value indicates that household vehicle value is a more influential predictor of accessibility in that specific location.

As illustrated in Figure 5a, the spatial distribution of this importance does not follow a simple, clear gradient. Instead, it presents a mosaic pattern. Subdistricts with moderate to high importance values (represented by warmer colors and the index values approximately above 4×10^{-6} to 5×10^{-6}) are interspersed with areas of lower importance (cooler colors) across multiple districts. For instance, pockets of relatively higher importance can be observed in parts of the southern Nanshan, Longgang, Yantian, and Luohu. Conversely, larger contiguous areas showing lower relative importance (e.g., index values predominantly in the 0 to 3×10^{-6} range) are evident in Bao'an District, Guangming New District, and the western sectors of Futian and Luohu.

This heterogeneous pattern is a key finding enabled by the GWRF approach. It demonstrates that the role of economic capacity (as proxied by vehicle value) is not uniform but highly context-dependent. The model does not identify a universal threshold of influence. This nuanced spatial variation suggests that policies aimed at improving accessibility through economic channels must be geographically targeted. The GWRF output effectively pinpoints specific subdistricts in which residents' economic capacity, as proxied by vehicle ownership, exhibits a stronger association with healthcare accessibility. This suggests that in these areas, transportation policies aimed at improving private vehicle mobility (e.g., optimizing road networks, mitigating congestion) could potentially yield significant co-benefits for healthcare access. This provides a refined, spatially targeted tool for policy prioritization.

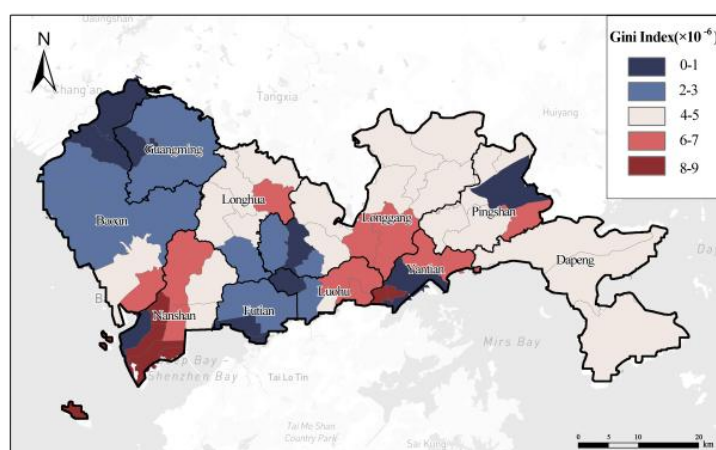


Figure 5. GWRF importance indicators for off-peak economic factors: Total Household Vehicle Value

Social Factor

The spatial heterogeneity in the importance of social factors is presented in Figure 6a (Percentage of Aged 60 and Above) and Figure 6b (Percentage of Population with a College Degree or Above). The maps illustrate the importance of each local variable in each factor.

The importance of the elderly population proportion exhibits a dispersed pattern. Areas with the highest relative importance (Gini Index values of 8×10^{-6} – 10×10^{-6} , represented by the darkest red) are concentrated in specific subdistricts, primarily found in the northern part of Dapeng New District, the eastern part of Luohu District, the southern part of Nanshan District, and the eastern part of Bao'an District. This pattern suggests that in these areas, the age demographic serves as a stronger predictor of healthcare accessibility, potentially reflecting a higher reliance on healthcare services by elderly residents or localized resource allocation patterns tailored to aging communities. Conversely, large contiguous areas, including most of Futian District, the northern part of Bao'an District, and the eastern part of Pingshan District, show low to very low importance values (Gini Index of 0 – 3×10^{-6} , in blue tones). This indicates that the issue of aging in these districts is less significant, or that these districts already have sufficient healthcare resources, resulting in less pressure on healthcare service demand due to aging.

The importance of the population with a college degree or above shows a distinct spatial concentration. The highest importance values (Gini Index of 8×10^{-6} – 10×10^{-6}) are predominantly clustered in the southern part of Nanshan District and the northern part of Longhua District, which suggests that individuals with higher education may prioritize health and the quality of healthcare services. Consequently, the allocation of healthcare resources in these districts is likely more rational and capable of meeting the needs of the highly educated population. Most other regions of the city display low to moderate importance values (0 – 5×10^{-6}). This suggests that the level of higher education is a key differentiating factor for healthcare accessibility in these specific, highly urbanized zones.

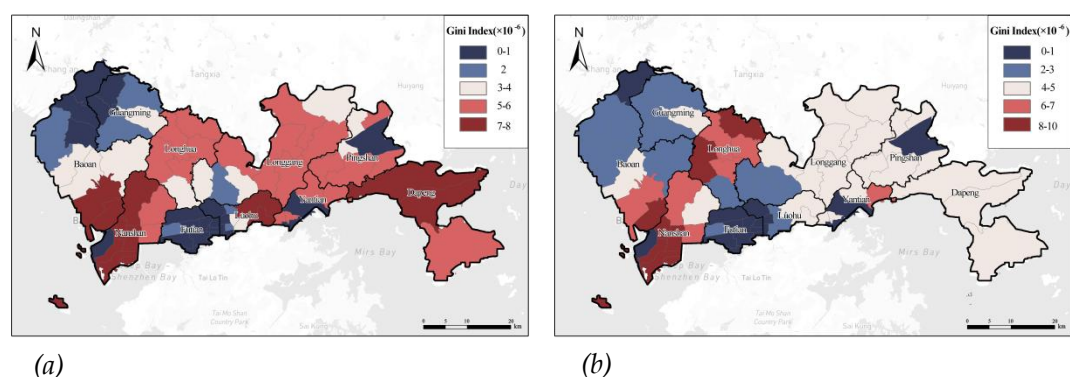


Figure 6. GWR importance indicators for off-peak social factors: (a) Elderly Population; (b) College-Educated Population;

