Covid-19 severity and urban factors: investigation and recommendations based on machine learning techniques

Saleh Qanazi, Ihab Hijazi, Anas Toma, Sohaib Abujayyab, Youness Dehbi, Shaker Zabadae, Xin Li

Palestinian Medical and Pharmaceutical Journal

Volume 7 | Issue 2 Article 14

4-16-2022

Covid-19 severity and urban factors: investigation and recommendations based on ma-chine learning techniques

Saleh Qanazi

Ihab Hijazi

Anas Toma

Sohaib Abujayyab

Youness Dehbi

See next page for additional authors

Recommended Citation

Qanazi, Saleh; Hijazi, Ihab; Toma, Anas; Abujayyab, Sohaib; Dehbi, Youness; Zabadae, Shaker; and Li, Xin (2022) "Covid-19 severity and urban factors: investigation and recommendations based on ma-chine learning techniques," *Palestinian Medical and Pharmaceutical Journal*: Vol. 7 : Iss. 2 , Article 14.
Available at: https://doi.org/10.59049/2790-0231.1095

Covid-19 severity and urban factors: investigation and recommendations based on ma-chine learning techniques

Abstract

Since March 5, 2020, the West Bank has faced a real crisis due to the Coronavirus dis-ease 2019 (COVID-19) pandemic. It has infected 581,678 people and caused 5,382 deaths so far, which has resulted in negative impacts on public health and other aspects of daily life. Based on the data provided by the Palestinian Ministry of Health, we inferred the spatial dis-tribution patterns of the pandemic condition in different communities using Geographic In-formation System (GIS) analysis for pattern and clustering by studying the impact of urban factors on the number of confirmed COVID-19 cases. Ten urban factors were selected (i.e., population, population density, aging ratio, the hierarchy of services, health services, land use, commercial services, road density, green areas, and open spaces) to check their relation to pandemic severity using a linear model, where five factors showed a globally-significant relation. Then, the Geographically Weighted Regression' model (GWR) was adopted to de-fine their unevenly distributed effects in the urban areas on the northwest bank. Among the five factors, the population factor has the most significant impact on the epidemic situation with a positive correlation. However, a negative correlation has been stated between the area of commercial services per person, population density, hierarchy of services, and health ser-vices. Finally, we provide recommendations that coordinate various urban factors to mitigate the pandemic spread. This paper will help decision-makers plan and develop different areas in Palestine and worldwide by better understanding the transmission, occurrence, and diffu-sion of the COVID-19 pandemic in urban areas.

Keywords

COVID-19, Urban factors, Epidemic analysis, Machine learning, Urban spatial patterns, Regression

Authors

Saleh Qanazi, Ihab Hijazi, Anas Toma, Sohaib Abujayyab, Youness Dehbi, Shaker Zabadae, and Xin Li

Palestinian Medical and Pharmaceutical Journal (PMPJ). 2022; 7(2): 237-254

Covid-19 severity and urban factors: investigation and recommendations based on machine learning techniques

Saleh Qanazi [1], Ihab Hijazi [1,], Anas Toma [2], Sohaib Abujayyab [3], Youness Dehbi [4], Shaker Zabadae [5] & Xin Li [6]*

[1] Department of Urban Planning Engineering, An-Najah National University, Nablus, Palestine. [2] Department of Computer Engineering, An-Najah National University, Nablus, Palestine. [3] Department of Geography, Karabuk University, Karabuk, Turkey. [4] Institute of Geodesy and Geoinformation, University of Bonn, Germany. [5] Department of Economy, An-Najah National University, Nablus, Palestine. [6] School of Urban Design, Wuhan University, Wuhan, China.

*Corresponding author: eehab@najah.edu

Received: (10/3/2022), Accepted: (16/4/2022)

ABSTRACT

Since March 5, 2020, the West Bank has faced a real crisis due to the **Coronavirus disease 2019** (COVID-19) pandemic. It has infected 581,678 people and caused 5,382 deaths so far, which has resulted in negative impacts on public health and other aspects of daily life. Based on the data provided by the Palestinian Ministry of Health, we inferred the spatial distribution patterns of the pandemic condition in different communities using **Geographic Information System** (GIS) analysis for pattern and clustering by studying the impact of urban factors on the number of confirmed COVID-19 cases. Ten urban factors were selected (i.e., population, population density, aging ratio, the hierarchy of services, health services, land use, commercial services, road density, green areas, and open spaces) to check their relation to pandemic severity using a linear model, where five factors showed a globally-significant relation. Then, the **Geographically Weighted Regression' model** (GWR) was adopted to define their unevenly distributed effects in the urban areas on the northwest bank. Among the five factors, the population factor has the most significant impact on the epidemic situation with a positive correlation. However, a negative correlation has been stated between the area of commercial services per person, population density, hierarchy of services, and health services. Finally, we provide recommendations that coordinate various urban factors to mitigate the pandemic spread. This paper will help decision-makers plan and develop different areas in Palestine and worldwide by better understanding the transmission, occurrence, and diffusion of the COVID-19 pandemic in urban areas.

Keywords: COVID-19, Urban factors, Epidemic analysis, Urban spatial patterns, Regression, Machine learning.

INTRODUCTION

Health concerns have always guided urban planning and city design. Recently, urban areas have considered disease prevention in response to previous epidemics that afflicted many urban areas at different times, such as cholera, tuberculosis, typhoid, dysentery, smallpox, and the old Spanish flu (1, 2). Nowadays, we are affected by a new pandemic, i.e., COVID-19 or COVID-19 disease, causing many adverse effects on all aspects of life. However, this provides an excellent opportunity to understand how pandemics might affect urban areas and which actions are needed to enhance their resilience and limit pandemics' impact. Throughout history, guidelines for managing infectious diseases were provided to minimize their damages and losses (3, 4, 5, 6).

Social, cultural, and age factors play an active role in influencing the pandemic spread pattern (7). Also, the socio-economic structure of the city could be affected by the COVID-19 pandemic (8). In order to control the spread of the epidemic in an urban area, we need to make urban management related to response and mitigation. The response phase includes different actions, like surveillance, response coordination, monitoring of illness trends, disease containment, mitigation, and delivery of countermeasures (9, 10). The physical aspect of response plans might include several actions, from controlling open and green areas to enhancing social distancing protocols and promoting walkability and biking. Moreover, social distancing, controlling population density inside neighborhood centers, and reducing the number of daily travels are significant in the response phase (6,10, 11, 12). Therefore, different urban factors can play a vital role in mitigation and adaptation to epidemics.

Significant urban factors related to physical dimensions can be classified into the following categories: urban form, access, infrastructure, and land use (6, 13). There are some procedures to follow for the role of the urban form if we want to mitigate the spread of the epidemics in the built environment of the urban areas (14, 15, 16). Furthermore, repurposing public and open spaces can foster resilience in epidemics by turning commercial and leisure areas like cafes, restaurants, and shops into plazas, squares, and streets and providing more pedestrian space (15). Moreover, urban transportation networks and road hierarchy significantly impact human mobility and movement as the central element in spreading the pandemic diseases (17). Bad network design and poor quality of transportation can lead to traffic jams, impeding quick access to health services for patients and doctors (18). Further, crowded public vehicles can increase the pandemic spread (6, 19).

Another important urban factor in building a city resilient to pandemics is self-sufficiency by dealing with land use. Therefore, the functions provided by the land use plan should achieve the plan's primary purpose, which should be designed to be self-sufficient so that the people can get all their needs from their neighborhoods, which means that the urban areas will work as a tree, where people movement and social interaction will be reduced (6, 19). In our globalized world, most services can end up in the heart of the urban centers, and viruses hitch a ride with them. Our urban areas may need to become more localized and self-sufficient in the future. However, factors that determine whether an urban area may have the ability to manage a pandemic can be described as follows: (1) reducing inequity and supporting equal well-being for all residents in the urban area, (2) designing and planning decentralized facilities, and services within city centers and different neighborhoods, (3) a good-distributed public open space, (4) distributing commercial areas within different areas and suburbs away from concentrating them in the urban center, (5) classifying land use according to the health conditions in pandemics, and finally (6) ensuring accessibility and quality of health services like hospitals, health centers and clinics (6, 19, 20, 21).

Many recent studies conducted a cross-sectional survey and analyzed data using multivariate regression analysis to determine the related factors to COVID-19 preventive health behaviors. This showed that the factors associated with preventive health behavior were gender, age, job, and the degree of engagement in regular physical exercise (22).

