Spatial modeling of COVID-19 transmission in Bangladesh

Showmitra Kumar Sarkar[1] · Khondaker Mohammed Mohiuddin Ekram[1] ·
Palash Chandra Das[1]

Abstract The purpose of the research was to investigate and identify the demographic risk factors behind the transmission of COVID-19 in Bangladesh based on spatial and statistical modeling. Number of COVID-19 confirmed cases per thousand population as the dependent variable and nine demographic explanatory variables were considered. Different spatial (i.e., Spatial Lag and Spatial Error Model) and non-spatial (Classic Model) regression techniques were employed in the research to detect the geographical relevance of potential risk factors affecting the transmission of COVID-19. Results indicate that population density was crucial for explaining the pattern of COVID-19 transmission in Bangladesh. Spatial Auto-correlation suggests that the spatial pattern of population density were significantly clustered at a confidence interval of 95%. Again, the regression analysis also shows that population density is an influential determinant for the propagation of COVID-19 in Bangladesh, with densely populated districts like Dhaka and Narayanganj also being among the worst affected areas. The findings of this research will help the government agencies and communities for effective and well-informed decision making in order to develop and implement strategies to contain the further spread of COVID-19 in Bangladesh.

✉ Showmitra Kumar Sarkar
mail4dhrubo@gmail.com

Khondaker Mohammed Mohiuddin Ekram
mohiuddinekram@gmail.com

Palash Chandra Das
chandra.palash46@gmail.com

[1] Department of Urban and Regional Planning, Khulna University of Engineering & Technology (KUET), Khulna 9203, Bangladesh

1 Introduction

Novel coronavirus (COVID-19) is rapidly infecting people around the globe [1]. As of June 11, 2020, 7,273,958 people were infected and 413,372 died throughout the world [2]. COVID-19 has first emerged in Wuhan (Hubei, China) in December 2019 [3, 4], and rapidly spread in the European countries in early 2020 [5]. The World Health Organization (WHO) declared the epidemic as a Public Health Emergency of international concern on 30 January 2020 [6] and as a pandemic on 12 March 2020 [7]. In recent times, like in the European countries, COVID-19 is spreading rapidly in the South-East Asian countries (i.e., India, Bangladesh, and Indonesia, etc.) with 407,414 confirmed cases as of 11 June 2020 [2]. The World Health Organization (WHO) confirmed community transmission of COVID-19 in two countries of this region (on 1 June 2020 in Bangladesh [2] and on 13 April 2020 in Indonesia [8]). Community transmission of such an infectious disease (i.e., COVID-19) in a densely populated country like Bangladesh might be severe for every citizen.

Bangladesh is already one of the world's most vulnerable country to climate change induced natural disasters [9, 10]. In terms of the risk of transmission of COVID-19, Haider et al. [11] have classified Bangladesh into the 4th quantile (highest risk) based on the destinations of the passengers travelling from four major cities of China from January onwards. The first COVID-19 patient was confirmed on 8 March 2020 [12] and total confirmed cases were 74,865 as of 11 June 2020. In early March 2020, immigrants (mostly from China, Italy, USA, UK, etc.)

returned to Bangladesh, allegedly carrying the coronavirus and eventually infecting more people. The unwillingness of immigrants with regards to following quarantine and social distancing rules could be one of the major causes of community transmission of COVID-19 [13]. The prevailing demographic (i.e., high population density) and socio-economic (i.e., extreme poverty) structure of the country meant that lockdown measures imposed by the government was not effective enough in containing the spread of COVID-19 countrywide [14]. The government of Bangladesh has taken several measures including the declaration of public holidays, nation-wide lockdown and massive awareness-raising campaigns, but with an increasing number of people being affected and a rising death toll, the overall COVID-19 scenario has shown little signs of improvement till date. It is not obvious which factors are actually aggravating the transmission of COVID-19 in Bangladesh as the causal factors for its spread have not yet been clearly identified.

Spatial modeling might be a useful tool for investigating COVID-19 transmission [15–18]. Sarkar [19] has measured the district-level susceptibility of Bangladesh to COVID-19 using multi-criteria evaluation technique. In this situation, spatial modeling based on demographic features might be a useful tool for investigating and explaining the pattern of COVID-19 transmission. This research investigates the relation of COVID-19 transmission with different demographic factors. Nine causal factors regarding demography (for e.g., population, population density, number of poor people, etc.) were selected based on literature review for generating spatial models.

Based on spatial correlation and model results, the most crucial causal factors behind COVID-19 transmission in Bangladesh were identified. Finally, the spatial and non-spatial impact of the identified causal factors on COVID-19 transmission were explained based on mapping and statistical analysis. The findings of the research might help Government agencies in effective decision-making for reducing the risk of COVID-19. The specific objective of this study was to investigate and identify the demographic risk factors behind the transmission of COVID-19 in Bangladesh based on spatial and statistical modeling.

2 Data and methods

2.1 Description of the study area

Bangladesh (containing 64 districts) was selected as the study area for the research (Fig. 1). The South Asian country lies in the Ganges–Brahmaputra–Meghna Delta connected with the Bay of Bengal. Bangladesh shares a land border with two countries (India and Myanmar) and is also surrounded by neighboring Nepal, Bhutan, and China. The country is ranked 8th based on population (161 million) [20] and 7th based on population density (1,107 people per square kilometer) [21]. It is one of the fastest-growing middle-income countries in the world [22].

2.2 Description of materials

Demographic and socio-economic data from Bangladesh Bureau of Statistics (BBS), COVID-19 data from the Institute of Epidemiology Disease Control and Research (IEDCR) and physical features such as boundary, river, etc. of Bangladesh from Geological Survey of Bangladesh (GSB) were collected as secondary data for the research (Table 1). The findings of the research were validated through online based stakeholder interviews including 30 key persons of government agencies.

2.3 Spatial autocorrelation and spatial modeling

Nine causal factors (i.e., population, population density, people aged ≥ 60, number of urban people, number of working-age people, number of industrial workers, number of poor people, access to television, and literacy rate) were selected for the research. The count data were transformed into scale data for better representation and further analysis. Spatial autocorrelation is the basic component in the field of spatial analysis [23]. Spatial autocorrelation helps to know the presence of systematic spatial variation based on both feature locations and feature values simultaneously [24]. Spatial autocorrelation evaluates spatial patterns (i.e., clustered, dispersed, or random) of individual entities [25].

The following binary pairing spatial weight matrix was used to demonstrate the spatial relationship of 64 districts:

$$W = \begin{bmatrix} \omega_{1,1} & \cdots & \omega_{1,64} \\ \vdots & \ddots & \vdots \\ \omega_{64,1} & \cdots & \omega_{64,64} \end{bmatrix}$$

where $\omega_{i,j}$ represents the proximity relationship between the ith and jth district, i, j = 1, 2, …, 64.

