Original Research

Determinants of COVID-19 Prevalence Rate in Asia: A study using Spatial Analysis

Shalini Chandra[1], Megha Sharma[2]

[1] PhD,[2] MSc, Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India.

How to cite this article: Shalini Chandra, Megha Sharma. Determinants of COVID-19 Prevalence Rate in Asia: A study using Spatial Analysis. Indian Journal of Public Health Research and Development/Volume 15 No. 2, April - June 2024.

Abstract

This study aims at finding out the important determinants of prevalence rate of COVID-19 in the Asian continent using spatial analysis. The impact of climatic, socioeconomic, demographic, and health status variables on the prevalence rate of COVID-19 is seen through various spatial models such as Spatial lag, Spatial error, Geographically Weighted regression model, and Multiscale Geographically Weighted regression model. The performance of the models is compared under different comparison criteria. It is found that among all, Multiscale Geographically Weighted regression model outperformed other competitive models. Findings also indicate that cardiovascular health, prevalence of smoking habit, human development index, and net migration rate played significant role in defining the prevalence rate of COVID-19 in Asia.

Keywords: Spatial analysis, COVID-19, global Moran's I, Multiscale Geographically Weighted regression.

Introduction

In March 2020, the World Health Organization classified the COVID-19 outbreak as a worldwide pandemic resulting from the SARS-COV-2 virus (an acute respiratory syndrome). Such pandemic outbreaks rapidly spread infectious diseases over large areas, thus, affecting more people in a short time. The novel coronavirus has contaminated millions of people and changed the lives of nearly every human being on the planet. COVID- 19 pandemic not just affected the health sector of the nations but also affected the social, economic, and political aspects worldwide.

Recent studies have identified several demographics, socioeconomic, and environmental factors that contribute to the spread of communicable dis- eases. Age, population density, poverty, average household income, temperature, air pollution, and smoking are some of the factors which regulated The intensity and speed of COVID-19 transmission and its impact pandemic.[1–5] Also, some analysts claimed that the most affected countries had a higher proportion of older people and poor medical facilities.[6,7] Recent studies on COVID-19 highlighted the importance of the native environment and also investigated the association between contaminated air and virus prevalence.[4] Moreover, some studies linked spatial patterns of COVID-19 channeling and mortality to levels of pollutants, as well as various weather conditions among different countries.[8] In addition,

Corresponding Author: Megha Sharma, MSc, Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India.

E-mail: meghasharma15aug@gmail.com

Phone: 7727049457

Submission date: May 22, 2023, **Revision date:** May 29, 2023, **Published date: Apr 4 2024**

various aspects such as poverty, urbanization, and low per capita income had significant impact on the COVID-19 outspread. Another significant factor, enforced by the government, for the COVID-19 ending strategy was the delivery of vaccines. According to numerous research, immunization has a considerable influence on lowering COVID-19 channeling.[9]

Asia is the agglomeration of developed and developing countries and accounts for nearly half of the world's population. It is a rapidly growing economy in terms of nominal GDP and purchasing power parity (refer to www.imf.org). However, its high population, overcrowded residences, and poor health facilities provide ample opportunities to analyze the effect of various climatic, socioeconomic, health, and demographic factors in the spread of COVID-19. Spatial analysis is an effective tool for analyzing the geographic relationship between variety of geographic relationship between variety of factors and infectious disease outbreaks like COVID-19. [10,2]

Spatial distribution of the COVID-19 incident and fatality rate of COVID-19 in Iran, Bangladesh and some European countries have been studied. [3, 5, 11-13] They used various spatial models, such as spatial lag model (SLM), spatial error models (SEM), geographical weighted regression (GWR), and multiscale geographical weighted regression (MGWR), to identify regions exhibiting either a high or low concentration of COVID-19 cases in various countries. Studies found that median household income, literacy rate, and population density directly affect the disease incident rate.[5,12] Thus in this study, statistical spatial analysis has been applied that aids in identifying country-level variation between different possible factors that facilitate the COVID-19 outbreak in the Asian continent.

Methods

Data sources

The present study includes all Asian countries except Turkmenistan, North Korea, and Palestine of which updated COVID-19 data were not available. The 46 Asian countries that are included in this study. Among the 46 Asian countries mentioned, five of them (Georgia, Azerbaijan, Russia, Kazakhstan, and Turkey) are considered transcontinental, as they extend into both Europe and Asia.

