Exploring COVID-19 incidence hotspot in Metropolitan area of Pakistan using geo-statistical approach: a study of Lahore city

Shakeel Mahmood

Abstract  Globally, COVID-19 is a top level public health concern. This paper is an attempt to identify and assess COVID-19 incidence hotspots in Metropolitan area of Pakistan using geo-statistical approach. The study is based on secondary data. The COVID-19 confirmed cases record (15/03/2020 to 15/04/2020) of entire Metropolitan area is obtained from hospitals and National Institute of Health website. Point-level geo-coding technique was applied on patient’s record and the relative location was converted into absolute location. Getis-Ord Gi* statistical model was applied in ArcGIS 10.3 to calculate Z-score and P values for each point location representing the COVID-19 incidence intensity. Then inverse distance weighted technique of spatial interpolation was applied on Z-score and spatial clusters of crime were geo-visualized in the form hotspot and cold spot. Spatially, more than 50% of land area of Allama Iqbal, Samanabad, Gulburg and Cantonment is covered by very high incidence zone which is surrounded by high incidence zone whereas Ravi, Shalimar, and north of Wagha and Aziz Bhatti towns are located in very low incidence zone. This study provides a suitable methodological framework for identification and analysis infectious disease hotspots. The study can also facilitate health and related authorities to fight war against COVID-19. Similarly, it can help policy makers to manage the movement of travelers and restrict social interaction.

Keywords  Coronavirus · COVID-19 · Incidence · Mapping · Intensity · Hotspot · GIS · Pakistan

1 Introduction

The COVID-19 is a humanitarian emergency, which is started in the Wuhan province of China in early December 2019 and was declared as an emergency in the third week of January 2020 [1–3]. The World Health Organization (WHO) declared COVID-19 as Public Health Emergency of International Concern (PHEIC) on 31th of January 2020, and finally a pandemic on 11th March 2020. As of 28th April, 2020, there are 3.06million confirmed Coronavirus cases, with 0.213 million deaths while 0.907million has been recovered. In the total Coronavirus patients died, very interestingly the highest number belongs to Italy i.e. 4,032 deaths. The death toll is followed by China (3248), Iran (1433) and Spain (1044). Worldwide, the impact of this pandemic has affected countries and communities socially, economically and psychologically as well as international relations [4]. At present, COVID-19 pandemic is hot topic globally [3].

Globally, COVID-19 pandemic became a biological disaster and tens of thousands of people have lost their lives; millions are infected and quarantined. Man has experienced various pandemics throughout the history with different intensity and impact [5]. Today we are observing and experiencing a very tough time by fighting once again with an invisible enemy; the COVID-19 [6]. The Coronavirus infection spread started in early January 2020 in China [7]. The impact of this public health emergency has affected countries and communities in terms of economic, socio-psychological issues, as well as international relations [4].

Currently, a variety of human diseases is prevailing with unknown etiology [1, 8, 9]. Viruses have been considered as cause of these diseases which has increased the continuous search for new viruses [10–12]. Corona viruses, a
The COVID-19 is still a biological hazard and the entire population is at high risk in Pakistan. Therefore, the aim of this study is to carry out spatio-temporal analysis of COVID-19 incidence hotspot in metropolitan area of Pakistan using Geo-statistical Approach.

2 Materials and methods

Lahore is selected for COVID-19 incidence hotspot mapping which is the second largest city in Pakistan with population more than 6.5 million. Administratively, Lahore is divided into nine towns; Ravi, Shahalim, Wagah, Aziz Bhatti, Data Ganj Bakhsh (DGB), Samanabad, Allama Iqbal, Nishtar and Johar Town (Fig. 1). This study is based on secondary data. The data regarding registered COVID-19 confirmed cases were acquired from National Institute of Health (NIH) of Pakistan website. The data were classified in MS Excel. Point-level geo-coding technique is applied to geo-code 492 confirmed COVID-19 cases of past 30 days i.e. 15/03/2020 to 15/04/2020. The hotspots is determined by applying Getis-OrdGi* (G-i-star) statistical model in GIS environment to quantify Z-score and P values. Finally, Inverse Distance Weighting (IDW) technique of spatial interpolation is implemented to generate a spatial layer depicting hotspots and coldspot zones by utilizing Z-score as input data. In the current study, Getis-OrdGi* geo-statistical model has utilized for spatial clustering after. The Gi* statistic is more interested in assessing incident intensity than in analyzing the spatial clustering of any particular value associated with COVID-19 incidence. Similarly, Gi* statistic examines the individual locations, enabling hotspots to be identified based on a comparison with the neighboring samples and has been successfully applied in hotspot identification and mapping.

The Getis-Ord local statistic is given in Eqs. 1, 2 and 3 [26–28].

