Spatio-temporal changes pattern in the hotspot’s footprint: a case study of confirmed, recovered and deceased cases of Covid-19 in India

Mohd Shamsh Tabarej1 · Sonajharia Minz1

Received: 8 October 2021 / Revised: 24 February 2022 / Accepted: 27 February 2022 / Published online: 14 May 2022
© The Author(s), under exclusive licence to Korean Spatial Information Society 2022

Abstract Hotspot detection and the analysis for the hotspot’s footprint recently gained more attention due to the pandemic caused by the coronavirus. Different countries face the effect of the virus differently. In India, very little research has been done to find the virus transmission. The paper’s main objective is to find changing pattern of the footprint of the hotspot. The confirmed, recovered, and deceased cases of the Covid-19 from April 2020 to Jan 2021 is chosen for the analysis. The study found a sudden change in the hotspot district and a similar change in the footprint from August. Change pattern of the hotspot’s footprint will show that October is the most dangerous month for the first wave of the Corona. This type of study is helpful for the health department to understand the behavior of the virus during the pandemic. To find the presence of the clustering pattern in the dataset, we use Global Moran’s I. A value of Global Moran’s I greater than zero shows the clustering in the data set. Dataset is temporal, and for each type of case, the value Global Moran’s I > 0, shows the presence of clustering. Local Moran’s I find the location of cluster i.e., the hotspot. The dataset is granulated at the district level. A district with a high Local Moran’s I surrounded by a high Local Moran’s I value is considered the hotspot. Monte Carlo simulation with 999 simulations is taken to find the statistical significance. So, for the 99% significance level, the p-value is taken as 0.001. A hotspot that satisfies the p-value threshold is considered the statistically significant hotspot. The footprint of the hotspot is found from the coverage of the hotspot. Finally, a change vector is defined that finds the pattern of change in the time series of the hotspot’s footprint.

Keywords Hotspot · Moran’s I · Footprint · Covid-19 · Change analysis · Global autocorrelation · Local autocorrelation · Monte Carlo simulation

1 Introduction

Numerous geographical areas have seen various diseases caused by pathogens such as the Ebola virus, Zika virus, Nipah virus, and Coronavirus (CoVs) in the last decade [1]. A new type of virus has emerged in Wuhan, Hubei Province, China, whose genome sequence does not match any of the previously reported CoVs. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is another name for the novel Coronavirus strain (2019-nCoV). The new Coronavirus (SARS-CoV-2) is thought to have originated in bats and then spread from humans to humans. Still, the possibility of another source cannot be ruled out [1]; current suspicions include pangolins and snakes [2]. The first case was discovered on December 8, 2019, and early death was recorded on January 9, 2020 [1]. On the same day, the World Health Organization announced that a new strain of Coronavirus had been discovered, which was rapidly spreading [3].

Coronavirus (CoVs) is an enveloped, positive-sense, single-stranded RNA virus with a wide range of mutations. Its surface contains a spike-like structure that gives it a crown-like look, therefore the name coronavirus [4]. The virus is divided into four types: Alpha, Beta, Gamma, and
has been produced and is already in use following a success-

develop a vaccination to combat it. A COVID-19 vaccine
research is ongoing to assess the potential for damage and
a slew of investigations into the virus’s features. A slew of
death rate of 1.42 percent.

2020. As of January 31, 2021, there were 10,758,551 con-
borders, with the first shutdown beginning on March 23,

people in jeopardy. India soon shuts down its international
would not be able to endure, putting the lives of millions of
the elderly and those who are already afflicted with diseases
ful study. Many studies have indicated that it is harmful to

Delta [4–6]. There have been two occasions in the last two
decades in which animal-to-animal beta-gene crossover has
resulted in various human disorders. The first such incidence
occurred in China’s Guangdong province in 2002–2003,
when a novel coronavirus that originated in bats crossed
over to people via palm civet cats as an intermediate host.
The virus is known as the severe acute respiratory syndrome
coronavirus (SARS-CoV), and it has infected 8422 people
and killed 916 of them, mostly in Hong Kong and China [2].
Another strain, Middle East respiratory syndrome coronavi-
irus (MERS-CoV), was discovered in Saudi Arabia in 2012,
over a decade later. It was again of bat origin, with camel as
the intermediate host. It affects 2499 people and kills 858 of
them, with a fatality rate of 34% [2]. SARS-CoV-2, the novel
virus, is indistinguishable from previous respiratory
diseases. In certain patients, the infection can proceed to
pneumonia, respiratory failure, and even death after just in
one week. The overall fatality rate is estimated to be between
2 and 3 percent [2].

The Chinese authorities did everything they could to
stop the virus from spreading. Airports, railways, highways,
public transportation, public gatherings, stores, and mass
activities that result in a social grouping are all put on hold
immediately. After all, the viral spread cannot be stopped by
these preventative measures. The virus spread from China
to the rest of the world, with the first corona case outside of
China being reported in Thailand on January 13, 2020 [7].
The World Health Organization (WHO) declares the illness
condition an emergency on January 31, 2020, and a global
pandemic on March 11, 2020 [3]. The virus began with a
single person in China and has since spread to clusters, com-
unities, and practically every country on the planet, with
the European continent being the most impacted region [8].