Physical environment factor

The analysis of physical environmental factors using the GWRF model reveals not only spatial patterns of variable importance but also provides insights into the underlying urban development dynamics affecting healthcare accessibility. For road network density, the importance mapping shows particularly high values (9×10^{-6} – 11×10^{-6}) in northern Longhua District and eastern Luohu District, indicating that transportation infrastructure serves as a critical constraint on healthcare accessibility in these rapidly developing areas. The concentration of high importance values in these locations suggests that transportation bottlenecks significantly limit healthcare access, where even minor improvements in road connectivity could yield substantial accessibility gains. The post-2000 housing stock importance presents a complementary pattern, with peak values (8×10^{-6} – 10×10^{-6}) concentrated in the southern Nanshan District and eastern Pingshan District. This distribution reflects the planned nature of newer residential developments, where healthcare infrastructure appears to have been integrated into neighborhood planning, resulting in better alignment between housing stock and healthcare accessibility.

When considering all physical environment indicators, Futian District demonstrates consistently low importance values for both factors (0 – 3×10^{-6}), suggesting that physical environment characteristics have limited power in explaining healthcare accessibility variations in this core urban area. This pattern indicates that healthcare resources in Futian District have reached a saturation point, where the dense concentration of medical facilities diminishes the marginal effect of further physical environment improvements. The district's comprehensive healthcare infrastructure appears to have established a baseline level of accessibility that overwhelms the influence of localized physical environmental factors.

Conversely, the northern part of Bao'an District shows moderate importance values but overall poor healthcare accessibility outcomes. This discrepancy suggests that while physical environmental factors do contribute to accessibility challenges in this region, the fundamental constraint lies in the inadequate provision of healthcare facilities. The lag in healthcare infrastructure development relative to population growth creates a situation where even improvements in road networks or housing conditions cannot compensate for the sheer shortage of medical resources.

Nanshan District presents the most balanced profile, with strong alignment between physical environment factors and healthcare accessibility. The district benefits from well-established transportation networks, modern community planning that integrates healthcare facilities from the early development stages, and demographic characteristics that support efficient healthcare utilization. This coordination between physical infrastructure and healthcare resources creates conditions in which accessibility is optimized and investments in the physical environment yield consistent returns.

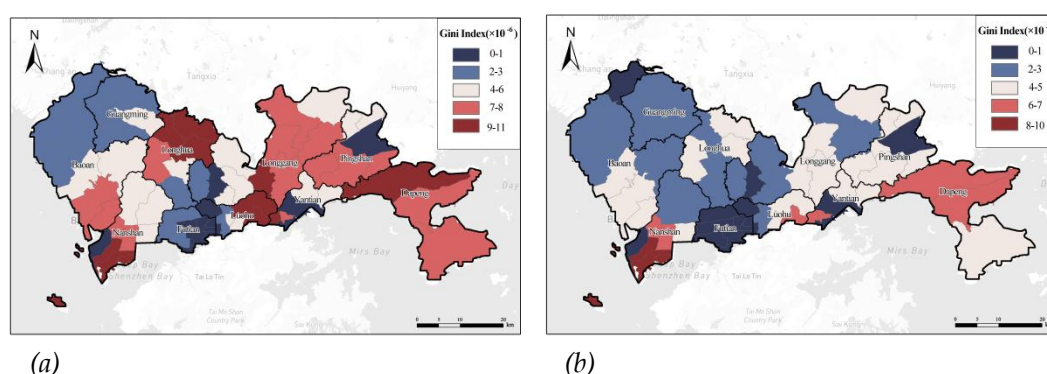


Figure 7. GWRF importance indicators for off-peak socioeconomic factors: (a) road network density; (b) Post-2000 Housing Stock.

Conclusion

When examining the comprehensive results from all indicators through the GWRF framework, distinct spatial patterns of variable importance emerge across Shenzhen's districts. The model reveals substantially lower variable importance across all factors in both Futian District and the northern part of Bao'an District, indicating that these predictors have limited explanatory power for healthcare accessibility variations in these areas.

In Futian District, this pattern suggests that healthcare resources have likely reached a saturation point, where the dense concentration of medical facilities diminishes the marginal effect of additional socioeconomic or infrastructural improvements. The consistently low importance values across all variables indicate that accessibility levels in this developed urban core are relatively stable and less influenced by the specific factors examined in our model.

Conversely, the northern Bao'an District presents a different scenario. While also demonstrating lower overall variable importance, this region simultaneously exhibits poor healthcare accessibility outcomes. The northern part of Bao'an District has poorer healthcare resources, with inadequate supporting facilities and infrastructure development (such as transportation, housing, and public service facilities). This lag in development relative to population growth and changing demands results in poor overall healthcare accessibility.