The Getis-Ord local statistic is given as:

$$G_{i}^{*} = \frac{\sum_{j=1}^{n} w_{i,j}x_{j} - \bar{x} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\left[ n \sum_{j=1}^{n} w_{i,j}^{2} - \left( \sum_{j=1}^{n} w_{i,j} \right)^{2} \right]}}$$

where $x_{j}$ is the attribute value for feature $j$, $w_{i,j}$ is the spatial weight between feature $i$ and $j$, $n$ is equal to the total number of features and:

$$\bar{x} = \frac{\sum_{j=1}^{n} x_{j}}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_{j}^{2}}{n} - \left( \bar{x} \right)^{2}}$$

In Pakistan, first case was reported on February 25, 2020. The total confirmed cases are 14,612, active cases 11,067, deaths 312 and recovered cases 3233 till 28th April, 2020. The main cause of COVID-19 pandemic in Pakistan is the entry of infected pilgrims from Iran, pilgrims and migrants from Saudi Arab. Similarly, infected people from UAE, Malaysia and China have also transmitted and spread the virus. These infected migrants travelled and meet different people over different locations and transmit corona virus to healthy people. This transmission continued till the complete lockdown. Currently, lockdown has minimized the interaction of infected and healthy persons. Similarly, the quarantine centers across the country have reduced the exposure of healthy citizens [25]. The COVID-19 is still a biological hazard and the entire population is at high risk in Pakistan. Therefore, the aim of this study is to carry out spatio-temporal analysis of COVID-19 incidence hotspot in metropolitan area of Pakistan using Geo-statistical Approach.
The $G^*_i$ statistic is a z-score so no further calculations are required.

2.1 Hotspot analysis

The geo-coded COVID-19 cases records were imported into ArcGIS10.2 and geo-visualized in the form of point feature class as spatial layer (Fig. 2). The layer was added as input data to perform hotspot analysis. The hotspot analyst, identifies spatial clusters of high values (hotspots) and low values (cold spots) by implementing Getis-Ord$G^*_i$ statistical model. It converts the geo-coded COVID-19 cases record into a new output feature having attributes of the z-score, $P$ value, and Gi-Bin confidence level ranging from $-3$ to $+3$. High counts of cases close together have $+ve$ z-score indicating hotspot and vice versa. The z-score range $\pm 3$ reflect statistical significance with 99-percent confidence level, $\pm 2$ bins reflect 95-percent confidence level, and $\pm 1$ bins reflect 90-percent confidence level, whereas 0 bin is not significant (Fig. 3). Then the IDW technique of spatial interpolation is implemented to generate the raster spatial layer depicting COVID-19 cases intensity as hotspot or coldspot. The z-score values are utilized as input data in spatial interpolation. The COVID-19 incidence intensity is categorized into five classes namely very low ($-2.1$ to $-3$), low ($-1.1$ to $-2$), moderate ($+1$ to $-1$), high ($1.1$ to $2$) and very high ($2.1$ to $3$). Linear regression analysis is applied on daily COVID-19 confirmed cases.

3 Result and discussion

The results obtained from processed data are explained in the following sections. The confirmed cases were 150 in Allama Iqbal town followed by Cantonment and DGB with less than 100 cases (Fig. 4). In Gulberg and Samanabad the number of confirmed cases was 60 and 64 respectively. Aziz Bhatti town is lowest number of confirm cases (Fig. 4). The spatial pattern of COVID-19 confirmed cases is variable across the district Lahore.
3.1 Spatial pattern of COVID-19 incidence

3.1.1 High to very high COVID-19 incidence zone (hotspots)

In the high to very high COVID-19 incidence zone has high to very high intensity of confirmed cases. Spatially, more than 50% of land area of Allama Iqbal, Samanabad, Gulberg and Cantonment is covered by hotspot (Fig. 5). Some areas in the north and west of Nishter Town and south of the DGB town are located in very high incidence zone. Whereas the remaining parts of the same towns are covered by high incidence zones. As a whole 63% land area of Lahore is covered by high to very high intensity zone.

3.1.2 Moderate COVID-19 incidence zones

The moderate COVID-19 incidence zone has moderate level of confirmed cases. Spatially, west of Wagha and south of Nishter Town has moderate incidence zone.

3.1.3 Low to very COVID-19 incidence zone (coldspots)

The low to very low COVID-19 incidence zone has low incidences. Spatially, Ravi, Shalimar, and north of Wagha and Aziz Bhatti towns are located in this zone.

Spatially, the COVID-19 incidence is very high (Hotspot) in central and south west of district Lahore whereas the surrounding of hotspots are covered by high incidence areas. The incidence is moderate in the south and west. Towards north incidence is decreasing forming coldspot with low incidence of COVID-19. The most affected (34%) age group is 31 to 40 years old followed (26%) by 21 to 30 years old (Fig. 6). The most vulnerable age group above 50 years old is also severely affected because 12% of
confirmed cases were of this group. No confirmed case is reported in age group less than 10 years old. The number of confirmed cases is growing very rapidly and still on rise (Fig. 7). This trend is depicting that the infected cases is likely to increase which will further grow the spatial extent of hotspot.

