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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Geospatial evaluation of COVID-19 mortality: Influence of socio-economic status and underlying health conditions in contiguous USA

Akinola S. Akinwumiju, Ph.D. a,*, Olawale Oluwafemi, M.Sc. b, Yahaya D. Mohammed, M.Sc. c, Jacob W. Mobolaji, Ph.D. d

a Department of Remote Sensing and GIS, Federal University of Technology, Akure, Ondo State, Nigeria
b Spatially Integrated Social Science Program, Department of Geography and Planning, University of Toledo, Toledo, OH, USA
c Centre for Geodesy and Geodynamics, Toro, Bauchi State, Nigeria
d Department of Demography and Social Statistics, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria

ARTICLE INFO

Keywords: COVID-19 mortality
Poverty
Case fatality rate

ABSTRACT

Since its outbreak, COVID-19 disease has claimed over one hundred thousand lives in the United States, resulting to multiple and complex nation-wide challenges. In this study, we employ global and local regression models to assess the influence of socio-economic and health conditions on COVID-19 mortality in contiguous USA. For a start, stepwise and exploratory regression models were employed to isolate the main explanatory variables for COVID-19 mortality from the ensemble 33 socio-economic and health parameters between January 1st and 16th of September 2020. Preliminary results showed that only five out of the examined variables (case fatality rate, vulnerable population, poverty, percentage of adults that report no leisure-time physical activity, and percentage of the population with access to places for physical activity) can explain the variability of COVID-19 mortality across the Counties of contiguous USA within the study period. Consequently, we employ three global and two local regression algorithms to model the relationship between COVID-19 and the isolated socio-economic and health variables. The outcomes of the regression analyses show that the adopted models can explain 61%–81% of COVID-19 mortality across the contiguous USA within the study period. However, MGWR yielded the highest R² (0.81) and lowest AICc values (4031), emphasizing that it is the most efficient among the adopted regression models. The computed average adjusted R² values show that local regression models (mean adj. R² = 0.80) outperformed the global regression models (mean adj. R² = 0.64), indicating that the former is ideal for modeling spatial causal relationships. The GIS-based optimized cluster analyses results show that hotspots for COVID-19 mortality as well as socioeconomic variables are mostly delineated in the South, Mid-West and Northeast of contiguous USA. COVID-19 mortality exhibited positive and significant association with black race (0.51), minority (0.48), and poverty (0.34). Whereas, the percentage of persons that attended college was negatively associated with poverty (−0.51), obesity (−0.50) and diabetes (−0.45). Results show that education is crucial to improve socio-economic and health conditions of the Americans. We conclude that investing in people’s standard of living would reduce the vulnerability of an entire population.

1. Introduction

The new scientific investigation trend in applied geography hinges on geospatial analysis to interpret and draw inferences from spatially referenced data in a problem-solving environment. The primary aim of the spatial analytical process is to measure geographic distributions, analyze patterns, map clusters, and model spatial relationships among observed variables. Hence, spatial analysis becomes vital in geographical characterization of infectious diseases because their distribution tends to be intrinsically linked with socio-economic, political, and environmental conditions that affect susceptibility.

