Building and Environment 269 (2025) 112330

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Neighborhoods renewal implications for the post-pandemic era: A study of COVID-19 infections in Central Shanghai, China

Poui Chong [a,c,1], Chao Liu [a,b,1,*], Wengin Chung [a,c,1]

[a] The College of Architecture and Urban Planning, Tongji University, Shanghai, China
[b] Shanghai Tongji Urban Planning and Design Institute Co. Ltd., Shanghai, China
[c] Key Laboratory of Spatial Intelligent Planning Technology, Ministry of Natural Resources of the People's Republic of China, China

ARTICLE INFO

Keywords:
Older neighborhood
Post-pandemic
Urban renewal
Community resilience

ABSTRACT

Since the onset of the Coronavirus Disease 2019 (COVID-19), urban older neighborhoods have faced increased vulnerability, prompting research into neighborhood renewal and resilience. However, research on COVID-19 infection and influencing factors in China's older neighborhoods remains relatively scarce. This study analyzed COVID-19 infections in central Shanghai to identify neighborhood-level factors affecting transmission. Using principal component analysis (PCA) and person correlation coefficients (PCC) to process the data, we established multiple linear regression (MLR) and geographically weighted regression models (GWR). To explore nonlinear relationships, we incorporated the random forest method (RF). Results indicated that older neighborhoods had higher infection rates compared to newer ones. Socioeconomic and built environment factors significantly influenced infection rates. Specifically, higher population density, road network density, and the number of subway stations were positively correlated with increased infection rates. RF analysis revealed a complex, nonlinear relationship between the number of high-income residents and infection rates. This study integrates built environment, socioeconomic, and population characteristics factors using multiple modeling approaches to better understand their impact on infection rates. It also introduces research on mainland Chinese cities as case studies, offering valuable insights for updating older urban neighborhoods to enhance community resilience. However, the study did not fully consider the impact of policies at the time, and its findings are primarily applicable to older neighborhoods in cities similar to Shanghai. Future research should examine the effectiveness of various intervention policies, the long-term effects of neighborhood renewal on community resilience, and the applicability of these findings to other urban environments.

1. Introduction

The pandemic that emerged in 2019 and rapidly spread worldwide precipitated far-reaching socioeconomic ramifications, including loss of life and trade disruptions. As the outbreak spread, there were significant changes occurred in people's behavioral patterns, lifestyles, and work practices. In the early stages of the pandemic, various regions implemented lockdowns and restrictive measures, leading to the stagnation of economic activities and trade interruptions [1]. Simultaneously, the pandemic exposed weaknesses in public health systems and prompted a reevaluation and increased attention to urban planning and management. Therefore, analyzing and researching the Coronavirus Disease 2019 (COVID-19), also known as SARS-CoV-2, infection rates in the central area is crucial to comprehend the impact of the pandemic on urban areas and proposing response strategies that could enhance the resilience of neighborhoods.

Urban planning is intricately linked with the proliferation of pandemics. Urbanization that resulted from continuous population growth, such as housing congestion and inadequate infrastructure, influenced disease transmission, particularly among vulnerable populations [2]. Infectious diseases spread rapidly with urbanization, current research focused on identifying and addressing key influencing factors to enhance pandemic prevention capabilities. [3]. In urban environments, certain common factors affected the transmission of infectious diseases, including population density, hygienic conditions, and transportation environments. High population density and frequent interaction of

* Corresponding author at: Department of Urban Planning, Room 203, Wenyuan Bulding, No. 1239, Siping Road, Shanghai, 200092, PR China.
E-mail addresses: chooong@tongji.edu.cn (P. Chong), liuchao1020@tongji.edu.cn (C. Liu), wchung@tongji.edu.cn (W. Chung).
1 Room 203, Wenyuan Bulding, No. 1239, Siping Road, Shanghai, 200092, China;

https://doi.org/10.1016/j.buildenv.2024.112330
Received 13 May 2024; Received in revised form 17 November 2024; Accepted 18 November 2024
Available online 19 November 2024
0360-1323/

people in places such as housing, public transportation, and offices, made it easy for pathogens to spread [4]. While influxes of migrants led to poor housing conditions and inadequate sanitation facilities, increasing the risk of pandemic outbreaks [5]. Enhancing the resilience of urban environments to public health emergencies necessitates strengthening not only the built environment but also the socioeconomic and cultural domains.

Neighborhood constituted the basic unit of the cities, and the level of respiratory infection prevention and control at the neighborhood-level affected the overall effectiveness of pandemic prevention and control in cities. [6]. Neighborhood resilience referred to the ability of neighborhoods to rapidly adapt, recover, and develop in response to disasters or diseases. During the COVID-19 pandemic, neighborhood resilience could effectively prevent outbreaks, with rapidly responded as the pandemic evolved to minimize cases, and planned with strategies tailored for the post-pandemic period. Based on general resilience principles, relatively independent multifunctional neighborhoods had advantages in enhancing neighborhood resilience [7]. Neighborhoods possessed a degree of self-sufficiency, enabling them to meet the residents' needs to some extent without being overly affected by external changes. Moreover, the multi-functionality of neighborhoods enabled them to better address diverse challenges, enhancing adaptability and flexibility. Additionally, during the pandemic, residents needed or were forced to conduct basic activities such as shopping and recreation at home or in the neighborhood, making neighborhoods better equipped to deal with special periods and policies [8].

Multiple studies have suggested that COVID-19 infection rates were elevated in areas such as slums and ethnic minority communities. This trend was primarily attributed to factors including aging infrastructure, high population density, and low-income levels. [6,9]. In China, a similar scenario occurred with older neighborhoods, constructed before the economic reforms, often lagging in keeping pace with modern developments. Older neighborhoods presented various issues such as small living spaces, shared kitchens and bathrooms, and poor living conditions and hygiene environments, potentially leading to significant risks of virus transmission. Therefore, researching older neighborhoods helps to fill in the existing research gaps regarding spatial factors and infection rates in China.

This study aims to analyze the relationship between COVID-19 prevalence and neighborhoods in the central areas of Shanghai, exploring the key factors affecting neighborhood resilience to guide the renewal of older neighborhoods. Additionally, it provides a dataset of COVID-19 samples from neighborhoods in the central urban areas of megacities in China, thereby addressing a data gap.

2. Literature review

2.1. Neighborhood resilience regarding COVID-19

Neighborhood resilience refers to the ability of a community to endure, adapt, and thrive in the face of various challenges, including natural disasters, economic shifts, and social changes [10]. This resilience is built upon several critical factors that contribute to the overall robustness and adaptability of the community [11].

Following the outbreak of the COVID-19 pandemic, research on enhancing resilience began to focus on preventing epidemic impacts. Identifying the influencing factors is crucial for improving neighborhood resilience. The infection rates of diseases could vary significantly among different neighborhoods within the same city, due to factors such as built environment, socioeconomic status, and population composition. In terms of the built environment, spatial density-related factors including population density, distribution of points of interest, and land use, significantly impacted disease transmission [12]. Regarding socioeconomic factors, income and occupation played crucial roles in determining infection rates [13]. High-income individuals reduced the infection rates during pandemics by choosing to work from home or commute by private vehicles [14]. For population composition, family structure has a significant impact on COVID-19 transmission, with crowded households exacerbating pandemic spread [15]. Moreover, neighborhoods with a higher proportion of elderly and low-income populations experienced an increasing rate of infection risks. In general, these studies provided evidence that the built environment, socioeconomic status, and population composition significantly affected COVID-19 infection rates in different neighborhoods.

Examining the factors influencing infectious disease transmission solely from the perspective of neighborhood resilience was not comprehensive, as it lacked an analysis of spatial environment and neighborhood behaviors from the viewpoint of disease transmission. Therefore, we further investigated the potential impacts of the built environment, socioeconomic status, and population composition on transmission from the mechanism of infectious disease spread.

2.2. affecting neighborhood resilience examined with SEIR model

The study concluded that understanding the dynamics and mechanisms of the transmission of respiratory infections is a key basis for identifying influencing factors. The SEIR model, transmission dynamics model widely embraced in the research field, is derived from the SIR model. The SEIR model, among the most frequently employed mathematical models for elucidating disease transmission, could act in predicting transmission scenarios and evaluating intervention effectiveness [16]. Following the outbreaks of SARS and COVID-19, many scholars used the SEIR model to predict transmission rates, vaccine efficacy, etc. [13,17]. However, there were a few studies that employed the SEIR modeling to study the factors influencing the spread of infectious diseases, especially at the urban environment level.

Transmission of respiratory infectious diseases is a complex systemic process involving various aspects such as populations and environments. Research analyzing cities based on the SEIR model allows for the delineation of relationships between disease transmission and built environment and population, deducing factors influencing disease transmission in cities. These factors could be categorized into population-level and spatial-level factors. At the population level, the study focused on factors such as population characteristics, socioeconomic and infection status. The quality of the population and economic conditions affected the probability of susceptible individuals and recovered patients contacting and infecting the virus [12]. The infection status of individuals or contacts also affected the spread of the virus. On the spatial level, key elements included the urban environment, neighborhood environment, and household environment. Urban environmental factors included land use types and transportation networks, which had significant impacts on disease transmission. Neighborhood environmental factors included green space ratio and building density [18,19], which indirectly affected residents' health by influencing physical activity and social interaction [20]. Household environmental factors included independent toilets, kitchens, and the number of rooms [21,22], which also influenced the transmission of the disease.

2.3. Current research at the neighborhood scale on infectious diseases

Current research at the neighborhood scale on infectious diseases is abundant. Fewer studies have considered the interaction of factors within cities. Therefore, it is necessary to study multiple factors that influence infection transmission rates within the neighborhood scale simultaneously. Table 1 summarized the current research at the neighborhood scale, including regions in North America, East Asia, and Europe. Typically, studies focused on communities within one or more cities or states, reflecting the geographical diversity. In North America, mainly in the United States, utilized data focusing on socioeconomic factors, built environment, and population characteristics, particularly in minority neighborhoods, owing to data availability [23,24]. Research in East Asia, predominantly in India, included socioeconomic and built

Table 1
Recent articles on neighborhood-level studies of elements of impact on COVID-19.