In Pakistan, first case was reported on February 25, 2020. Initially, District Sukker and Karachi were most affected with 273 and 154 confirmed cases respectively till
March, 26th, 2020. The district Sukker has highest number of cases because of returning pilgrims from Iran to Taftan (a town located in district Chaghai, Baluchistan) and then Sukker. This has resulted in high incidence in Sindh province but later on the travelling of unaware infected persons has transferred to Punjab province and transmitted in healthy people. Currently, Punjab has the highest number of confirmed cases followed by Sindh. The hotspot

Fig. 5 District Lahore: COVID-19 hotspot and coldspots zones

Fig. 6 COVID-19 incidence in different age groups

| Age Groups | COVID-19 Incidence (%) |
|------------|------------------------|
| <10        | 5                      |
| 11 to 20   | 10                     |
| 21 to 30   | 25                     |
| 31 to 40   | 35                     |
| 41 to 50   | 30                     |
| >50        | 15                     |
spatially extended towards Punjab and seems to grow further.

Prediction of infected persons can be made on the bases of existing data. The number of infected persons is likely to be increased but most importantly, it will require intensive health care units. This prediction is of crucial importance to plan for new facilities in hospitals. The very high incidence zones (Hotspots) are spatially located in the center of the city where population density is very high (> 1000 persons/Km²), crossed by number of major and minor roads, dense economic activities and high level of people’s interaction. These conditions have further increased the probability of new cases which will grow the spatial extent of hotspot. In such predictions, travelling restrictions towards hotspots need to be implemented and extraordinary community measures must be taken within and outside the hotspots. Currently, lock down has minimized the interaction of infected and healthy persons. Similarly, the quarantine facility across the country has decrease the exposure of healthy citizens.

4 Conclusion

It is concluded from the study that Pakistan is still in a good position than many other countries to react to the current outbreak. The aggressive and systematic approach taken by the Government of Pakistan has played a crucial role and resulted in containing the spread. Similarly, good health facilities have recovered more than 2500 infected persons. The quarantine centers are supporting the existing health system. It is predicted in the study that if the exponential trend continues for the next few weeks then the health system will be not able to respond. The measures taken by the government are in the right direction, but prediction tells us that they need to be implemented without delay. Otherwise, a substantial number will become inevitable.

Spatially, the COVID-19 incidence is very high (Hotspot) in central and south west of district Lahore. The incidence is moderate in the south and west. Towards north incidence is decreasing forming cold spot with low incidence of COVID-19. The spatial extent of hotspot seems to extend towards the borders of Lahore. In this regard, research on emergency bases is highly needed to minimize the spread of COVID-19 pandemic.

The study further concludes that the maximum number of patients will require intensive health care units. The spatio-temporal pattern and prediction of COVID-19 is of crucial importance to plan for new facilities in Pakistani hospitals and to calculate the time period in which they need to be available. Finally, lock down has minimized the interaction of infected and healthy persons. Similarly, the quarantine facility has decreased the exposure of healthy citizens to COVID-19. But still, the entire population in main cities particularly in Lahore is at high risk due to the rapidly growing COVID-19 confirmed cases.

Declarations

Conflict of interest There is no conflict of interest.

Ethical approval There are no ethical issue related to the data and manuscript.

References

1. Li, Q., Guan, X., Wu, P., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New England Journal of Medicine. https://doi.org/10.1056/NEJMoa2001316
2. Shankar, K., Mohanty, S. N., Yadav, K., Gopalakrishnan, T., & Elmisery, A. M. (2021). Automated COVID-19 diagnosis and classification using convolutional neural network with fusion
based feature extraction model. *Cognitive Neurodynamics*. https://doi.org/10.1007/s11571-021-09712-y

3. Satpathy, S., Mangla, M., Sharma, N., Deshmukh, H., & Mohanty, S. (2021). Predicting mortality rate and associated risks in COVID-19 patients. *Spatial Information Research, 19*(2), 455–464. https://doi.org/10.1007/s41324-021-00379-5

4. Hua, J., & Shaw, R. (2020). Corona Virus (COVID-19) “Infodemic” and emerging issues through a data lens: The case of China. *International Journal of Environmental Research and Public Health, 17*, 2309. https://doi.org/10.3390/ijerph17072309

5. Remuzzi, A., & Remuzzi, G. (2020). COVID-19 and Italy: What next? *The Lancet, 395*(10231), 1225–1228. https://doi.org/10.1016/S0140-6736(20)30627-9

6. Bastola, A., Sah, R., Rodriguez-Morales, A. J., Lal, B. K., Jha, R., Ojha, H. C., et al. (2020). The first 2019 novel coronavirus case in Nepal. *The Lancet Infectious Diseases, 20*, 279–280.