This research aims to study spatial distribution patterns of the epidemic condition in the selected Palestinian communities using pattern and clustering analysis by studying the impact of urban factors on the number of confirmed COVID-19 cases using machine learning techniques, including linear and geographically weighted regressions. These factors include population density, aging ratio, a hierarchy of services, health services, land use, commercial services, road density, green areas, and open spaces. The rest of this paper is structured as follows: Section 2 describes the study area, data sources, and hypothesis. Section 3 describes the analysis results using *Ordinary Least Square* (OLS) and Geographic Weighted Regression (GWR) techniques; in section 4, a discussion on the results is provided, whereas section 5 draws a conclusion and gives suggestions for future research.

METHODS

Study Area

This study was conducted in the West Bank - Palestine, which was named so due to its location in the west of the Jordan River, in the northern region, which includes 115 communities located in 8 governorates where 787478 people live according to the Palestinian Central Bureau of Statistics (23) on an area of 1478.3 km^2 (24). (Kur village is the less-populated community with 305 inhabitants, while Nablus city has the highest population of 164758). The map shown in (Figure 1) illustrates our study area, which is in the north part of the west bank. Many Palestinian communities have witnessed the development of urban centers and their urban environment. The development includes many economic, social, and cultural aspects. The West Bank has been facing a real crisis due to the spread of the COVID-19 disease since March 5, 2020, which affected public health and other aspects of life. In this research, we studied the spatial distribution patterns of the epidemic conditions for different communities from the northern part of the West Bank by considering the impact of urban factors on the number of confirmed COVID-19 cases (i.e., the spread of the disease). (Figure 2) shows the summary of these factors classified into three categories. Within each category, we further differentiate between a range of urban factors subject of our conducted study.

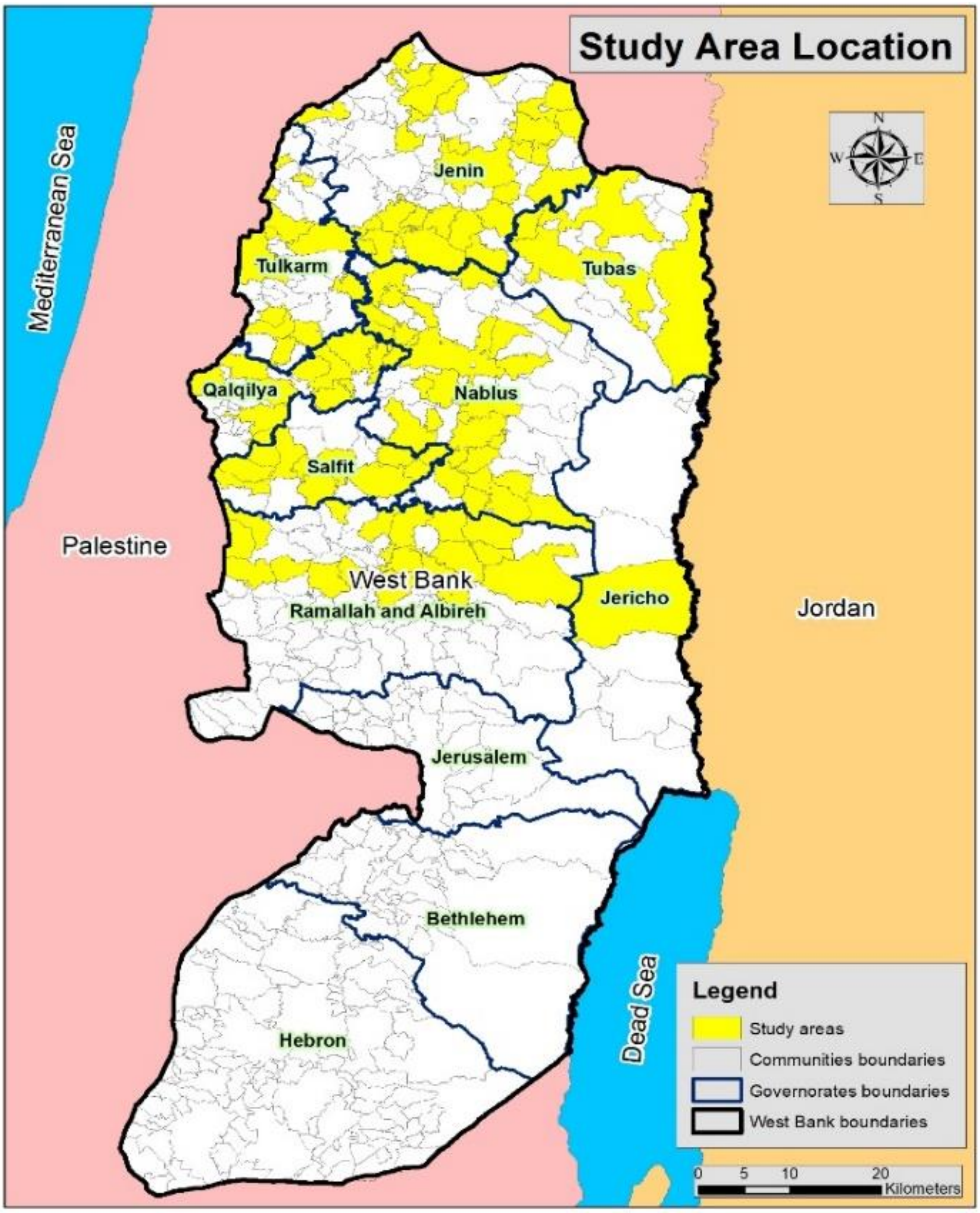

Figure 1: Study area location.

Categories	Urban Factors	Description	Source of data
Socio-demographic factors	Population	The number of populations live in each of the study communities.	Palestinian Central Bureau of Statistics, 2020
	Population density	The ratio between the population of each community divided by the built-up area (Person/Dunum).	Source of data: Palestinian Central Bureau of Statistics, 2020; Geomolg, Palestinian Local Governorate Ministry, 2020
	Aging ratio	The ratio between the number of residents over the age of 65 divided by the population of each locality.	Palestinian Central Bureau of Statistics, 2020
Urban facilities	Hierarchy of services	This variable is an indication of the size of the community and its importance in terms of the services it provides, where the localities were classified into 5 main categories as follows: 1. Normal community 2. Neighborhood center 3. Local center 4. Sub-regional center 5. Regional center	Conceptual Framework for Defining Spatial Structure for Public Service Centers in the WBGS, 2007
	Health Services	The indication was calculated by giving a mark to each health service in the community and then summing the values obtained (each hospital is given 10 marks, each health center 6 marks, and each health clinic 2 marks)	Geomolg, Palestinian Local Governorate Ministry, 2020
	Commercia l services	The ratio between the area of the areas classified as commercial use in the master plan divided by the population of each community.	Source of data: Palestinian Central Bureau of Statistics, 2020; Geomolg, Palestinian Local Governorate Ministry, 2020
	Road density	The percentage of roads area in the master plan	Geomolg, Palestinian Local Governorate Ministry, 2020
	Green areas	The percentage of green areas in the master plan	Geomolg, Palestinian Local Governorate Ministry, 2020
Land uses & Urban growth	Land use	the percentage of each land use in the master plan of each community multiplying by land use factor (1- agricultural, 2- industrial, 3-residential, 4- commercial, 5- public) and then summing the values obtained.	Geomolg, Palestinian Local Governorate Ministry, 2020
	Open spaces	The ratio between the total building area in each community divided by the built-up area	Geomolg, Palestinian Local Governorate Ministry, 2020

Figure (2): Summary of the studied urban factors.

Data Source

The data were collected from official sources as follows:

- The number of coronavirus cases 2019: Ministry of Health and Good Shepherded engineering (REF) published the number of confirmed cases daily. Data published until November 21, 2020, was considered in this study (25).
- The urban factors: Ministry of local government geospatial portal (24) was used to extract urban factors. Also, the Palestinian Central Bureau of Statistics publication was used to extract information about population and aging ratio (23, 24).

Research Hypothesis

This paper aims to study the most influential urban factors in the number of COVID-19 cases. Before the analysis, null and alternative hypotheses were accepted or rejected depending on the results.

Null hypothesis (H0). In this hypothesis, we suppose there is no relation between COVID-19 casesin the study area and the urban factors.