Either adjacency criterion or distance criterion can be used to measure this relationship [26]. This study used Rook neighborhood to calculate statistics for local indicators of the number of COVID-19 confirmed cases in the study area. The adjacency binary matrix for this study is:

$$\omega_{i,j} = \begin{cases} 1, & i \text{ is adjacent to } j. \\ 0, & \text{otherwise.} \end{cases} \quad i,j = 1, 2, 3, \ldots, 64, \quad i \neq j$$

The ith district has no adjacency relationship with itself, which means $\omega_{i,i} = 0$. [27]

The research used local univariate and bivariate Moran's I statistics for spatial pattern analysis (i.e., spatial

Fig. 1 The study area (Bangladesh, highlighting the location of districts)
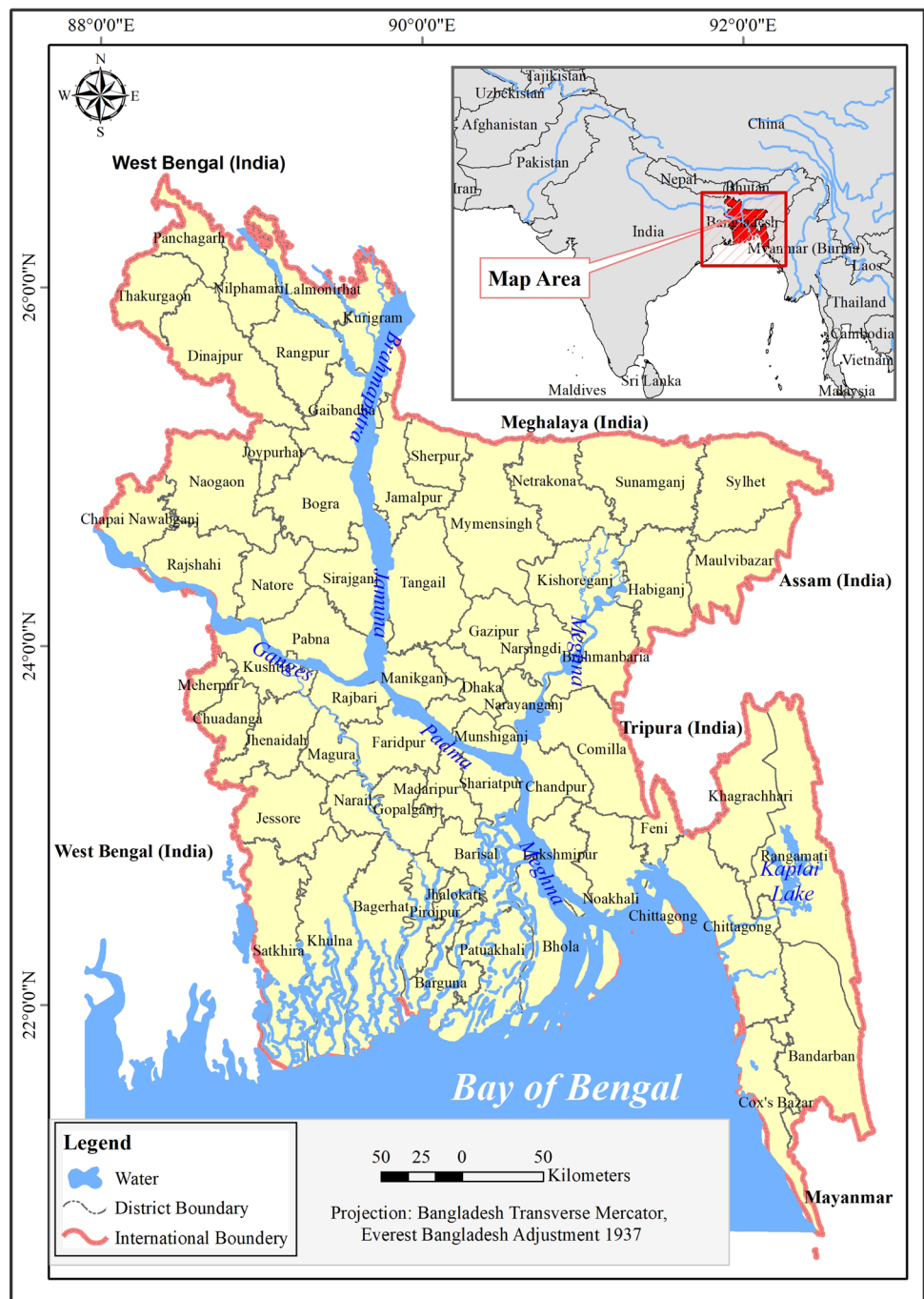
autocorrelation). The univariate Moran's I was calculated (Eq. 1) according to Moran [28].

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} \quad (1)$$

where n = the number of spatial entities indexed by i and j; w_{ij} = their spatial weight; and S_0 = aggregation of all the spatial weights.

The bivariate Moran's I is a generalized format of univariate Moran's I used to measure the spatial correlation of two variables (Eq. 2) [29].

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(z_i^A\right) \left(z_j^B\right)}{\sum_{i=1}^{n} \left(z_i^A\right)^2} \quad (2)$$

where, z^A and z^B = the deviations from the mean for variable A and B.

Table 1 Description of data used in the research

Data with units	Mean	Minimum	Maximum	Standard deviation	Source with year
COVID-19 confirmed cases (1/100,000)	12.32	1.29	151.30	21.20	Institute of Epidemiology Disease Control and Research (2020)
Population (1/100,000)	22.51	3.88	120.44	17.29	Bangladesh Bureau of Statistics (2011 Census)
Population density (person/km^2)	1107.69	86.00	8111.00	1030.07	
People age $\geq$ 60 (1/10,000)	499.03	270.04	656.81	81.45	
No. of urban people (1/10,000)	1784.38	880.00	7710.00	1015.40	
No. of working-age people (1/10,000)	5996.13	5281.12	7069.64	385.97	
No. of industrial workers (1/10,000)	312.21	64.39	2278.14	346.40	
No. of poor people (1/10,000)	3225.15	358.67	6372.18	1194.75	
Access to television (1/10,000)	3891.55	427.46	7074.08	1305.79	
Literacy rate (percentage)	54.67	37.50	73.70	7.70	
District boundary (shapefile)	–	–	–	–	Geological Survey of Bangladesh (2020)

Bivariate Moran's I was used for investigating spatial autocorrelation among causal factors with COVID-19 confirmed cases and univariate Moran's I COVID-19 confirmed cases.

