Socio-Economic and Demographic Factors Associated with the Spatial Distribution of COVID-19 in Africa

Asiyeh Abdollahi1 · Saeed Behzadi2

Received: 6 July 2022 / Revised: 30 October 2022 / Accepted: 7 November 2022 / Published online: 17 November 2022
© W. Montague Cobb-NMA Health Institute 2022

Abstract
Corona is a disease that affects the whole world. Countries with weak economies are specifically more vulnerable. A proper understanding of COVID-19 spreading, identifying the high-risk areas, and discovering factors influencing the spread of the disease are crucial to improving disease control. This study evaluates the geo-statistical distribution of COVID-19 to identify critical areas of Africa using spatial clustering pattern analysis. In addition, the spatial correlation between infected cases and variables such as the unemployment rate, gross domestic product (GDP), population, and vaccination rate is calculated using Geographically Weighted Regression (GWR) analysis. The hot-spot map showed a statistically significant cluster of high values in southern and northern Africa. Moreover, the outcome of the GWR analysis revealed the GDP and population had the most significant correlation with the spreading of COVID-19, with Local R2 values of (0.01–0.99) and (0–0.89), respectively.

Keywords Health GIS · Spatial distribution · Geographically weighted regression

Introduction
Infectious diseases are one of the noticeable health concerns nationally and globally. Early disease detection and improved management strategies are essential in addressing infectious disease problems. On December 31, 2019, the first case of COVID-19 was reported to the World Health Organization (WHO) in Wuhan, China. The virus then spread rapidly worldwide on February 11, 2020 [24]. By December 21, 2021, the number of confirmed cases and death reached 274,628,461 and 5,358,978, declared by WHO (https://covid19.who.int/).

Africa is the world’s second most populous continent, with widespread diseases managed for years. Considering the special conditions in Africa, it can be assumed that Corona disease will remain in this continent for many years, like many previous infections [31]. The rapid mutation and widespread of the coronavirus have caused it to be different from other viruses. This virus will not remain exclusively in the African continent and will spread again to worldwide [33]. Therefore, protection against the coronavirus should be done equally on all continents [1]. Otherwise, the people of the world will not be relieved from this virus for many years. According to the WHO, Africa has the lowest number of infections and deaths, which shows the inverse relationship between poverty and COVID-19. However, this cannot prove that poverty confers immunity against COVID-19. It reveals that several criteria are involved in the spread of the coronavirus, and this is a complex issue [25]. There are also little details about the outbreak of COVID-19 in Africa. Accordingly, Africa was considered as the study area in this research. The first confirmed case in this continent was in Egypt on February 14, 2020 (Olusola, [23]).

The emergence of COVID-19 is a challenge that requires new responses from public health and medical care systems [15]. Poor medical infrastructure in underdeveloped countries is an essential obstacle to preventing the spread of this disease. Understanding the patterns of disease spread in space is a critical issue in monitoring infectious diseases
Population density, international airport, and literacy ratio had the most vulnerable places to exposure to COVID-19. In a dataset. This research indicated that important states Error (SER), are implemented for testing multicollinearity. Two models, Ordinary Least Square (OLS) and Spatial, were used to assess the spatial distribution of COVID-19 risk factors within the first quarter (March–May). In research titled “The Epidemiological and Spatiotemporal Characteristics of the 2019 Novel Coronavirus Disease (COVID-19) in Libya,” investigated the epidemiological parameters and spatiotemporal patterns of COVID-19 by exploring dynamic spatial trends. They indicated considerable growth in the outbreak in the middle of July 2020, specifically in the west and south of Libya. In the paper titled “A snapshot of space and time dynamics of COVID-19 risk in Malawi. An application of spatial–temporal model,” implemented a spatiotemporal model for the weekly confirmed cases for 24th June to 20th August. They concluded the city area were under higher threat than rural ones. Although the outbreak’s intensity has fluctuated in most cities, it was constant in the rural. In a paper entitled “Spatial variability of COVID-19 and its risk factors in Nigeria: A spatial regression method,” investigated the COVID-19 risk factors within the first quarter (March–May). Two models, Ordinary Least Square (OLS) and Spatial Error (SER), are implemented for testing multicollinearity in a dataset. This research indicated that important states had the most vulnerable places to exposure to COVID-19. Population density, international airport, and literacy ratio were influential predictors. In the paper titled “Mapping vulnerability to Covid-19: Supplementary material to the March 2020 Map of the Month,” prepared two maps for Gauteng City in South Africa. The first one shows the map of risk factors for maintaining hygiene and primary prevention. The second map shows the risk factors map about the lockdown and the possibility of a disease outbreak. These maps indicated areas that are most vulnerable to the spread of COVID-19. In research titled “Spatial analysis of COVID-19 and traffic-related air pollution in Los Angeles,” investigated the relationship between air pollution and the spreading of Covid-19 in Los Angeles. The result showed a significant relationship between exposure to air pollution and the increase in the number of infected cases. In addition, the infected population often has lower income so boosting the mortality rate of these people may be due to exposure to air pollution. In a paper titled “A vulnerability index for COVID-19: spatial analysis at the subnational level in Kenya,” calculated the Social Vulnerability Index (SVI), Epidemiological Vulnerability Index (EVI), and a combination of them for Kenya spatially. As a result, vulnerability priorities for sub-counties were obtained.

