

The impact of Urban Spatial Environment on COVID-19: a case study in Beijing

Zhen Yang¹, Jiaxuan Li¹, Yu Li^{1*}, Xiaowen Huang¹, Anran Zhang¹, Yue Lu¹, Xu Zhao¹, Xueyan Yang¹

¹Beijing University of Civil Engineering and Architecture, China

Submitted to Journal:
Frontiers in Public Health

Specialty Section:
Disaster and Emergency Medicine

Article type:
Original Research Article

Manuscript ID:
1287999

Received on:
04 Sep 2023

Journal website link:
www.frontiersin.org

Scope Statement

Epidemics are dangerous and difficult to prevent and control, especially in urban areas. In this study, we combined the frequency of COVID-19 outbreaks with the external spatial environmental elements of the city, and investigated the correlation between the frequency of COVID-19 outbreaks and the internal spatial environmental elements of different grades of neighbourhoods. Clarifying the correlation between the frequency of COVID-19 outbreaks and the urban spatial environment may help improve cities' ability to respond to such public health emergencies.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

CRediT Author Statement

Yue Lu: Data curation, Formal Analysis, Investigation, Software, Writing - original draft. Xu Zhao: Formal Analysis, Supervision, Writing - original draft. Xueyan Yang: Formal Analysis, Writing - original draft. Anran Zhang: Data curation, Formal Analysis, Investigation, Software, Writing - original draft. Xiaowen Huang: Data curation, Formal Analysis, Investigation, Software, Writing - original draft. Yu Li: Conceptualization, Funding acquisition, Methodology, Writing - review & editing. Jiaxuan Li: Conceptualization, Formal Analysis, Methodology, Software, Writing - original draft. Zhen Yang: Conceptualization, Funding acquisition, Methodology, Writing - review & editing.

Keywords

machine learning, epidemic, Urban Spatial Environment, COVID-19, BP neural network

Abstract

Word count: 237

Epidemics are dangerous and difficult to prevent and control, especially in urban areas. Clarifying the correlation between the frequency of COVID-19 outbreaks and the urban spatial environment may help improve cities' ability to respond to such public health emergencies. In this study, we first analysed the spatial distribution characteristics of COVID-19 epidemic outbreaks by correlating the geographic locations of COVID-19 epidemic-affected neighbourhoods in the city of Beijing with the time point of onset. Second, we created a geographically weighted regression model and an unordered multi-categorical logistic regression model combining the frequency of COVID-19 outbreaks with the external spatial environmental elements of the city. Different grades of epidemic-affected neighbourhoods in the study area were classified according to the clustering analysis results. The correlation between the frequency of COVID-19 outbreaks and the internal spatial environmental elements of different grades of neighbourhoods was investigated using a binomial logistic regression model. Finally, a back propagation neural network was used to test the correlation between epidemic outbreak frequency and the city's internal and external spatial environments. The study yielded the following results. (i) Epidemic outbreak frequency was correlated with the urban external spatial environment, and this correlation was affected by the classification of neighbourhoods. (ii) The correlation between epidemic outbreak frequency and the internal environmental elements of neighbourhoods of different grades differed. (iii) Machine learning results further verified that epidemic outbreak frequency was correlated with both internal and external urban spatial environments.

Funding information

This work was supported by the National Natural Science Foundation of China (Grant NO.52178002, 52008015, 51608021, and 82070320), the Research Capacity Enhancement Program for Young Teachers of Beijing University of Civil Engineering and Architecture (Grant NO. X22018), and the Quality Improvement Project of Postgraduate Education and Teaching of Beijing University of Civil Engineering and Architecture (Grant NO. J2023012)

Funding statement

The author(s) declare financial support was received for the research, authorship, and/or publication of this article.

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated Statement: No human studies are presented in the manuscript.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable images or data are presented in this study.

Data availability statement

Generated Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

In review

The impact of Urban Spatial Environment on COVID-19: a case study in Beijing

1 **Zhen Yang¹, Jiaxuan Li¹, Yu Li^{1*}, Xiaowen Huang¹, Anran Zhang¹, Yue Lu¹, Xu Zhao¹,**
2 **Xueyan Yang¹**

3 ¹School of Architecture and Urban Planning, Beijing University of Civil Engineering and
4 Architecture, Beijing 100044 , China

5 *** Correspondence:**
6 Yu Li
7 liyu@bucea.edu.cn

8 **Keywords:** Machine Learning; Epidemic; Urban Spatial Environment; COVID-19; BP neural
9 network

10 Abstract

11 Epidemics are dangerous and difficult to prevent and control, especially in urban areas. Clarifying the
12 correlation between the frequency of COVID-19 outbreaks and the urban spatial environment may
13 help improve cities' ability to respond to such public health emergencies. In this study, we first
14 analysed the spatial distribution characteristics of COVID-19 epidemic outbreaks by correlating the
15 geographic locations of COVID-19 epidemic-affected neighbourhoods in the city of Beijing with the
16 time point of onset. Second, we created a geographically weighted regression model and an
17 unordered multi-categorical logistic regression model combining the frequency of COVID-19
18 outbreaks with the external spatial environmental elements of the city. Different grades of epidemic-
19 affected neighbourhoods in the study area were classified according to the clustering analysis results.
20 The correlation between the frequency of COVID-19 outbreaks and the internal spatial
21 environmental elements of different grades of neighbourhoods was investigated using a binomial
22 logistic regression model. Finally, a back propagation neural network was used to test the correlation
23 between epidemic outbreak frequency and the city's internal and external spatial environments. The
24 study yielded the following results. (i) Epidemic outbreak frequency was correlated with the urban
25 external spatial environment, and this correlation was affected by the classification of
26 neighbourhoods. (ii) The correlation between epidemic outbreak frequency and the internal
27 environmental elements of neighbourhoods of different grades differed. (iii) Machine learning results
28 further verified that epidemic outbreak frequency was correlated with both internal and external
29 urban spatial environments.

30 1 Introduction

31 Human beings are now living in a risk society, and significant public health events can seriously
32 affect a country's social, economic and political order and endanger national security and
33 development. Furthermore, the rapid population movement brought about by economic globalisation
34 has sharply increased the public health and safety risks facing human society. Approximately 70% of
35 public health risks come from major infectious disease epidemics (Liu & Liang, 2022). While public

36 health risks have already attracted significant attention globally, a new round of attention to and
37 reflection on epidemic-related research has been initiated by all sectors of society. It is urgently
38 necessary to carry out epidemic-related prevention and control work as well as analysis and summary
39 work.

40 Exploration of models related to epidemics, which are dangerous and difficult to prevent and control,
41 began in the fields of medicine and mathematics. As early as 1760, the Swiss mathematician Daniel
42 proposed a mathematical model for a smallpox inoculation strategy. Since the 20th century, with the
43 increased complexity and accuracy of modelling and the development of science and technology, the
44 role of computers and complex science and other disciplines in developing epidemic-related models
45 has become increasingly significant. For example, Lee et al. (2013) constructed a new model based
46 on the SEIAR (susceptible (S), exposed (E), symptomatic infected (I), asymptomatic infected (A),
47 recovered (R)) model and studied the effects of control measures using the optimal control approach
48 to identify the best control measures. Prosper et al. (2011) developed a two-strain model to analyse
49 the optimal treatment and isolation strategies under the constraint of minimal total incidence. Chao &
50 Zhang (2022) developed a class of epidemic diseases in polluted environments and modelled factors
51 such as the boundedness of air pollutant concentrations and the kinetic nature of respiratory disease
52 transmission. In terms of model considerations and elements of concern, most recent studies related
53 to epidemics have started with population and disease characteristics and thus predicted the duration
54 of epidemics and the number of infections; few have explored the relevant models affecting epidemic
55 outbreaks.

56 The various factors associated with epidemic outbreaks are complex, and different urban form and
57 design factors can influence epidemic dynamics. In a study of more than 900 metropolitan areas in
58 the United States, scholars found that virus-related mortality was slightly lower in high-density areas
59 than in spreading areas (Hamidi et al., 2020), and research in China has found that connectivity,
60 especially to the outbreak city, is a major factor influencing the spread of the epidemic in the early
61 stages of the outbreak (LIN et al., 2020). Yu et al., (2021) investigated the mechanism of COVID-19
62 transmission within and between cities in different periods in China to better understand the nature of
63 the outbreak. Other scholars have argued that to maintain reasonable physical distances during a
64 pandemic, cities need to allocate more public space to better meet the needs of pedestrians and
65 cyclists, and to provide sufficient green and open space to meet citizens' need for outdoor exercise
66 and recreation (Honey-Rosés et al., 2020). This reconfiguration may also provide opportunities to
67 capture additional health and climate-appropriate benefits from further integration of urban greenery
68 into the city, helping to counteract the effects of other stressors and adverse events (Sharifi, 2020),
69 etc. However, recent research on urban factors influencing epidemic outbreaks has focused on
70 specific aspects of urban space, without a detailed and comprehensive consideration of the factors
71 influencing epidemic outbreaks. Such research has rarely explored the relevance of overall urban
72 spatial environment, which still needs to be considered.

73 In conclusion, recent domestic and international studies related to epidemics have mainly focused on
74 the characteristics of the populations and the diseases themselves; studies have less often linked
75 epidemic outbreaks to the macroscopic urban level and thus explored the interactions with the overall
76 urban spatial environment. COVID-19 pneumonia, however, posed a massive threat to human health,
77 as the most widespread and globally prevalent infectious disease encountered by humans in the last
78 100 years (Chen et al., 2020). The spread of COVID-19 can affect people's physical and mental
79 health (Eastin & Eastin, 2020). The measures carried out to prevent the spread of the disease have
80 affected the economy, with potentially catastrophic long-term impacts (Cash & Patel, 2020).
81 According to the WHO, as of February 9, 2023, more than 672 million confirmed cases of COVID-

82 19 and more than 6.85 million deaths had been reported worldwide. Thus, the COVID-19 pneumonia
83 epidemic posed a significant challenge to global public health security, economic development and
84 social stability (Huang et al., 2020), and thus has particular research value and research significance.

85 Therefore, we took COVID-19 as a case study to clarify the correlation between the frequency of
86 COVID-19 outbreaks and the urban spatial environment. Our findings have important implications
87 for improving the capacity of cities to respond to public health emergencies.