Again, this research used local G coefficient which can identify the aggregated area more accurately [26]. Moreover, the spatial distribution pattern of G coefficient can detect if a regional unit is in a high-value cluster or a low-value cluster compared with the Moran index [27]. So, to detect spatial dependence in a small area local G statistic G_i for the COVID-19 confirmed cases was calculated. The formula to calculate G_i for the ith district is as follows (Eq. 3).

$$G_i = \frac{\sum_{j=1}^{n} \omega_{i,j} x_j}{\sum_{j=1}^{n} x_j} \quad i,j = 1, 2, 3, \ldots, n \quad (3)$$

where n = 64, x_j denotes the number of confirmed COVID-19 cases in jth district, $i \neq j$, $\omega_{i,j}$ represents the neighboring relationship between the ith and the jth districts.

The formula to calculate the local G coefficient G_i normalized statistic Z (G_i) is in Eq. 4.

$$Z(G_i) = \frac{G_i - E(G)}{\sqrt{Var(G)}} \quad (4)$$

In this formula, E(G) indicates the mean of G and Var(G) indicates the variance of G.

Spatial regression analysis was employed in order to further explore the governing factors of the COVID-19 cases of the study unit. Spatial regression analysis is normally accompanied by three models: Classical Regression Model, Spatial Error Model and Spatial Lag Model [27]. Classical Regression Model is expressed in Eq. 5 [30, 31].

$$Y = \alpha + \beta X + \varepsilon, \quad \varepsilon \sim iid\,N(0, \sigma^2) \quad (5)$$

where α is the intercept term, β represents the regression coefficients, ε denotes the error term (independent identical distribution (iid) is assumed), Y and X represent the dependent and independent variables respectively.

From the Classical Regression test diagnostic Lagrange multiplier test (LM test) was used to test the spatial dependence [32]. Moreover, this study incorporated different spatial study variables into a Spatial Lag Model (SLM) and a Spatial Error Model (SEM) to compensate/account for further spatial variation and establish the ideal spatial regression model.

SLM includes spatial lag variables to justify the spatial dependence triggered by externalities and spillover effects [32]. The formula used for SLM in this study is expressed in Eq. 6 [33].

$$Y = \alpha + \rho WY + \beta X + \varepsilon, \quad \varepsilon \sim iidN(0, \sigma^2 I) \quad (6)$$

where α denotes the intercept term, ρ represents spatial autoregressive coefficient, WY represents the spatial lag variable, β is the regression coefficient, X denotes the explanatory regression variable and ε is the error term vector.

An additional spatial lag variable is considered in the SLM process. A relationship matrix of the research and neighboring samples are incorporated into the regression model as well, where ρ stands as the spatial autoregressive coefficient. Finally, whether or not the variable equals 0 ($\rho \neq 0$) is assessed to find out if SLM is affected by any form of spatial autocorrelation [34].

Again, to remove the interference of spatial autocorrelation and get accurate estimation SEM is used. In Spatial

Error Model (SEM), the spatial weight matrix is multiplied by the spatial error coefficient λ to calculate the error term. Then, whether or not the spatial error coefficient λ is significant and equals 0 (λ, 0) are assessed to find out if there is any spatial autocorrelation in the SEM. The model is expressed in Eq. 7 [35].

$$Y = \alpha + \beta X + \varepsilon, \quad with \varepsilon = \lambda W \varepsilon + \xi, \quad \xi \sim iidN(0, \sigma^2) \tag{7}$$

where α represents the intercept term, β denotes the regression coefficient, X is the explanatory regression variable, ε represents the error term vector, λ represents the spatial error coefficient, W is the spatial weight matrix and ξ is the modified error term.

Significant causal factors of COVID-19 transmission were selected based on a best fit model and spatial autocorrelation of causal variables. Finally, spatial relations of causal factors and COVID-19 confirmed cases were investigated by mapping.

3 Results and discussion

Data cleaning, data transformation, and necessary treatment for missing values were conducted before performing the aforementioned analyses. Spatial autocorrelation using Local Moran's I (univariate and bivariate) was assessed to determine the distribution pattern of COVID-19 and the aforementioned spatial regression models were formulated to ascertain the spatial relationship of the demographic risk factors of COVID-19 based on the data of all 64 administrative districts of Bangladesh.

Table 1 provides a statistical description of the parameters considered in this study. The maximum number of COVID-19 confirmed cases was found in Dhaka and the minimum in Sirajganj. Dhaka might have been affected more due to a higher population (1st rank with 12,043,977 population), population density (1st rank with 8,111 people per square kilometer), and the fact that this district is the center of national and international collaborations of Bangladesh. Senior citizens (i.e., people age ≥ 60) are more vulnerable to COVID-19, but the Dhaka has the minimum number of senior citizens per 10,000 population (i.e., 270 approx.) while Pirojpur has the maximum (i.e., 657 approx.). The number of urban people, the number of working-age people, and the number of industrial workers per 10,000 population are maximum in Dhaka and minimum in Brahmanbaria. The minimum number of poor people per 10,000 population (defined as those living below the poverty line) is in the Kushtia (i.e., 359 approx.) and the maximum is in Kurigram (i.e., 6372 approx.).

Pirojpur has the highest literacy rate (i.e., 73.7%) while Bandarban has the lowest (i.e., 37.5%).

The result for univariate local Moran's statistic of confirmed COVID-19 cases was found to be 0.261 which is significant at a 5% level of significance, indicating a positive spatial autocorrelation across the country (Fig. 2j). So, there is some sort of spatial relationship of a particular district with the neighboring districts in terms of COVID-19 confirmed cases.

Again, spatial Auto-correlation of the above-mentioned parameters using bivariate local Moran's I statistic is illustrated in Fig. 2 to identify the parameters that have significantly affected the number of COVID-19 confirmed cases spatially. According to the p value, two parameters (i.e., population density and number of industrial workers) were found to be significant at a confidence interval of 95%. Spatial patterns of these two parameters were clustered and the parameters are spatially auto-correlated with COVID-19 cases of the neighboring districts (Fig. 1b, f). The rest of the parameters were random in spatial patterns and not significant at a confidence interval of 95%.

Figure 3 exhibits that there are four districts in HH (High High) region which are statistically significant ($p < 0.05$), including Dhaka, Munshiganj, Narayanganj and Gazipur. It indicates that the number of COVID-19 confirmed cases in these districts is high and this number in neighboring districts is still high. In LL (Low Low) region, there are six districts with statistical significance ($p < 0.05$), including Khulna, Jessore, Barguna, Kushtia, Natore and Jamalpur. It represents that the number of confirmed COVID-19 cases in these districts is low and the same holds true for the neighboring districts as well.

