(RESEARCH ARTICLE)

Check for updates

Spatial clustering analysis of COVID-19 using LISA: A case study of the 2022 winter Day in Hanoi, Vietnam

Thi-Quynh Nguyen [1, *], Thi-Thuy Ngo [2] and Thi-Binh Than [3]

[1] Faculty of Nursing, East Asia University of Technology, Hanoi, Vietnam.
[2] Faculty of Nursing, Phenikaa University, Hanoi, Vietnam.
[3] Department of Training, Lang Son Medical College, Lang Son, Vietnam.

Abstract

Background: A novel coronavirus disease outbreak in 2019 (COVID-19) caused by he emergence of severe acute respiratory syndrome coronavirus 2 (SARS CoV 2) in China quickly spreads throughout the world. This study aimed to analyze the spatial clustering of COVID-19.

Methods: The global and local Moran's I statistic (LISA) was used to investigate the spatial clusters of COVID-19 including spatial clusters (high-high and low-low) and spatial outliers (low-high and high-low).

Results: A case study of COVID-19 locally transmitted cases reported in a 2022 winter day in Hanoi city has indicated that high-high spatial clusters were totally concentrated in 6 urban districts in the Hanoi metropolitan including such as districts of Dong Da, Gia Lam, Thanh Tri, Hai Ba Trung, Cau Giay, and Long Bien. Whereas, low-low spatial clusters were mainly in sub-urban districts such as Ba Vi, Thach That, Phuc Tho, and Son Tay town (0 cases) in the northwest and Ung Hoa district in the south of Hanoi.

Conclusions: The study results indicated the effectiveness of LISA in analysis of spatial clustering of COVID-19. Findings in this study make great contributions to the fight of the COVID-19 pandemic.

Keywords: Spatial Clustering; Moran's I statistic; LISA; COVID-19; Hanoi; Vietnam

1. Introduction

The COVID-19 pandemic has been a serious global danger to public health since early 2020. SARS-CoV2, the coronavirus that causes severe acute respiratory syndrome, is the cause of COVID-19 (1). The virus spread fast to additional nations in eastern Asia, Europe, and the rest of the world (2). As of 16 August 2023, the World Health Organisation had received reports of more than 769.9 million confirmed cases of COVID-19, including more than 6.9 million deaths (3). It is, therefore, attempts have been made to the study of the COVID-19 pandemic, particularly spatial clustering of COVID-19 (4,5).

Geographical factors play crucial roles in the fight against the COVID-19 pandemic (6). When the COVID-19 pandemic were discovered to be thoroughly spatial in nature (7). Lots of efforts have been made to study the COVID-19 pandemic from a geographical perspective such as the assessment of the effects of living environment deprivation on COVID-19 hotspot in Kolkata megacity, India using Getis-Ord G statistic and geographically weighted principal component analysis (8). Exploratory spatial data analysis and the geodetector method have been also employed to analyse the spatial and temporal differentiation characteristics and the influencing factors of the COVID-19 epidemic spread in mainland China

* Corresponding author: Thi-Quynh Nguyen: ORCID: https://orcid.org/0009-0001-9868-4695; E-mail: quynhnt@eaut.edu.vn

(9). Also in China, spatial autocorrelation analysis was used to investigate the spatial clustering characteristics of the COVID-19 pandemic in Beijing (10). In addition, spatial autocorrelation analysis has been also successfully to COVID-19 studies in many countries such as to investigate the spatiotemporal interaction effect of COVID-19 transmission in the United States (11), to identify long-term exposure to air-pollution and COVID-19 mortality in England (12), to spatial autocorrelation patterns across five waves of COVID-19 in Catalonia, Spain (13), and to analyse spatio-temporal COVID-19 outbreak in Italy (14).

This study aimed to analyze the spatial clustering of COVID-19. The global and local *Moran's I* statistic (LISA) was used to investigate the spatial clusters of COVID-19 in Hanoi city. The spatial clusters including spatial clusters (high-high and low-low) and spatial outliers (low-high and high-low) will be identified using the first order and second of contiguity.