Alternative hypothesis (H1). In this hypothesis, the research supposes that the COVID-19 cases in the study area are affected by one or more urban factors suggested in the previous subsection 2.1.

Method and Regression Analysis Variables

Regression analysis estimates the relationship between one or more independent (or explanatory) variables and a dependent variable. Based on the relationship, a model can be built to predict the dependent variable based on the independent input variables. Our study uses it to model, check, and explore spatial relationships to better understand the factor's degree of influence. In our problem, the regression model estimates the number of COVID-19 cases (i.e., the dependent variable) based on the urban factors described in subsection 2.1 (i.e., the independent variables). The regression model (or equation) can be described as follows:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + . . . + b_nx_n, \quad (1)$$

where; $\hat{y}$ is the dependent variable, x_i, $i = 1, . . . n$ are the explanatory variables, b_0 is the estimate of the regression intercept, and b_i, $i = 1 . . . n$ coefficients are the estimate of the regression slope, i.e., the explanatory variable coefficients. The following two regression techniques were used in our analysis:

– Ordinary Least Squares regression (OLS) as a global regression method: It calculates the parameters of Equation 1 above so that the mean squared differences (errors) between the predicted and the actual values in the training set (29). The OLS is used in our study to analyze the number of COVID-19 cases and the urban factors and a prediction model that can be used to estimate the number of COVID-19 cases.

– Geographically Weighted Regression (GWR) as a local and spatial regression method Wheeler andPa'ez: It is a developed technique of the OLS, where it considers the locality variations for the relationships between the independent and dependent variables, i.e., considers the spatial dependency in regression analysis (26). Therefore, considering spatial information, the GWR technique is suitable for our study.

To evaluate the performance of regression models, the coefficient of determination metric or R^2 score is used in our paper. It can be calculated as follows:

$$R^2 = \frac{\sum(y_i - \hat{y_i})^2}{\sum(y_i - \bar{y})^2} \quad (2)$$

where; y_i is the actual value of sample i, and $\bar{y}$ is the mean of all actual samples. R^2 score evaluates the model's goodness, where the calculated value ranges from 0 to 1. The value 0 indicates that the model predicts just the mean of the training data, and the value 1 indicates that the model performs a perfect prediction. The collected data was analyzed and visualized using the ArcGIS Desktop software (version 10.7, ESRI) (27). The epidemic is assumed to be related to many urban factors in different urban environments. Therefore, 115 samples representing the northern communities in the West Bank were collected to examine the effect of the 10 factors on the epidemic situation.

RESULTS & DISCUSSION

This section introduces our conducted experiments and discusses their outcomes and findings. We used ArcGIS, Ordinary Least Square Model, and Geographic (OLS) Weighted Regression (GWR) Models for our experiments.

Ordinary Least Square Model (OLS Model)

(Figure 3) presents the final OLS map, which visually quantifies the performed model's quality. The red areas are under predictions (i.e., the correct number of COVID-19 cases is greater than the model's prediction), and the blue areas are over predictions (i.e., the correct number of COVID-19 cases is lower than the prediction). After studying the relationship between urban factor variables andthe number of COVID-19 cases, the results are given in Figure 3.

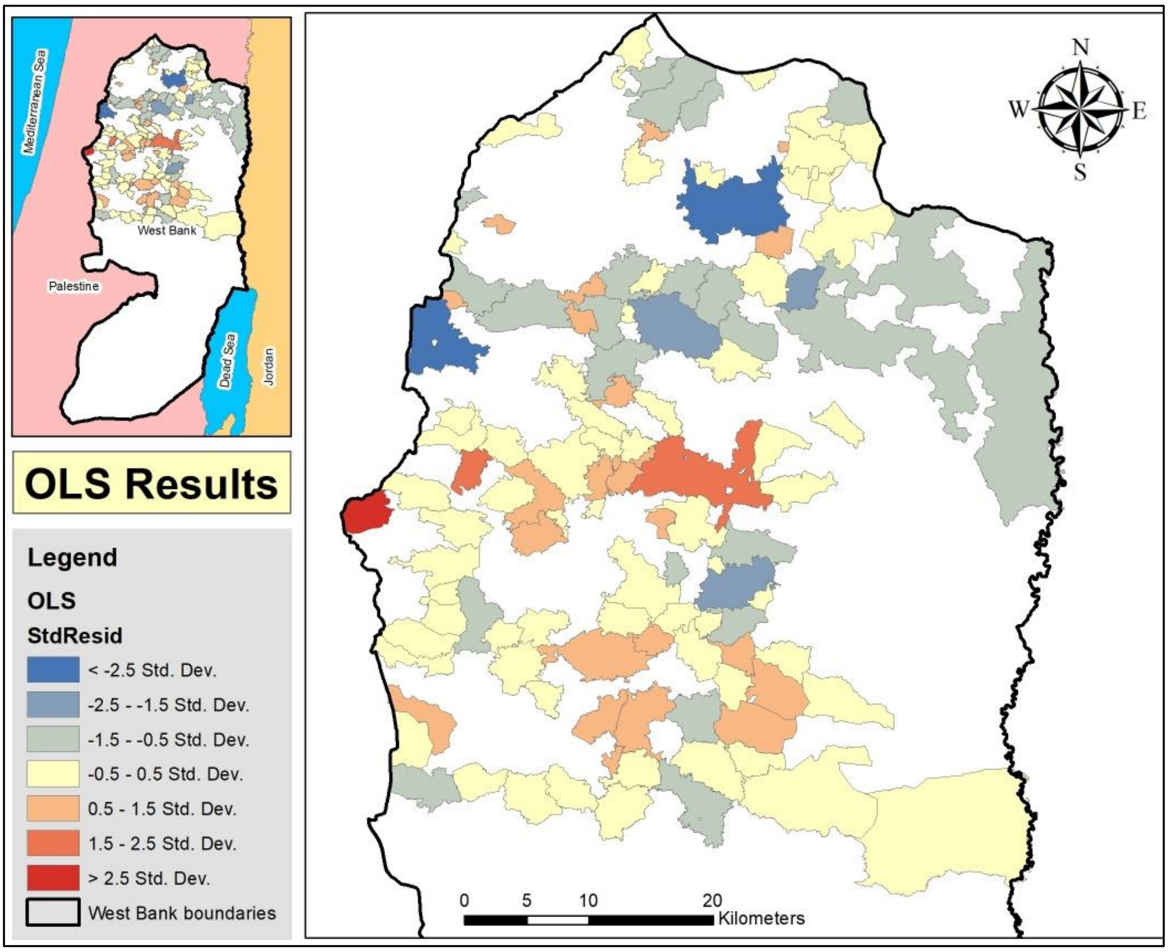

Figure (3): The results of the Ordinary Least Square– Standard Residuals.

The model represents 95.6% of COVID-19 cases in communities, and the results in Figure 4 show that the factors of population, population density, services hierarchy, health services, and commercial services are statistically significant, which will be explained in detail now. Note that the Adjusted R-Squared value is 95%, which indicates that using the five significant factors (population, population density, hierarchy of services, health services, and commercial services), the model explains 95% of the COVID-19 cases in the study area. The variance inflation factor (VIF) value detects multicollinearity in regression analysis. So, it is an indication to know if one or more variables are redundant or strongly correlated. In our regression model, there is no redundancy among explanatory variables. Therefore, we do not need to remove any variable because the values are insignificant.

The acquired probability measures the statistical significance coefficient. The probability of significant variables, i.e., < 0.05, are highlighted by an asterisk. This is also good for a variable selection were correlated or redundant variables that smear the expected model can be removed. Statistically significant variables, however, are maintained. They correspond to those having low probabilities of less than 5%. In our context, the following variables have been identified: population, population density, services hierarchy, health services, and commercial services.

A sign + or preceding each coefficient variable denotes whether the relationship is positive or negative. A positive relationship indicates that both explanatory and dependent variables have the same monotony behavior, while a negative relationship characterizes an opposite monotony. Among the five factors, the population factor was the most significant impact on epidemic situations with a strong positive relationship. Then the area of commercial services per person, population density, hierarchy of services, and health services are characterized by a negative relationship.

Both Figures 4 and 5 show the relationship between number of COVID-19 cases and statistical probability which gives the following regression line (2) that estimates the COVID-19 cases:

$$y = 258.553963 + 0.033691x_1 - 14.487682x_2 - 36.438013x_3 - 2.939463x_4 - 1446.66765x_5, \quad (3)$$

The following variables served as explanatory variables: x_1: population, x_2: population density, x_3: services hierarchy, x_4: health services, and x_5: commercial services.

