Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A study on geospatially assessing the impact of COVID-19 in Maharashtra, India

Saneev Kumar Das a,⁎, Sujit Bebortta b,⁎

a Department of Computer Science and Engineering, Centurion University of Technology and Management, Bhubaneswar, Odisha 752050, India
b School of Information and Computer Sciences, Department of Computer Science, Ravenshaw University, Odisha 753003, India

ARTICLE INFO

Article history:
Received 11 May 2021
Revised 21 December 2021
Accepted 27 December 2021
Available online 13 January 2022

Keywords:
COVID-19
Geographic information systems (GIS)
Choropleth rendering
Spatial clustering
Spatial autocorrelation
Getis-Ord Gs

ABSTRACT

The emergence of 2019 novel corona virus disease (COVID-19) raised global health concerns throughout the world. It has become a major challenge for healthcare personnel and researchers throughout the world to efficiently track and prevent the transmission of this virus. In this paper, the role of geographic information system (GIS) based spatial models for tracking the spread of COVID-19 and discovery of testing centres in Maharashtra, India was studied. The datasets collected from diverse sources were geocoded to make it geospatially compatible. A three-tiered framework was proposed to practically realize the impact of COVID-19 in a cartographic fashion. Initially, choropleth maps labeled with testing centres, number of confirmed cases and casualties were visualized in a district-wise manner. Heatmaps for visualizing the spatial density of confirmed cases and casualties were presented. The visualization of spatial clustering and density of confirmed cases was done using K-means clustering for optimal value of "k" estimated using the heuristic-based Elbow method was provided along with zonal analysis of the districts. Map showing spatial autocorrelation was also presented to identify spatial hotspots and coldspots. The districts of Pune and Thane reported respective z-scores of 3.424 and 3.347 along with p-values of 0.006 and 0.001 respectively. It was inferred from the generated results that Pune and Thane districts in Maharashtra were identified as COVID-19 hotspots. Based upon this analysis, certain effective mitigation strategies can be devised in order to check the uncontrolled spread of COVID-19 in the identified hotspot areas.

1. Introduction

The 2019 novel coronavirus disease (COVID-19) has vulnerably led to a highly infectious pneumonia type disease, which has abruptly affected a large population of the world (Gilbert et al., 2020; Chen et al., 2020). The World Health Organization (WHO), affirmed the outbreak of this virus as an international public health emergency. The incidence of this virus was initially reported in Hubei province of China, being claimed to have found its origin from the sea food market of Huanan. Several affected countries have witnessed a deplorable decline in their social and economic developments following the impact of COVID-19 crisis (Rodriguez-Morales et al., 2020). Considering the global outbreak of COVID-19, John Hopkins Resource Centre for coronavirus had suggested several modelling techniques for monitoring and tracking the spread of this virus globally (Sarwar et al., 2020). Towards this end, the efficacy of geographic information system (GIS) based techniques have found large-scale applications in evidencing the spatial patterns observed in tracking the spread and demographic trends of COVID-19 (Vasantha and Patil, 2020; Radanliev et al., 2020).

The concept of GIS in healthcare domain has improved a lot during COVID-19 times and the spatio-temporal analysis of the distribution of COVID-19 has played a key role in controlling its spread (Franch-Pardo et al., 2020). The role of visualizing real-time GIS-based tracker for COVID-19 and corresponding dashboard has helped health personnel a lot (Rosenkrantz et al., 2021; Mollalo et al., 2020). Many researches have shown the efficacy of convolutional neural networks in scrutinizing the reason for rapid transmission of COVID-19 (Jadhav et al., 2021). The time-series analysis to predict rise and fall of COVID-19 cases has been scrutinized with the assistance of GIS and machine learning techniques (Khan et al., 2021). Decision makers were educated through proper geospatial analysis in the entire COVID-19 period in various...
regions across the world. The role of GIS in presenting the vulnerability mapping, risk analysis, and hotspot identification has helped a lot in decision making for further actions (Shadeed and Alawna, 2021; Arab-Mazar et al., 2020; Ahasan and Hossain, 2021).

1.1. Motivation

The rapidly growing cases of COVID-19 in Maharashtra province of India has been a matter of concern and several mitigation strategies require to be adopted in order to check the uncontrolled spread of the deadly corona virus. In the domain of computer science, geographic information systems (GIS) has been playing a key role in constantly tracking and monitoring the spread of COVID-19. The efficacy of GIS in mapping such a rife pandemic motivated us to geospatially assess the impact of COVID-19 in Maharashtra province of India. Our prime focus was to determine COVID-19 hotspots as well as coldspots in order to educate the public authorities to take necessary actions for controlling the further spread of COVID-19 in those regions. Since COVID-19 has been communicating rapidly, mitigation strategies should focus in breaking the chain of spread. Geospatial assessment of COVID-19 can aid the common public in constantly monitoring the number of confirmed cases, number of recovered cases and number of casualties which would further create a sense of awareness among the common public. Following regulatory guidelines to safeguard oneself can lead to the checking of uncontrolled spread of COVID-19.