2. Material and methods

2.1. Material

In this study, a dataset of COVID-19 cases in the 2022 Winter Day in Hanoi was used to analysis of spatial clusters. These COVID-19 locally transmitted cases were reported on 31 January 2022. The spatial distribution of these COVID-19 cases is shown in Figure 1. Data from Figure 1 illustrate that the COVID-19 cases were mainly reported in the northeast of Hanoi city. In particular, high numbers of COVID-19 case were mainly confirmed in the Hanoi metropolitan where the population density is dense. Data from Table 1 shows that high numbers of COVID-19 cases were reported mostly in urban districts such as Dong Da (298 cases), Hoang Mai (266 cases), Ha Dong (260 cases), and Nam Tu Liem (196 cases). On the other hand, low and very low number of COVID-19 cases were reported in suburban districts such as Ba Vi, My Duc, Phuc Tho, Son Tay and Thach That (0 cases), Chuong My (1 case), Quoc Oai (2 cases), and Ung Hoa (3 cases).
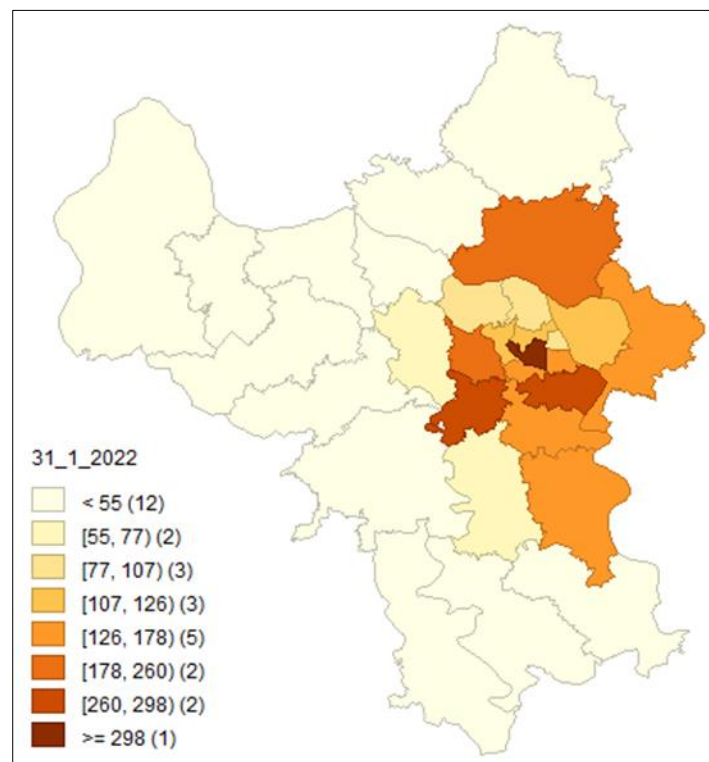

Figure 1 Map of new COVID-19 confirmed cases in Hanoi city on 31 January 2022

2.2. Methods

This study employed global *Moran's I* statistic to identify the spatial clustering of the COVID-19 pandemic at global scale (15,16). The definition of the global *Moran's I* statistic is expressed in equation (1):

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}\sum_{j=i}^{n} W_{ij} \sum_{i=1}^{n}(x_i - \bar{x})^2} \ldots\ldots\ldots(1)$$

where x_i and x_j are the number of new COVID-19 confirmed cases for district i and district j; $\bar{x}$ is the mean of COVID-19 cases and be given by $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$; n is the total number of districts in the whole study area; and W_{ij} is a $(n \times n)$ spatial weight matrix (5).

