



Spatial prediction and mapping of the COVID-19 hotspot in India using geostatistical technique

Farhana Parvin¹ · Sk Ajim Ali¹ · S. Najmul Islam Hashmi¹ · Ateeque Ahmad¹

Received: 3 August 2020 / Revised: 13 November 2020 / Accepted: 19 November 2020 / Published online: 4 January 2021
© Korean Spatial Information Society 2021

Abstract The world has now facing a health crisis due to outbreak of novel coronavirus 2019 (COVID-19). The numbers of infection and death have been rapidly increasing which result in a serious threat to the social and economic crisis. India as the second most populous nation of the world has also running with a serious health crisis, where more than 8,300,500 people have been infected and 123,500 deaths due to this deadly pandemic. Therefore, it is urgent to highlight the spatial vulnerability to identify the area under risk. Taking India as a study area, a geospatial analysis was conducted to identify the hotspot areas of the COVID-19. In the present study, four factors naming total population, population density, foreign tourist arrivals to India and reported confirmed cases of the COVID-19 were taken as responsible factors for detecting hotspot of the novel coronavirus. The result of spatial autocorrelation showed that all four factors considered for hotspot analysis were clustered and the results were statistically significant (p value < 0.01). The result of Getis-Ord G_i^* statistics revealed that the total population and reported COVID-19 cases have got high priority for considering hotspot with greater z -score (> 3 and > 0.7295 respectively). The present analysis reveals that the reported cases of COVID-19 are higher in Maharashtra, followed by Tamil Nadu, Gujarat, Delhi, Uttar Pradesh, and West Bengal. The spatial result and geospatial methodology

adopted for detecting COVID-19 hotspot in the Indian subcontinent can help implement strategies both at the macro and micro level. In this regard, social distancing, avoiding social meet, staying at home, avoiding public transport, self-quarantine and isolation are suggested in hotspot zones; together with, the international support is also required in the country to work jointly for mitigating the spread of COVID-19.

Keywords COVID-19 · Foreign tourist arrivals · Hotspot analysis · India · Population pattern · Spatial autocorrelation

1 Introduction

The newly discovered 2019 coronavirus (COVID-19 or 2019-nCoV) is an example of severe acute respiratory syndrome coronavirus or commonly known as SARS-CoV, which was rapidly spread over 37 countries of the world during 2002–2003 and caused more than 8000 confirmed cases and 800 deaths [1]. On 11th March 2020, the COVID-19 was declared as a pandemic by the world health organization [2]. The new coronavirus is a respiratory tract infection that started in the Wuhan city of China. But presently spread over many countries of the world. By 25th July 2020, about 15,676,604 confirmed cases and about 638,577 deaths across 188 countries had reported [3].

In India, the first 2019-nCoV case was reported in Kerala on 30th January 2020, the positive COVID-19 patient was a student returning from Wuhan, China [4]. Since the beginning of the COVID-19 in India, major strategies for enhanced vigilance and preparedness were initiated at both state and district level to fight with 2019-nCoV. But the situation was not hopeful and 360

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s41324-020-00375-1>) contains supplementary material, which is available to authorized users.

✉ Farhana Parvin
farhanaparvin93@gmail.com

¹ Department of Geography, Faculty of Science, Aligarh Muslim University (AMU), Aligarh, UP 202002, India

more confirmed cases were reported on 22nd March from 23 states/union territories throughout India, which resulted to announce nationwide lockdown by Prime Minister till 14th April. However, the cases rapidly increased after 22nd March and currently reach to 13,37,022 confirmed cases in all 33 states/union territories as on 25th July 2020 [5].

In this crisis of public health, the Government of India continuously making new initiatives to fight with the COVID-19 and extend the timeline of nationwide lockdown till 3rd May. To reinforce the government's initiatives, all districts of the country have now categorised into three zones, naming red, orange, and green. More than 170 districts have declared as red zone (a large number of positive cases) and 207 as green zone (limited number of positive cases) based on the number of confirmed cases reported [6]. With constantly increasing the number of cases, maximum districts have a chance to come under the red zone. In this regard, the present study aimed to identify major hotspot (red zone) of the COVID-19 by considering responsible factors and using the geospatial technique.

The hotspot is an emerging and dynamic concept which changes with time-based on diseases spread. A hotspot is a spatial tool that helps in ensuring the centre of emerging cases. Presently, different geospatial techniques have been frequently applied for tracing and monitoring the 2019-nCoV pandemic. A recent study showed the application of geographic information system and spatial tools for mapping the incidences of SARS-CoV-2 globally [7]. Their findings suggest that how spatial methods can be helpful to identify locations of new health facilities to the affected groups. As a critic of this work due to lack of available resources and time to build new health facilities, the present emphasized to identify locations of new cases and hotspot which can more help to make emergency measures in existing facilities available.

As an aim to deal with a communicable and infectious disease, it is difficult to decide a particular factor for the COVID-19 based on which hotspot can be declared. That's why the present study employed three more responsible factors towards the outbreak of this disease besides with reported 2019-nCoV cases. Data from the Registrar General & Census Commissioner, Government of India was used to show spatial pattern and cluster of the district-wise total population and population density. The foreign tourist arrivals data was collected from the Ministry of Tourism, Government of India to spatial interpolate and recognize the high percentage of foreigner's arrivals. Additionally, district-wise total infected persons were mapped to identify the hotspot throughout India. The locations of hotspot were recognized and suggested for critical preparedness, readiness and response actions towards this health crisis of the COVID-19.

2 Materials and methods

The present study aimed to map COVID-19 hotspot in India. To reach such aim different conditioning factors had been considered and geostatistical assessment performed. The detail of methodology adopted in this study is shown in Fig. 1.

2.1 Study area

India is the second-largest populous country of the world, located in South Asia between 08°04'N to 37°06'N and 68°7'E to 97°25'E. Presently, India has 28 states and 8 union territories. India is the home of 1,369 million populations. Many health risks can arise due to overpopulation. Human health is at risk as to the denser the population [8]. The rapid growth of population has created many environmental challenges and the emergence of various airborne and infectious diseases. In the past, several diseases like malaria, cholera, influenza, dengue, and tuberculosis were also spread in India and affected a large population [9].

Now, India is running by another health problem, it is a novel coronavirus. Even after taking many initiatives, after 1st July constantly more than 20,000 coronavirus cases have been reporting every day. No medical treatment, medicine, and the vaccine can fight to decrease the rate. Based on numbers and frequency of cases, many cities like New Delhi, Mumbai, Pune, Thane, Ahmedabad, Chandigarh, Bengaluru, Kanpur, Kasaragod, Indore, Bhopal, Kochi, Bhubaneswar, Mohali, Jaipur, Chennai, Kolkata, Tirunelveli, Guntur, Gurugram, and Noida are now treated as the COVID-19 hotspot (Fig. 2). These areas can also be considered as hotbeds for spreading this community infection to other areas shortly. The Government of India has enhancing rapid vigilance and preparedness measures to fight against the COVID-19 in every corner of the country. Besides, the present study, hotspot analysis would offer a new finding for increasing spatial surveillance for the country.

2.2 Conditioning factors for mapping COVID-19 hotspot

In the present study, total population, population density, total foreign tourist arrivals (FTAs) to India and total confirmed cases of the COVID-19 were considered as determining factors for hotspot mapping. The COVID-19 is a novel coronavirus and researches are still going on to understand more about this fatal virus. About little information of the COVID-19 says that this disease spreads through close contact by breathing, coughing, sneezing and