In this paper, an extensive apprehension for understanding the role of GIS based spatial analytics in tracking the spread of COVID-19 and discovery of nearest testing services has been provided. The performance of several spatial data analysis models has been reported. The experimentation has been performed on the state of Maharashtra, which is the second most populous state in India. A district-level analysis of the nearest COVID-19 Testing Centres (CTCs) is provided along with a choropleth analysis indicating the confirmed cases for each district. We have also implemented heatmap analysis for depicting the density points pertaining to the confirmed cases and number of casualties over the provinces of Maharashtra. The K-means clustering approach was employed for creation of spatial clusters corresponding to number of confirmed COVID-19 cases. The Elbow method was leveraged to estimate the optimal number of spatial clusters corresponding to minimum value of Intra-cluster variation (ICV). Further, the use of spatial autocorrelation technique for hotspot analysis of the areas at highest risk of being affected by considering different statistical metrics such as p-value, z-score, and so on were reported. Hence, our proposed framework provides a comprehensive visualization and monitoring strategy for realizing the spatial effects of COVID-19. This ranges from mobility monitoring of the population for an identified hotspot zone to tracing nearest health services. This framework can leverage policy makers and local healthcare authorities to develop strategic models by combining interactive mapping schemes for imposing authoritative and restrictive measures in ruling the spread of COVID-19.

1.2. Contributions

We scrutinized similar works that assessed the impact of COVID-19 using diverse technologies in the domain of computer science. We presented a generalized GIS-based framework which entails three modules for generating diverse maps to visualize COVID-19 statistics over a specified region. The testing zones in Maharashtra were provided as data points labeled with a distinct symbol over a choropleth map representing district wise number of confirmed cases. Choropleth map epitomizing casualties due to COVID-19 in a district-wise manner was presented. Heatmaps corresponding to regions with higher density of confirmed cases as well as casualties were provided. Considering, the NP-hardness of K-means clustering approach, a heuristic based technique i.e., the Elbow method was used to estimate the optimal number of spatial clusters pertaining to the area of study which provides fast convergence to a local optimum. We performed spatial K-means clustering along with zonal analysis pertaining to optimal value of k estimated using heuristic-based Elbow method. Spatial autocorrelation analysis is presented with the aid of Getis-Ord Gi/C3 to identify COVID-19 hotspots as well as coldspots based on z-scores.

2. Related works

The role of GIS has been extensively studied in healthcare sector for geographically classifying patient demographics, monitoring trends related to health for specific regions, tracking the outbreak and spread of infectious diseases, and so on. Wang et al. (2019), studied the epidemiological traits of COVID-19 cases in Jiangsu province. They performed analysis over COVID-19 data collected from 22 January to 20 February, 2020. A peak was observed around 31 January, 2020 specific to their area of study mostly resulting from travellers to Hubei province. Al-Ahmadi et al. (2019), performed spatiotemporal analysis upon the cases of Middle East respiratory syndrome coronavirus (MERS-CoV) in Saudi Arabia. They employed Kulldorff’s spatial scan statistic for identifying the time period as well as spatial regions at high risk of being affected by MERS-CoV. From their analysis, ten spatiotemporal annual clusters and three monthly clusters were identified for Riyadh province which were considered as high risk areas.

Lakhanì (2020), conducted a spatial analysis for discovery of palliative care services in the Melbourne metropolitan area for aging adults with COVID-19. In this study, the primary objective was to identify high priority areas in need of palliative care services with highest population of adult citizens over 65 years of age using GIS techniques. Several epidemiological hypotheses comprise of climatological analysis for understanding the survival and spread of infectious diseases. Briz-Redón and Serrano-Aroca (2020), explored the impact of temperature on the spread of COVID-19 in Spain. Several factors like population density of the provinces, age group, and number of travellers to those regions were analyzed. It was observed that there was no impact of warmer temperature on the cases of COVID-19.

Rashed et al. (2020), analyzed the impact of population, temperature, and absolute humidity on the spread of COVID-19 cases and derived a correlation between the pattern of spread and decay in COVID-19 cases. Their analysis was conducted in a multi-prefecture setting in Japan. A promising correlation result was obtained for observing decay in the duration of cases with its spreading frequency. However, exceptions were observed for regions with more number of foreign returnees and travellers. Pani et al. (2020), performed a meteorological analysis upon the transmission and spread of COVID-19 in Singapore. They employed Spearman and Kendall correlation tests for analyzing the associations observed between COVID-19 cases and the meteorological parameters.

Defries et al. (2020), proposed a GIS based disease spread model for assessing the spread of COVID-19 post lockdown in 32 administrative districts of Central India. They employed the susceptible-exposed-infected-recovered (SEIR) model for hypothetically studying the epidemiological spread of COVID-19. Guliyev (2020), extensively studied spatial panel models for analyzing the spatial effects of COVID-19 in Mainland China. Models such as spatial autocorrelation, spatial autoregression (SAR), spatial linear regression model (SLM), spatial Durbin model (SDM), spatial error model (SEM), spa-
tial weight matrix (SWM), and spatially lagged-X (SLX) model were employed over the COVID-19 data acquired for 31 regions from 22 January to 10 March, 2020. The spatiotemporal impact over the death rates, recovery rates, and confirmed cases for this time duration was analyzed by employing the above models.

