ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

Analysis of temporal and spatial distribution characteristics and influencing factors of COVID-19 epidemic in 285 cities in China

Yihua He[1,a], Chongjing Zhu[1,b], Shiyin Chen[1,c], Ya Yang[1,d]

[1]School of Geographical Sciences, Southwest University, Chongqing, China
[a]hyh0027@163.com, [b]zhucj211@gmail.com, [c]3102207914@qq.com, [d]ya1213156@163.com

Abstract: *China is dealing with a serious public health crisis caused by the new coronavirus (COVID-19k). In China, it is essential to investigate the impact of several factors on the diagnosis of COVID-19. Based on data related to the propagation of the COVID-19 epidemic from January 24 to February 28, 2020, this article selected 285 cities in China that have been impacted by the epidemic and analyzed the spatiotemporal distribution characteristics of the epidemic. Additionally, multi-source data such as meteorological environment, social economy, and geographic information were comprehensively considered. On the data, I utilized geographic detectors and random forest algorithms to investigate the most important elements influencing the epidemic's spread. The findings revealed that: (1) The COVID-19 epidemic in China has phase time sequence, spatial aggregation, and distance attenuation features, and the distribution is concentrated in Wuhan; (2) There are two types of COVID-19 infections, non-linear enhancement, and two-factor enhancement. The interaction of natural factors and socioeconomic factors has a stronger impact on the epidemic than the interaction of natural factors. (3) The distance from Wuhan and the number of beds in medical facilities are the two most important elements influencing the epidemic's spread. As a result, timely control of personnel flow, sensible allocation of medical resources, and other actions can successfully stop the epidemic from spreading and have a favorable impact on epidemic prevention and control.*

Keywords: *coronavirus pandemic; spatiotemporal pattern; influencing factors; geographical detector; random forest*

1. Introduction

Since the novel coronavirus (COVID-19) outbreak in Wuhan City, Hubei Province, in December 2019, the pandemic had quickly expanded to other parts of the country, posing a severe risk to public health. The spread of the disease has put people's lives in jeopardy and posed serious difficulties to public health, social activities, and living circumstances. All provinces, municipalities, and autonomous areas around the country began first-level responses to all key public health emergency institutions since January 23, 2020. Governments at all levels have firmly adopted prevention and control measures such as city closures and isolation, thereby curbing the epidemic's spread and offering outstanding contributions to the prevention and control of the epidemic situation in the world. Cities are the primary targets for the transmission of the epidemic since they are the focal point of population and economy within a geographical area. However, as urban populations have become more mobile and trade has expanded, the epidemic's influence has been more widespread. Studying the transmission mode of the COVID-19 epidemic in China, as well as the temporal and spatial evolution characteristics of the rapid development in the early phases of the epidemic, and exploring the driving factors of the epidemic's spread and development among cities, is critical for epidemic prevention, control, and future urban public health security.

Several researches are currently being conducted to investigate the spatiotemporal aggregation characteristics and transmission driving variables of the COVID-19 pandemic in China. Through epidemiological investigations, Lung-Chang Chien et al. [1] established a link between meteorological parameters in high-risk areas in the United States and the COVID-19 pandemic. The research of M.S.I.Mozumder et al. [2] showed that community prevention and control measures have an inhibitory effect on virus infection. Wang Jiao'e et al. [3] summarized the inter-city spatial spread pattern of the epidemic, revealing that factors such as geographic proximity, population, transportation, and epidemic prevention and control have an impact on the epidemic. The impact of spatial diffusion by Gong

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

Shengsheng et al. [4] studied the stage characteristics of the epidemic evolution in Hunan Province, and explored the factor interaction of the epidemic. Chen Xiao et al. [5] explored the spatiotemporal clustering characteristics of cases in Chongqing at the street scale. Based on the data released by Sina Weibo in the core area of Wuhan, Li Xin et al. [6] revealed the spatial distribution pattern of the COVID-19 epidemic and its impact on different urban areas of Wuhan.

In summary, 285 prefecture-level cities in China are selected as the study area. Based on the analysis of the Spatio-temporal distribution characteristics of the epidemic from January 24 to February 28, 2020, the urban COVID-19 diagnosis and new data was combined with meteorological environment data, social and economic data, and geographic information data to explore the key factors that affect the spread of the epidemic. This was done to offer a necessary basis for strengthening the work of epidemic prevention and control and to aid in formulating the policy of epidemic prevention.

2. Materials and methods

2.1. Research materials

Table 1: Indicators and data sources.