Nanshan District presents a contrasting profile, with higher relative importance values across multiple indicators. This pattern demonstrates that the selected variables have stronger predictive power in this region, where healthcare accessibility levels show clearer relationships with local economic conditions, social factors, and physical infrastructure. The alignment between variable importance and accessibility outcomes suggests a more balanced urban development approach, where healthcare resources are better integrated with transportation networks, community planning, and demographic characteristics.

The GWRF methodology effectively captures the spatial heterogeneity in how different factors influence healthcare accessibility across the urban landscape. The regional variability in variable importance, as quantified by GWRF, provides valuable insights for evidence-based decision-making. By identifying where specific factors have the greatest impact on accessibility, policymakers can prioritize interventions that address the most binding constraints in each subdistrict, ultimately leading to more efficient and equitable healthcare resource distribution throughout Shenzhen.

4 Discussion

4.1 Policy recommendations

Following over four decades of rapid progress, Shenzhen's healthcare infrastructure has largely stabilized, and the city's core region has established a relatively comprehensive healthcare network. However, due to traditional economic, population, and other criteria for healthcare facility placement, there is a high rate of functional overlap and low utilization of healthcare resources in existing facilities. Additionally, there is a need for more healthcare facilities in areas outside the core region, resulting in an uneven spatial distribution. Therefore, there is an urgent need to optimize the layout of healthcare facilities. The GWRF importance analysis reveals a complex tapestry of spatially heterogeneous factors influencing healthcare accessibility across Shenzhen. This necessitates a move beyond one-size-fits-all policy solutions towards a sophisticated, geographically differentiated strategy that acknowledges the unique constraints and opportunities of each area. Policy must be informed by whether a district's primary constraint is resource saturation (e.g., Futian), fundamental resource scarcity (e.g., Northern Bao'an), or mediating infrastructure and socio-economic factors (e.g., Nanshan, Longhua).

Implementing a differentiated facility strategy based on local constraints

The analysis demonstrates that the drivers of accessibility vary dramatically across the city. Therefore, optimizing healthcare facilities requires a dual approach tailored to these specific local contexts. In Core Saturated Districts (e.g., Futian, Central Luohu), the consistently low importance values for all factors indicate that accessibility is less sensitive to further incremental improvements in infrastructure or socioeconomics due to existing resource density. Here, the policy focus must shift from expansion to optimization. The goal is to enhance the efficiency of the existing high-capacity network without exacerbating congestion.

Enhancing the design of healthcare facilities

When it comes to optimizing healthcare facilities, there are generally two approaches: constructing new facilities and restructuring the current geographical distribution system. Regarding investment, the previous investment surpasses the expenses associated with transferring healthcare facilities (Liu X., 2017). Hence, when the government optimizes the arrangement of healthcare facilities, it should prioritize restructuring existing facilities and aim to achieve comprehensive coverage of top-tier healthcare security services while minimizing costs.

In the central districts, such as Futian and Luohu, where the GWRF model indicates resource saturation and low variable importance, the policy objective shifts from physical expansion to the sophisticated optimization of the existing high-capacity network. This involves promoting advanced integration through healthcare consortia to create seamless referral systems between tertiary hospitals and community health centers, thereby reducing duplication and improving coordination. Furthermore, significant investment in "Internet + Healthcare" and telemedicine solutions is crucial to managing non-essential consultations remotely, reducing physical travel to overcrowded central hospitals, and freeing up capacity for critical cases. Conversely, in peripheral districts such as northern Bao'an and Dapeng, where the fundamental constraint is resource scarcity, the policy must prioritize strategic, incremental investment. This entails establishing new Tier 2 and Tier 3 medical centers to create anchor institutions that drastically reduce long-distance travel to the core for essential services. The GWRF importance maps provide critical guidance for these investments; for instance, new facilities in subdistricts where road network density is a high-importance factor should be paired with concurrent road upgrades to maximize the accessibility return on investment.

Improvement of the transportation road network

The application of random forest geographically weighted regression reveals that road network density has a notable influence on the accessibility of healthcare services. Enhancing transportation infrastructure is crucial for enhancing healthcare accessibility. Healthcare accessibility in the area north of Xixiang-Minzhi-Pinghu exhibits significant variation across different periods. The primary issue hindering accessibility in these areas is the traffic conditions, not the scale and layout of healthcare resources. Enhancing the efficiency of road networks can significantly enhance healthcare accessibility. Consequently, transportation infrastructure budgets and projects should be prioritized for these high-importance zones, where improvements in road connectivity, such as adding key arteries or optimizing traffic flow, will yield the most substantial and immediate gains in healthcare accessibility.

To achieve this goal, Shenzhen should focus on improving the coordination between transportation planning and healthcare service system planning. This can be done by optimizing the current road network and enhancing road construction in Longhua, Bao'an, and Guangming districts. Additionally, efforts should be made to improve the connection between Shenzhen hospitals and densely populated urban areas, ensuring people have quick access to healthcare when needed.