Source	Research Factors	Analysis method	Major findings	Study area
[25]	Type of dwelling, age, sex, family size, years of education, floor space per capita, shared toilets, etc.	Logistic regression, linear regression	Despite crowded living conditions having facilitated widespread transmission, the variability in prevalence in localities that were in geographical proximity indicates a heterogenous spread of infection	Pune, India
[27]	Social Vulnerability Index (SVI): socioeconomic status, household composition, and disability, racial or ethnic minority status and language, and housing type and transportation	Descriptive and Poisson regression analyses	People living in the poorest neighborhoods have a nearly 40% higher risk of COVID-19 infection	Louisiana, USA
[28]	Individual: age, sex, race/ethnicity, preferred language, insurance ensus tract-level: demographics, insurance, income, education, employment, occupation, household crowding, and occupancy, built home environment, and transportation	linear mixed model	At the individual level, age, sex, race, language, and type of insurance were associated with the likelihood of SARS-CoV-2 infection; at the census tract level, population density, household occupancy, and education level were also associated with the likelihood of SARS-CoV-2 infection	Massachusetts, USA
[29]	Regional deprivation index (ADI), population size, age, gender, and ethnic distribution	Descriptive analysis, correlation analysis	Socioeconomic characteristics of communities seemed to be associated with their susceptibility to COVID-19.	Arizona, Florida, Illinois, Maryland, North Carolina, South Carolina Virginia, USA
[24]	Built environment: number of dwelling units per building and average assessed value (per square foot) Economic status of the neighborhood: median household income, poverty rate, unemployment rate, population density, household members (number per household), and household crowding (percentage of households with >1 person per room)	logistic regression model with two variables	SARS-CoV-2 transmission among pregnant women in New York City has been associated with large numbers of family members at the neighborhood and building levels, household crowding, and markers of low socioeconomic status	New York City, USA
[15]	Educational disadvantage, unemployment, overcrowded housing, mobility and population density	Multilevel logistic regression model	There was a pattern of socioeconomic inequality in the pandemic, and living in areas characterized by social and economic disadvantage increased the risk of transmission	Milan & Lodi, Italy
[26]	Indicators of socioeconomic deprivation: occupation, education, median income, median rent, unemployment rate and nationality, etc	Spatio-temporal clustering, regression modeling	The significantly longer duration of SARS-CoV-2 clustering in socioeconomically disadvantaged communities may also contribute to the increased risk of infection in disadvantaged individuals	Geneva, Switzerland
[12]	Socio-demographic data: race, gender, age, education, income, unemployment, etc. Travel behavior data: cars, trucks, vans, cabs, buses, motorcycles, subway/elevated rail, bicycling and walking Built environment data: street connectivity, regional auto center index, land use and diversity	Poisson regression, spatial correlation, descriptive statistics	Auto-oriented built environment design (greater auto accessibility) is positively correlated with COVID-19 fatality rate. Sedentary (auto) travel is associated with a greater COVID-19 fatality rate	Washington, D.C., USA
[30]	Census data,Residential buildings and food access data,Mobility and transit data	BWQS regression analysis	A significant association between social disadvantage and new crown pneumonia infection and mortality rates	New York City, USA
[14]	Housing quality, living conditions, travel patterns, race/ethnicity, household income	Multivariate regression model	Combined architectural and social environment variables were the strongest and most significant predictors of COVID-19 deaths. Congestion rate resulted in the most significant effect, followed by work commute time and the percentage of African Americans.	Washington, D.C., USA
[31]	Social Vulnerability Index (SVI): socioeconomic status, household composition, and disability, racial or ethnic minority status and language, and housing type and transportation	Multivariate negative binomial regression	Overall social vulnerability and vulnerability themes significantly associated with increased COVID-19 case rates	Alabama & Louisiana, USA
[6]	Spatial resilience,Capital resilience,Social resilience,Governance resilience	Qualitative Comparative Analysis (QCA)	Vulnerable, alienated, and inefficient communities are three types of communities that are less resilient to risk	Wuhan,China
[32]	Social Vulnerability Index (SVI): socioeconomic status, household composition and disability, racial or ethnic minority status and language, and housing type and transportation	Mixed effects logistic regression model	COVID-19 hospitalized patients from socially disadvantaged communities showed greater disease severity	Michigan, USA
[13]	Social Vulnerability Index (SVI): socioeconomic status, household composition, and disability, racial or ethnic minority status and language, and housing type and transportation	Pearson correlation analysis, distributed lag nonlinear models, standard two-stage meta-analytic models, machine learning	Environmental factors (e.g., population mobility, temperature, and relative humidity) have different effects on the spread of COVID-19	Brazil, Brazil

environment data [6,25]. Conversely, research in Europe were relatively scarce, primarily discussing the impact of socioeconomic factors on disease transmission [15,26]. Overall, there was a lack of comprehensive studies considering factors such as built environment, socioeconomic status, and population structure, while a few cases studies on mainland Chinese cities.

Studies conducted at the neighborhood scale have employed various methods to assess factors influencing infectious diseases, which can be categorized into three types. The first type comprised descriptive statistical models, including principal component analysis or Pearson

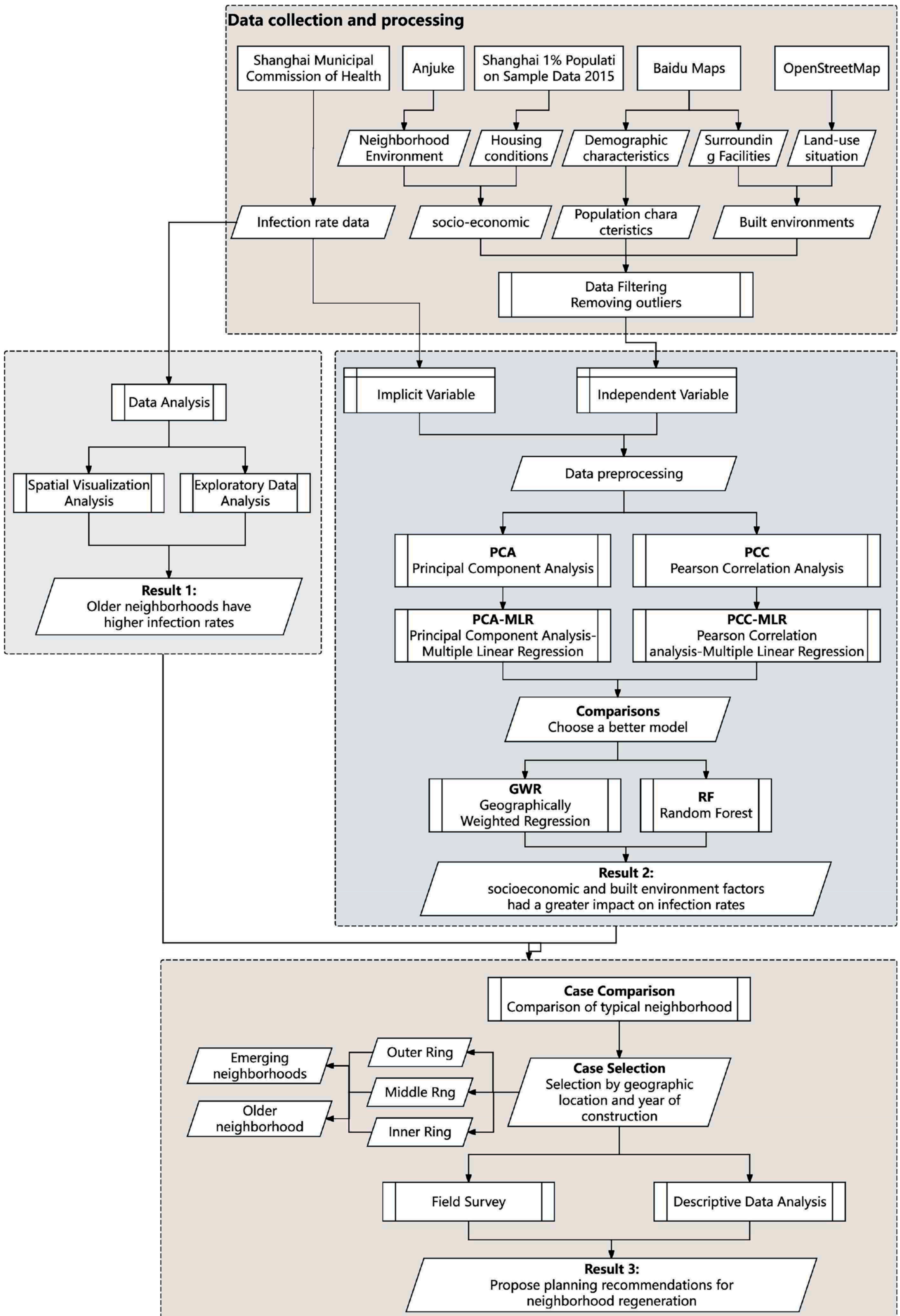


Fig. 1. Technical route.

analysis, multiple regression analysis, and Poisson regression analysis [14,30]. The second type involved spatial statistical models, including spatial regression models and geographically weighted regression (GWR) [33]. The third type encompassed machine learning methods, such as random forests and decision trees (Cazzolla [34]). In urban and environmental research, multivariate statistical methods extract valuable information from extensive urban data to identify key factors [35].

PCC(Pearson Correlation Analysis) and PCA(Principal Component Analysis) are commonly used tools for data preprocessing in model processing [36]. PCC can identify and eliminate variables that are highly correlated and redundant with each other, thereby improving the efficiency and accuracy of the model [37]. PCA can identify variables that account for the largest variance in the data, which typically make a significant contribution to enhancing the accuracy of predictive models [38,39].

Descriptive statistical models can capture the contribution of various influencing factors. However, both the PCC-MLR(Pearson Correlation Analysis-Multiple Linear Regression) and PCA-MLR (Principal Component Analysis-Multiple Linear Regression)models operated under the assumption of linear relationships between variables. [40]. The relationship between various variables in the city and disease infection rates may resulted in a more complex non-linear relationship, therefore, a machine learning approach was needed to accurately capture the complex associations [41].

Random forest is a commonly used method in machine learning modeling, widely applied in research related to urban planning [42,43]. This technique is a data mining practice based on random sampling learning, incorporating random feature selection. Random forests offer several advantages, including high accuracy, avoidance of overfitting, and moderate complexity [44]. Additionally, they can internally generate out-of-bag accuracy estimates to assess model performance and return evaluations of the importance of input variables.

Despite numerous studies on infectious disease transmission, rare studies had considered the interaction of multiple factors within cities, and fewer had used large cities in mainland China as case studies. To fill the gap in neighborhood-level infectious disease research, this study aimed to identify the neighborhood-scale factors influencing COVID-19 transmission by employing various statistical methods to explore infection patterns and influencing factors at the neighborhood level. This study addressed two key questions: whether morbidity is higher in older neighborhoods? Second, what are the factors that influence the spread of infectious diseases at the neighborhood level? The findings were crucial for developing effective neighborhood renewal and pandemic prevention policies, providing references for renewal strategies in older neighborhoods. Furthermore, the methods used in this study could be applied to other cities similar to Shanghai worldwide, aiding in the formulation of effective response policies and the coherent allocation of urban resources.

3. Materials and methods

This study applied a multi-method approach to explore infection patterns and influencing factors at the neighborhood level, and the research process is outlined in Fig 1.

For the first research question, whether morbidity is higher in older neighborhoods, the investigation focused on whether morbidity rates are higher in older neighborhoods. Data visualization and correlation analysis were employed to systematically compare infection rates across neighborhoods constructed during different periods. This approach facilitated the identification of spatial clustering trends and outbreaks within the study area.

For the second research question, the study examined factors influencing the spread of infectious diseases at the neighborhood level, focusing on socioeconomic, built environment, and population characteristics. Principal Component Analysis (PCA) and Pearson Correlation Coefficient (PCC) were used to analyze these factors. Multiple Linear Regression (MLR) models were then employed to assess linear relationships between these factors and infection rates within the research area.

In summary, this study employed these two methods for data processing in conducting multiple linear regression analysis. PCC-MLR and PCA-MLR are capable of thoroughly exploring the linear relationships between variables. Considering the effect of spatial autocorrelation, this study used GWR to study the effect of spatial geographic factors and visualized the results of the GWR model. Additionally, to further investigate the nonlinear relationships between variables, the study employed RF methods for analysis. Case comparisons of typical neighborhoods were made to corroborate model results and validate study findings, identifying key factors such as population density, road network density, land use mix, income levels, and housing prices. The detailed research methodology and data processing methods were shown in the Chapter 3.3.

3.1. Study area

Downtown Shanghai, China, was chosen as the research area for two reasons: (1) Shanghai, as an international metropolis with a dense population and developed economy, possesses unique urban characteristics and complex population mobility patterns, making it an ideal subject for study. (2) Shanghai has relatively abundant data, compared to other cities, providing a reliable foundation for research.

The urban structure of Shanghai is delineated into three zones based on ring roads: the Outer Ring, Middle Ring, and Inner Ring areas. The Outer Ring connects urban and suburban areas, promoting satellite town growth with its abundant land and lower housing prices, characterized by industrial parks and logistics centers. The Middle Ring, developed from the 1980s onwards, features residential neighborhoods, business districts, and industrial parks, with lower housing prices and spacious living environments. The Inner Ring, the earliest developed zone, forms the city's core with dense skyscrapers, prosperous commercial activities, and higher housing prices. These three rings collectively foster Shanghai's coordinated growth. The study area of this research is located in downtown Shanghai, specifically within the area bounded by the Outer Ring of Shanghai. Due to the lack of available data for Yangpu District during the pandemic, Yangpu District is not considered in this study.

As shown in Fig. 2, there are 2255 neighborhood1 in the study area. In this study, older neighborhoods referred to those built between year of 1960 to 1999 (before the prevalence of commercial housing), while emerging neighborhoods referred to those built between year of 2010 to 2019.