Data Category	Data Name	Data Source
Basic Data	Epidemic data	Chinese center for disease control and prevention (http://www.chinacdc.cn/)
	City Administrative Boundaries Vector Data	National Basic Geographic Information System National million database (http://www.ngcc.cn/ngcc/)
Meteorological and environmental data	Average daily temperature(℃)	China Meteorological Data Network (http://data.cma.cn/)
	Air quality index	Ministry of Environmental Protection (http://www.mee.gov.cn/)
Socioeconomic data	GDP per capita (yuan/person)	Chinese Urban statistical yearbook (https://www.tongjinianjian.com)
	Population density (person/km^2)	
	Road network density in built-up areas (km/km^2)	
	Green coverage rate of built-up area (%)	
	Built-up area (km^2)	
	Urban population (10,000 people)	
	Number of beds in medical and health institutions	Seventh National Bureau of Statistics Bulletin of the National Population Census (http://www.stats.gov.cn/)
	Population with high school education per 100,000 people (person/100,000)	
	Distance to Wuhan (km)	Calculated with ArcGIS
Remote sensing data and its products	Night light data	NOAA (https://www.ngdc.noaa.gov/)
	Urban average altitude	DEM Data Acquisition

Wuhan took efforts to close the city on January 23, 2020, to prevent population migration. Except for Wuhan, there have been no additional confirmed cases in cities across the country as of February 28. As a result, data for this article was collected from January 24, 2020, to February 28, 2020. On the 28th, daily COVID-19-related data for 285 Chinese cities (excluding Hong Kong Special Administrative Region of China, Macao Special Administrative Region of China, and Taiwan Region of China) were obtained from the Chinese Center for Disease Control and Prevention's daily data and the provincial and municipal health committees' official websites. Report on the current state of the epidemic. The China Meteorological Data Network and the official website of China's Ministry of Ecology and Environment provided meteorological and environmental data. The socioeconomic data were collected from provincial

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

statistical yearbooks, municipalities directly under the Central Government, and autonomous areas, as well as the National Bureau of Statistics' Seventh National Census Bulletin. ArcGIS software was used to calculate the geographic information data. Table 1 shows the major metrics, data sources, and other essential information for constituencies.

2.2. Analysis method

2.2.1. Spatial analysis of COVID-19 cases

The potential interdependence between observations of some variables at a location and observations at adjacent locations, which can show the resemblance between attribute values of cells in spatially adjacent regions, is referred to as spatial autocorrelation [7]. The Moran index (I) is used in this study to determine the degree of spatial aggregation of the COVID-19 pandemic. Its formula is as follows:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{i,j}(x_i-\bar{x})(x_j-\bar{x})}{s^2\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{i,j}} \tag{1}$$

In the formula, $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$, x_i and x_j represent the observed number of people diagnosed in the i and j regions, and i≠j, $\omega_{i,j}$ represents the binary adjacent spatial weight. Where I∈[-1, 1], I>0 means that the spatial entities in adjacent areas are clustered. I<0 means that the spatial entities are discretely distributed. I=0 means that the spatial entities are randomly distributed, and the larger the absolute value of I, the more significant the correlation of the spatial distribution. The Moran index is calculated using the Z value and the P-value. It is assumed that the epidemic has some degree of spatial clustering if the p-value passes the confidence test and the Z value exceeds the threshold value for rejecting the null hypothesis. If the z value is high and the p-value is small, it indicates that there is high-valued spatial clustering and that the epidemic condition is more dangerous. If the z-value is low, negative, and the p-value is low, it indicates that the spatial clustering is low-valued. The higher (or lower) the z number, the more clustering there is. Based on the development trend of the epidemic situation, this paper separates the research period from January 24 to February 28 into four stages. The LISA cluster map of the confirmed number of cases in cities across the country from January 24th to February 28th was generated using ArcGIS software.

2.2.2. Analysis of Influencing Factors

Geographic detectors can be used to identify the influencing factors and mechanisms of spatial heterogeneity. The phenomenon of Spatial heterogeneity occurs when the sum of variance within layers is less than the total variance between layers. If the spatial distributions of the two independent variables tend to be consistent, then both are in Statistical correlation. When geographic detectors are used for impact factor analysis, the independent variables are usually categorical [8]. The geographic detector includes 4 modules which are, factor detector, risk detector, interaction detector, ecological detector.

The factor detector detects the dependent variable's spatial heterogeneity as well as the independent variable's explanatory power for the dependent variable. It is measured by the q value. The stronger the explanatory power, the greater the q value. The following is the formula:

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

In formula: h=1,2..., L is the stratification of factor X. N_h and N are the number of cells in the h-th layer and region. σ_h^2 and σ^2 are the variances of Y values in the h-th layer and the whole region, respectively. SSW is the sum of intra-layer variances. SST is the total regional variance. q∈[0, 1], the dominant factor affecting the propagation of COVID-19 can be identified according to the size of the q value, and the significance of the q value can be checked by the geographic detector software.

The risk detector uses the t statistic to judge whether there is a significant difference in the number of confirmed cases of the epidemic among the strata of the influencing factors and to identify high-risk areas of the epidemic. The interaction detector can be used to detect whether the interaction of two independent variables will increase or decrease the effect of the independent variable on the influence of the dependent variable. The ecological detector compares whether the influence of the two independent variables on the spatial distribution of the COVID-19 epidemic is significantly different through the F statistic [9].

Random Forest is a decision tree-based statistical learning theory. It extracts numerous samples from the original sample using the bootstrap resampling method, and performs decision tree judgment and

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

classification on each bootstrap sample. Each decision tree will get a classification result, and the category with the most votes is the final prediction result. [10]. It can process high-dimensional data and has excellent accuracy and noise immunity [11]. In this paper, the training set and the test set are divided according to the ratio of 7:3, and the root means square error (RMSE) model accuracy (R^2) of the test set is used to measure the influence of the impact factor on the number of newly diagnosed patients.