India is the world’s second-most populous country of the
World. Its population is 1.3 billion people. The proportion
of the population is dispersed over the many states, each with
its cultural values. The World Economic Forum ranks the
country 150th in the world in terms of health care. On Janu-
ary 31, 2021, the first instance of COVID-19 was discovered
[9]. Because of its grave problems, it was expected that India
would not be able to endure, putting the lives of millions of
people in jeopardy. India soon shuts down its international
borders, with the first shutdown beginning on March 23,
2020. As of January 31, 2021, there were 10,758,551 con-

confirmed cases, with a recovery rate of 96.98 percent and a
death rate of 1.42 percent.

The hazardous effect of the Coronavirus have prompted
a slew of investigations into the virus’s features. A slew of
research is ongoing to assess the potential for damage and
develop a vaccination to combat it. A COVID-19 vaccine
has been produced and is already in use following a success-
ful study. Many studies have indicated that it is harmful to
the elderly and those who are already afflicted with diseases
such as lung disease, heart attack, and so on. The authority
faces a significant task in monitoring the virus to ensure that
it causes the least amount of harm [10].

Numerous studies have been carried out to determine the
virus’s possible impact. The most notable studies are stoch-
astic simulation [11], exponential growth [12], Weibull
distribution [13], log-normal distribution [3], and others
[14]. The study was able to identify the average incubation
period as well as a 14-day total quarantine period to evaluate
the virus within the human. Various studies have been
carried out in China and other countries; very little work has
been done in India to identify the Covid-19 dissemination
pattern.

Further investigations are done on the study of the iden-
tification of vulnerable area have included hotspot analysis
[15], entropy-based hotspot detection [16], forest fire
hotspot detection [17], risk assessment related to climate
change [18], and identification of high-risk malaria areas
[19]. Cheshmida et al. [20] used the Dolphi technique to
prioritize the severity of floods in Iran’s Ivar watershed. They
analyzed using six criteria and discovered that the Dolphi
approach is useful for determining the intensity of
flooding. Shojaei et al. [21] finds the appropriate location for
the Astragalus hypsogeton Bunge. The relative importance of
each criterion was determined using an analytical hier-
archy procedure, and the cases for each were calculated. To
locate the proper site, ArcGis is utilized to map the collected
results. The method examines the footprint of the hotspot as
well as the pattern of the footprint. Issa et al. [22] have done
a spatial assessment of the suitable are for the medicinal spe-
cies of Astragalus. Forozan et al. [23] identified land cover
changes by applying CA–Markov model for simulation. Sat-
ellite photos of Yazd city were taken in the years 2000, 005,
2010, and 2016. Researchers next utilized a support vector
machine to classify the land that was used. Then there are
changes in land cover over time. They discovered that the
residential area has grown while the dry land and vegeta-
tion area has shrunk. They also forecasted land use develop-
ment for the year 2040 and discovered a rise in the amount
of land utilized. Huang et al. [24] proposes a method that
automatically detects statistically significant clusters in the
point data. First, they use the Voronoi diagram to detect the
high-density point and then use density-based clustering to
find the hotspot. Shiode et al. [25] uses network-based scan
statistics to find the hotspot’s exact location and the extent of
the hotspot. Mondal et al. [26] uses a geostatistical technique
to find the crime hotspot.

Barboza et al. [27] worked on spatiotemporal analysis. In
the city of Los Angeles, they employed negative binomial
regression to detect a link between social distancing pro-
tocol and child abuse and neglect (CAN). During the pan-
demic, spatial–temporal analysis pinpoints the location of
the growing hotspot and cold spot. The relationship between
neighborhood structural factors and hotspots and cold spots was investigated. The CAN report reveals a significant drop in the Covid-19 period. Abulibdeh [28] examined water and electricity use in Doha, Qatar during the Covid-19 epidemic. The hotspot and cold spot of water and electricity consumptions were determined using Moran’s I and Getis-Ord G*.

According to the findings, there is a difference in water and electricity consumption at block level over time. Aral et al. [29] finds the spatiotemporal pattern of the Covid-19 in Turkey. They also employ spatial regression to uncover the relevant factors affecting the Covid-19 cases. In the Covid-19 model, they discovered that population density and the elderly dependency ratio are essential. A Covid-19 distribution pattern based on Hotspot and spacetime cube is given by Purwanto et al. [30]. Paul et al. [31] studied the spatiotemporal behavior of the spread of the Covid-19 from urban to rural area. Spatiotemporal analysis of the Covid-19 is given in [32–34]. Mostly the methods discussed in the literature considered the number of cases as well as the population density for the detection of the hotspot. Then the temporal analysis is considered on the detected hotspot.

As a result, the goal of this research is to determine the footprint of the COVID-19 hotspot, as well as the temporal pattern of the detected hotspot’s footprint. This study adds to our understanding of the virus’s transmission and temporal pattern. It will also help people understand and put preventative measures in place in the event of future pandemics. The paper’s main contributions are determining the hotspot of the Covid-19 cases district by district, determining the footprint of the hotspot, and finally determining the changes pattern of the footprint.

The document is divided into five sections, the first of which comprises the above-mentioned introduction. The Study Area and Methodology are covered in the second part. The third section is devoted to the results of the findings. The conclusion and Recommendation of the research is found in section four.