The Severe Acute Respiratory Syndrome Coronavirus 2019 SARS-CoV-19, also known as COVID-19, belongs to the family of respiratory disease (Sobral et al., 2020). The SARS-CoV-19 was first reported on December 30, 2019, in Wuhan, China, as a pneumonia-related diagnosis (Xie et al., 2020). On January 8, 2020, a group of experts in the China
Health Commission affirmed the outbreak as a novel coronavirus originally circulating in wild animals (Zhao et al., 2020). According to Velavan and Meyers (2020), COVID-19 taints both humans and animals, which can result in a range of clinical manifestations, which are principally respiratory, starting from mild symptoms to lethal (symptomatic) and asymptomatic (no symptoms) among patients. The symptomatic patients are often characterized by fever, cough, headache, fatigue, sour throat, and difficulty breathing, to mention a few. Studies have shown that COVID-19 disease can be studied in space and time (Ai et al., 2020; Liu et al., 2020; Xie et al., 2020). The World Health Organization (WHO) officially tagged coronavirus “2019-ncov” as 2019 novel coronavirus and estimated its incubation period to be about 2–14 days. The transmission routes of COVID-19 disease can be grouped into direct and indirect routes (Xie et al., 2020). These transmission routes can either be through respiratory droplets from close direct contact with symptomatic, pre-symptomatic, asymptomatic people, indirect contact via contaminated objects, or through polluted aerosols over longer distances (Zhang & Schwartz, 2020). Thus, COVID-19 incidence rate is influenced by the rate of person-to-person contact and interactions among the humans (Gao et al., 2021; Huang et al., 2021; Huang & Kwan, 2021; Kann et al., 2021). COVID-19 could be regarded as one of the most dreaded diseases that have ever plagued the world because it can easily be contracted and kills very fast especially when a victim is having underlying health conditions. COVID-19 pandemic is characterized by suffering and a high human-to-human transmission rate. From a social perspective, COVID-19 brings about social isolation and impacted mobility because of its ability to transmit faster at proximity with the infected persons.

As at 6th of January 2022, there are over 293 million COVID-19 cases worldwide and over 5 million deaths has been attributed to this novel disease (World Health Organization (WHO), 2022). Latest global reports show that over 821 thousand COVID-19 deaths have been recorded in the United States (WHO, 2022). The challenge of COVID-19 has been of global concern because of its “unknowns” and the impacts of its emergence on virtually all aspects of life. Early studies on the COVID-19 have suggested that preexisting health conditions, air pollution, and socio-economic variables could be precursors to COVID-19 incidence and mortality (Petroni et al., 2020; Xie et al., 2020; Wu & Jennifer, 2020; Pansini & Fornacca, 2020). However, understanding the spatial distribution of COVID-19 cases, comparing mortality rates, and determining the predictors of mortality in the United States relies largely on big data. The quality of the big data would probably determine the possible level of digital manipulations that could be performed using spatially-inclined software that allows exploratory spatial analysis.

Countiests has been established that the spread and fatality of an infectious disease largely depend on physical and behavioral environmental settings of the affected area (Omran, 1971; Verhasselt, 1993; Phillip, 1995). Thus, the incidence rate and fatality of an infectious disease would vary with the dynamics of human lifestyles and behavior at any point in time and space. Also, the mortality rate of an infectious disease could be explained in terms of differences in geographic areas while incidence rate is a function of a set of physical and behavioral environmental factors (May, 1958; McGlashan, 1985; Amstrong, 1996). Where there is an outbreak of infectious disease, humans are usually exposed when interacting in high-risk areas (Shannon & Spurlock, 1975). Thus, incidence and mortality rates largely depend on the socio-economic and health status of the people (Meades et al., 1986). Thus, the socio-economic and health predictive factors of COVID-19 incidences and mortality across the US’s counties could potentially be observed.

Since the emergence of COVID-19 pandemic, many scientists have attempted to study the environmental conditions that favor the spatio-temporal spread of the viral infection (Desjardins et al., 2020; Dong et al., 2020; Xie et al., 2020). The Incidence of COVID-19 has been found to vary over space at global, continental, national, and local scales depending on the risk factors (Adekunle et al., 2020; Desjardins et al., 2020; Franch-Pardo et al., 2020; Petroni et al., 2020; Sobral et al., 2020). In particular, there is a notable geographical variation in the distribution of COVID-19 cases across the US’s counties (Desjardins et al., 2020). In addition, COVID-19 mortality also varies across counties in the US (Zhang & Schwartz, 2020).