88 **2 Materials and Methods**

89 **2.1 Data sources**

90 According to the Prevention and Control Program for COVID-19 Pneumonia (9th Edition) issued by
91 China's National Health and Health Commission, to better prevent and control the source of infection,
92 home health surveillance was required to be conducted for 7 days after COVID-19 patients had
93 recovered and been discharged from hospital. In addition, the absence of new infections for 7
94 consecutive days in medium- and high-risk areas was considered one of the requirements for the
95 release of risk areas. Therefore, this study used a 7-day cycle to record the location of new confirmed
96 cases living in neighborhoods within Beijing's fifth ring road for a total of 53 cycles over
97 approximately one year since November 1, 2021, by searching the Beijing Municipal Government
98 Data Resource Network(<https://data.beijing.gov.cn/index.htm>), the Beijing Municipal Health and
99 Wellness Commission(<http://wjw.beijing.gov.cn/wjwh/tztl/xxgzb/gzbdyqtb/> index.html), and the
100 Beijing Daily(<https://www.bjd.com.cn/index.shtml>). Using ArcGIS, the above statistical information
101 corresponded to 2,769 grids within the study area of 500 m × 500 m. The frequency of outbreaks in
102 each grid was calculated as the mean ratio of the cycles in which new cases were recorded or
103 classified as medium- to high-risk areas to all 53 cycles in each grid.

104 The urban spatial environment was divided into two dimensions: external environmental elements
105 and internal environmental elements. The external environment was characterised in terms of six
106 elements: building density, volume ratio, water and greenery coverage, density of commercial
107 facilities, density of public service facilities and density of transportation facilities. The internal
108 environment was characterised in terms of four elements: housing price, building age, number of
109 buildings and number of households. Data on building density, volume ratio, density of commercial
110 facilities, density of public service facilities, density of transportation facilities, number of buildings
111 and number of households were obtained from Gaode Map, and data on water and greenery coverage
112 were obtained from the Landset-8 satellite remote sensing map in the Chinese Academy of Sciences
113 Geospatial Data Cloud for July 2020.

114 **2.2 Evaluation Methodology**

115 **2.2.1 Geographically weighted regression**

116 Geographically weighted regression (GWR) is a spatial analysis method commonly used in
117 geography and related disciplines. The GWR model is an extended model of multiple linear
118 regression, which can create a local regression equation for each point in the range of the model. The
119 GWR model introduces a spatial weight function to estimate the different relationships between
120 variables in different regions based on spatial variability to better characterise the quantitative
121 relationships' spatial variation (Brunsdon et al., 1996). The GWR model equation is as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^m \beta_k(u_i, v_i)X_{ik} + \varepsilon_i$$

122 where y_i is the response variable at spatial position (u_i, v_i) , X_{ik} denotes the observed value of the
 123 independent variable at spatial position (u_i, v_i) , $\beta_0(u_i, v_i)$ is the intercept term of the regression
 124 relationship, $\beta_k(u_i, v_i)$ is the regression coefficient of the k th independent variable at spatial position
 125 (u_i, v_i) , which is a continuous function of spatial position (u_i, v_i) , and ε_i is a mutually independent
 126 random error term.

127 In the GWR model, the estimated value of the coefficient $\beta_k(u, v)$ is reflected in a graph, which can
 128 visualise the spatial distribution of the intensity of the influence of the independent variable on the
 129 dependent variable. We aimed to investigate the influence of urban spatial environmental elements
 130 on the frequency of epidemic outbreaks at different spatial locations through geographically weighted
 131 regression.

132 **2.2.2 Unordered multi-categorical logistic regression**

133 The unordered multi-categorical logistic regression model is used in cases where the dependent
 134 variable is unordered and multi-categorical. In unordered multi-categorical logistic regression, it is
 135 necessary to first define one level of the dependent variable as the reference level (Turan et al., 2022).
 136 all of the other levels are compared with it to build a generalised logit model with the number of
 137 levels $n-1$. As an example, we fit three generalised logistic models with four levels of the dependent
 138 variable.

$$\text{logit}\left(\frac{\rho_b}{\rho_a}\right) = \alpha_1 + \beta_{11}X_{11} + \dots + \beta_{1\rho}X_{1\rho}$$

$$\text{logit}\left(\frac{\rho_c}{\rho_a}\right) = \alpha_2 + \beta_{21}X_{31} + \dots + \beta_{2\rho}X_{1\rho}$$

$$\text{logit}\left(\frac{\rho_d}{\rho_a}\right) = \alpha_3 + \beta_{31}X_{31} + \dots + \beta_{3\rho}X_{3\rho}$$

139 where $\rho_a + \rho_b + \rho_c = 1$. We intended to explore the relationship between the frequency of epidemic
 140 outbreaks and urban spatial environmental elements for different classifications through unordered
 141 multi-categorical logistic regression.

142 **2.2.3 Binomial logistic regression model**

143 The binomial logistic regression model is a classification model represented by a conditional
 144 probability distribution $P(Y|X)$ in the form of a parametric logistic distribution. Here, the random
 145 variable X takes the value of a real number, and the random variable Y takes the value of 1 or 0. We
 146 classified the presence or absence of an epidemic as 1 or 0. The binomial logistic regression model
 147 was the following conditional probability distribution.

$$P(Y = 1|x) = \frac{\exp(w \cdot x + b)}{1 + \exp(w \cdot x + b)}$$

$$P(Y = 0|x) = \frac{1}{1 + \exp(w \cdot x + b)}$$

148 Here, $x \in R^n$ is the input, $Y \in \{0,1\}$ is the output, $w \in R^n$ and $b \in R$ are the parameters, w is
 149 called the weight vector, b is called the bias, and $w \cdot x$ is the inner product of w and x .

150 For a given input instance x , $P(Y=1|x)$ and $P(Y=0|x)$ can be found according to Eq. 7 and Eq. 8. The
151 logistic regression compares the magnitude of the two conditional probability values and assigns the
152 instance x to the class with the larger probability value.

153 For convenience, the weight vector and the input vector can be expanded; they are still denoted as w ,
154 x , i.e. $w = (w(1), w(2), \dots, w(n), b)^T$, $x = (x(1), x(2), \dots, x(n), 1)^T$. The logistic regression model is then as
155 follows:

$$P(Y = 1|x) = \frac{\exp(w \cdot x)}{1 + \exp(w \cdot x)}$$

$$P(Y = 0|x) = \frac{1}{1 + \exp(w \cdot x)}$$

156 We aimed to investigate whether epidemic outbreak frequency is related to urban spatial
157 environmental elements through a binomial logistic regression model and simultaneous analysis of
158 multiple covariates using a dichotomous logistic model (Goldstein, 2003). We hoped to find the key
159 environmental elements that influence the frequency of epidemic outbreaks and the specific
160 mechanisms of action of this influence.

161 **2.2.4 Cluster analysis**

162 Cluster analysis is an important computational method in data mining, which uses the relationships
163 between sample data variables to represent the relationships between samples. Through clustering,
164 the same or similar objects can be classified into a class or cluster, and the average centre of objects
165 belonging to the same class can be taken as the centre of the class. The class centres can reflect the
166 common properties of the objects in the class. The relationships between class centres can be
167 calculated to determine the difference between different classes. Using cluster analysis, researchers
168 can take advantage of in-depth phenotyping to reveal unique patterns of association among
169 phenotypic variables (Ahmad et al., 2014).

170 A clustering algorithm classifies objects with the same or similar properties into the same class and
171 objects with different properties into different classes through different algorithms. In this study,
172 clustering analysis was used to study the division of different classes of epidemic-affected plots to
173 subsequently study the relationship between the frequency of epidemic outbreaks in different classes
174 of plots and the environmental elements within each class of plots.

175 **2.2.5 Machine learning through a back propagation (BP) neural network**

176 BP neural networks are among the most widely used neural network models. It is a multi-layer feed-
177 forward neural network that propagates the signal forward and the error backwards, and after
178 repeated cycles, the error information is adjusted by forward and backward propagation of the signal
179 to obtain the best results within the error range. Unlike traditional information and data processing
180 methods, a BP neural network is able to process distributed stored information in parallel and
181 collaboratively, which is a nonlinear dynamical process; thus, it has a very obvious advantage in
182 dealing with complex multidimensional nonlinear problems.

183 The structure of a BP neural network is shown in the following figure (Figure 1); it is divided into an
184 input layer (Input), a hidden layer (Hidden), and an output layer (Output). The number of nodes in
185 the input layer depends on the number of features in the input, and the number of nodes in the output
186 layer is determined by the type of classification. The number of hidden layers and the number of

187 nodes in each hidden layer are artificially set by the training engineer's experience and are generally
188 taken as the squared value of the number of nodes in the input layer.

189 In this study, a BP neural network for machine learning was used to validate the results of the GWR
190 and binary logistic regression.

191 **3 Results and Analysis**

192 **3.1 Analysis of urban spatial characteristics**

193 **3.1.1 Analysis of urban spatial environmental characteristics**

194 In the analysis of urban spatial environmental characteristics, the main urban spatial elements studied
195 comprised building density, volume ratio, water and greenery coverage, density of commercial
196 facilities, density of public service facilities, and density of transportation facilities. ArcGIS was used
197 to establish a 500 m × 500 m grid covering the study area, and the building density and volume ratio
198 within the grid were calculated using building outline and building height data from Gaode Map . In
199 addition, we calculated the density of commercial facilities, service facilities, and transportation
200 facilities. Based on a Landset-8 satellite remote sensing map for July 2020 from the Geospatial Data
201 Cloud of the Chinese Academy of Sciences, surface cover extraction was performed using the
202 supervised classification method to calculate the coverage of waters and greenery within the grid.

203 The building density and volume ratio of the study area showed a spatial pattern of high in the middle
204 and low in the surrounding areas. The areas with the highest building density were mainly
205 concentrated within the second ring road, with the highest building density in the northeastern and
206 northwestern parts of Beihai Park; there were more points with high building density scattered
207 around the southern fifth ring. The areas with higher volume ratios were scattered between the
208 second and fourth rings, while the overall volume ratios in the old city within the second ring and
209 outside the fourth ring were lower.