2 Materials and methods

The methodology for the change pattern of the hotspot’s footprint is shown in Fig. 10 in the supplementary section. The first step of the algorithm is the pre-processing. In this phase firstly the granulation of the study area is done at the district level. A polygon representing the district boundary is taken as the lower-level granule. The data of the covid-19 cases for each district is taken and mapped with the district polygon. After mapping the two we will get the database for the covid-19 for the analysis. The dataset is the temporal so, we have granulated the dataset for each month in the specified range for the temporal analysis. Following the creation of the data the next step is to check for the presence of the spatial dependance in the data set. The dataset having the spatial dependance is considered for the detection of the hotspot. Next the hotspot is detected and then the footprint of the hotspot is found. Temporal pattern of the footprint is found which gives the change pattern of the footprint.

2.1 Study area

India is the world’s second-most populous country. It lies between 08°04’0 N and 37°06’0 N, and 68°7’0 E and 97°25’0 E in South Asia. India is organized into 36 entities, including 28 states and 8 union territories. Districts are created inside each state and union territory. There are 718 districts in all. The population of the country is dispersed over many states, health inequalities, social diversity, and expanding economic and cultural values. The international economic forum ranks the country 150th in the world for health care, and the World Bank ranks the country 112th out of 154 nations for household income. There are numerous health complications as a result of all of these conditions [15]. The rapid rise of the human population has resulted in a slew of problems, including the spread of many airborne and infectious diseases. Several diseases, including malaria, cholera, influenza, dengue fever, and tuberculosis, were once common in India and afflicted a vast population [15]. Aside from these vital circumstances, the first instance of COVID-19 in India was discovered on January 31, 2021 [9]. Because of its catastrophic problems, it was expected that India would not be able to endure, putting the lives of millions of people in jeopardy.

2.2 Data source

This dataset for the study was gathered from the site Covid-19 India API [35]. The data is temporal and includes information on COVID-19 cases in India from April 24, 2020, to January 31, 2021. The dataset includes information on the overall number of confirmed, recovered, deceased patients, and the test done on a given day. The raw temporal data is delivered on a daily basis. For spatial processing, the dataset is aggregated monthly. The administrative boundaries of each district, as well as additional factors such as district area, population, and so on, are included in this collection of data and can be found at DIVA-GIS [36].

2.3 Preprocessing and data cleaning

Geographical processing necessitates the use of spatial data such as geographic coordinates, administrative borders, and so on. District coordinates (latitude, longitude) are included in the dataset since data is aggregated at the district level. The dataset contains information about railway quarantine, airport quarantine, foreign evacuees, Italians, BSF camp,
and other states people who were quarantined and identified at these places but their respective district was not identified. These people have been removed and are not considered for further analysis. Using python and its Geospatial package geopandas for processing of spatial data, the two data sets stated above were combined. Because the data is temporal, a new data file is produced for each month. The study area is depicted in Fig. 1 along with the district center’s coordinates. Each coordinate is linked to several instances of the corona case.

### 2.4 Global spatial autocorrelation

Spatial data is autocorrelated in space [37], in contrast to the often assumed independently identically distributed (I.I.D.) data. The analysis of geographical data is more difficult than I.I.D. because it follows the first rule of geography, which states that "everything is related to everything else, but nearby things are more related than distant things". The global spatial autocorrelation [38] technique is utilized to identify the spatial dependency among the data, and the pysal module from Python is used for the processing. Equation (1) defines the Global Moran’ I index as a typical measure of spatial autocorrelation in the data.

\[ I = \frac{N}{S_0} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \]  

(1)

where, \( N \) is the number of observations, \( z_i \) is the standardized value of the interest, \( w_{ij} \) is the value of the weight matrix \( W \) corresponding to the \( i \)th row and \( j \)th column.

\[ S_0 = \sum_i \sum_j w_{ij} \]  

(2)

The \( Z_i \) score for the statics is given by:

\[ Z_i = \frac{I - E[I]}{\sqrt{V[I]}} \]  

(3)

where \( E[I] \) is the expected value of \( I \) and is given as \( E[I] = -\frac{1}{(n-1)} \) and \( V[I] \) is the variance of \( I \) and is given as

---

**Fig. 1** District coordinate of all Covid-19 affected district
Spatio-temporal changes pattern in the hotspot’s footprint: a case study of confirmed, recovered…

\[ V[I] = E[I^2] - E[I]^2 \]. The value of Moran’s I index ranges between -1 to +1. A positive value indicates spatial dependence i.e., clustering of high density and a negative value indicates a dispersion i.e., a clustering of dissimilar values, and a value zero indicates spatial randomness.

2.5 Hypothesis test

The null hypothesis \((H_0)\) in global spatial autocorrelation says that data points are distributed randomly throughout the study region according to a random Poisson process. The alternative hypothesis \((H_1)\) asserts that the data points are not distributed randomly, and that data is concentrated at some places in the study region. The significance of the data set is estimated using the test statistics Global Moran’s I and the distribution of the test statistics obtained by Monte Carlo Simulation. The significance test determines whether to accept or reject the null hypothesis for a certain level of significance.