The range of values of global *Moran's I* coefficient is in the interval [-1, +1] (5). Positive values of *Moran's I* result from the data's positive spatial autocorrelation., whereas *Moran's I* values are negative when there is a negative spatial autocorrelation (17). The absence of spatial autocorrelation or random COVID-19 epidemic distribution is shown by values of the global *Moran's I* coefficients that are near to zero.

The global *Moran's I* reflects the presence or lack of spatial autocorrelation as a whole. The regional *Moran's I* statistic was used to quantify the spatial clustering of low and high COVID-19 pandemic levels in each district (5). The local *Moran's I* statistic (I_i) for COVID-19 pandemic at district i is given by the following equation (18):

$$I_i = \frac{(x_i - \bar{x})}{\sigma^2} \sum_{j\#i, j \in J_i}^{N} W_{ij}(x_j - \bar{x}) \ldots\ldots\ldots(2)$$

where x_i, x_j, $\bar{x}$, and W_{ij} are defined in equation (1); N is the total number of neighborhood districts (5); J_i denotes the neighborhood set of COVID-19 confirmed cases at district i; $j\#i$ implies that the sum of all ($x_j - \bar{x}$) of nearby neighbourhood districts of district i but not including x_i; and σ^2 is the variance of x, given in equation (3). W_{ij} defines neighbor connectivity and can be constructed using first order and second of contiguity (Figure 2).

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N}(x_i - \bar{x}) \ldots\ldots\ldots(3)$$

The level of spatial clustering of the COVID-19 pandemic at each district is indicated by local *Moran's I* statistic. Similar to the global *Moran's I* statistic, the local *Moran's I* value at district i (I_i) also ranges between -1 and +1. There is no spatial autocorrelation of COVID-19 casses if the local *Moran's I* coefficient at district i equals zero ($I_i = 0$). If $I_i > 0$ then there will be a positive spatial autocorrelation of COVID-19 cases (5). If $I_i < 0$ then there will be a negative spatial autocorrelation of COVID-19 cases. A high positive I_i shows the district i has a similarly high or low number of COVID-19 cases as its neighbors and called the "spatial cluster"(17). In this case, when there is a positive local spatial autocorrelation, the local *Moran's I* statistic indicates two types of spatial clusters for COVID-19 cases, including: high-high spatial clusters and low-low spatial clusters. Low-high and high-low clusters are also two forms of spatial outliers that are identified using the local Moran's *I* statistic when there is a negative local spatial autocorrelation.
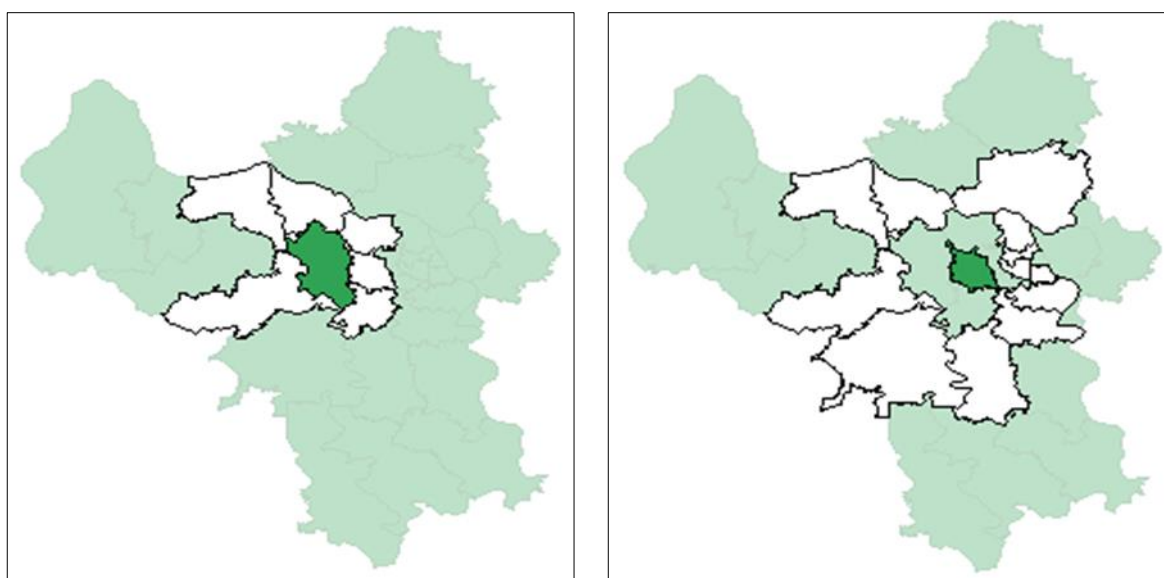