210 The spatial pattern of vegetation coverage in the study area was low in the central part and high in the
211 surrounding areas, which was opposite to the spatial distribution pattern of building density and
212 volume ratio in the study area. The areas with the highest levels of vegetation and water coverage
213 were the Summer Palace and Olympic Forest Park near the North 5th Ring Road, the Nanyuan Forest
214 Wetland Park and several country parks near the South 5th Ring Road, followed by the Chaoyang
215 Park area near the Northeast 4th Ring Road, the Lize Financial and Business District near the
216 Southwest 3rd Ring Road, and the Temple of Heaven Park and the Six Seas area near the 2nd Ring
217 Road.

218 Commercial facilities and public service facilities in the study area showed a spatial pattern of high
219 density in the northwest, a scattered distribution in the southeast, and low density in other areas.
220 Commercial facilities and service facilities were mainly concentrated in Xicheng District, Dongcheng
221 District and Haidian District, and the areas with the highest facility rates were scattered within the
222 second ring road and the northwest section between the third and fourth ring roads. The next highest
223 density of service facilities was in the Chaoyang District near the East 4th Ring Road, which showed
224 a distribution pattern extending eastward. Areas with high-density commercial and service facilities
225 were scattered in a dotted pattern near the fifth ring road in the southeast. The spatial density pattern
226 of transportation facilities was also high in the northwest, low in other areas, and scattered in the
227 southeast, but these facilities were mainly concentrated on either side of the northwest ring road and

228 near the main transportation space. Their distribution pattern was circular, following the circular road
229 network (Figure 1).

230 Figure 1. Results of analysis of urban spatial environmental characteristics (building density (A),
231 volume ratio (B), water and green coverage (C), density of commercial facilities (D), density of
232 public service facilities (E), density of transportation facilities (F)

233 **3.1.2 Analysis of spatial characteristics of urban epidemic outbreaks**

234 In the analysis of the spatial characteristics of urban epidemic outbreaks, the main urban spatial
235 element under study was the frequency of epidemic outbreaks. A 500 m × 500 m grid covering the
236 study area was established using ArcGIS, and the outbreak frequency within the grid was calculated
237 based on the Beijing Municipal Government Data Resource Network, the Beijing Municipal Health
238 and Wellness Commission, and Beijing Daily Public by recording information on the locations of the
239 residential neighbourhoods with new daily confirmed cases within Beijing's fifth ring road from
240 November 2021 to October 2022, as well as the locations of medium- and high-risk areas.

241 In terms of its spatial pattern, outbreak frequency was low in the central and peripheral segments and
242 high in the remaining segments of the study area. The highest outbreak frequencies were mainly
243 concentrated in the areas outside the South Second Ring Road, with the highest outbreak frequency
244 in Chaoyang District, followed by Shijingshan and Daxing Districts, and scattered in the northwest
245 part of Xicheng District and Haidian District (Figure 2).

246 Figure 2. Spatial distribution of the frequency of epidemic outbreaks in the 5th Ring Road area of
247 Beijing

248 **3.2 Correlation analysis of epidemic outbreak frequency and external spatial environment**

249 **3.2.1 Geographically weighted regression analysis**

250 The frequency of COVID-19 outbreaks in a total of 2,769 grids of 500 m × 500 m within the study
251 area was used as the dependent variable. Six indicators, namely building density, volume ratio,
252 density of commercial facilities, density of public service facilities, and density of transportation
253 facilities, were used as independent variables to construct a geographically weighted regression
254 model. The measured coefficients, R^2 and adjusted R^2 , of the geographically weighted regression
255 model were 0.79 and 0.74, respectively. Both were greater than 0.7 and the absolute values of the
256 regression coefficients of each element were large, indicating that the geographically weighted
257 regression model had a strong explanatory effect (Wang et al., 2022). There was a significant
258 correlation between the frequency of epidemic outbreaks and the external spatial environment (Table
259 1).

260 Table 1. GWR Model Determination Coefficients.

261 The regression coefficients obtained from the geographically weighted regression model are shown
262 in Table 1. The regression coefficients represented the degree of influence of the indicators on
263 COVID-19 epidemic outbreak frequency. Among them, the regression coefficients of building
264 density, volume ratio, commercial facility density, the density of public service facilities, and the
265 density of transportation facilities were positive overall, indicating that there was an overall positive
266 relationship between these five elements and the frequency of COVID-19 outbreaks in the study area.
267 The regression coefficients of water and greenery coverage were negative, indicating a negative

268 correlation between this element and the frequency of COVID-19 outbreaks in the study area (Figure
269 3).

270 Figure 3. Geographically weighted regression model local R² (A), building density regression
271 coefficients (B), volume ratio regression coefficients (C), water and greenery coverage (D), density
272 of commercial facility density (E), density of public service facilities (F), density of transportation
273 facilities (G).

274 **3.2.2 Unordered multi-categorical logistic regression analysis**

275 In the unordered multi-categorical logistic regression analysis, first, the epidemic outbreak frequency
276 Y1 was taken as the dependent variable. Y1 was divided into four groups according to the quartile
277 method, i.e. $Y1 < 0.021$, $0.021 \leq Y1 < 0.024$, $0.024 \leq Y1 < 0.027$, and $Y1 \geq 0.027$, for building density,
278 volume ratio, water and greenery coverage, density of commercial facility density, density of public
279 service facilities, and density of transportation facilities. The six representative urban external spatial
280 environmental elements, namely building density, volume ratio, water and greenery coverage, density
281 of commercial facilities, density of public service facilities, and density of transportation facilities,
282 were used as covariates in the logistic regression analysis. Second, SPSS was used to set the group
283 with the smallest value of $Y1 < 0.021$ as the reference term, while the other three groups were
284 classified as high frequency, medium frequency and low frequency in the order of high to low. The
285 above data were subjected to unordered multi-categorical logistic regression analysis to calculate
286 their validity and further investigate the relationships between changes in the classification of
287 outbreak frequency Y1 in different groups and changes in building density X1, volume ratio X2,
288 water and greenery coverage, and the density of transportation facilities. The correlations between the
289 six representative external spatial environmental elements, namely building density X1, volume ratio
290 X2, water and greenery coverage X3, density of commercial facilities X4, density of public service
291 facilities X5, and density of transportation facilities X6, were further investigated.

292 When $0.021 \leq Y1 < 0.024$, an unordered multi-categorical logistic regression analysis was conducted
293 with $Y1 < 0.021$ as the benchmark. After model testing and screening, the significance of building
294 density X1, volume ratio X2, and water and green coverage X3 was <0.05 , passing the significance
295 test (Table 4). This implied that the latter three elements had a significant impact on the classification
296 of epidemic outbreak frequency Y1, as follows: an increase in building density X1 caused epidemic
297 outbreak frequency Y1 to move from the $Y1 < 0.021$ group to the $0.021 \leq Y1 < 0.024$ group; while
298 an increase in volume ratio X2 and water and greenery coverage X3 caused epidemic outbreak
299 frequency Y1 to move from the $0.021 \leq Y1 < 0.024$ group to the $Y1 < 0.021$ group.

300 Table 2. Introduction of covariates to the model.

301 When $0.024 \leq Y1 < 0.027$, an unordered multi-categorical logistic regression analysis was conducted
302 with $Y1 < 0.021$ as the benchmark. After model testing and screening, the significance of volume
303 ratio X2, water and greenery coverage X3, and density of public service facilities X5 was <0.05 ,
304 passing the significance test (Table 3). This implied that the latter three elements had a significant
305 effect on the classification of epidemic outbreak frequency Y1. An increase in volume ratio X2,
306 water and greenery coverage X3, and density of public service facilities X5 would result in a shift in
307 epidemic outbreak frequency Y1 from the $0.024 \leq Y1 < 0.027$ group to the $Y1 < 0.021$ group.

308 Table 3. Introduction of covariates to the model.

309 When $Y_1 \geq 0.027$, an unordered multi-categorical logistic regression analysis was conducted with Y_1
310 < 0.021 as the benchmark. After model testing and screening, the significance of building density X_1 ,
311 volume ratio X_2 , water and greenery coverage X_3 , and density of public service facilities X_5 was
312 < 0.05 , passing the significance test (Table 4). This implied that these four elements had a significant
313 impact on the classification of epidemic outbreak frequency Y_1 , as follows: an increase in building
314 density X_1 caused epidemic outbreak frequency Y_1 to move from the $Y_1 < 0.021$ group to the $Y_1 \geq$
315 0.027 group, while an increase in volume ratio X_2 , water and greenery coverage X_3 , and density of
316 public service facilities X_5 caused epidemic outbreak frequency Y_1 to move from the $Y_1 \geq 0.027$
317 group to the $Y_1 < 0.021$ group.

318 Table 4. Introduction of covariates to the model

319 **3.3 Correlation analysis of epidemic outbreak frequency and internal spatial environment**

320 **3.3.1 Cluster analysis**

321 Using SPSS software, K-means cluster analysis was conducted for 344 disease-affected
322 neighbourhoods in the study area. We used four elements, namely housing price, building age,
323 number of buildings, and number of households, to classify the disease-related neighbourhoods.

324 The clustering results were obtained by K-means calculation. The final clustering centres for housing
325 price, building age, number of buildings, and number of households in the first cluster were at 54,892,
326 1,991, 18, and 1,904, respectively; in the second cluster, the clustering centres of housing price,
327 building age, number of buildings, and number of households were 81,800, 1,986, 14, and 1,392,
328 respectively; in the third cluster, the clustering centres of housing price, building age, number of
329 buildings, and number of households were 107,399, 1,986, 16, and 1,204, respectively. In the three
330 clusters, the mean differences were significantly different and showed an increasing trend, indicating
331 that clusters 1, 2, and 3 characterised low-end, mid-end, and high-end neighbourhoods, respectively.
332 Based on the number of cases in each cluster, 185, 101, and 58 low-end, mid-end, and high-end plots,
333 respectively, were involved in the epidemic within the study area (Table 5).

334 Table 5. Final clustering centres.

335 **3.3.2 Binomial logistic regression analysis**

336 Based on the results of the above cluster analysis, epidemic-affected neighbourhoods in Beijing were
337 classified into three classes: low, medium, and high. Their outbreak frequencies were analysed
338 separately to investigate the correlation between outbreak frequency in different classes of residential
339 neighbourhoods Y and four representative internal environmental factors, namely housing price X_1 ,
340 number of buildings X_2 , number of households X_3 , and building age X_4 .