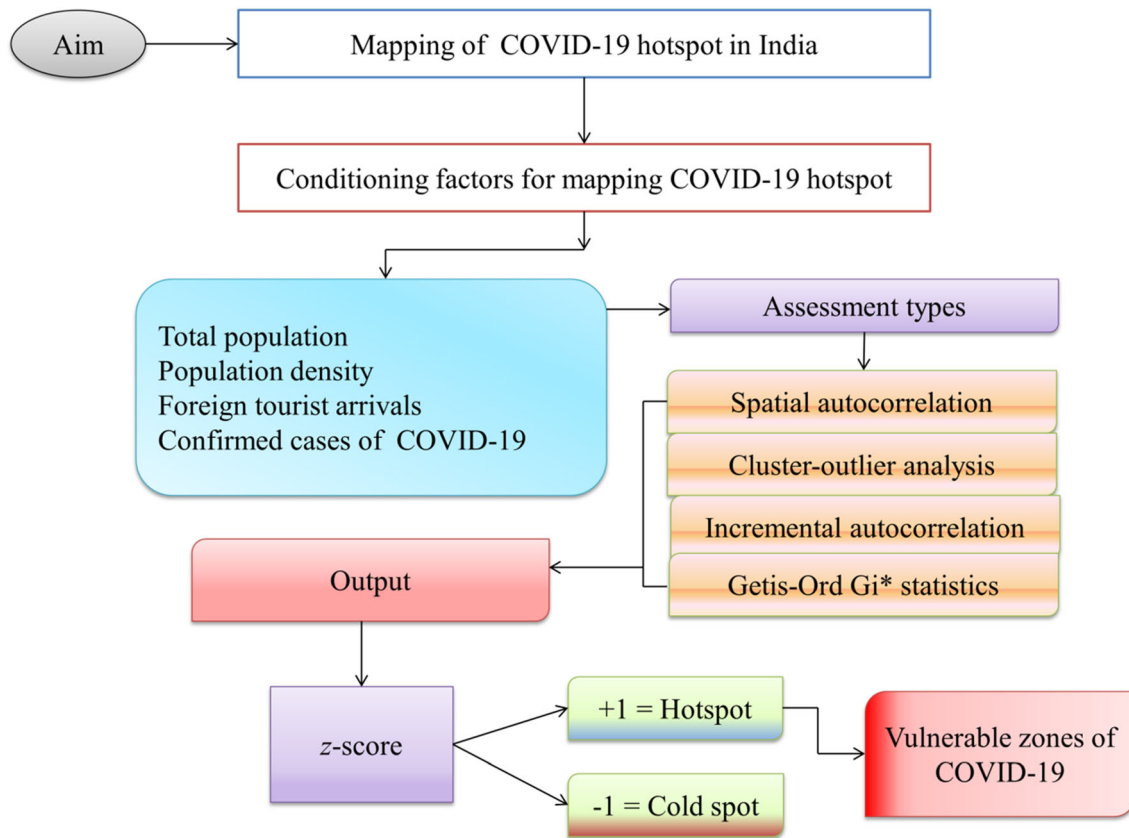


Fig. 1 Flow chart showing details of methodology adopted in the present study

even after touching an object or materials that has this virus. There are still not invented any measure or treatment to fight with this virus, but the rate of spread can be slowed by keeping social distance, avoiding social meet, staying at home, avoiding transport, managing for self-quarantine and isolation.

Thus, the areas with a high proportion of the population can be considered as vulnerable areas of spreading this disease. Along with those places that are recently visited by outer country’s people that has this virus can also be vulnerable towards the outbreak of this disease. So, in the present study, four factors i.e. total population, population density, foreign tourist arrivals (FTAs) and last one month confirm cases of the COVID-19 were considered for spatial mapping and hotspot analysis in Indian districts.

A large number of population and high population density is a responsible factor for spreading the COVID-19 pandemic. Many innovative and productive cities with large clustering of people have also vulnerable to this infectious disease. Population density is a key factor that determines the vulnerability of a place. It is evidenced by the reported cases of the COVID-19 that has hit hard in many places across the world. A densely populated city like New York, London, and Mumbai with a large

population, dense residential areas, and diverse global flow of foreign tourists have reported the worse situation to this pandemic and it is important to differentiate hotspots by considering the characteristics of places from where it first cropped and spread [10].

In the case of India, the novel coronavirus has spread because of many factors that have found by analysing many cases of highly affected states. Indian state like Maharashtra, Delhi, Chennai, Kolkata, Bangalore, Hyderabad etc. are highly urbanised, densely residential areas and a high proportion of the floating population. These cities have also busiest airport service in the country which supports maximum foreign tourist inflow inside India from different countries like USA, China, UK, Russia, Germany, France, Italy and many others. Indeed, this kind of foreign passenger’s inflow inside India is the warning of high risk to this fatalities infection in this situation [11]. The first confirmed case of COVID-19 in India was reported on 30th January 2020 in a south-western state. The affected person had travelled from Wuhan city, China [12]. But till 1st March 2020, either confirmed or deceased cases were raising, although Mumbai, New Delhi, Bengaluru, Hyderabad, Kolkata, and Patna were in the high suspected list. After 20th March, both confirmed and deceased cases are



Fig. 2 The COVID-19 hotspot areas based on number and frequency of reported cases

continuously raising and now reached to more than 158,093. Not only large population and high density are responsible for such spreading, one of the important reason behind this crisis is that the passengers arrived from abroad were initially not checked in the busiest airport or other check post.

Now India is going a nationwide lockdown since mid-night 23rd March and ordered of Prime Minister to continue till 31st May, which may extend further after that dateline. But even after completely lockdown no expected result appeared and constantly increasing both confirmed and deceased cases. Thus, in this worse situation of the

COVID-19, it is relevant to the spatial analysis of this novel coronavirus in India.

Based on reported cases, the Government of India has declared many districts as a hotspot to COVID-19, while others are classified as non-hotspot and non-infectious districts. But this was done only based on reported cases. Those districts have reported a large number of confirmed cases, were declared as hotspot district. Indeed, it is correct for analysing hotspot, the number of infected cases is important but this will not predict the future areas of the hotspot to this pandemic. Thus, in the present study, the relevant factors mentioned above (i.e. the total population, population density, foreign tourist arrivals (FTAs) in India between December and February, and reported COVID-19 cases) were considered for hotspots detection which not only shows the present hotspot areas but also predicts more vulnerable areas to novel coronavirus for future.

For micro-level analysis, district wise-population data, total FTAs in the last three months and confirmed cases of COVID-19 were taken into consideration. The district-wise total population and population density are shown in the supplementary file (Table S1). As per the data of the Ministry of Tourism, a total number of 377,092, 738,803, and 381,267 foreign tourists arrived in India during December 2019, January 2020, and February 2020 respectively. The average percentage of shares of topmost countries has mapped (Fig. 3). The district-wise COVID-19 confirmed cases were also considered as an important factor for analysing the hotspot.

2.3 Data source

Three types of numeric data like number and density of population, the percentage of foreigners arrived, and the latest data on confirmed COVID-19 cases were used in this study. The district-wise population data of 2011 was obtained from the official website of the Registrar General & Census Commissioner, Government of India. The total population of each district of India were divided with the respective area to calculate the density. The data on foreign tourist inflow to India for December, January, and February was collected from the Ministry of Tourism, Government of India. Moreover, the data on COVID-19 from 30th January to 17th May 2020 was collected from the Ministry of Health and Family Welfare (MoHFW), Government of India and data of 17th May 2020 was used for spatial mapping (Table S1). For preparing GIS database and spatial distribution, the district-level shapefile was created from district administrative division (Scale-1:25,000,000) and numeric data were joined in attribute field of each district by geo-coding and matching with district code in a GIS environment.

2.4 Spatial autocorrelation

Spatial autocorrelation had been used to show the degree of similarity in respect of other nearby objects. Generally, Global Moran's I statistics had been used to measure the spatial correlation in which three types of classification statistics have appeared as positive, negative and no autocorrelation [13]. The positive and negative autocorrelation shows respectively similar and dissimilar values cluster in a map. The Global Moran's I statistics are comparable to the Karl Pearson coefficient [14], which is expressed as:

$$M_I = \frac{N}{S_x} \sum_l \sum_k W_{lk} \frac{(y_l - p)(y_k - p)}{\sum_l (y_l - p)^2}$$

where M_I is the Global Moran's I statistics, N is the No. of districts; and S_x is the spatial attribute, which is further defined as:

$$S_x = \sum_l \sum_k W_{lk}$$

W_{lk} is the component in the spatial weight equivalent to the observation pair l, k , y_l and y_k for areas l and k with mean y . The summation of spatial weights are equal to 1 ($\sum W_{lk} = 1$). Primarily in spatial autocorrelation, a spatial matrix needs to construct that contains evidence about the neighbouring location. The non-neighbouring locations are assumed weight of 0.

The spatial autocorrelation produces five statistical values, naming Moran's I index, expected index, variance, z-score, and p value. Where z-score and p value indicate statistical significance. The value of Moran's I index ranges between +1 and -1; where a positive value points toward clustering and a negative value points toward dispersion (Table 1).

2.5 Cluster-outlier analysis

The Global Moran's I statistics generally specify the tendency of clustering and dispersing of similar and dissimilar values. While, Cluster-outlier distinguishes between statistically significant ($p < 0.01$) and insignificant ($p > 0.01$) cluster of low values (low-low) and high values (high-high), the outlier of low value which is surrounded by high values (low-high), an outlier of high value surrounded by low values (high-low). Additionally, Anselin Local Moran's I was used here to observe the spatial outlier. A positive and negative value of ' I ' indicates that the selected feature is respectively enclosed with similar and dissimilar values. The Anselin Local Moran's I is interpreted with the context of returned z-score and p value. A p value less than 0.01 (99% significance level) was taken for local autocorrelation (Table 2).