2.6 Hotspot detection

The presence of spatial autocorrelation, i.e., the presence of clustering of the spatial phenomena that is a hotspot \([39, 40]\) or cold spot, is shown by the global spatial autocorrelation, but it does not reveal the location of the cluster. Local spatial autocorrelation is employed for this purpose. The relation between each observation and its surroundings is defined by local spatial autocorrelation. Local Moran’s I \([36]\) is one of the local measurements of spatial autocorrelation. Equation \((4)\) defines the local Moran’s I:

\[
I_i = \frac{n}{(n-1)\sigma^2} (\eta_i - \bar{\eta}) \sum_j w_{ij} (\eta_j - \bar{\eta})
\]

where \(\eta_i\) represent the number of Covid-19 cases with units \(i\), \(\bar{\eta}\) denotes the average number of cases over the study region, \(w_{ij}\) is the spatial weight between the spatial unit \(i\) and the neighboring unit \(j\), \(n\) represent the total number of spatial units in the dataset, and \(\sigma^2\) denotes the variance of the observed cases.

The Moran’s I plot is used to comprehend the foundations of local spatial autocorrelation. This is done by first determining the spatial lag of the variable of interest. We have chosen the confirmed cases of Covid-19 for January 2021 for the graphical representation. The graph is plotted using the standardized values of the parameters, i.e., confirmed cases, and the spatial lag. The plot of confirmed cases vs Spatial Lag of the confirmed cases is shown in the Fig. 2. The plot is partitioned on the basis of mean of the confirmed cases and the Spatial Lag.

Therefore the map is divided into four quadrants I, II, III, and IV. These four quadrants are named as High–high (HH), Low–high (LH), Low–low (LL), and High–low (HL). The quadrant HH represents a high value of confirmed cases and a high Spatial Lag. As Spatial Lag represents the spatial neighbors of a spatial entity. A high Spatial Lag means high value of the spatial neighbors. So, the quadrants HH represents the values corresponding to the Hotspot. Similarly, LL represents low value surrounded by low neighbors and it is corresponding to a cold spot. In the same manner LH, and HL represent doughnuts and diamond respectively.

Local Moran’s I is determined as specified in Eq. (4) for the identification of the covid-19 hotspot. The choropleths plot of the Moran’s I value is presented in Fig. 3a. The quadrant plot of the covid-19 dataset is shown in Fig. 3b based on the value of the quadrant analysis as mentioned in Fig. 2. The Monte Carlo Simulation is now used to determine the statistical significance of the values in the quadrants. The district that passes the statistical test i.e., has a p-value greater than 0.05, is regarded as significant. Figure 3c depicts a plot of the significant district. Finally, the district in the first quadrants that passes significant test is regarded as the hotspot. The plot of the hotspot district is highlighted in red color is shown in the Fig. 3d.

2.7 Information of the hotspot

After the detection of the hotspot as described in Sect. 2.6 for the hotspot detection. The measurements form the hotspot are taken. The measurements are used for the analysis of the hotspot. The measurements that are taken for the analysis of the hotspot are the number of hotspots i.e., the number of districts that are qualified as the hotspot, the number of
cases that are found in the hotspot i.e., that is the number of confirmed, deceased and recovered cases.

2.8 The footprint of the hotspot

For a given phenomenon $\Phi$ such as confirmed, deceased and recovered cases of Covid-19 the footprint is defined as the geographical coverage. Therefore, the footprint \( \xi_H \) of the hotspot is defined as the geographical coverage of the hotspot. Let the hotspot of the phenomenon $\Phi$ is $H$ then the footprint $\xi_H$ of the hotspot is defined in Eq. (5):

$$\xi_H = area(H)$$

(5)

where, \( area(H) \) is the total area of the hotspot district.

Let the study region $S$ is divided into $n$ number of sub regions $\alpha_1, \alpha_2, \ldots, \alpha_n$ and let out of total $n$ sub-region $m$ number of sub-regions compose the hotspot. Then the total area covered by the hotspot is described by Eq. (6):

\[ \text{Total Area} = m \times area(H) \]
\[ \xi_H = \sum_{i=1}^{m} a_i, \forall \alpha_i \in H \]  

That is the area of the hotspot is the sum of the area of all sub-regions belonging to the hotspot. For the analysis of the footprint, the footprint is taken itself. In our case footprint of the hotspot is taken in square km (km²).

2.9 Spatio-temporal analysis of the hotspot’s footprint

Spatio-temporal change footprint pattern identification is the process of determining the location and/or time of such changes in data, given a description of change and a dataset concerning a spatiotemporal phenomenon [41]. Change can be defined in many ways like change in the statistical parameter, change in the actual value, change in the model fitted to the data, and change in the derived attributes. Depending on the definition of change, data model used and the time, spatiotemporal change footprint is defined differently. For example, space can be represented by point, line, polygon and a combination of these i.e., spatial network. The time could be single snapshot, set of snapshots, point in time series, and interval in time series. In our case we are using the polygon to represent a district (a data point) and time is point in time series. So, the Spatiotemporal change footprint for any phenomenon Φ is the process of identifying the area where the phenomenon Φ is either growing or shrinking. The Spatio-temporal analysis is applied to many other areas, including biology, ecology, metrology, medicine, transportation, and forestry. To investigate the temporal changes in the footprint of the hotspot a new algorithm that represent the change in form of bit has been developed. The proposed algorithm has several advantages in the first it is used to find the common structure between the different object. For example, for the footprint of hotspot of confirmed and the recovered cases of corona. This algorithm is able to find the common structure i.e., an increase in the footprint of the two are not if the number of the cases in the hotspot is increasing and footprint in not increasing then there is increase in the density of the hotspot.