In order to study the prevalence rate, data of country-level confirmed cases of COVID-19 in Asia is extracted from the site of the World Health Organization for the period of March 2020 to August 2022. The whole duration considered in this study is divided into three intervals (March 2020 to December 2020, March 2020 to October 2021, March 2020 to August 2022) which are roughly covering the first, second, and third waves cumulatively (see figure 1).[5]

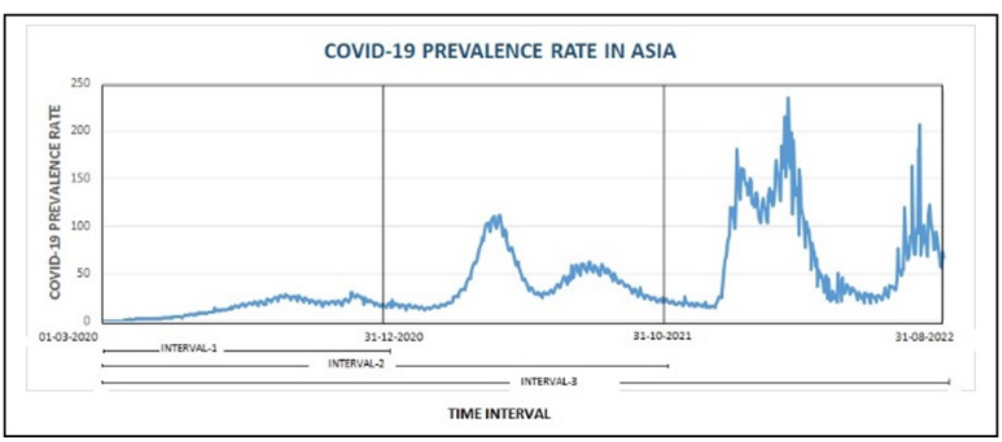


Figure 1: COVID-19 prevalence rate in Asian countries from March 2020 to August 2022

Eighteen different socioeconomic, environmental, health, and demographic variables have been selected in this study based on the previous studies and availability of data. Table 1 describes these variables, and data collected for these variables are from Our World in Data repository (https://ourworldindata.org). These variables help to evaluate a country's ability to determine and retaliate to such crisis.

Table 1: List of explanatory variables, and their description

S.No	Indicators	Abbreviation	Description
1	CO_2 emission per capita	CO_2	CO_2 emission (metric tons per capita)
2	Healthcare expenditure	HE	Total healthcare spending expressed as a percentage of national gross domestic product (GDP).
3	Hospital beds per thousands	HB	Total number of hospitals beds for one thousand population.
4	Basic Hand washing facilities	BHWF	Proportions of populations of having basic hand washing facilities.
5	Basic Drinking water	BD	Proportions of populations of having basic drinking water facilities.
6	Basic Sanitation facilities	BSF	Proportions of populations of having basic drinking water
7	Dependency Ratio	DR	Proportion of dependents (young and elderly) to working-age population (15-64 years).
8	Net migration rate	NMR	(Immigrants - Emigrants) / Population person-years lived.
9	Prevalence of smoking	PS	The share of men and women aged 15 and older who smoke.
10	Extreme poverty	EP	Share of population in extreme poverty (as state by the World Bank, living on less than 1.90 int.-$).
11	Share of urban population	SUP	Percentage of the total population living in urban areas
12	Prevalence of undernourishment	PU	Percentage of population lacking adequate calories for minimum energy requirements.
13	Vaccination rate	VR	Share of population with at least one dose of vaccination
14	Population density	PD	The number of people per km2 of land area.
15	GDP per capita	GDP	Gross domestic product per capita for any country
16	Cardiovascular health	CVH	Annual cardiovascular disease deaths per million population.
17	Diabetes prevalence	DP	Proportion of adults aged 20-79 with type 1 or type 2 diabetes.
18	Human Development Index	HDI	HDI is a summary measure of human development, combining living standards, education, and health.

