Article

Methods for Infectious Disease Risk Assessments in Megacities Using the Urban Resilience Theory

Hao Wang [1], Changhao Cao [1,2,*], Xiaokang Ma [1] and Yao Ma [1]

1   Institute of Photogrammetry and Remote Sensing, Chinese Academy of Surveying and Mapping,
    Beijing 100830, China; wanghao@casm.ac.cn (H.W.); 202283020074@sdust.edu.cn (X.M.);
    raladyomo@gmail.com (Y.M.)
2   College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China
*   Correspondence: 19863815306@163.com

Abstract: Since the 20th century began, the world has witnessed the emergence of contagious diseases such as Severe Acute Respiratory Syndrome (SARS), H1N1 influenza, and the recent COVID-19 pandemic. Conducting timely infectious disease risk assessments is of significant importance for preventing the spread of viruses, safeguarding public health, and achieving sustainable development. Most current studies on epidemic risk assessments focus on administrative divisions, making it challenging to reflect the risk disparities within these areas. Taking Shanghai as an example, this research introduces the concept of urban resilience frameworks and identifies the risk factors. By analyzing the interactions among different risk factors using geographic detectors, this study establishes the distribution relationship between the risk factors and newly reported cases using Geographically Weighted Regression. A risk assessment model is constructed to evaluate the infection risk within different regions of the administrative area. The results demonstrate that the central area of Shanghai exhibits the highest infection risk, gradually decreasing toward the periphery. The Spearman's correlation coefficient (p) between the predicted and actual distribution of new cases reaches 0.869 ($p < 0.001$), and the coefficient of determination (R^2) is 0.938 ($p < 0.001$), indicating a relatively accurate assessment of infection risk in different spatial areas. This research methodology can be effectively applied to infectious disease risk assessments during public health emergencies, thereby assisting in the formulation of epidemic prevention policies.

Keywords: geographic big data; GWR; risk assessment; data-driven

1. Introduction

Since the 20th century began, infectious diseases such as Severe Acute Respiratory Syndrome (SARS), H1N1 influenza, and novel coronaviruses have rapidly spread worldwide. These diseases have significantly impacted economic and social development, persisting on a global scale. Particularly in megacities, which represent densely populated regions integrating geographical, political, economic, social, and cultural functions, there exists a complex interplay of abundant resources and immense pandemic pressures and risks. Ensuring the safety, health, and stability of these megacities is of paramount importance. Confronted with infectious diseases, timely and comprehensive risk assessments play a crucial role in preventing virus transmission, safeguarding public health, ensuring the security of megacities, and realizing sustainable development. The dynamic interconnections between various factors necessitate a meticulous evaluation to effectively combat the challenges posed by these diseases [1].

Infectious disease risk assessment refers to the use of existing information by health institutions to assess the level of threat posed by an epidemic and provide risk warnings. However, most current studies on epidemic risk assessments are based on administrative regions [2], which makes it difficult to reflect the differences in infection risks within

these regions [3]. Therefore, conducting fine-grained infectious disease risk assessment studies is essential for the precise management of epidemics within administrative regions, safeguarding public health, and achieving sustainable development.

Researchers have proposed various models for infectious disease risk assessment, such as the Susceptible–Exposed–Infectious–Recovered (SEIR) model, which uses the number of cases and population contact to construct differential equations [4]. The Pressure–State–Response (PSR) model combines multiple risk factors to assess the epidemic risk [5]. The Long Short-Term Memory (LSTM) model has been utilized to assess risks by exploring time-series information on disease infections [6,7]. However, these models often evaluate risks at the administrative level and pay less attention to the spatial distribution of risks within administrative regions [8,9]. Different areas within administrative regions often exhibit varying risks [10], such as the risk differences between densely populated and sparsely populated areas in terms of infection distribution [11].

The concept of "urban resilience" has opened up new avenues for epidemic risk assessments. Urban resilience refers to the ability of a city or urban system to absorb and withstand external shocks, maintaining its key features and functions without significant impact. When dealing with infectious diseases, different risks are often observed within urban areas due to varying external impacts and resistance capabilities. Using the "urban resilience" theory to construct models for calculating epidemic risks helps clarify the mechanisms underlying epidemic risks, enabling the scientific calculation of the impact and resistance of a city when facing infectious diseases, which in turn determines the accuracy of the model.

Previous studies have indicated that the impact force on a city during an epidemic is mainly determined by the number of newly infected individuals, while resistance is primarily influenced by the population, transportation, and aggregation in proximity to the patients [12,13]. The fine-grained representation of the spatial distribution of new infections and the density of surrounding populations are crucial for utilizing the "urban resilience" theory to assess epidemic risks in a granular manner [14,15]. With the advent of geospatial big data, these data can effectively represent the population, transportation, and aggregation in various areas within a city, making them widely used in spatial health analysis and research [16].

Researchers such as MF [17] have used geospatial big data to construct epidemic tree models to determine the basic reproduction numbers of different spatial epidemics. Xia Jizhe et al. [18] used geospatial big data to correct the transmission parameters of population dynamics models. Yao Xiao et al. [10] employed geospatial big data and random forest models to classify the risk of epidemic transmission in different areas within administrative regions, yielding favorable results.

Furthermore, in recent years, there has been an increasing amount of research utilizing the addresses of new patients and geocoding techniques for fine-grained spatial localization of epidemic patients [19–21]. For instance, Hu Tao et al. used geocoding techniques to map the distribution of liver diseases at a fine-grained city level [22], and Peng Ming Jun employed weighted geocoding techniques to map the community distribution of COVID-19 patients within a city [23].

According to the laws of geography, resilience indicators of the same size often have different effects in different spatial contexts. Geographically Weighted Regression (GWR) models have achieved good results in modeling with spatially varying effects. The GWR model explores the spatial variations and related influencing factors of diseases at a certain scale by establishing local regression equations at each point within the spatial extent and can be used to assess the future development of diseases. Due to its consideration of the local influences and effects of spatial objects, it exhibits higher accuracy.

Therefore, to address the problem of the difficulty in reflecting the differences in risk within administrative regions, this study introduces the concept of "urban resilience". Using Shanghai as an example and utilizing geocoding techniques to pinpoint the fine-grained distribution data of patients, this study characterizes the impact force indicators

faced by the city during an epidemic. Furthermore, this study combines grid-level data on diagnosed patients (GLD) obtained using geocoding techniques with geospatial big data such as population density (PD), points of interest (POI), and road network (RD) data to comprehensively construct risk factors (RFS).

This study establishes an epidemic infection risk assessment framework and analyzes the interaction between RFS and geographic detectors. Finally, by using the GWR model, the relationship between RFS and the distribution of new cases is modeled to construct the risk assessment model. The assessment results are then correlated with the actual distribution of cases to validate the model.

2. Materials and Methods

The pandemic risk within different regions of a city is intricately linked to its geographical characteristics. Large urban centers experience variations in infection rates among different areas due to differences in population size, the presence of gathering places such as supermarkets and public squares, and disparities in transportation infrastructure. Considering these pivotal factors, our research employs grid-level data on diagnosed patients (GLD), population density (PD), points of interest (POI) data, and road network (RD) data to create pandemic risk factors.

2.1. Study Area

Shanghai is located at the mouth of the Yangtze River on the central coast of mainland China and is divided into 16 districts. Since March 2022, Shanghai has experienced a sharp increase in the cumulative number of confirmed COVID cases, which was significantly impacted by the pandemic. Therefore, Shanghai was chosen as the study area due to its representative nature regarding the outbreak.

2.2. Data Sources

The data used in this study include geospatial big data and grid-level data on newly diagnosed patients.

2.2.1. Geospatial Big Data

The spatial distribution of populations and factors such as transportation and clustering hotspots are highly correlated. By combining corresponding geospatial data, it is helpful to accurately characterize the density of populations at the grid level and quantify population clustering characteristics.