There are two districts with statistical significance ($p < 0.05$) in LH (Low High) region, including Tangail and Faridpur. It indicates that the number of COVID-19 cases in these districts is low, but this number in the neighboring districts is high. There is only one statistically significant ($p < 0.05$) district in HL (High Low) region named Rangpur, indicating that the number of COVID-19 cases in this district is high surrounded by the neighboring districts with a low COVID-19 cases. Among these regions, High–High and Low–Low regions represent positive spatial autocorrelation and Low–High and High–Low regions represent negative spatial autocorrelation.

Again, in Fig. 4, Dhaka, Munshiganj, Narayanganj, Gazipur, Tangail and Faridpur are clustered areas with high number of COVID-19 cases. Khulna, Jessore, Barguna, Kushtia, Natore, Rangpur and Jamalpur are clustered areas with a low number of COVID-19 cases.

The High-Low clusters and Low–High clusters in LISA map became Low cluster and High clusters respectively in the Local G Coefficient map. It is because unlike Local Moran's I, local G coefficient does not consider spatial

Fig. 2 Local Moran's I scatter
plots. Bivariate local Moran's I
of COVID-19 confirmed cases
with (**a** population, **b** population
density, **c** people age $\geq$ 60,
d number of urban people,
e number of working-age
people, **f** number of industrial
workers, **g** number of poor
people, **h** access to television,
i literacy rate) **j** univariate local
Moran's I of COVID-19
confirmed cases
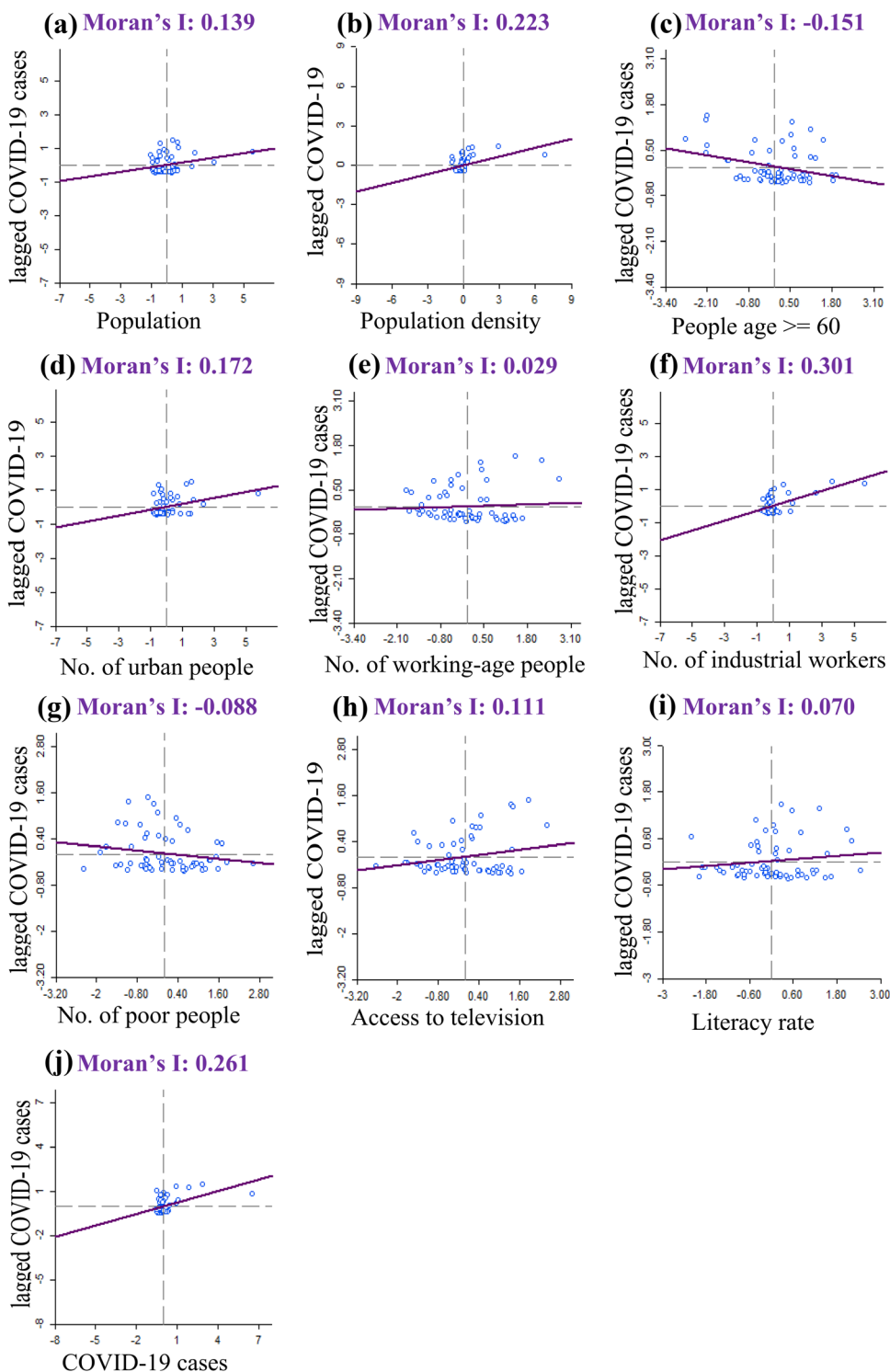
outlier. Again, Gi index works on the principle of spatial aggregation to identify hot spots where concentrations of high and low values result in high and low Gi values respectively. That is why instead of spatial variation in cluster (High–High, Low–Low, Low–High and High–Low)

spatial aggregation of High and Low areas are displayed in Gi cluster map [36].