Spatial Data Analysis

Global Moran's I statistic captured the overall spatial dependence among COVID-19 prevalence rates in Asia. Further, the local indicators of spatial dependence (LISA) tool were applied to obtain the local indicators of spatial association (LISA). After standardizing (normalizing) all variables in this study, the forward stepwise regression approach was used to select a group of variables by removing nonsignificant variables. Subsequently, Pearson's correlation analysis was employed to examine the associations among the chosen variables. Following the identification of multicollinearity using the variance inflation factor (VIF), independent variables were selected for the models. Then spatial models were applied to capture the spatial dependence using the chosen variables. SLM and SEM were fitted using GeoDa 1.14, and MGWR 2.2 was used to obtain GWR and MGWR. The weight matrix was generated using the inverse distance (the impact of one feature on another feature decreases with distance) which calculated using the centroid latitude and longitude locations. To assess the effectiveness of different models in explaining COVID-19 incidence rates across Asia, R2 and AIC metrics were utilized for performance comparison. R-square measures the goodness of fit; its values range from 0 to 1. Furthermore, AIC is a model performance measure that can compare predictive models while accounting for model complexity. The model with lower AIC value and higher value of R^2 better fits the observed data.

Spatial autocorrelation

Spatial autocorrelation refers to the correlation between the observation of variable at specific location and its corresponding observation at a neighboring location within the same geographic region. The examination of spatial autocorrelation can be conducted on two levels: global and local. Global Moran's I measure spatial autocorrelation and ranges from–1 to 1. Negative values indicate clustered dissimilar value, positive indicate clustered similar values, and values close to zero indicate no spatial pattern or randomness.

Nonetheless, the ability of the Moran's I statistic to detect structural instability within the dataset is limited. Hence, the LISA tool was employed to compute the local spatial autocorrelation instead. It helps to detect spatial non-stationarity or locations of outliers. It describes significant correlations at specific locations as local spatial clusters (hotspots) or correlations between observations and neighboring observations.[14]

The SLM model

The SLM effectively estimates the influence of independent variables on the dependent variable, incorporating spatial dependency between the dependent variable. The SLM model with n number of observations, and m number of independent variables presented in equation 1.

$$y = \rho W y + X\beta + \in \quad \text{...(1)}$$

where y as the n × 1 vector of dependent variable, χ as the $n \times m$ matrix of independent variables, β as the vector of regression coefficients, the spatial autocorrelation coefficient of y represented by ρW as spatial weight matrix and $\in$ is random error.

The SEM model

The SEM is an expanded version of the traditional regression model that includes spatial dependence within the disturbance term. Equation 2 presented SEM model with μ vector of spatially dependent disturbance terms, and λ its spatial autocorrelation coefficient.

$$y = X\beta + \mu,\ \mu = \lambda W \mu + \in \quad \text{...(2)}$$

The GWR model

SLM and SEM models assume spatial stationarity (association between dependent and independent variables do not vary over space). In contrast, the geographically weighted regression model estimates local interactions (estimating the value of regression parameters by fitting a regression model to each feature in the dataset) among the dependent and independent variables.

The GWR model is presented by equation 3 given below

$$y = \sum_{j=1}^{m} X_{ij}\beta_{ij} + \epsilon_i,\ i = 1,2,\ldots,n. \quad \text{...(3)}$$

Parameters estimates for each independent variable at i^{th} location is given by equation 4.

$$\hat{\beta}(i) = (X'W(i)X)^{-1}X'W(i)y \quad \text{...(4)}$$

Where $\hat{\beta}(i)$ is m ×1 vector of parameter estimates, W(i) is spatial weight matrix calculated by the Gaussian kernel function and the bandwidth which is based on Euclidean distance.

The MGWR model

The MGWR model is an extension of GWR that study the relationships of independent and dependent variables at different spatial scales by using the varying bandwidth (used to define the neighborhood around each feature) rather than a single, constant bandwidth for the entire study area. [14]

MGWR model presented in equation 5 with β_{bwj} as the bandwidth used for calibration of the j^{th} relationship.

$$y_i = \sum_{j=0}^{m} X_{ij}\beta_{bwj} + \epsilon_i, \qquad \text{...(5)}$$

$$i = 1,2,\dots,n$$

Results

Quantitative spatial distribution and autocorrelation of COVID-19 prevalence rate in Asian countries

Spatial COVID-19 prevalence in Asian countries (Intervals 1, 2, 3 - Figures 2a, 2b, 2c): Western Asia consistently had the highest prevalence. Mongolia and South Korea were highly affected in intervals 2 and 3. East Asia showed low prevalence initially, but with variability. China had low rates, while Mongolia had very high rates. Russian Federation suffered greatly. Southeast Asia generally had medium rates, except India and Nepal with high rates in interval 1, improving later.