The study period ranged from March 1st to May 31st, 2022, more than 11,000 cases were reported during the study period. The development of the pandemic in Shanghai could be summarized in four phases (Fig 3), The incubation period of the pandemic (March 1st −31st), the outbreak stage of the pandemic (April 1st −28th), the pandemic control and management stage (April 29th -May 31st), and the period of returning to normal life (June 1st -present). During the outbreak stage, the number of infections steadily increased, and the data was more spontaneous because citywide nucleic acid screening was not conducted. By April 1st, there was a cumulative total of 1809 locally confirmed infections and 41,384 asymptomatic infections. During the control and management stage, the pandemic was further brought under control, normalized prevention and control were gradually promoted, the number of newly diagnosed cases continued to decline, and some districts and counties achieved a social surface of zero cases. On June 1st, work resumed at full capacity in Shanghai, and the situation of the pandemic was stable to favorable. Given the decisive role these three phases played in the evolution of the pandemic, they were chosen as the focus of this study.

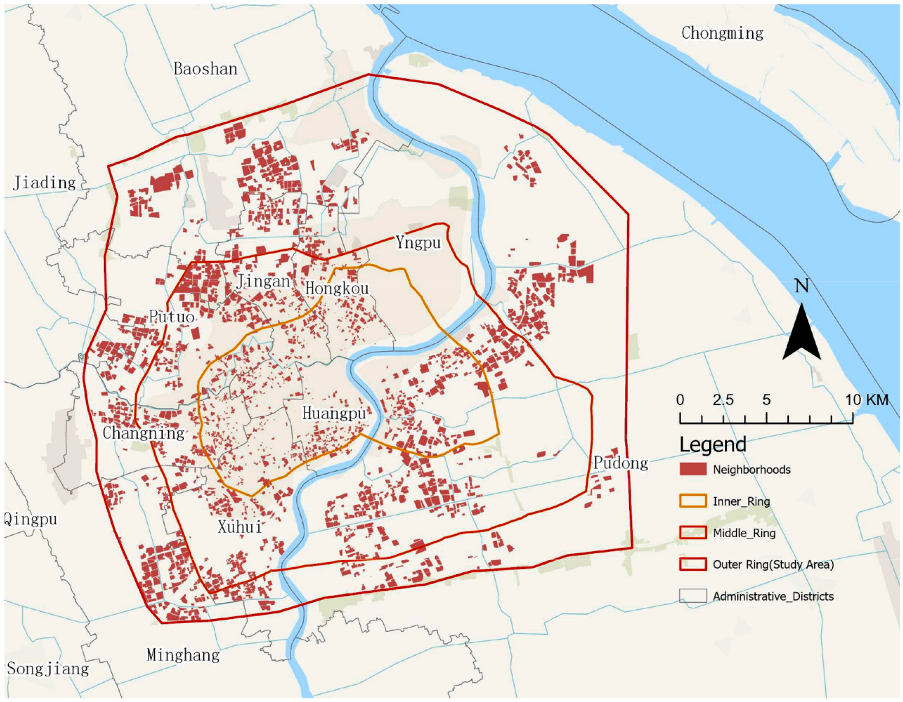


Fig. 2. Map of the scope of the study neighborhood in downtown Shanghai.

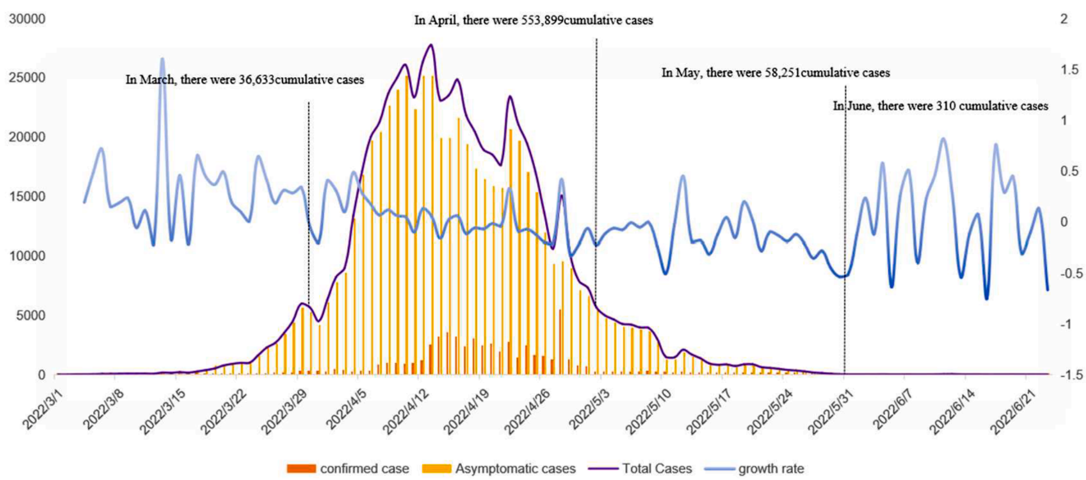


Fig. 3. Incidence of COVID-19 in Shanghai.

3.2. Modeling methods

This study employed a variety of modeling methods, including data processing methods, linear regression models spatial analysis methods, and machine learning methods. The study used a simple linear regression and plotted box plots to explore the relationship between the year of building at each neighborhood and the rate of infection in the neighborhood. Data visualization visualized a comparative assessment of infection trends in different neighborhoods and revealed spatial clustering patterns and outbreaks. Correlation analysis and box plots assessed whether older neighborhoods were more susceptible to infection than newer neighborhoods. The study found that the year the neighborhood was built was negatively related to the infection rate, and older neighborhoods were more susceptible to infectious diseases to newer neighborhoods.

To further explore the relationship between other variables and infection rate, the independent variables were processed by using PCA and PCC, and two sets of models were constructed: PCA-MLR model and PCC-MLR. Meanwhile, the independent variables included demographic characteristics, built environment, and socioeconomics, and the dependent variables, which referred to the number of cases, per capita incidence rate, and geographic average incidence rate, respectively. By comparing the results of PCA-MLR and PCC-MLR, the study found that PCC-MLR had a better degree of explanation.

Based on the results of PCC-MLR, the study built a GWR model to visualize the spatial and temporal distribution of the results, and considering the complexity of the city and the possible nonlinear relationship between the influencing elements and the infection rate, the study also introduced a RF for analysis. Random forest, widely used in urban planning research, is a data mining technique that offers high accuracy, avoids overfitting, and evaluates input variable importance, making it ideal for studying factors impacting urban areas.

In addition, typical neighborhoods were selected for case comparisons to explore in depth whether there were specific factors affecting their incidence rates, and to further validate the reliability of the study results, corroborating with the findings from the model analysis. Through these analyses, the study identified several key factors that influence neighborhood infection rates, including population density, road network density, land mix (land mixing degree), number of high-income people, and housing prices.

3.3. Model setting

3.3.1. Dependent variable

The dependent variables included the total number of cases in neighborhoods, per capita infection rate, and per area infection rate. The total number of cases was the cumulative number of positive reports in the neighborhood during the study period. The per capita infection rate was calculated using the total number of people in the plot as the denominator and the cumulative number of people reporting positive as the numerator. The per-area infection rate was calculated using the total area of the neighborhood as the denominator and the cumulative number of reported positive cases as the numerator.

The data were sourced from the official WeChat account of the General Office of the Shanghai Municipal People's Government and reports published by Shanghai.These reports provided daily updates on the number of new infections and the residential addresses of infected individuals. The study used Baidu Maps to convert these geographic addresses into spatial coordinates, obtaining spatial locations for each infected individual. Subsequently, the total number of infected people in each neighborhood within the Outer Ring were calculated over a three-month period.The data of the population living in the neighborhood was generated using Baidu Maps Urban Population Geographic Big Data Platform.

3.3.2. Independent variables

Based on the SEIR model and previous studies, three categories of indicators were screened, demographic characteristics, socioeconomic, and built environment, where the built environment includes the urban environment, neighborhood environment, and household environment. Indicators related to demographic characteristics and the built environment were related to the probability of infection in the population,

Table 2
Types of variables used in the study.

Category	Source	Name	Meaning	Mean	STD
Built Environ-ment	OSM	Road_Densi	Road Network Density	0.11	0.04
		G_area	Green Space Area	4.43	506,681.75
		SiteMix	Land Mixing Degree	0.8	0.25
		FAR	Floor Area Ratio	2.46	1.46
		build_area	Building area	0.06	48,014.86
		AREA	Land Area	0.03	24,557.92
		Medical_Services	Number of medical service facilities	50	28.77
		Leisure	Number of leisure facilities	195	88.9
		Education	Number of educational facilities	12	4.84
		Shopping	Number of shopping facilities	520	288.53
		Public_Tranportation	Number of public transport facilities	33	7.27
		Metro	Number of subway facilities	2	1.42
		Dining	Number of Food and Beverage Establishments	271	151.87
Socioeco-nomic	Baidu Maps Urban Population Geography Big Data Platform,anjuke,	Income_Lv_2499	The monthly income is less than 2499 yuan	40	37.3
		Income_Lv_20,000	The monthly income is more than 20,000 yuan	143	124.69
		Low_consumption	Low consumption level	364	305.45
		High_consumption	High consumption level	372	303.79
		Owns_Car	Own a car	419	420.11
		POPdens	Population density	32,240	0.01
	Anjuke website	T_price	Total Property Price	707.6	477.96
		P_price	Price per Square Meter of Housing	75,157	25,040.39
		floor	Number of Floors	11	8.08
		room	Number of Rooms	4	1.51
	Shanghai 1% Population Sample Data 2015	Kitchen	Have a separate kitchen (2), shared kitchen (1), no kitchen (0)		
		Toilet	Have a separate toilet (2), share a toilet with others (1), no toilet (0)		
Populati-on Character-istics	Baidu Maps Urban Population Geography Big Data Platform	F_18	Number of Individuals Under the Age of 18	10	9.7
		F55_64	Number of Individuals aged 55–64	104	91.65
		F_65	Number of Individuals Over the Age of 65	72	62.86
		High_School_Below	Education Level High School or Below	622	511.39
		Bachelor_s_Degree__above	Education Level Bachelor's Degree or Higher	126	115.08
		OCH	Occupational Heterogeneity	1.19	0.06
		High_School_Student	Number of Individuals in High School Stage	49	43.78
		College_Student	Number of Individuals in College Stage	20	19.83
		Graduate_Student	Number of Individuals in Graduate School Stage	4	6.2
		Pregnancy_Period	Number of Individuals in Pregnancy Stage	24	24.24
		Parenting_Stage	Number of Individuals in Child-Rearing Stage	1	2.69
		Pregnant_at_Home	Number of Individuals with Pregnant Women at Home	16	16.91
		Has0_1YearOld	Number of Individuals with Children Aged 0–1 at Home	51	47.44
		Has1_3YearsOld	Number of Individuals with Children Aged 1–3 at Home	32	31.5
		Has3_6YearsOld	Number of Individuals with Children Aged 3–6 at Home	85	75.28
		Has_Young_Children	Number of Individuals with Elementary School Students at Home	341	286.12
		Has_Elementary_Children	Number of Individuals with Middle School Students at Home	119	103.59
		Has_HighSchool_Children	Number of Individuals with High School Students at Home	69	59.62

and indicators related to the built environment were related to the spread of the virus.

Built environment data including vector road network, neighborhood contour data, and green space data were from OpenStreetMap, and facility POI data such as medical services were from 2020 Baidu map web crawling. Socioeconomic data including the number of rooms, number of floors, and house prices were from the 2020 Anjuke website, kitchen and toilet ownership were from the Shanghai 2015 1% population sample data, income level and consumption level were from the Baidu Maps Urban Population Geo-Big Data Platform. Data on demographic characteristics including age, education, and occupational composition were from the Baidu Data Urban Population Geography Big Data Platform. The variables were shown in Table 2.

3.3.3. Data pre-processing

First, in order to investigate the relationship between the age at which the neighborhood was built and the infection rate, the data were grouped according to the age at which the neighborhood was built. Each decade was grouped into six groups of neighborhoods with completion years from 1960 to 2019, thus observing the infections in neighborhoods with different completion years.