Bangladesh is the world's eighth-most densely populated country (population density is 1,108 people per square kilometer) [37]. Sarkar [19] identified that population density has an impact on COVID-19 transmission in
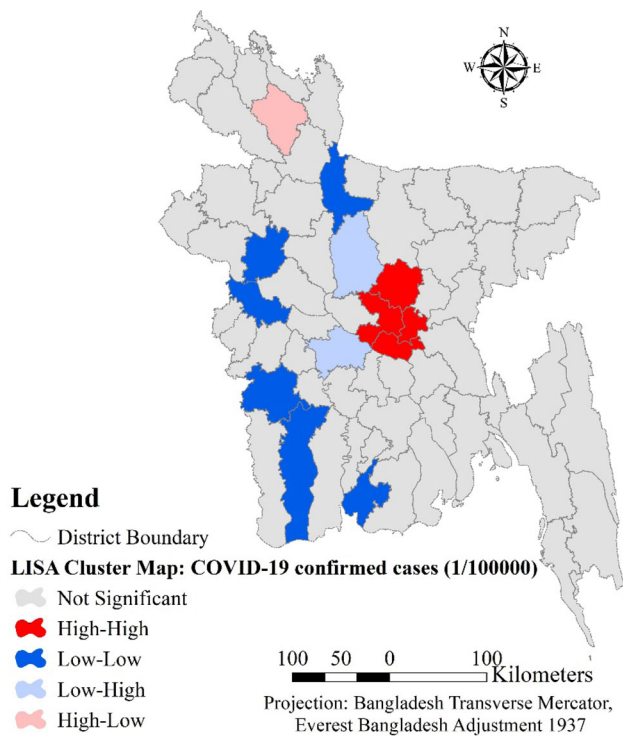
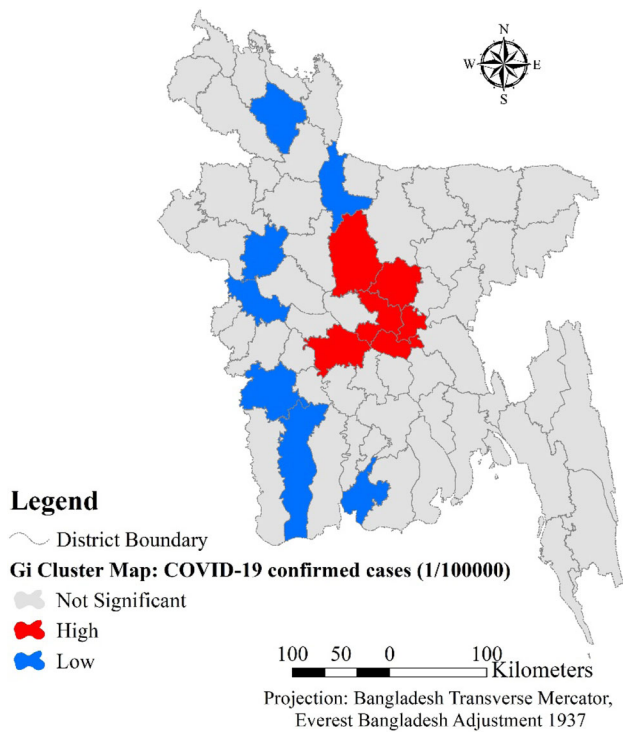Fig. 3 Results of local Moran index visualization analysis

Fig. 4 Results of local G coefficient visualization analysis

Bangladesh. Difficulty in maintaining social distance in densely populated areas like Dhaka might have fostered transmission. On the other hand, community transmissions

in densely populated areas might force the rapid spreading of COVID-19. Based on Figs. 3 and 4 and results associated with them, it can be claimed that the central zone of Bangladesh (Dhaka and adjacent districts) is more vulnerable to COVID-19 due to its higher population density.

Dhaka being the center of majority of the economic activities in Bangladesh has always attracted a huge population to it and its neighboring districts. The sheer number of people that work together in various industries in and around Dhaka, and failure to maintain proper hygiene there, might have been responsible for COVID-19 transmission among them and consequently among other people of these regions. On the other hand, most of the large and medium industries (for e.g., the apparel industries) of this region are located together in a cluster and as a result, maintaining social distance is difficult in those areas. Industrial workers working in Dhaka and Narayanganj (most affected districts during early COVID-19 transmission) eventually left for their native homes following a nation-wide lockdown. However, miscommunication between Bangladesh Garment Manufacturers and Exporters Association (BGMEA) and government bodies throughout April regarding reopening the industries, only worsened the situation. A vast number of workers had to return to their workplace twice amidst a nationwide lockdown, using local public transport and even on foot from fear of being laid off and uncertainty over receiving their due wages [38, 39]. These unwanted and avoidable migrations might have influenced the spread of Covid-19 in the Dhaka and its adjacent areas.

To add to the misery, the lack of awareness regarding safety measures resulting from the low level of education, unhealthy and unhygienic lifestyle owing to living in slums and squatters, inability to effectively maintain social distancing due to denser and poorer living conditions and a general lack of access and inability to afford basic hygiene materials and Personal Protective Equipment (PPE) have worsened the situation for the vulnerable poor population living in Dhaka and its adjoining districts.

The p-value of Bartlett's Test of Sphericity is greater than 0.05 which indicates the absence of multicollinearity among the explanatory variables. Hence, all of these parameters were used to model the causal effects on the number of COVID-19 confirmed cases. The modeling results are illustrated in Table 2 and it is found that only population density has significantly (p value less than 0.05) influenced the number of COVID-19 confirmed cases when a classic regression model is applied on the dataset.

Besides, the univariate local Moran's I result shows a positive spatial autocorrelation with a p value of less than 0.05 indicating rejection of the null hypothesis of spatial independence of residuals of the classic method. Again, the low probabilities ($p < 0.05$) of Breusch–Pagan test,

Table 2 Variables and modeling results

Variables		Classical regression model (CRM)	Spatial lag model (SLM)	Spatial error model (SEM)
Constant	Coefficient	0.626072	− 18.3936	− 13.5801
	Probability	0.9782	0.3751	0.5933
Population	Coefficient	− 0.0775117	− 0.0392796	− 0.0648459
	Probability	0.4506	0.6582	0.4236
Population density	Coefficient	0.016737	0.0159396	0.0171707
	Probability	0.0000	0.0000	0.0000
People age ≥ 60	Coefficient	− 0.0356168	− 0.039213	− 0.0461551
	Probability	0.0949	0.0303	0.0549
No. of urban people	Coefficient	0.0029676	0.00264053	5.99E−05
	Probability	0.1312	0.1159	0.9745
No. of working age people	Coefficient	0.000895814	0.00406054	0.00481118
	Probability	0.8330	0.2821	0.2972
No. of industrial workers	Coefficient	− 0.000259399	− 0.00575691	− 0.00254137
	Probability	0.9538	0.1824	0.5808
No. of poor people	Coefficient	− 0.00103537	− 0.000795051	− 0.00125877
	Probability	0.3094	0.3677	0.1914
Access to television	Coefficient	− 0.00215033	− 0.00246411	− 0.00112741
	Probability	0.0930	0.0260	0.2994
Literacy rate	Coefficient	0.252333	0.270138	0.209675
	Probability	0.1788	0.0918	0.2793
Lag coeff. (Rho\ρ)	Coefficient	–	0.284915	–
	Probability	–	0.0060	–
Lambda (λ)	Coefficient	–	–	0.519562
	Probability	–	–	0.0001

Koenker–Bassett test and White test indicate presence of heteroskedasticity in the residual of the used dataset. This error variance might well be caused by the spatial dependence in the data. So, it is obvious that the classic model is not appropriate to explain the spatial relationship of the problem under investigation. Again, as Moran's I do not necessarily give a precise impression of the spatial structure, it is indispensable to find out which one of the spatial lag or spatial error process is responsible for the autocorrelation existent in the residuals.