This targeted approach ensures fiscal efficiency and maximizes impact. This focus on strategic road enhancements should be complemented by a continued commitment to multimodal transport integration around all hospitals, including improvements to pedestrian access, cycling infrastructure, and connectivity between public transit hubs and healthcare facilities, ensuring benefits for the non-driving population. Optimizing last-mile access around major hospitals through improved pedestrian infrastructure, dedicated pick-up/drop-off zones, and better public transit connectivity to mitigate local congestion.

Improving fairness in healthcare

The GWRF results indicate that healthcare accessibility in Shenzhen's old urban regions and low-education districts is currently inadequate. To promote healthcare fairness, planning departments should refrain from excessively concentrating healthcare resources in newly developed urban areas or in high-concentration areas of highly educated individuals. Instead, they should focus on improving healthcare services in older urban areas, expanding the number of healthcare facilities in economically and socially disadvantaged areas, and ensuring equitable access to high-quality, comprehensive healthcare resources for all population groups. This will help to enhance the fairness of public healthcare resources in the city. Government agencies should promote the implementation of a resource-sharing mechanism among healthcare institutions. This mechanism would involve transferring resources from densely populated areas to resource-poor locations where disadvantaged people are concentrated. In addition, medical professionals should be rationally allocated according to the population needs of a specific region, such as adding pediatricians in areas with more children and gynecologists in areas with more women. Rather than vaguely targeting "old urban regions," policies can now be directed at specific subdistricts where the importance of the elderly population factor is calculated to be high. In these locations, ensuring that facilities are equipped with geriatric care units and are highly accessible by public transport becomes a primary equity objective. The resource-sharing mechanisms, such as mobile medical clinics or specialist rotation programs, can be strategically routed from well-served core areas to these precisely identified underserved subdistricts. This ensures that support is directed based on measurable need rather than general assumption, making the transfer of resources from "densely inhabited areas to resource-poor locations" a more effective and evidence-based process. Similarly, the allocation of medical professionals can be informed by local demographic data and the GWRF results, ensuring that staffing meets the specific needs of each community.

4.2 Limitations and perspectives

This study has made progress in quantifying and examining the availability of healthcare resources in Shenzhen. However, there are still certain deficiencies that need to be addressed:

1. Assessment of hospital service capability. Obtaining reliable and comprehensive hospital-related information takes time and effort. The current study primarily relies on the number of beds and specialized doctors reported by each hospital. However, it needs to consider other important factors such as healthcare equipment, funding levels, and more. This limitation may impact on the overall evaluation of hospital service capacity.
2. Fitting the distance decay function for the single-trip mode. Due to the bus usage pattern in Shenzhen, where only boarding is recorded by card swiping, it is impossible to form a complete Origin-Destination (OD) matrix with IC card data. Additionally, there needs to be more research in the academic community about data mining for walking to the hospital. Thus, this study solely focused on modeling the distance decay function in the context of driving, disregarding the availability of healthcare in other means of transportation. Although these non-driving modes of travel account for a smaller proportion of healthcare visits, they may still affect the accuracy of the accessibility analysis to some extent.
3. Insufficient data coverage. This study's analysis at the street-town level primarily relies on data from the seventh national population census, whereas reliable, detailed 1km grid-level microdata are much scarcer. To some extent, this restricts the study's depth and breadth, as well as the variety of indicators chosen.
4. Balancing efficiency and equity in optimization modeling. While this study has begun to address the critical tension between serving large populations efficiently and ensuring spatial equity, the methodological approach primarily focuses on diagnostic analysis. The current framework identifies disparities and influencing factors, but does not provide a prescriptive optimization model to formally quantify the trade-offs between these two objectives.

To address the mentioned shortcomings, future research should actively explore and utilize additional data sources. It is recommended that cell phone signaling, Baidu navigation big data, and other relevant information be incorporated to construct a more comprehensive multi-travel mode distance attenuation function. This will enhance the accuracy and applicability of the accessibility analysis. Furthermore, obtaining more detailed data on hospitals from the health department, such as patient satisfaction and the level of modernization of healthcare equipment, will enable the creation of a multi-dimensional hospital service capacity evaluation system. This will improve the timeliness of the study and the comprehensiveness of the indicators.

Building on this foundation, future research can take a significant step forward by developing a comprehensive optimization model. Such a model should not only aim to maximize overall spatial accessibility but also formally incorporate constraints such as 'minimizing the population service loss' or 'ensuring that a specific percentage of the population has accessibility above a certain threshold'. This type of multi-objective optimization framework would provide decision-makers with a precise tool for quantitatively weighing the trade-offs between efficiency (serving the maximum number of people) and equity (ensuring a minimum level of access for all), leading to more informed and balanced healthcare resource allocation strategies.