Secondly, the data were screened to ensure the accuracy of data analysis. After the correlation was found between the age of completion and the infection rate, modeling analysis was conducted in order to explore the influencing factors affecting the infection situation in the neighborhood. In order to guarantee the accuracy of the data, the variable data were tested for normal distribution, data other than three times the standard deviation were filtered to remove extreme values, and the dependent variable was logarithmically processed, and finally 1699 samples were retained to construct the model. In the subsequent research model, in order to improve the accuracy and explanatory power of the model, the built-up era element was excluded and attention was focused on other potential factors such as demographic characteristics, socioeconomic factors and built environment.

The study adopted different variable screening methods for different models. For the PCA-MLR, the study obtained a total of seven new components by PCA, and the new components were renamed according to their main constituents (the first five) for clear discrimination. As shown in Table 3 For the PCC-MLR and RF, the study analyzed the correlation of variables to prevent multicollinearity. When the PCC value was greater than 0.8, the variable see was considered to be strongly correlated and only one comparison variable was retained. After filtering, 19 variables were selected for the final model, as shown in Fig. 4. For the RF model, the study selected 80% of the data as the training dataset, while the remaining 20% was used as the test dataset to verify the model fit by comparing the R^2 and Mean Square Error (MSE). On the training set, $R^2 = 0.95$, MSE = 0.071. On the test set, $R^2 = 0.63$, MSE = 0.506. The model was considered to be a good fit, and on the test set there was an error, but it was within the acceptable range, so the model was able to capture the features of the training data.

In order to identify the independent variables that have a significant effect on the dependent variable, we used forward stepwise multiple regression. In order to obtain the best model where all variables have a significant effect on the dependent variable, we added variables one by one. F-tests were performed on the new variables introduced, while the existing variables were subjected to *t*-tests, and the variables that would no longer be significant would be eliminated. This process was repeated until there were no significant variables to model and all non-significant independent variables had been eliminated from the regression equation. The study used SPSS, GWR4, and python for data cleaning and modeling, and Geographic Information System (GIS) was used to analyze and visualize the results.

Table 3
Selected variables for PCA-MLR modeling.

Name	variant
Factor 1: Socioeconomic	High_School_Below,Owns_Car,Low_consumption, High_consumption,Has_Elementary_Children
Factor 2: POI	Shopping,Dining,Leisure,Medical_Services, Road_Densi
Factor 3: Income level	T_price,room,P_price,build_area,POPdens
Factor 4: Kitchen and Toilet	Toilet,Kitchen,Metro,Public_Transportation, Shopping
Factor 5: Public Transportation	Public_Transportation,Education,POPdens,FAR, floor
Factor 6: Neighborhood Environment	G_area,FAR,floor,OCH,POPdens
Factor 7: Urban Mix	Metro,SiteMix,Graduate_Student,OCH,room,P_price

4. Results

4.1. Distribution of COVID-19 neighborhoods

According to the COVID-19 dataset of Shanghai, as of May 30th, 2022, the cumulative number of confirmed positive cases exceeded 70,000 cases. The map of the year of built of the neighborhoods and the map of neighborhood infection rates (Fig. 5) showed that neighborhoods generally concentrated in the central Huangpu, Jing'an, Xuhui and Hongkou districts, while neighborhood's within the peripheral administrative districts of Putuo, Changning, Minhang and Baoshan districts had lower infection rates. Neighborhoods along the river also had higher infection rates, and most of the neighborhoods with high infection rates had smaller land areas.

Linear regression analyses were performed on the year of plot construction and per area infection rate. The adjusted $R^2 = 0.104$ and standardized coefficient of -3.23 showed that the year of plot construction was negatively correlated with the infection rate, indicating that the older the building, the higher the infection rate in the plot.

Further analysis using box plots (Fig. 6) showed that older neighborhoods were more susceptible to infectious diseases relative to emerging neighborhoods, and the data of older neighborhoods have greater volatility. On the one hand, the median of the older neighborhoods resulted higher rate than the emerging neighborhoods in all three data sets, which indicated that the older neighborhoods were more impacted by infectious diseases than the emerging neighborhoods in three aspects: total cases, per capita infection rate, and per area infection rate. Emerging neighborhoods have a higher concentrated total, per capita infection rate, and per area infection rate, while older neighborhoods have longer box lengths. The box plots suggested greater volatility and more extremes in the data for older neighborhoods, which reflected the greater challenges and uncertainties faced by older neighborhoods in responding to outbreaks.

4.2. PCA-MLR modeling results

As shown in Table 4, in Model 1–1, where the dependent variable was the total number of cases in the neighborhood, the model included six factors and the adjusted R^2 was 0.281. In Model 1–2, where the dependent variable was the per capita infection rate in the neighborhood, the model included four factors and the adjusted R^2 was 0.510. In Model 1–3, where the dependent variable was the per-area infection rate in the neighborhood, the model included six factors and the adjusted R^2 was 0.577. It could be demonstrated that Model 1–3 have the best result. In the overall three models, factors 1, 2, 3, and 7 had positive effects on all three dependent variables simultaneously. Among them, factor 7 has a higher degree of influence on the dependent variable, and factor 1 has the lowest degree of influence.

4.3. PCC-MLR modeling results

The 19 variables data PCC-MLR model and the results are shown in Fig. 4 and Table 5. In Model 2–1, where the dependent variable was the total number of cases in the neighborhood, the model included ten

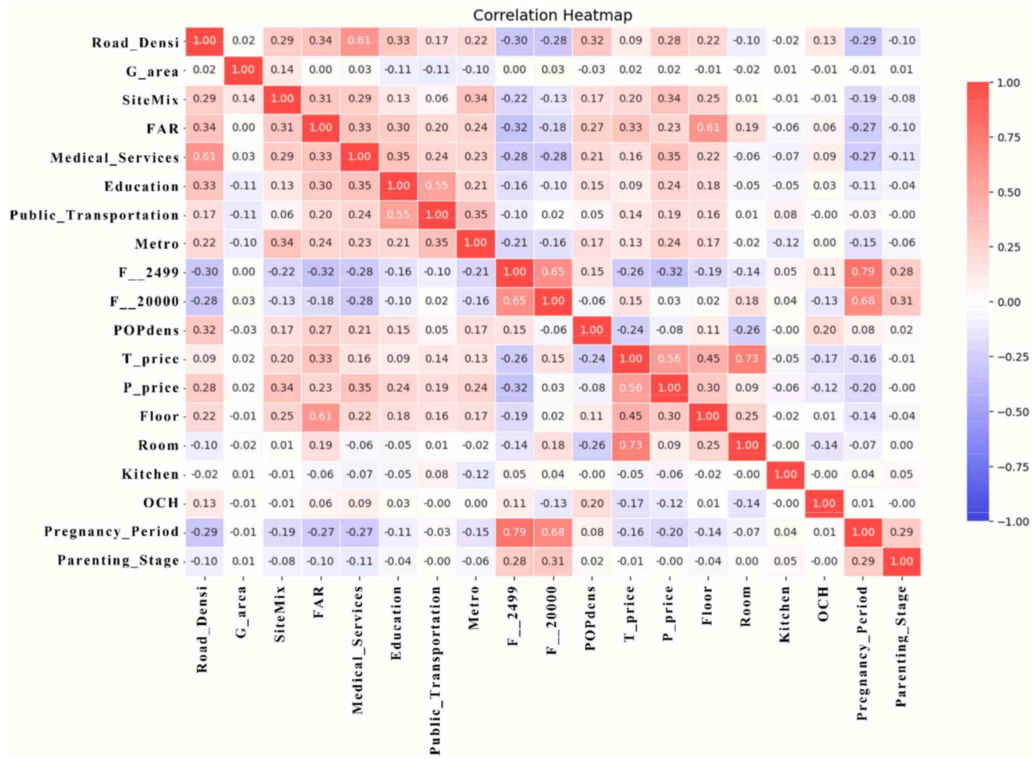


Fig. 4. Results of Pearson analysis.

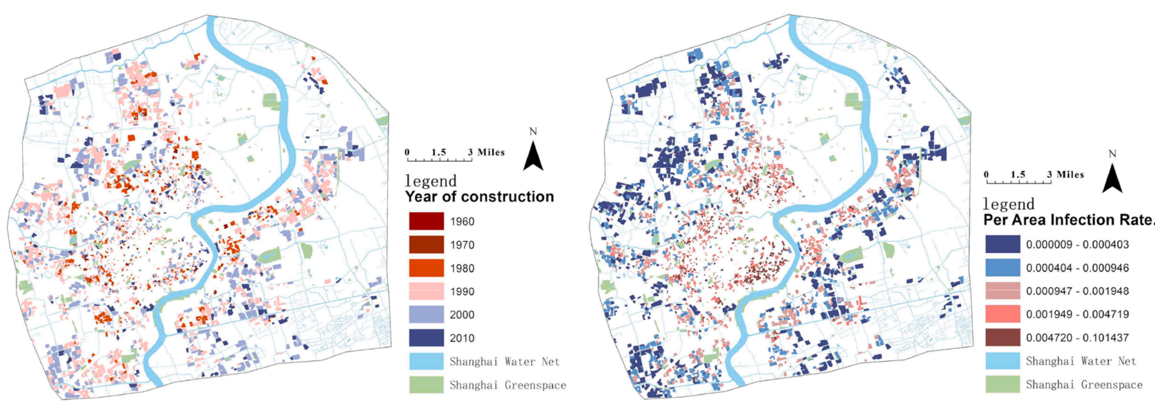


Fig. 5. Shanghai city center by neighborhood (a: Year of construction, b: Neighborhood prevalence rate).

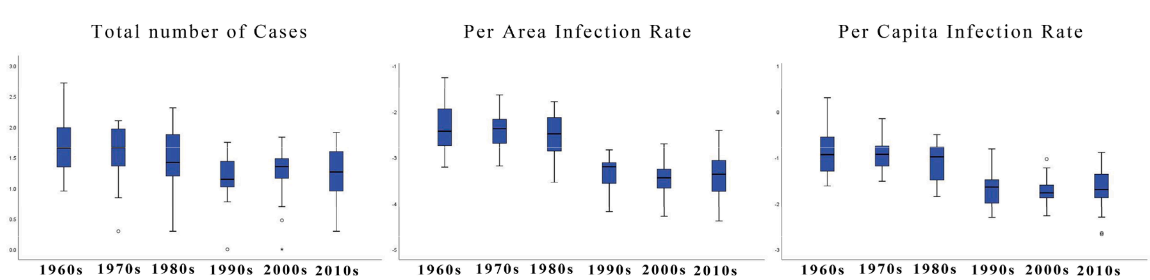


Fig. 6. Infections in neighborhoods by age of completion.

factors, and the adjusted R^2 was 0.340. In Model 2–2, where the dependent variable was the per capita infection rate in the neighborhood, the model included 13 factors, and the adjusted R^2 was 0.545. In Model 2–3, where the dependent variable was the per-area infection rate in the neighborhood, the model included 13 factors, and the adjusted R^2 was 0.624. Overall, nine variables were included in all three models simultaneously. Among these, only the variable floor exhibited a negative correlation with all three dependent variables.

Based on the observations from Table 5, it was revealed population density, the number of subway stations, and the number of people with a

Table 4
PCA-MLR modeling results.

Variable	Model 1–1: Total number of cases		Model 1–2: Per capita infection rate		Model 1–3: per area infection rate	
	Stand	*Sig.*	*Stand*	*Sig.*	*Stand*	*Sig.*
(Constant)	0.018	<0.001	0.019	<0.001	0.02	<0.001
Factor 1: socioeconomic	0.004	<0.001	0.004	<0.001	0.004	<0.001
Factor 2: POI	0.009	<0.001	0.009	<0.001	0.009	<0.001
Factor 3: Income level	0.011	<0.001	0.011	<0.001	0.012	<0.001
Factor 4: Kitchen and Toilet	0.013	0.045				<0.001
Factor 5: Public transportation					0.016	<0.001
Factor 6: Neighborhood y Environment	0.016	<0.001			0.017	<0.001
Factor 7: Urban Mix	0.017	<0.001	0.018	<0.001	0.018	<0.001
Adjusted R square	0.281		0.51		0.577	

monthly income higher than 20,000 yuan have a high correlation with the number of cases and the infection rate, respectively. For Model 2–1, the highest degree of influence was found in population density, followed by the number of metro stations and the number of people with a monthly income of less than 2499 yuan, all of which were positively correlated with the total number of cases. For Model 2–2, the highest level of influence was the number of people with a monthly income higher than 20,000 yuan, which was negatively correlated with the per capita prevalence rate, followed by the number of subway stations, which was positively correlated with the prevalence rate. For Model 2–3, population density and the number of metro stations were positively correlated with the per-area infection rate, and the number of people with a monthly income higher than 20,000 yuan was negatively correlated with the per-area infection rate. In the Model 2–2 and the Model 2–3, the number of subway stations was positively correlated with infection while the number of bus stops was negatively correlated with the infection rate, which might be because of the discontinuation of some bus routes as a result of the policy at the time.