This study selected POI data, road network data, and population density data from geospatial big data. POI data are highly correlated with population clustering hotspots. Supermarkets, public places, and public transportation hubs still attracted typical population clusters during the epidemic. Therefore, this study obtained POI data from Baidu Maps, including public services, shopping services, and transportation services categories, for Shanghai in 2022, totaling 109,237 records.

Additionally, through a grid-based analysis, hotspot areas of population clustering were divided into units, and each grid value represented the number of clustering hotspots in that area, indicating the attractiveness of geographical grid regions for population clustering.

Population density directly reflects the degree of population aggregation and is closely related to disease transmission. The data were obtained from the Land Scan Global Population Database (https://landscan.ornl.gov/, accessed on 1 May 2022), which aims to provide high-precision spatial population data for risk assessments. In this study, it was aligned with the data from the seventh national census for calibration purposes. The distribution of road networks exhibits a strong spatial correlation with population distribution [2].

Road network data were sourced from OpenStreetMap (https://www.openstreetmap.org, accessed on 1 May 2022). To meet the requirements of quantitative analysis, primary,

secondary, and urban arterial roads were selected, and a line density analysis was carried out to convert them into grid format.

2.2.2. Grid-Level Data on Diagnosed Patients

The data were obtained from the daily announcements by the Shanghai Municipal Health Commission (sh.gov.cn, accessed on 1 May 2022) regarding the residential information of the cases. This study utilized web scraping to obtain a total of 150,546 records of patients' residential information, with a higher number of newly infected individuals between 1 April and 14 May 2022.

Furthermore, the study utilized the geocoding technology available in the Baidu Maps API interface to obtain high-precision spatial location information for the cases. This technology converts the distribution addresses of the cases into spatial coordinates. Finally, the ArcGIS tool was utilized to add XY coordinates to spatialize the case data at a finer granularity. For quantitative analysis at the grid level, the patient community distribution data were divided into 1 km grids using the geographical grid method, and generated GLD data using geographic grid sampling, with each grid value representing the number of cases in that area.

Additionally, the distribution of new cases within each grid was used to indicate the risk of infection. Taking April 1 as an example, the resulting case distribution data are shown in Figure 1.
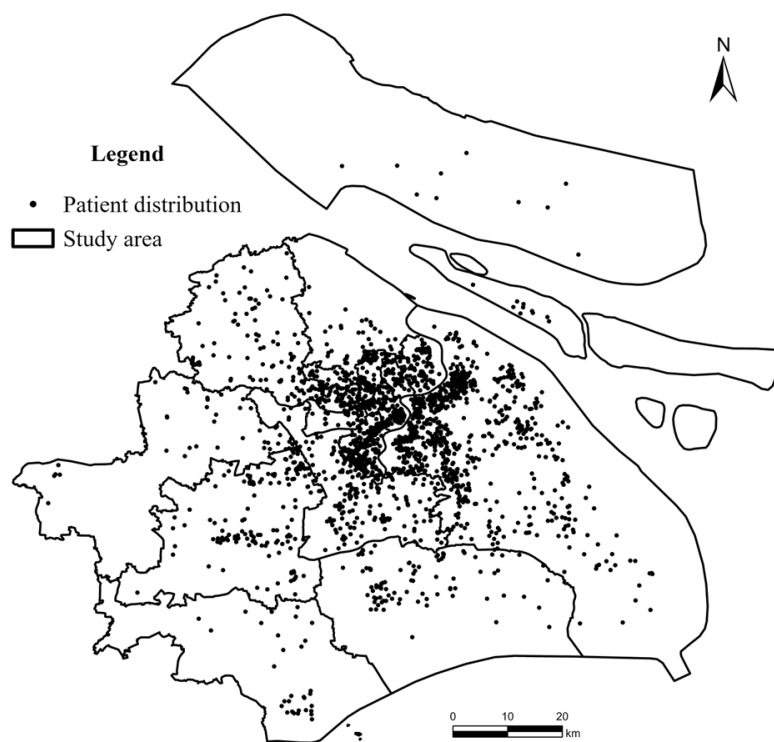

Figure 1. Fine-scale case distribution data.

According to the provided text, the incubation period of a general coronavirus infection is typically around 14 days. Therefore, it is possible to designate a 14-day period as an analytical cycle for studying the distribution of new cases.

In this study, the obtained epidemiological data from Shanghai are divided into three periods: April 1 to April 15, April 16 to April 30, and May 1 to May 14 in 2022. The first two periods are used for detecting an interaction and establishing evaluation models. The third period is used for model validation.

Additionally, we conducted a multicollinearity test on the selected indicators using the Variance Inflation Factor (VIF). The results of the test revealed VIF values of 6.7, 3.8, 4.4,

and 2.7 for the indicators GLD, PD, POI, and RD, respectively. All these values were found to be less than 10, indicating the absence of severe multicollinearity issues at a tolerance level of 0.1.

2.3. Risk Assessment Model Establishment Methods

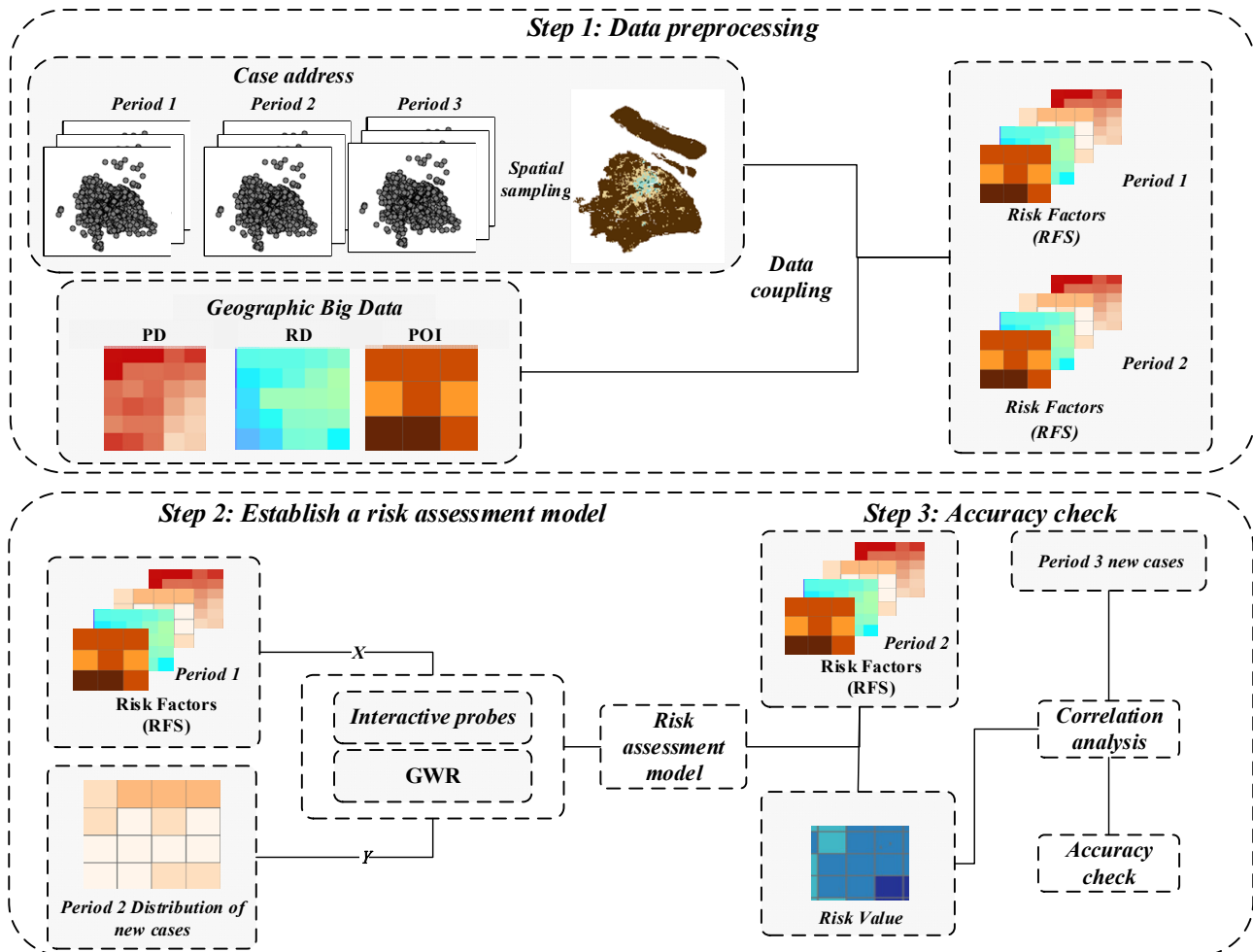The experimental flowchart is shown in Figure 2.