Franch-Pardo et al. (2020) affirmed that one of the first and most cited works in spatiotemporal analysis of COVID-19 in China is that of Guan et al. (2020) that makes use of geographic information systems (GIS) and spatial analysis. Several published literatures have identified a paucity of information on the spatiotemporal analysis of COVID-19 incidence and mortality when comparing articles from the field of medicine, mathematics, and basic medical sciences across regions in the United States (Ai et al., 2020; Franch-Pardo et al., 2020; Wang et al., 2020). Liu et al. (2020) also suggested a few geographical discussions on the predictive factors of COVID-19 incidence and mortality across regions. Earlier in the United States, Roy and Ghosh (2020) have attributed COVID-19 deaths to some socio-economic and health factors, but could not put its findings in spatial perspective. Moreover, Franch-Pardo et al. (2020) argued that COVID-19 should be adequately studied in geographical and across all the geographic themes to reveal the “unknowns” associated with the pathogens and COVID-19 disease diffusion. Hence, this study examines the spatial pattern of COVID-19 mortality rate, as well as its predictors across the counties in the contiguous United States. The study further explains the relationship between COVID-19 mortality and social-economic and health variables using spatial exploratory statistics and regression models to provide a framework for monitoring COVID-19 mortality and indeed, to improve on human development and public health. In-between, we also attempt to assess the association among the dependent (COVID-19 mortality) and predictor (socio-economic and health) variables.

2. Methodology

2.1. Study area

The United States is one of the North American continent countries; it is believed to be the most powerful nation globally in terms of Gross Domestic Product (GDP). According to the United States Census Bureau (2019) projection, the population is 329, 256,465 million, with the Capital city in Washington DC. The contiguous United States has 3,143 counties as at 2020 and 5 administrative regions (Fig. 1). The contiguous United States occupy a land area of 3,796,725 square miles (9,833,517 square kilometers) with 50 states within the contiguous United States (ThoughtCo.). The temperature is mostly temperate, tropical in Florida, semi-arid along the Mississippi River, and arid in the southwest’s Great Basin (ThoughtCo). This study area includes all counties in the contiguous United States, where COVID-19 incidence and mortality data are available for the study period from January 21 to September 16, 2020. The data also includes socio-economic and health variables for the study area.

2.2. Data collection and pre-processing

Secondary data was used for this study. The dataset comprises over 40 variables, which includes COVID-19 cases count (7 and 14 days), mortality rate, COVID-19 incidence, Race (Black, White, Hispanic), % male, % female, household income, community vulnerability index, population density, % insured, % uninsured, age over 65, poverty, diabetes, obesity among others. The datasets were collected from Emory University COVID-19 Health Equity Interactive Dashboard and County health rankings and roadmaps collective initiatives of the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute. Details on the datasets and the variables are given in the supporting document.

In this study, we collated COVID-19 incidence and mortality data for 8 months (January 21 to September 16, 2020). The choice of the study...
The study period was informed by two reasons. First, the first (March–June) and second (June–September) waves of COVID-19 infection took place within the study period (Kim & Kwan, 2021). Second, the study period was characterized by uninterrupted increasing trend of COVID-19 deaths in the United States: an indication that COVID-19 mortality had managed to defile various intervention within the study period in contiguous USA. Thus, our aim is to examine the influence of the explanatory variables on COVID-19 deaths across the counties in contiguous USA. Thus, we collated 33 socio-economic and health variables as well as the supporting document. We subjected the data to cleaning and consistency had managed to defile various intervention within the study period (Kim & Kwan, 2021). Typically, Case fatality rate is not constant; it varies spatially over a specific period of time; the resulting ratio is then multiplied by 100 to derive a percentage (Khafafie & Rahim, 2021).