For further investigation, spatial lag and spatial error models were employed on the dataset and two parameters (i.e., population density, number of people aged ≥ 60) were found to be statistically significant at a confidence interval of 95% (i.e., p value < 0.05) in Spatial Lag Model. Among them, population density and number of people aged ≥ 60 have positively and negatively influenced the number of COVID-19 cases respectively. In the spatial lag model, the lag coefficient for confirmed COVID-19 confirmed cases (Rho/ρ) was estimated as 0.284915 (p value 0.0060) which is significant at a 5% level of significance, indicating that the number of COVID-19 confirmed cases

in one district depends on the cases of its neighboring districts. A significant coefficient for the variables implies that the number of COVID-19 confirmed cases in a given district depends on the change in explanatory variables in the same district controlling the spatial lag of the dependent variable. This result supports the result obtained using Moran's I statistics where some sort of spatial relationship was observed.

Again, the spatial error model comes in handy to justify the spatial variation that the explanatory variables in this study couldn't explain due to the presence of error terms. Population density was found to be positively significant at a confidence interval of 95% (i.e., p value < 0.05) in Spatial Error Model. In the spatial error model, the spatial autoregressive term (λ) was estimated as 0.51956 with a p value of 0.0001 (significant at 5% level of significance). This suggests that the spatial dependence present in the residuals of the classic model was due to a geographic clustering of overleaped variables or variables which are not considered in the modeling of the risk factors for the number of COVID-19 confirmed cases using the classic regression model.

Furthermore, to identify the best fit model regression results for the above three models are shown in Table 3. The values of Akaike Information Criterion (AIC) and Schwarz Criterion (SC) of Spatial Error Model are all smaller and the values of R^2 and Log L are all bigger than the Spatial Lag Model. Hence, Spatial Error Model is a better fit to the study dataset. Again, spatial dependence test was employed to find out the appropriate spatial process to explain the spatial dependence of the modelled residuals (Table 4).

Among Lagrange Multiplier (lag) and Lagrange Multiplier (error) test, Lagrange Multiplier (lag) test is found significant ($p < 0.05$). Here, Spatial Lag Model (SLM) is found to be more appropriate to model the spatial relationship of the study dataset [40].

This means that these two models do explain the spatial autocorrelation found in the residuals of the classic model. In both of these spatial models, population density was found positively significant at a confidence interval of 95% (i.e., p value < 0.05). Hence, population density is spatially correlated (positively) with COVID-19 transmission, meaning that, higher the values of this parameter, higher the number of COVID-19 confirmed cases.

Bangladesh is the world's eighth-most densely populated country [21]. Higher population density was found to be responsible for higher COVID-19 transmission (Fig. 5). The research classified the 64 districts of Bangladesh into 4 classes according to population density (Fig. 6). Dhaka has 8111 people per square kilometer and for Narayanganj, the figure is 4139. On the other hand, COVID-19 confirmed cases per 100,000 were higher in these two districts as well (i.e., 151 for Dhaka district and 73 for Narayanganj district). COVID-19 spread rapidly in these two districts and community transmissions raised the severity. Gazipur, Narsingdi, and Comilla districts have higher population density (i.e., 1500–3000 people per square kilometer) and consequently higher COVID-19 confirmed cases. The

Table 3 Comparison of regression results of three models

Index	CRM[a]	SEM[b]	SLM[c]
R^2	0.839357	0.881285	0.878209
Log L[d]	− 222.812	− 220.335267	− 219.497
AIC[e]	465.623	460.671	460.994
SC[f]	487.212	482.259	484.742

[a]Classical regression model

[b]Spatial error model

[c]Spatial lag model

[d]Log likelihood

[e]Akaike information criterion

[f]Schwarz criterion

Table 4 Results of the spatial dependence test of COVID-19 confirmed cases

TEST	MI/DF	Value	Probability
Lagrange multiplier (lag)	1	6.5722	0.01036
Robust LM (lag)	1	4.5941	0.03208
Lagrange multiplier (error)	1	2.051	0.1521
Robust LM (error)	1	0.0728	0.78723

Chittagong district (the port city and influential financial center) has 33 COVID-19 confirmed cases per 100,000 with a population density of 1421 people per square kilometer. The southern coastal districts (i.e., Satkhira, Khulna, Bagerhat, Barguna, and Patuakhali) have a lesser number of COVID-19 confirmed cases with less population density (less than 1000 people per square kilometer). Population density is relatively higher (1001–1500 less than 1000 people per square kilometer) in the northern region of Bangladesh and COVID-19 confirmed cases were also found to be higher in those districts. Spreading of COVID-19 was lower in the hill tract districts (i.e., Khagrachhari, Rangamati, and Bandarban) having lower population density.

4 Conclusion

This research aimed at identifying the demographic factors that were responsible for the transmission of COVID-19 in Bangladesh. Three spatial and non-spatial models (i.e., Classic Model, Spatial Lag Model, and Spatial Error Model) were employed to study the nature of the relationship of these factors with COVID-19 transmission. As found by Spatial Auto-correlation, spatial patterns of factor (population density) were significantly clustered at a confidence interval of 95%. The regression models illustrate that population density has influenced the propagation of COVID-19 in Bangladesh. This study analyzed the demographic data pertaining to the factor for the 64 districts of Bangladesh and showed that a higher value of each of the factors was associated with a higher number of COVID-19 confirmed cases. For instance, results show that Dhaka and Narayanganj, the two most densely populated districts of Bangladesh are also placed 1st and 3rd respectively in terms of COVID-19 confirmed cases. It can be claimed that the central zone of Bangladesh (Dhaka and adjacent districts) is more vulnerable to COVID-19 due to its higher population density.