5 Conclusion

This work uses Shenzhen taxi OD data to enhance the conventional two-step mobile search approach for analyzing the accessibility of healthcare resources in Shenzhen during different periods. Additionally, it investigates the factors that influence healthcare availability using GWRF. The primary discoveries are as follows:

1. The availability of healthcare resources in Shenzhen varies significantly by location and time. The presence of spatial heterogeneity is evident in the disparity of healthcare resource accessibility between city center areas like Futian and Luohu, which generally have high accessibility, and the marginal regions like Guangming and Dapeng, which have relatively low accessibility. Temporal heterogeneity is observed in the expansion of low-value areas from peripheral regions during off-peak hours to the central part of the city during peak hours. This creates a band of high-value regions in a southwesterly-northeasterly direction, with a significant reduction in the size of the hotspot area in the city center.
2. The findings of the GWR model show that economic, physical environment, and social factors have varying levels of impact on the accessibility of healthcare resources in different places. Enhancing the accessibility of healthcare resources requires a thorough evaluation of multiple issues rather than focusing solely on a single aspect of improvement. Overall, the two indicators, aging degree and road network density, exhibit the broadest spatial range of connection with healthcare accessibility, suggesting a universal relationship.
3. Policy implications for equitable healthcare planning. The results of this study demonstrate that for Shenzhen, the optimal allocation of healthcare resources should not be a simple spatial redistribution but rather a differentiated and synergistic strategy. This approach recognizes the distinct roles of various urban contexts: in the resource-saturated central districts, the policy emphasis should be on enhancing the utilization efficiency of existing resources through managerial and technological innovations, such as promoting healthcare consortia and telemedicine. Conversely, in the resource-scarce peripheral districts, the priority must be addressing the structural deficit through new investments in medical facilities and infrastructure to promote spatial equity. Future urban planning should therefore aim to achieve a Pareto improvement in overall urban healthcare welfare. This would be accomplished by acknowledging the value of high service efficiency in the central core while simultaneously implementing precise, targeted investments in the outer regions to create a more balanced and equitable system.

Additionally, future research should focus on the accessibility of walking to hospitals, particularly considering differences in walking speed and accessibility among vulnerable groups (such as children, the elderly, pregnant women, or people with disabilities). This will help to more comprehensively assess the accessibility of medical resources and ensure that all populations can fairly obtain medical services.

Data Access Statement

Data supporting this publication is available upon request.

Author and Contributor Statement

Yunying Liu: Conceptualization, Methodology, Project Administration, Software, Formal Analysis, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing; Huayang Liu: Software, Formal Analysis, Validation, Writing – Review & Editing; Zhaoqi Xu: Data Curation, Investigation, Visualization, Writing – Review & Editing; Yiming Song: Visualization, Investigation; Yixin Wan: Visualization, Investigation; Xinyue Gu: Supervision, Resources.

Use of AI

During the preparation of this work, the authors used ChatGPT and DeepSeek to assist with code writing and debugging, and to aid in language polishing and formatting consistency during manuscript drafting, respectively. After using these tools, the authors reviewed, edited, made the content their own and validated the outcome as needed, and take full responsibility for the content of the publication. The AI tools were not involved in formulating the research questions, designing the methodology,

interpreting results, or drawing conclusions. All AI-generated content was rigorously reviewed and adapted by the authors.

Conflict Of Interest (COI)

There is no conflict of interest.

References

- Chen, B. Y., Cheng, X.-P., Kwan, M.-P., & Schwanen, T. (2020). Evaluating spatial accessibility to healthcare services under travel time uncertainty: A reliability-based floating catchment area approach. *Journal of Transport Geography*, 87, 102794. <https://doi.org/10.1016/j.jtrangeo.2020.102794>
- Chen, Y., Jia, B., Wu, J., Liu, X., & Luo, T. (2022). Temporal and spatial attractiveness characteristics of Wuhan urban Riverside from the perspective of traveling. *Land*, 11(9), 1434. <https://doi.org/10.3390/land11091434>
- Cheng, M., & Lian, Y. (2018). 基于改进潜能模型的城市医疗设施空间可达性——以上海市杨浦区为例 [Spatial accessibility of urban medical facilities based on improved potential model: A case study of Yangpu District in Shanghai]. *Progress in Geography*, 37(2), 266-275. <https://doi.org/10.18306/dlkxjz.2018.02.009>
- Fayaz, J., & Galasso, C. (2024). Interpretability and spatial efficacy of a deep-learning-based on-site early warning framework using explainable artificial intelligence and geographically weighted random forests. *Geoscience Frontiers*, 15(5), 101839. <https://doi.org/10.1016/j.gsf.2024.101839>
- Gao, J., Zhang, X., Hu, G., & Sun, J. (2022). 信阳市区医疗设施可达性评价及空间优化研究 [The accessibility and spatial optimization of medical facilities in Xinyang urban area]. *Journal of Xinyang Normal University (Natural Science Edition)*, 35(2), 224-229. <https://doi.org/10.3969/j.issn.1003-0972.2022.02.009>
- Gao, Y., Yang, Q., & Li, J. (2023). 城市医疗机构分级可达性与空间公正研究——以西安市为例 [Research on hierarchical accessibility and spatial justice of urban medical institutions: Taking Xi'an as an example]. *Area Research and Development*, 42(1), 68-74. <https://doi.org/10.3969/j.issn.1003-2363.2023.01.012>
- Geng, T., Chen, H., Zhang, H., Shi, Q., & Liu, D. (2020). 基于 GWR 的陕西省生态系统服务价值时空演变特征及影响因素分析 [Spatiotemporal evolution of land ecosystem service value and its influencing factors in Shaanxi province based on GWR]. *Ziran Ziyuan Xuebao (Journal of Natural Resources)*, 35(7), 1714 - 1727. <https://doi.org/10.31497/zrzyxb.20200715>
- Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuyse, S., Mboga, N., Wolff, E., & Kalogirou, S. (2021). Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 36(2), 121-136. <https://doi.org/10.1080/10106049.2019.1595177>
- Georganos, S., & Kalogirou, S. (2022). A forest of forests: A spatially weighted and computationally efficient formulation of geographical random forests. *ISPRS International Journal of Geo-Information*, 11(9), 471. <https://doi.org/10.3390/ijgi11090471>
- Guagliardo, M. F. (2004). Spatial accessibility of primary care: Concepts, methods and challenges. *International Journal of Health Geographics*, 3, 1-13. <https://doi.org/10.1186/1476-072X-3-3>
- Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2), 73-76. <https://doi.org/10.1080/01944365908978307>
- Haynes, R., Jones, A. P., Sauerzapf, V., & Zhao, H. (2006). Validation of travel times to hospital estimated by GIS. *International Journal of Health Geographics*, 5, 1-8. <https://doi.org/10.1186/1476-072X-5-40>