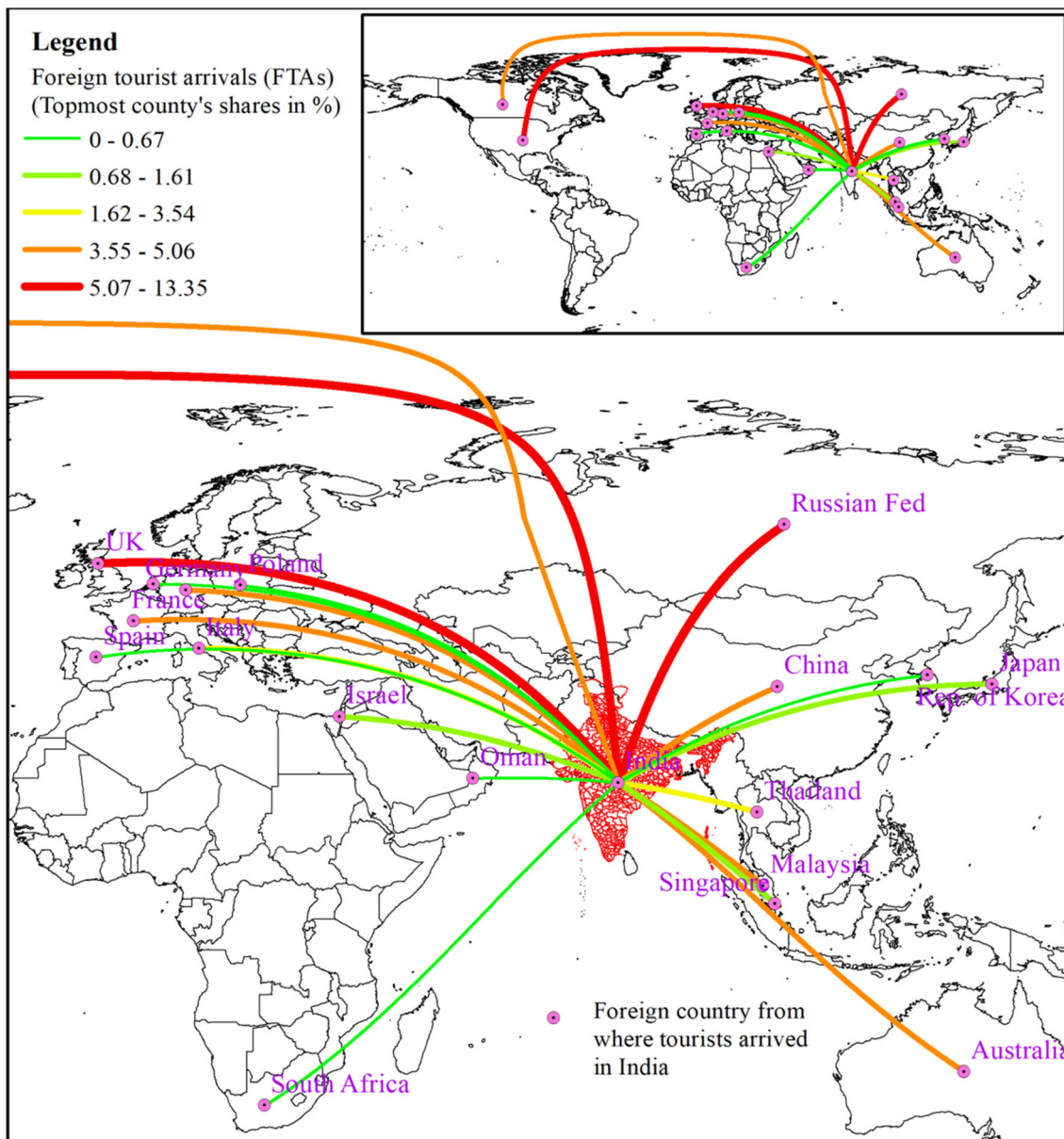


Fig. 3 The source countries from where foreigners came during last three months (based on e-Tourist Visa)

Table 1 Spatial autocorrelation of different factors considered for hotspot analysis of the COVID-19 using Moran’s *I* statistics

Factors	Observed (Moran’s <i>I</i>)	Expected	Variance	z-score	<i>p</i> value	Pattern
Total population	0.6075	- 0.0013	0.000268	37.167	0	Clustered
Population density	0.1035	- 0.0013	0.000241	6.7563	0	Clustered
Foreign tourist arrivals (FTAs)	0.11	- 0.0013	0.000257	6.956	0	Clustered
Reported 2019-nCoV cases	0.0879	- 0.0013	0.000258	5.5559	0	Clustered

The value of Moran’s *I* index (observed) ranges between + 1 and - 1. The positive values imply more clustering and significant. *p* < 0.01 indicates statistical significance of the cluster

Table 2 Cluster-outlier analysis of different factors considered for hotspot analysis of the COVID-19 using *Getis-Ord Gi** statistic

Factors	Observed (Moral’s I)	Expected	Variance	z-score	p value	Remarks
Total population	0.375776	− 0.00132	0.000068	45.59242	< 0.01	HC
Population denisty	0.129985	− 0.00132	0.00005	18.49292	< 0.01	LC
Foreign tourist arrivals (FTAs)	0.180797	− 0.00132	0.000086	19.65839	< 0.01	LC
Reported 2019-nCoV cases	0.263732	− 0.00132	0.000054	36.07419	< 0.01	HC

HC highly considered for COVID-19 hotspot, LC least considered for COVID-19 hotspot

2.6 Incremental spatial autocorrelation

For hotspot analysis in a GIS environment, the scale of analysis is important which specify the distance value. Incremental spatial autocorrelation is an autocorrelation by distance. It runs the Global Moran’s *I* (spatial autocorrelation) for measuring the intensity of clustering with increasing distances. The output z-score is used to determine the intensity of clustering. Usually, z-score indicates the rate of clustering that intensified with increasing distance. But at a certain distance, z-score become peaks, which reveal distances where clustering are most prominent (Fig. 10).

2.7 Hotspot analysis and detection of vulnerable zones (Getis-Ord *Gi** statistics)

A hotspot is defined as a region which has a greater concentration of incidences as compared to the expected number of random distribution [15]. Hotspot indicates the form of clustering in spatial distribution [16]. The main difference between density and hotspot is that both analyse the cluster but density is not statistically significant, while hotspot uses features to identify statistically significant locations in the data.

Hotspot generates z-scores and p values. A significant hotspot is indicated by a high z-score and low p value, while a negative and low z-score indicates a significant cold spot. In the present study, Getis-Ord *Gi** (G_i^*) was used to separate the clusters of high values (hotspot) from clusters of low values (cold spots). The Getis-Ord *Gi** statistics is expressed as:

$$G_i^* = \frac{\sum_{l=1}^n W_{l,k}y_k - \bar{Y} \sum_{l=1}^n W_{l,k}}{\sqrt{\frac{\pi}{n-1} \sum_{l=1}^n W_{l,k}^2 - (\sum_{l=1}^n W_{l,k})^2}}$$

where y_k is the attribute value of the feature k , $W_{l,k}$ is the spatial weight of l and k , n is the total number of features, and;

$$\bar{X} = \frac{\sum_{k=1}^n y_k}{n}$$

$$S = \sqrt{\frac{\sum_{k=1}^n y_k^2}{n} - (\bar{X})^2}$$

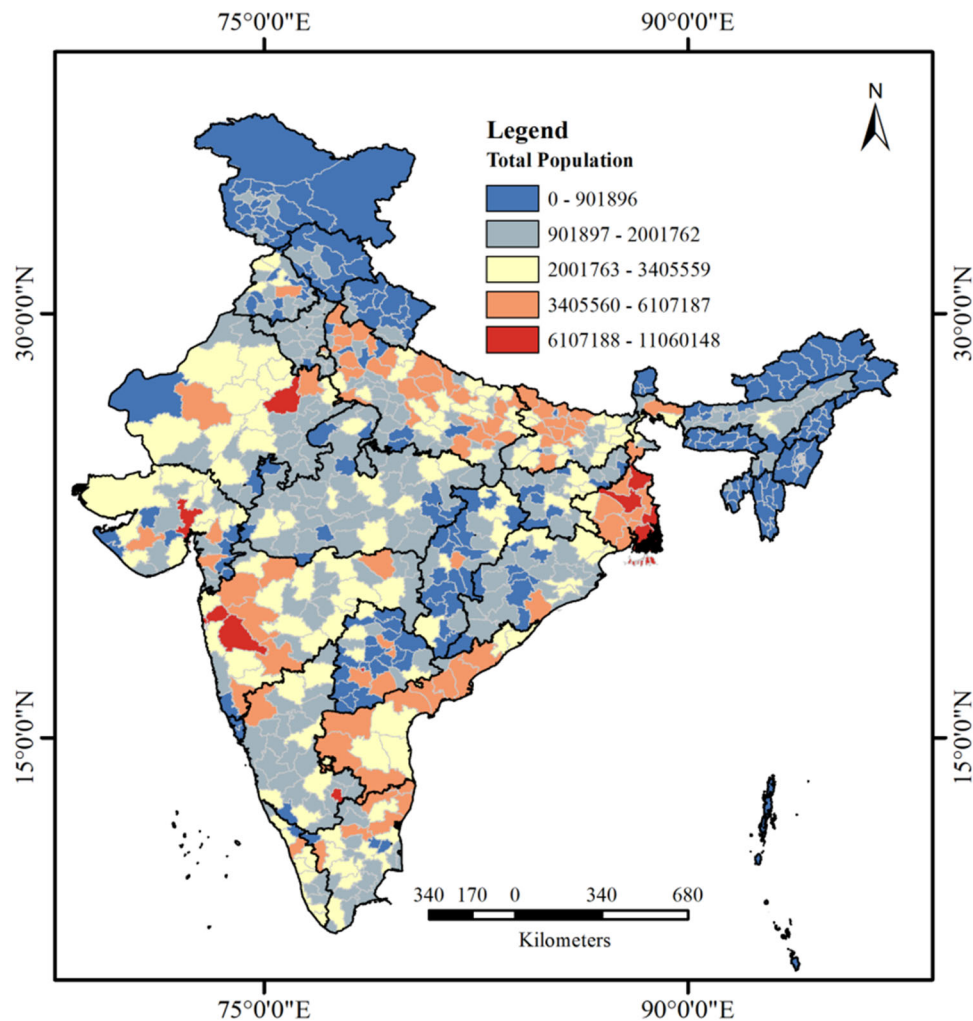
The result expresses the high z-score and low p value of computed G_i^* which represent the statistical significance of spatial clusters. In the present study, G_i^* was used to detect the vulnerable and non-vulnerable areas of the COVID-19 by representing hotspot (spatial clusters of high values) and cold spots (spatial clusters of low values) respectively.