7. The Economist. (2020). Tourism flows and death rates suggest covid-19 is being under-reported. https://www.economist.com/graphicdetail/2020/03/07/tourism-flows-and-death-rates-suggest-covid-19-is-being-under-reported.

8. Rodriguez-Morales, A. J., Bonilla-Aldana, D. K., Balbin-Ramon, G. J., Paniz-Mondolfi, A., Rabaan, A., Sah, R., et al. (2020). History is repeating itself, a probable zoonotic spillover as a cause of an epidemic: The case of 2019 novel Coronavirus. *Infezioni Med Le, 28*, 3–5.

9. Callaway, E. (2020). Time to use the p-word? Coronavirus enter dangerous new phase. *Nature, 579*, 12.

10. Stohlman, S. A., & Hinton, D. R. (2001). Viral induced demyelination. *Brain Pathology, 11*, 92–106.

11. Jubelt, B., & Berger, J. R. (2001). Does viral disease underlie ALS? Lessons from the AIDS pandemic. *Neurology, 57*, 945–946.

12. Shingadia, D., Bose, A., & Booy, R. (2002). Could a herpesvirus be the cause of Kawasaki disease? *Lancet Infectious Diseases, 2*, 310–313.

13. Rota, P. A., et al. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science, 300*, 1394–1399.

14. Martina, B. E., et al. (2003). Virology: SARS virus infection of cats and ferrets. *Nature, 425*, 915.

15. Harapan, H., Itoh, N., Yufika, A., Winardi, W., Kearn, S., Te, H., Megawati, D., Hayati, Z., Wagner, A. L., & Mudaris, M. (2020). Coronavirus disease 2019 (COVID-19): A literature review. *Journal of Infection and Public Health, 13*(5), 667–673.

16. Covid, C. D. C., Team, R., COVID, C., Team, R., COVID, C., Team, R., ... & Walters, M. (2020). Characteristics of health care personnel with COVID-19—United States, February 12–April 9, 2020. *Morbidity and Mortality Weekly Report, 69*(15), 477.

17. Van der Hoek, L., Pyrc, K., Jubbink, M. F., Vermeulen-Oost, W., Berkhout, R. J., Wolthers, K. C., Wertheim-van Dillen, P. M. E., Kaandorp, J., Spaargaren, J., & Berkhout, B. (2004). Identification of a new human coronavirus. *Nature medicine, 10*(4), 368–373.

18. Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D. Y., Chen, L., & Wang, M. (2020). Presumed asymptomatic carrier transmission of COVID-19. *JAMA, 323*(14), 1406–1407.

19. Leung, G., & Wuhtpps, J. (2020). Real-time nowcast and forecast on the extent of the Wuhan CoV outbreak, domestic and international spread. https://www.med.hku.hk/f/news/3549/7418/Wuhan coronavirus-outbreak_AN-UPDATE_20200127.pdf. Accessed April 21, 2020.

20. Rodriguez-Morales, A. J., MacGregor, K., Kanagarajah, S., Patel, D., & Schlagenhauf, P. (2020). Going global - travel and the 2019 novel coronavirus. *Travel Medicine and Infectious Disease, 33*, 101578.

21. Arab-Mazar, Z., Sah, R., Rabaan, A. A., Dhama, K., & Rodriguez-Morales, A. J. (2020). Mapping the incidence of the COVID-19 hotspot in Iran-Implications for Travellers. *Travel Medicine and Infectious Disease*. https://doi.org/10.1016/j.tmaid.2020.101630

22. Khadidos, A., Khadidos, A. O., Kannan, S., Natarajan, Y., Mohanty, S. N., & Tsaramirsis, G. (2020). Analysis of COVID-19 infections on a CT image using deepsense model. *Frontiers in Public Health*. https://doi.org/10.3389/fpubh.2020.599550

23. Dash, S., Chakravarty, S., Mohanty, S. N., Pattanaik, C. R., & Jain, S. (2021). A deep learning method to forecast COVID-19 outbreak. *New Generation Computing, 39*(2), 437–461. https://doi.org/10.1007/s00354-021-00129-2

24. Parvin, F., Ali, S. A., Hashmi, S., & Ahmad, A. (2021). Spatial prediction and mapping of the COVID-19 hotspot in India using geostatistical technique. *Spatial Information Research, 29*(4), 479–494.

25. Government of Pakistan. (2020). National Institute of Health, Islamabad. Ministry of National Health Services, Regulation and Coordination, Pakistan.

26. Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*(3), 189–206.

27. Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis, 27*(4), 286–306.

28. Getis, A., & Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. In *Perspectives on spatial data analysis* (pp. 127–145). Springer, Berlin, Heidelberg.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.