2.3. Model descriptions

We employ five different regression models (OLS, SLM, SEM, GWR and MGWR) to examine the relationship between COVID-19 mortality (as the dependent variables) and five explanatory variables (case fatality rate (CFR), vulnerable population (VPP), physical inactivity (PInactive), exercise opportunities (EOportunity), and poverty). CFR is the proportion of deaths that resulted from a certain disease in relation to the total number of infected individuals over a specific period of time. Case fatality rate is computed by dividing the number of deaths from a given infection by the number of individuals diagnosed with the disease over a specific period of time; the resulting ratio is then multiplied by 100 to derive a percentage (Khafafie & Rahim, 2021). Typically, Case fatality rate is taken as a measure of disease severity and is mostly utilized to predict disease course or outcome (prognosis), where comparatively high rates are indicative of relatively poor outcomes. It also can be used to evaluate the effect of new treatments, with measures decreasing as treatments improve. Case fatality rate is not constant; it varies spatially and temporally between populations, depending on the interaction among disease causative agent, the host, ecological settings as well as both pharmaceutical and non-pharmaceutical interventions (Abdollahi et al., 2020). VPP comprises of total percentage of Black, Hispanic and Native (i.e., minority inhabitants) at county level. Available data show that some counties (e.g., Claiborne, Jefferson and Holmes in Mississippi; Zapata and Zavala in Texas; Navajo and Apache in Arizona) are dominated by Black Americans, Hispanics or Natives while these races constitute the minority in some counties. In our study, Black, Hispanic and Native are captured as minority (i.e., vulnerable population) in white-dominated counties. Equally, vulnerable population consists of Black, Hispanic and Native in counties where they dominate. Therefore, VPP represents the combined population of Black, Hispanic and Natives at county level. PInactive is the percentage of adults that report no leisure-time physical activity while EOpportunity is the percentage of the population with access to places for physical activity, at county level. Poverty represents the percentage of persons living below poverty line. The adopted regression models were implemented with both GIS embedded and stand-alone algorithms. The comprehensive definitions of the regression models are presented below.

2.3.1. Global regression models

One of the assumptions of OLS is that parameter values are not dependent on each other. In this case, OLS assumes that COVID-19 mortality and the purported predictor variables vary independently from one county to another. In contrast, COVID-19 mortality and socio-economic variables exhibit spatial autocorrelation. Thus, the inability to account for spatial interactions in regression analysis makes OLS unfit for modeling the relationship between COVID-19 mortality and socio-economic and health variables (Anselin & Arribas-Bel, 2013; Mollalo et al., 2020). Thus, we also employed SLM and SEM that are both variants of OLS (Anselin, 2003; Ward & Gleditsch, 2018) and both take spatial dependence into account, but model it differently.

2.3.1.1. Ordinary least squares regression. OLS is defined as:

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + \epsilon_i$$

Where,

$y_i$ is the dependent variable, $x_{i1}, x_{i2}, x_{i3}, \ldots x_{i4}$ are the predictor variables, $\beta_0$ is the intercept, $\beta_1, \ldots, \beta_4$ are the partial regression coefficients.
and \( \epsilon_i \) is the error term. In this study, COVID-19 mortality is the dependent variable while socio-economic and health parameters (such as poverty, case fatality rate, population density, household income, education, obesity, diabetes among others) are the predictor variables. In OLS analysis, the prediction errors of sum of squared is minimized in order to optimize (Anselin & Arribas-Bel, 2013; Mollalo et al., 2020). The validity of OLS hinges on the assumptions that, variable counts are not interdependent and are fixed over the entire study site. Also, it is expected that error terms are uncorrelated (Anselin & Arribas-Bel, 2013; Mollalo et al., 2020; Oshan et al., 2020).

### 2.3.1.2. Spatial lag model. SLM is defined as:

\[ y_i = \beta_0 + x_i \beta + \rho W y_i + \epsilon_i \]

(2)

Where, 
- \( \rho \) is the spatial autoregressive variable (i.e. the spatial lag parameter), 
- \( W_i \) is a row of the matrix of spatial weights (that is, vector of the spatial weights). The origin of equation (2) is rooted in the decomposition of the error term in equation (1) (Mollalo et al., 2020; Ward & Gleditsch, 2018). Here, \( W \) indicates the neighbors around county \( i \) and, thus, accounts for the influence of the predictor variables on the dependent variable at the boundaries around county \( i \) (Anselin & Arribas-Bel, 2013; Mollalo et al., 2020).