Figure 2 Map of neighbor connectivity using first order (left) and second (right) of contiguity

In this work, with the help of the spatial statistics software, GeoDA, developed by (19), a randomization test was used to test the significance of spatial autocorrelation statistics. Spatial autocorrelation statistics were generated and tested at the significance of 0.05 using 999 permutations.

3. Results and discussion

3.1. Spatial distribution of the COVID-19 cases

A total of 2638 new COVID-19 cases were detected in the community in Hanoi city. The data was summarized in Table 1. The spatial distribution of new COVID-19 cases was shown in Figure 2. Data from Table 1 and Figure 2 shows that the high number of cases was mainly concentrated in urban districts such as in the districts of Dong Da (298 cases), Hoang Mai (266 cases), Ha Dong (260 cases), and Nam Tu Liem (196 cases). Areas with a low number of new cases were concentrated in suburban districts such as Thach That and Phuc Tho, Ba Vi and My Duc (0 cases), Chuong My (1 case), and Quoc Oai (2 cases).

Table 1 Summary table of LISA map of new COVID-19 confirmed cases

No	New cases	Districts	Community types	First order of contiguity			Second order of contiguity		
				LISA	Clusters	P value	LISA	Clusters	P value
1	107	Ba Dinh	Urban	0.13	0	0.08	0.18	1	0.00
2	0	Ba Vi	Rural	0.98	2	0.02	0.97	2	0.04
3	77	Bac Tu Liem	Urban	-0.02	0	0.27	-0.03	0	0.16
4	116	Cau Giay	Urban	0.22	1	0.04	0.18	1	0.02
5	1	Chuong My	Rural	0.26	0	0.23	0.06	0	0.37
6	4	Dan Phuong	Rural	0.25	0	0.22	-0.08	0	0.40
7	178	Dong Anh	Rural	-0.24	0	0.27	0.40	0	0.08
8	298	Dong Da	Urban	0.76	0	0.16	1.97	1	0.00
9	166	Gia Lam	Rural	0.85	1	0.02	0.04	0	0.40
10	260	Ha Dong	Urban	-0.08	0	0.46	0.07	0	0.31
11	127	Hai Ba Trung	Urban	0.46	1	0.01	0.34	1	0.00
12	56	Hoai Duc	Rural	-0.01	0	0.48	0.06	0	0.23
13	95	Hoan Kiem	Urban	0.06	0	0.06	0.06	1	0.02
14	266	Hoang Mai	Urban	1.09	0	0.05	1.43	1	0.00
15	110	Long Bien	Urban	0.16	1	0.04	0.03	0	0.32
16	10	Me Linh	Rural	0.38	0	0.15	0.26	0	0.13
17	0	My Duc	Rural	0.95	0	0.06	0.09	0	0.43
18	196	Nam Tu Liem	Urban	0.56	0	0.10	0.24	0	0.16
19	4	Phu Xuyen	Rural	0.28	0	0.28	-0.14	0	0.39
20	0	Phuc Tho	Rural	0.84	2	0.00	-0.18	0	0.34
21	2	Quoc Oai	Rural	0.27	0	0.23	0.34	2	0.03
22	17	Soc Son	Rural	-0.05	0	0.42	0.13	0	0.31
23	0	Son Tay	Town	0.98	2	0.00	0.78	2	0.02
24	84	Tay Ho	Urban	-0.01	0	0.21	-0.02	0	0.05
25	0	Thach That	Rural	0.97	2	0.00	0.24	0	0.26
26	55	Thanh Oai	Rural	-0.01	0	0.49	-0.12	0	0.17
27	143	Thanh Tri	Rural	0.56	1	0.01	0.08	0	0.26