3 Results

3.1 Spatial results

The hotspot mapping of 2019-nCoV in the present study was based on a number of the total population, population density, foreign tourists arrival in the last three months, and reported confirmed cases of the COVID-19. Hotspot and cold spot were identified by district-wise data. The presumption was areas with high population, density, and high rate of tourists from abroad could be more vulnerable to the COVID-19. The spatial autocorrelation was then run to prove the presumption. The spatial results based on total population revealed that the ‘Thane’ district of Maharashtra with 11,060,148 populations is the highest populous district and Dibang Valley of Arunachal Pradesh is the lowest populous district in India. Many other districts of Maharashtra, Gujarat, Bihar, Tamil Nadu, Haryana, Uttar Pradesh, Andhra Pradesh, and West Bengal have a high population (Fig. 4). In term of population density, about 80% districts of Uttar Pradesh, Bihar, West Bengal, Telangana, Kerala, and Tamil Nadu have greater than 500/Km² population density. Additionally, maximum COVID-19 confirmed cases have been reporting from some districts of these states like Mumbai, Kolkata, Hyderabad, Chennai, East Delhi, North East Delhi, Central Delhi, Nizamabad, and Kamareddy, where population density is more than 20,000/km² (Fig. 5).

Fig. 4 District-wise total population of India



In the beginning, foreign tourist's arrival was also played an important role in spreading COVID-19 throughout India. Result of the last three months data (December to till the international flight ban to India) revealed that USA, UK, Russian Federation, and China constantly shared a large percentage of FTAs to India. Others major countries like Malaysia, Australia, Canada, France, Italy, Germany, Singapore, and Japan also shared a countable percentage (Fig. 6). Three international airport naming, Delhi airport, Mumbai airport, and Dabolim (Goa) airport received about 10–35% of the total incoming foreigners. Others important airports like Chennai, Bengaluru, Kolkata, Kochi, Hyderabad, and Ahmedabad also received denumerable percentage (Fig. 7).

The attribute data of foreign tourist arrivals have mapped using inverse distance weighting (IDW) interpolation technique (Fig. 8). The result revealed that Delhi, Punjab, Haryana, Rajasthan, and maximum districts of Madhya Pradesh and Maharashtra are major states that have maximum spatial coverage of FTAs and the maximum confirmed cases of COVID-19 have been reported from these

states. District-wise reported confirmed cases of the COVID-19 from 30th January to 15th July was mapped to show the spatial distribution of this pandemic in India (Fig. 9). The spatial result showed that there are more than 110 districts of different states like Delhi, Maharashtra, Punjab, Haryana, Rajasthan, Gujarat, Andhra Pradesh, Kerala, Tamil Nadu, Telangana, Uttar Pradesh, Chhattisgarh, and West Bengal where greater than 300 per/district COVID-19 positive cases have been reported. Some districts have reported higher confirmed cases like Mumbai (98,980), Chennai (82,198), Thane (87,781), Pune (66,614), Ahmadabad (20,097), Kolkata (13,808), Surat (8421), Indore (5182), Jodhpur (4311), Bhopal (3894) and Jaipur (3902).

3.2 Spatial autocorrelation of the COVID-19

The Global Moran's I statistics of spatial autocorrelation showed that all four factors considered for hotspot analysis of the COVID-19 were clustered and the results were also highly significant (p value < 0.01) (Table 1, Fig. 10). The

Fig. 5 District-wise population density (persons/km²)