The PCA-MLR and PCC-MLR models shared several indicators. Firstly, socioeconomic factors, especially population density and the price per square metre of housing in the neighbourhood, showed a positive correlation with infection. Second, the number of metro stations in the built environment also increased the infection rate to some extent.

4.4. GWR modeling results

Due to the presence of spatial dependence and spatial autocorrelation, neglecting the spatial relationship in the analysis of COVID-19 would diminish the accuracy of the model. Therefore, to ascertain the spatial distribution of the variables, the study established a GWR model (using first-order QUEEN to construct the spatial weight matrix) based on the findings of the PCC-MLR model. As depicted in Table 6, in Model 3–1, the adjusted R^2 was 0.442 when the dependent variable was the total number of cases in the neighborhood. In Model 3–2, the dependent variable was the per capita infection rate in the neighborhood, and the adjusted R^2 was 0.625. In Model 3–3, the dependent variable was the per-area infection rate in the neighborhood, and the adjusted R^2 was 0.693. Because the GWR model is one coefficient per polygon, these coefficients will have a maximum, minimum, and mean, and the mean and standard deviation of the coefficients are provided in the table.

Combining the previous models, the model had the most significant level of explanation when the dependent variable was the per area infection rate, thus the results of Model 3–3, which contained a total of 13 factors, were visualized and analyzed. As shown in Fig. 7, according to the Standardized Residual (Std. Resid) plot, most communities were within 2.5 standard deviations, indicating that the model fitted better. According to the Local R^2 plot, the low value aggregation area appeared in the center of the region, around the low value aggregation area from the inside to the outside of the trend of increasing and decreasing. In the southern part of the region, the high value aggregation area appeared on both sides of the river. Based on the local R^2 plot, areas with low values were clustered within the inner ring and the outer ring, while the status of clustering was indicated around the middle ring. And high-value clusters were also present in the southern part of Pudong District.

The spatial distribution of the coefficients of the 13 factors included in Model 3–3 is illustrated in Fig. 8, where the population density and road network density shown a significant positive correlation, and the number of people with a monthly income of less than 2499 yuan resulted in a negative correlation with the number of neighborhood floors, total neighborhood house price, average neighborhood house price, and land mix. In addition, the coefficients of the number of subway stations, land mix, and floor area ratio shown large spatial variations.

4.5. RF modeling results

The RF model was chosen to model the rate of infection per unit area as the dependent variable and the independent variables as the 19 variables that passed the correlation test. Five major influencing factors were eventually derived (showen in Fig. 9), including Number of

Table 5
PCC-MLR model results.

Variable	Model 2–1: Total number of cases		Model 2–2: Per capita infection rate		Model 2–3: per unit area infection rate	
	beta	sign	beta	sign	beta	sign
Intercept		0.630		<0.001		<0.001
POPdens	0.290	<0.001	0.043	0.032	0.332	<0.001
Metro	0.288	<0.001	0.229	<0.001	0.210	<0.001
Income_Lv_2499	0.246	<0.001	−0.109	<0.001	−0.089	<0.001
Education	0.142	<0.001	0.155	<0.001	0.146	<0.001
Medical_Services	0.090	<0.001	0.125	<0.001	0.111	<0.001
floor	−0.087	0.001	−0.064	0.004	−0.063	0.002
OCH	0.061	0.004				
Per_Meter_price	0.060	0.014	0.080	<0.001	0.068	0.001
Income_Lv_20,000	0.102	0.001	−0.357	<0.001	−0.312	<0.001
FAR	−0.056	0.046	0.079	0.001	0.077	<0.001
Total_price			−0.071	0.003	−0.083	<0.001
Road_Densi			0.073	0.001	0.070	0.001
SiteMix			0.062	0.001	0.052	0.003
Public_Transportation			−0.048	0.023	−0.044	0.022
Adjusted R square	**0.340**		**0.545**		**0.624**	

Table 6
GWR model results.

Variable	Model 3–1: Total number of cases		Model 3–2: Per capita infection rate		Model 3–3: per area infection rate	
	Mean	STD	Mean	STD	Mean	STD
Intercept	0.368	0.963	−3.722	0.588	−8.233	0.594
POPdens	20.451	6.587	3.189	7.115	34.118	7.519
Metro	0.116	0.114	0.116	0.097	0.127	0.105
Income_Lv_2499	0.008	0.003	−0.002	0.002	−0.002	0.002
Education	0.025	0.026	0.026	0.026	0.028	0.029
Medical_Services	0.005	0.004	0.005	0.005	0.005	0.005
floor	−0.007	0.012	−0.006	0.009	−0.008	0.01
OCH	1.102	0.695				
Per_Meter_price	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Income_Lv_20,000	<0.001	<0.001	−0.004	0.001	−0.004	0.001
FAR	−0.068	0.076	0.041	0.045	0.043	0.05
Total_price			<0.001	<0.001	<0.001	<0.001
Road_Densi			4.26	3.159	4.534	3.216
SiteMix			0.298	0.566	0.27	0.625
Public_Transportation			−0.009	0.018	−0.01	0.018
Adjusted R square	**0.438**		**0.625**		**0.693**	

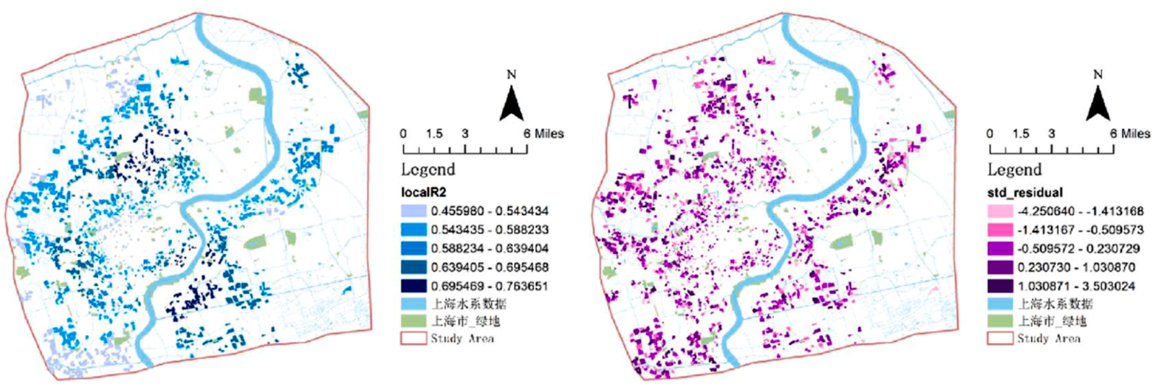


Fig. 7. Visualization of Model 3–3 results.

persons with a monthly income of more than 20,000 yuan, population density, road network density, the number of metro stations around the neighborhood, and the number of healthcare service facilities. The model demonstrated strong concordance, specifically, the R-squared of the model was 0.63, which indicated that the model explained 63% of the variance in the target variables. the Income_Lv_20,000 feature had the highest weight, accounting for 25.89%, and played a critical role in the model. Following closely was the POPdens feature, which accounted for 20.32%, also playing an important role in the model construction. Additionally, the Road_Densi feature had a significant weight of 18.83%. In contrast, the Medical_Services and Metro features had lower weights, accounting for 4.40% and 3.85%, respectively. Nevertheless, these five features collectively accounted for 73.30% of the total weight, significantly impacting the model's construction.

Since the R^2 value of the random forest model was significantly higher than that of the PCA-MLR, PCC-MLR, and GWR models, the study suggested that these RF models offered a better fit, indicating that there might have been nonlinear relationships between the influencing factors and the infection rate. Some variables exhibited similar behavior in both linear and nonlinear models. Population density, road network density, and the number of subway stations showed a significant positive correlation in both linear models and the random forest model. Meanwhile, the number of people with a monthly income exceeding 20,000 yuan and the number of medical facilities around the residential area displayed higher correlations in the random forest model, suggesting that these two variables had a nonlinear relationship with the infection rate.

As shown in the partial dependence plots (PDP) in Fig. 10, the nonlinear relationships between the variables and the infection rate were evident. Variables related to the built environment and socioeconomic factors exhibited significant nonlinear patterns. Among the built environment variables, the PDP curves for Road Density, G_area, and Public Transportation displayed nonlinear trends. Road Density showed a monotonically increasing relationship, indicating that areas with denser road networks, due to improved accessibility, experienced increased human mobility and higher transmission risks. G_area demonstrated a fluctuating pattern, initially decreasing and then increasing, while SiteMix showed a linear upward trend, suggesting a positive effect of mixed land use on the target variable. For socioeconomic variables, Medical Services, Education, POPDens, and T_price exhibited diverse and segmented patterns in their PDPs. The PDP for Income_Lv_20,000 revealed that as the proportion of high-income populations increased, the infection risk declined sharply. In contrast, the PDP for POPDens showed a strong positive correlation, indicating that densely populated areas were high-risk zones for infection transmission.

These results suggested that the effects of these factors on the incidence of infectious diseases were significant at the neighborhood-level in the center of Shanghai. Population density, road network density, and the number of healthcare facilities might reflect the mobility of people, transportation, and the distribution of healthcare resources in the neighborhood, which were closely related to the spread of infectious diseases. In addition, the plot ratio of the neighborhoods and the number of metro stations around the neighborhood reflected the density of buildings and accessibility within the neighborhoods, which also resulted in a significant impact on the neighborhood morbidity rate.

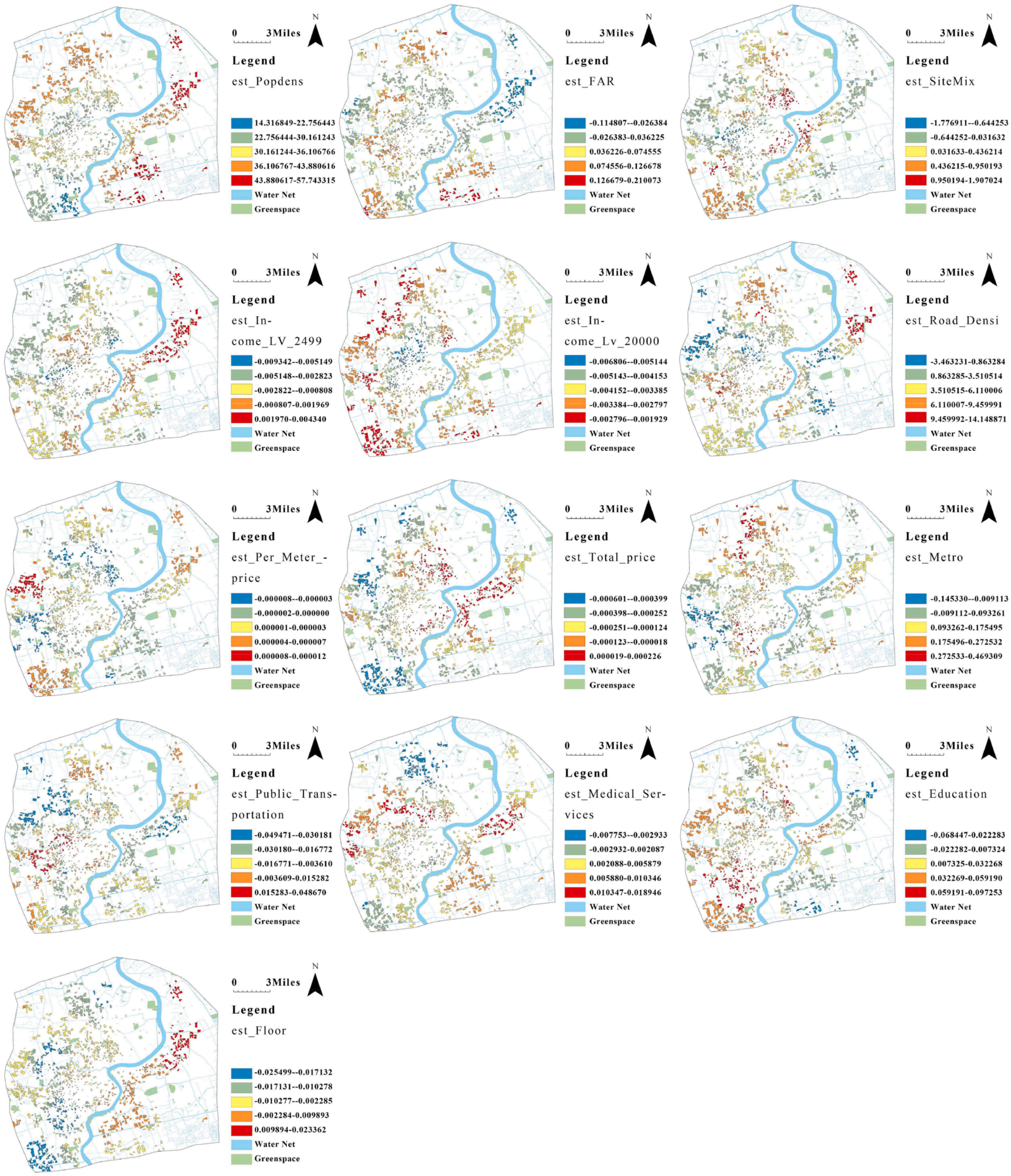


Fig. 8. GWR coefficients for variables in Model 3–3.