The dataset is temporal and aggregated monthly. The hotspot identified in the dataset is represented as \( H_n \). The area covered by the hotspot is called the footprint of the hotspot and is denoted by \( \xi \).

1. Let’s suppose at time \( t_i \) dataset \( D_i \) contains the \( H_{ni} \) the number of the hotspot. Let the footprint of each hotspot is \( \xi_{i1}, \xi_{i2}, \ldots, \xi_{im} \) where, \( \xi_{i1} \) represent the footprint of the hotspot at timestamp \( t_i \).
2. Let the footprint of the hotspot at timestamp \( t_{i+1} \) is denoted by \( \xi_{(i+1)1}, \xi_{(i+1)2}, \ldots, \xi_{(i+1)m} \).
3. The difference vector is defined as:

\[ \sigma_i = \xi_{(i+1)n} - \xi_{in} \]  

4. Now we define the change vector \( \lambda \) such that:

\[ \lambda_i = \begin{cases} 1, & \text{if } \sigma_i > 0 \\ 0, & \text{if } \sigma_i = 0 \\ -1, & \text{otherwise} \end{cases} \]  

If the value of \( \lambda \) is 1 then there is an increase in the footprint, if it is −1 then there is a decrease in the footprint, if the value is 0 then there is no change in the footprint of the hotspot.

Example: suppose the footprint of the hotspot for the time interval 01-04-2020, 01-05-2020, 01-06-2020, 01-07-2020, 01-08-2020, 01-09-2020 the footprint is taken itself. In our case footprint of the hotspot is taken in square km (km²).

3 Results

3.1 Analysis of global Moran’s I

The COVID-19 hotspot analysis in India is based on the number of confirmed cases, deceased patients, and recovered cases between April 1, 2019, and January 30, 2020. Before detection of the hotspot, we need to identify the spatial dependence in the dataset i.e., the presence of the clustering in the dataset. The method for the detection of the spatial dependence in the dataset is described in Sect. 0. The result of the Global Spatial Autocorrelation is presented in Table 1. This table displays the Global Moran’s I index, as well as the z-score and p-value produced from Monte Carlo simulation. The number of simulations for the Monte Carlo simulation is set to be 999 runs. Therefore, for the 99% significance level the pre-defined p-value is 0.001 and the z-score is between −2.58 and +2.58. We can see from the table that all of the Global Moran’s I index values are positive, suggesting the presence of spatial autocorrelation. This means the data was clustered at certain points, resulting in a non-uniform distribution. The datasets that meet the aforementioned criteria are evaluated for spatial dependence. The result of the Global Moran’s I shows that all the temporal dataset for each moth shows the clustering pattern.
3.2 Hotspot detection of the COVID-19

The hotspot detection of the Covid-19 data set is done as per the description of the hotspot detection described in Sect. 0. The Fig. 4 shows the hotspot of the Confirmed, Recovered, and the Deceased patient on January 2021 in the red color. The red-colored area also depicts the footprint of the hotspot area. Similarly, the hotspot of each month was found and measurements for the hotspot as well as footprints are recorded, but the plot for each month is not shown here. The measurements as described in Sect. 0 were taken. These measures include the number of cases in the hotspot and the total number of districts in the hotspot. They are used for the analysis of the hotspot as well as footprints.

3.3 The footprint of the hotspot

The hotspot’s footprint is defined as the area covered by the hotspot as stated in Eq. (5). As a result, the footprint of