3. Materials and methods

In this section, deals with the study area along with the specification of dataset. The three-tiered proposed framework is explained in detail along with some background knowledge presented on spatial clustering and spatial autocorrelation techniques. The hardware specification of the local engine over which the implementation of the proposed framework was performed includes 1 TB HDD (hard disk drive), 256 GB SSD (solid state drive), 8 GB RAM (random access memory) and Intel(R) Core(TM) i5-8265U CPU. The tool used to perform the geospatial assessment over COVID-19 data is QGIS 3.12.3 Bucuresti.

3.1. Study area and dataset specification

The study area selected for assessing the impact of COVID-19 was Maharashtra province situated in the western division of India. The geographical extent of Maharashtra is spread over 307,713 sq km with the GPS coordinates 19°39'47.808"N and 75°18'1.054"E. The state of Maharashtra has been adversely affected by COVID-19 with the highest number of confirmed cases as well as casualties. The study area comprised of 35 districts in Maharashtra province of India. Due to unavailability of geospatial data for COVID-19 spread in Maharashtra, we have created two significant datasets manually which is summarized in Table 1. The data was collected initially from https://covidindia.org/maharashtra and https://www.icmr.gov.in/pdf/covid/labs. The description of attributes in the considered datasets i.e., “Maharashtra-COVID” and “Testing Zones” is summarized in Table 2 and Table 3 respectively. We observed that the datasets collected were not geographically significant and thus, we performed geocoding operations on the aggregated datasets to allocate geographical coordinates to each instance. Also, the testing centres in Maharashtra differentiated by type of tests were geocoded to make it geographically significant and compatible for performing GIS operations. The datasets after preprocessing constituted of 35 districts and 95 testing centres with latitude and longitude values allocated to each instance.

3.1.1. Dataset description

The vector shape file of Maharashtra region was used in order to get the polygons of each district over which geospatial analysis was performed. There were options to represent the same over OpenStreet Maps or any raster layer but spatial autocorrelation over any raster becomes difficult. Thus, we have considered the use of vector layer for Maharashtra region. Further, the Maharashtra-COVID dataset was overlaid on the polygons so as to identify the status of COVID-19 in each region based on its corresponding geospatial coordinates. Further, Testing Zones dataset was used in order to correlate the number of cases, recoveries, and the testing zones nearby and simultaneously plot the point-cluster overlay.

3.2. Proposed framework

The contributions of our proposed technique can be realized from Fig. 1. The architecture is three-tiered which assists in appraisal of COVID-19 hotspots for policy makers and healthcare authorities. The initial phase deals with collection of COVID-19 dataset from diverse sources and processing the acquired datasets to satisfy our domain of study. The processed data is then employed for creation of spatial database which could be appropriately handled using GIS tools. Further, we perform geocoding, which is a spatial mapping operation over the spatial datasets. We transform the geographical coordinates pertaining to our area of study for accomplishing a descriptive view of the vital spatial epidemiological parameters considered in this study. The obtained coordinate pairs and dataset attributes are overlaid upon the respective shape

### Table 1

| Dataset Name | Rows | Columns | Instances | Source |
|--------------|------|---------|-----------|--------|
| Maharashtra-COVID | 35   | 6       | 210       | https://covidindia.org/maharashtra |
| Testing Zones | 95   | 5       | 475       | https://www.icmr.gov.in/pdf/covid/labs |

### Table 2

Attribute specification for the dataset “Maharashtra-COVID”.

| Attributes | Description |
|------------|-------------|
| District   | This parameter comprises names of considered 35 districts in Maharashtra which in turn constitutes the study area. |
| Confirmed  | This attribute epitomizes the district wise total number of confirmed cases of COVID-19 as of 4 August, 2020 in Maharashtra. |
| Recovered  | This column signifies the number of COVID-19 confirmed cases that are tested COVID-19 negative and discharged as of 4 August, 2020 in Maharashtra. |
| Death      | This parameter epitomizes the total number of casualties occurred due to COVID-19 in a district wise manner as of 4 August, 2020 in Maharashtra. |
| Latitude   | After performing the geocoding operations, this attribute is added to the dataset which specifies the geographical latitude of each district in Maharashtra. |
| Longitude  | As the geocoding operations are performed to make the dataset geographically significant, this attribute provides the longitude coordinates of each district in Maharashtra. |

### Table 3

Attribute specification for the dataset “Testing Zones”.

| Attributes | Description |
|------------|-------------|
| Testing Zones | This attribute provides the name of various organizations that are dedicated towards testing for COVID-19 among suspected patients. |
| Organization | This parameter is of binary type and indicates whether a testing organization is of government type or private type. |
| Type | This attribute determines the types of test performed by the testing zones in Maharashtra. The type of tests included in the dataset are RT-PCR (reverse transcription polymerase chain reaction), TestNat and CB-NAAT (cartridge based nucleic acid amplification test). Each instance of this attribute specifies the type of test performed by each organization. |
| Test Type | The geocoding operation performed over the dataset created this attribute to indicate latitude values for each testing organization in Maharashtra. |
| Latitude | The geocoding operation performed over the dataset created this attribute to indicate latitude values for each testing organization in Maharashtra. |
| Longitude | This attribute epitomizes the geocoded longitude values allocated to each testing organization in Maharashtra. |
files to procure a complete spatial visualization of the epidemiological impact of COVID-19.