4.6. Comparison of typical neighborhoods

In this study, some older and emerging neighborhoods were selected for comparative analysis in each of the outer, middle, and inner rings of Shanghai to further investigate the impact of socioeconomic and demographic characteristics on the transmission of infectious diseases. Each neighborhood selected for comparison was geographically close to each other to ensure the consistency of the surrounding environment and the implemented pandemic prevention policies (shown in Table 7).

After the fieldwork, it was observed that the older neighborhoods generally presented a lower floor structure, usually six floors. The spacing between these buildings was relatively small, resulting in a high density distribution of buildings with relatively little green area or plaza space. The distinguishing feature of this environment was clutter and the frequent movement of people within the neighborhood, which presented an open neighborhood pattern. In contrast, the emerging neighborhoods mainly used high-rise buildings as the main structure, and there was a wealth of green space and activities within the district, providing residents with places for recreation and leisure, but also improving the overall environment of the district. These Emerging neighborhoods also paid higher attention to the rational layout of the main and secondary entrances in planning, so that the entrances and exits of the neighborhood have been clearly and effectively divided.

Comparing the three groups of neighborhoods (Table 8), data

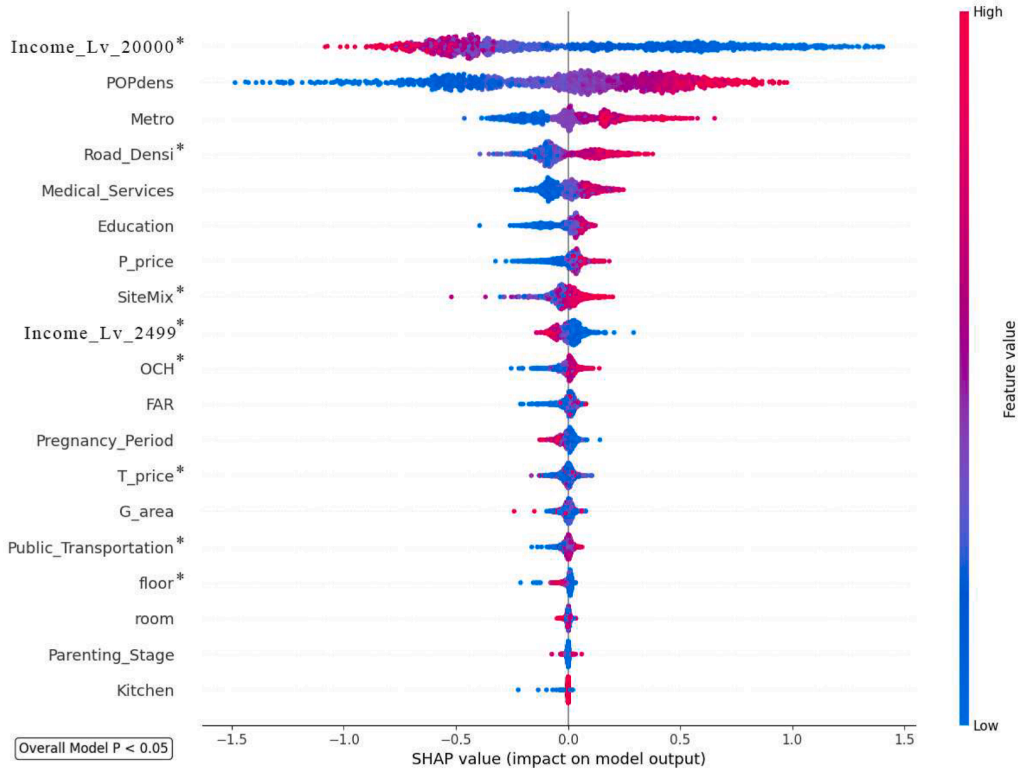


Fig. 9. SHAP value (impact on model output).

analysis revealed that, with similar built environment factors, older neighborhoods had higher total case numbers, cases per unit area, average neighborhood cases, and per capita cases than emerging neighborhoods. Additionally, older neighborhoods located in the middle ring had more severe infection rates compared to those in the inner and outer rings. The results indicated that higher population density, a larger elderly population, and lower educational attainment levels might have increased the risk of disease infection. Socioeconomically, having more rooms and higher household incomes were beneficial in reducing the incidence rate.

5. Discussions

The study initially employed regression models to analyze the relationship between the construction period of neighborhoods and the infection rate of diseases. It was found that there was a negative correlation between the construction period and the infection rate, indicating that older neighborhoods were more susceptible to diseases compared to emerging ones. This result supported research on building age and infection rates, with previous studies finding that historic buildings present higher infection rates than apartments [45]. Complemented by analyses of typical neighborhoods, factors such as the long construction period, fewer rooms, a higher proportion of elderly residents, lower average income, and lower education levels in older neighborhoods contributed to their higher vulnerability to infectious diseases [29,46].

To further explore the factors influencing COVID-19 transmission at the neighborhood level, a range of modeling approaches was employed, including traditional regression models (PCA-MLR, PCC-MLR), a spatial statistical model (GWR), and a machine learning method (Random Forest, RF). The results from these models highlighted key differences in their ability to capture the complexity of the relationships between various neighborhood characteristics and infection rates.The traditional regression models (PCA-MLR and PCC-MLR) identified linear relationships between socioeconomic factors, such as population density and housing prices, and infection rates. However, these models were limited in their ability to capture more complex, nonlinear interactions. In contrast, the Random Forest model, with its significantly higher R^2 value, demonstrated a better fit and revealed nonlinear relationships, such as stronger associations between residents' income and medical facility density with infection rates. The GWR model, which accounts for spatial heterogeneity, provided localized insights into the variation in the strength of relationships across neighborhoods, but still relied on linear assumptions.

Both linear and nonlinear models consistently highlighted the importance of factors such as population density, road network density, and the number of subway stations in explaining infection rates. However, the Random Forest model additionally identified variables such as the proportion of residents earning over 20,000 yuan per month and the density of medical facilities as having stronger and more intricate associations with infection rates. These findings underscore the need for a combination of linear and nonlinear modeling approaches to fully capture the complexity of neighborhood-level factors influencing disease transmission, offering a more comprehensive understanding of the underlying dynamics.

At the level of the built environment, the study revealed that road network density and land-use mix were critical factors affecting disease transmission. Road network density and the number of metro stations were positively correlated with the infection rate, as areas with higher road network density are often located in bustling districts with more public transport stations, leading to increased population mobility and contact, thus elevating the risk of infection [47]. Land-use types influenced residents' activity intensity, frequency, and duration, thereby impacting residents' health [48]. Properly increasing land-use mix, green space, and outdoor recreational areas, optimizing urban land use, and spatial layout were conducive to reducing the risk of infection [33]. An unexpected finding was that a higher number of medical facility points led to a higher infection rate, as individuals tended to visit these places for medical services after infection, thereby increasing the risk of cross-infection. Therefore, when old neighbourhoods are renovated, it is essential to enhance the infrastructure of older neighborhoods. This

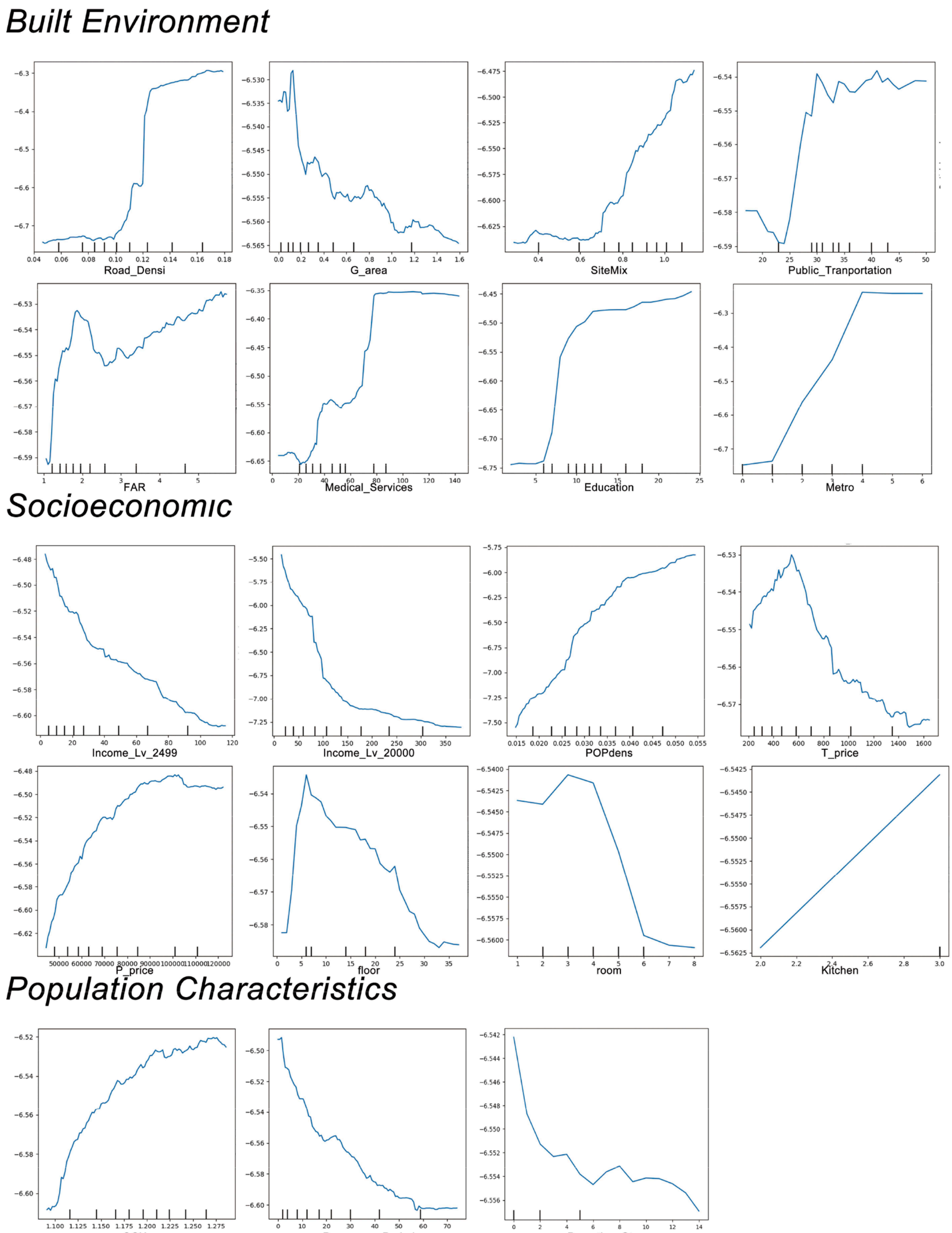
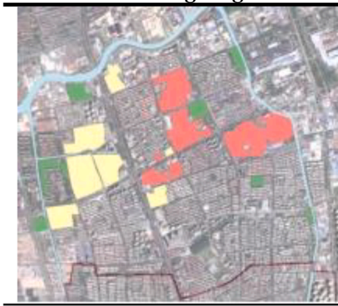
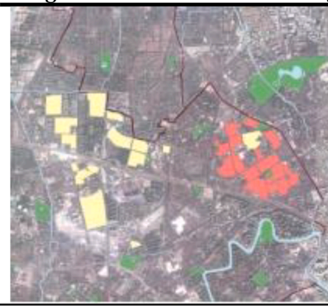
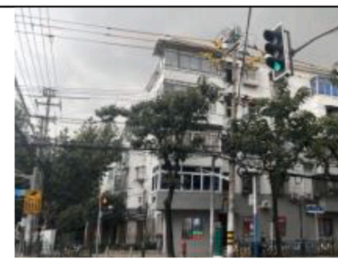
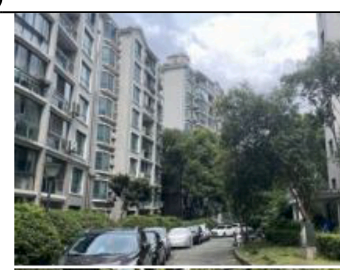
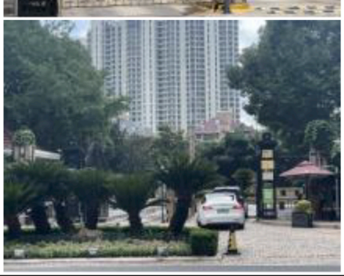
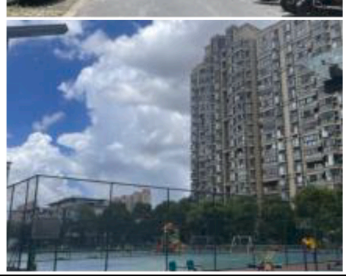
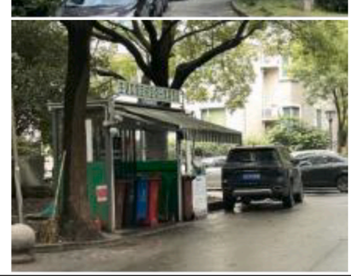