28	137	Thanh Xuan	Urban	0.70	1	0.00	-0.05	0	0.38
29	126	Thuong Tin	Rural	-0.10	0	0.36	0.24	0	0.06
30	3	Ung Hoa	Rural	0.78	2	0.02	-0.48	0	0.18

3.2. Analysis of spatial clustering of new COVID-19 cases

The Moran scatter plots in Figure 3 shows the degree of globally spatial autocorrelation of COVID-19 cases in the case of using two different types of spatial weight matrices. The global Moran's I values were 0.244 and 0.413 corresponding with the use of the first order and second order of contiguity method, respectively. This result shows that the first order method produces results with higer correlation coeffient when comparing with that of the first order method. This shows a higher degree of spatial autocorrelation between COVID-19 cases for method of the first order of contiguity. In the analysis of spatial clustering, both methods were used to idenfity LISA of COVID-19 cases.
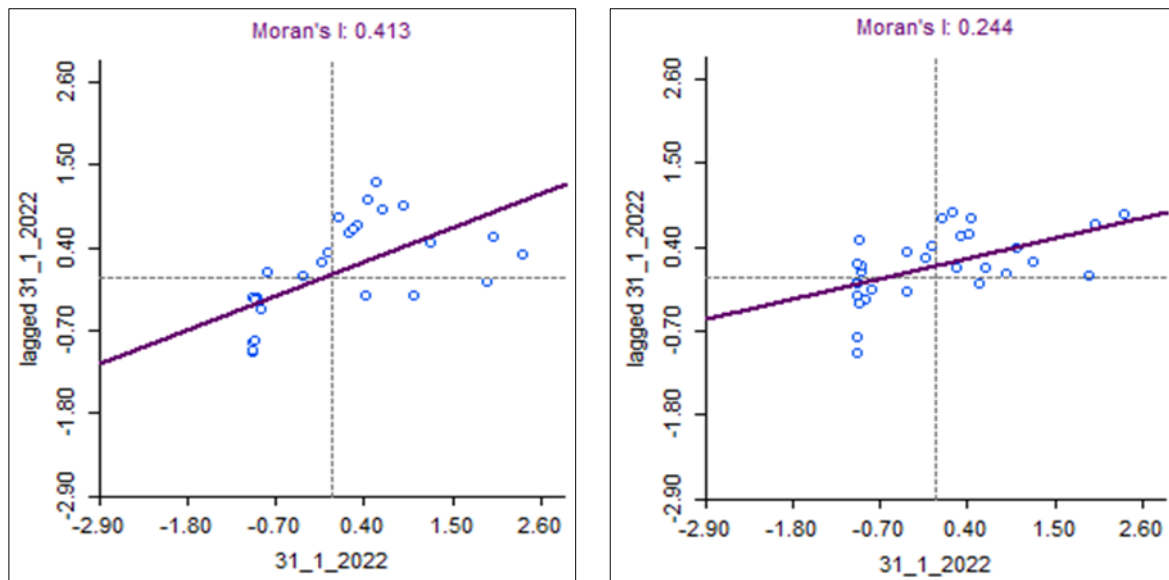

Figure 3 Moran scatter plots of new COVID-19 confirmed cases using first order (left) and second (right) of contiguity