341 (1) Correlation between frequency of epidemic outbreaks and the internal spatial environment in low-
342 end neighbourhoods

343 First, the epidemic frequency Y_2 was assigned to the low-end neighbourhoods according to the
344 dichotomous method, and the neighbourhoods where the epidemic frequency Y_2 was greater than
345 0.019 was assigned a value of 1, and the opposite was assigned a value of 0. Second, the assigned Y_2
346 was logically regressed with the above four representative internal environmental factors, and the
347 results showed a significance of 0.730. This was greater than the comparative value of 0.05,
348 indicating that there was no significant difference between the predicted and true values and the
349 model fit was good (Chowdhuri et al., 2019).

350 Next, the regression equation model was established using the best predictive elements. A total of
351 four urban spatial environmental factors affecting the probability of events were selected. After the
352 model had been tested and screened, the significance of housing price X1 and household number X3
353 in the model was <0.05 , which passed the significance test and was entered into the equation (Table
354 6). We then concluded that the frequency of epidemic outbreaks in low-end neighbourhoods was
355 positively correlated with housing prices and the number of households: the higher the housing prices
356 and the larger the number of households, the more frequent the epidemic outbreaks.

357 Table 6. Hosmer-Lemeshaw test and Introduction of covariates in the model.

358 (2) Correlation between the frequency of epidemic outbreaks and the internal spatial environment in
359 mid-range neighbourhoods

360 First, the epidemic frequency Y3 was assigned to the middle-grade neighbourhoods according to the
361 dichotomous method, and the neighbourhoods where the epidemic frequency Y3 was greater than
362 0.019 was assigned a value of 1, and the opposite was assigned a value of 0. Second, the assigned Y3
363 was logically regressed against the above four representative internal environmental factors, and
364 the results showed a significance of 0.853 . This is greater than the comparative value of 0.05,
365 indicating that there was no significant difference between the predicted and true values and the
366 model fit was good.

367 Next, the regression equation model was established using the best predictive elements. A total of
368 four urban spatial environmental elements that affected the probability of events were selected. After
369 the model had been tested and screened, the significance of the number of households X3 in the
370 model was <0.05 , which passed the significance test and was entered into the equation (Table 7). We
371 concluded that the frequency of epidemic outbreaks in mid-range neighbourhoods was positively
372 correlated with the number of households: the more households, the more frequent the epidemic
373 outbreaks.

374 Table 7. Hosmer-Lemeshaw test and Introduction of covariates to the model.

375 (3) Correlation between the frequency of epidemic disease outbreaks and the internal spatial
376 environment in high-end neighbourhoods

377 First, epidemic outbreak frequency Y5 was assigned to cluster 1 if it was greater than 0.019 and 0 if
378 it was greater than 0.019. Second, the assigned Y5 was logically regressed against the above four
379 representative internal environmental factors, and the results showed a significance of 0.036 (Table
380 8). As this is less than the comparative value of 0.05, we concluded that there was a significant
381 difference between the predicted and true values and the model fit was poor. Therefore, there was no
382 significant correlation between the frequency of epidemic outbreaks in upscale neighbourhoods and
383 the housing price, building age, number of buildings, and number of households in the
384 neighbourhoods.

385 Table 8. Hosmer-Lemeshaw test.

386 **3.4 Test of correlation between epidemic outbreak frequency and internal and external
387 spatial environments**

388 **3.4.1 Correlation between epidemic outbreak frequency and external spatial environment**

389 The Matlab data processing tool was used to construct a three-layer BP neural network model of the
390 frequency of COVID-19 outbreaks and external spatial environmental elements. Indexes for the six
391 spatial environmental elements in 2,769 grids within the fifth ring road of Beijing were used as input
392 layers. The training set, validation set and test set were set according to the ratios of 70%, 15% and
393 15%. Next, the frequency of COVID-19 outbreaks was used as the output layer, and 10 hidden layers
394 were set to construct the topology (Figure 4).

395 Figure 4. BP neural network node diagram.

396 Using the Levenberg-Marquardt algorithm, the optimal training result was obtained at the 42nd
397 training iteration (Figure 6). The overall fit probability parameter was about 0.32, the fit probability
398 parameters for the validation and test sets were 0.25 and 0.30, respectively, and the fit probability
399 parameter R for the training set was 0.34 (Figure 7). This indicated that there was a correlation
400 between the frequency of COVID-19 outbreaks and the external spatial environmental elements.

401 To measure the importance of external spatial environmental elements to epidemic outbreak
402 frequency, this study used the SPSS data processing tool to construct a BP neural network multi-layer
403 perceptron. First, epidemic outbreak frequency and external spatial environmental elements were
404 input, and the training set, validation set and test set were set according to the ratios of 70%, 15% and
405 15%. The results showed that the importance ratios of water and greenery coverage and density of
406 commercial facilities were 0.232 and 0.203, respectively, which were ranked relatively high; the
407 importance ratios of public service facility density and volume ratio were 0.179 and 0.155,
408 respectively, which were ranked in the middle; and the importance ratios of building density and
409 transportation facility density were the lowest, at 0.144 and 0.088, respectively. According to the BP
410 neural network calculation, the importance ranking of the elements based on the BP neural network
411 converged with the results of the geographically weighted regression analysis (Figure 5).

412 Figure 5. BP neural network training effect (A), Regression of neural network (B) and Importance
413 ranking of external spatial environmental elements (C).

414 **3.4.2 Test of correlation between epidemic outbreak frequency and internal spatial
415 environment**

416 Based on the results of the cluster analysis, a three-layer BP neural network model was constructed
417 using the Matlab data processing tool for the frequency of epidemic outbreaks in the two classes of
418 epidemic-affected neighbourhoods in the study area, low-end and mid-end, and the internal
419 environmental elements of the neighbourhoods, respectively. No significant correlation was found
420 between the frequency of epidemic outbreaks and housing price, building age, number of buildings or
421 number of households in the high-end neighbourhoods, so the test was not performed.

422 (1) Correlation between the outbreak situation in low-end neighbourhoods and the internal
423 environment of the neighbourhood

424 The topology was constructed by taking four internal environmental elements of 185 low-end
425 neighbourhoods in Beijing's fifth ring road as the input layer, setting the training set, validation set
426 and test set according to the ratios of 70%, 15% and 15%, and then taking the COVID-19 outbreak
427 situation as the output layer and setting 10 hidden layers (Figure 4).

428 Using the Levenberg-Marquardt algorithm, the optimal training result was obtained at the 24th
429 training iteration. The overall fit probability parameter was about 0.56, the fit probability parameters
430 for the validation and test sets were 0.29 and 0.45, respectively, and the fit probability parameter R
431 for the training set was 0.71. These results indicated that the training effect was good, and they
432 verified that there was a strong correlation between the frequency of COVID-19 outbreaks in low-
433 end neighbourhoods and the internal environmental elements of the neighbourhoods.

434 To calculate the importance of environmental elements within the low-end cell for the frequency of
435 epidemic outbreaks, we used SPSS data processing tools to construct a BP neural network multi-
436 layer perceptron. Inputted low-end neighbourhood outbreak frequency and neighbourhood internal
437 environmental elements were used to set the training set, validation set and test set according to the
438 ratios of 70%, 15% and 15%. The results showed that the importance of the number of households
439 and housing price were 0.432 and 0.342, respectively, which were ranked relatively high; and the
440 importance of the number of buildings and building age were 0.157 and 0.097, respectively, which
441 were ranked low. The importance ranking of configuration elements calculated based on the BP
442 neural network was basically consistent with the results of our binomial logistic regression analysis
443 (Figure 6).

444 Figure 6. BP neural network training effect (A), Regression of neural network (B) and Importance
445 ranking of external spatial environmental elements (C).

446 (2) Correlation between the outbreak situation in mid-range neighbourhoods and internal
447 neighbourhood environmental elements

448 The topology was constructed by taking the four focal internal environmental elements of 101 mid-
449 range neighbourhoods in Beijing's 5th Ring Road as the input layer, setting the training set,
450 validation set and test set in the ratios of 70%, 15% and 15%, and then taking the COVID-19
451 outbreak situation as the output layer and setting 10 hidden layers (Figure 4).

452 Using the Levenberg-Marquardt algorithm, the optimal training result was obtained at the 10th
453 training iteration. The overall fit probability parameter was about 0.65, the fit probability parameters
454 for the validation and test sets were 0.56 and 0.09, respectively, and the fit probability parameter R
455 for the training set was 0.84. These results indicated that the training effect was good, and that the
456 frequency of COVID-19 outbreaks in mid-range neighbourhoods had a significant correlation with
457 the internal environmental elements of the neighbourhoods.

458 To calculate the importance of the internal environmental elements of mid-range neighbourhoods to
459 the frequency of epidemic outbreaks, we used SPSS data processing tools to construct a BP neural
460 network multi-layer perceptron. First, the mid-range neighbourhood outbreak frequency and
461 neighbourhood internal environmental elements were inputted, and the training set, validation set and
462 test set were set according to the ratios of 70%, 15% and 15%. The results showed that the
463 importance ratio of the number of households was 0.531, which was relatively high in the ranking;
464 and the importance ratios of building age, the number of buildings and housing price were 0.311,
465 0.104 and 0.054, respectively, which were low in the ranking. The importance ranking of
466 configuration elements calculated based on the BP neural network was basically consistent with the
467 results of our binomial logistic regression analysis (Figure 7).

468 Figure 7. BP neural network training effect (A), Regression of neural network (B) and Importance
469 ranking of external spatial environmental elements (C).

470 **4 Discussion**

471 **4.1 Correlation of epidemic outbreak frequency with the external spatial environment**

472 The correlation between epidemic outbreak frequency and the external urban spatial environment
473 was analysed, and the results revealed a certain correlation between urban spatial environmental
474 elements and epidemic outbreak frequency. Building density, volume ratio, density of commercial
475 facilities, density of public service facilities, and density of transportation facilities showed a positive
476 correlation with epidemic outbreak frequency, and water and greenery coverage showed a negative
477 correlation with epidemic outbreak frequency. Most previous studies have focused on these six
478 aspects of urban spatial conditions to investigate the objective factors that exacerbate the spread of
479 epidemics, seeking to enhance cities' attention and efforts to ensure safety.