Variable	Coefficient	Robust_SE	Robust_t	Probability	VIF	Model Fit	
Intercept	258.553963	213.218127	1.212627	0.177541	-	**F**	225.549517
Population	0.033691	0.002375	14.185215	0.000000***	3.907803	**Wald**	562.836680
Population density	-14.487682	7.756588	-1.867791	0.049662*	1.588201	**R2**	0.955923
Aging ratio	308.217650	358.948593	0.858668	0.537231	1.044385	**Adjusted R2**	0.951685
Hierarchy of services	-36.438013	26.907227	-1.354209	0.045789*	2.004619	**AICc**	1434.286093
Health Services	-2.939463	3.384065	-0.868619	0.030754*	3.974895	**Prob(>F), (10,104) degrees of freedom**	0.000000*
Land use	-1.546425	11.192945	-0.138161	0.941740	1.154968	**Prob(>chi-squared), (10) degrees of freedom**	0.000000*
Commercial services	-1446.66765	594.810570	-2.432149	0.002377**	1.169275	**Prob(>chi-squared), (10) degrees of freedom**	0.000079*
Road density	112.042246	130.866545	0.856157	0.559918	1.088149	**Prob(>chi-squared), (2) degrees of freedom**	0.000000*
Green areas	224.867688	96.531268	2.329480	0.390488	1.238851		
Open spaces	-307.640830	198.526413	-1.549622	0.099140	1.392675		
Note: * p < 0.05; ** p < 0.01; *** p < 0.001							

Figure (4): Summary of OLS Results.

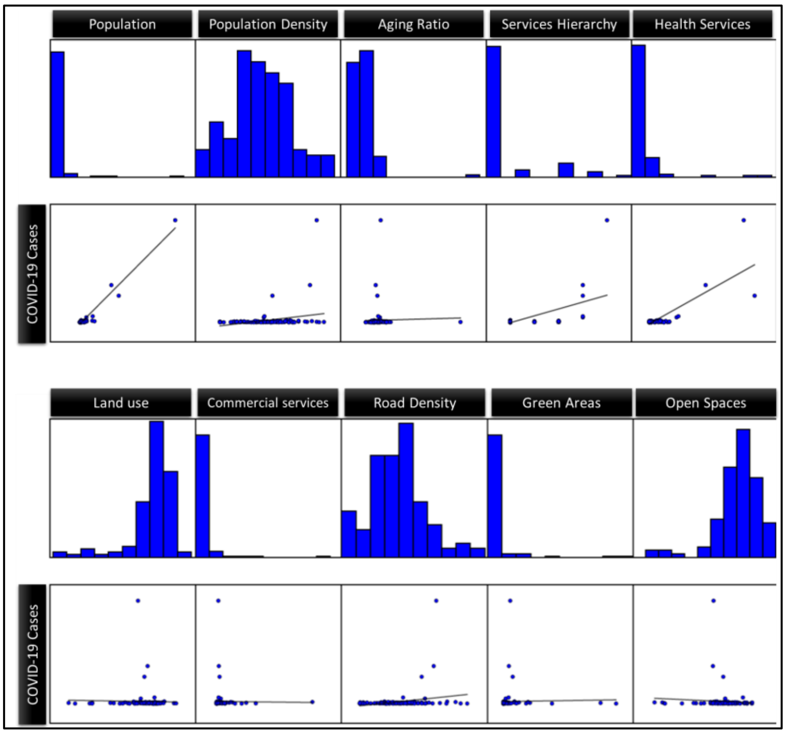

Figure (5): Graphical illustrations of the expected relationship between COVID-19 cases and urban factors.

Figure 5 shows the graphical representation of the expected relationship between COVID-19 cases and different urban factors, which indicate the distribution of samples and the type of correlation between them (positive, negative, or on relationship).

Since we are highly interested in urban factors that are primarily location-based, Spatial Autocorrelation was used to test whether the residuals show a random spatial pattern. We found that residuals are randomly distributed; we accept the null hypothesis as the z-score of 1.63 is not statistically significant (Figure 6).

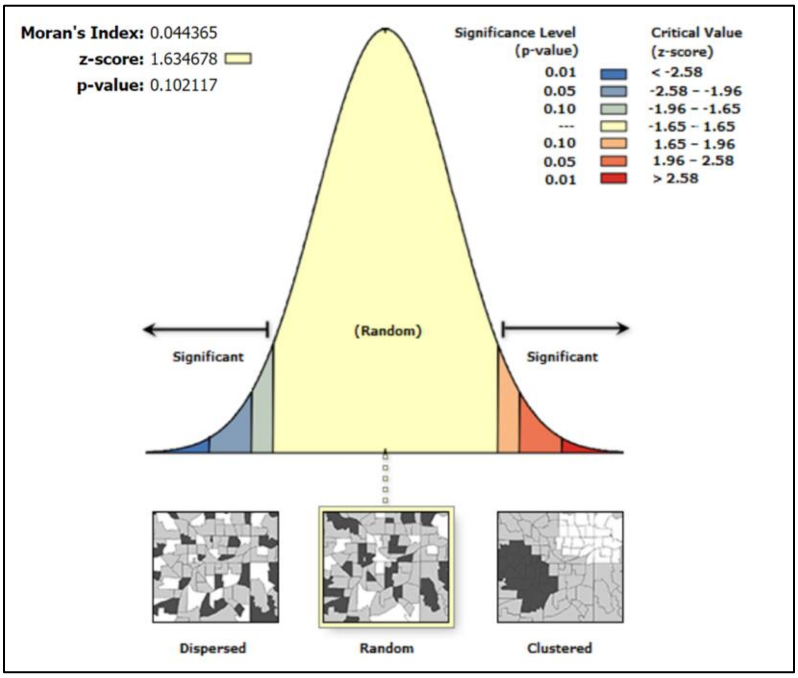

Figure (6): Spatial Autocorrelation result for the OLS regression model.

Geographic Weighted Regression Model (GWR Model)

Following the minimum description length paradigm, the factors that showed a non-significant impact on the pandemic in the OLS model have been removed from the GWR model. This allows for simple but more accurate models. The results show that the GWR model has an average adjusted R^2 of 0.98, indicating a very high explanatory power (Figures 7 and 10).

According to the mathematical model, we are now interested in choosing one model among different choices taking some observations l into account. To do that, a Bayesian approach is used in order to select a model M having the most significant posterior probability $P(M \mid l)$ out of several models M_m:

$$M = \text{argmax}_{\text{m}}\ p(l \mid M_m)P(M_m). \quad (4)$$

The model is then defined following the Akaike Information Criterion (AIC) (28) according to the following equation:

$$M_{\text{AIC}} = \text{argmax}_{\text{m}} - \log\ p(l \mid M_m) + U_m. \quad (5)$$

U_m denotes the number of the parameters of the model M_m characterizing the model complexity.

(Figure 7) also shows Akaike Information Criterion used to perform a model selection and measure the model performance. If we have more than one model with the same dependent variable, we can assess which model is the best by finding the lowest AIC value according to Equation 5, where any decrease of more than 3 points indicates a real improvement in model performance (27). By comparing the results in (Figures 4 and 7), we can note that the AIC value is lower for the GWR model than for the OLS model (OLS is 1434 while GWR is 1361). Furthermore, we can note that the Adjusted R^2 value is much higher than in the OLS model (OLS is 95.2%, whereas GWR is almost 98%), indicating that this model explains 98% of the COVID-19 case's story. This is a good improvement over the OLS model as the GWR model studies the factors locally for each community while the OLS studies them globally. The GWR map (Figure 8) presents the quality of the performed model using the urban factors variables to explain COVID-19 cases. (Figure 10) shows the Local R^2 results of the Geographic Weighted Regression model.