Despite the growing literature, there seems to be a need to investigate the spread of the coronavirus across the continent based on data collected over a more extended period. In this study, statistical and spatial analyses are applied to determine the spatial distribution and spatial clustering patterns of the COVID-19 incidence and mortality rate by the end of December 21, 2021, in Africa. This information can be beneficial for identifying the high-risk areas infected by COVID-19. This study aims to show the correlation between the spatial distribution of COVID-19 and different possible variables using the Geographically Weighted Regression (GWR) method.

**Methodology**

In this research, geo-statistical analysis is done in three stages. In the first step, the spatial data is evaluated to determine whether the data meet the necessary conditions for geo-statistical analysis or not. If the spatial data passes this test, the analysis of the two next steps can be done on the data. In the second step, clustering analysis is applied to the data, and hot-spot locations are identified in the study area. Ultimately, the correlation between the effective criteria and the data set of patients is examined. At this stage, the impact of different criteria is determined on the data set of patients. Figure 1 shows the flowchart of the method in this research.

**Data**

The continent of Africa is considered the study area of this research. This continent has an area of 29,648,481 km².
This continent contains 54 fully recognized sovereign states along with three dependent territories. Western Sahara is another country that is known as a disputed territory. Since this territory does not have political independence, some parts of the data (unemployment rate and gross domestic product (GDP) growth) of this region are not available. In this research, 55 countries (54 independent countries and the Western Sahara) are examined. The data required for these countries was obtained from the worldometers website. In this research, the rate of the infected (cases—cumulative per 100,000 populations) and the rate of deaths (death—cumulative per 100,000 populations) are used to perform hot-spot and outlier analysis. The cumulative number of infected is also used for GWR analysis. This data is extracted from the site the Johns Hopkins University (JHU) and WHO. The data period of this research is considered from the emergence time of Corona until December 21, 2021. Until this date, 6,830,309 infected and 225,780 cases of death have been identified. Therefore, all input data are collected in this period. The variables used as input in this research are as follows:

**Cumulative Incidence Rate (CIR)**

CIR consists of the number of persons who experience the disease during a specified period divided by the total population at risk. This rate is often used to predict the risks associated with the disease outbreak over shorter or longer periods [35].

**Cumulative Mortality Rate (CMR)**

CMR is the sum of the mortality rates in a defined period. It is based on probability theory. CMR is also an indicator of death probability, commonly expressed as a percentage.

---

1. https://www.worldometers.info/world-population/africa-population/
2. https://www.worldometers.info/geography/how-many-countries-in-africa/
3. https://coronavirus.jhu.edu/map.html
4. https://covid19.who.int/data
and used as an approximation of the cumulative mortality risk [26] [32].

**Unemployment Rate**

The unemployment rate is the portion of the labor force without a job. When the economy is in an unsatisfactory condition, the country faces a shortage of jobs. In this case, it can be predicted that the unemployment rate will increase. On the other hand, when the economy has positive growth, jobs are relatively abundant. Therefore unemployment rate can be expected to descend [21]. In this study, unemployment rate values are tabulated for Africa from the Trading Economics website\(^5\) on December 21, 2021. Unemployment rate values range from 0.7 (for South Africa with the highest jobless rate) to 34.4 (for Niger with a growing economy) (Fig. 2a).