Figure 2. Experimental flowchart.

Within the framework of resilient cities theory, the risk faced by different regions within a city in dealing with infectious diseases primarily consists of two elements: shocks and resilience. Using the following examples depicted in Figure 3a,b, the methodology for analyzing epidemic risks under the theory of urban resilience can be elucidated. In Figure 3a, which depicts a region with a low resilience level, a higher risk is often manifested when facing the same shocks compared to the region depicted in Figure 3b, which exhibits a high resilience level and consequently shows lower risk. Furthermore, within the same region, when confronted with different shocks, a greater risk is generated when the impact is stronger.

Therefore, this study characterizes the impact indicators of different regions within a city in the face of an epidemic by utilizing patient distribution data at the grid scale (grid-level data). Additionally, geospatial big data such as PD, POI data, and RD are employed as resilience indicators within the framework. The combination of impact and resilience constructs the RFS.
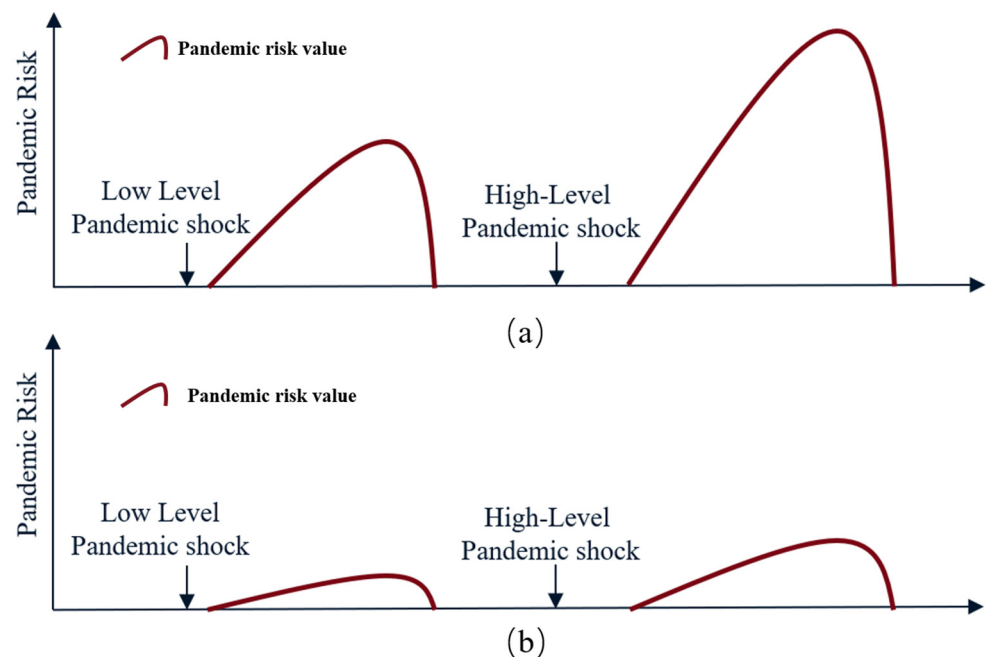
Figure 3. Epidemic risk analysis chart based on urban resilience theory: (**a**) Area with low level of resistance; (**b**) Area with high level of resistance.

2.3.1. RFS Interaction Detection Method

The geographic detector technique allows for the exploration of the interaction between RFS [6]. It is used to assess the coupling relationship between RFS and the distribution of new cases. One advantage of the geographic detector is that it does not assume linearity and has clear physical interpretations. The quantitative evaluation of the results is represented by the q-value, which reflects the similarity of spatial patterns among different factors. The change in q-values before and after RFS interactions is used to evaluate the coupling relationship between various indicators. The q-value is calculated using the following formula:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} = 1 - \frac{SSW}{SST}$$
$$SSW = \sum_{h=1}^{L} N_h \sigma_h^2 \tag{1}$$
$$SST = N\sigma^2$$

Here, h = 1, 2, ..., L represents the stratification of the independent variable X or the dependent variable Y. N_h and N are the number of units in stratum h and the entire region, respectively. σ_h^2 and σ^2 are the variances of the Y values in stratum h and the entire region, respectively.

In this study, the "GD" package in the R language is used to perform the geographic detector analysis. The RFS are treated as explanatory variables (X) and the distribution of new cases is the variable of interest (Y). The variables are stratified according to the optimal stratification scheme provided. After calculating the q-value for individual factors, "q(X1∩X2)" is computed to analyze the interaction between factors in space. If "q(X1∩X2)" > Max(q(X1), q(X2)), this indicates an enhanced interaction between the two factors. If "q(X1∩X2)" < Min(q(X1), q(X2)) or Min(q(X1), q(X2)) < "q(X1∩X2)" < Max(q(X1), q(X2)), this suggests a weakened interaction between the two factors.

2.3.2. Establishment Method of Risk Factors and Distribution of New Cases

Establishing the relationship between RFS and the distribution of new cases involves the use of Geographically Weighted Regression (GWR) models, which are essential tools for explaining the spatial distribution of diseases [7–9]. These models analyze the spatial heterogeneity of the impact through the distribution of regression coefficients and perform

a risk assessment based on the fitting relationship. By incorporating a spatial weighting function, GWR models link grid points with neighboring areas and perform regression modeling in each partition.

Compared to the Ordinary Least Squares (OLS) model, GWR models can more effectively consider the influence of geographic neighbors and the heterogeneity of the impact factors. By using the GWR model, the neighborhood case distribution and population characteristics, as well as the heterogeneous influence levels of the factors in different regions, can be adequately considered. This provides a better explanation of the spatial distribution of RFS and new cases.

To eliminate the influence of data dimensionality, RFS are standardized using the following formula:

$$
\begin{aligned}
Risk_l(u_l, v_l) &= \sum_{i=1}^{n} \beta_{gwi}(u_l, v_l) \cdot x_{il} + \varepsilon_l \\
\beta_{gwi}(u_l, v_l) &= \left(X^T W(u_l, v_l) X\right)^{-1} X^T W(u_l, v_l) Y \\
x &= \frac{x_i - \overline{x}}{\sigma^2}
\end{aligned}
\tag{2}
$$

Here, (u_l, v_l) represents the spatial location of the $l-$th sample, and $Risk_l$ and x_{il} represent the risk and RFS value at the l-th spatial location, respectively. $\beta_{gwi}(u_l, v_l)$ represents the regression coefficient of the i-th independent variable for the $l-$th sample in space. ε_l is the random error, following a normal distribution.

2.4. Accuracy Test Method

In order to assess the infection risk in the subsequent period 3 of the study area, the RFS during mid-term 2 of the research area were used as explanatory variables in the evaluation model.

The relative magnitude of the risk index obtained from the model was used to assess the level of infection risk among different regions within the administrative area [10].