- He, X., Zhou, Y., Yuan, X., & Zhu, M. (2024). The coordination relationship between urban development and urban life satisfaction in Chinese cities-An empirical analysis based on multi-source data. *Cities*, 150, 105016. <https://doi.org/10.1016/j.cities.2024.105016>
- Hu, L., He, S., & Su, S. (2024). 深圳市公共住房居民就医可达性——基于互联网地图服务和改进两步移动搜索法 [Accessibility of healthcare resources to public housing in Shenzhen, China: Indirect map service and optimized two-step floating catchment area method]. *Tropical Geography*, 44, 226–235. <https://doi.org/10.13284/j.cnki.rddl.003824>
- Hu, S., Lu, Y., Hu, G., & Sun, J. (2021). 基于多源大数据的深圳市医疗设施可达性与公平性测算 [Measuring accessibility and equity of medical resources in Shenzhen based on multi-source big data]. *Economic Geography*, 41(11), 87 – 96. <https://doi.org/10.15957/j.cnki.jjdl.2021.11.010>
- Hu, Y., & Liang, C. (2024). Study on the spatial relationship between road network and the diversity of urban public facilities: The case of the central area of Changsha City. *Journal of Engineering and Applied Science*, 71(1), 1–19. <https://doi.org/10.1186/s44147-024-00484-w>
- Jian, I. Y., Chan, E. H., Xu, Y., & Owusu, E. K. (2021). Inclusive public open space for all: Spatial justice with health considerations. *Habitat International*, 118, 102457. <https://doi.org/10.1016/j.habitatint.2021.102457>
- Jiao, W., Huang, W., & Fan, H. (2022). Evaluating spatial accessibility to healthcare services from the lens of emergency hospital visits based on floating car data. *International Journal of Digital Earth*, 15(1), 108–133. <https://doi.org/10.1080/17538947.2021.2014578>
- Khan, S. N., Li, D., & Maimaitijiang, M. (2022). A geographically weighted random forest approach to predict corn yield in the US Corn Belt. *Remote Sensing*, 14(12), 2843. <https://doi.org/10.3390/rs14122843>
- Kyaw, A. K., Hamed, M. M., Kamruzzaman, M., & Shahid, S. (2023). Spatiotemporal changes in population exposure to heat stress in South Asia. *Sustainable Cities and Society*, 93, 104544. <https://doi.org/10.1016/j.scs.2023.104544>
- Li, C., Bu, P., Fang, J., Ma, K., & Chen, A. (2018). 基于改进引力模型的医疗服务可达性评价 [Accessibility evaluation of medical service in Xiangtan City based on improved gravity model]. *Economic Geography*, 38(12), 83 – 88. <https://doi.org/10.15957/j.cnki.jjdl.2018.12.011>
- Li, Y., & Derudder, B. (2022). Dynamics in the polycentric development of Chinese cities, 2001–2016. *Urban Geography*, 43(2), 272–292. <https://doi.org/10.1080/02723638.2020.1847938>
- Li, Y., & Liu, X. (2018). How did urban polycentricity and dispersion affect economic productivity? A case study of 306 Chinese cities. *Landscape and Urban Planning*, 173, 51–59. <https://doi.org/10.1016/j.landurbplan.2018.01.007>
- Liao, Z., & Jiang, H. (2018). Urban medical facilities spatial accessibility based on improved potential model: A case study in Cangshan, Fuzhou. *Journal of Fujian Normal University (Natural Science Edition)* 34: 38–49. <https://opaj.napstic.cn/periodicalArticle/0120180401674314>
- Liu, S., Qin, Y., & Xu, Y. (2019). Inequality and influencing factors of spatial accessibility of medical facilities in rural areas of China: A case study of Henan Province. *International Journal of Environmental Research and Public Health*, 16(10), 1833. <https://doi.org/10.3390/ijerph16101833>
- Liu, X. (2017). 基于可达性的医疗服务均等性评价与优化布局——以武汉市中心城区为例 [Evaluation and optimization of medical service equity based on accessibility: A case study of Wuhan city center] [Master's thesis, Wuhan University]. Wuhan University Open Institutional Repository. https://openir.whu.edu.cn/AchievementDetail/degree_712ac856-855e-48dc-8147-3108da0c1c15
- Liu, Y., Gu, H., Lai, X., & Qiao, X. (2023). 基于城市轨道交通的医疗设施可达性与供需匹配度分析 [Analysis of medical facility accessibility and supply – demand matching based on urban rail