The findings of this research can be beneficial for policymakers as well as communities in order to effectively design strategies to prevent the further spreading of

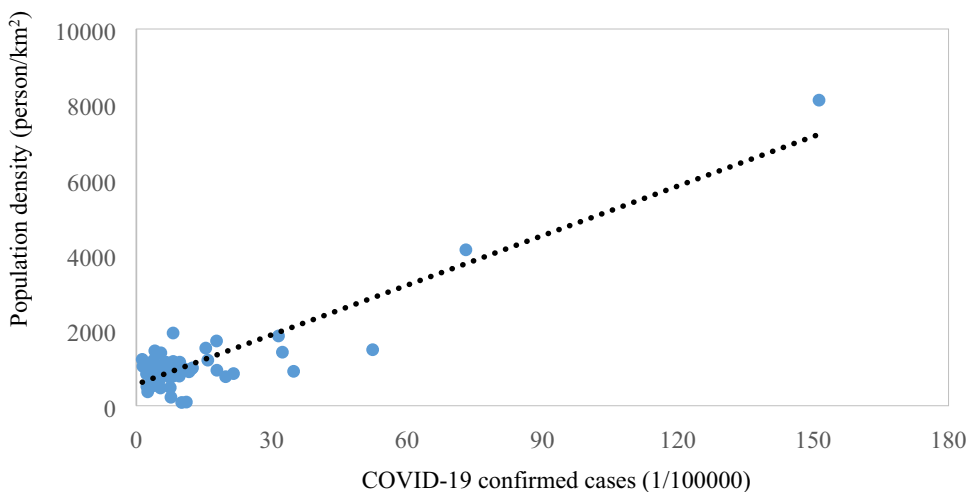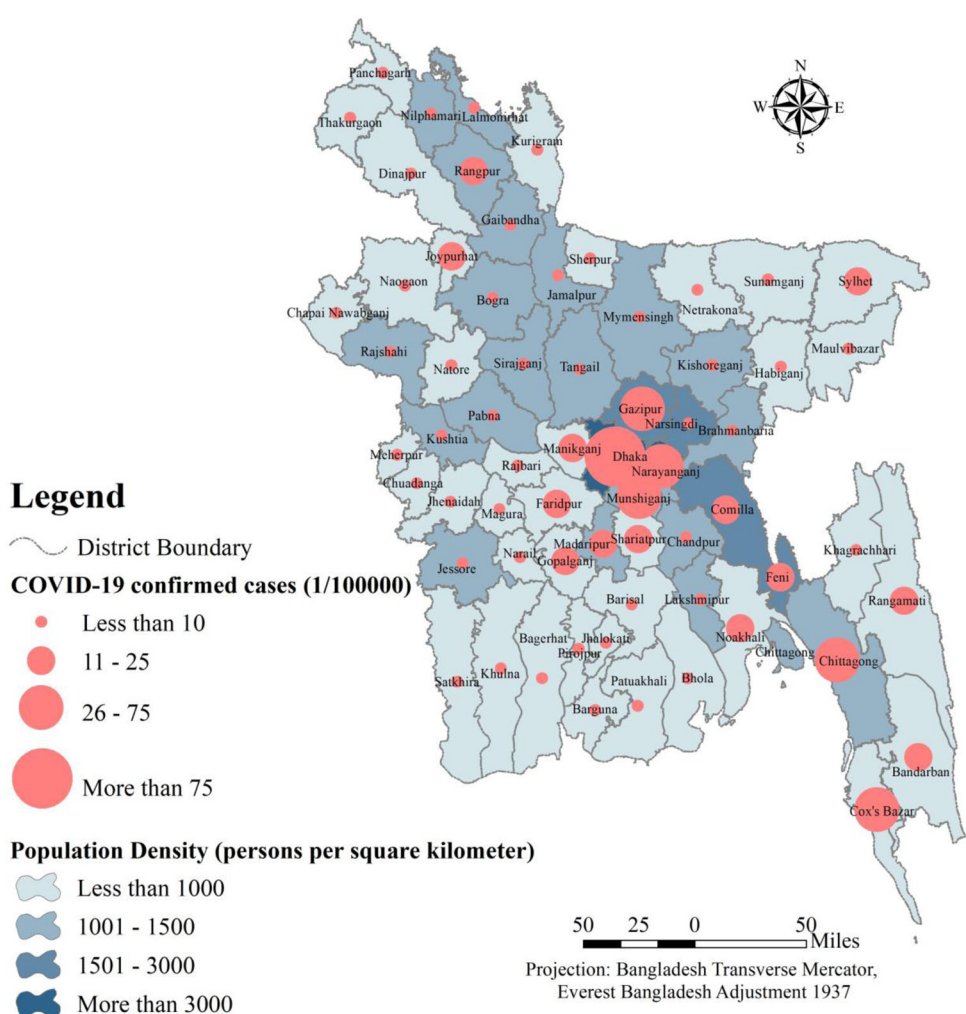**Fig. 5** Population density and COVID-19 confirmed cases

Fig. 6 Map showing population density and COVID-19 confirmed cases

COVID-19 in Bangladesh. The government of Bangladesh has already withdrawn the country-wide shut down and is now trying to contain the virus by imposing zone-wise lockdown measures throughout the country. The spatial models presented in this study can providing crucial insights to the decision making and implementation process, such as, making better-informed decisions in identifying transmission hotspots based on population density

and streamlining resources and efforts to implement prevention measures in those areas for gradually abating and eventually terminating the infection chain.

This study ignores if there are spillover effects in one area to a certain extent. Besides; only nine explanatory variables were considered in this study, although there are a lot of influencing factors. Further research endeavors should address the issues above to get a better idea about the problem under study. The research, however, is limited only to facts and figures obtainable contemporarily of this global pandemic. Also, the number of suspected cases wasn't taken into consideration in this research, which would have yielded more aggravating results, albeit with limited reliability. Future research can examine a bigger dataset and can also factor in such crucial variables as the spatial prevalence of pre-existing health conditions like diabetes, cardiovascular diseases, etc. in the population under study.

Availability of data and materials Not applicable.

Code availability Not applicable.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interest to declare.

Ethical Approval The authors have no Ethical Approval to declare.