Additionally, to validate the accuracy of the risk assessment model, the evaluation results were subjected to correlation analysis with the actual distribution of cases, and the Spearman correlation coefficient (p) and the coefficient of determination (R2) for the linear regression relationship between the two were calculated. The Spearman correlation coefficient (p) quantitatively evaluates the ordinal relationship between two sets of data distributions [11], determining whether there is a higher number of new cases in areas with higher risk indexes.

The coefficient of determination (R2) assesses the explanatory power of the heterogeneous distribution of risk indexes on the heterogeneous distribution of actual new cases by calculating the extent to which the variation of the independent variable explains the variation of the dependent variable [12]. The calculation formulas are provided below:

$$
\begin{aligned}
p &= 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \\
R^2 &= 1 - \frac{\sum_{i=1}^{n} (\hat{y}_1 - y_i)^2}{\sum_{i=1}^{n} (\overline{y}_1 - y_i)^2}
\end{aligned}
\tag{3}
$$

d_i represents the difference between the risk index of the grid region i and the ordinal distribution of new populations and n represents the sample size. y_i represents the actual distribution of cases and $\hat{y}_1$ is the regression-fitted value using the evaluated risk index.

3. Results

3.1. Analysis of Risk Assessment Model Results

In this study, the *GLD* was considered as x_1, *PD* as x_2, *POI* as x_3, and *RD* as x_4. The results of the single-factor explanatory power (q-value) and its significance (*p*-value) are presented in Table 1, while the results of the RFS interactions are shown in Table 2.

Table 1. q-value of single factor (GLD was considered as x_1, PD as x_2, POI as x_3, and RD as x_4).

X	x_1	x_2	x_3	x_4
q	0.813	0.720	0.536	0.111
p	<0.01	<0.01	<0.01	<0.1

Table 2. Interaction analysis (GLD was considered as x_1, PD as x_2, POI as x_3, and RD as x_4).

C	A + B	Result
x1∩x2 = 0.912	x1(0.813) + x2(0.720)	C > Max (A, B)
x1∩x3 = 0.911	x1(0.813) + x3(0.536)	C > Max (A, B)
x1∩x4 = 0.822	x1(0.813) + x4(0.111)	C > Max (A, B)
x2∩x3 = 0.840	x2(0.720) + x3(0.536)	C > Max (A, B)
x2∩x4 = 0.810	x2(0.720) + x4(0.111)	C > Max (A, B)
x3∩x4 = 0.871	x3(0.536) + x4(0.111)	C > A + B

Based on the q-values of single factors (Table 1), the highest explanatory power is observed for patient distribution, reaching 0.813. This indicates that the spatial distribution of cumulative cases is the main factor influencing the spatial distribution of future new cases. The greater the number of cumulative case distributions in a region, the higher the number of future new cases.

The population density factor follows, with a q-value of 0.72, which is slightly lower than the patient distribution factor but still at a relatively high level. It reflects a high similarity between areas with high/low population density and areas with high/low numbers of new cases. Therefore, in areas with higher population density, there are more patient distributions and higher risks.

The q-value of cluster hotspots POI reaches 0.536, indicating that regions with more clustering hotspots generally have a higher number of case distributions.

The factor with the lowest explanatory power is road network density, with a q-value of 0.111, and it also exhibits lower significance.

An analysis of the interaction results (Table 2, Figure 4) reveals that after interacting with population density, patient distribution exhibits a higher explanatory power (0.912) compared to its individual factor (0.813).
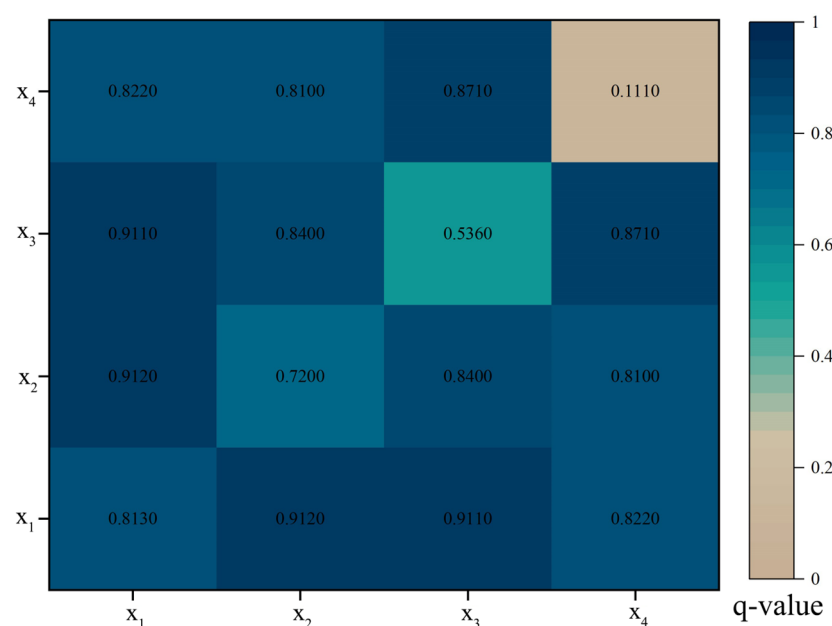

Figure 4. RFS interaction results (GLD was considered as x_1, PD as x_2, POI as x_3, and RD as x_4).

Moreover, when interacting with road network density and cluster hotspot indicators, the explanatory power for the distribution of new patients is enhanced, reaching 0.911 and 0.822, respectively.

Although road network density alone shows lower explanatory power, its interaction with cluster hotspots demonstrates significant non-linear enhancement. The dense transportation network facilitates population flow toward clustering hotspots, leading to a substantial increase in regional infection risk through interaction. The interaction between various indicators of the risk index enhances their explanatory power, demonstrating a synergistic effect. Therefore, combining patient distribution data with geographical big data can better explain the spatial heterogeneity of patient distribution.

3.2. Analysis of the Relationship between RFS and the Distribution of New Patients

The relationship between RFS and the distribution of new cases (Figure 5) was fitted using the Geographically Weighted Regression (GWR) model. All variables of the RFS passed the significance test at a confidence level of 0.05. The fitted coefficient of determination (R^2) was 0.903 ($p < 0.001$).
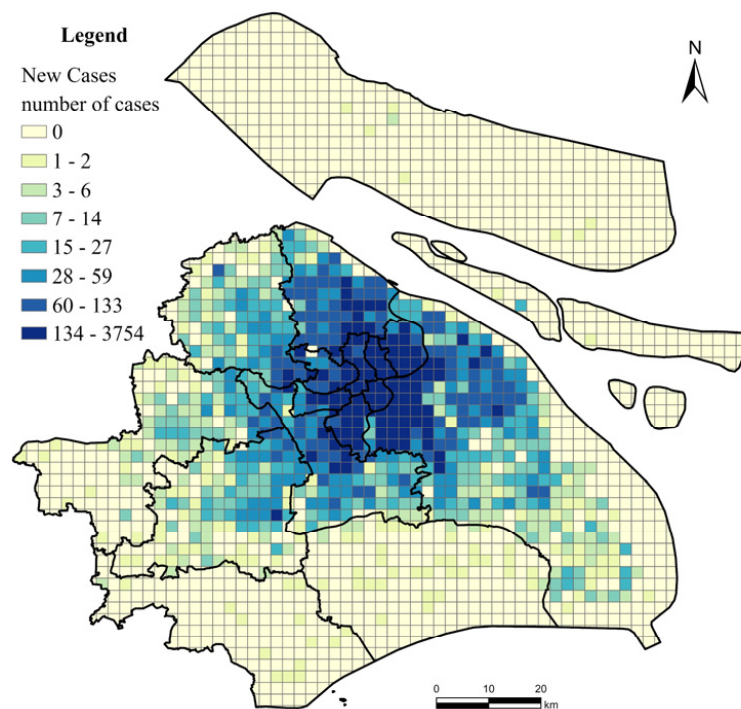

Figure 5. New cases.