<table>
<tr><th>Variable</th><th>Est. (Lower)</th><th>Est. (Upper)</th><th>S.E. (Lower)</th><th>S.E. (Upper)</th><th colspan="2">Model Fit</th></tr>
<tr><td>Intercept</td><td>-81.619914</td><td>-971.870613</td><td>33.580066</td><td>84.671941</td><td rowspan="2">Local R2</td><td rowspan="2">0.330486 - 0.994516</td></tr>
<tr><td>Population</td><td>0.003766</td><td>0.049186</td><td>0.000981</td><td>0.010929</td></tr>
<tr><td rowspan="2">Population density</td><td rowspan="2">-23.474538</td><td rowspan="2">6.627001</td><td rowspan="2">6.476184</td><td rowspan="2">16.534438</td><td>R2</td><td>0.986626</td></tr>
<tr><td>Adjusted R2</td><td>0.979551</td></tr>
<tr><td>Hierarchy of services</td><td>-199.531027</td><td>23.271044</td><td>16.981113</td><td>75.540778</td><td>AICc</td><td>1361.193118</td></tr>
<tr><td rowspan="2">Health Services</td><td rowspan="2">-11.357525</td><td rowspan="2">8.866308</td><td rowspan="2">1.089281</td><td rowspan="2">6.796668</td><td>Bandwidth</td><td>11246.764884</td></tr>
<tr><td>Residual Squares</td><td>421319.972803</td></tr>
<tr><td>Commercial services</td><td>-1092.64349</td><td>1604.483208</td><td>368.953576</td><td>3343.77399</td><td>Effective Number</td><td>40.443832</td></tr>
</table>

Figure (7): Summary of GWR Results.

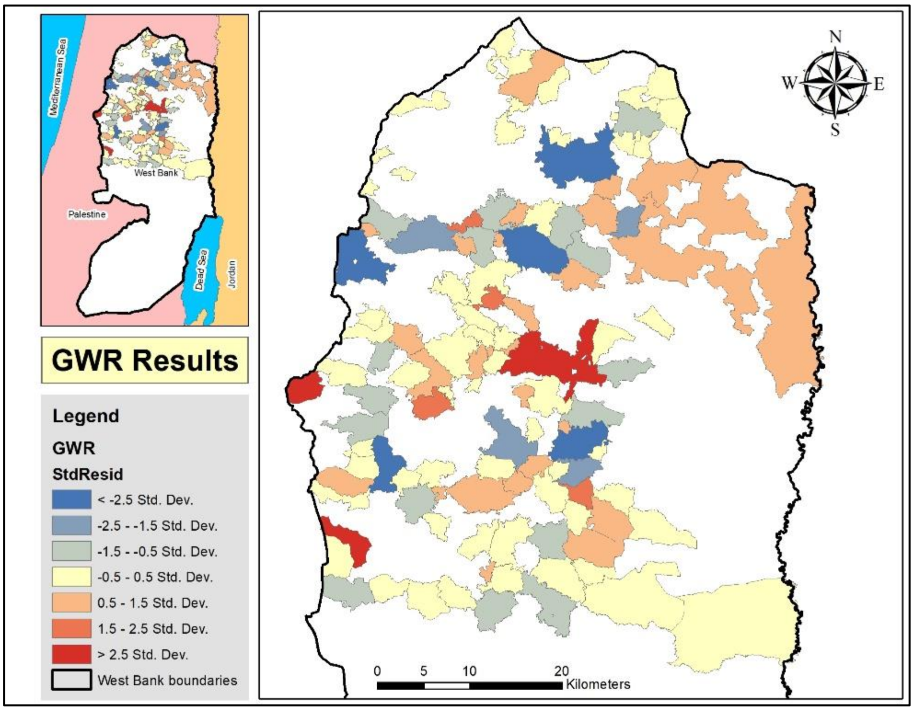

Figure (8): Geographic Weighted Regression results – (Standard Residuals).

Notice from (Figure 9) that the residuals (the model under/over predictions) show less clustering than in the OLS model. We can also observe that GWR output is a map of standardized model residuals like the OLS tool. In this context, we will make sure that the under and over predictions appear random by running the Spatial Autocorrelation tool on the Standardized Residuals; the z-score of 0.54 is not statistically significant, that is why we reject the alternative hypothesis and accept the null hypothesis of complete spatial randomness (Figures 11 and 12). (Figure 12) shows that most of the Standardized Residuals appear within the normal distribution, while very few residual values appeared in the abnormal areas of the histogram.

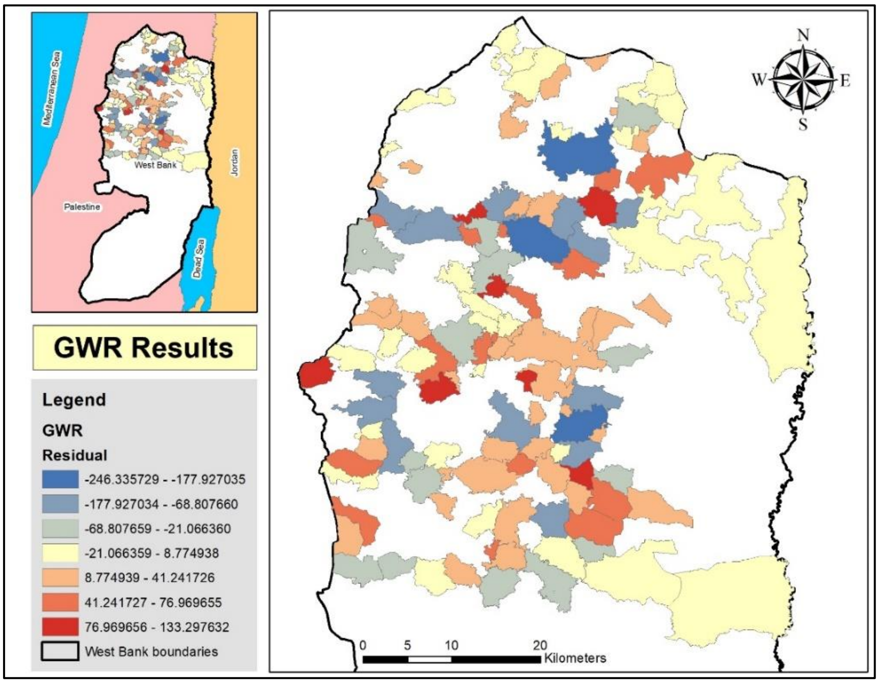

Figure (9): Geographic Weighted Regression results – (Residuals).

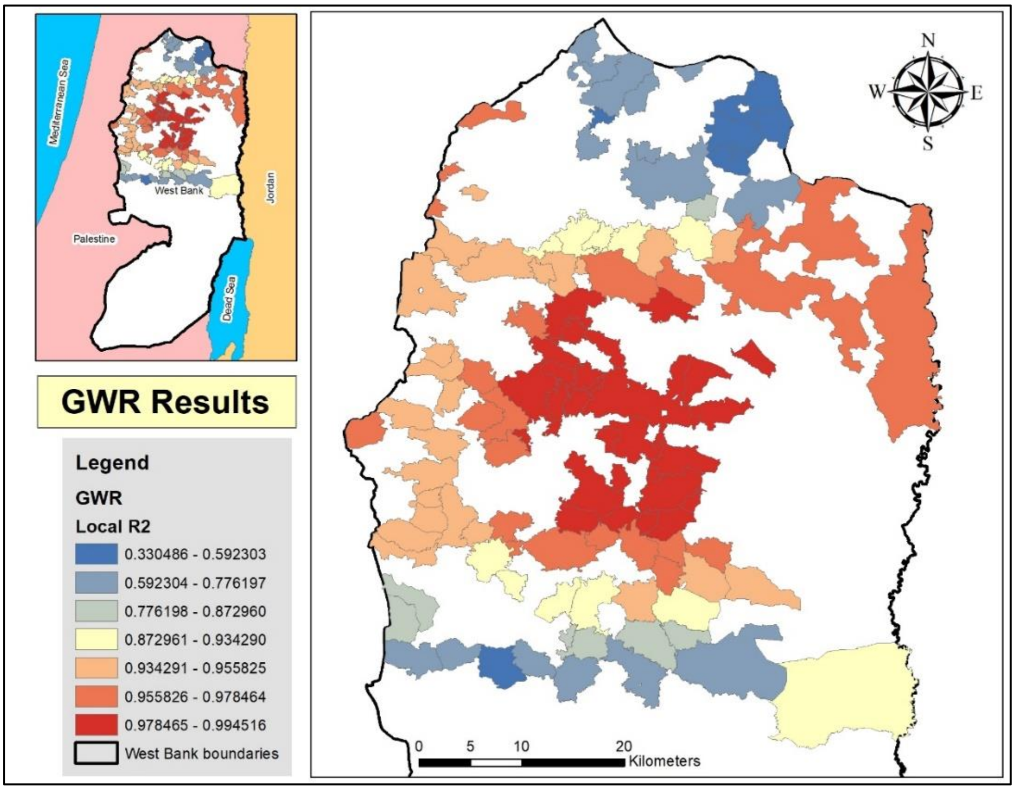

Figure (10): Geographic Weighted Regression results – (Local R^2).