**Gross Domestic Product (GDP)**

GDP is an index measuring the total economic value of all products and services within a country’s borders for a specified period [4]. GDP provides an economic snapshot of a country, which is used to assess economic health and growth rate. The GDP of African countries is collected based on official data from the Trading Economics website\(^6\) in a table form (Fig. 2b).

**Population**

Africa is the second most populated continent in the world, whose population is equal to 16.72% of the total world population based on Worldometers. Since the number of infections and death is obtained from the population, the population can be a fundamental variable for Corona disease. The

---

\(^5\) https://tradingeconomics.com/country-list/unemployment-rate?continent=africa

\(^6\) https://tradingeconomics.com/country-list/gdp?continent=africa
latest population data of African countries is collected in tabular form from Worldometers.7

**Vaccinated People**

The number of vaccinated people in Africa until December 21, 2021 is obtained from the WHO. Values were expressed as the cumulative number of people fully vaccinated per 100. These data are also obtained from the WHO website.8

### Preprocessing

In this stage, data preparation and integration are done. First, a spatial map of the African continent is prepared, and then all the collected data are added to each country of this continent as descriptive data. Finally, a spatial layer with descriptive data is created. Table 1 shows the descriptive data of 20 sample countries out of 55 countries. In this table, ISO3 is the abbreviation of the countries. The rest of the table column names are also determined based on the data required for this research.

Table 2 also shows the statistical information of the collected data. In this table, the minimum, average, maximum, and standard deviation values of each data are displayed. The spatial map produced in this step, along with the attached

---

7 https://www.worldometers.info/geography/how-many-countries-in-africa/
8 https://covid19.who.int/who-data/vaccination-data.csv
The spatial distribution maps of the unemployment rate, GDP, population, and vaccination rate are shown in Fig. 3. The color change of each map is defined based on their impact. In some maps, higher values indicate their severity, and in some others, the lower value. Therefore, the map whose lower values are more effective is used in reverse legend.

**Analysis**

**Step 1: Spatial Autocorrelation**

Before the main analyses are applied to the data, it is necessary to check the distribution pattern of the input data. In this research, spatial autocorrelation is used for this analysis. Spatial autocorrelation is an essential concept in spatial statistics. There are two primary reasons to measure spatial autocorrelation. First, it indexes the violation degree of a fundamental statistical assumption. Secondly, it indicates the degree to which conventional statistical inferences are compromised when non-zero spatial autocorrelation is overlooked [6]. Autocorrelation complicates statistical analysis by altering the variance of variables. It also changes the probabilities that statisticians commonly assign to incorrect statistical decisions (e.g., positive spatial autocorrelation results in an increased tendency to reject the null hypothesis when it is true). It quantifies the extent of redundant information in geo-referenced data, affecting the geo-referenced observation [13].

The spatial correlation mechanism indicates whether related data behave similarly or not. In spatial autocorrelation, apart from examining the relationship between the two elements, the neighborhood of the data is also checked [10]. If the z score or p value indicates statistical significance, a positive Moran’s I index value indicates a tendency toward clustering. In contrast, a negative Moran’s I index value indicates a dispersion tendency [11]. In this case, the spatial distribution of infection and death rates is estimated for each country by the global Moran’s I based on Eqs. (1)–(2) [20]. This analysis revealed the clustered patterns.

\[ I = \frac{n}{S_0} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}Z_i Z_j \]

(1)
\[ S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} \]  

where \( W_{ij} \) is a spatial weight, \( n \) is the number of features, \( Z \) is the variable of interest, and \( S_0 \) is the sum of all \( W_{ij} \).

**Step 2: (Hot-Spot and Cluster Analyses)**

If the distribution pattern of data is clustering, the analysis of this step can be done; otherwise, the analysis cannot be continued. The analysis of this stage is done in two parts. The first part is hot-spot analysis to identify the clustering of spatial phenomena. The clustering method is one of the most common methods for identifying and extracting patterns in big data [11]. Clustering algorithms are classified as the cluster model. One of the model-based clustering methods is the degree of clustering used in the Getis-ord-Gi* Statistics for High/Low Values. This method is inferential statistics, so the analysis results are based on the null hypothesis and values density [29]. The Getis-Ord Gi* index is measured for each country. The analytic output provides \( z \) scores, \( p \) values, and confidence level bin for each country, which statistically shows significant spatial clusters of high values (hot spots) and low values (cold spots). This index is obtained according to Eq. (3):

\[ G_i^* = \frac{\sum_{j=1}^{n} W_{ij}x_j}{\sum_{j=1}^{n} x_j} \]  

where \( x_j \) is the value of variable \( X \) at location \( j \), \( W_{ij} \) is the spatial weight of variable \( x_j \), and \( n \) is the number of features.