The influence coefficients of the RFS variables were categorized using the natural break classification method and visualized for analysis (Figure 6). The parameter estimation results of each indicator in the grid units exhibited distinct variations across different regions. Overall, most indicators showed positive regression coefficients, indicating a strong spatial variation in the impact of RFS on the spatial distribution of new patients.

The high-value areas of the fitted coefficient between the infected population distribution and population density were primarily concentrated in the city center. The impact decreased gradually from the center to the surrounding areas.
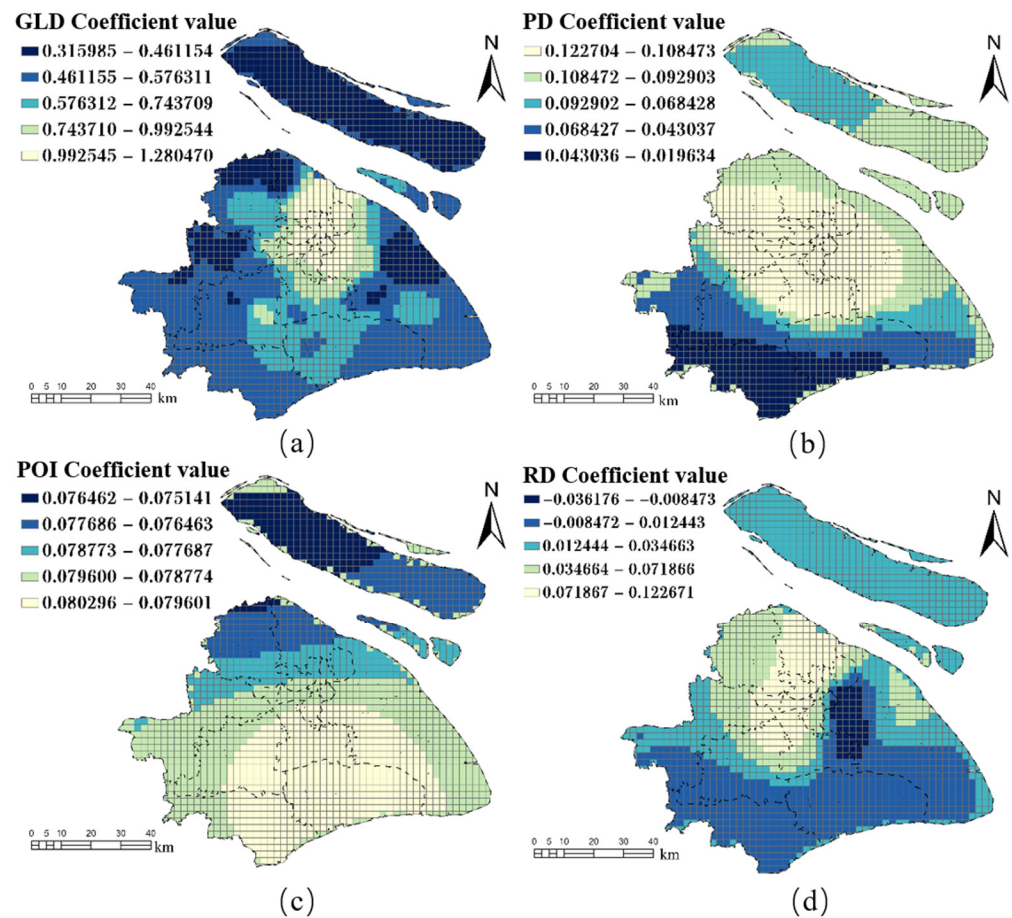
Figure 6. Spatial distribution of GWR regression coefficients: (**a**) GLD coefficient; (**b**) PD coefficient; (**c**) POI coefficient; (**d**) RD coefficient.

Table 3 illustrates the spatial distribution statistics of coefficients. It is evident that the highest coefficient corresponds to the distribution of patients from the previous time period, with a maximum value of 1.28. In contrast, the coefficients for the factors PD, POI, and RD exhibit close statistical values.

Table 3. GWR results.

Variable	Mean	STD	Min	Media	Max
Intercept	−0.161	0.129	−0.993	−0.145	0.451
GLD	0.632	0.298	0.315	0.622	1.280
PD	0.097	0.007	0.019	0.062	0.122
POI	0.078	0.002	0.075	0.077	0.080
RD	0.036	0.054	−0.361	0.031	0.071

The spatial distribution of the influence coefficient of POI displayed a zonal pattern from south to north, with relatively small overall variations and almost no significant spatial heterogeneity. The impact of aggregated hotspots was not strongly associated with whether they were located in the city center or suburban areas.

The population in both suburban and central areas of the city resided in environments with a higher risk of susceptibility. Regarding the influence of road network density, the coefficient was largest in the city center and decreased toward the surrounding areas. However, negative values appeared in areas closer to the city center, which could be attributed to the proximity of these regions to the city center and the influx of population predominantly concentrated in the central area.

3.3. Model Accuracy Evaluation

In the evaluation model constructed by inputting the RFS within period 2 as explanatory variables, the risk of new COVID-19 infections in various regions during the next period, period 3, was assessed. Based on the assessment of the risk of infection (Figure 7a), the spatial distribution of the risk index exhibited a spatial pattern of decreasing intensity from the center to the periphery. The Huangpu District, situated in the central region, had the highest infection risk index, surpassing 7.
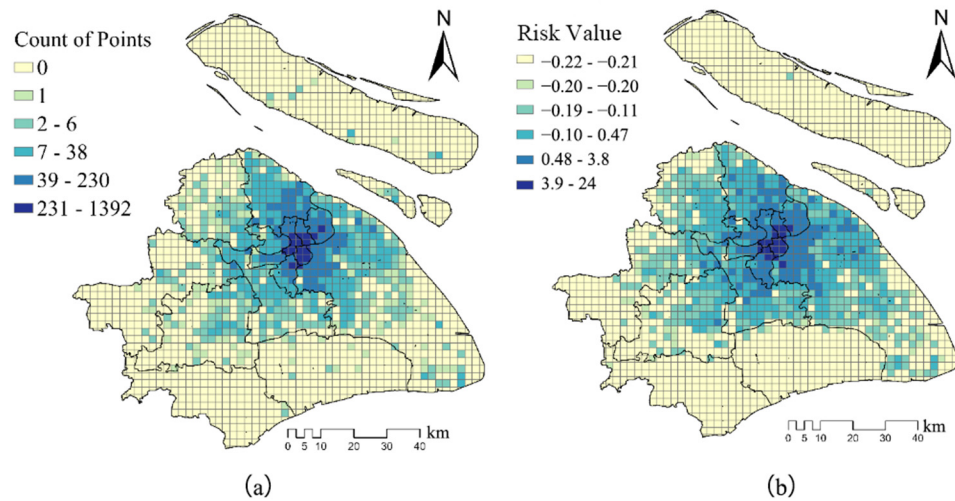

Figure 7. GWR results vs. actual new cases: (**a**) GWR risk value results; (**b**) Spatial distribution of actual new cases.