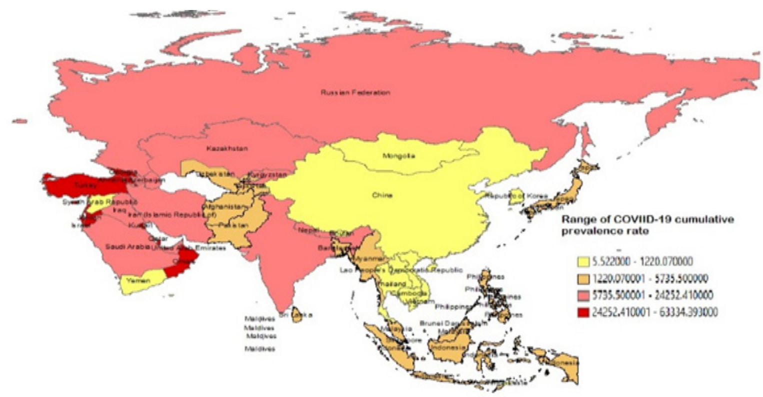


Figure (a): Quantitative spatial distribution (March 2020 to December 2021)

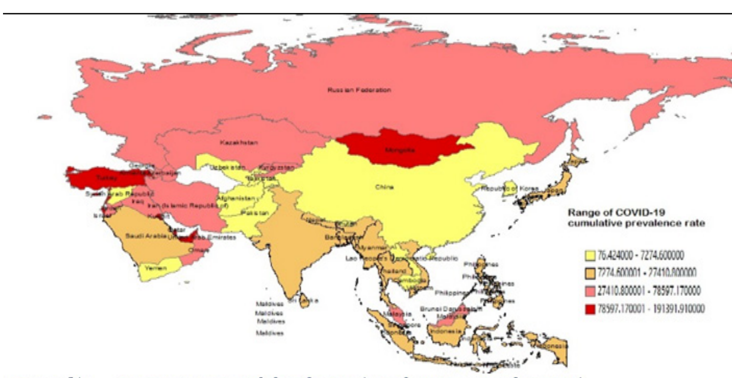


Figure (b): Quantitative spatial distribution (March 2020 to October 2021)

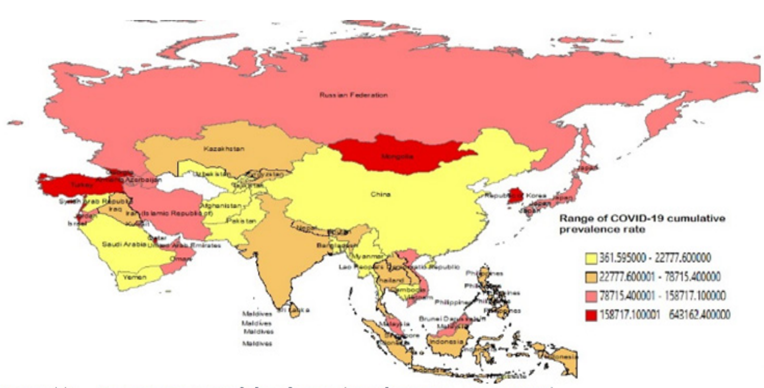
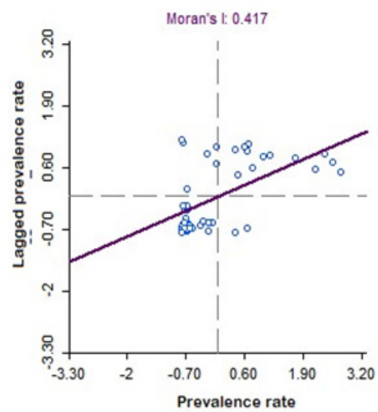


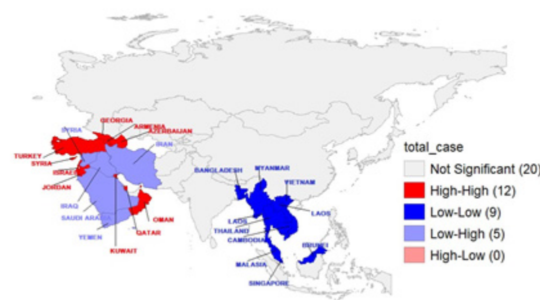
Figure (c): Quantitative spatial distribution (March 2020 to August 2022)