The formula for calculating RMSE and R^2 are as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_{is}-y_{it})^2}{n}} \quad (3)$$

$$R^2 = 1 - \sum_{i=1}^{n}(y_{is} - y_{it})^2 \,/\, \sum_{i=1}^{n}(y_{is} - \bar{y}_{is})^2 \quad (4)$$

In the formula, y_{is} represents the simulated number of newly diagnosed patients in the i-th city. y_{it} represents the actual number of newly diagnosed patients in the i-th city. $\bar{y}_{is}$ is the average of the simulated newly diagnosed numbers in the corresponding data set. n denotes the number of samples in the corresponding data set. Throughout this study, SPSS was used to perform Pearson correlation analysis between newly confirmed cases and various influencing factors (13 in total) during the study period. Then the factors with a strong correlation with epidemic spread and no collinearity were selected as independent variables for random forest analysis.

3. Results and Analysis

3.1. Spatio-temporal characteristics of epidemic development

From January 24 to February 28, 2020, the epidemic continued to develop, affecting more than 290 regions nationwide. According to the national accumulative number of confirmed cases, the cumulative number of confirmed cases from Hubei province, the national number of cured cases, the number of cured cases from Hubei province, the national number of new cases, the number of new cases from Hubei province, were collected to make a combination diagram (Figure 1). Combined with the spread of the epidemic process and major events, the early development process of the epidemic is divided into four phases.

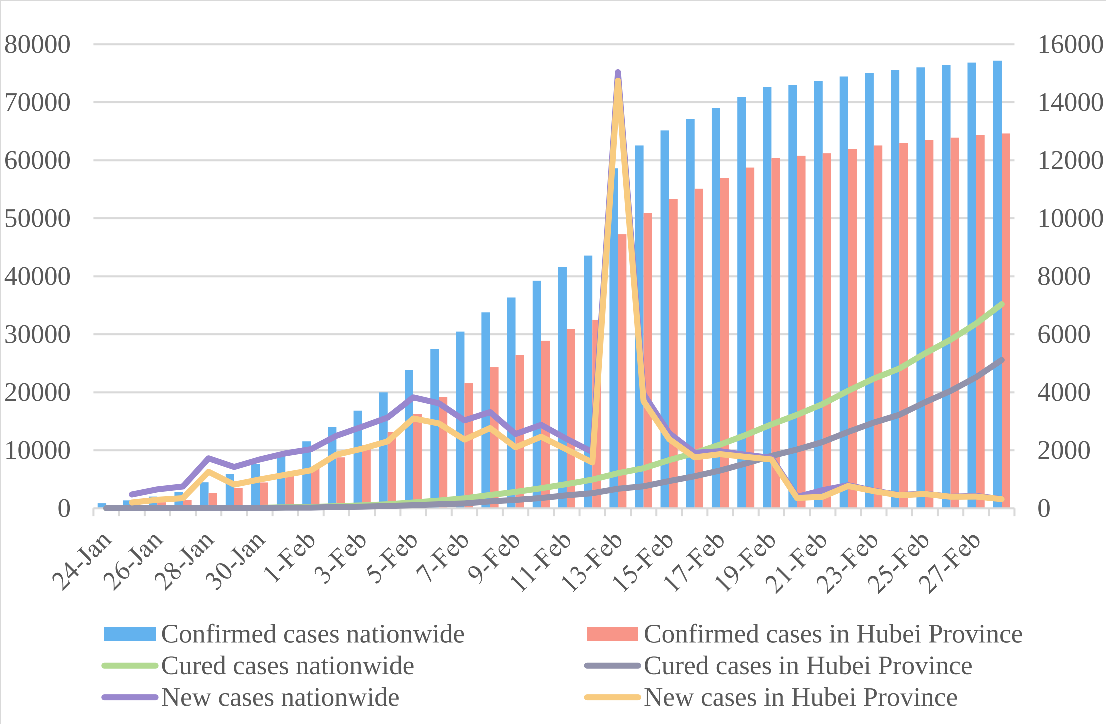

Figure 1: Daily variation of the epidemic situation in China and Hubei Province.

The first stage: the spread of the epidemic (January 24-February 1). Before the announcement of the lockdown in Wuhan, a widespread mass movement of people caused the epidemic to spread rapidly across the country, coinciding with New Year's Eve (January 24) and the Spring Festival (January 25),