Using the first order of contiguity, spatial clusters of COVID-19 identified by LISA is shown in Figure 4. Data from Figure 4 (left) shows that there were six high-high clusters, three low-low spatial clusters, no low-high and high-low spatial outliers, and 19 districts and town with unsignificant spatial clustering. In which, six high-high clusters were mainly distributed in the central area of Hanoi city where the population density is dense. Three low-low spatial clusters were distributed in the northwest and south of the city in rural areas with lower population densities. Data Figure 4 (right) and table 1 illustrate that the high-high and low-low spatial clusters were statistically significant with at the level of 0.05, of which five spaital clusters at the significant level of 0.05, four spatial clusters clusters at the significant level of 0.01 and only two spatial clusters at the significant level of 0.001. Both the left and right images in Figure 4 show that high-high spatial clusters were statistically significant at level of 0.01. These high-high spatial clusters were totally concentrated in 6 urban districts where there were high COVID-19 cases reported on January 31, 2022, including Dong Da (298 cases), Gia Lam (166 cases), Thanh Tri (143 cases), Hai Ba Trung (127 cases), Cau Giay (116 cases) and Long Bien (110 cases). Five low-low clusters include Ba Vi, Son tay town, Thach That, and Phuc Tho (0 cases) in the northwest and Ung Hoa district (3 cases) in the south of Hanoi.
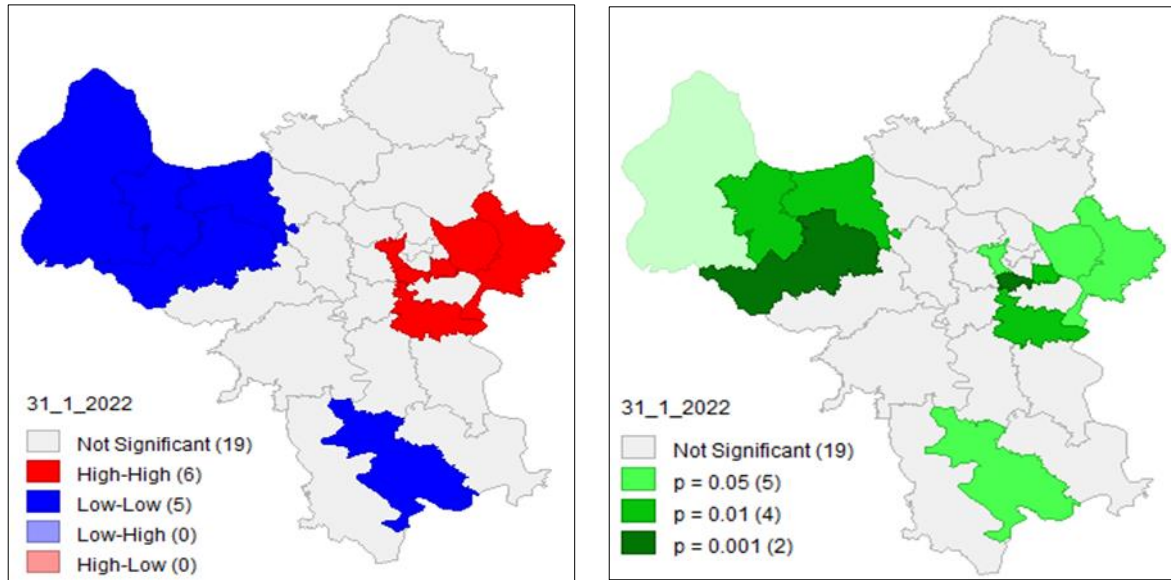
Figure 4 Spatial clustering of new COVID-19 cases using first order of contiguity: LISA map (left) and significant map (right)