Figure 2: Quantitative spatial distribution of COVID-19 prevalence rate in Asia

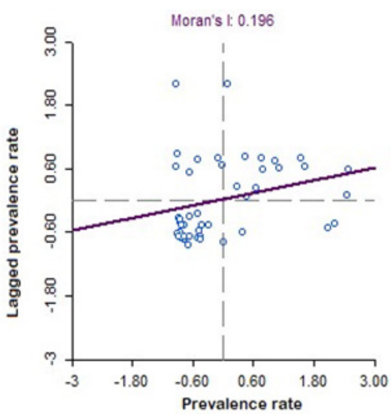
Figure- 3a, 3c, 3e show the estimates of Moran's I statistic between the prevalence rate of COVID-19 and their lagged value. The value of Moran's I were 0.417, 0.196 and 0.099 suggesting marginal but significant (p – value < 0.05 spatial dependency in prevalence rate in Asia in interval-1,2,3, respectively. The positive values of the Global Moran's I indicate that the prevalence of COVID-19 in one Asian country may have been spatially associated with that of neighboring countries, particularly in the first wave.



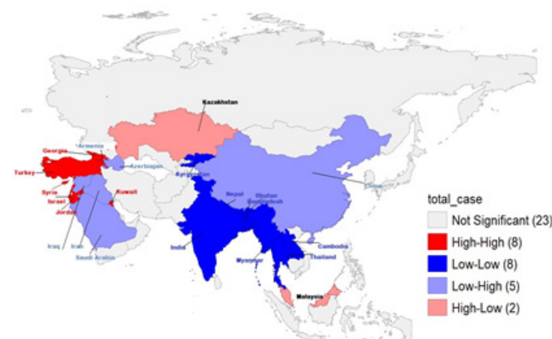
(a) March 2020 to December 2020



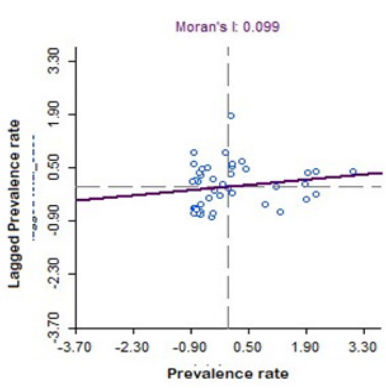
(b) March 2020 to December 2020



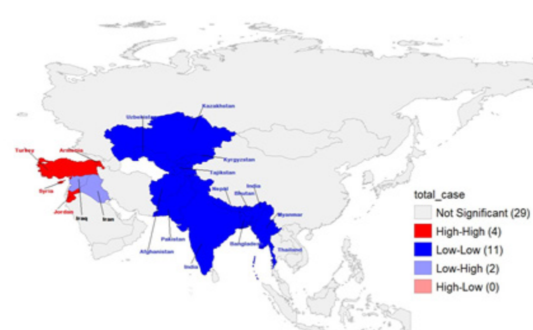
(c) March 2020 to October 2021



(d) March 2020 to October 2021



(e) March 2020 to August 2022



(f) March 2020 to August 2022

Figure 3: Plot for Global Moran's I statistics and LISA cluster maps for COVID-19 prevalence rates in Asia

The LISA tool was used to calculate the local spatial autocorrelation. LISA clusters map shown in Figure- 3b, 3d, 3f exhibit the High-High, Low-Low, High-Low, and Low-High clusters of the prevalence rate of COVID-19. In Figure 3b, 3d, and 3f, the LISA cluster maps reveal interesting patterns. Iran and Iraq consistently exhibit low COVID-19 prevalence rates throughout the entire duration, while their neighboring countries experience higher rates. During the second wave, Kazakhstan and Malaysia have high prevalence rates, contrasting with their neighboring countries showing low rates. Southeast Asian countries consistently fall into low prevalence rate clusters across the three waves of the pandemic. The LISA cluster maps also indicate a significant number of countries with insignificant spatial clustering, explaining the relatively small values of Global Moran's I. Specifically, during the second and third intervals, 23 and 29 out of 46 countries respectively formed insignificant spatial clusters.