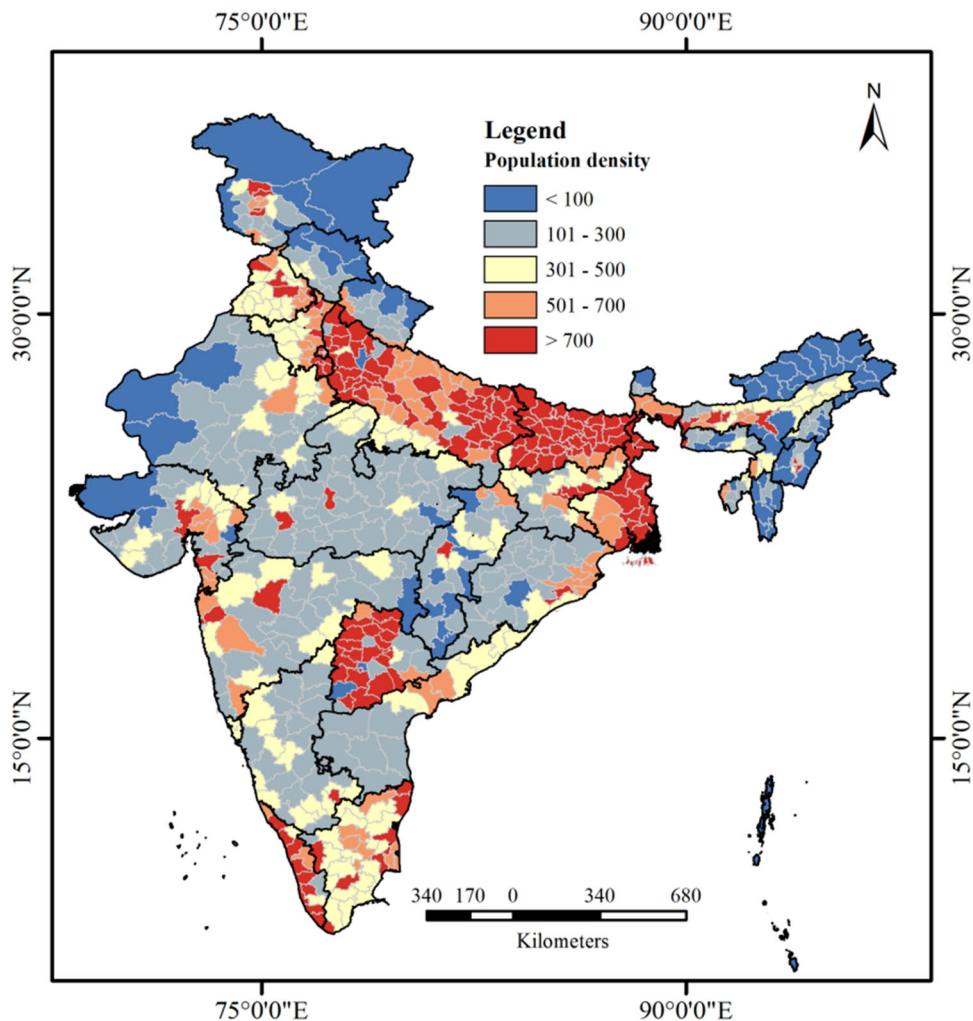


Fig. 6 The % shares of topmost source countries availing e-Tourist Visa

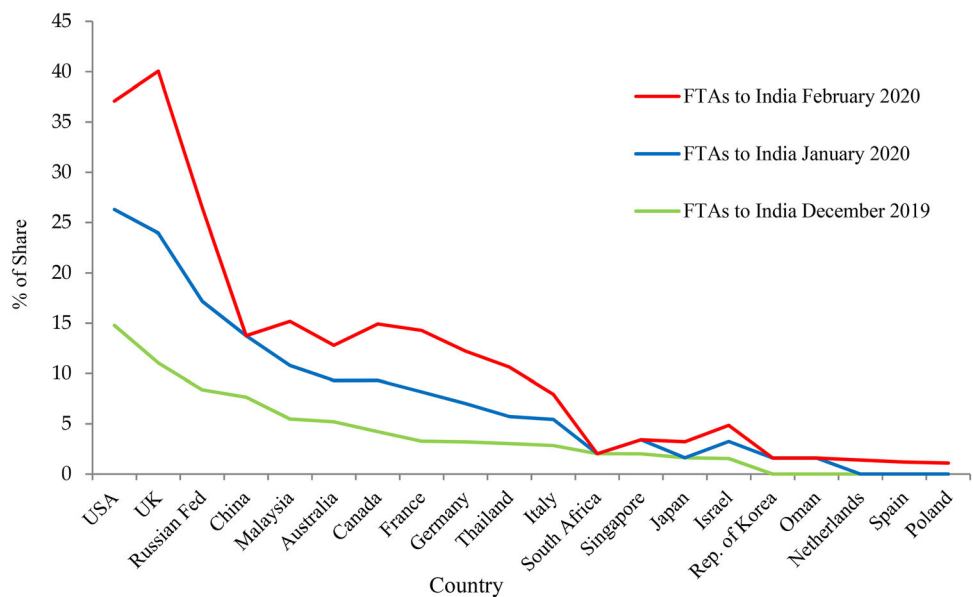


Fig. 7 The % shares of topmost airports in tourist arrivals on e-Tourist Visa

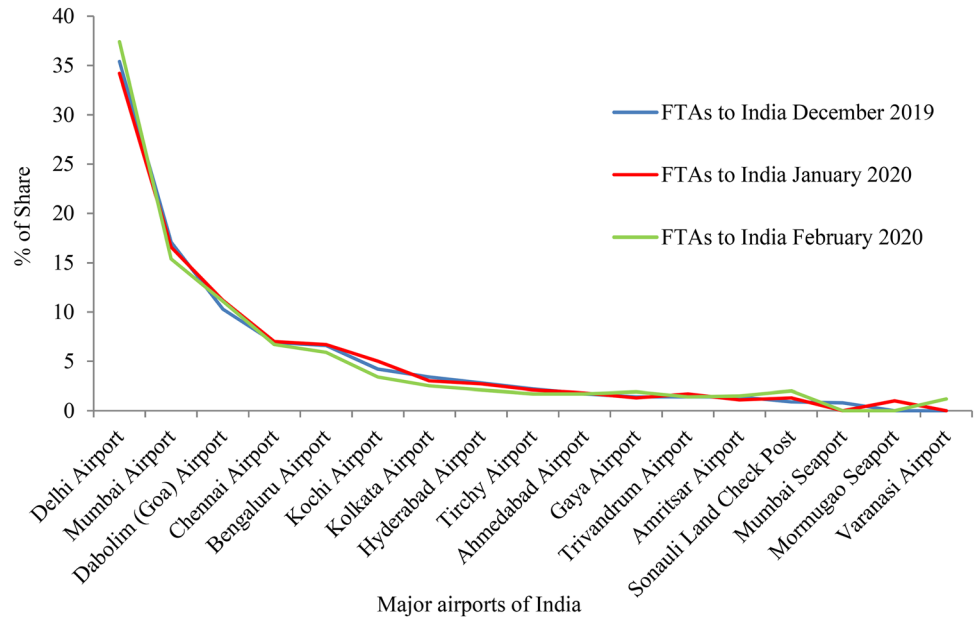


Fig. 8 The interpolation map showing foreign tourist arrivals in India through major airports

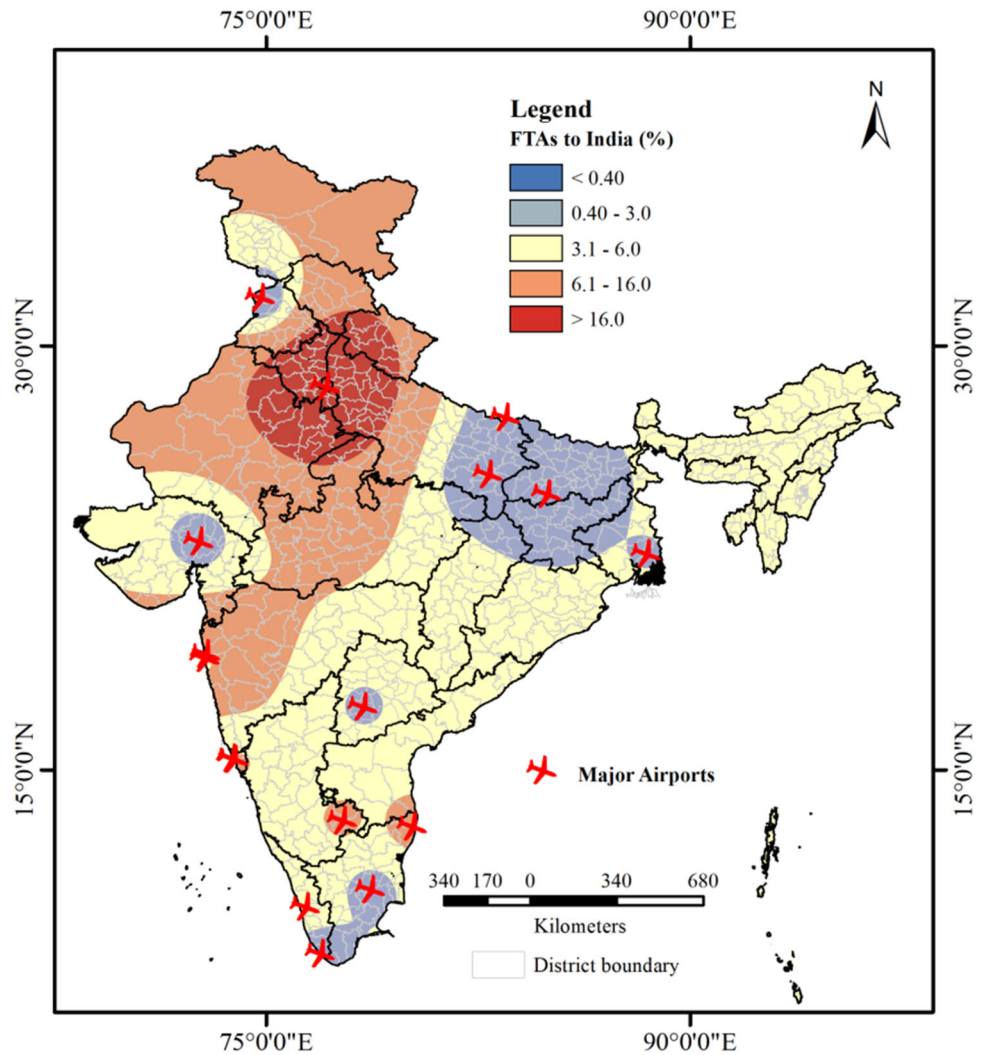
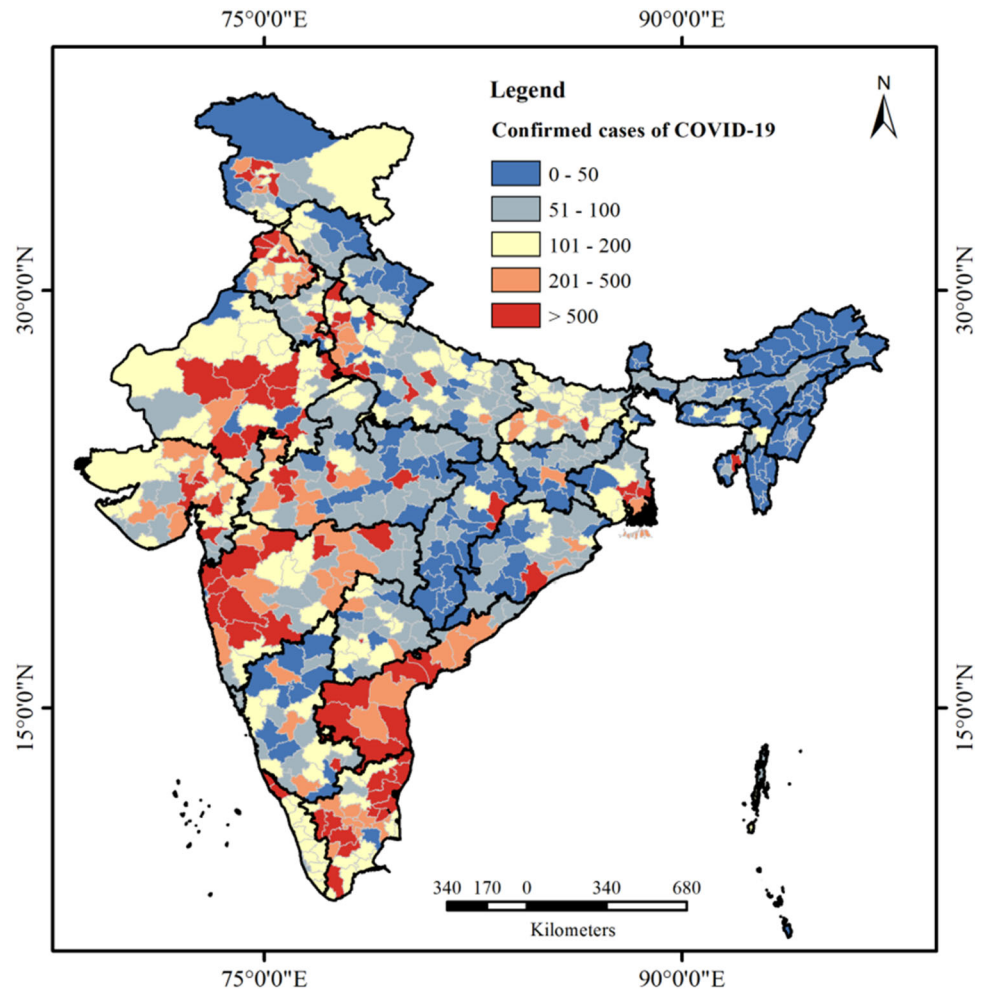