The second part is Cluster and Outlier Analysis to identify concentrations of high (low) values and spatial outliers. This analysis is based on Local Moran statistics. Local Moran statistics is decomposed from General Moran statistics. Local Moran’s \( I \) statistics calculate \( z \) score and pseudo-\( p \) value. It also displays cluster types along with statistical properties. The results of the local Moran statistics are tested with a \( z \) score, which determines the confidence level. If cell \( i \) has a positive sign, the value of cell \( i \) is similar to its neighbor cells’ value. If the value of \( i \) is a large positive number, it indicates a substantial clustering range. If the value of \( i \) is negative and significant, the amount of surface cell property \( i \) has a high difference from its neighbor cells, which indicates a negative spatial correlation [5]. The possible cluster and outliers of COVID-19 distribution are applied by Anselin Local Moran’s \( I \) analysis. Based on the obtained values, four types of spatial patterns are presented: the High–High (HH) cluster, High–Low (HL) outlier, Low–Low (LL) cluster, and Low–High (LH) outlier. These four types of spatial patterns show the spatial structure of the epidemic risk in the study area. Specifically, the HH (LL) cluster indicates some adjacent areas with relatively high (low) values of epidemic incidence, showing a high (low) risk of the epidemic in these areas. The HL (LH) outlier indicates a high (low) value primarily surrounded by low (high) values of incidence, which may be caused by a unique mechanism [30]. The Anselin Local Moran’s \( I \) is calculated based on Eqs. (4)–(5):

\[ I_i = \frac{X_i - \bar{X}}{S_i^2} \sum_{j=1,j\neq i}^{n} W_{ij} (X_j - \bar{X}) \]  

\[ S_i^2 = \frac{\sum_{j=1,j\neq i}^{n} (X_j - \bar{X})^2}{n-1} \]  

where \( x_i \) is the value of feature \( i \), \( \bar{X} \) is the mean of features value, \( W_{ij} \) is a spatial weight between features \( i \) and \( j \), and \( S_i^2 \) is the total number of features.

**Step 3: Geographically Weighted Regression**

In this step, the correlation between the cumulative number of patients and input variables (such as population, GDP, etc.) is calculated. This analysis is based on geographically weighted regression (GWR). GWR is a local non-stationarity for modeling spatial variable relationships. Stationarity is a condition in which the mean, variance, and location dependence are not adjusted in space. GWR allows a connection between independent and dependent variables to vary by location [27]. Equation (6) shows the relationship of GWR:

\[ y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \epsilon_i, i = 1, \ldots, n \]  

where \( y_i \) is the dependent variable at location \( i \), \( \beta_0 \) is the intercept coefficient at location \( i \), \( x_{ik} \) is the k-th explanatory variable at location \( i \), \( \beta_k \) is the k-th local regression coefficient for the k-th explanatory variable at location \( i \), and \( \epsilon_i \) is the random error term associated with location \( i \).

In GWR, the independent variables are selected locally. Then, a linear model is locally fitted to the data. Therefore, the criteria are generally not considered if they are linear or non-linear. If the data are locally linear, the GWR method can identify the appropriate model and determine the impact of the criteria. Otherwise, the model declares no significant relationship between the desired criterion and the dependent variable [36]. In this paper, GWR explores the relationship between the dependent variable (cumulative infected) and explanatory variables (vaccination rate, unemployment rate, population, and GDP).
Results

Spatial Autocorrelation Analysis

Spatial autocorrelation was calculated for CIR and CMR based on the feature locations and attribute values (Fig. 4). Since spatial autocorrelation is an inferential statistic, its result is interpreted in the context of the null hypothesis. The null hypothesis states the values are randomly distributed in the study area. If the $p$ values are small and the $z$ score is very high or very low, the clustering is spatial and can reject the null hypothesis [12]. As seen in Fig. 4a, CIR has a small $p$ value and a positive $z$ score of 2.205600, which indicates a spatial cluster. Therefore, with a probability of less than 5%,
this cluster pattern could result from random chance. Also, in Fig. 4b, with a small \( p \) value and positive \( z \) score, CMR has a cluster pattern. This means that the spatial pattern could result from random chance with a probability of 1%.