| Table 1 Autocorrelation analysis |
|----------------------------------|
| Data   | Factor   | Expected | Variance | z-score | Global Moran’s I | p-value | Pattern |
|--------|----------|----------|----------|---------|------------------|---------|---------|
| Apr-2020 | Confirmed  | -0.0008  | 0.000531 | 1.6127  | 0.1364           | 0.001   | Clustered |
|         | Recovered | -0.0027  | 0.000497 | 1.917   | 0.14             | 0.001   | Clustered |
|         | Deceased  | -0.0031  | 0.000539 | 2.5454  | 0.156            | 0.001   | Clustered |
| May-2020 | Confirmed | -0.0012  | 0.000408 | 1.8493  | 0.1361           | 0.001   | Clustered |
|         | Recovered | -0.0019  | 0.000416 | 1.54    | 0.1295           | 0.001   | Clustered |
|         | Deceased  | -0.0024  | 0.000414 | 2.506   | 0.1486           | 0.001   | Clustered |
| Jun-20  | Confirmed | -0.0018  | 0.000369 | 1.0218  | 0.1178           | 0.001   | Clustered |
|         | Recovered | -0.0038  | 0.000363 | 1.7325  | 0.1292           | 0.001   | Clustered |
|         | Deceased  | -0.0012  | 0.000389 | 2.5782  | 0.2097           | 0.001   | Clustered |
| Jul-20  | Confirmed | -0.0014  | 0.000388 | 1.3143  | 0.1245           | 0.001   | Clustered |
|         | Recovered | -0.0014  | 0.000386 | 0.2807  | 0.1142           | 0.001   | Clustered |
|         | Deceased  | -0.0016  | 0.000394 | 3.8122  | 0.174            | 0.001   | Clustered |
| Aug-20  | Confirmed | -0.0021  | 0.000371 | 8.6656  | 0.1649           | 0.001   | Clustered |
|         | Recovered | -0.0022  | 0.000365 | 4.6598  | 0.1868           | 0.001   | Clustered |
|         | Deceased  | -0.0013  | 0.000405 | 7.8198  | 0.1561           | 0.001   | Clustered |
| Sep-20  | Confirmed | -0.0022  | 0.00035  | 8.6309  | 0.1592           | 0.001   | Clustered |
|         | Recovered | -0.0006  | 0.000376 | 6.5403  | 0.1262           | 0.001   | Clustered |
|         | Deceased  | -0.0008  | 0.000395 | 13.7366 | 0.2721           | 0.001   | Clustered |
| Oct-20  | Confirmed | -0.0023  | 0.000397 | 12.6236 | 0.2493           | 0.001   | Clustered |
|         | Recovered | -0.0013  | 0.000401 | 10.503  | 0.2091           | 0.001   | Clustered |
|         | Deceased  | -0.002   | 0.000389 | 17.1654 | 0.3364           | 0.001   | Clustered |
| Nov-20  | Confirmed | -0.0016  | 0.000368 | 8.3259  | 0.158            | 0.001   | Clustered |
|         | Recovered | -0.0035  | 0.00038  | 8.1427  | 0.1553           | 0.001   | Clustered |
|         | Deceased  | -0.0018  | 0.000371 | 11.8834 | 0.227            | 0.001   | Clustered |
| Dec-20  | Confirmed | -0.0015  | 0.000367 | 8.8278  | 0.1675           | 0.001   | Clustered |
|         | Recovered | -0.0026  | 0.000404 | 8.0221  | 0.1587           | 0.001   | Clustered |
|         | Deceased  | -0.0016  | 0.000383 | 12.9425 | 0.2518           | 0.001   | Clustered |
| Jan-21  | Confirmed | -0.0016  | 0.000401 | 8.3418  | 0.1655           | 0.001   | Clustered |
|         | Recovered | -0.0016  | 0.000387 | 7.9515  | 0.1547           | 0.001   | Clustered |
|         | Deceased  | -0.0017  | 0.000403 | 12.094  | 0.2411           | 0.001   | Clustered |
increase after that. The footprint’s highest point was discovered in October. Following then, there is a drop in the footprint. In addition, the number of hotspot districts peaked in October.

### 3.4 Temporal analysis of the footprint

A temporal analysis of the footprint will be performed using the method described in Sect. 2.9. The relative pattern of the hotspot’s footprint is shown in Fig. 8. The monthly variation in the three cases of the footprint is depicted in the graph. We can see that the footprint of confirmed cases is less than the footprint of recovered cases in April 2020. Similarly, for each month, we can examine the entire plot. Another thing to notice from the graph is that the footprint of all the cases is growing over time, peaking in October 2021. In October, the maximum area covered by dead cases is also discovered. The hotspot’s footprint pattern is depicted in Fig. 9. The graph depicts the shift in the three classifications of the hotspot’s footprint, namely confirmed cases, recovered cases, and deceased cases. Initially, all of the footprints were reduced.
in May. This indicates that the hotspot’s coverage region has shrunk. All of them rose after June and continued to do so until October. Only in July and September did the number of recovery instances drop. All of the imprints were reduced again in November and December. In January, there was still another increase in the number of footprints.

4 Conclusions

In this paper, we have considered India as the study region. To find the change pattern of the hotspot’s footprint, Covid-19 data of India is taken. The study region is granulated at the district level, and the dataset is temporal, from April 24, 2020, to January 31, 2021. The Global Moran’s I is used to find the presence of clustering in the dataset, while Local Moran’s I find location of the spatial cluster, i.e., hotspot in the dataset.

A Local Moran’s I is calculated for each district. A district having the high Moran’s I surrounded by a district with high Moran’s I is considered the hotspot. Finally, statistical significance is done by using Monte Carlo Simulation. Only the statistically significant hotspot is considered as the significant hotspot.

This paper also defines the footprint mathematically, according to the definition coverage of the hotspot defines the footprint. The footprint of the hotspot is calculated from the identified hotspot. Following the identification of the footprint, change analysis is performed. The footprint of confirmed, deceased, and recovered cases was calculated from the identified hotspot of the respective type. These are shown in Figs. 7, 8, and 9. From the analysis of these three, we can see that initially, footprint increases and decreases after reaching peaks. Also, the number of districts identified as the hotspot increases in the same fashion as the footprint. Only the exception is the footprint of Confirmed cases where the peak for the number of districts is found in September. Change analysis for all the cases reveals that the footprint is highest in October. So, October is the most dangerous month. Also, change analysis shows an increase and decrease in the footprint. To find the pattern of change vector is defined. Change vector shows whether the footprint is increasing or decreasing.