480 At the level of urban spatial development intensity, the main urban spatial elements studied included
481 building density and volume ratio, both of which showed a positive correlation with the frequency of
482 epidemic outbreaks. The higher the degree of aggregation of urban space, the more vulnerable it was
483 to various external shocks, especially sudden shocks, and the higher the risk of urban security (Pan,
484 2020). The main manifestation of this was that the higher the degree of aggregation, the higher the
485 risk was bound to be, while for infectious epidemics, the higher the intensity of construction meant
486 the stronger the connection to the outside world and the wider the coverage of transmission. Chen et
487 al. (2020) used a Bayesian spatiotemporal model to determine the distribution of new coronary
488 pneumonia cases and its relationship with population migration in Wuhan. After mining data on
489 confirmed cases in the United States and Italy, Kuchler (2020) determined the correlation between
490 the level of communication and infection rates in two cities, New York and Lombardia, based on
491 Facebook data. They showed that a social connectivity index could help epidemiologists predict the
492 spread of infectious diseases. It has been argued that the spread of epidemics in China shows macro
493 and meso-micro spatial aggregation characteristics, with more aggregated infections occurring in
494 public spaces such as markets and shopping malls within cities (Qu et al., 2021). High spatial
495 development intensity, high population density, mobility, and economic activity lay the groundwork
496 for the spread of epidemics.

497 At the level of urban watershed greenery construction, watershed and greenery coverage showed a
498 negative correlation with the frequency of epidemic outbreaks. Urban green space planning is
499 particularly important for cities to inhibit the spread and proliferation of epidemics, where green
500 spaces that are more aggregated and larger in size have a relatively stronger inhibitory capacity (Ca et
501 al., 1998). Establishing a holistic green open space system can improve the urban environment and
502 enhance the ability of cities to cope with natural disasters (Lei, 2022). The proportion of green space
503 is positively correlated with the health of residents (Maas et al., 2006). Second, analysis of the
504 characteristics of epidemic-prone people and their lifestyles, behavioural habits, and living
505 environments has shown that traditional green space design emphasises ornamental functions. To
506 effectively control health risks and improve the physical and mental sub-health of community
507 residents, improve the public conditions of the community and enhance the quality of life of residents
508 (Larson, 2006), space should be differentiated according to the population density and surrounding
509 environment within the service area, enabling designers to determine the scale and functions of
510 different green areas and use the landscape environment to improve public health.

511 At the level of urban service facility construction, the main urban spatial elements analysed included
512 the density of commercial facilities and public service facilities, both of which showed a positive
513 correlation with the frequency of epidemic outbreaks. From the perspective of epidemic prevention,

514 in the construction of service facilities, it is necessary to dismantle, implant, and reorganise spatial
515 use functions according to different behavioural needs to enhance regions' capacity to handle public
516 health emergencies. Important nodes such as hospitals, public places, and households and related
517 factors may have a direct impact on the spread and diffusion of infectious diseases. Using data from
518 public institutions in China, Zhou et al. (2020) analysed the spatial manifestations of disease,
519 population and psychosocial conditions at three scales: individual (spatial surveillance of the
520 epidemic and spatial and temporal movement of infected individuals), group (population movement
521 and spatial distribution) and regional (spatial risk level). The geographic supply and demand of
522 resources. Kuupiel et al. (2020) found that improving the spatial accessibility of health services and
523 control and care points played a significant role in reducing patient access times and diagnosis times.
524 According to Ren et al. (2003) and other scholars, the spatial distribution of these public places is
525 influenced by urban development, and in order to improve the utilisation of resources, they are often
526 clustered in more densely populated areas, increasing the economic level and the frequency of
527 interpersonal activities in the surrounding areas, and also increasing the risk of epidemic transmission.

528 At the level of urban transportation facility construction, the main urban spatial elements involved
529 included the density of transportation facilities, which showed a positive correlation with the
530 frequency of epidemic outbreaks. Urban expansion and development were increasingly dependent on
531 urban rail transportation, but convenient and efficient urban rail transportation was also a fast track
532 for epidemic transmission. It has been suggested that spatial elements such as transportation, travel
533 and commuting may also have an impact on epidemic development under the effect of migration of
534 various elements and change the spatial association of neighbourhoods to some extent (Li et al.,
535 2021).

536 **4.2 Correlation of epidemic outbreak frequency with the internal spatial environment**

537 Correlation analysis between epidemic outbreak frequency and inner urban spatial environmental
538 elements revealed a correlation between elements of the inner urban spatial environment and
539 epidemic outbreak frequency. Epidemic outbreak frequency was significantly correlated with the
540 inner environment of low- and mid-end neighbourhoods, but there was no significant correlation with
541 the inner environment of high-end neighbourhoods. Some studies have shown that the characteristics
542 of the age of the community affect the spatial openness of the community to a certain extent (Yu et
543 al., 2020), and that the spatial exposure of old communities is much higher than that of new
544 communities (Shao&Fan, 2020). The age, quality, property level, and spatial environment of the
545 community were used to categorise community grade levels, indicating a correlation between
546 neighbourhoods in different grades of the community and the frequency of epidemic outbreaks.

547 The frequency of epidemic outbreaks in low-end neighbourhoods was positively correlated with
548 housing price and number of households, probably due to the fact that the low-end neighbourhoods
549 were mostly located in more remote locations and were home to low-income residents with a longer
550 commuting time. The more distant a location is, the lower the housing price and the greater the
551 commuting distance and time, so the number of households, i.e. population density, will be relatively
552 low, resulting in a decrease in the chance of epidemic infection among residents. The frequency of
553 outbreaks in mid-range neighbourhoods was positively correlated with the number of households: the
554 more households in the neighbourhood, the higher the population density, and COVID-19 followed
555 an aggregated transmission route, mainly in households and various types of public spaces (Wang et
556 al., 2020). Areas with high population densities and mobility are more vulnerable to virus
557 transmission (Chen et al., 2020), so with an increase in the number of households, the frequency of
558 population movement and aggregation in public places increases, and the frequency of epidemic

559 outbreaks increases. In contrast, the frequency of outbreaks in upscale neighbourhoods was not
560 significantly correlated with the internal environment, probably due to the fact that upscale
561 settlements are usually more strictly controlled and have relatively good medical and health facilities,
562 which can effectively prevent and control epidemic outbreaks, thus showing no significant
563 correlation between the frequency of epidemic outbreaks in upscale neighbourhoods and the internal
564 environment.

565 **5 Conclusion**

566 This study used geographically weighted regression, unordered multi-analytic logistic regression,
567 cluster analysis, binomial logistic regression, and a BP neural network with machine learning to
568 investigate the correlation between epidemic outbreak frequency and the internal and external spatial
569 environment of cities. We drew the following conclusions.

570 First, the results of the geographically weighted regression showed that the frequency of epidemic
571 outbreaks was correlated with the external spatial environment of the city. Elements of the urban
572 external spatial environment, such as building density, volume ratio, density of commercial facilities,
573 density of public service facilities, and density of transportation facilities, were positively correlated
574 with epidemic outbreak frequency, while water and greenery coverage was negatively correlated with
575 epidemic outbreak frequency.

576 Second, the unordered multi-analytic logistic regression analysis showed that there was a correlation
577 between the outbreak frequencies in different groups, their patterns of classification and the external
578 spatial environment. When Y1 was in the low-frequency group, the building density increased to a
579 certain value, and epidemic outbreak frequency moved from the reference group to the low-
580 frequency group. When the volume ratio and water and green coverage increased to a certain value,
581 epidemic outbreak frequency moved from the low-frequency group to the reference group. When Y1
582 was in the medium-frequency group, the volume ratio, water and green coverage, and density of
583 public service facilities increased to a certain value, and epidemic outbreak frequency moved from
584 the medium-frequency group to the reference group. When Y1 was in the medium frequency group
585 and the building density increased to a certain value, the epidemic outbreak frequency moved from
586 the reference group to the high-frequency group; when the volume ratio, water and greenery coverage,
587 and density of public service facilities increased to a certain value, epidemic outbreak frequency
588 moved from the high-frequency group to the reference group.

589 Third, the correlation between the frequency of epidemic outbreaks and the internal environmental
590 elements of the neighbourhood varied among different grades of neighbourhoods. The frequency of
591 epidemic outbreaks in low-end neighbourhoods was significantly correlated with the internal
592 environmental factors, among which house price and the number of households were positively
593 correlated with the frequency of outbreaks. The higher the house prices and the more households in
594 low-end neighbourhoods, the more frequent the epidemic outbreaks. The frequency of epidemic
595 outbreaks in mid-range neighbourhoods was significantly correlated with internal environmental
596 factors. The number of households in mid-range neighbourhoods was positively correlated with the
597 frequency of outbreaks: the larger the number of households, the more frequent the epidemic
598 outbreaks. There was no significant correlation between the frequency of epidemic outbreaks in high-
599 end neighbourhoods and the internal environmental factors of these neighbourhoods.

600 Fourth, our machine learning test revealed that epidemic outbreak frequency was correlated with both
601 the internal and the external spatial environments of the city. The test results showed that there was a

602 correlation between epidemic outbreak frequency and the external spatial environment, among which
603 water and greenery coverage and density of commercial facilities had the greatest influence on
604 epidemic outbreak frequency, followed by density of public service facilities and volume ratio;
605 building density and density of transportation facilities were the least influential. The frequency of
606 epidemic outbreaks was thus correlated with the external spatial environment. The number of
607 households and housing price had the greatest influence on the frequency of epidemic outbreaks in
608 low-end neighbourhoods, followed by the number of buildings and building age.