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

which are the peak periods of population movement during the Spring Festival. The cumulative number of confirmed cases nationwide surged. On February 1, the total number of confirmed cases in the country had surpassed 10,000. During this time, the number of cities affected by the pandemic increased rapidly, reaching 290. The second stage: the peak period of the epidemic (February 2-February 11). The flow of patients throughout the incubation period aided the spread of the pandemic due to the time lag between case confirmation and actual infection time. The cumulative number of confirmed cases across the country continued to rise. At the same time, outside of Hubei province, the spread of the pandemic was at its peak in the Yangtze River Delta, Pearl River Delta, and Beijing-Tianjin-Hebei regions. The pandemic demonstrated a trend of expanding from province capital cities to prefecture-level cities during this time. The epidemic's spatial structure of multi-point dissemination was developed. The third stage is the period of epidemic recession (February 12-February 19). Governments around the country enacted the toughest home isolation and joint prevention and control measures possible. The pandemic prevention and control efforts had yielded impressive results, and the virus's progress was effectively halted. Over time, the number of cities with no new cases grew from 154 to 240. With the exception of Hubei, the number of newly diagnosed cases declined in a "ten-point continuous drop" on February 14. On February 18, the number of newly added cured cases surpassed the number of newly added confirmed cases across the country. The fourth stage: the extinction period of the epidemic (February 20-February 28). The overall number of new confirmed cases in the country, excluding Hubei, has gradually decreased to single digits at this point. The number of cities with no new cases had risen to 276 by February 28. In Hubei Province, the number of new confirmed cases has fallen by 85.8% from the previous stage, and 88.8% of newly diagnosed cases are centered in Wuhan City. A stage-by-stage by stage victory was accomplished in the national pandemic prevention and control effort.

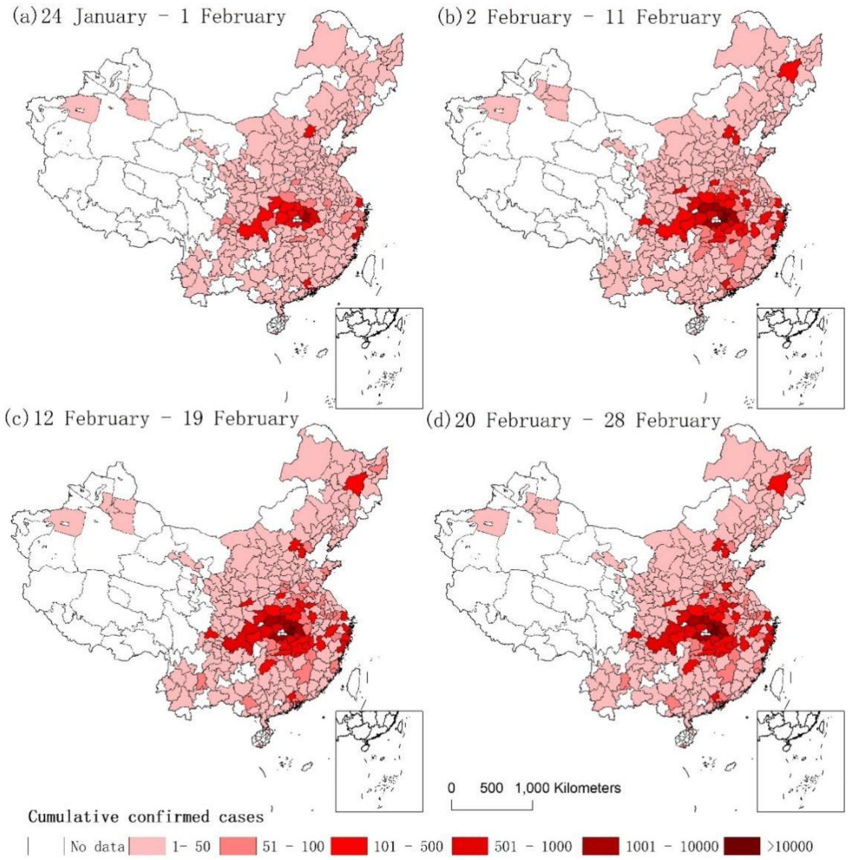

Figure 2: Temporal and spatial variation of confirmed cases nationwide.

The Moran's I of each stage was calculated and tested by Z and P values in order to more intuitively assess the clustering characteristics of the four stages of the epidemic (Table 2).

Moran's I value for Stage 1 (January 24 to February 1, 2020) was 0.256. For the second stage (February 2 to February 11), the Moran's I value was 0.158. The Moran's I value for the third stage (February 12 to February 19) equaled 0.084 and the Moran's I value was 0.079 in the fourth stage (February 20th to February 28th). P-value was less than 0.001 in all four stages, Z value was more than

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

1.96, which indicated that there was Spatio-temporal aggregation in the confirmed cases. The degree of Spatio-temporal aggregation gradually weakened with the progression of time. According to the LISA cluster diagram (Figure 3), it can be seen that in the four stages, except for Shiyan City in the northwest of Hubei Province, most cities in Hubei Province have high-high clustering types, indicating that the epidemic was primarily concentrated in Hubei Province. The most seriously impacted areas in Hubei Province were Wuhan and its neighboring areas. The spatial clustering types of Luan and Anqing cities in Anhui Province and the neighboring cities of Hubei Province were of low-high clustering, and the spatial clustering types of Xinyang and Jiujiang cities near to Hubei Province were gradually changed from high-high clustering to low-high clustering. As the speed and scope of the spread of the epidemic decreased, the number of low-to-high clusters in the surrounding cities of Wuhan gradually increased, and the degree of clustering of the epidemic weakened significantly, indicating that the epidemic in the surrounding areas of Wuhan was gradually controlled over time. In terms of spatial scope, the degree of clustering of urban cases displays characteristics of spreading from high-high clustering to low-high clustering and then to low-low clustering as the spatial distance from Wuhan increases, with Wuhan as the center and radiating outward.