Fig. 9 District-wise confirmed cases of COVID-19



spatial autocorrelation result of the distribution of total population produces the highest Moran's I value as 0.6075 which reveals that the probability of random chances is $> 1\%$. On the contrary, reported 2019-nCoV cases to offer lowest Moran's I value as 0.0879. We can, therefore, conclude that the similar values of each factor tend to have a similar distribution, can be high or low. Also, the lower value of Moran's I , indicating that the random probability of cluster occurs in a very close or small district. This finding specifies that the COVID-19 is being occurred in the cluster which is randomly distributed throughout the country. Based on this result, the output locations may be considered a hotspot for the future days, where strategies should strictly execute for controlling this crisis.

3.3 Cluster analysis and incremental spatial autocorrelation of the COVID-19

The high and low value of cluster analysis indicates the pattern of COVID-19 distribution with respect to distance. The result of G_i^* statistics showed that every factor

considered for hotspot analysis of the COVID-19 is highly clustered (Table 2). In this table, the p value highlights the possibility of the observed spatial distribution of points generated by a random process. The low p value means it is less possibility that observed value is the result of random processes. Contrary, high value shows more chances of random processes. A higher value of z -score and lower p value indicates the best result and highly cluster. The values of z -score related to G_i^* statistics revealed that district-wise total population and reported COVID-19 cases have high z -score. Thus, for analysing hotspot, these two factors were given more priority.

3.4 Hotspot areas

Hotspot analysis helps in identifying statistically significant locations of an incidence and mapping the areas of hotspots and cold spots with the help of z -score. A minus value of z -score indicates cold spots, while plus value means hotspot. In the present study, there were found some locations with high spatial clusters that can be considered

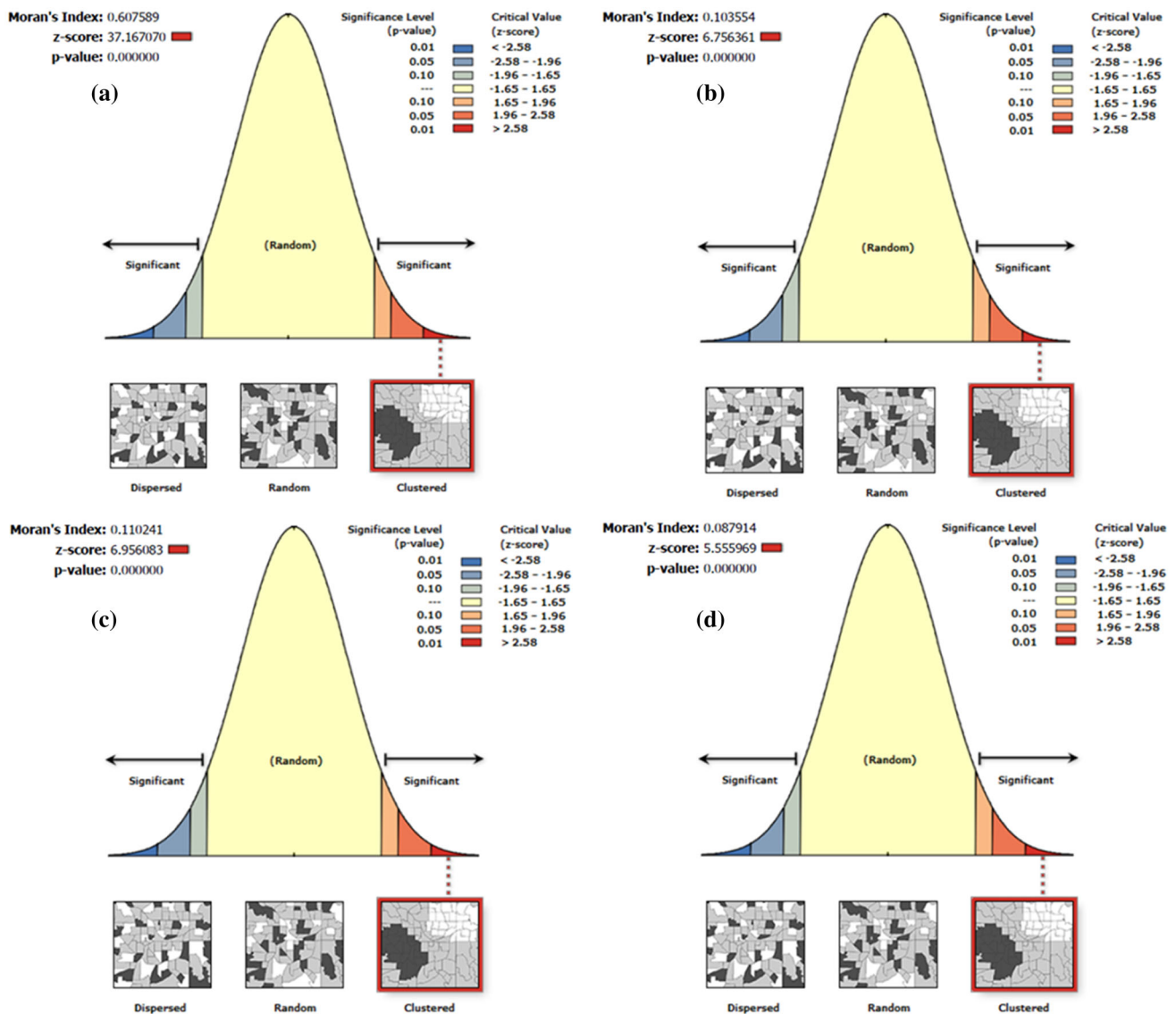


Fig. 10 Graphical presentation of the result of spatial autocorrelation **a** total population, **b** population density, **c** foreign tourist arrivals and **d** confirmed cases of COVID-19

for the hotspot of COVID-19 in India. In Fig. 11, the red and green spots indicate statistically significant hotspot and cold spot respectively. The figure shows spatial locations of novel coronavirus that spread over maximum districts of India. Among four factors, distribution of total population and the COVID-19 cases have got more priority based on z -score of Getis-Ord G_i^* statistic. Based on the total population of respective districts of each state, the map shows that Maharashtra, Andhra Pradesh, Uttar Pradesh, West Bengal, and Rajasthan state have more hotspot with a high value of z -score (> 3). Based on district-wise reported COVID-19 cases, the result portrayed that excepting the districts of the north-eastern state of India and Union territory (UT) like Puducherry, Andaman & Nicobar Islands, Lakshadweep, Daman Diu, and Dadra & Nagar Haveli,

more or fewer districts of all other states and UTs are hotspot to the COVID-19 with a high value of z -score (> 0.8745). While, two other factors i.e. population density and foreign tourist arrivals to India also produce statistically significant hotspot with a positive value of z -score as > 0.7836 and > 0.7698 respectively, but these were least considered due to lower z -score of Getis-Ord G_i^* statistic.