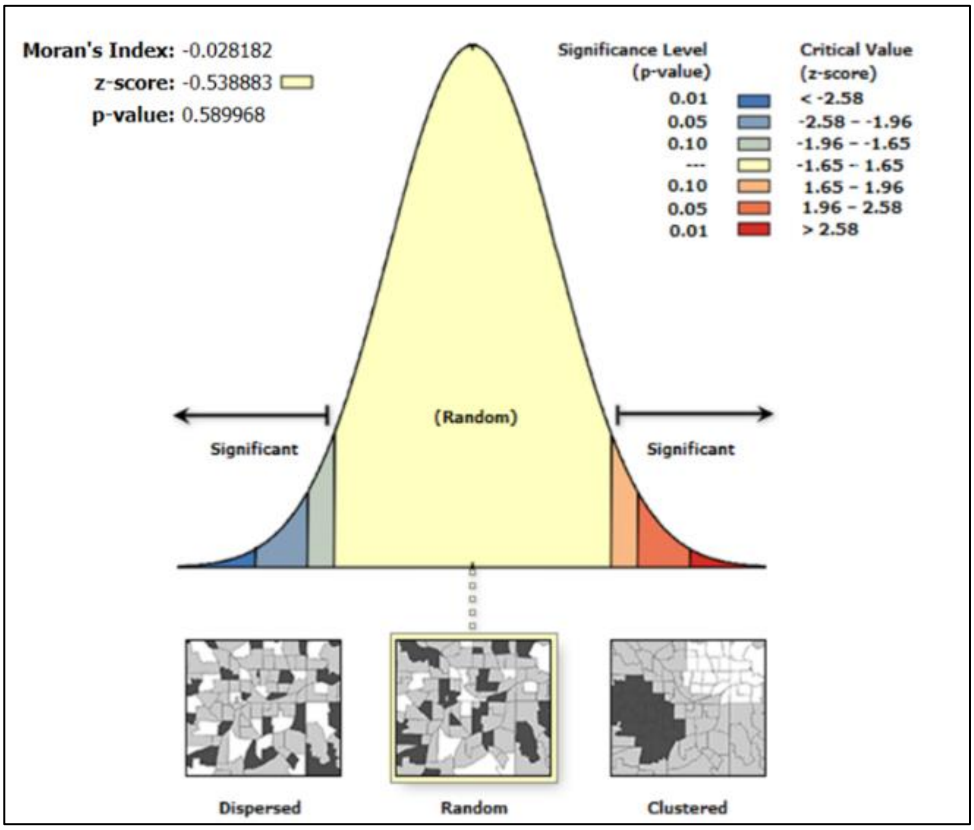

Figure (11): Spatial Autocorrelation result for the GWR regression model.

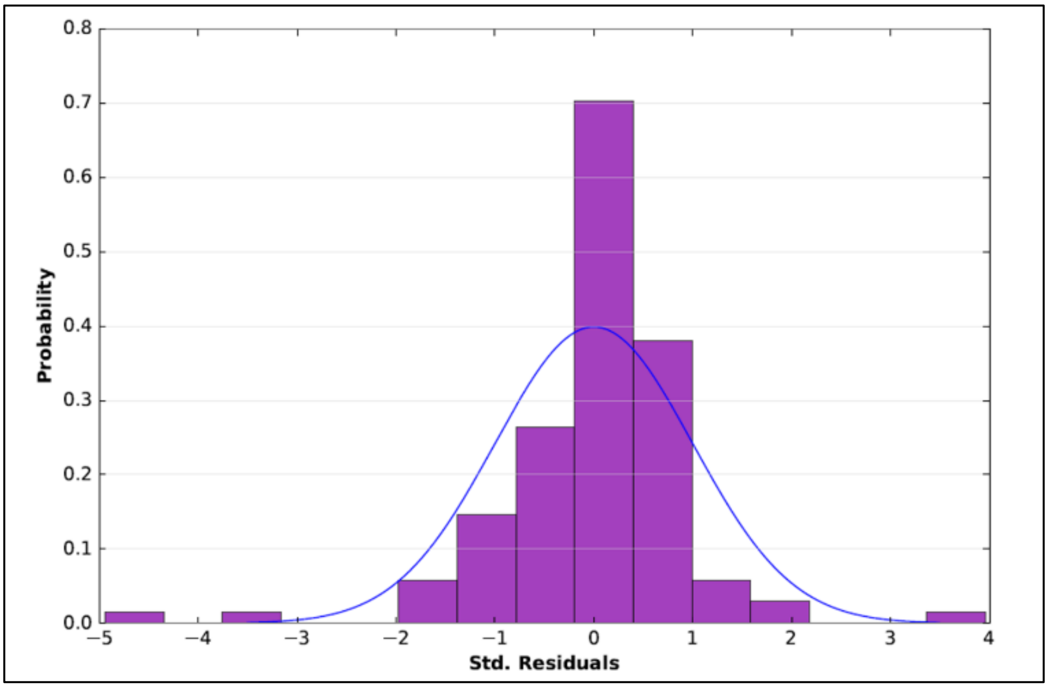

Figure (12): Histogram of standardized residuals.

Urban Factors Influence

This section quantifies the impact of urban factors on the spreading of Covid19 cases. In this context, five factors have been identified as candidates subject to closer investigation.

The Influence on Population

As we mentioned before, the coefficient variable's signs (+/-) indicate whether the relationship is positive or negative (27). We note that the coefficient for the population variable is positive,which means that as the number of people increases, the number of COVID-19 cases increases. Therefore, there is a strong positive relationship between COVID-19 cases and the population, with a probability ratio of less than 0.001, indicating that the number of populations can be considered an essentialfactor for the distribution of COVID-19 cases. For example, Nablus has a population of 164758 people,and the number of COVID-19 cases is 5199, while Qabatia has 25773 people and 55 has COVID-19 cases means that the number of COVID-19 cases is expected to increase as a result of the increase in the population. The population factor has the most explanation in the central area. For example, Al-Yamounis characterized by a high population but is not affected by this factor which may be due to the type andthe way of living in this community. In the map (Figure 13), the red areas are locations where the population variable is a solid indicator for the number of COVID-19 cases, while the blue areas are locations where the population is not a very strong indicator.

The Influence of Population density

As mentioned, the population factor has an important impact on the registered coronavirus cases. A further related factor to investigate is population density. The coefficient for the Population density variable is negative, which means that the number of COVID-19 cases decreases as population density increases. So, such a negative relationship and the probability ratio of less than 0.05 indicate that high-density areas may have a limited social life (social distancing). Furthermore, people may realize the danger of being in these areas (people's awareness), and they may feel the crowdedness in such public community facilities, which leads to more significant commitment to preventive measures, unlike in low-density areas like the rural areas. Moreover, the government restrictions are more severe and functional in high-density areas like big cities.

The population density is more explainable when we go to the south, as seen in (Figure 13). For example, Saer, a Palestinian village, has a high density but is not affected by this factor. This refers to the type of the community and the living style. The number of COVID-19 cases is not very massive, which may refer to social distancing, people's awareness, and the region's significant health and educational services distribution.

The Influence of Services hierarchy

The services hierarchy variable has a strong negative relationship with COVID-19 cases, which means that as the center hierarchy increases, the number of COVID-19 cases decreases. This indicates that the higher the center hierarchy, the lower the COVID-19 cases. A higher hierarchy means a more urbanized community serving itself and the neighboring communities. In contrast, a lower hierarchy means more rural communities, which have a different way of living and different activities for leisure time.

Moreover, large societies (urban societies) have fewer social relations. All educational, health and commercial services are available in several places. Then the number of infection cases is relatively small and vice versa for the small societies (such as rural communities). The number of COVID-19 cases in rural communities is relatively huge due to the significant social relations there. This is also attributed to the lack of health, educational and commercial services. These circumstances force rural residents to move to big cities. For example, extended families in rural areas lead to more interaction and social life between relatives. For example, Jayyous, classified as a typical rural community, has 76 corona cases, while Qabalan is classified as a local center and has only 10 cases.