- transit]. *Resource Development & Market*, 39(5), 600 – 608. <https://doi.org/10.3969/j.issn.1005-8141.2023.05.011>
- Liu, Z., Yang, H., Xiong, W., & Chen, G (2017). 基于改进两步移动搜索法的县域医疗卫生服务空间可达性研究 [Spatial accessibilities of medical services at county level based on optimized two-step floating catchment area method]. *Scientia Geographica Sinica*, 37, 728–737. <https://doi.org/10.13249/j.cnki.sgs.2017.05.010>
- Luo, J., Chen, G., Li, C., Xia, B., Sun, X., & Chen, S. (2018). Use of an E2SFCA method to measure and analyse spatial accessibility to medical services for elderly people in Wuhan, China. *International Journal of Environmental Research and Public Health*, 15(7), 1503. <https://doi.org/10.3390/ijerph15071503>
- Luo, W., & Qi, Y. (2009). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, 15(4), 1100–1107. <https://doi.org/10.1016/j.healthplace.2009.06.002>
- Luo, Y., Yan, J., McClure, S. C., & Li, F. (2022). Socioeconomic and environmental factors of poverty in China using geographically weighted random forest regression model. *Environmental Science and Pollution Research*, 1–13. <https://doi.org/10.1007/s11356-021-17513-3>
- Lü, F., Chen, M., & Wei, X. (2022). Analysis of influencing factors of medical facility accessibility based on decision tree model: A case study of Suzhou City. *Shanghai Urban Planning Review*, 5, 40 – 44. <https://doi.org/10.11982/j.supr.20220507>
- Lyu, G., Bertolini, L., & Pfeffer, K. (2020). How does transit-oriented development contribute to station area accessibility? A study in Beijing. *International Journal of Sustainable Transportation*, 14(7), 533–543. <https://doi.org/10.1080/15568318.2019.1578841>
- Pengfei, C., Jian, I. Y., Zhang, L., Siu, K. W. M., Jianxiao, L., Zhewei, L., & Xiaojian, C. (2024). Towards a smart healthy city: A generalised flow-based 2SFCA method for incorporating actual mobility data in healthcare accessibility evaluation. *Travel Behaviour and Society*, 34, 100706. <https://doi.org/10.1016/j.tbs.2023.100706>
- Polzin, P., Borges, J., & Coelho, A. (2014). An extended kernel density two-step floating catchment area method to analyze access to health care. *Environment and Planning B: Planning and Design*, 41(4), 717–735. <https://doi.org/10.1068/b120050p>
- Ren, H., & Wang, Y. (2021). 不同交通方式下居民城际出行网络结构特征——以“春运”为例 [Characteristics of intercity trip network structure of residents under different traffic modes: A case study of Spring Festival travel rush]. *Progress in Geography*, 40(5), 774 – 783. <https://doi.org/10.18306/dlkxjz.2021.05.004>
- Santos, F., Graw, V., & Bonilla, S. (2019). A geographically weighted random forest approach for evaluate forest change drivers in the Northern Ecuadorian Amazon. *PloS One*, 14(12), e0226224. <https://doi.org/10.1371/journal.pone.0226224>
- Shen, Y., & Li, L. (2021). The influence of medical facility accessibility on residents' health seeking behaviors from the perspective of age stratification: A case study of shanghai chongming island. *Human Geography*, 36, 46–54. <https://doi.org/10.13959/j.issn.1003-2398.2021.02.007>
- Song, Z., & Chen, W. (2009). 基于潜能模型的医疗设施空间可达性评价方法 [Measuring spatial accessibility to health care facilities based on potential model]. *Progress in Geography*, 28(6), 848 – 854. <https://doi.org/10.11820/dlkxjz.2009.06.003>
- Song, Z., Chen, W., Che, Q., & Zhang, L. (2010). 基于改进潜能模型的就医空间可达性度量和缺医地区判断——以江苏省如东县为例 [Measurement of spatial accessibility to health care facilities and defining health professional shortage areas based on improved potential model: A case study of Rudong County, Jiangsu Province]. *Scientia Geographica Sinica*, 30(2), 213 – 219. <http://geoscienc.neigae.ac.cn/CN/10.13249/j.cnki.sgs.2010.02.213>