### Clustering Map Based on Hot-Spot Analysis

The hot-spot maps were provided for CIR and CMR. The infected map showed a statistically significant cluster of high values in southern Africa (Fig. 5a). The clustering of high values is demonstrated in Namibia, Botswana, and Lesotho with a confidence level of 90%, while a confidence level of 95% is seen in South Africa. No statistically significant cluster was observed in the rest of the continent. The map for death rate indicated substantial spatial clusters of high values in the north and south of Africa (Fig. 5b). Seven countries were represented with a confidence level of 99%, and there were two countries with a confidence level of 95%. In addition, the clustering index shows low values (cold spots) for the Ivory Coast and Burkina Faso with a confidence level of 90%. In general, as seen in Fig. 5a–b, the spatial patterns of COVID-19 have more severe conditions in the northern and southern regions of Africa.

### Evaluation of Anselin Local Moran’s I

Anselin Local Moran’s \( I \) analysis represented the map with hot and cold spots considering neighborhood similarities or differences. Two high-high clusters, including Namibia and Botswana, were identified in the infected map. These clusters had high values (positive \( z \) score). They were also surrounded by similar values (Fig. 6a). Moreover, the death rate map identified six countries located in the north and south of Africa as high-high clusters (Fig. 6b).

### Assessment of Geographically Weighted Regression

The GWR map shows the correlation between the infected rate and other variables. The darker areas represented the stronger significant relationship. GWR calibrated the regression equations for each country; it displays the map for each variable (Fig. 7). The unemployment rate was one of the explanatory variables. According to Fig. 7a, there was not a significant correlation between the unemployment rate and cumulative infected.

In Fig. 7b, the GWR analysis is shown for the GDP and incidence data. In this figure, a significant correlation can be seen on the coastal edges of the continent (from the east through the south and part of the west). In the GDP distribution map (Fig. 3e), it was shown that the underdeveloped countries are located in the interior of the continent. Cumulative incidence in these countries has a lower correlation with GDP (Fig. 7b). On the other hand, the developed countries that are in contact with the international community show the highest correlation with COVID-19 cases, which are shown in the darker color. This could emphasize the relationship between the development and the spread of COVID-19 [24].

In Fig. 7c, the GWR analysis is shown for the population and incidence data. According to map (Fig. 7c), the highest correlation between population and COVID-19 infection was almost similar to the GDP map. The countries situated...
in the coastal south, west, and east of the continent have the highest correlation. These countries have Local R2 values larger than 0.6, which represents a significant correlation between the two variables. Although countries like Tanzania, DR Congo, and Algeria have large populations, their Local R2 values are almost zero. This showed that population alone could not be an effective measure of corona expansion. Therefore, other variables should be considered along with the population.

In Fig. 7d, the GWR analysis is shown for the vaccination distribution and incidence data. The vaccination distribution map (Fig. 3f) shows that most areas of the continent are not vaccinated. The biggest correlation between vaccination rate and COVID-19 infection was displayed in northwestern Africa, according to Fig. 7d.

**Discussion**

The present research is comparable with Shariati, Mesgari et al. (2020) [28]. In that research, the hot-spot analysis was implemented for CIR worldwide. Their study demonstrated the hot spot in the north of Africa with a 99% confidence level and no significant cluster in the south of Africa. Noting
that their analysis was on March 31 and April 30, 2020, 1 month after the emergence of Covid-19 in the African continent (February 14, 2020). Their study was at the beginning of the disease outbreak. Onafeso et al. [24] implemented a spatio-temporal model for the first 61 days of the coronavirus outbreak in Africa. Their study has presented the stringent response of African countries, such as Border closure, Airport Closure, Partial Lockdown, and Mobile Test Centers. In addition, they analyzed emerging hot spot of COVID-19 cases in South Africa, Egypt, Morocco, Tunisia, Burkina Faso, Côte d’Ivoire, Senegal, Ghana, Cameroon, and Nigeria with the highest level of confidence.