609 **Conflict of Interest**

610 *The authors declare that the research was conducted in the absence of any commercial or financial
611 relationships that could be construed as a potential conflict of interest.*

612 **Reference**

- 613 [1] Ahmad, T., Pencina, M. J., Schulte, P. J., O'Brien, E., Whellan, D. J., Piña, I. L., Kitzman, D. W.,
614 Lee, K. L., O'Connor, C. M., & Felker, G. M. (2014). Clinical Implications of
615 Chronic Heart Failure Phenotypes Defined by Cluster Analysis. *Journal of the American College
616 of Cardiology*, 64(17), 1765–1774. <https://doi.org/10.1016/j.jacc.2014.07.979>
- 617 [2] Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted
618 regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281–298.
619 <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- 620 [3] Ca, V. T., Asaeda, T., & Abu, E. M. (1998). Reductions in air conditioning energy caused by a
621 nearby park. *Energy and Buildings*, 29(1), 83–92. [https://doi.org/10.1016/S0378-7788\(98\)00032-2](https://doi.org/10.1016/S0378-7788(98)00032-2)
- 622 [4] Cash, R., & Patel, V. (2020). Has COVID-19 subverted global health? *The Lancet*, 395(10238),
623 1687–1688. [https://doi.org/10.1016/S0140-6736\(20\)31089-8](https://doi.org/10.1016/S0140-6736(20)31089-8)
- 624 [5] Chao, J., & Zhang, Q. (2022). Study on the dynamic behaviour of epidemic models affected by
625 pollution environment. *Journal of Henan Normal University (Natural Science Edition)* (04), 54–59.
626 doi:10.16366/j.cnki.1000-2367.2022.04.008.
- 627 [6] Chen, S., Yang, J., Yang, W., Wang, C., & Bärnighausen, T. (2020). COVID-19 control in China
628 during mass population movements at New Year. *The Lancet*, 395. [https://doi.org/10.1016/S0140-6736\(20\)30421-9](https://doi.org/10.1016/S0140-
629 6736(20)30421-9)
- 630 [7] Chen, Z.-L., Zhang, Q., Lu, Y., Guo, Z.-M., Zhang, X., Zhang, W.-J., Guo, C., Liao, C.-H., Li,
631 Q.-L., Han, X.-H., & Lu, J.-H. (2020). Distribution of the COVID-19 epidemic and correlation with
632 population emigration from Wuhan, China. *Chinese Medical Journal*, 133(9), 1044–1050.
633 <https://doi.org/10.1097/CM9.0000000000000782>
- 634 [8] Chen, X., Liao, B., Cheng, L., Peng, X., Xu, X., Li, Y., Hu, T., Li, J., Zhou, X., & Ren, B. (2020).
635 The microbial coinfection in COVID-19. *Applied Microbiology and Biotechnology*, 104(18), 7777–
636 7785. <https://doi.org/10.1007/s00253-020-10814-6>
- 637 [9] Chowdhuri, I., Pal, S. C., & Chakrabortty, R. (2020). Flood susceptibility mapping by ensemble
638 evidential belief function and binomial logistic regression model on river basin of eastern India.
639 *Advances in Space Research*, 65(5), 1466–1489. <https://doi.org/10.1016/j.asr.2019.12.003>

- 640 [10] Goldstein, H (2003) Multilevel Statistica Models, 3rd ed, London: Arnold.
- 641 [11] Eastin, C., & Eastin, T. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China.
642 The Journal of Emergency Medicine, 58(4), 711–712.
643 <https://doi.org/10.1016/j.jemermed.2020.04.004>
- 644 [12] Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does Density Aggravate the COVID-19
645 Pandemic?: Early Findings and Lessons for Planners. Journal of the American Planning Association,
646 86, 1–15. <https://doi.org/10.1080/01944363.2020.1777891>
- 647 [13] Honey-Rosés, J., Anguelovski, I., Chireh, V. K., Daher, C., Konijnendijk van den Bosch, C., Litt,
648 J. S., Mawani, V., McCall, M. K., Orellana, A., Oscilowicz, E., Sánchez, U., Senbel, M., Tan, X.,
649 Villagomez, E., Zapata, O., & Nieuwenhuijsen, M. J. (2021). The impact of COVID-19 on public
650 space: An early review of the emerging questions – design, perceptions and inequities. Cities &
651 Health, 5(sup1), S263–S279. <https://doi.org/10.1080/23748834.2020.1780074>
- 652 [14] Huang, L., Chen, Y., Xiao, J., Luo, W., Li, F., Wang, Y., Wang, Y., & Wang, Y. (2020).
653 Progress in the Research and Development of Anti-COVID-19 Drugs. Frontiers in Public Health, 8,
654 365. <https://doi.org/10.3389/fpubh.2020.00365>
- 655 [15] Kuchler, T., Russel, D., & Stroebel, J. (2020). The Geographic Spread of COVID-19 Correlates
656 with the Structure of Social Networks as Measured by Facebook (Working Paper No. 26990).
657 National Bureau of Economic Research. <https://doi.org/10.3386/w26990>
- 658 [16] Kuupiel, D., Adu, K. M., Bawontuo, V., Adogboba, D. A., Drain, P. K., Moshabela, M., &
659 Mashamba-Thompson, T. P. (2020). Geographical Accessibility to Glucose-6-Phosphate
660 Dioxygenase Deficiency Point-of-Care Testing for Antenatal Care in Ghana. Diagnostics, 10(4),
661 Article 4. <https://doi.org/10.3390/diagnostics10040229>
- 662 [17] Larson, J. (2006). A comparative study of community garden systems in Germany and the
663 United States and their role in creating sustainable communities. Arboricultural Journal, 29.
664 <https://doi.org/10.1080/03071375.2006.9747450>
- 665 [18] Lee, J., Kim, J., & Kwon, H.-D. (2013). Optimal control of an influenza model with seasonal
666 forcing and age-dependent transmission rates. Journal of Theoretical Biology, 317, 310–320.
667 <https://doi.org/10.1016/j.jtbi.2012.10.032>
- 668 [19] Lei, W. (2022). Urban spatial planning response strategies under sudden public health
669 events. Anhui Architecture (04), 179–180. doi:10.16330/j.cnki.1007-7359.2022.04.075.
- 670 [20] Li, X., Zhou, L., Jia, T., et al. (2021). Decoding the impact of potential urban risk factors on the
671 COVID-19 situation: A case study of WuHan. Urban Planning (08), 78–86.
- 672 [21] Lin, C., Lau, A. K. H., Fung, J. C. H., Guo, C., Chan, J. W. M., Yeung, D. W., Zhang, Y., Bo,
673 Y., Hossain, M. S., Zeng, Y., & Lao, X. Q. (2020). A mechanism-based parameterisation scheme to
674 investigate the association between transmission rate of COVID-19 and meteorological factors on
675 plains in China. Science of The Total Environment, 737, 140348.
676 <https://doi.org/10.1016/j.scitotenv.2020.140348>

- 677 [22] Liu, M., & Liang, W.(2022). Improving the prevention and control capacity of urban outbreaks -
678 an example of prevention and control of the new crown pneumonia outbreak in the capital. *Frontline*
679 (10),28-31.
- 680 [23] Maas, J., Verheij, R. A., Groenewegen, P. P., de Vries, S., & Spreeuwenberg, P. (2006). Green
681 space, urbanity, and health: How strong is the relation? *Journal of Epidemiology and Community*
682 *Health*, 60(7), 587–592. <https://doi.org/10.1136/jech.2005.043125>
- 683 [24] Pan, J.(2020). Security risks of urban spatial agglomeration and their prevention and control-
684 based on the prevention and control of the COVID-19 epidemic. *People's Frontier* (04),11-17.
685 doi:10.16619/j.cnki.rmltxsqy.2020.04.002.
- 686 [25] Prosper, O., Saucedo, O., Thompson, D., Torres-Garcia, G., Wang, X., & Castillo-Chavez, C.
687 (2010). Modeling control strategies for concurrent epidemics of seasonal and pandemic H1N1
688 influenza. *Mathematical Biosciences & Engineering*, 8(1), 141–170.
689 <https://doi.org/10.3934/mbe.2011.8.141>
- 690 [26] Qu, C., Zhang, J., & Jia, W.(2021). Analysis of the strategy of urban space environment
691 optimization under the thought of epidemic prevention. *Architecture and Culture* (04),166-168.
692 doi:10.19875/j.cnki.jzywh.2021.04.056.
- 693 [27] Ren, Z., Jin, J., Bao, S., et al.(2003). Summary of the Forum on SARS and City. *Urban*
694 *Development Studies* (04),1-7.
- 695 [28] Shao, Q., & Fan, X.(2020). Renovation opportunities of Old Communities after the
696 Epidemic. *Construction Technology* (14),68-70. doi:10.16116/j.cnki.jskj.2020.14.015.
- 697 [29] Sharifi, A. (2020). Urban Resilience Assessment: Mapping Knowledge Structure and Trends.
698 *Sustainability*, 12(15), Article 15. <https://doi.org/10.3390/su12155918>
- 699 [30] Turan, B., Ayataç, H., & İnce, E. C. (2022). Estimating the Modal Split Ratios of Square
700 Visitors in the Case of Istanbul's Historical Peninsula: Evidence from Multinomial Logistic
701 Regression Model. *Journal of Urban Planning and Development*, 148(4), 05022037.
702 [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000880](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000880)
- 703 [31] Wang, L., Jia, Y., Li, X., & Yang, X.(2020). Urban Spatial Intervention Strategies for Infectious
704 disease prevention and control. *Urban Planning* (08),13-20+32.
- 705 [32] Wang, Z., Zhao, Y., & Zhang, F. (2022). Simulating the Spatial Heterogeneity of Housing Prices
706 in Wuhan, China, by Regionally Geographically Weighted Regression. *ISPRS International Journal*
707 *of Geo-Information*, 11(2), 129. <https://doi.org/10.3390/ijgi11020129>
- 708 [33] Yu, H., Li, J., Bardin, S., Gu, H., & Fan, C. (2021). Spatiotemporal Dynamic of COVID-19
709 Diffusion in China: A Dynamic Spatial Autoregressive Model Analysis. *ISPRS International Journal*
710 *of Geo-Information*, 10(8), 510. <https://doi.org/10.3390/ijgi10080510>
- 711 [34] Yu, Q., He, D., Huang, J., & Ran, J.(2020). Practice and reflection on the renovation of old
712 communities under the epidemic. *China Survey and Design* (04),68-72.

713 [35] Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Cao, Z., Wang, J., Yuan, W., Zhu, Y., Song,
 714 C., Chen, J., Xu, J., Li, F., Ma, T., Jiang, L., Yan, F., Yi, J., Hu, Y., ... Xiao, H. (2020). COVID-19:
 715 Challenges to GIS with Big Dat. *Geography and Sustainability*, 1(1), 77–87.