A correlation analysis was conducted between the assessment results and the spatial distribution of actual new cases within the corresponding time period (Figure 7b), resulting in a scatter plot of the correlation (Figure 8). Overall, both the coefficient of determination (R^2) and the Spearman correlation coefficient were found to be at a relatively high level. With an R^2 value of 0.938 ($p < 0.01$), the heterogeneous distribution of the assessed risk index can effectively explain the spatial heterogeneity of newly infected individuals. According to the Spearman correlation coefficient of 0.869 ($p < 0.01$), there is a good correlation between the risk index and the number of patient distributions, indicating that in the high-value assessment areas, the number of new cases also tends to be high.
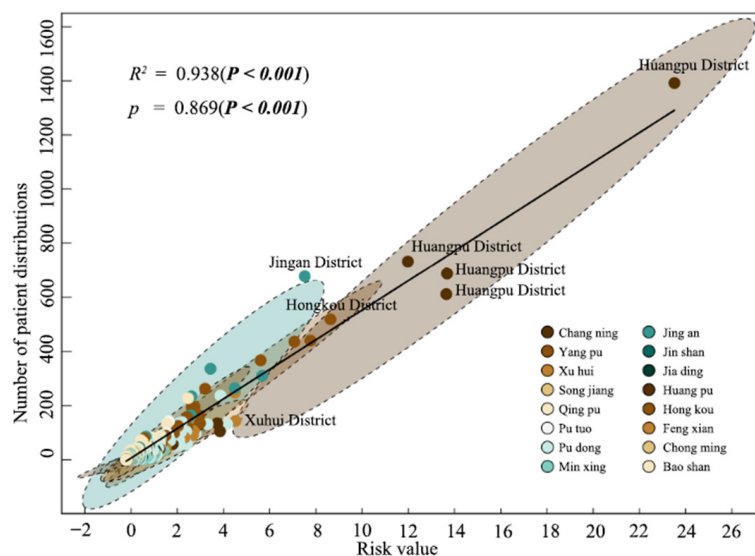

Figure 8. Correlation scatter plot.

Specifically, several grid cells in the Huangpu District had standardized risk indices exceeding 10, and the number of actual new cases in residential areas was also the highest, all of which fell within the 95% confidence ellipse, indicating a strong correlation in the high-value areas. However, in some low-value areas, the risk assessment appeared to be overestimated for certain regions, which was possibly due to higher road network density and population density. Nevertheless, most of the low-value areas also fell within the 95% confidence ellipse. In general, both the model fit and the ordinal correlation were quite good. Therefore, the model achieved good results by integrating patient distribution data with geographic big data related to population aggregations and mobility patterns.

To comprehensively investigate the variations in model performance across different regions, our study employed a geographical division of Shanghai based on national standards. We conducted a detailed analysis of the model's accuracy discrepancies within the urban central areas and other regions. The central urban areas, as defined, comprise seven distinct administrative districts, namely, Huangpu, Xuhui, Changning, Yangpu, Hongkou, Putuo, and Jing'an. In contrast, the remaining regions were classified as non-core areas. Notably, these central areas are characterized by a significantly higher population density, while the non-core areas exhibit a comparatively lower population density.

As shown in Figure 9a,b, it is evident that in Shanghai's core areas, the model achieved a coefficient of determination (R2) of 0.943. In contrast, in other areas, the model's R2 was 0.826, which is noticeably lower than the 0.943 in core areas. This clearly indicates that the model exhibits higher precision in high-population density core areas.
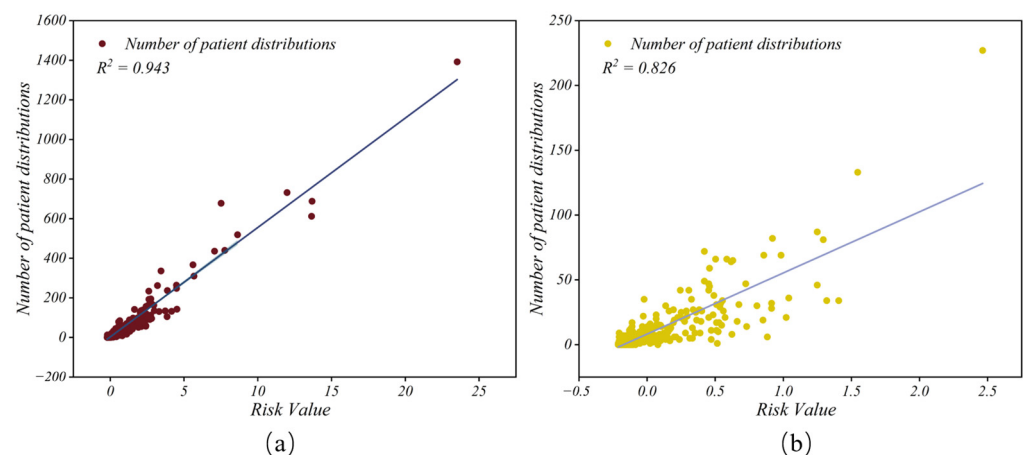

(a)

(b)

Figure 9. Central area and other area correlation scatter plot: (**a**) Central area scatter plot; (**b**) Other area scatter plot.

4. Discussion

4.1. Model Advantages and Potential for Large-Scale Applications

We found that GLD is the most critical factor for epidemic risk generation (Table 3), enhancing the model's reliability by incorporating GLD indicators. In contrast to numerous prior studies [12,14–19], detailed data on patient distribution are frequently overlooked in risk research. In this study, the residential addresses of patients were geographically coded, and fine-grained patient distribution data were obtained as a risk factor, significantly enhancing the scientific foundation of our risk factor analysis [19–22].

In Figure 9, the model demonstrates superior accuracy in the central areas of the study region, particularly in the core. This heightened precision can be attributed to the more concentrated distribution of patients in these central areas, aligning with Yao Xiao's perspective [10]. This suggests that the model may be particularly well-suited for urban core regions.

This study uses a 1 km × 1 km spatial scale for risk assessment, providing a more nuanced representation of risk variations in various areas within administrative regions.

Unlike the numerous studies typically concentrated on a macro scale, as illustrated in Table 4, this study excels in delineating risk variations within urban areas, contributing to the formulation of specific prevention and control policies [24–31]. Therefore, the risk assessment of our model significantly aids in the meticulous management of epidemic risks at a micro level.

Table 4. Spatial scale comparison.

Related Research	Spatial Scale Level
Xu et al. [30]	Provincial level
Wei et al. [31]	District level
Our study	1 km × 1 km grid level

4.2. Implications of Research Results for Epidemic Prevention and Control

As shown in Table 1, the impact of new cases in the past 14 days emerges as the most influential factor on epidemic risk. These findings align with our existing understanding of the fundamental spatial distribution pattern of COVID-19 [32–34]. This emphasizes the need to address dynamic risks associated with the geographical locations of infected individuals for effective epidemic risk management. It helps reduce substantial risks posed by these individuals to the epidemic. Population density is the second-most influential factor, indicating a higher susceptibility to disease outbreak and spread in densely populated areas [35,36]. Consequently, the control of densely populated areas becomes a critical focal point for enhanced prevention and management of epidemics.

Table 2 reveals stronger interactions among various risk factors, particularly the nonlinear enhancement between road networks and gathering hotspots' Points of Interest (POI). This highlights the imperative to prioritize the control of highly interconnected areas within road networks as a strategic measure to mitigate the risk of disease spread during epidemics.

In Figure 5, GWR model regression coefficients show significant spatial heterogeneity, with notable differences in GLD and Population Density (PD) coefficients. To better understand the spatial heterogeneity of GLD and PD regression coefficients, a statistical analysis of the coefficients' means was conducted, categorizing the region into core and non-core areas (Figure 10). The statistical results indicated higher regression coefficients in the core areas. This phenomenon can be attributed to the increase in population density [15], the presence of densely populated spaces [18], and the extensive distribution of complex transportation networks in urban center areas [27]. The synergistic interaction of these interconnected factors creates favorable conditions for the spatial spread of diseases [37]. Consequently, due to increased population density and numerous gathering places, there is an increased infection risk in core areas [38]. This emphasizes the urgent need, especially during pandemics, to implement targeted intervention measures meticulously designed to address the escalating risks within the urban core zones [39,40].