When this study performed hot-spot analysis for incidence rates up to December 21, 2021, it identified southern African countries, including South Africa, Namibia, Botswana, and Lesotho, as critical areas with a confidence level between 90 and 95%. The same regions were also detected as hot spots for death rate analysis with a higher confidence level. Moreover, other countries such as Zimbabwe in the south and Libya, Tunisia, and Algeria in the north of Africa were added to critical countries, while Burkina Faso and Côte d’Ivoire were in the cold-spot cluster. These cold-spot clusters had been emerging hot spots in previous research. This comparison can show the effect of appropriate government response to prevent the spread of COVID-19.

Understanding how hot spots are created in countries led us to do the GWR analysis to determine the relationship between the infected cases with some independent factors. This analysis examined the correlation between COVID-19 incidence and other variables such as unemployment rate, GDP, population, and vaccination rate in each country. Figure 7a, d showed the correlation between unemployment and vaccination rates with infected cases. Local R2 values (0.09–0.21) and (0.003–0.38) of these two maps indicated that there was no significant relationship between these variables. GWR analysis between GDP and COVID-19 cases showed a strong relationship between these variables. The countries such as Ethiopia and Somalia, and the countries of southern Africa have a significant correlation between their GDP and the spreading of COVID-19, with Local R2 values larger than 0.7. Examining the geographic distribution of GDP and population (Fig. 3e–d) showed that densely populated countries with high GDP remarkably correlate with the outbreak. This significant relationship led to the analysis of GWR for population and COVID-19 cases. The obtained results emphasized the correlation between the outbreaks of COVID-19 in developed countries with high populations. The Local R2 values for the twenty most infected countries are shown in Table 3. In this table, ISO3 is the abbreviation of the countries. The columns R2_GDP, R2_Population, R2_unemployment, and R2_Vaccination, respectively, deliver the value of Local R between the infected cases and the GDP, population, unemployment, and vaccination data.

To ensure the efficiency of GWR analysis, ordinary least squares (OLS) regression is also tested on the available data. OLS is a common technique for estimating coefficients of

| Country name | ISO3 | Cumulative infected | R2_GDP | R2_population | R2_unemployment | R2_vaccination |
|--------------|------|---------------------|--------|---------------|-----------------|----------------|
| South Africa | ZAF  | 2,666,964           | 0.99506| 0.859487      | 0.204777        | 0.019379       |
| Morocco      | MAR  | 799,775             | 0.13474| 0.263292      | 0.143825        | 0.380819       |
| Tunisia      | TUN  | 640,897             | 0.04386| 0.084526      | 0.155247        | 0.277971       |
| Ethiopia     | ETH  | 293,737             | 0.75561| 0.822945      | 0.094742        | 0.033534       |
| Libya        | LBY  | 293,532             | 0.01067| 0.006156      | 0.128321        | 0.171827       |
| Egypt        | EGY  | 285,995             | 0.30891| 0.186516      | 0.112722        | 0.118864       |
| Kenya        | KEN  | 227,077             | 0.66437| 0.492796      | 0.112374        | 0.025404       |
| Zambia       | ZMB  | 203,978             | 0.77293| 0.04053       | 0.153869        | 0.024927       |
| Algeria      | DZA  | 190,656             | 0.02008| 0.000192      | 0.139564        | 0.272556       |
| Nigeria      | NGA  | 182,503             | 0.64046| 0.60907       | 0.100061        | 0.113779       |
| Botswana     | BWA  | 146,461             | 0.97225| 0.538854      | 0.180738        | 0.023452       |
| Mozambique   | MOZ  | 142,013             | 0.86435| 0.11899       | 0.175518        | 0.01605        |
| Namibia      | NAM  | 123,483             | 0.95563| 0.530339      | 0.173056        | 0.029527       |
| Zimbabwe     | ZWE  | 122,288             | 0.94356| 0.26104       | 0.176333        | 0.019896       |
| Uganda       | UGA  | 118,673             | 0.5413 | 0.354124      | 0.107841        | 0.03365        |
| Ghana        | GHA  | 112,928             | 0.62700| 0.562596      | 0.100864        | 0.141919       |
| Cameroon     | CMR  | 82,454              | 0.67724| 0.68706       | 0.096967        | 0.083363       |
| Rwanda       | RWA  | 82,215              | 0.30054| 0.094237      | 0.115782        | 0.032974       |
| Senegal      | SEN  | 71,628              | 0.35070| 0.236165      | 0.09602         | 0.276093       |
| Malawi       | MWI  | 59,249              | 0.65830| 0.033733      | 0.159215        | 0.01893        |
linear regression equations, which describe the relationship between independent variables and a dependent variable. In Fig. 8, the scatter diagram of this analysis is shown for each of the variables. In this analysis, R2 of OLS is also calculated between the variables. Then, this R2 value was compared with the R2 value obtained from the GWR method. Table 4 shows the results of this comparison. In this table, “OLS R-Squared” and “GWR R-Squared” are the calculated R-squared for OLS and GWR analyses, respectively. Since the R2 value of OLS analysis is lower than GWR analysis, it can be concluded that GWR analysis is more suitable for modeling input variables. The main reason for this result is that GWR analysis can calculate the spatial distribution of a variable locally, while this capability is ignored in the OLS method. This makes GWR analysis model non-stationary variables more correctly [37].