The maps obtained from the first module are then processed using appropriate GIS operations like choropleth rendering, heatmap analysis, spatial cluster analysis, and spatial autocorrelation analysis. This comprises the second module of our proposed framework. Statistical parameters such as mean, standard deviation, variation coefficient, and minimum and maximum values have been analyzed for inevitable attributes in the considered datasets. Further, essential statistics such as p-value, z-score, and Getis-Ord G* statistics for individual districts of the study area have been analyzed and their respective values obtained. These statistics also play a crucial role in identifying the COVID-19 infection hotspots, a detailed description of which is provided in the later sections.

Finally, the third layer deals with the storage, processing, visualization, and dissemination of maps for COVID-19 affected regions and the nearest COVID-19 Testing Centres (CTC). These results can be combined with interactive maps obtained from National Spatial Data Infrastructures (NSDI) by local policy makers and health service providers to efficiently track the spread of disease and impose regulatory measures. This also assists in monitoring some non-meteorological parameters like monitoring mobility of citizens, geographical distribution of affected regions, monitoring health demographics, and daily trends in the spread of COVID-19.

3.2.1. Spatial cluster analysis

The convergence of machine learning algorithms with GIS technology provides predictive competency to the GIS based model (Das and Bebortta, 2020; Das et al., 2020). Spatial clustering is an unsupervised learning task which works on unlabeled data in a spatial environment. Spatial clustering technique clusters similar data points based on features further differentiating the clusters (Shekhar et al., 2011). Existing spatial clustering algorithms include DBSCAN (Density Based Spatial Clustering of Applications with Noise) clustering and K-means clustering. The spatial clustering algorithms generate clusters in a spatial setting after specifying certain parameters. The number of clusters to be generated is user-defined and it can further be modified by manipulating the corresponding attribute table. In this paper, we have used the well-known K-means clustering technique in order to create different clusters of importance. After specifying the number of user-defined clusters, each data point is allocated a cluster ID. The spatial proximity based on which K-means clustering is performed is estimated using Euclidean or Manhattan distance between two data points. Also, as per user convenience, cluster IDs can be allocated by modifying the attribute table.

Unlike DBSCAN clustering, the K-means clustering approach proves to be more versatile towards handling large datasets and remains unaffected for varying densities of data points (Rodriguez et al., 2019). Thus, this article uses K-means clustering technique since the dataset which we considered for analysis was a user-specified dataset and the amount of noise in the dataset was negligible (Rodriguez et al., 2019). However, in case of noisy data, DBSCAN clustering is considered to be more efficient.

Considering, each spatial point to be a real valued vector, the Euclidean distance between samples a and b can be given as, \( \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2} \), where \( i = 1, 2, 3, \ldots, n \) represents \( i \)th samples. Following Tîrnăucă et al. (2018) and Kosovac et al. (2020), the ICV for finding the appropriate value of \( k \) corresponding to spatial clusters \( C_k \) is obtained as, \( ICV = \sum_{b_i \in C_k} (b_i - \mu_k)^2 \), where \( \mu_k \) represents the mean value for the clusters. The different spatial clusters can be heuristically chosen by leveraging the Elbow method computed for minimum value of ICV as, \( \arg \min_k \left( \sum_{b_i \in C_k} (b_i - \mu_k)^2 \right) \).

3.2.2. Spatial autocorrelation analysis

In this study, we employed an efficient spatial hotspot identification technique known as Getis-Ord \( G'_* \). This technique comes under local indicator of spatial association (LISA) framework which has been often employed for identifying spatial hotspots. Unlike the Getis-Ord global G model (Getis, 1996; Getis and Ord, 2010; Shekhar and Xiong, 2007), Getis-Ord \( G'_* \) leverages the problem for identifying neighboring data points within certain specific search bands as it includes the reference data point as a neighbor of itself. Hence, all the data points for a given observation have a minimum of one neighbor. Therefore, by employing the Getis-Ord \( G'_* \) technique, we can achieve the statistically significant hotspot regions as Getis (1996) and Getis and Ord (2010),

\[
G'_* = \frac{\sum_j (\omega_j x_j - \bar{x} \omega_j)}{\sqrt{\sum_j (\omega_j (n-\omega_j))^2 / (n-1)}}
\]

In Eq. (1), \( x_j \) represents the characteristic value corresponding to data point \( j \) where \( j = 1, 2, \ldots, n \). Spatial weights \( \omega_j \) and \( \omega_j \) correspond to the weights distributed among data points \( i \) and \( j \) respectively, where \( \omega_j \) and \( \omega_j \) account for \( n \) data points. The mean of \( x_j \) attributes for \( n \) data points is represented as \( \bar{x} \) and can be given as follows,

\[
\bar{x} = \frac{\sum x_j}{n}
\]

Finally, the spatial weights \( s \) for entire data points considered for evaluation is provided as,

\[
s = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}
\]

4. Results

We have visualized cartographically various choropleth maps epitomizing confirmed cases as well as fatalities due to COVID-19 in Maharashtra. Heatmap analysis was performed in order to
identify COVID-19 hotspots based on confirmed cases and casualties in a district-wise manner. With the aid of spatial clustering techniques, various clusters were generated according to confirmed case counts. Finally, spatial autocorrelation analysis was performed to epitomize the hotspots as well as coldspots of COVID-19 in Maharashtra province of India. In this section, practical realization of the proposed framework is presented in detail.