The paper’s subsequent work will include expanding the data set to include the second wave of the Covid-19 pandemic in India and determining the second wave pattern and the relationship between the two waves of the epidemic, which will aid in the modeling of the third wave. As a result, the third wave’s impact on human life should be minimized. This analysis is also beneficial for fire, flood, drought, deforestation, urbanization, etc.

Author contributions Sonajharia Minz offers comprehensive direction for early manuscript drafting and preparation, as well as revisions of the paper via consecutive improvements of the article in various revision versions. The portions of the paper are written by Mohd Shamsh Tabarej. He also assumed responsibility for the corresponding authors, as well as handling the paper’s modification and re-submission.

Funding Maulana Azad National Fellowship (MANF) funded by ministry of Minority Affairs was received to assist with the preparation of this manuscript.

Data availability (Data transparency) COVID-19 data is obtained for the website http://api.covid19india.org

Code availability (Software application or custom code) QGIS Software is used for the map plotting and Python 3 is used for implementation of the algorithms.

Declaration

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.
Ethics approval  The material in the study has been ethically approved by all of the paper’s authors.

Consent for publication  All the authors of the paper give their consent to publish the material presented in the paper.

References

1. El-Dairi, M., & House, R. J. (2019). Coronavirus disease 2019- COVID-19. In Handbook of Pediatric Retinal OCT and the Eye-Brain Connection (pp. 285–287). https://doi.org/10.1007/B978-0-323-60984-5-00062-7

2. Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). The Indian Journal of Pediatrics, 87, 281–286.

3. Salgotra, R., Gandomi, M., & Gandomi, A. H. (2020). Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. Chaos, Solitons and Fractals. https://doi.org/10.1016/j.chaos.2020.109945

4. Zumla, A., Chan, J. F. W., Azhar, E. I., Hui, D. S. C., & Yuen, K. Y. (2020). Novel Coronavirus – Thailand (ex-China). (2020). https://www.who.int/medicines-documents/world-health-day-2020-

5. Chan, J. F. W., Lau, S. K. P., & Woo, P. C. Y. (2013). The emerg- ing novel Middle East respiratory syndrome coronavirus: The “knowns” and “unknowns.” Journal of the Formosan Medical Association, 112(7), 372–381. https://doi.org/10.1016/j.jfma.2013.05.010

6. Chan, J. F. W., To, K. K. W., Tse, H., Jin, D. Y., & Yuen, K. Y. (2013). Interspecies transmission and emergence of novel viruses: Lessons from bats and birds. Trends in Microbiology, 21(10), 544–555. https://doi.org/10.1016/j.tim.2013.05.005

7. Riou, J., & Althaus, C. L. (2020). Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance, 25(4), 1–5. https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058

8. Lancet, T. (2020). India under COVID-19 lockdown. The Lancet, 395(10233), 1315. https://doi.org/10.1016/S0140-6736(20)30938-7

9. Reid, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A., & Jewell, C. P. (2020). Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. MedRxiv. https://doi.org/10.1101/2020.01.23.20018549

10. Boldog, P., Tekeli, T., Vizi, Z., Dénès, A., Bartha, F. A., & Röst, G. (2020). Risk assessment of novel coronavirus COVID-19 outbreaks outside China. Journal of Clinical Medicine, 9(2), 571. https://doi.org/10.3390/jcm9020571

11. Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. Nature, 429(6988), 180–184. https://doi.org/10.1038/nature02541

12. Akhmetzhanov, A. R., Mizumoto, K., Jung, S., Linton, N. M., Omori, R., & Nishiura, H. (2020). Epidemiological characteristics of novel coronavirus infection: A statistical analysis of publicly available case data. MedRxiv.

13. Liu, T., Hu, J., Xiao, J., He, G., Kang, M., Rong, Z., Lin, L., Zhong, H., Huang, Q., Deng, A., Zeng, W., Tan, X., Zeng, S., Zhu, Z., Li, J., Gong, D., Wan, D., Chen, S., Guo, L., et al. (2020). Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China. BioRxiv. https://doi.org/10.1101/2020.01.25.919787

14. Parvin, F., Ali, S. A., Hashmi, S. N. I., & Ahmad, A. (2021). Spatial prediction and mapping of the COVID-19 hotspot in India using geostatistical technique. Spatial Information Research, 29(4), 479–494. https://doi.org/10.1007/s41324-020-00375-1

15. Ha, J. H., Yoon, D. H., & Koh, J. H. (2016). Evidence for correlation between land use and PM10 hotspot explored by entropy weight. Spatial Information Research, 24(5), 599–606. https://doi.org/10.1007/s41324-016-0056-4

16. Ahmad, F., Goparaju, L., & Qayum, A. (2018). Himalayan forest fire characterization in relation to topography, socio-economy and meteorology parameters in Arunachal Pradesh. India. Spatial Information Research, 26(3), 305–315. https://doi.org/10.1007/s41324-018-0175-1

17. Goparaju, L., & Ahmad, F. (2019). Analyzing the risk related to climate change attributes and their impact, a step towards climate-smart village (CSV): A geospatial approach to bring geoponics sustainability in India. Spatial Information Research, 27(6), 613–625. https://doi.org/10.1007/s41324-019-00258-0