Table 2: Moran's I and test values by stages in the country.

	Moran’s I	Z value	P value
Stage 1	0.256	9.498472	<0.001
Stage 2	0.158	12.7814	<0.001
Stage 3	0.084	15.12844	<0.001
Stage 4	0.079	13.82313	<0.001

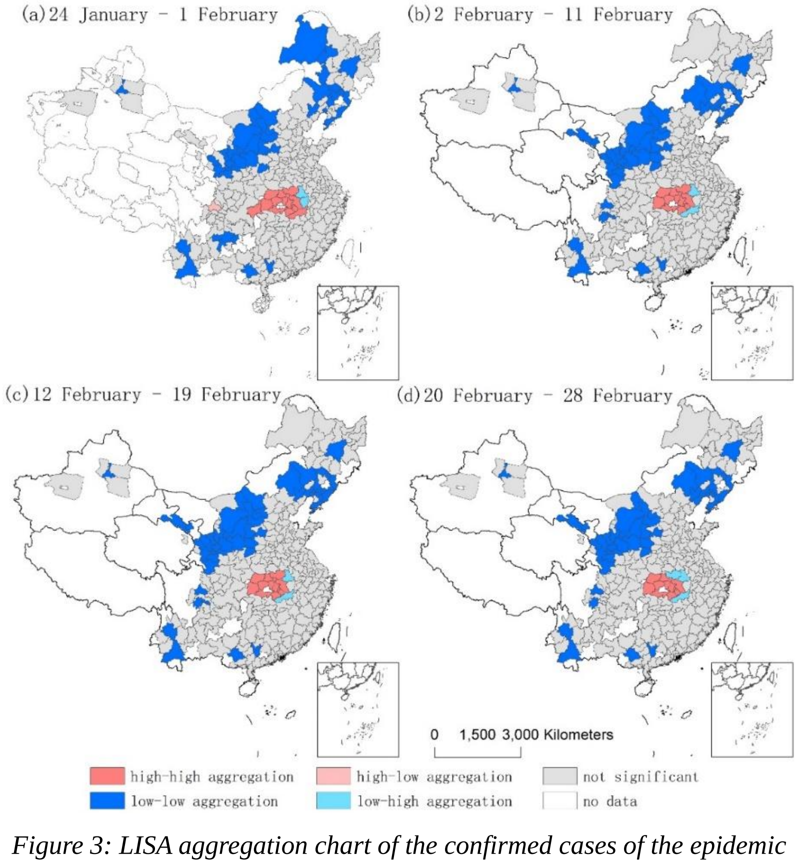

Figure 3: LISA aggregation chart of the confirmed cases of the epidemic

3.2. Geographical detector

The independent variables must be of type variable data when using geographic detectors to investigate the impact of various influencing factors on the epidemic. K-Means cluster analysis was used to discretize the numerical independent variable data and convert it to type variable data. The results of the factor detector (Table 3) show that different impact factors have different effects and significance. Among the social impact factors, the number of beds in medical institutions has the highest q value, followed by the number of urban population q value. It shows that among these variables, the level of

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

medical facilities and the number of the urban population are the most important impact factors which determine the spatial pattern of the number of confirmed cases. The most important environmental component is the built-up area. The q value for population density and road network density is the lowest, indicating that they have the least impact on the pandemic. The distance to Wuhan, the green coverage of the built-up region, the air quality index, the average temperature, and the altitude are among the natural influence elements with q values ranging from large to low. For a single risk factor, the risk detector table displays whether there is a statistical difference between the findings of the risk area detection and the number of confirmed epidemic cases in each type of zone. "Y" indicates a significant difference, while "N" indicates no significant difference, according to a t-test with a significance level of 0.05.

Table 3: Statistical values of q for different impact factors.

Natural factors	q value	Social factors	q value	Social factors	q value
Green coverage rate of built-up area	0.04	Built-up area	0.14	GDP per capita	0.05
Air quality index	0.03	Urban population	0.19	Number of beds in medical and health institutions	0.20
Urban average altitude	0.01	Road network density in built-up areas	0.04	Population with high school education	0.05
Average daily temperature	0.03				
Distance to Wuhan	0.11	Population density	0.03	Night light data	0.06