As the initial phase of the proposed framework suggests the creation of spatial database and geocoding, we collected relevant data of COVID-19 for 35 districts in Maharashtra. The aggregated data was then transformed into a spatial database by performing geocoding operations which allocated latitude and longitude values to each instance of the considered datasets. The datasets in our case, were imported as delimited text into the GIS application. After validating the geospatial compatibility of the datasets, each feature attribute was plotted as data points over the vector shape file of Maharashtra province. The vector shape file is a collection of polygons for each sub-region within a region. The data points when plotted over a vector, the operation is called a point-vector overlay operation (Shekhar and Xiong, 2007). The primary phase of the proposed framework thus comprised of the geocoded datasets indulged within a spatial database and point-vector overlay operation. Generating labels to each data point is an indispensable task while working with GIS technology and thus, we properly labeled each data point by the corresponding district names.

The further stages of the proposed model recommends the implementation and visualization of geospatial and statistical operations over the output generated from the initial phase i.e., point-vector overlay. In this module, initially the statistical summary of the feature attributes in the considered dataset i.e., “Confirmed”, “Recovered” and “Death” is presented in Table 4. As the confirmed cases of COVID-19 is rapidly increasing, the number of testing per day needs to be improved. As per the collected data of testing centres, we have classified the number of government as well as private testing agencies as per the type of testing i.e., RT-PCR, TrueNat and CB-NAAT. The plot epitomizing the classification of testing centres is presented in Fig. 2.

Fig. 2 shows some difference in the government and private testing zones. The difference is due to lack of resources as well as the investment cost. Since, COVID-19 was considered to be a novel virus, the testing mechanisms were initially not so trustworthy and private organizations thus were reluctant to provide testing resources initially (Agarwal et al., 2019).

After obtaining the point-vector overlay, we performed join operation on the vector shape file to merge the attributes in the imported delimited text with the attributes of shape file. On successful join operation, the choropleth maps were generated by weighting data points in adherence with the confirmed cases which creates a graduated map of diverse colors specifying range of equal intervals of confirmed cases in Maharashtra. Further, we labelled each data point by the number of confirmed cases in the defined district. A distinct symbol is used to define COVID-19 testing zones in different regions of Maharashtra. The initial choropleth map with defined testing centers is epitomized in Fig. 3. A choropleth map created in accordance with number of casualties in a district wise manner with labeled death counts was presented in Fig. 4. The visualization using point-cluster analysis was presented in order to identify number of testing centres in each region compared to the number of positive cases which in turn can be analyzed to identify the regions requiring more testing centres. The corresponding map was presented in Fig. 5. In order to identify dense regions where number of confirmed cases were maximum, we performed heatmap analysis by using pseudo-color codes whose intensity gradually decreases as the density of weighting factor decreases. By selecting the weighting factor to be number of confirmed cases and number of fatalities, two corresponding heatmaps were visualized in Fig. 6 and Fig. 7 respectively.

The efficacy of machine learning in GIS technology is well-known and spatial clustering techniques aid in analyzing pandemic situations. The number of clusters defined using spatial K-means clustering approach over the latitude and longitude attributes were tested for k = 5, k = 4, and k = 3 which was represented in respective number of diversified colors in Fig. 8. Further, the estimates from the Elbow plot indicated the optimal value of k corresponding to minimum ICV value as presented in Fig. 9. K-means clustering technique by defining three clusters with cluster IDs ranging from 0 to 2 is implemented in this assessment. The number of clusters were set to three since the optimal value of k was found to be three after incorporating heuristic-based Elbow method over the dataset. The executed spatial K-means clustering was visualized in Fig. 10. The outcome of the performed spatial K-means clustering was a zonal classification of Maharashtra in accordance with the number of confirmed cases. Spatial autocorrelation analysis was then performed by specifying the LISA parameter to be Getis-Ord G, and the weighting type to be “Queen”. The weighting factor was selected as the number of confirmed cases based on which significant p-values, z-scores and G, values were generated. As the z-score becomes extreme positive, the region was identified as the hotspot and the regions for which z-scores were extreme negative were identified coldspots. The spatial autocorrelation map with labeled confirmed cases and casualties for each district along with diverse color codes specifying range of z-scores was provided in Fig. 11.