### Conclusion

Corona disease is an infectious disease that has killed millions of people in the last few years. Identifying the influencing factors can play a significant role in controlling the disease. Since treatment costs are limited in the African continent, identifying the factors affecting Corona can approximately compensate for the weakness of treatment. Therefore, in this research, the influential factors such as unemployment rate, GDP, population, and vaccination were investigated on Corona in the African continent. In this research, first, the suitability of the data was examined for spatial analysis. Then, sensitive areas in the continent were identified for infected and deceased by hot-spot analysis and clustering. Finally, factors affecting the disease were investigated by GWR analysis.

In this research, the spread of Corona was analyzed spatially in the entire continent of Africa. Therefore, the spatial analysis of the spread of Corona and its spatial correlation was carried out with several independent variables in this article. This issue is one of the innovations of this research. On the other hand, in previous research, statistical analyses were performed on data, while in this research, spatial factor such as neighborhood analysis was also considered on the disease. This issue is the main difference between the current research and the previous one.

According to the obtained results, it was determined that the southern part of the African continent is a hot spot for the infected. Hot-spot areas of the deceased were also observed in the southern and northern parts of this continent. According to the result, the populated developed countries were being exposed to the outbreak of COVID-19 more than the other countries. A significant relationship was observed between the outbreak of COVID-19 and explanatory variables in these countries. Some of these countries are located in hot-spot clustering as well. The countries in the central part of the African continent are economically weak. It should be noted that the absence of these countries in the hot-spot map does not indicate the absence of disease in these areas. Due to the underdeveloped public health infrastructure, Corona tests are less likely to be performed in these areas. Therefore, it is impossible to have accurate statistics on the number of infected in these areas. The ideal situation is to sample the number of infected and deaths in the same conditions in all countries. In this case, more accurate analyses can be obtained. In this research, influential factors (such as GDP) were investigated, while environmental factors, such as temperature, altitude, humidity, etc., can also be studied on Corona disease in Africa.
outcome of these sorts of research could moderately provide an appropriate perspective of diseases outbreak. However, disease management needs accurate statistics of real people’s conditions to access reasonably public health services.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40615-022-01453-w.