According to the results of the interaction detector (Table 3), there are two types of COVID-19 epidemics in China: linear enhancement and two-factor enhancement. The impacts of the interactions between the two factors outweigh the effects of the single factors. The other is the interaction between natural factors and social-economic factors, which are greater than the interaction between natural factors. The results of 91 interactions were detected by 13 impact factors of the 2020 epidemic. The maximum value of (X1∩X2) was 0.99 and the minimum value was 0.01. In the synergistic effect of the two factors, the interaction of the number of beds in medical and health institutions, the urban population, the distance from Wuhan, and other social and economic factors has a strong impact on the number of confirmed cases of the epidemic. Ecological detection is used to analyze the impact of two factors on the increase of confirmed cases. The results showed that the number of beds in medical institutions and the population in urban areas was obviously different from other variables in terms of the effect on the spatial distribution of confirmed cases. The factors that affect the spatial heterogeneity of the confirmed number of cases are: the number of beds of medical and health institutions > urban population > built-up area > distance to Wuhan > nighttime light data >educated population per 100,000 people in 2020 > GDP per capita > road network density of built-up area > average air temperature > population density > air quality index > green coverage rate of built-up area > altitude.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
X1	0.14												
X2	0.33	0.04											
X3	0.50	0.25	0.02										
X4	0.24*	0.49	0.99	0.19									
X5	0.25	0.28	0.15	0.25	0.03								
X6	0.49	0.99	0.33	0.99	0.99	0.05							
X7	0.99	0.99	0.50	0.49	0.49	0.50	0.20						
X8	0.50	0.99	0.25	0.99	0.25	0.34	0.99	0.05					
X9	0.33	0.25	0.33	0.49	0.49	0.17	0.50	0.33	0.06				
X10	0.33	0.25	0.17	0.99	0.50	0.20	0.99	0.33	0.50	0.03			
X11	0.25	0.14	0.09	0.49	0.20	0.07	0.49	0.13	0.10	0.09	0.01		
X12	0.33	0.19	0.13	0.49	0.22	0.20	0.99	0.26	0.20	0.08	0.06	0.03	
X13	1.00	0.29	0.36	0.99	0.56	0.99	0.99	0.53	0.99	0.22	0.31	0.13*	0.11

X1: Number of beds
X2: Urban population
X3: Built-up area
X4: Distance to Wuhan
X5: Night light
X6: education level
X7: GDP per capita
X8: Road network density
X9: Average daily temperature
X10: Population density
X11: Air quality index
X12: Green coverage rate
X13: Urban average altitude
* indicates two-factor enhancement
the rest is linear enhancement

Figure 4: Interaction detection of different influencing factors.

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

3.3. Random forest

According to the correlation analysis results, the two factors of built-up area and the urban population had no substantial impact on the number of confirmed epidemics, and there was collinearity between them. Factors including height, green coverage, and air quality index were found to be irrelevant by the factor detector, so they were eliminated from the random forest model.

Table 4: Significance and collinearity of influencing factors.

	β	t	Significance	tolerance	VIF
Road network density in built-up areas	17.856	2.038	.043	.914	1.094
Green coverage rate of built-up area	-28.394	-5.560	.000	.763	1.311
Urban population	.049	.310	.757	.145	6.877
Built-up area	-.063	-.283	.778	.123	8.105
Population density	.010	1.204	.230	.855	1.170
GDP per capita	.001	.880	.380	.401	2.497
Population with high school education	.001	.776	.439	.195	5.117
Number of beds in medical and health institutions	.014	1.907	.058	.601	1.665
Night light data	-2.032	-.861	.390	.323	3.096
Air quality index	-1.285	-1.425	.155	.602	1.660
Urban average altitude	-.066	-1.656	.099	.712	1.404
Average daily temperature	-3.793	-1.035	.302	.399	2.509
Distance to Wuhan	-.168	-4.429	.000	.517	1.932

After several tests, the decision tree n-tree was set at 150, mtry as 3. Randomly, 30% of the points were selected as the test set, and the rest 70% of the points were selected as the training set, by comparing the test set accuracy (R^2 and RMSE) to determine the best regression model (Figure 5). The fitting accuracy (R^2) of random forest regression in the test set reaches 0.7192 and has a minimal RMSE, as shown in Figure 5. Except that the predicted values and the actual values of the individual data have a large deviation, from the fitting line between the predicted values and the actual values, the fitting effect is good, which shows that the model constructed by using random forests has a good prediction accuracy.

Under the random forest technique, Figure 6 depicts the feature relevance of the 9 influencing elements. The distance to Wuhan is the most relevant variable among the nine contributing factors, with a relative significance of 0.92. It demonstrates that the distance from the outbreak's epicenter is the most important element influencing the epidemic's spread. The epidemic had a huge impact on the cities surrounding Wuhan, and the closer the cities were to Wuhan, the worse the epidemic became. As a result, the epidemic's spatial pattern expanded out from Wuhan and gradually faded. A minor characteristic variable with a relative value of 0.56 is the number of beds in a medical and health institution. The magnitude, grade, and ability of hospitals in a region to deliver health services were reflected in the number of beds in medical and health facilities. Cities can better manage health resources and detect and diagnose the epidemic faster as the number of beds increases, hence it is positively connected with the cumulative confirmed cases. Other factors such as GDP per capita, road network density in built-up areas, night light index, air quality, and population education level all had some impact on the outbreak, although only to a minor degree. These variables can represent a region's socio-economic progress and scope of urban construction. In general, the higher the per capita GDP and the density of the road network in the built-up region, the higher the city's development level, attractiveness, and people flow. This also suggests that the closer the population's activities are clustered, the greater the chance of infection from mutual contact.