Fig. 10. Partial dependence plots.

includes improving building ventilation and lighting systems, adding independent bathrooms and kitchens, and optimizing building density and internal space to enhance the community's ability to prevent disease. Additionally, it is necessary to consider the location conditions of the neighborhood, planning public facilities and transportation infrastructure appropriately. During an outbreak, selectively closing certain facilities to limit the flow of people while not affecting residents' daily lives is crucial.

The socioeconomic dimension affected the infection rate in a number of ways, mainly related to factors such as population density, per capita income, and house prices. Results from PCC-MLR and RF analyses indicated a positive correlation between population density and the infection rate. However, GWR coefficient maps revealed a significant positive correlation between population density and the infection rate in suburban areas, gradually diminishing closer to the city center. While direct control of population density may not be feasible, it is possible to reduce population density in local areas by rationally planning community spaces, such as increasing green spaces, parks, etc., and optimizing residential layout. This may also reduce the risk of infection. In terms of income and housing prices, the number of individuals with a monthly income exceeding 20,000 yuan and the housing prices in neighborhoods were negatively correlated with the infection rate, corroborating findings from previous studies on the association between disease spread and economic inequality [24]. It is recommended that

Table 7
Typical neighborhood comparison - based on field survey.

Comparison Group 1 Inner Outer Ring neighborhood	*Comparison Group 2 Neighborhoods in the middle ring*	*Comparison Group 3 Inner Ring neighborhood*
		
Older neighborhoods (pre-1990)		
		
		
Emerging neighborhoods (after 2000)		
		
		

neighborhoods implement strategies to promote residents' employment and entrepreneurship while enhancing their income levels. For example, initiatives could include creating additional job opportunities, providing vocational training programs, or offering support for small and micro enterprises as well as entrepreneurial ventures. In neighborhoods renewal plans, it is recommended to introduce different types of housing, including detached houses, apartments, and social housing. This approach can help disperse population density and provide more housing options. Adjusting the proportion of various types of housing within residential areas to balance income distribution and reduce socioeconomic disparities is also advised to decrease the concentration of high-risk groups. Furthermore, implementing monetary subsidy policies for low- to middle-income groups, particularly during epidemic periods, can improve residents' income levels and enhance their ability to

Table 8
Typical neighborhood comparison - based on data.

	Comparison Group 1 Inner Outer Ring neighborhood		Comparison Group 2 middle ring Neighborhoods		Comparison Group 3 Inner Ring neighborhood	
	Emerging	older	Emerging	older	Emerging	older
total cases	163	548	265	888	279	366
per unit area cases	0	0.001	0	0.001	0	0.002
Cases per capita	23.29	78.29	14.72	49.33	15.36	40.67
per capita infection rate	0.009	0.022	0.01	0.022	0.029	0.055
F_65	7.47%	7.95%	8.54%	10.09%	6.26%	7.14%
High_School_Below	75.58%	85.42%	71.50%	82.21%	63.72%	76.18%
Income_Lv_2499	5.00%	7.26%	4.54%	6.91%	2.72%	4.81%
Income_Lv_20,000	15.05%	9.02%	20.39%	9.87%	31.49%	13.52%
High_consumption	46.32%	36.68%	49.81%	35.41%	61.59%	41.27%
Owns_Car	63.62%	63.98%	55.09%	61.88%	49.45%	48.16%
FAR	2.241	1.653	2.285	1.585	2.804	1.616
floor	6	6	18	6	32	6
room	7	2	5	2	5	2
POPdens	0.028	0.035	0.023	0.046	0.016	0.036

prevent disease.

Potential demographic indicators influencing the infection rate also included factors related to population characteristics. Although demographic indicators had a minimal impact on the infection rate in this study, the positive correlation between occupational heterogeneity and the number of cases was still observed through the Models 2–1 and 3–1. Individual occupational types largely influenced behavioral activities and thus affected the likelihood of infection. Given that different occupational types were more concentrated among individuals of lower economic status, the risk of infection increased [21]. Therefore, it is recommended that health education activities be carried out within the community, targeting different occupational groups. The educational content should incorporate how to adopt appropriate protective measures in accordance with the occupational characteristics to minimize the infection risk. Simultaneously, collaboration can be forged with local health authorities to evaluate the infection risk of various occupational types within the community. Targeted intervention measures, such as regular testing and preferential vaccination, can be offered to high-risk occupational groups. Thus, the infection risk of different occupational types within the community can be mitigated, and the health level and quality of life of the residents can be enhanced. It is also proposed to increase elderly care services. For neighborhoods with a higher proportion of elderly residents, providing more elderly care services, including regular physical examinations, senior canteens, psychological support, and social activities, would help elderly people maintain their health and reduce the risk of infection.

Additionally, incorporating smart governance services into urban renewal projects can significantly enhance the resilience of neighborhoods. Smart governance involves utilizing digital technologies and data analytics to improve decision-making, resource allocation, and service delivery [49]. By implementing smart governance systems, city planners can monitor and manage public health data, identify hotspot areas, and deploy resources more effectively. This approach aids in quicker responses to outbreaks and better management of public health measures. Moreover, smart governance services can enhance neighborhood engagement [50]. Digital platforms can be used to inform residents, collect feedback, and involve them in decision-making processes. This participatory approach ensures that the needs and preferences of the neighborhoods are considered in renewal projects. Adopting these smart governance strategies not only strengthens the neighborhood’s ability to withstand infectious diseases but also improves the overall efficiency and sustainability of urban renewal efforts.

6. Conclusion

Following the discovery of the first local infection case in Shanghai in March 2022, the city’s pandemic experienced an exponential increase within a brief timeframe of one and a half months. This study, utilizing data from Shanghai and incorporating the SEIR model, examined the influencing factors at the neighborhood-level in downtown Shanghai, and analyzed the relationship between 19 factors including the socioeconomic, built environment, and demographic characteristics and the transmission of infectious diseases. The study demonstrated a higher incidence rate in older neighborhoods compared to emerging ones and further explored the factors influencing the infection rate, indicating a greater impact of socioeconomic and built environment-related factors on the infection rate. Further research should explore the impact and effectiveness of various intervention policies on infection rates and the long-term effects of neighborhood renewal on enhancing public health resilience. Investigating different community renewal strategies and their influence on infection dynamics can provide valuable insights for urban planning and public health policy.

This study had some limitations, which did not adequately consider the impact of factors such as policy, weather and household structure on infection rates. Existing research suggests that factors such as family structure and education level can influence the infection rate at the neighborhood level. Crowded households, insufficient rooms, and shared kitchen and toilet facilities can exacerbate the spread of pandemics [14,31]. According to a study in Spain, populations with higher levels of university education could reduce the infection rate [51]. Meteorological factors such as temperature, wind, humidity, and air quality have different effects on the spread of COVID-19 [13]. And, the study’s findings are primarily applicable to the renovation of older neighborhoods in cities similar to Shanghai around the world.

The innovative contributions of this study were as follows: firstly, it filled the gap in international research on mainland Chinese cities and conducted an analysis integrating built environment, socioeconomic, and demographic characteristics. Secondly, it established different models to describe the spatial distribution of results. Finally, the research findings could provide references for neighborhood pandemic prevention policies, enhance neighborhood pandemic prevention capabilities, and guide urban healthy development.

In conclusion, this study found that in urban planning and renewal, it is necessary to comprehensively consider various factors such as built environment, socioeconomic, and demographic characteristics to improve urban resilience more effectively. Differentiated urban renewal strategies should be adopted according to the characteristics of different regions. For suburban neighborhoods, population density has the highest impact on the infection rate, and measures should be taken to control population density. For downtown neighborhoods, attention should be paid to balancing income distribution, reducing socioeconomic disparities, and minimizing the aggregation of high-risk populations. Secondly, efforts should be made to strengthen urban infrastructure construction and improve public health levels. Increasing green space and outdoor recreational areas, optimizing living environments, and improving residents' quality of life can effectively reduce disease transmission. Neighborhood renewal should comprehensively consider factors such as built environment, socioeconomic, and demographic characteristics and adopt multi-level, multi-directional comprehensive prevention and control strategies to enhance neighborhood pandemic prevention resilience and ensure the health and safety of residents.

CRediT authorship contribution statement

Poui Chong: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chao Liu:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition. **Wengin Chung:** Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (No. 52108060); the Science Foundation for the Science and Technology Commission of Shanghai Municipality (No.20210247900522dz1207800); the Shanghai Rising-Star Program (No.22QB1404900).

Thanks to Shanghai Key Laboratory of Urban Renewal and Spatial Optimization Technology and Key Laboratory for High-density Habitat Ecology and Energy Conservation of Ministry of Education" for the support of Shanghai Population Geography Big Data. Thanks to Metrodata Technology Co. for the support of 1 % population sampling microdata within the outer ring of Shanghai and the technical guidance of Chief Economist Chen Qin and Researcher Qian Junjie.

Data availability

The authors are unable or have chosen not to specify which data has been used.