References
1. AbouGhayda R, et al. The global case fatality rate of coronavirus disease 2019 by continents and national income: a meta-analysis. J Med Virol. 2022;94(6):2402–13.
2. Arashi M, et al. Spatial analysis and prediction of COVID-19 spread in South Africa after lock down. 2020.
3. Bayode T, et al. Spatial variability of COVID-19 and its risk factors in Nigeria: a spatial regression method. Appl Geogr. 2021;138:102621.
4. Callen T. Back to basics: what is gross domestic product? Finan Dev. 2008;45:004.
5. Chen Y. New approaches for calculating Moran’s index of spatial autocorrelation. PLoS ONE. 2013;8(7):e68336.
6. Chun Y, Griffith DA. Spatial statistics and geostatistics: theory and applications for geographic information science and technology. Sage; 2013.
7. Cromley EK, McLafferty SL. GIS and public health. Guilford Press; 2011.
8. Daw MA, et al. The epidemiological and spatiotemporal characteristics of 2019 novel coronavirus diseases (COVID-19) in Libya. Front Public Health. 2021;9:586.
9. de Kadt J, et al. Mapping vulnerability to Covid-19: supplementary material to the March 2020 Map of the Month. Johannesburg: Gauteng City-Region Observatory (GCRO); 2020.
10. Dormann CF, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography. 2007;30(5):609–28.
11. Ghashghaie S, Behzadi S. Spatial statistics analysis to identify hot spots using accidental event calls services. J Stat Res Iran JSRI. 2019;16(1):121–41.
12. Goovaerts P, Jacquez GM. Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: the case of lung cancer in Long Island, New York. Int J Health Geogr. 2004;3(1):1–23.
13. Griffith DA (1987). “Spatial autocorrelation.” A Primer (Washington, DC, Association of American Geographers).
14. Hafner CM. The spread of the Covid-19 pandemic in time and space. Int J Environ Res Public Health. 2020;17(11):3827.
15. Lin X, et al. Challenges and strategies in controlling COVID-19 in mainland China: lessons for future public health emergencies. J Soc Health. 2021;4(2):57–61.
16. Lipsit J, et al. Spatial analysis of COVID-19 and traffic-related air pollution in Los Angeles. Environ Int. 2021;153:106531.
17. Liu R, Li GM. Hesitancy in the time of coronavirus: temporal, spatial, and sociodemographic variations in COVID-19 vaccine hesitancy. SSM-population Health. 2021;15:100896.
18. Lone SA, Ahmad A. COVID-19 pandemic—an African perspective. Emerg Microb Infect. 2020;9(1):1300–8.
19. Macharia PM, et al. A vulnerability index for COVID-19: spatial analysis at the subnational level in Kenya. BMJ Glob Health. 2020;5(8):e003014.
20. Martins-Melo FR, et al. Spatiotemporal patterns of schistosomiasis-related deaths, Brazil, 2000–2011. Emerg Infect Dis. 2015;21(10):1820.
21. Murphy KM, Topel R. Unemployment and nonemployment. Am Econ Rev. 1997;87(2):295–300.
22. Ngwira A, et al. A snap shot of space and time dynamics of COVID-19 risk in Malawi. An application of spatial temporal model. medRxiv. 2020;1–12.
23. Olusola A, et al. Early geography of the coronavirus disease outbreak in Nigeria. Geojournal. 2020;87:1–5.
24. Onafeso OD, et al. Geographical trend analysis of COVID-19 pandemic onset in Africa. Social Sci Human Open. 2021;4(1):100137.
25. Osayomi T, et al. A geographical analysis of the African COVID-19 paradox: putting the poverty-as-a-vaccine hypothesis to the test. Earth Syst Environ. 2021;5(3):799–810.
26. Porta M. “Mortality rate, morbidity rate; death rate; cumulative death rate; case fatality rate.” A Dictionary of Epidemiology. 5th ed. Oxford UK: Oxford University Press; 2014.
27. Schug K, Hildenbrand Z. Environmental issues concerning hydraulic fracturing. Academic Press; 2017.
28. Shariati M, et al. Spatiotemporal analysis and hotspots detection of COVID-19 using geographic information system (March and April, 2020). J Environ Health Sci Eng. 2020;18(2):1499–507.
29. Sheikholeslami G, et al. Wavecluster: a multi-resolution clustering approach for very large spatial databases. VLDDB. 1998;428–39.
30. Shi Y, et al. A spatial anomaly points and regions detection method using multi-constrained graphs and local density. Trans GIS. 2017;21(2):376–405.
31. Siraj A, et al. Early estimates of COVID-19 infections in small, medium and large population clusters. BMJ Glob Health. 2020;5(9):e003055.
32. Sun Z, Xu Y. Health medical statistics. People’s Health Publishing House; 2006. p. 377–81.
33. Tkachenko A, et al. Development of a simulation model for the spread of COVID-19 coronavirus infection in Kaluga region. E3S Web Conf EDP Sci. 2021;270:01003.
34. Vafaeinezhad A, et al. A new approach for modeling spatiotemporal events in an earthquake rescue scenario. J Appl Sci. 2009;9(3):513–20.
35. Vandembroucke JP, Pearce N. Incidence rates in dynamic populations. Int J Epidemiol. 2012;41(5):1472–9.
36. Wang N, et al. Local linear estimation of spatially varying coefficients: an improvement on the geographically weighted regression technique. Environ Plan A. 2008;40(4):986–1005.
37. Wang Q, et al. Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. Glob Ecol Biogeogr. 2005;14(4):379–93.

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