Interaction results and GWR regression coefficients indicate that higher population density, concentrated urban spaces, and complex transportation networks create favorable conditions for the rapid spread of infectious sources. These findings have implications for public health policies and intervention measures, emphasizing the need for nuanced approaches to protect population health, especially in densely populated core regions of urban centers [37–40].
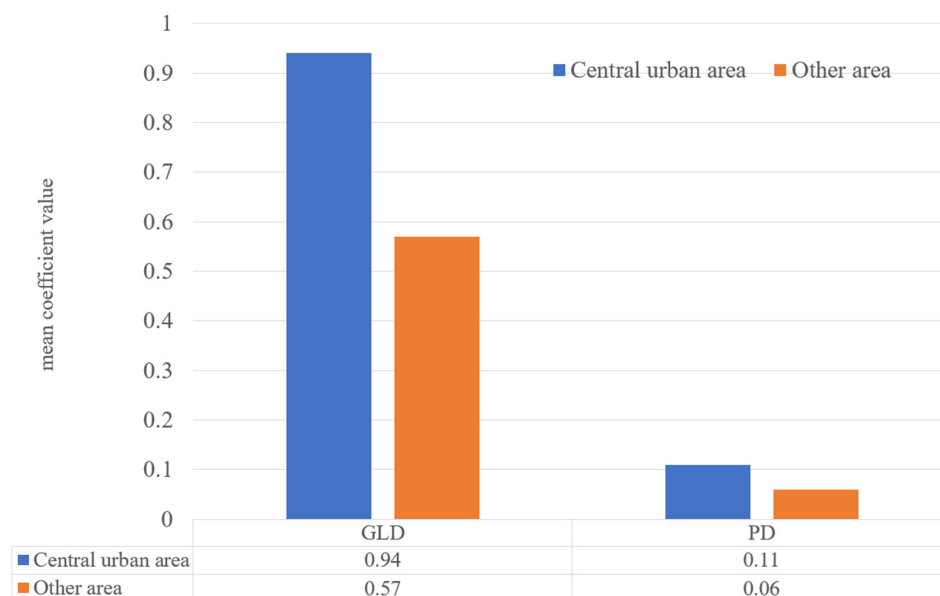
	GLD	PD
■ Central urban area	0.94	0.11
■ Other area	0.57	0.06

Figure 10. The mean coefficients of the central region and other regions.

4.3. Shortcomings and Prospects

This study achieved a high level of accuracy in establishing an epidemic risk assessment model using geographic detector and GWR models. However, it is crucial to acknowledge that local economic conditions often influence epidemiological risks to a certain extent [41]. Although variables such as economic conditions are interconnected, obtaining them at a fine spatial resolution is challenging. Furthermore, data on fine-grained economic conditions typically introduce significant measurement errors. Therefore, this study did not incorporate these data as risk factors. Subsequent research could benefit from integrating high-precision data related to these risk factors to further enhance the model. Additionally, considering that this study only reflects spatial characteristics and does not account for temporal aspects, future efforts could improve the robustness of this analytical framework by integrating time-series models with the GWR model.

5. Conclusions

In examining spatial variations in COVID-19 risks within a city through the lens of urban resilience, we applied geographic coding techniques to gridify the distribution data of COVID-19 cases. These were then integrated with geographic big data, including points of interest, population density, and road network density, as risk factors. Utilizing the Geographically Weighted Regression model, we developed a risk assessment model to evaluate infection risks across different areas within the administrative regions over a 14-day period. Subsequently, we conducted a correlation analysis between the assessment results and the actual distribution of cases to gauge the model's precision. Our study led to several key conclusions:

1. The model crafted in this study accurately simulates the spatial variation in COVID-19 infection risks within diverse areas of the administrative regions. This underscores its reliability in assessing infection risks across different spatial units within the administrative regions;
2. By accounting for the interplay among risk factors, the explanatory power for the spatial distribution of new cases is heightened, revealing a synergistic effect;
3. The assessment of infection risks in Shanghai reveals a spatial pattern characterized by a gradual decrease from the city center towards the periphery. This indicates that the core areas of Shanghai provide favorable conditions for the spatial spread of diseases, resulting in elevated risks in the central regions.

In essence, this research enhances our understanding of the intricate interplay between urban resilience factors and COVID-19 risks, providing valuable insights for targeted interventions and public health strategies within urban environments.

Author Contributions: Conceptualization, H.W. and X.M.; Methodology, H.W. and Y.M.; Investigation, C.C.; Data curation, H.W.; Writing—original draft, C.C.; Writing—review & editing, H.W.; Visualization, C.C.; Project administration, H.W.; Funding acquisition, H.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, S.; Tu, W.; Jin, M.; Feng, Y.; Xie, Y.; Tang, L.; Shi, G.; Xiang, N. Risk assessment of sudden public health emergencies to be concerned in mainland China in November 2022. *Dis. Surveill.* **2022**, *37*, 1389–1392.
2. Fang, Y.; Gu, K. Exploring the risk assessment of geographical spatial epidemic based on multiple data: A case study of COVID-19 epidemic data from January 1 to April 11, 2020 in China. *J. Geo-Inf. Sci.* **2021**, *23*, 284–296.
3. Liu, S.; Zhou, Y.; Yang, X.; Yin, J. COVID-19 positive cases prediction based on LSTM algorithm and its variants. In Proceedings of the Asia Conference on Algorithms, Computing and Machine Learning (CACML), Hangzhou, China, 25–27 March 2022; pp. 268–271.
4. Piovella, N. Analytical solution of SEIR model describing the free spread of the COVID-19 pandemic. *Chaos Solitons Fractals* **2020**, *140*, 110243. [CrossRef]
5. Jin, S.; Wang, R. Construction of city epidemic risk assessment index system based on PSR model: A case study of novel coronavirus pneumonia. In Proceedings of the 2020/2021 China Urban Planning Annual Meeting and 2021 China Urban Planning Academic Season, Chengdu, China, 25–30 September 2021; pp. 136–143.
6. Li, Z.; Gao, H.; Dai, X.; Sun, H. Epidemic transmission risk prediction model coupling LSTM algorithm and cloud model. *J. Geo-Inf. Sci.* **2021**, *23*, 1924–1935.
7. Sunjaya, B.A.; Permai, S.D.; Gunawan, A.A. Forecasting of COVID-19 positive cases in Indonesia using long short-term memory (LSTM). *Procedia Comput. Sci.* **2023**, *216*, 177–185. [CrossRef]
8. Annas, S.; Pratama, M.I.; Rifandi, M.; Sanusi, W.; Side, S. Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. *Chaos Solitons Fractals* **2020**, *139*, 110072. [CrossRef]
9. Shen, J. Research on Data-Driven Modeling and Application of COVID-19 Epidemic. Master's Thesis, Jiangnan University, Wuxi, China, 2022.
10. Yao, Y.; Yin, H.; Li, X.; Guo, Z.; Ren, S.; Wang, R.; Guan, Q. Risk assessment of the spread of Wuhan novel coronavirus pneumonia based on fine-scale multi-source geographical data. *Acta Ecol. Sin.* **2021**, *41*, 7493–7508.
11. Li, X.; Zhou, L.; Jia, T.; Wu, H.; Zou, Y.; Qin, K. The impact of urban factors on COVID-19 epidemic: A case study of Wuhan. *J. Wuhan Univ. (Inf. Sci. Ed.)* **2020**, *45*, 826–835.
12. Yu, Z.; Tian, R.; Sun, X.D. Research progress on risk assessment indicators of novel coronavirus pneumonia. *Front. Public Health Control* **2022**, *27*, 6–10.
13. Du, F.Y.; Wang, J.E.; Jin, H.T. Construction of spatial interaction network theory based on individual "movement-contact" and risk assessment of epidemic. *J. Geogr. Sci.* **2022**, *77*, 2006–2018.
14. Pei, T.; Wang, X.; Song, C.; Liu, Y.; Huang, Q.; Shu, H.; Chen, X.; Guo, S.; Zhou, C. Advances in COVID-19 Spatiotemporal Analysis and Modeling. *J. Geo-Inf. Sci.* **2021**, *23*, 188–210.
15. Qu, X.; Yuan, W.; Yuan, W.; Hu, J.; Meng, Q. Application of Spatiotemporal Big Data Analysis Techniques in Infectious Disease Prediction and Early Warning. *Chin. Digit. Med.* **2015**, *10*, 36–39.
16. Zou, Y. Research on Population Spatialization with Multi-source Data Support. Master's Thesis, China University of Mining and Technology, Beijing, China, 2020.
17. Li, M.; Shi, X.; Li, X.; Ma, W.; He, J.; Liu, T. Epidemic Forest: A Spatiotemporal Model for Communicable Diseases. *Ann. Am. Assoc. Geogr.* **2019**, *109*, 812–836. [CrossRef]
18. Xia, J.; Zhou, Y.; Li, Z.; Li, F.; Yue, Y.; Cheng, T.; Li, Q. Risk Assessment of Novel Coronavirus Transmission Driven by Urban Spatiotemporal Big Data: A Case Study of the Guangdong-Hong Kong-Macao Greater Bay Area. *Acta Geod. Cartogr. Sin.* **2020**, *49*, 671–680.