716 **Table**

717 Table 1. GWR Model Determination Coefficients.

R²	R²Adjusted	AICc	Residual squares	Bandwidth (km)
0.792374	0.735484	-4862.462265	19.40998	1717.829009
Projects	Average value	Median	Maximum value	Minimum value
Local R2	0.178944	0.168504	0.574433	0.008505
Building density	0.093273	-0.061378	1.984351	-1.200134
Volume ratio	0.023199	0.024738	0.444783	-0.4181
Water and greenery coverage	-0.022804	-0.023266	0.584145	-0.543087
Density of commercial facilities	0.292249	0.053153	1.420713	-0.98085
Density of public service facilities	0.036014	0.029701	3.324738	-3.515115

718 Table 2. Introduction of covariates to the model.

	B	Significance		Significance
Enter the equation elements	X1	1.386	0.038	Not entered
	X2	-0.338	0.002	into the
	X3	-1.669	0.000	equation
	Constants	0.914	0.000	elements

719 Table 3. Introduction of covariates to the model.

	B	Significance		Significance
Enter the equation elements	X2	-0.407	0.000	Not entered
	X3	-2.342	0.000	into the
	X5	-3.361	0.006	equation
	Constants	1.377	0.000	elements

720 Table 4. Introduction of covariates to the model.

	B	Significance		Significance
Enter the equation elements	X1	2.130	0.001	Not entered
	X2	-0.488	0.000	into the
	X3	-2.150	0.000	equation
	X5	-5.412	0.000	elements
	Constants	1.235	0.000	

721 Table 5. Final clustering centres.

	Clustering		
	1	2	3
Housing price	54892	81800	107399
Building Age	1991	1986	1986
Number of buildings	18	14	16
Number of households	1904	1392	1204
	Distance		
	1	2	3
1		26912.940	52511.492
2	26912.940		25599.441
3	52511.492	25599.441	
	Case items		
1		185.000	
2		101.000	
3		58.000	
Effective		344.000	
Missing		0.000	

723 Table 6. Hosmer-Lemeshaw test and Introduction of covariates in the model.

Card side	Degree of freedom			Significance		
	5.258	8		0.730		
	B	Significance			Significance	
Enter the equation elements	X1	0.000	0.034	Not entered into the equation elements	X2	0.777
	X3	0.000	0.037		X4	0.916
					Constants	0.936

724 Table 7. Hosmer-Lemeshaw test and Introduction of covariates to the model.

Card side	Degree of freedom			Significance		
	4.043	8		0.853		
	B	Significance			Significance	
Enter the equation elements	X3	0.001	0.035	Not entered into the equation elements	X1	0.718
					X2	0.552
					X4	0.543
					Constants	0.555