Table 4
Summary of fundamental statistics for the parameters “Confirmed”, “Recovered” and “Death” as in the considered dataset.

| Statistics                  | Confirmed | Recovered | Death |
|-----------------------------|-----------|-----------|-------|
| Total Count                 | 35        | 35        | 35    |
| No. Of Unique instances     | 35        | 35        | 35    |
| Null values                 | 0         | 0         | 0     |
| Min. Value                  | 0         | 0         | 0     |
| Max. Value                  | 134356    | 105134    | 7515  |
| Range                       | 134356    | 105134    | 7515  |
| Sum                         | 612289    | 439905    | 20745 |
| Mean                        | 17008.028 | 12219.584 | 7576.25 |
| Median                      | 3793.5    | 2193.0    | 127.5 |
| Std. dev.                   | 34008.686 | 26155.445 | 1391.708 |
| Coeff. of variation         | 1.999     | 2.140     | 2.415 |
| Rarest Occurring Value      | 0         | 0         | 0     |
| Frequent Occurring Value    | 0         | 0         | 0     |
| First quartile              | 1248      | 762.5     | 37    |
| Third quartile              | 14693     | 10051     | 446.5 |
| Inter-quartile range        | 13445     | 9288.5    | 409.5 |
A tabulation of various Getis-Ord $G'_i$ attributes in a district wise manner is presented in Table 5.

5. Discussions

India being the second most populous country is adversely affected by the COVID-19 pandemic. Due to huge population density, the rate of spread of SARS-CoV-2 virus is getting uncontrolled in India. Many provinces in India are striving to combat the spread of COVID-19. But, the province of Maharashtra is currently in a state of havoc due to the widespread nature of the ongoing pandemic. Our proposed framework deals with the geospatial assessment of COVID-19 in 35 districts of Maharashtra. Maharashtra being the second highest populous state in India bearing a popula-
Fig. 4. Choropleth map labeled with district names and district-wise casualties due to COVID-19.

Fig. 5. Point-cluster analysis performed to identify number of testing zones in each region compared to the number of cases represented in black labels, the number of recovered represented in green labels, and the number of testing centres represented within red circles.
Fig. 6. Heatmap analysis using pseudo color codes to identify COVID-19 hotspots in Maharashtra based on confirmed cases.

Fig. 7. Heatmap analysis using pseudo color codes to identify COVID-19 death hotspots in Maharashtra.
tion of 112,374,333 people as reported in Census Survey of India for the year 2011.

This study explored the role of geostatistical techniques to analyze the impact of COVID-19 spread over Maharashtra. Initially, the work provided number of confirmed cases of COVID-19 overlaid with testing centres corresponding to different districts. Choropleth map for the casualties reported was then presented. Overlay map for number of confirmed cases, recoveries, and testing centres were presented pertaining to each district. The heatmap analysis was performed to illustrate the spatial density of COVID-19 confirmed cases and death hotspots. We then exploited the K-means technique to obtain spatial pattern of COVID-19 cases for the considered study area. We leveraged the Elbow method to estimate the optimal number of clusters in convergence to minimum value of ICV. Furthermore, the Getis-Ord spatial autocorrelation technique was adopted to predict spatial hotspots and coldspots for COVID-19.

The results obtained from K-means clustering approach to observe spatially clustered patterns for spread of COVID-19 is consistent with some recent findings (Azarafza et al., 2020; Yahya et al., 2021). However, to our knowledge, in view of COVID-19 outbreak this study for the first time leveraged Elbow method towards estimating optimal number of clusters and minimizing the ICV to visualize spatial pattern of spread across different districts of Maharashtra. The studies in Bherwani et al. (2021), Kodge (2021), and Roy et al. (2021), illustrated the impact of COVID-19 over Maharashtra and other regions of India through Thiessen Polygon and choropleth maps. However, there was a lack of correlation in deriving relationship between number of confirmed cases, recoveries, deaths, and testing facilities associated with each area. The present study successfully derived a correlation between the above mentioned attributes for assessing the outbreak of COVID-19 in Maharashtra. It was observed from our analysis that the regions with higher number of testing centres facilitated better detection of COVID-19 cases and hence the recovery rate for those regions also significantly showed higher numbers as compared to other regions under the study area.

After the practical realization of the proposed framework, we observed from the generated results that Pune and Thane districts of Maharashtra recorded the highest number of confirmed cases and were considered as COVID-19 hotspots with respective z-scores of 3.424 and 3.347, and p-values of 0.006 and 0.001. The recovery rate is the highest in Mumbai i.e., the capital city of Maharashtra. Further, the results specified that the highest rise in death toll is recorded in Mumbai city with 7,515 casualties out of 130,232 confirmed cases. We inferred from the statistical plot that 44 government and 32 private organizations are conducting RT-PCR tests whereas 7 government and 3 private agencies are conducting TestNat. Also, there exists 3 government and 6 private organizations that are conducting CB-NAAT COVID-19 testing.

The objective of implementing the proposed framework is to educate the governing bodies and public authorities by geospatially assessing the spread of COVID-19 which shall aid in executing control measures in identified hotspots. The proposed framework can be extended to perform a Web-GIS based real-time analysis of the ongoing pandemic further leading to constant

Fig. 8. Spatial cluster analysis for K-means pertaining to different k values. (a) Scatter plot for \( k = 5 \). (b) Scatter plot for \( k = 4 \). (c) Scatter plot for \( k = 3 \) (optimal \( k \)).
tracking and monitoring of the spread of COVID-19 in Maharashtra. By adhering to our proposed framework, effective mitigation strategies can be developed to mitigate the increasing rate of spread of COVID-19. The real-time visualization of COVID-19 can be practically realized by using advanced cloud GIS frameworks further providing a multi-client visualization. Also, with the assistance of serverless computing paradigms, constant tracking and monitoring of COVID-19 with lesser developer effort can be made possible (Bebortta et al., 2020). The geospatial assessment of COVID-19 in Maharashtra is studied and the point-cluster analysis performed in this implementation shows the correlation between the number of testing zones, confirmed cases, and recovered count in that corresponding region.