References

1. Tsai, J., & Wilson, M. (2020). COVID-19: A potential public health problem for homeless populations. *The Lancet Public Health, 5,* e186–e187. https://doi.org/10.1016/S2468-2667(20)30053-0.
2. Culp, W. C. (2020). Coronavirus disease 2019. *A and A Practice, 14,* e01218. https://doi.org/10.1213/xaa.0000000000001218.
3. Lai, C. C., Shih, T. P., Ko, W. C., et al. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents, 55,* 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924.
4. Sun, P., Lu, X., Xu, C., et al. (2020). Understanding of COVID-19 based on current evidence. *Journal of Medical Virology, 92,* 548–551. https://doi.org/10.1002/jmv.25722.
5. Verelst, F., Kuylen, E., & Beutels, P. (2020). Indications for healthcare surge capacity in European countries facing an exponential increase in coronavirus disease (COVID-19) cases. *Eurosurveillance, 25,* 1–4. https://doi.org/10.2807/1560-7917.ES.2020.25.13.2000323.
6. Sohrabi, C., Alsafi, Z., O'Neill, N., et al. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery, 76,* 71–76. https://doi.org/10.1016/j.ijsu.2020.02.034.
7. Oxford Analytica (2020). The WHO's COVID-19 pandemic declaration may be late. Emerald Expert Briefings.
8. World Health Organization (2020). Coronavirus disease 2019 (COVID-19) situation report-84, 13th April 2020.
9. USAID (2018). Bangladesh country risk profile.
10. Sarkar, S. K., Rahman, M. A., Esraz-Ul-Zannat, M., & Islam, M. F. (2020). Simulation-based modeling of urban waterlogging in Khulna City. *Journal of Water and Climate Change.* https://doi.org/10.2166/wcc.2020.256.
11. Haider, N., Yavlinsky, A., Simons, D., et al. (2020). Passengers' destinations from China: Low risk of Novel Coronavirus (2019-nCoV) transmission into Africa and South America. *Epidemiology and Infection.* https://doi.org/10.1017/S0950268820000424.
12. World Health Organization (2020). Coronavirus disease 2019 (COVID-19)-situation report-49 (31 March 2020). WHO Bull 2019 (pp. 1–19).
13. Kazi Abdul, M., & Khandaker Mursheda, F. (2020). The COVID-19 pandemic: Challenges and reality of quarantine, isolation and social distancing for the returnee migrants in Bangladesh. *International Journal of Migration Research and Development, 6,* 27–43.
14. Shammi, M., Bodrud-Doza, M., Towfiqul Islam, A. R. M., & Rahman, M. M. (2020). COVID-19 pandemic, socioeconomic crisis and human stress in resource-limited settings: A case from Bangladesh. *Heliyon,.* https://doi.org/10.1016/j.heliyon.2020.e04063.
15. Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics, 38,* 100443. https://doi.org/10.1016/j.spasta.2020.100443.
16. Desjardins, M. R., Hohl, A., & Delmelle, E. M. (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography, 118,* 102202. https://doi.org/10.1016/j.apgeog.2020.102202.
17. Adekunle, I. A., Onanuga, A. T., Akinola, O. O., & Ogunbanjo, O. W. (2020). Modelling spatial variations of coronavirus disease (COVID-19) in Africa. *Science of the Total Environment, 729,* 138998. https://doi.org/10.1016/j.scitotenv.2020.138998.
18. Sarwar, S., Waheed, R., Sarwar, S., & Khan, A. (2020). COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions? *Science of the Total Environment, 730,* 139089. https://doi.org/10.1016/j.scitotenv.2020.139089.
19. Sarkar, S. K. (2020). COVID-19 Susceptibility mapping using multicriteria evaluation. *Disaster Medicine and Public Health Preparedness.* https://doi.org/10.1017/dmp.2020.175.
20. United Nations (2019). World population prospects 2019. In *United Nations Department of economic and social affairs population division 18.% 09.* United Nations Department of Economic and Social Affairs, Population Division.
21. Bangladesh Bureau of Statistics (2011). Population and housing census 2011.
22. The World Bank Economic Reforms Can Make Bangladesh Grow Faster. In: World Bank Press Release. https://www.worldbank.org/en/news/press-release/2018/10/02/economic-reforms-can-make-bangladesh-grow-faster.
23. Getis, A. (2008). A history of the concept of spatial autocorrelation: A geographer's perspective. *Geographical Analysis, 40,* 297–309. https://doi.org/10.1111/j.1538-4632.2008.00727.x.
24. Haining, R. (2003). Spatial data analysis. *Spatial Data Analytics.* https://doi.org/10.1017/cbo9780511754944.
25. Chou, Y. H. (1995). Spatial pattern and spatial autocorrelation. Lecture Notes in Computer Science (including its subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 988, pp. 365–376). https://doi.org/10.1007/3-540-60392-1_24.

26. Wang, T., Xue, F., Chen, Y., et al. (2012). The spatial epidemiology of tuberculosis in Linyi City, China, 2005–2010. *BMC Public Health, 12,* 2005–2010. https://doi.org/10.1186/1471-2458-12-885.

27. Zhang, Y., Wang, X. L., Feng, T., & Fang, C. Z. (2019). Analysis of spatial-temporal distribution and influencing factors of pulmonary tuberculosis in China, during 2008–2015. *Epidemiology and Infection.* https://doi.org/10.1017/S0950268818002765.

28. Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society: Series B (Methodological), 10,* 243–251.

29. Anselin, L., Syabri, I., & Smirnov, O. (2002). Visualizing multivariate spatial correlation with dynamically linked windows. *Urbana, 51,* 61801.

30. Rosen, S. (2019). Hedonic prices and implicit markets: Product differentiation in pure competition. In *Revealed preference approaches to environmental valuation volumes I and II* (pp. 5–26). https://doi.org/10.1086/260169

31. Crespo, R., & Grêt-Regamey, A. (2013). Local hedonic house-price modelling for urban planners: Advantages of using local regression techniques. *Environment and Planning B Planning and Design, 40,* 664–682. https://doi.org/10.1068/b38093.

32. Shekhar, S., & Xiong, H. (Eds.). (2007). *Encyclopedia of GIS.* Springer.

33. Anselin, L. (2007). Spatial econometrics in RSUE: Retrospect and prospect. *Regional Science and Urban Economics, 37,* 450–456. https://doi.org/10.1016/j.regsciurbeco.2006.11.009.

34. Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics, 22,* 453–474. https://doi.org/10.1016/0166-0462(92)90039-4.

35. Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics, 17,* 99–121. https://doi.org/10.1023/A:1007707430416.

36. Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24,* 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

37. Borst, R. A., & McCluskey, W. J. (2008). Using geographically weighted regression to detect housing submarkets: Modeling large-scale spatial variations in value. *The Journal of Property Tax Assessment and Administration, 5,* 21–54.

38. The Daily Star (2020). BGMEA issues reopen timetable. Retrieved from https://www.thedailystar.net/front, https://www.thedailystar.net/frontpage/news/bgmea-issues-reopen-timetable-1896880.

39. The Financial Express (2020). Many of apparel, other factories reopen. Job insecurity, denial of April wages haunt RMG workers. Retrieved from https://thefinancialexpress.com.bd/national/many-of-apparel-other-factories-reopen1587957283?fbclid=IwAR2DYvA81s9g1z9sw9myGyRvbykHixZx6EVU7SnVZbEF2Hp8ueHiz_yzH9Y

40. Anselin, L. (2005). Exploring spatial data with GeoDa: A workbook. Center for spatially integrated social science (pp. 1–244).