725 Table 8. Hosmer-Lemeshaw test.

Card side	Degree of freedom	Significance
16.518	8	0.036

In review

Figure 1.JPG

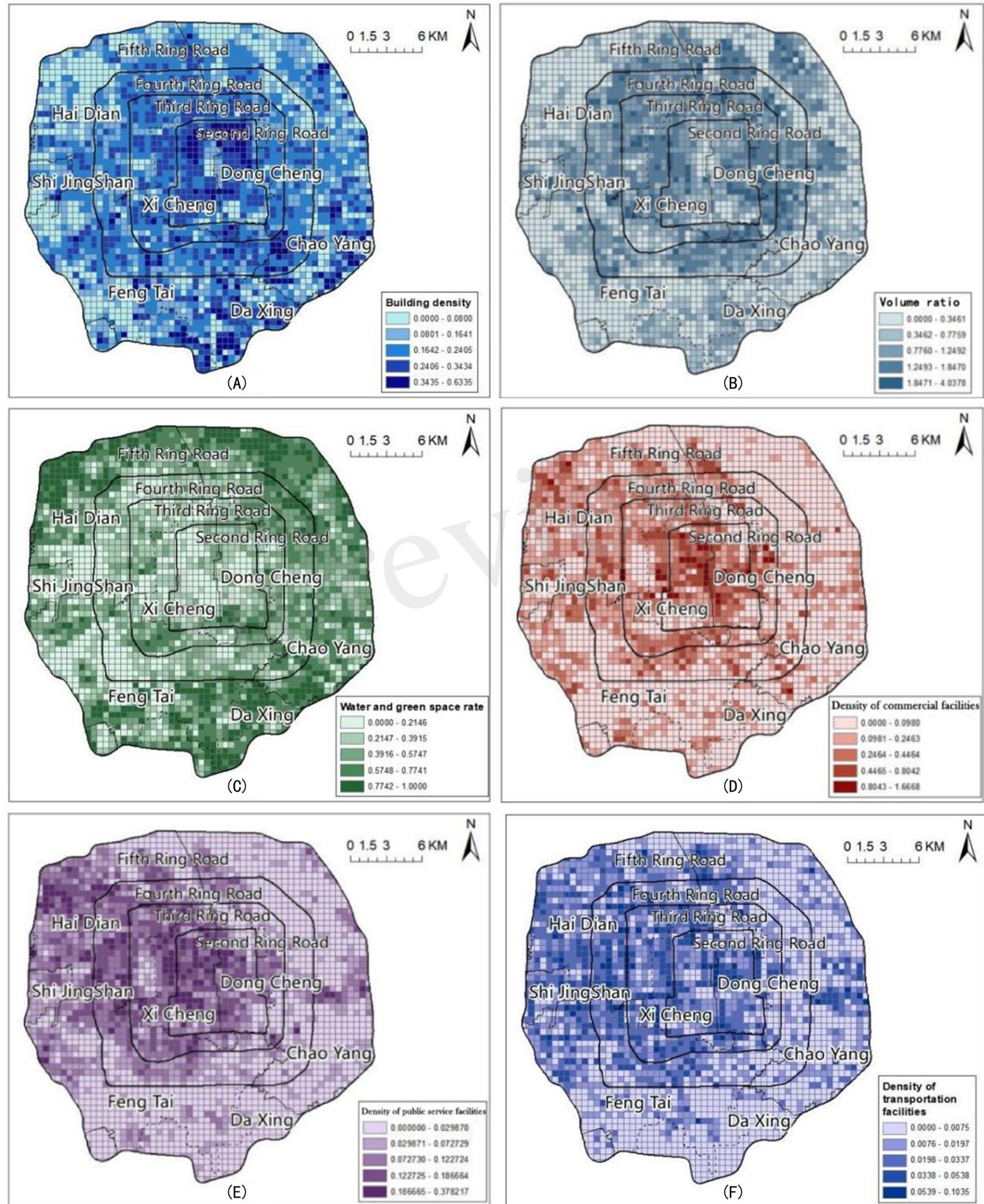


Figure 2.JPG

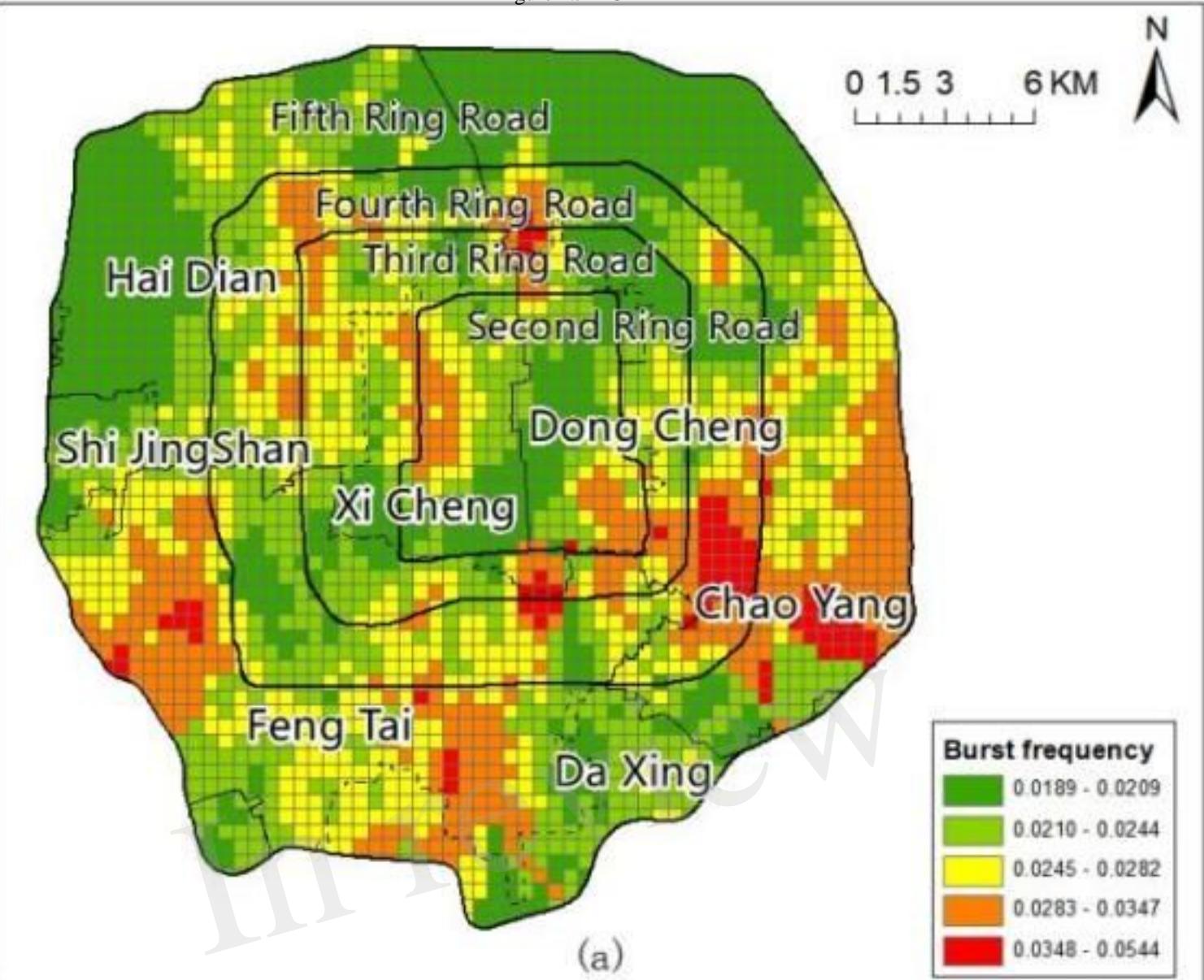


Figure 3.JPG

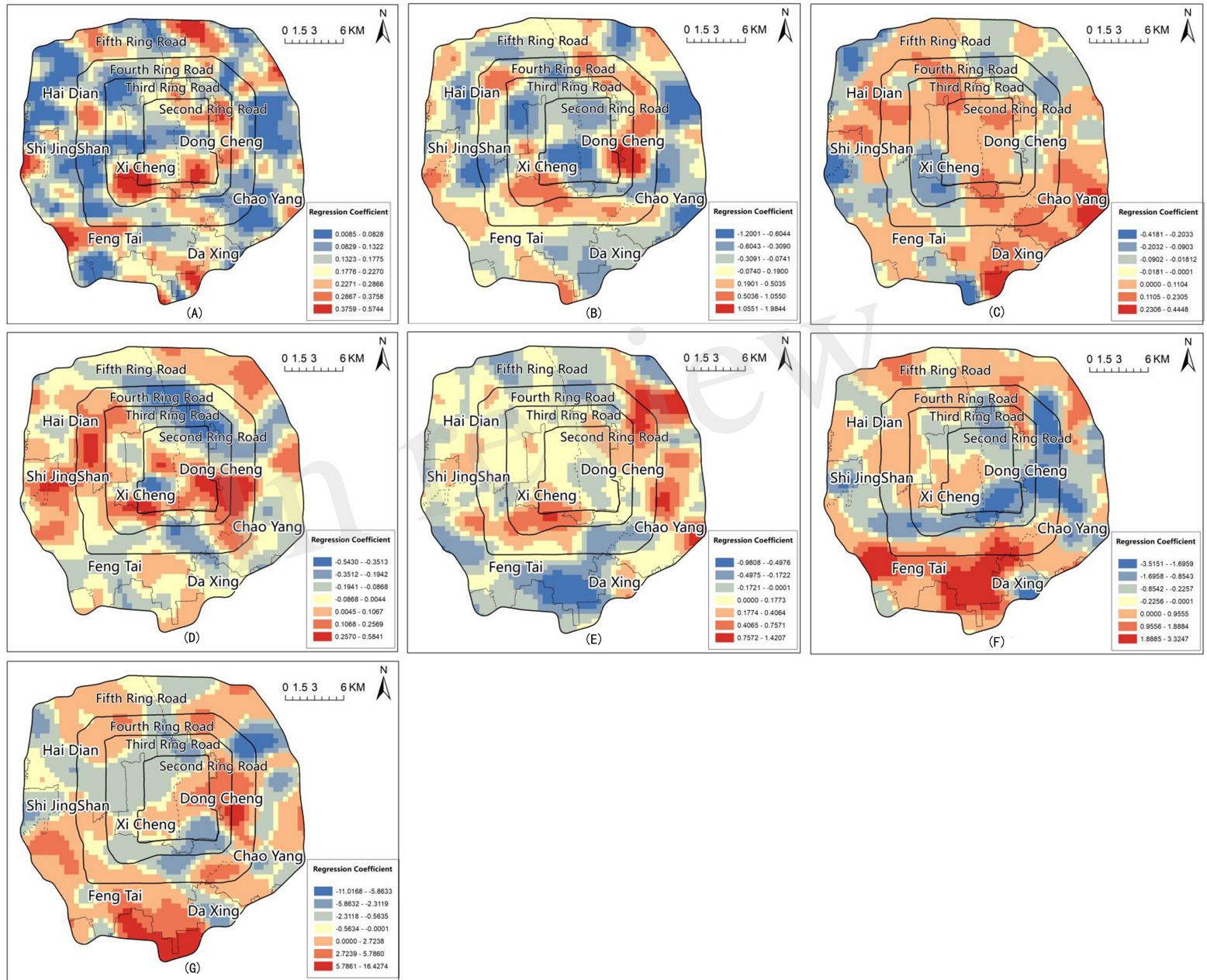


Figure 4.JPG

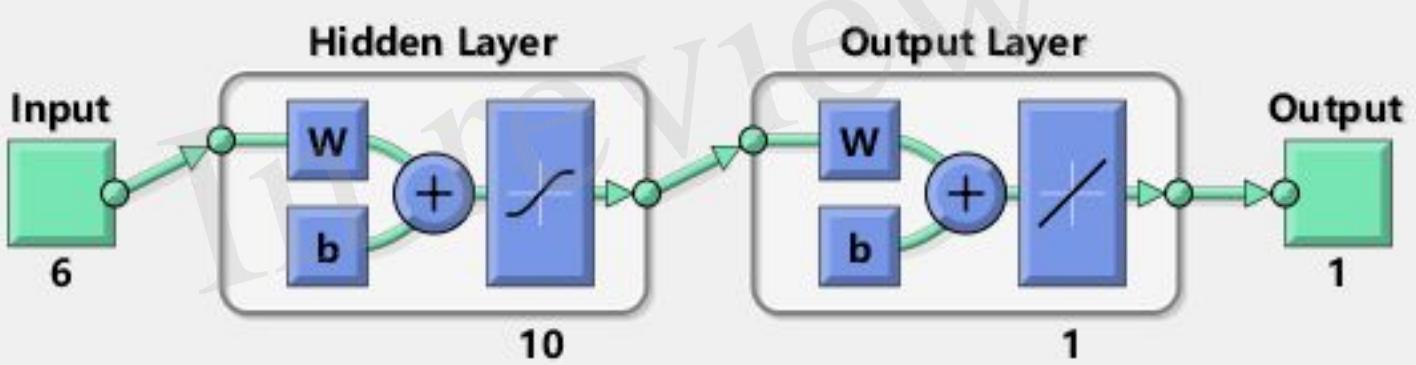


Figure 5.JPG

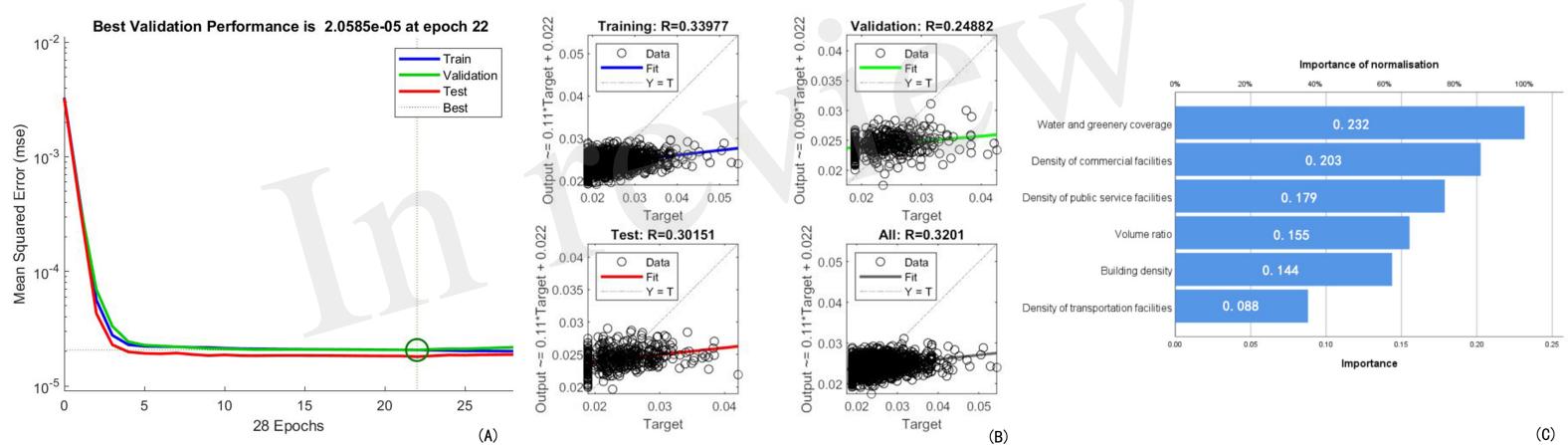


Figure 6.JPG

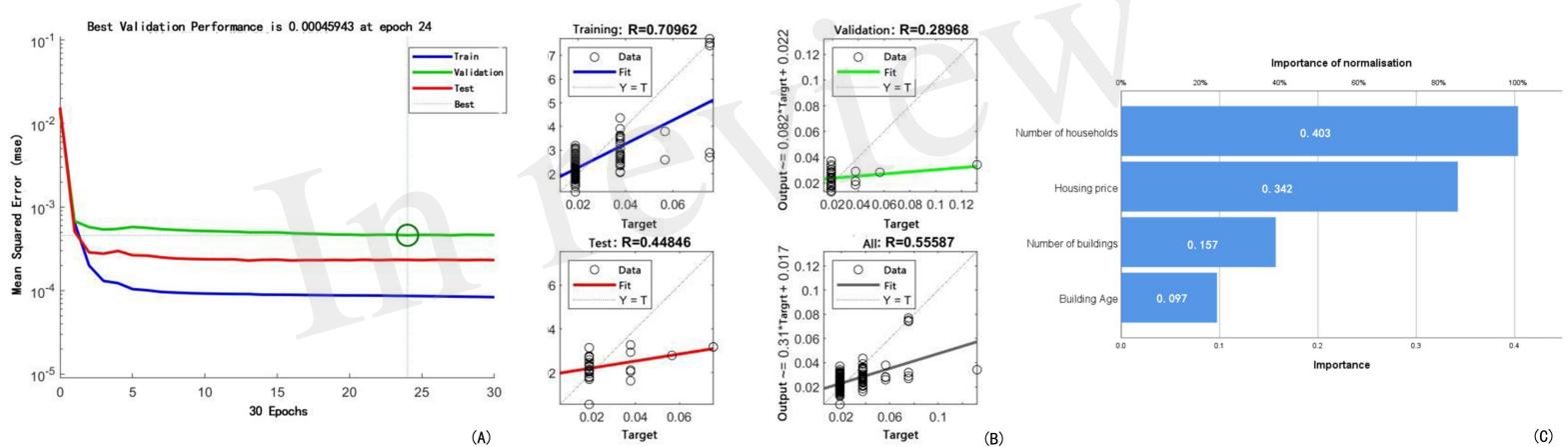


Figure 7.JPG