6. Concluding remarks

The rife nature of COVID-19 pandemic in India has led towards creating a sense of havoc among common public. It therefore becomes inevitable to keep the public updated about the ongoing pandemic. To create public awareness regarding the impacts of COVID-19 as well as preventive measures, GIS technology plays a pivotal role. The constant tracking and monitoring of such a pandemic is practically made possible with the assistance of advanced GIS technologies. Nowadays, many real-time trackers are in action which plots the increasing number of confirmed cases as well as casualties due to COVID-19. This paper suggested a three-tiered framework to assess the impact of COVID-19 in Maharashtra province of India using advanced geospatial approaches. A statistical plot epitomizing the number of government as well as private agencies which are conducting COVID-19 tests classified as per the types of testing was provided. Certain choropleth maps weighted by inevitable parameters (like confirmed cases and deaths from COVID-19) were shown. The focus of the choropleth map with labeled confirmed case counts was on presenting available COVID-19 testing centres which further visually aided in identifying the nearest testing centres to each district. Also, a choropleth map showing casualties in a district-wise manner was presented. Further, in order to identify the adversely affected districts in Maharashtra, heatmap analysis was performed based on number of confirmed cases as well as number of casualties. The efficacy of spatial

![Fig. 9. Optimal number of clusters using Elbow method corresponding to minimum value of ICV.](image)

![Fig. 10. Spatial K-means clustering with zonal analysis and labeled cluster IDs for Maharashtra with labeled confirmed cases region-wise.](image)
clustering techniques was shown by performing a heuristic Elbow method based estimation of optimal value of “k” followed by zonal analysis of the rapid spread of COVID-19 in Maharashtra. Spatial autocorrelation map visually aids in identifying the hotspots as well as coldspots. In this article, the LISA parameter was selected to be Getis-Ord $G_i^C$ and spatial autocorrelation map was generated which further presented a map graduated based on $z$-scores. The tabulation of obtained Getis-Ord $G_i^C$ parameters was presented in a district-wise manner along with summarized cluster IDs. Finally, it can be inculcated that Pune and Thane districts of Maharashtra recorded the extreme positive $z$-scores of 3.424 and 3.347, and $p$-values of 0.006 and 0.001 respectively. Thus, the two districts can be called as COVID-19 hotspots in Maharashtra. So, mitigation strategies to control the spread of COVID-19 in Pune and Thane districts need to be executed.

This paper presented a geospatial analysis of increasing COVID-19 cases in Maharashtra and the regions which were identified as hotspots can be focused more by the healthcare authorities. In order to maintain the tested and confirmed ratio, the healthcare authorities can increase the number of testing zones as presented by the point-cluster analysis. Further, the regions which showed extremely positive $z$-scores as visualized using spatial autocorrelation can be considered as regions of high vulnerability and corrective measures can be taken in those regions.

![Spatial autocorrelation map](image.png)

Fig. 11. Spatial autocorrelation using Getis-Ord $G_i^C$ with labeled confirmed cases represented using blue text labels, deaths represented using green text labels along with corresponding $z$-scores in the legend.
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agarwal, N., Raheja, A., Suri, A., 2019. Guidelines for preoperative testing for neurosurgery in coronavirus disease 2019 (covid-19) era: Indian viewpoint amidst global practice. World Neurosurg. 146 (2021), 103–112.

Ahasan, R., Hassain, M.M. Leveraging gis and spatial analysis for informed decision-making in covid-19 pandemic. Health Policy Technol.

Al-Abdul, M., Al-Husaini, A., Al-Hassani, A., 2019. Spatiotemporal clustering of middle east respiratory syndrome coronavirus (mers-cov) incidence in saudi arabia. 2012–2019. Int. J. Environ. Res. Public Health 16 (14), 2520.

Arab-Mazar, Z., Sah, R., Rabaan, A.A., Dhma, K., Rodriguez-Morales, A.J., 2020. Mapping the incidence of the covid-19 hotspot in iran-imlications for travellers. Travel Med. Infect. Dis. 34, 101630.

Azarafza, M., Azarafza, M., Akgun, H. Clustering method for spread pattern analysis of coronavirus (covid-19) infection in iran. medRxiv.

Bebortta, S., Das, S.K., Kandpal, M., Barik, R.K., Dubey, H., 2020. Geospatial serverless computing: Architectures, tools and future directions. Int. J. Geo-Inf. 9 (5), 311.

Bheriwi, H., Anjum, S., Kumar, S., Gautam, S., Gautam, A., Kumbhare, H., Anshul, A., Kumar, R., 2021. Understanding covid-19 transmission through bayesian probabilistic modeling and gis-based voronoi approach: a policy perspective. Environ. Dev. Sustain. 23 (4), 5846–5864.