Spatial models summary

Following the standardization (normalization) of all variables in this study, a group of independent variables was chosen by removing non-significant variables using the forward stepwise regression approach. Pearson's correlation analysis investigated the relationships between the selected variables and detecting multi-collinearity with the variance inflation factor (VIF < 7.5).

Spatial models (SLM, SEM, GWR, MGWR) were applied to COVID-19 prevalence rates in Asian countries for all intervals. A significant positive global Moran's I value was observed. The MGWR model demonstrated superior performance, exhibiting low AIC and high R-squared values across all intervals, surpassing other spatial models (refer to Table 2). Tables 3 provide a comprehensive summary of the MGWR model's results for all intervals.

Table 2: Comparison of results of spatial models

Time interval	Model comparison	SLM	SEM	GWR	MGWR
Intrval-1	R^2	0.61	0.59	0.62	0.67
	AIC	111.11	107.17	109.52	102.75
Intrval-2	R^2	0.64	80.64.	0.70	0.84
	AIC	99.85	101.56	98.06	76.80
Intrval-3	R^2	0.56	0.61	0.66	0.71
	AIC	111.5	106.04	105.50	99.74

Results demonstrate that the HDI, a measure of critical aspects of human development, is significantly related to the prevalence rate of COVID-19 in Asian countries. This study consider unit increase in HDI is associated with 0.49% to 0.70% increase in the COVID-19 prevalence rate in all intervals. Despite having a long-life expectancy, a good education, and a good standard of living, Asian countries such as Georgia and Jordan (HDI -0.81 and 0.73) have struggled to deal with the COVID-19 pandemic. Results demonstrate that the Human Development Index (HDI), a measure of critical aspects of human development, is significantly related to the prevalence rate of COVID-19 in Asian countries. Affect of COVID-19 in Asian countries. This study consider unit increase in HDI is associated with 0.49% to 0.70% increase in the COVID-19 prevalence rate in all intervals. Despite having a long-life expectancy, a good education, and a good standard of living, Asian countries such as Georgia and Jordan (HDI -0.81 and 0.73) have struggled to deal with the COVID-19 pandemic. The prevalence of COVID-19 is positively related to smoking prevalence (as the coefficient estimates are 0.233, 0.270 and 0.214 for interval-1, 2, and 3, respectively).

Table 3: Multiscale geographically weighted regression model summary for all selected significant variables.

Interval-1			Interval-2			Interval-3		
Variables	Coefficient	SE	Variables	Coefficient	SE	Variables	Coefficient	SE
Constant	0.008	0.246	Constant	0.185	0.156	Constant	0.045	0.012
CO_2	0.307	0.015	CO_2	0.185	0.156	CO_2	0.19	0.234
BHWF	0.176	0.077	PD	0.377	1.01	BHWF	0.420	0.063
PS	0.233	0.001	PS	0.270	0.001	PS	0.214	0.147
HCE	0.348	0.001	HCE	0.215	0.00	EP	0.449	0.001
CVH	0.059	0.00	GDP	-0.597	0.00	CVH	-0.339	0.003
HDI	0.426	0.001	HDI	0.749	0.00	HDI	0.741	0.032
NMR	0.163	0.057	NMR	0.358	0.433	NMR	0.170	0.053
HB	-0.400	0.00	HB	-0.451	0.00	BSF	-0.415	0.002

This result supports the World Health Organization's statement that smoking, tobacco, or alcohol exposure can increase the risk of infection from COVID-19. Net migration is another important factor across all time intervals, as an 1% increase in any country's migration rate makes it 0.16 to 0.35% more vulnerable to COVID-19. This study also discovers that the level of pollutants CO_2 emission) in Asian countries has a positive significance on COVID-19 prevalence in all intervals. Moreover, the total number of hospital beds per thousand people is inversely related to COVID-19 prevalence. A significant number of hospital beds reduces an area's resilience and aids in combating the effects of COVID-19. An epidemic or pandemic management depends on easy and affordable access to well-capable healthcare systems and health security. Population density and GDP per capita was found to be significantly impacting prevalence rate in the second interval which covers the first and second wave of COVID-19 in Asia. Although the vaccination rate is a crucial factor in eliminating the mortality of COVID-19 in the third wave, it became an insignificant factor in this study because it was based on cumulative confirmed cases of all three waves and did not consider the mortality rate before and after the allocation of vaccination.