4 Discussion

Coronavirus is a newly discovered virus and infectious disease that rapidly spreading in every corner of the world and causes more than 1,220,500 deaths globally. Presently no treatment or vaccine invented for this fatal disease. The

Fig. 11 The COVID-19 hotspot map of India **a** based on district-wise total population; **b** based on based on district-wise population density; **c** based on percentage of foreign tourists arrivals to India; **d** based on district-wise confirmed cases of the COVID-19

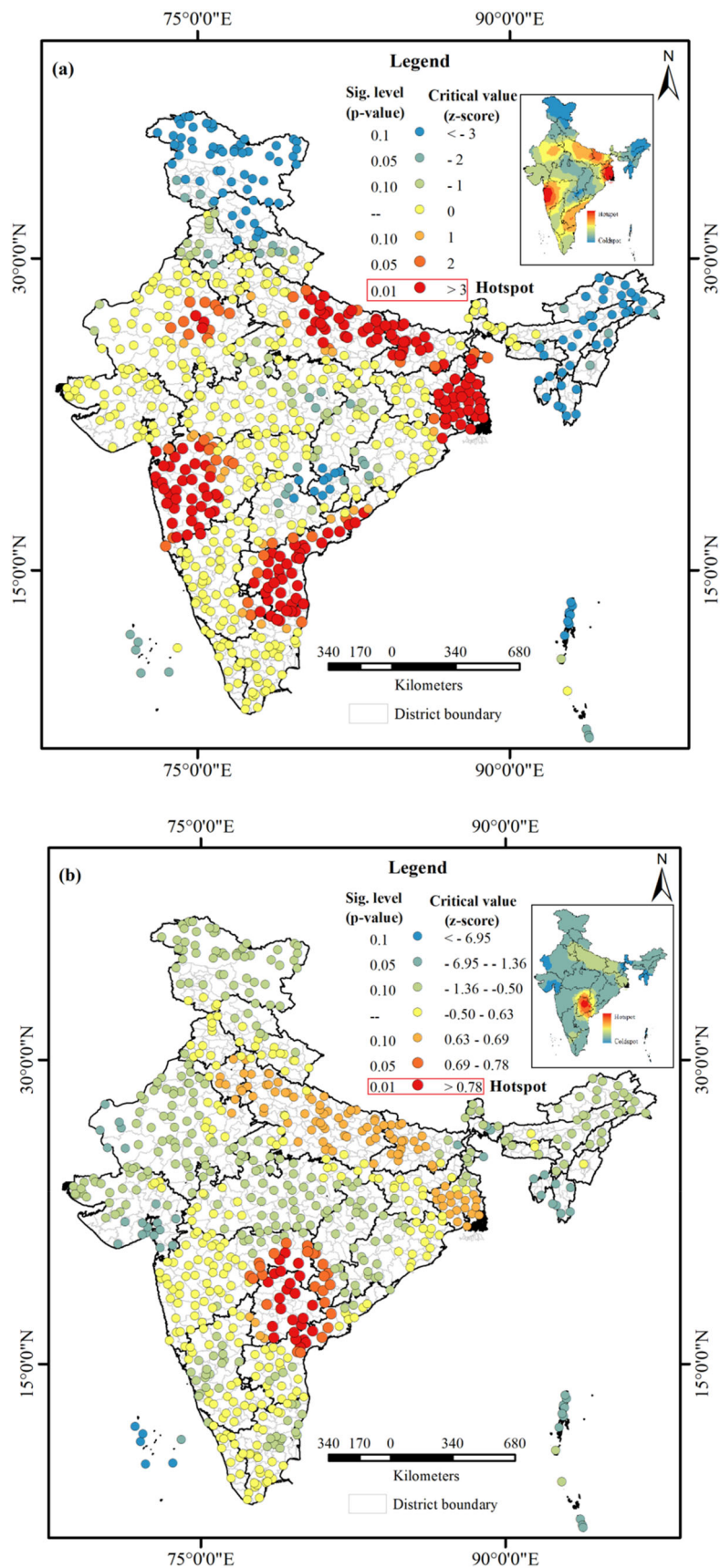
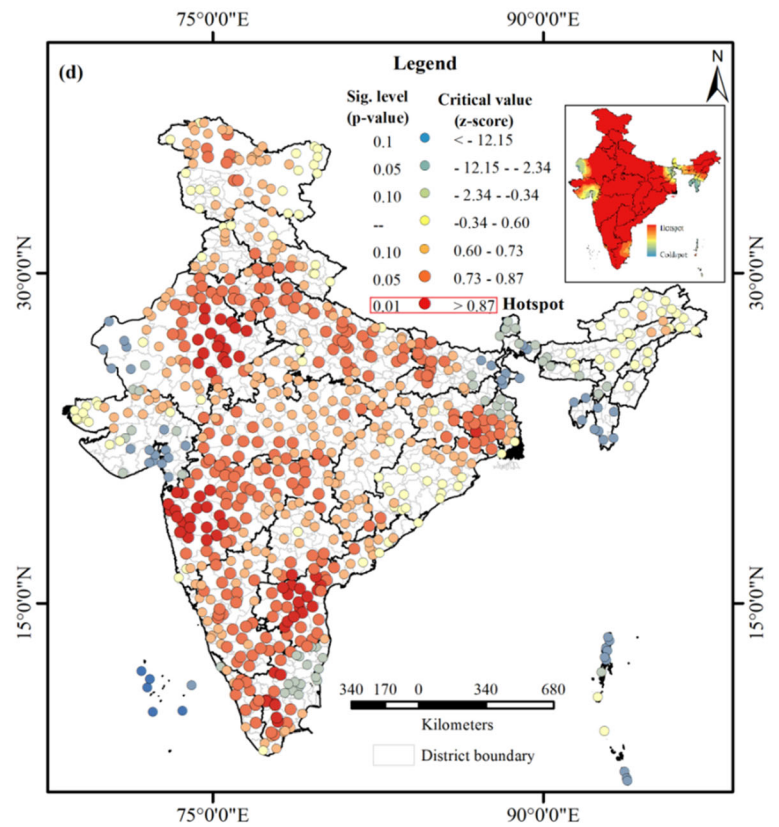
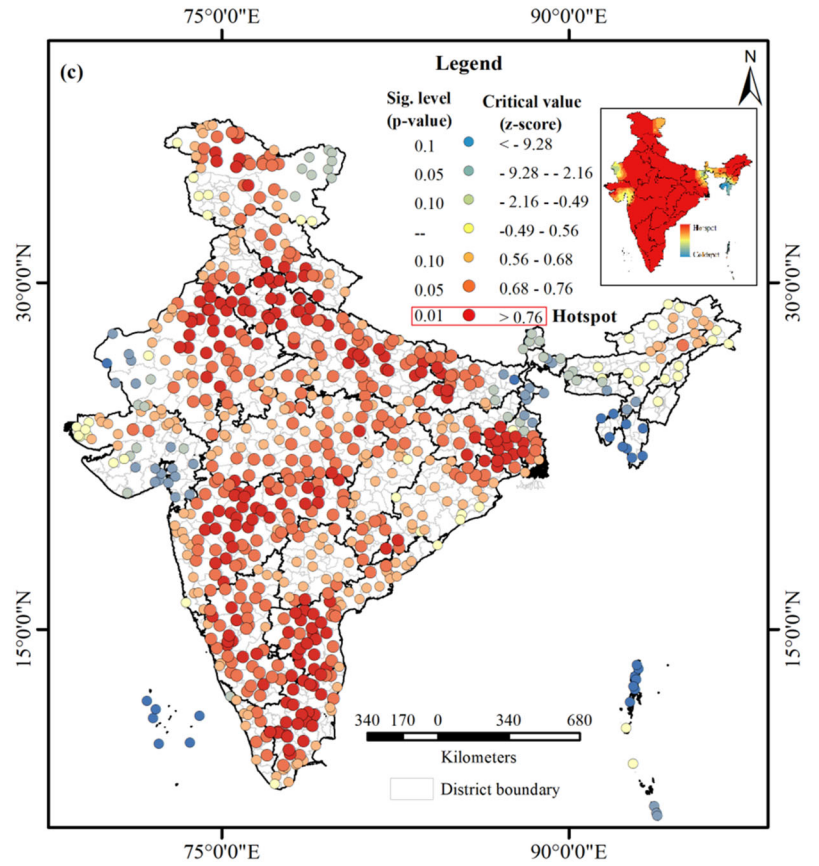


Fig. 11 continued



only solution to protect individual is self-quarantined, social distancing, stay home, and avoid social gathering. In this respect, hotspot detection will play an important role to identify risk-prone regions and practice health safety strategies. Many countries like India have been trying to detect hotspots of the COVID-19 for taking reckonable actions, but these are based on reported confirmed and positive cases. However, the present study used geospatial and geostatistical technique and other relevant factors instead of only the COVID-19 cases to identify the most significant factor which can be considered for detecting hotspots of novel coronavirus in India. Currently, many studies have been continuously focused on different aspects of novel coronavirus throughout the world, like non-respiratory symptoms of the COVID-19 patients [17], control and special attention towards the COVID-19 [18], predicting the possible domestic and international spread of 2019-nCoV outbreak [1], modelling and forecasting of the 2019 coronavirus outbreak [19], geographic information system (GIS) and big data technologies in tracking, predicting, and effective assessment of COVID-19 [20], risk map of COVID-19 infection in mainland Portugal using an ensemble of geostatistical realizations [21]. But in their study the application to geospatial technique and spatial analysis of the COVID-19 have been not found. However, this study is an attempt to apply the geospatial technique and spatial statistics to examine and detect the hotspot of novel coronavirus pandemic in India. The result of other studies on different aspects had also shown the applicability of geospatial analysis and its accuracy for decision making [22–28].