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

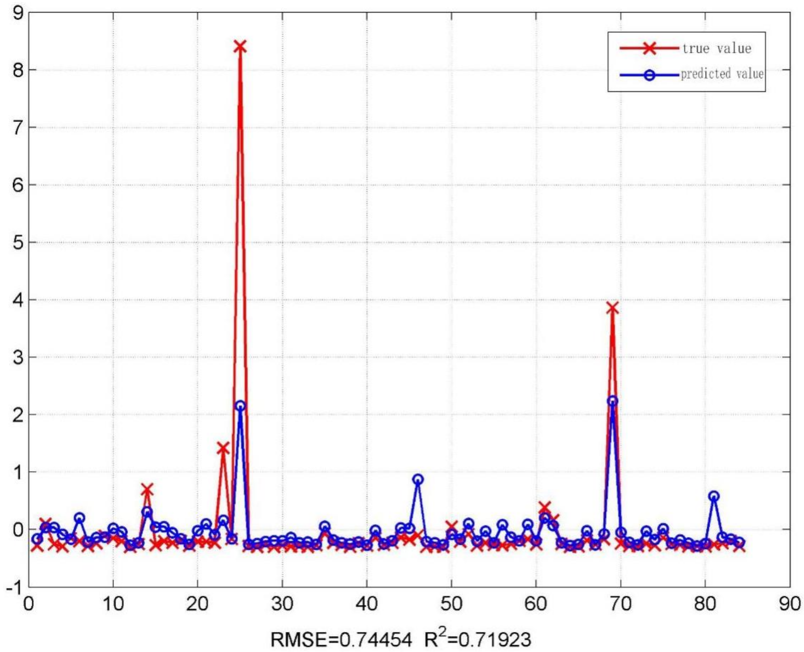

Figure 5: Prediction results of random forest test set.

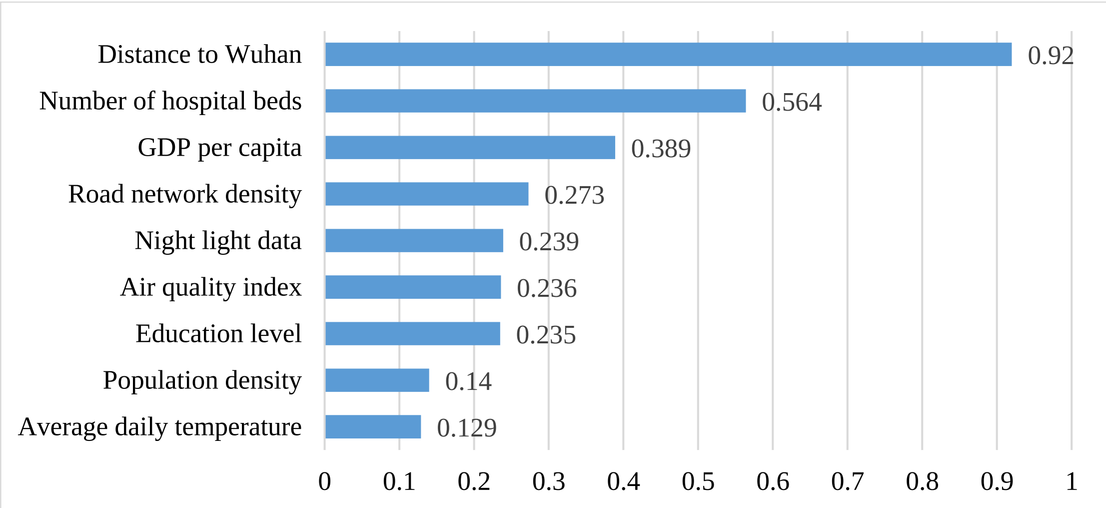

Figure 6: Relative importance ranking of influencing factors.

4. Conclusion

In this study, based on the confirmed and newly added data of 285 prefecture-level cities in China, the spatial and temporal distribution characteristics of the epidemic from January 24 to February 28, 2020, were analyzed from the two dimensions of time and space by combining spatial autocorrelation and Moran index. The index system was established by considering the meteorological environment, social economy, geographical information, and other influencing factors. The dominant factors affecting the spread of the pandemic and their mechanism were discussed by using the geographical detector and random forest method. After conducting the study, the following conclusions are drawn:

1) The COVID-19 epidemic in Chinese cities has obvious stage characteristics. The development of the epidemic has experienced four stages: the diffusion period, the peak period, the recession period and the extinction period. In the first and second stages, the epidemic broke out in Wuhan and quickly spread to 290 cities across the country in a short period of time; In the second and third stages, the number of newly diagnosed patients continued to decline, and the epidemic prevention and control policies achieved remarkable results.

2) Cities have a spatial aggregation of the epidemic. There is a spatial clustering phenomenon of

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

confirmed cases, according to the spatial autocorrelation analysis of each stage, but the degree of clustering is attenuated to some amount. The regions with the worst epidemic performance are mostly in Hubei Province and its adjacent cities. High-high clustering is the type of spatial clustering observed in this region. The degree of clustering of the epidemic lessens as the distance from Wuhan increases.