Discussion and Conclusion

This research applied statistical spatial analysis to observe the effect of various climatic, socioeconomic, health, and demographic factors in the spread pandemic's outbreak and may help identify highly vulnerable areas. Global Moran's I value was small but significant. Moreover, LISA clusters also appeared to be significant in 50% of the Asian countries considered in the study, which justified the use of spatial models. In this study, spatial models (i.e., SLM, SEM, GWR, and MGWR) were fitted, MGWR model outperformed all spatial models in the analysis of COVID-19 prevalence in Asian countries, with a low AIC and a high coefficient of determination in all intervals. The spatial models illustrate that the prevalence of smoking, the level of pollutants, HDI, and the migration rate is positively associated with the COVID19 prevalence rate in Asian countries in all intervals. Furthermore, the availability of hospital beds per thousand people is inversely related to COVID-19 prevalence. Expectedly, population density and GDP per capita were found to be significantly impacting the prevalence rate of COVID-19 in Asian countries. We can conclude from this analysis that spatial models are beneficial for monitoring COVID-19 prevalence and its influencing factors. The MGWR model explained significant variations in COVID-19 prevalence rates across Asia. It offered insights to policymakers and communities, aiding the development of effective strategies to prevent disease outbreaks. The spatial models used in this study helped identify transmission hotspots, enabling informed decision-making and targeted prevention measures. While the study focused on national-level data, analyzing regional groupings could have provided further valuable insights if data limitations were not present.

Acknowledgments: The authors are grateful to the Department of Mathematics and Statistics, Banasthali Vidyapith for providing all the necessary resources for completing this report.

Ethical clearance: Not required

Source of funding: Self

Conflict of Interest: Nil

References

1. Liu J, Zhou J, Yao J, et al. Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. Science of the total environment. 2020;726:138513.
2. Dutta I, Basu T, Das A. Spatial analysis of COVID-19 incidence and its determinants using spatial modeling: A study on India. Environmental Challenges. 2021;4:100096.
3. Sarkar SK, Ekram KMM, Das PC. Spatial modeling of COVID-19 transmission in Bangladesh. Spatial Information Research. 2021;29(5):715– 726.
4. Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Exposure to air pollution and COVID- 19 mortality in the United States: A nationwide cross-sectional study. MedRxiv. 2020.
5. Ram´ırez-Aldana R, Gomez-Verjan JC, Bello-Chavolla OY. Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. PLoS neglected tropical diseases. 2020;14(11):e0008875.
6. Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. Bmj. 2020;368.
7. Gardner W, States D, Bagley N. The coronavirus and the risks to the elderly in long-term care. Journal of Aging & Social Policy. 2020;32(4- 5):310–315.
8. Pavel MRS, Salam A, Yesmin M, Ahsan N, Zaman SU, Jeba F. Impact and correlation of air quality and climate variables with COVID-19 morbidity and mortality in Dhaka, Bangladesh. medRxiv. 2020.
9. Mukandavire Z, Nyabadza F, Malunguza NJ, Cuadros DF, Shiri T, Musuka G. Quantifying early COVID-19 outbreak transmission in South Africa and exploring vaccine efficacy scenarios. PloS one. 2020;15(7):e0236003.
10. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Science of the total environment. 2020;728:138884.
11. Appiah-Otoo I, Kursah MB. Modelling spatial variations of novel coronavirus disease (COVID-19): evidence from a global perspective. GeoJournal. 2021:1–15.
12. Adekunle IA, Onanuga AT, Akinola OO, Ogunbanjo OW. Modelling spatial variations of coronavirus disease (COVID-19) in Africa. Science of the Total Environment. 2020;729:138998.
13. Manda SO, Darikwa T, Nkwenika T, Bergquist R. A Spatial Analysis of COVID-19 in African Countries: Evaluating the Effects of Socio- Economic Vulnerabilities and Neighbouring. International Journal of Environmental Research and Public Health. 2021;18(20):10783.
14. Anselin L. Local indicators of spatial association—LISA. Geographical analysis. 1995;27(2):93–115.
15. Fotheringham AS, Yang W, Kang W. Multiscale geographically weighted regression (MGWR). Annals of the American Association of Geographers. 2017;107(6):1247–1265.