The present work has some strength in this worse situation of the COVID-19 crisis, like (1) this is a GIS-based study on novel coronavirus in India, which could offer a foundation for further geospatial analysis and risk factor determination of this deadly pandemic, (2) the findings of this study reveal that geospatial mapping and spatial statistics could be applied for rapid display and produce maps by highlighting spatial and temporal variation of risk-prone areas, (3) the results of the present study may be helpful for a quick presentation of the COVID-19 hotspot which further help for preparing effective control and preventive strategies, (4) the COVID-19 hotspot is shown in this study are distributed at smallest geographical areas of India and which may be taken for emergency planning in those districts come under hotspot zone.

The present study showed that geospatial and statistical analysis plays a significant role in hotspot detection and analysis of the COVID-19 pandemic. We used Moran's I and G_i^* statistics revealed that the spatial structure of the COVID-19 was statistically significant and highly

clustered, and detected hotspot throughout the country. This proved that spatial autocorrelation is a significant tool for studying the spatial structure of an incidence. Our analysis suggested that district-wise total population is more responsible for high numbering and out breaking of the COVID-19 positive cases in the Indian subcontinent, because of the high possibility of crowding and the social contract. Another factor like a large number of confirmed cases also revealed the possibility of more hotspot generation in the future.

5 Conclusions

In this worse situation of a health crisis due to the COVID-19 pandemic, more researches are required to analyse risk factor and spatial location. This initiative will support in decreasing the doubling rate of confirmed cases. The present study highlighted alike objective. A very high populous country like India needs to prepare an action plan immediately to react, respond and recover the affected groups and make strategies for vulnerable areas. The hotspot analysis of the present study can be helpful to reach this target that which location of state or district should give more attention, which locations are more effected and which locations highly are more vulnerable in near future. Hotspot analysis is a technique through which potential diseases outbreak areas can be identified based on responsible factors. Looking towards the nature of the COVID-19, district-wise total population, population density, reported confirmed COVID-19 cases, and foreign tourist arrivals in India were taken as a responsible factor for detecting hotspot areas. The geospatial analysis and spatial statistics suggested that these all factors are spatially clustered (p value < 0.01). The hotspot generated using Getis-Ord G_i^* statistic also revealed that the total population and reported COVID-19 cases produced more accurate hotspot as they produced higher z -score value (z -score > 3 and > 0.7295 respectively). Thus, the resultant hotspot generated using these two factors will be considered as the COVID-19 hotspot for India. The spatial locations of the hotspot in our study will help the Government, district administrative, local authorities, as well as residents of hotspot location for taking actions and precautions against this novel infectious disease.

Acknowledgements We thankfully acknowledge the anonymous reviewers and the Editor in Chief for their valuable time, productive comments and suggestions for improving the overall quality of our manuscript.

Funding No fund was received from anywhere.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author declares that there is no conflict of interest.

References

- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395, 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
- WHO. (2020a). WHO Director-General's opening remarks at the media briefing on COVID-19. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020>. Retrieved July 10, 2020.
- Johns Hopkins. (2020). CSSE coronavirus COVID-19 global cases (dashboard). <https://coronavirus.jhu.edu/map.html>. Retrieved July 24, 2020.
- WHO. (2020b). Novel coronavirus (2019-nCoV), situation report-1. https://www.who.int/docs/default-source/wrindia/india-situation-report1.pdf?sfvrsn=5ca2a672_0. Retrieved July 15, 2020.
- Ministry of Health and Family Welfare. (2020). India ramps up efforts to contain the spread of novel coronavirus. <https://www.who.int/india/emergencies/novel-coronavirus-2019>. Retrieved July 25, 2020.
- BBC. (2020). India coronavirus: All major cities named Covid-19 'red zone' hotspots. <https://www.bbc.com/news/world-asia-india-52306225>. Retrieved July 14, 2020.
- Boulos, M. N., & Geraghty, E. M. (2020). Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal of Health Geographics*, 19, 8.
- Kasthuri, A. (2018). Challenges to Healthcare in India—*The Five A's*. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive and Social Medicine*, 43(3), 141–143. https://doi.org/10.4103/ijcm.IJCM_194_18.
- SRS Statistical Report. (2016). *Office of the registrar general & census commissioner*. New Delhi: Ministry of Home Affairs, Government of India.
- Florida, R. (2020). The geography of coronavirus. CITYLAB, 3rd April. <https://www.citylab.com/equity/2020/04/coronavirus-spread-map-city-urban-density-suburbs-rural-data/609394/>.
- Pawar, Y. (2020). Maharashtra's population density, international travellers led to its COVID-19 crisis. *The CARAVAN*, 28 March. <https://caravanmagazine.in/health/maharashtra-population-density-international-travellers-led-to-covid-crisis>.
- Unnithan, P. S. G. (2020). Kerala reports first confirmed coronavirus case in India. INDIA TODAY, 31st January. <https://www.indiatoday.in/india/story/kerala-reports-first-confirmed-novel-coronavirus-case-in-india-1641593-2020-01-30>.
- Boots, B. N., & Getis, A. (1998). *Point pattern analysis*. Newbury Park: Sage Publications.
- Tsai, P.-J., Lin, M.-L., Chu, C.-M., & Perng, C.-H. (2009). Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health*. <https://doi.org/10.1186/1471-2458-9-464>.
- Osei, F. B., & Duker, A. A. (2008). Spatial and demographic patterns of Cholera in Ashanti region-Ghana. *International Journal of Health Geographics*, 7, 44.
- Bhunia, G. S., Kesari, S., Chatterjee, N., Kumar, V., & Das, P. (2013). Spatial and temporal variation and hotspot detection of Kala-Azar disease in Vaishali district (Bihar), India. *BMC Infectious Diseases*, 13(1), 10. <https://doi.org/10.1186/1471-2334-13-64>.
- Lu, S., Lin, J., Zhang, Z., Xiao, L., Jiang, Z., Chen, J., et al. (2020). Alert for non-respiratory symptoms of Coronavirus Disease 2019 (COVID-19) patients in epidemic period: A case report of familial cluster with three asymptomatic COVID-19 patients. *Journal of Medical Virology*. <https://doi.org/10.1002/jmv.25776>.
- Bedford, J., Enria, D., Giesecke, J., Heymann, D. L., Ihekweazu, C., Kobinger, G., et al. (2020). COVID-19: Towards controlling of a pandemic. *The Lancet*. [https://doi.org/10.1016/s0140-6736\(20\)30673-5](https://doi.org/10.1016/s0140-6736(20)30673-5).
- Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE*, 15(3), e0230405. <https://doi.org/10.1371/journal.pone.0230405>.
- Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., et al. (2020). COVID-19: Challenges to GIS with big data. *Geography and Sustainability*. <https://doi.org/10.1016/j.geosus.2020.03.005>.
- Azevedo, L., Pereira, M. J., Ribeiro, M. C., & Soares, A. (2020). Geostatistical COVID-19 infection risk maps for Portugal. *International Journal of Health Geographics*, 19(1), 1–8.
- Hamidy, N., Alipur, H., Nasab, S. N. H., Yazdani, A., & Shojaei, S. (2016). Spatial evaluation of appropriate areas to collect runoff using Analytic Hierarchy Process (AHP) and Geographical Information System (GIS) (case study: the catchment "Kasaf" in Bardaskan. *Modeling Earth Systems and Environment*, 2(4), 1–11.
- Shojaei, S., Alipur, H., Ardakani, A. H. H., Nasab, S. N. H., & Khosravi, H. (2018). Locating Astragalus hypsogeton Bunge appropriate site using AHP and GIS. *Spatial Information Research*, 26(2), 223–231.
- Forozan, G., Elmi, M. R., Talebi, A., Mokhtari, M. H., & Shojaei, S. (2020). Temporal-spatial simulation of landscape variations using combined model of markov chain and automated cell. *KN-Journal of Cartography and Geographic Information*, 70(2), 45–53.
- Ali, S. A., & Ahmad, A. (2018). Using analytic hierarchy process with GIS for Dengue risk mapping in Kolkata Municipal Corporation, West Bengal, India. *Spatial Information Research*, 26(4), 449–469. <https://doi.org/10.1007/s41324-018-0187-x>.
- Ali, S. A., & Ahmad, A. (2019). Mapping of mosquito-borne diseases in Kolkata Municipal Corporation using GIS and AHP based decision making approach. *Spatial Information Research*, 27(3), 351–372. <https://doi.org/10.1007/s41324-019-00242-8>.
- Ali, S. A., Khatun, R., Ahmad, A., & Ahmad, S. N. (2019). Application of GIS-based analytic hierarchy process and frequency ratio model to flood vulnerable mapping and risk area estimation at Sundarban region, India. *Modeling Earth Systems and Environment*, 10, 10. <https://doi.org/10.1007/s40808-019-00593-z>.
- Parvin, F., Ali, S. A., Hashmi, S. N. I., & Khatoun, A. (2020). Accessibility and site suitability for healthcare services using GIS-based hybrid decision-making approach: A study in Murshidabad, India. *Spatial Information Research*. <https://doi.org/10.1007/s41324-020-00330-0>.

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