19. Jacquez, G.M. A research agenda: Does geocoding positional error matter in health GIS studies? *Spat. Spatio-Temporal Epidemiol.* **2012**, *3*, 7–16. [CrossRef]

20. Zimmerman, D.L.; Li, J.; Fang, X. Spatial autocorrelation among automated geocoding errors and its effects on testing for disease clustering. *Stat. Med.* **2010**, *29*, 1025–1036. [CrossRef]

21. Han, D.; Bonner, M.R.; Nie, J.; Freudenheim, J.L. Assessing bias associated with geocoding of historical residence in epidemiology research. *Geospat. Health* **2013**, *7*, 369–374. [CrossRef]

22. Hu, T. Research on Address Matching and Its Application in Urban Disease Mapping and Spatial Analysis. Master's Thesis, Wuhan University, Wuhan, China, 2015.

23. Peng, M.; Li, Z.; Liu, H.; Meng, C.; Li, Y. Application of Weighted Geocoding Based on Chinese Word Segmentation in the Spatial Localization of COVID-19 Epidemic Prevention and Control. *J. Wuhan Univ. (Inf. Sci. Ed.)* **2020**, *45*, 808–815.

24. Xiao, X.; Yang, C.; Tan, K.; He, H.; Zhang, T.; Li, X. Application of Geographically Weighted Regression Model in Spatial Analysis of Infectious Diseases. *Chin. J. Health Stat.* **2013**, *30*, 833–836+841.

25. Wang, Z.; Liu, L.; Yang, K. Research Progress on Spatiotemporal Geographically Weighted Regression Model and its Application in Epidemiology. *Chin. J. Schistosomiasis Control* **2023**, *35*, 199–205. [CrossRef]

26. Fu, H.; Sun, Y.; Wang, B.; Chen, L.; Zhang, H.; Gao, S.; Mao, J.; Jing, Y.; Shao, S. Estimation of PM(2.5) concentration in Beijing-Tianjin-Hebei region based on AOD data and GWR model. *China Environ. Sci.* **2019**, *39*, 8.

27. Chakraborty, T.; Ghosh, I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos Solitons Fractals* **2020**, *135*, 109850. [CrossRef]

28. Asteris, P.G.; Douvika, M.G.; Karamani, C.A.; Skentou, A.D.; Chlichlia, K.; Cavaleri, L.; Daras, T.; Armaghani, D.J.; Zaoutis, T.E. A Novel Heuristic Algorithm for the Modeling and Risk Assessment of the COVID-19 Pandemic Phenomenon. *Comput. Model. Eng. Sci.* **2020**, *125*, 815–828. [CrossRef]

29. Zhou, Z.; Zhao, Y.; Shen, C.; Wang, Z. Construction of a Whole Process Prevention and Control System for Major Public Health Risks in Cities. *J. Xi'an Jiaotong Univ. (Soc. Sci. Ed.)*. 1–20. Available online: http://kns.cnki.net/kcms/detail/61.1329.C.20230717.1227.002.html (accessed on 23 November 2023).

30. Xu, L. Spatio Temporal Quantitative Modeling and Risk Assessment of the Spread of COVID-19. Master's Thesis, Anhui University of Finance and Economics, Bengbu, China, 2021.

31. Wei, Y.; Jiang, N.; Chen, Y.; Li, X.; Yang, Z. Modeling and application of epidemic risk assessment considering spatiotemporal and spatial interactions. *J. Earth Inf. Sci.* **2021**, *23*, 274–283.

32. Tessema, Z.T.; Azanaw, M.M.; Bukayaw, Y.A.; Gelaye, K.A. Geographical variation in determinants of high-risk fertility behavior among reproductive age women in Ethiopia using the 2016 demographic and health survey: A geographically weighted regression analysis. *Arch. Public Health = Arch. Belg. Sante Publique* **2020**, *78*, 74. [CrossRef]

33. Danny, W. Spatial risk adjustment between health insurances: Using GWR in risk adjustment models to conserve incentives for service optimisation and reduce MAUP. *Eur. J. Health Econ. HEPAC Health Econ. Prev. Care* **2019**, *20*, 1079–1091.

34. Banerjee, K.S.; Andrews, B. Analysis of SEIR Epidemic Model Engraft with Incompatible Incidence Rate. *Biomed. Stat. Inform.* **2023**, *8*, 37–41. [CrossRef]

35. Zhou, X.; Ma, X.; Gao, S.; Ma, Y.; Gao, J.; Jiang, H.; Zhu, W.; Hong, N.; Long, Y.; Su, L. Measuring the worldwide spread of COVID-19 using a comprehensive modeling method. *BMC Med. Inform. Decis. Mak.* **2023**, *21* (Suppl. 9), 384. [CrossRef]

36. Yan, Y.; Zhu, H.; Li, X.; Li, X.; Zhang, X.; Li, D. Improvement of SEIR Epidemic Dynamics Model and Its Matlab Implementation. *MEDS Public Health Prev. Med.* **2023**, *3*, 49–58. [CrossRef]

37. Moosavi, S.; Namdar, P.; Moghaddam Zeabadi, S.; Akbari Shahrestanaki, Y.; Ghalenoei, M.; Amerzadeh, M.; Kalhor, R. Healthcare workers exposure risk assessment in the context of the COVID-19: A survey among frontline workers in Qazvin, Iran. *BMC Health Serv. Res.* **2023**, *23*, 155. [CrossRef] [PubMed]

38. Ye, Y.; Fan, Y.; Hou, S.; Zhang, Y.; Qian, Y.; Sun, S.; Peng, Q.; Ju, M.; Song, W.; Loparo, K. Community Mitigation: A Data-driven System for COVID-19 Risk Assessment in a Hierarchical Manner. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management, ACM 2020, Virtual, 19–23 October 2020. [CrossRef]

39. Cont, R.; Kotlicki, A.; Xu, R. Modelling COVID-19 contagion: Risk assessment and targeted mitigation policies. *R. Soc. Open Sci.* **2021**, *8*, 201535. [CrossRef]

40. Picard, M. Why Do We Care More About Disease than Health? *Phenomics* **2022**, *2*, 145–155. [CrossRef]

41. Kalang, L.A.; Eboy, O.V. Mapping of Population Behaviour during the Early Phase of COVID-19 Disease Spread In Kota Kinabalu, Sabah Using PCA-GIS. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *1064*, 012005. [CrossRef]