18. Ahmad, F., Goparaju, L., Arakali, A. H., Alipur, H., & Shojaei, S. (2017). Applying Delphi method in prioritizing intensity of flooding in Ivar watershed in Iran. Spatial Information Research, 25(2), 173–179. https://doi.org/10.1007/s41324-017-0086-6

19. Shojaei, S., Alipur, H., Arakali, A. 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.

20. Piri, I., Moosavi, M., Taheri, A. Z., Alipur, H., Shojaei, S., & Mousavi, S. A. (2019). The spatial assessment of suitable areas for medicinal species of Astragalus (Astragalus hypsogeton Bunge) using the Analytic Hierarchy Process (AHP) and Geographic Information System (GIS). Egyptian Journal of Remote Sensing and Space Science, 22(2), 193–201. https://doi.org/10.1016/j.ejrs.2018.02.003

21. 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 cellular. KN Journal of Cartography and Geographic Information, 70(2), 45–53. https://doi.org/10.1007/s42489-020-00037-0

22. Huang, J., & Tang, J. (2021). Discovery of arbitrarily shaped significant clusters in spatial point data with noise. Applied Soft Computing, 108, 107452. https://doi.org/10.1016/j.asoc.2021.107452

23. Shioide, S., & Shioide, N. (2020). A network-based scan statistic for detecting the exact location and extent of hotspots along urban streets. Computers, Environment and Urban Systems, 83, 101500. https://doi.org/10.1016/j.compenvurbys.2020.101500

24. Mondal, S., Singh, D., & Kumar, R. (2022). Crime hotspot detection using statistical and geospatial methods: a case study of Pune City, Maharashtra, India. Geojournal. https://doi.org/10.1007/s10708-022-02173-z

25. Barboza, G. E., Schiamberg, L. B., & Pachl, L. (2021). A spatiotemporal analysis of the impact of COVID-19 on child abuse and neglect in the city of Los Angeles, California. Child Abuse and Neglect, 116(2), 104740. https://doi.org/10.1016/j.chiabu.2020.104740

26. Abulideh, A. (2021). Spatiotemporal analysis of water-electricity consumption in the context of the COVID-19 pandemic across six socioeconomic sectors in Doha City, Qatar. Applied Energy, 304, 117864. https://doi.org/10.1016/j.apenergy.2021.117864
29. Aral, N., & Bakir, H. (2021). Spatiotemporal analysis of COVID-19 in Turkey. *Sustainable Cities and Society*, 76, 2. https://doi.org/10.1016/j.scs.2021.103421

30. Purwanto, P., Utaya, S., Handoyo, B., Bachri, S., Astuti, I. S., Sastro, K., Utomo, B., & Aldianto, Y. E. (2021). Spatiotemporal analysis of COVID-19 spread with emerging hotspot analysis and space-time cube models in East Java, Indonesia. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi10030135

31. Paul, R., Arif, A. A., Adeyemi, O., Ghosh, S., & Han, D. (2020). Progression of COVID-19 from urban to rural areas in the United States: A spatiotemporal analysis of prevalence rates. *Journal of Rural Health*, 36(4), 591–601. https://doi.org/10.1111/jrh.12486

32. Liu, W., Wang, D., Hua, S., Xie, C., Wang, B., Qiu, W., Xu, T., Ye, Z., Yu, L., Yang, M., Xiao, Y., Feng, X., Shi, T., Li, M., & Chen, W. (2021). Spatiotemporal analysis of COVID-19 outbreaks in Wuhan. *China. Scientific Reports*, 11(1), 1–9. https://doi.org/10.1038/s41598-021-93020-2

33. Spassiani, I., Sebastiani, G., & Palù, G. (2021). Spatiotemporal analysis of covid-19 incidence data. *Viruses*, 13(3), 1–13. https://doi.org/10.3390/v13030463

34. Ye, Y., Wang, C., Zhang, H., Yang, J., Liu, Z., Wu, K., & Deng, Y. (2020). Spatiotemporal analysis of COVID-19 risk in Guangdong Province based on population migration. *Journal of Geographical Sciences*, 30(12), 1985–2001. https://doi.org/10.1007/s11442-020-1823-7

35. COVID19-India API. (2020). Data. Retrieved January 10, 2021. https://data.covid19india.org/

36. Hijmans, R. (2009, August 12). Download data by country|DIVA-GIS. DIVA-GIS. Retrieved January 10, 2021. https://www.diva-gis.org/gdata

37. Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27(4), 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

38. Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

39. Tabarej, M. S., & Minz, S. (2019). Rough-set based hotspot detection in spatial data. *Communications in Computer and Information Science*, 1046, 655–665. https://doi.org/10.1007/978-981-13-9942-8

40. Tabarej, M. S., & Minz, S. (2020). Change footprint pattern analysis of crime hotspot of Indian Districts. *Advances in Intelligent Systems and Computing*, 1141, 325–335. https://doi.org/10.1007/978-981-15-3383-9_30

41. Zhou, X., Shekhar, S., & Ali, R. Y. (2014). Spatiotemporal change footprint pattern discovery: An inter-disciplinary survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(1), 1–23. https://doi.org/10.1002/widm.1113

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