Briz-Rodríguez, A., Serrano-Aroca, Á. 2020. A spatio-temporal analysis for exploring the effect of temperature on covid-19 early evolution in Spain. Sci. Total Environ. 138811.

Chen, H., Guo, J., Wang, C., Luo, F., Yu, X., Zhang, W., Li, J., Zhao, D., Xu, D., Gong, Q., et al., 2020. Clinical characteristics and intraventricular vertical transmission potential of covid-19 infection in nine pregnant women: a retrospective review of medical records. Lancet 395 (10226), 809–815.

Das, S.K., Bebortta, S. Geospatial data analytics-a deep learning perspective. Das, S.K., Pant, M., Bebortta, S. Geospatial data analytics: A machine learning perspective. Available at SSRN 3599656.

DeFries, R., Agarwala, M., Baquie, S., Choksi, P., Dogra, N., Preetha, G., Khanwilkar, S., Mondal, P., Nagendra, H., Urdal, J., Paris, J.-F., Zhang, Z.-J., Cheng, Y., 2019. Epidemiology of 2019 novel coronavirus in jiangsu province, china. Computer & Electrical Science: A survey of methods. Wiley Interdiscip. Rev. Data Min. Knowl. Disc. 9 (1), 193–214.

Duijvestijn, P.J.J., de Graaf, P., van der Pal, M.N., van Dijk, R., van Dijk, C.R., van der Zee, K.A., 2020. Potential of covid-19 infection in nine pregnant women: a retrospective review and clinical characteristics study. Lancet 395 (10226), 809–815.

Guliyev, H., 2020. Determining the spatial effects of covid-19 using the spatial panel data model. Spatial Stat. 100443.

Jadhav, J., Surampudi, S.R., Agarirsamy, M. Convolution neural network based infection transmission analysis on covid-19 using gis and covid data materials. Mater. Today: Proc. 102389.

Khan, F.M., Kumar, A., Puppala, H., Kumar, G., Gupta, R. Projecting the criticality of covid-19 transmission in india using gis and machine learning methods. J. Saf. Sci. Resilience.

Kong, B., Chai, B., Zhao, J., 2021. A review on current status of covid-19 cases in maharashtra state of india using gis: a case study. Spat. Inform. Re., 21, 223–229.

Kosovac, A., Muharemovic, E., Begovic, M., Simic, E., 2020. Determining the location of postal centers in bhikk using machine learning clustering method and gis. In: 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, pp. 1318–1322.

Lakhani, A. Which melbourne metropolitan areas are vulnerable to covid-19 based on age, disability and access to health services? using spatial analysis to identify service gaps and inform delivery. J. Pain Symptom Manage.

Mollalo, A., Vallesi, B., Rivera, K.M., 2020. Gis-based spatial modeling of covid-19 incidence rate in the continental united states. Sci. Total Environ. 728, 138884.

Pani, S.K., Lin, N.-H., Ravindraabu, S., 2020. Association of covid-19 pandemic with meteorological parameters over singapore. Sci. Total Environ. 140112.

Radanliev, P., De Roure, D., Walton, R. Data mining and analysis of scientific research data on covid 19 mortality, immunity, and vaccine development-in the first wave of the covid-19 pandemic. Diabetes Metabolic Syndrome: Clin. Res. Rev.

Rashed, E.A., Kodera, S., Gomez-Tames, J., Hirata, A. 2020. Influence of absolute humidity, temperature and population density on covid-19 spread and decay durations: Multi-preference study in japan. Int. J. Environ. Res. Public Health 17 (15), 5354.

Rodriguez, M.Z., Comin, C.H., Casanova, D., Bruno, O.M., Amancio, D.R., Costa, L.d.F., Yahya, B.M., Yahya, F.S., Thannoun, R.G., 2021. Covid-19 prediction analysis using deep convolution neural network. Mater. Today: Proc.

Shekhar, S., Evans, M.R., Kang, J.M., Mohan, P., 2011. Identifying patterns in spatial information: A survey of methods. Wiley Interdiscip. Rev. Data Min. Knowl. Disc. 1 (3), 193–214.

Tirnuc, C., Gómez-Pérez, D., Balčiūnas, J.I., Montaña, J.L., 2018. Global optimality in k-means clustering. Inf. Sci. 439, 79–94.

Vasanth, R.N., Patil, S. Indian publications on sars-cov-2: A bibliometric study of research production. Mater. Today: Proc.

Wang, K.W., Gao, J., Wang, H., Wu, X.-L., Yuan, Q.-F., Cao, Y.-F., Zhang, Z.-J., Cheng, Y., 2019. Epidemiology of 2019 novel coronavirus in jiangsu province, china. Computer & Electrical Science: A survey of methods. Wiley Interdiscip. Rev. Data Min. Knowl. Disc. 9 (1), 193–214.

Yaah, B.M., Yahiya, F.S., Thannoun, R.G., 2021. Covid-19 prediction analysis using artificial intelligence procedures and gis spatial analytic: a case study for iraq. Appl. Geomatics. 1–11.