The result reveals that the service hierarchy in most communities significantly affects the epidemic situation. This is because higher-level communities are considered to be more urbanized in Palestine. A second point that could be highlighted here is the change, like the services provided by each community. Thus, the pressure on the services and the number of daytime visitors are higher in high-level service centers, as shown on the map (Figure 13). The red areas in a map represent large communities (such as urban communities). Also, the area of land inhabited by city residents has a small number of cases of infection with the COVID-19. In small communities, as is the case in rural communities, the number of cases of COVID-19 infections is large.

The Influence of Health Services

The Health Service factor represents the number of health centers in each sub-area. The coefficient for the health services variable is negative, which means that the number of COVID-19 cases decreases when health services increase. Hence, the correlation between them is negative with a probability ratio of less than 0.05, which could be explained by the presence of health centers and health awareness inside communities that lead to reduce the number of cases. For example, according to its health services, Azzon has a health score of 18 with 67 COVID-19 cases, while Deir Al-Goson has a health score of 8 with 100 COVID-19 cases. Noticing the locations of significant hospitals (level 4 of health services) with the epidemic distribution, we found that these hospitals are all located within or close to the epidemic hotspots. A possible explanation is the presence of these hospitals in major cities. The health services factor is not an explanation in central west communities. For example, although Tulkarm has a hospital but is not significantly affected by that, which could be referred to the fact that people tend to depend on health services from other communities (Figure 13).

The Influence of Commercial Services

We can note that the coefficient for the commercial services variable is negative, which means that as the area of commercial services per person increases, the number of COVID-19 cases decreases. So, the correlation between COVID-19 cases and commercial services is negative—the smaller the commercial area, the more densities, clusters, and interactions between people. Furthermore, less social distancing is taken into account. For example, Tulkarm and Al-Yamoun correlate 0.007 and 0.23 dun/person and 1353 and 80 COVID-19 cases. The results reveal that the epidemic significantly affected limited communities by commercial density, especially cities considered a commercial center for the surrounding areas like Nablus and Rawabi. The commercial activities that increase the crowds vary between closed commercial centers (which reduces air circulation), shops spread on the roads, and commercial stalls, especially in cities, suffer from narrow sidewalks. In contrast, the commercial factor in Tulkarm has a non-significant effect even if it represents a commercial center. This can be attributed to lower-density commercial activities and more open areas with fresh air (Figure 13).

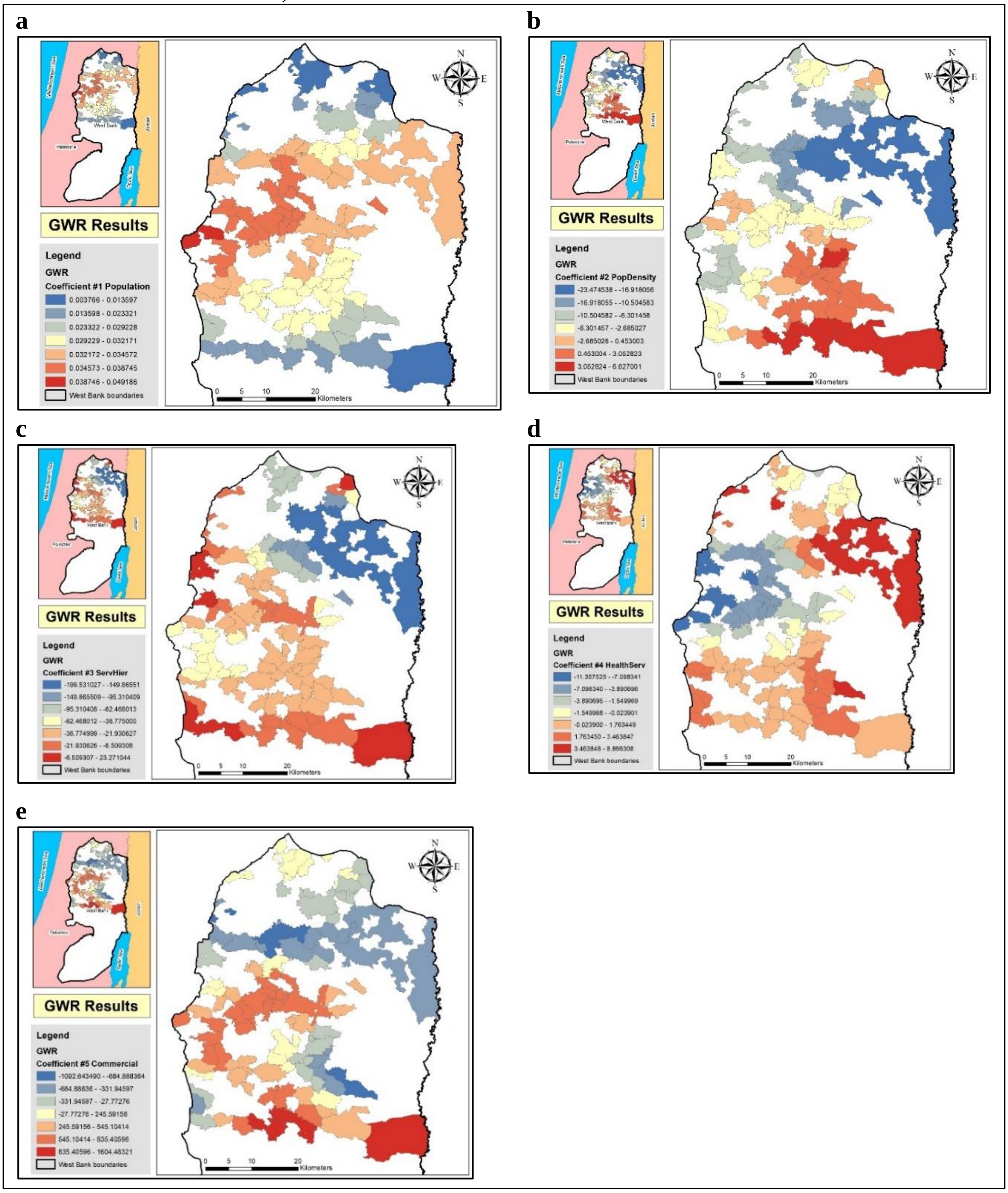

Figure (13): Geographic Weighted Regression results (Influence of Population, Population Density, Services hierarchy, Health Services, and Commercial Services).

RECOMMENDATIONS AND POSSIBLE FUTURE STRATEGIES

Based on the reported results and the discussion in the previous sections, this section provides recommendations for possible future strategies. The previous analysis showed that five urban factors affect and increase the cases of COVID-19. Therefore, we suggest the following future strategies and recommendations mitigate the impact of the urban factors on the spread of COVID-19 and avoid the worsening of the epidemiological situation in the West Bank governorates:

1. We note that the relationship between the population and the incidence of COVID-19 cases is positive. In this context, the greater the population in a particular area, the greater the possibility of infection with the virus. Therefore, housing and quarantine isolation policies must be applied, and mixing should be avoided, especially for people at high risk of COVID-19.

2. They are establishing restrictions for the expansion of areas so that they do not exceed the permissible limit in each gathering and that the number of inhabitants should be studied according to the absorptive capacity, which limits the spread of the epidemic.

3. They provide construction regulations and set restrictions on buildings to decrease population and control population density inside them, such as fixed floors and the number of apartments on each floor. Furthermore, addressing slums, camps, and crowded residential areas to avoid mixing and crowdedness and work to balance the distribution of population density in urban areas.

4. They are activating the role of the police in small societies to remind people of the danger and limit their continuous movement.

5. It prevents crowded areas from growing in residential areas, decreases social interaction by isolating buildings using gardens and green covers, and provides sufficient open spaces in a neighborhood that increase social distancing.

6. We are providing a suitable distribution for healthcare facilities within communities and sufficient health centers to provide different services to limit the spread of the epidemic.

7. For health services, we also recommend increasing the number of health service centers in all communities and raising people's awareness about how to deal with diseases.

8. Regarding the hierarchy of services, we recommend increasing cultural awareness about the spread of an epidemic among people, especially in small and rural communities.

9. The responsible authorities can also perform a new redistribution of service centers' hierarchy by putting new services in local and ordinary communities. This is expected to reduce the pressure on regional and sub-regional centers.

10. We are moving towards self-sufficiency and resource-efficient centralized neighborhoods. Rural development could be a suitable alternative to reduce the density in urban areas.