3) There are multi-factor interactions and different impact factors have varied effects on the pandemic. The number of beds in medical and health institutions is the most important factor impacting the spatial pattern of the number of confirmed epidemic cases, according to the factor detector. The interaction between two factors is greater than that of a single factor, and the interaction between natural factors and socioeconomic factors is greater than the interaction between natural factors.

4) The number of beds in medical and health institutions, as well as the distance from the epidemic core, are crucial factors impacting the epidemic's spread. According to the characteristic importance of random forests to different impact factors, the relative importance of the distance to Wuhan is 0.92, followed by the number of beds in medical institutions, which is 0.56.

In this study, spatial autocorrelation was used to analyze the spatial and temporal distribution of COVID-19. Geographic detectors and random forest models were used to explore the driving factors that influence the distribution of confirmed cases. The results showed that the interaction of multiple factors and the distance from the center of the epidemic, urban health care resources have a significant impact on the spread of the epidemic. These analyzes can further deepen the understanding of the COVID-19 epidemic pattern and influencing factors in different geographical conditions, and provide a theoretical basis for the epidemic prevention and control work in China. They are helpful for city decision-makers to take scientific epidemic prevention and control measures in time, such as timely blocking the spread of the epidemic in space and planning the allocation of urban medical and health resources reasonably, to effectively contain the spread of infectious diseases.

Acknowledgements

Fund Project: College Students' Innovation and Entrepreneurship Training Program *Detection of factors affecting the transmission and development of COVID-19* ; Project Number: 202110635090.

References

[1] Chien Lung Chang,Chen Lung Wen(2020). Meteorological impacts on the incidence of COVID-19 in the U.S.[J]. Stochastic environmental research and risk assessment, 34(10), 1675-1680.

[2] Mozumder Salatul Islam, Amin Mohammad Shaiful Alam, Uddin Mohammad Rakib, Talukder Musabbir Jahan(2020). Coronavirus COVID-19 outbreak and control: effect of temperature, relative humidity, and lockdown implementation[J]. Archives de Pediatrie, 28(2), 111-116.

[3] WANG Jiaoe, DU Delin, WEI Ye, YANG Haoran(2020). The development of COVID-19 in China: Spatial diffusion and geographical pattern[J]. Geographical Research, 39(07), 1450-1462.

[4] Gong Shengsheng, Mo Hui(2021). Geographical Analysis of the COVID-19 Epidemic in Hunan Province, China in 2020[J]. Tropical Geography, 41 (4), 708-722.

[5] CHEN Xiao, HUANG Yujin, LI Jiahui, WANG Shiyang, PEI Tao(2020). Clustering characteristics of COVID-19 cases and influencing factors in Chongqing Municipality[J]. Progress in Geography, 39(11), 1798-1808.

[6] LI Xin, ZHOU Lin, JIA Tao, WU Hao, ZHOU Yuliang, QIN Kun(2020). Influence of Urban Factors on the COVID-19 Epidemic: A Case Study of Wuhan City [J]. Geomatics and Information Science of Wuhan University, 45(6), 826-835.

[7] Getis A, Ord J K(2010). The analysis of spatial association by use of distance statistics. [J]. Geographical Analysis, 24(3), 189-206.

[8] WANG Jinfeng, XU Chengdong(2020). Geodetector: Principle and prospective [J]. Journal of Geographical Sciences, 72(01), 116-134.

[9] WANG Huan, GAO Jiangbo, HOU Wenjuan(2019). Quantitative attribution analysis of soil erosion in different geomorphological types in karst areas: Based on the geodetector method [J]. Journal of Geographical Sciences, 29(2), 271-286.

[10] Breiman L(2001). Random forests. Machine Learning, 45(1), 5-32.

[11] FANG Kuang-nan, WU Jian-bin, ZHU Jian-ping, SHIA Bang-chang (2011). A Review of Technologies on Random Forests [J]. Statistics &Information Forum, 26(03), 32-38.

[12] Pei T, Wang X, Song C, et al (2021). Review on spatiotemporal analysis and modeling of COVID-

ISSN 2616-7433 Vol. 4, Issue 2: 52-62, DOI: 10.25236/FSST.2022.040210

19 pandemic [J]. Journal of Geo-information Science, 23(2), 188-210.
[13] Liao Wenyue, Sun Meiwei, Yu Chuying, Deng Yingbin, Li Miao, Yang Ji, Li Yong, Xu Jianhui, Chen Yuchan and Yan Yingwei(2021). Impact Factors of COVID-19 Epidemic Spread in Hubei Province Based on Multi-Source Data[J]. Tropical Geography, 41(1), 12-24.
[14] ZHANG Xin, LIN Hui, ZHU Changming, WANG Jinfeng, LI Manchun, XU Chengdong, HU Maogui, MENG Bin, CAO Chunxiang(2020). Spatiotemporal Characteristics and Dynamic Process Analysis of Epidemic Diagnosis Time of COVID-19 at Early Stage in Chinese Mainland [J]. Geomatics and Information Science of Wuhan University, 45(06), 791-797.
[15] Wang L,Zhao J S(2020). Spatiotemporal distribution pattern of the COVID-19 epidemic and geographical detection. Acta Ecologica Sinica, 40(19), 6788-6800.