- Su, Y., Ma, Y., Chang, J., Liu, J., & Long, Y. (2023). 急诊急救就医可达性和公平性的研究——以急性心肌梗死就医为例 [Research on accessibility and equity of emergency medical care: A case study of acute myocardial infarction]. *Journal of Human Settlements in West China*, 38(2), 1 - 7. <https://doi.org/10.13791/j.cnki.hsfwest.20230201>
- Tao, Y., & Shen, Y. (2018). 医疗设施可达性空间差异及其影响因素——基于上海市户籍与流动人口的对 比 [Spatial difference of medical facility accessibility and influencing factors: A comparative study of registered and floating populations in Shanghai]. *Progress in Geography*, 37(8), 1075 - 1085. <https://doi.org/10.18306/dlkxjz.2018.08.007>
- Tao, Z., Cheng, Y., & Dai, T. (2014). 北京市养老设施空间可达性评价 [Measuring spatial accessibility to residential care facilities in Beijing]. *Progress in Geography*, 33(5), 616 - 624. <https://doi.org/10.11820/dlkxjz.2014.05.003>
- Tebeje, T. M., Gelaye, K. A., Chekol, Y. M., Tesfie, T. K., Gelaw, N. B., Mare, K. U., & Seifu, B. L. (2024). Geographically weighted regression analysis to assess hotspots of early sexual initiation and associated factors in Ethiopia. *Heliyon*, 10(9). <https://doi.org/10.1016/j.heliyon.2024.e30535>
- Tian, L., Zhang, J., Wang, F., Li, X., Zheng, W., & Luo, J. (2019). 公平与效率导向下农村公共医疗资源的 空间优化研究——以湖北省仙桃市为例 [Rural public medical resources allocation spatial optimization towards equity and efficiency access: A case study on Xiantao City, Hubei Province]. *Scientia Geographica Sinica*, 39(9), 1455 - 1463. <https://doi.org/10.13249/j.cnki.sgs.2019.09.011>
- Tong, D., Sun, Y., & Xie, M. (2021). 基于改进高斯两步移动搜索法的深圳市公园绿地可达性评价 [Evaluation of green space accessibility based on improved Gaussian two-step floating catchment area method: A case study of Shenzhen City, China]. *Progress in Geography*, 40(7), 1113 - 1126. <https://doi.org/10.18306/dlkxjz.2021.07.004>
- Tong, S., & Chen, H. (2017). Evaluation of accessibility of medical facilities based on street scale: A case study in Shenzhen City. *Health Economics Research*, 12, 31 - 34. <https://doi.org/10.14055/j.cnki.33-1056/f.2017.12.009>
- Wang, J., Du, F., Huang, J., & Liu, Y. (2020). Access to hospitals: Potential vs. Observed. *Cities*, 100, 102671. <https://doi.org/10.1016/j.cities.2020.102671>
- Wang, J., & Xu, C. (2017). 地理探测器：原理与展望 [Geodetector: Principle and prospective]. *Acta Geographica Sinica*, 72(1), 116 - 134. <https://doi.org/10.11821/dlxb201701010>
- Xing, J., & Ng, S. T. (2022). Analyzing spatiotemporal accessibility patterns to tertiary healthcare services by integrating total travel cost into an improved E3SFCA method in Changsha, China. *Cities*, 122, 103541. <https://doi.org/10.1016/j.cities.2021.103541>
- Yin, X., Li, P., Feng, Z., Yang, Y., You, Z., & Xiao, C. (2021). Which gridded population data product is better? Evidences from mainland Southeast Asia (MSEA). *ISPRS International Journal of Geo-Information*, 10(10), 681. <https://doi.org/10.3390/ijgi10100681>
- Yu, W., Ai, T., Li, J., Yang, M., & Shuai, Y. (2018). Potential change of spatial accessibility to health services with the opening of private streets in Shenzhen, China. *IEEE Access*, 6, 72824 - 72835. <https://doi.org/10.1109/ACCESS.2018.2881654>
- Zeng, W., Xiang, L., Li, H., Zhang, X., & Zhang, H. (2017). 南京市医疗服务设施可达性的空间格局及其 形成机制 [Spatial pattern of access to health care facilities and its formation mechanism in Nanjing]. *Economic Geography*, 37(6), 136 - 143. <https://doi.org/10.15957/j.cnki.jjdl.2017.06.019>
- Zhang, C., Li, X., & Joyce, Y. M. (2017). 北京城市保障性住房居民的就医可达性研究——基于 GIS 网络分 析方法 [Accessibility to health care service of the affordable housing residents in Beijing: Based on GIS network analysis]. *Human Geography*, 32(2), 59 - 64. <https://doi.org/10.13959/j.issn.1003-2398.2017.02.009>

- Zhang, T., Cao, L., & Liang, Z.-J. (2017). 基于 GIS 的城市医疗设施空间可达性研究——以哈尔滨市为例 [GIS-based spatial accessibility study of urban medical facilities: Taking Harbin City as an example]. *Journal of Epoch World*, 2017(4), 36 - 40. <https://doi.org/10.3969/j.issn.1673-7563.2017.04.014>
- Zhang, X., Chang, L., Wang, M., & Stein, A. (2023). Measuring polycentric urban development with multi-temporal Sentinel-1 SAR imagery: A case study in Shanghai, China. *International Journal of Applied Earth Observation and Geoinformation*, 121, 103340. <https://doi.org/10.1016/j.jag.2023.103340>
- Zhang, Z., Sun, S., Wang, X., Xiao, R., & Gao, J. (2022). 特大城市医疗公共服务可达性的空间格局及其影响因素分析——以上海市为例 [Spatial pattern of medical public services accessibility in megacities and its influencing factors: A case study of Shanghai]. *Scientia Geographica Sinica*, 42(4), 622 - 630. <https://doi.org/10.13249/j.cnki.sgs.2022.04.007>
- Zhong, S., Yang, X., & Chen, R. (2016). 层级性公共服务设施空间可达性研究——以北京市综合性医疗设施为例 [The accessibility measurement of hierarchy public service facilities based on multi-mode network dataset and the two-step 2SFCA: A case study of Beijing's medical facilities]. *Geographical Research*, 35(4), 731-744. <https://doi.org/10.11821/dlyj201604011>