The spatial cluster determined by the second order method is shown in Figure 5 and also summarized in Table 1. Figure 5 data (left) shows that there are six high-high clusters, three low-low clusters, and no low-high and and high-low cluster, and 19 districts with statistical unsignificance. In which, six high-high clusters were also mainly distributed in the central area of Hanoi city with high population density is dense. Five low-low clusters were distributed in the northwest areas of the city in the suburban area of Hanoi. Data from Figure 5 (right) and Table 1 show that the nine high-high and low-low clusters were statistically significant at the level of 0.05, in which three clusters at the sinificance of 0.01, one cluster at the significance of 0.001. Almost high-high spatial clusters were statistically significant at the level of 0.01. The spatial distribution of the LISA map shows that high-high spatial clusters were still concentrated in inner-city districts where most of COVID-19 cases were confirmed January 31, 2022, including Dong Da (298 cases), Hoang Mai (266 cases), Hai Ba Trung (127 cases), Cau Giay (116 cases), Ba Dinh (107 cases) and Hoan Kiem (95 cases). Three low-low spatial clusters were in districs of Ba Vi and Son tay town (0 cases), and Quoc Oai (2 cases) in the northwest of Hanoi.
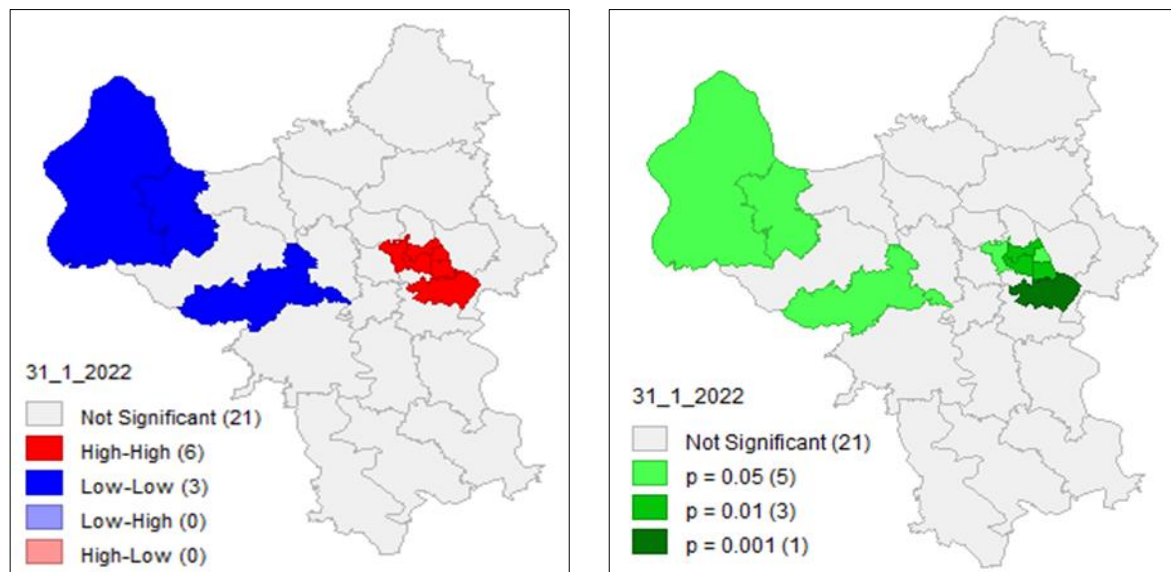

Figure 5 Spatial clustering of new COVID-19 cases using second order of contiguity LISA map (left) and significant map (right)

4. Conclusion

In this study aimed to analyze the spatial clustering of COVID-19. The global and local *Moran's I* statistic (LISA) was used to investigate the spatial clusters of COVID-19 including spatial clusters (high-high and low-low) and spatial outliers (low-high and high-low). It was found from a case study of COVID-19 locally transmitted cases in a 2022 winter day in Hanoi city has indicated that high-high spatial clusters were totally concentrated in 6 urban districts in the Hanoi metropolitan including such as districts of Dong Da, Gia Lam, Thanh Tri, Hai Ba Trung, Cau Giay, and Long Bien. Whereas, low-low spatial clusters were mainly in sub-urban districts such as Ba Vi, Thach That, Phuc Tho, and Son Tay town (0 cases) in the northwest and Ung Hoa district in the south of Hanoi. This study has demonstrated the effective use of LISA in the investigation of COVID-19 spatial clustering. The results of this investigation have a significant impact on the fight against the COVID-19 pandemic.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Shi Z, Chen H, Fan K, Chen P. Some thoughts and strategies of planning for the impact of "COVID-19" epidemic in Yunnan plateau basin. In: E3S Web of Conferences. EDP Sciences; 2020. p. 3044.