11. There should be a fair distribution of spaces for commercial services space in the master plans. This will help make them sufficient and appropriate for people and avoid concentrating them in a particular area. So this can prevent the concentration of people in one area and lead to a decrease in infection with COVID-19.

12. By increasing commercial services, the incidence of epidemics decreases by reducing the movement of people from one place to another, thus reducing infection rates and improving the ventilation system.

13. They consider the commercial centers the highest priority in the master plan operation process by increasing their area because they represent hot spots for spreading viruses.

14. We are setting restrictions on the number of people allowed to access commercial areas simultaneously, i.e., the

carrying capacity of the place.

CONCLUSION

This research provides a statistical study for modeling and analyzing COVID-19 spread in many different communities on the West Bank. Using regression models and analysis tools, we have investigated the relation between COVID-19 cases and many urban factors such as population, population density, aging ratio, services hierarchy, health care services, land uses, commercial uses, road density, and green areas and open spaces. The results show that the West Bank communities face a significant challenge in spreading the COVID-19, which reaches the red spot, i.e., the last danger stage. Furthermore, we provided thorough analyses to determine the extent of the impact of urban variables on the number of cases of COVID-19. We determined five factors globally significant to epidemic severity: population, population density, services hierarchy, health care services, and commercial services. The results show that the correlation between such factors and the number of COVID-19 cases is significant, which provides an overall picture of COVID-19 according to different urban factors.

Moreover, we provide beneficial and valuable recommendations for good urban management and planning. This will significantly help prevent the spread of epidemics in urban areas and cope with them. Furthermore, the results highlight the critical role of local governments as front-line responders in crisis response, recovery, and rebuilding. This is due to their leading role in service delivery, infrastructure investments, and mobilization of urban residents. The actions taken by local governments have been essential in addressing immediate health risks and putting in place life-saving measures. These include monitoring and tracing contacts, establishing additional health quarantine and isolation facilities, and delivering supplies and food to vulnerable communities and households. Finally, based on the provided analyzes, the communities should be prepared for any potential disaster, which should be considered in any plans. We also induce the responsible authorities to consider and implement our recommendations, which comprehensively cover all sectors and aspects of the urban variables.

Confirmation

This is to confirm that the methods carried out in this research follow relevant guidelines

Ethics approval and consent to participate

Not applicable

Consent for publication

The authors give the Publisher the Author's permission to publish the Work.

Competing interest

The authors declare that they have no competing interests.

Availability of data and materials

The data supporting this study's findings are available from the corresponding author, [I.H. H], upon reasonable request.

ACKNOWLEDGMENTS

The authors would like to thank Eng. Michael Younan from Good Sphered Engineering for providing the Covid 19 cases data.

Author's contribution

Saleh Qanazi: conceptualization, writing-original draft, data curation, formal analysis, investigation, methodology, software, supervision, validation, visualization, and writing review & editing. **Ihab H. Hijazi:** conceptualization, writing-original draft, data curation, formal analysis, investigation, methodology, software, supervision, validation, visualization, and writing review & editing. **Anas Toma:** conceptualization, formal analysis, investigation, methodology, software, supervision, validation, visualization, and writing review & editing. **Sohaib Abujayyab:** Analysis, writing review & editing. **Younes Dehbid:** Analysis, writing review & editing. **Shaker Zabadae:** Analysis, writing review & editing. **Xin Li:** conceptualization, writing review & editing.

REFERENCES

1) Lak A, Hasankhan F, Garakani SA. Principles in practice: Toward a conceptual framework for resilient urban design. Journal of environmental planning and management. 2020b Oct 14; 63(12): 2194-226.

2) Lazarević EV, Keković Z, Antonić B. In search of the principles of resilient urban design: Implementability of the principles in the case of the cities in Serbia. Energy and buildings. 2018 January 1; 158: 1130-8.

3) Musa GU. COVID-19: The Global Pandemic and Cities' Resilience| Godswill Unekwuojo Musa (GUMS). 2020.

4) Zhou C, Su F, Pei T, Zhang A, Du Y, Luo B, Cao Z, Wang J, Yuan W, Zhu Y, Song C. COVID-19: challenges to GIS with big data. Geography and sustainability. 2020 Mar 1; 1(1): 77-87.

5) Reyes R, Ahn R, Thurber K, Burke TF. Urbanization and infectious diseases: general principles, historical perspectives, and contemporary challenges. InChallenges in infectious diseases 2013 (pp. 123-146). Springer, New York, NY.

6) Lak A, Asl SS, Maher A. Resilient urban form to pandemics: Lessons from COVID-19. Medical journal of the Islamic Republic of Iran. 2020a; 34:71.

7) Komarova NL, Schang LM, Wodarz D. Patterns of the COVID-19 pandemic spread around the world: exponential versus power laws. Journal of the Royal Society Interface. 2020 Sep 30;17(170): 20200518.

8) Tricarico L, De Vidovich L. Proximity and post-COVID-19 urban development: Reflections from Milan, Italy. Journal of Urban Management. 2021 Sep 1; 10(3): 302-10.

9) Bell DM, Weisfuse IB, Hernandez-Avila M, Del Rio C, Bustamante X, Rodier G. Pandemic influenza as 21st century urban public health crisis. Emerging infectious diseases. 2009 Dec; 15(12): 1963.

10) Brizuela NG, García-Chan N, Gutiérrez Pulido H, Chowell G. Understanding the role of urban design in disease spreading. Proceedings of the Royal Society A. 2021 January 27; 477(2245): 20200524.

11) Lee VJ, Aguilera X, Heymann D, Wilder-Smith A, Lee VJ, Heymann DL, Bausch DG, Briand S, Bruschke C, Carmo EH, Cleghorn S. Preparedness for emerging epidemic threats: a Lancet Infectious Diseases Commission. The Lancet Infectious Diseases. 2020 January 1; 20(1): 17-9.

12) Lee VJ, Ho M, Kai CW, Aguilera X, Heymann D, Wilder-Smith A. Epidemic preparedness in urban settings: new challenges and opportunities. The lancet infectious diseases. 2020b May 1;20(5):527-9.

13) Golkar K. Creating sustainable place: Reflections on urban design theory. Uni Shahid Beheshti Publ. 2011: 112-46.

14) Neiderud CJ. How urbanization affects the epidemiology of emerging infectious diseases. Infection ecology & epidemiology. 2015 January 1; 5(1): 27060.

15) Chandra, S., Kassens-Noor, E., Kuljanin, G., Vertalka, J.. A geographic analysis of population density thresholds in the influenza pandemic of 1918-19. *International Journal of Health Geographics*, 2013.

16) Lin S, Qiao Y, Huang J, Yan N. Research on the Influence of Effective Distance Between Cities on the Cross-regional Transmission of COVID-19. medRxiv. 2020 January 1.

17) Hackl J, Dubernet T. Epidemic spreading in urban areas using agent-based transportation models. Future internet. 2019 Apr;11(4):92.

18) Meyer, M., Elrahman, O.. Transportation and Public Health:An Integrated Approach to Policy, Planning, and Implementation, 2019.

19) Sharifi A. Resilient urban forms: A macro-scale analysis. Cities. 2019 Feb 1; 85: 1-4.

20) World Health Organization. Infection prevention and control of epidemic-and

pandemic-prone acute respiratory infections in health care. World Health Organization; 2014.

21) Zaria Gorvett. BBC. Timber skyscrapers are sprouting up across the globe, from Vancouver to Vienna, 2017.

22) Sánchez-Arenas R, Doubova SV, González-Pérez MA, Pérez-Cuevas R. Factors associated with COVID-19 preventive health behaviors among the general public in Mexico City and the State of Mexico. PloS one. 2021 Jul 23; 16(7): e0254435.

23) Palestinian Central Bureau of Statistics, 2020.

24) Geomolg, Palestinian Local Governorate Ministry, 2020.

25) Palestinian Health Ministry. COVID19 Surveillance System, 2020.

26) Wheeler DC, Páez A. Geographically weighted regression. In Handbook of applied spatial analysis 2010 (pp. 461-486). Springer, Berlin, Heidelberg.

27) ESRI. Environmental Systems Research Institute, Inc., Regression Analysis Tutorial ArcGIS10, 2021, Redlands, CA, USA.

28) Akaike H. A new look at the statistical model identification. IEEE transactions on automatic control. 1974 Dec; 19(6): 716-23.

29) Müller AC, Guido S. Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc."; 2016 September 26.