References

[1] M. Sakurai, H. Chughtai, Resilience against crises: COVID-19 and lessons from natural disasters, Eur. J. Inform. Syst. 29 (5) (2020) 585–594, https://doi.org/10.1080/0960085X.2020.1814171. Available at:.
[2] J. Corburn, Confronting the Challenges in Reconnecting Urban Planning and Public Health, Am. J. Public Health 94 (4) (2004) 541–546, https://doi.org/10.2105/AJPH.94.4.541. Available at:.
[3] V.J. Lee, et al., Epidemic preparedness in urban settings: new challenges and opportunities, Lancet Infect. Dis. 20 (5) (2020) 527–529, https://doi.org/10.1016/S1473-3099(20)30249-8. Available at:.
[4] M. Hadei, et al., Presence of SARS-CoV-2 in the air of public places and transportation, Atmos. Pollut. Res. 12 (3) (2021) 302–306, https://doi.org/10.1016/j.apr.2020.12.016. Available at:.
[5] K. Azuma, et al., Environmental factors involved in SARS-CoV-2 transmission: effect and role of indoor environmental quality in the strategy for COVID-19 infection control, Environ. Health Prev. Med. 25 (1) (2020) 66, https://doi.org/10.1186/s12199-020-00904-2. Available at:.
[6] C. Shi, et al., Which urban communities are susceptible to COVID-19? An empirical study through the lens of community resilience, BMC Public Health 22 (1) (2022) 70, https://doi.org/10.1186/s12889-021-12419-8. Available at:.
[7] M. Amirzadeh, et al., Towards building resilient cities to pandemics: a review of COVID-19 literature, Sustain. Citi. Soc. 89 (2023) 104326, https://doi.org/10.1016/j.scs.2022.104326. Available at:.
[8] C. Moreno, et al., Introducing the "15-minute city": sustainability, resilience and place identity in future post-pandemic cities, Smart Citi. 4 (1) (2021) 93–111, https://doi.org/10.3390/smartcities4010006. Available at:.
[9] M. Das, et al., Habitat vulnerability in slum areas of India – What we learnt from COVID-19? Int. J. Disast. Risk Reduct. 65 (2021) 102553 https://doi.org/10.1016/j.ijdrr.2021.102553. Available at:.
[10] M. Breton, Neighborhood resiliency, J. Commun. Pract. 9 (1) (2001) 21–36, https://doi.org/10.1300/J125v09n01_02. Available at:.
[11] G. Huang, et al., Influencing factors and their influencing mechanisms on urban resilience in China, Sustain. Citi. Soc. 74 (2021) 103210, https://doi.org/10.1016/j.scs.2021.103210. Available at:.
[12] B. Wali, L.D. Frank, Neighborhood-level COVID-19 hospitalizations and mortality relationships with built environment, active and sedentary travel, Health Place 71 (2021) 102659, https://doi.org/10.1016/j.healthplace.2021.102659. Available at: .
[13] J. Li, et al., Natural and socio-environmental factors contribute to the transmissibility of COVID-19: evidence from an improved SEIR model, Int. J. Biometeorol. (2023), https://doi.org/10.1007/s00484-023-02539-8 [Preprint]. Available at:.
[14] M. Hu, et al., The role of built and social environmental factors in Covid-19 transmission: a look at America's capital city, Sustain. Citi. Soc. 65 (2021) 102580, https://doi.org/10.1016/j.scs.2020.102580. Available at:.
[15] D. Consolazio, et al., Assessing the impact of individual characteristics and neighborhood socioeconomic status during the COVID-19 pandemic in the provinces of Milan and Lodi, Int. J. Health Serv. 51 (3) (2021) 311–324, https://doi.org/10.1177/0020731421994842. Available at:.
[16] A. Godio, F. Pace, A. Vergnano, SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence, Int. J. Environ. Res. Public Health 17 (10) (2020) 3535, https://doi.org/10.3390/ijerph17103535. Available at:.
[17] S. He, Y. Peng, K. Sun, SEIR modeling of the COVID-19 and its dynamics, Nonlin. Dyn. 101 (3) (2020) 1667–1680, https://doi.org/10.1007/s11071-020-05743-y. Available at:.
[18] B. Li, et al., Built environment and early infection of COVID-19 in urban districts: a case study of Huangzhou, Sustain. Citi. Soc. 66 (2021) 102685, https://doi.org/10.1016/j.scs.2020.102685. Available at:.
[19] C. Liu, Z. Liu, C. Guan, The impacts of the built environment on the incidence rate of COVID-19: a case study of King County, Washington, Sustain. Citi. Soc. 74 (2021) 103144, https://doi.org/10.1016/j.scs.2021.103144. Available at:.
[20] N.I. Heredia, et al., The neighborhood environment and hispanic/latino health, Am. J. Health Promot. 36 (1) (2022) 38–45, https://doi.org/10.1177/08901171211022677. Available at:.
[21] M. Almagro, A. Orane-Hutchinson, JUE Insight: the determinants of the differential exposure to COVID-19 in New York city and their evolution over time, J. Urban Econ. 127 (2022) 103293, https://doi.org/10.1016/j.jue.2020.103293. Available at:.
[22] J.T. Chen, N. Krieger, Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: US County versus zip code analyses, J. Public Health Manag. Pract. 27 (Supplement 1) (2021) S43–S56, https://doi.org/10.1097/PHH.0000000000001263. Available at:.
[23] Cromer, S.J. et al. (no date) 'Geospatial Analysis of Individual and Community-Level Socioeconomic Factors Impacting SARS-CoV-2 Prevalence and Outcomes'.
[24] U.N. Emeruwa, et al., Associations between built environment, neighborhood socioeconomic status, and sars-cov-2 infection among pregnant women in New York City, JAMA 324 (4) (2020) 390, https://doi.org/10.1001/jama.2020.11370. Available at:.
[25] A. Ghose, et al., Community prevalence of antibodies to SARS-CoV-2 and correlates of protective immunity in an Indian metropolitan city, preprint. Infectious Diseases (except HIV/AIDS) (2020), https://doi.org/10.1101/2020.11.17.20228155. Available at.
[26] D. De Ridder, et al., Socioeconomically disadvantaged neighborhoods face increased persistence of SARS-CoV-2 clusters, Front. Public Health 8 (2021) 626090, https://doi.org/10.3389/fpubh.2020.626090. Available at:.
[27] M. KC, E. Oral, S. Straif-Bourgeois, A.L. Rung, E.S. Peters, The effect of area deprivation on COVID-19 risk in Louisiana, PLoS One 15 (12) (2020) e0243028, https://doi.org/10.1371/journal.pone.0243028.
[28] S.J. Cromer, C.M. Lakhani, D.J. Wexler, S.M. Burnett-Bowie, M. Udler, C.J. Patel, Geospatial Analysis of Individual and Community-Level Socioeconomic Factors Impacting SARS-CoV-2 Prevalence and Outcomes, medRxiv [Preprint] 30 (2020), https://doi.org/10.1101/2020.09.30.20201830, 2020.09.30.20201830.
[29] E. Hatef, et al., Assessing the impact of neighborhood socioeconomic characteristics on COVID-19 prevalence across seven states in the United States, Front. Public Health 8 (2020) 571808, https://doi.org/10.3389/fpubh.2020.571808. Available at:.
[30] D. Carrión, et al., Neighborhood-level disparities and subway utilization during the COVID-19 pandemic in New York City, Nat. Commun. 12 (1) (2021) 3692, https://doi.org/10.1038/s41467-021-24088-7. Available at:.
[31] G.R. Oates, et al., The Association Between Neighborhood Social Vulnerability and COVID-19 Testing, Positivity, and Incidence in Alabama and Louisiana, J. Commun. Health 46 (6) (2021) 1115–1123, https://doi.org/10.1007/s10900-021-00998-x. Available at:.
[32] R. Tipirneni, M. Karmakar, M. O'Malley, H.C. Prescott, V. Chopra, Contribution of Individual- and Neighborhood-Level Social, Demographic, and Health Factors to COVID-19 Hospitalization Outcomes, Ann. Intern. Med. 175 (4) (2022) 505–512, https://doi.org/10.7326/M21-2615.
[33] D. Schmiege, et al., Associations between built environment factors and SARS-CoV-2 infections at the neighbourhood level in a metropolitan area in Germany, J. Urban Health 100 (1) (2023) 40–50, https://doi.org/10.1007/s11524-022-00708-5. Available at:.

[34] R. Cazzolla Gatti, et al., Machine learning reveals that prolonged exposure to air pollution is associated with SARS-CoV-2 mortality and infectivity in Italy, Environ. Pollut. 267 (2020) 115471, https://doi.org/10.1016/j.envpol.2020.115471. Available at:.

[35] L. Meng, et al., Apportionment and evolution of pollution sources in a typical riverside groundwater resource area using PCA-APCS-MLR model, J. Contam. Hydrol. 218 (2018) 70–83, https://doi.org/10.1016/j.jconhyd.2018.10.005. Available at:.

[36] C. Jayaweera, N. Aziz, Reliability of principal component analysis and pearson correlation coefficient, for application in artificial neural network model development, for water treatment plants, IOP Conferen. Ser.: Mater. Sci. Eng. 458 (2018) 012076, https://doi.org/10.1088/1757-899X/458/1/012076. Available at: .

[37] T. Ioanna, et al., Factors that influence disease-specific quality of life or health status in patients with COPD: a systematic review and meta-analysis of Pearson correlations, Prim. Care Respirat. J. 20 (3) (2011) 257–268, https://doi.org/10.4104/pcrj.2011.00029. Available at:.

[38] H.T. Davis, et al., Identifying natural and anthropogenic sources of metals in urban and rural soils using GIS-based data, PCA, and spatial interpolation, Environ. Pollut. 157 (8) (2009) 2378–2385, https://doi.org/10.1016/j.envpol.2009.03.021. Available at:.

[39] K. Faisal, A. Shaker, An investigation of gis overlay and pca techniques for urban environmental quality assessment: a case study in Toronto, Ontario, Canada, Sustainability 9 (3) (2017) 380, https://doi.org/10.3390/su9030380. Available at: .

[40] F. Carozzi, Urban Density and Covid-19, *SSRN Electron. J.* [Preprint] (2020), https://doi.org/10.2139/ssrn.3643204. Available at:.

[41] G. Grekousis, et al., Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: a geographical random forest approach, Health Place 74 (2022) 102744, https://doi.org/10.1016/j.healthplace.2022.102744. Available at:.

[42] V. Chaturvedi, W.T. de Vries, Machine learning algorithms for urban land use planning: a review, Urban Sci. 5 (3) (2021) 68, https://doi.org/10.3390/urbansci5030068. Available at:.

[43] N. Xiong, et al., Spatial distribution and influencing factors of litter in urban areas based on machine learning – A case study of Beijing, Waste Manag. 142 (2022) 88–100, https://doi.org/10.1016/j.wasman.2022.01.039. Available at:.

[44] J. Ma, et al., Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques, Land Use Policy 94 (2020) 104537, https://doi.org/10.1016/j.landusepol.2020.104537. Available at:.

[45] H. Tong, M. Li, J. Kang, Relationships between building attributes and COVID-19 infection in London, Build. Environ. 225 (2022) 109581, https://doi.org/10.1016/j.buildenv.2022.109581. Available at:.

[46] E. Goldstein, M. Lipsitch, M. Cevik, On the effect of age on the transmission of SARS-CoV-2 in households, schools, and the community, J. Infect. Dis. 223 (3) (2021) 362–369, https://doi.org/10.1093/infdis/jiaa691. Available at:.

[47] L. Von Seidlein, et al., Crowding has consequences: prevention and management of COVID-19 in informal urban settlements, Build. Environ. 188 (2021) 107472, https://doi.org/10.1016/j.buildenv.2020.107472. Available at:.

[48] A. Ulubaş Hamurcu, M. Yılmaz, Geostatistical assessment of the built environment and spatio-temporal distribution patterns of COVID-19 cases in Istanbul, Türkiye, Build. Environ. 243 (2023) 110666, https://doi.org/10.1016/j.buildenv.2023.110666. Available at:.

[49] G. Huang, et al., A methodology for assessing supply-demand matching of smart government services from citizens' perspective: a case study in Nanjing, China, Habit. Int. 138 (2023) 102880, https://doi.org/10.1016/j.habitatint.2023.102880. Available at:.

[50] G. Huang, et al., Factors affecting sustainability of smart city services in China: from the perspective of citizens' sense of gain, Habit. Int. 128 (2022) 102645, https://doi.org/10.1016/j.habitatint.2022.102645. Available at:.

[51] R. Mogi, G. Kato, S. Annaka, Socioeconomic inequality and COVID-19 prevalence across municipalities in Catalonia, *Spain*. preprint. SocArXiv (2020), https://doi.org/10.31235/osf.io/5jgzy. Available at:.