[2] Spagnuolo G, De Vito D, Rengo S, Tatullo M. COVID-19 outbreak: an overview on dentistry. Int J Environ Res Public Health. 2020; 17(6):2094.

[3] WHO. WHO Coronavirus (COVID-19) Dashboard [Internet]. 2023. Available from: https://covid19.who.int/

[4] Hoang A, Nguyen T. Identifying Spatio-Temporal Clustering of the COVID-19 Patterns Using Spatial Statistics: Case Studies of Four Waves in Vietnam. Int J Appl Geospatial Res. 2022; 13(1):1–15.

[5] Vu D-T, Nguyen T-T, Hoang A-H. Spatial clustering analysis of the COVID-19 pandemic: A case study of the fourth wave in Vietnam. Geogr Environ Sustain. 2021; 14(4).

[6] Shepherd M. Why Geography is a key-part of fighting the COVID-19 Coronavirus outbreak. Forbes. 2020; 5:2020.

[7] Rose-Redwood R, Kitchin R, Apostolopoulou E, Rickards L, Blackman T, Crampton J, et al. Geographies of the COVID-19 pandemic Dialogues in Human Geography 10 (2): 97–106. 2020.

[8] Das A, Ghosh S, Das K, Basu T, Dutta I, Das M. Living environment matters: Unravelling the spatial clustering of COVID-19 hotspots in Kolkata megacity, India. Sustain Cities Soc. 2021; 65:102577.

[9] Alves JD, Abade AS, Peres WP, Borges JE, Santos SM, Scholze AR. Impact of COVID-19 on the indigenous population of Brazil: A geo-epidemiological study. Epidemiol Infect. 2021; 149:e185.

[10] Han Y, Yang L, Jia K, Li J, Feng S, Chen W, et al. Spatial distribution characteristics of the COVID-19 pandemic in Beijing and its relationship with environmental factors. Sci Total Environ. 2021; 761:144257.

[11] Liu L, Hu T, Bao S, Wu H, Peng Z, Wang R. The spatiotemporal interaction effect of COVID-19 transmission in the United States. ISPRS Int J Geo-Information. 2021; 10(6):387.

[12] Huangfu P, Atkinson R. Long-term exposure to NO2 and O3 and all-cause and respiratory mortality: A systematic review and meta-analysis. Environ Int. 2020; 144:105998.

[13] Belvis F, Aleta A, Padilla-Pozo Á, Pericàs J-M, Fernández-Gracia J, Rodríguez JP, et al. Key epidemiological indicators and spatial autocorrelation patterns across five waves of COVID-19 in Catalonia. Sci Rep. 2023; 13(1):1–11.

[14] Ghosh P, Cartone A. A Spatio-temporal analysis of COVID-19 outbreak in Italy. Reg Sci Policy Pract. 2020; 12(6):1047–62.

[15] Cliff AD, Ord JK. Spatial processes: models & applications. (No Title). 1981;

[16] Getis A, Ord JK. Local spatial statistics: An overview. Spatial analysis: Modeling in a GIS environment. Longley, P., and M. Batty. Wiley, New York; 1996.

[17] Nguyen TT, Vu TD. Identification of multivariate geochemical anomalies using spatial autocorrelation analysis and robust statistics. Ore Geol Rev. 2019; 111.

[18] Anselin L. Local indicators of spatial association—LISA. Geogr Anal. 1995; 27(2):93–115.

[19] Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. In: Handbook of applied spatial analysis: Software tools, methods and applications. Springer; 2009. p. 73–89.