Spatial lag is an indicator of a potential diffusion process (Kostov, 2010; Mollalo et al., 2020). Thus, SLM presumes interdependence between the predictor variables and predicted variable. Hence, the inclusion of spatially-lagged dependent variable in the regression model is in order to account for the inherent spatial interactions and dependence (Anselin, 2003; Mollalo et al., 2020; Ward & Gleditsch, 2018).

### 2.3.1.3. Spatial error model. SME is defined as (Mollalo et al., 2020; Ward & Gleditsch, 2018):

\[ y_i = \beta_0 + x_i \beta + \lambda W \epsilon_i + \epsilon_i \]

(3)

Where, 
- \( \epsilon_i \) represents the spatial component of the error, \( \lambda \) connotes the existing correlation rate among the components, and \( \epsilon_i \) denotes the non-correlated spatial error term.

The unique assumption in SEM is that spatial dependency in the error term of OLS and puts the error term in Eq. (1) into two separate terms (\( \lambda W \epsilon_i \) and \( \epsilon_i \)) (Anselin, 2003; Chen et al., 2016; Mollalo et al., 2020).

### 2.3.2. Local regression model

The general assumption of the Global regression models (OLS, SEM, and SLM) is that the dependent variable exhibits constant relationships with the predictor variables across the entire study area (Brunsdon et al., 1996, 1998; Mollalo et al., 2020). Besides, OLS measures linear relationships and serve as diagnostic tool for SLM and SEM. However, this assumption does not hold when Locations of phenomena and neighborhood criteria are considered. To accommodate spatial variability of both predictor and dependent variables, a local regression model (GWR) was proposed by Brunsdon et al. (1996) on the basis of kernel-weighted regression. In place of area-wide estimation of parameter values, GWR permits independent estimation of regression parameters for individual entities (i.e. counties); thereby, inculcating the context of spatial variability into regression model (Mollalo et al., 2020; Oshan et al., 2020).

### 2.3.2.1. Geographically weighted regression. GWR is defined as (Fotheringham & Oshan, 2016):

\[ y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon \quad i = 1, 2, \ldots, n \]

(4)

Where,
- \( y_i \) is the value for the COVID-19 incidence rate, \( \beta_0 \) is the intercept, \( \beta_j \) is the jth regression parameter, \( x_{ij} \) is the value of the jth explanatory parameter, and \( \epsilon \) is a random error term. Parameter estimates for each explanatory variable and at each county in matrix form is given by (Fotheringham & Oshan, 2016):

\[ \hat{\beta}(i) = (X'W(i)X)^{-1}X'W(i)y \]

(5)

where \( \hat{\beta} \) denotes the parameter estimates’ vector (m x 1), X stands for the selected explanatory variables’ matrix (n x m), \( W(i) \) denotes the spatial weights’ matrix (n x n), and \( y \) is the dependent variable observations’ vector (m x 1) (Fotheringham & Oshan, 2016).

The diagonal matrix (\( W(i) \)) is a function of the weights of individual parameter observation, which is proportional to its positional distance from location \( i \) and the calibration is on the basis of a locally weighted regression (Brunsdon et al., 1998; Fotheringham & Oshan, 2016; Mollalo et al., 2020). The computation of \( W(i) \) requires the specification of both kernel function and bandwidth. Among the available kernel functions, Gaussian and bi-square are mostly preferred and the functions’ bandwidth is a derivative of the Euclidean distance on population of the nearest neighbors. Nevertheless, the preferred type of bandwidth will certainly influence the local weighting occurrence neighborhood.

Despite the improvement of GWR over the global regression models as touching the variability of relationships in space, the assumption of spatial homogeneity of relationships’ scale is still hold. Consequently, GWR does not recognize local variability of spatial relationships (Mollalo et al., 2020; Oshan et al., 2019; Fotheringham et al., 2017). In contrary, the assumption of constant relationships does not always hold in space, particularly in terms of the cross-boundary spread of infectious diseases, such as the case of COVID-19 pandemic. To do away with this assumption in spatial regression analysis, multiscale geographically weighted regression (MGWR) model was introduced. As an extension of GWR, MGWR permits the spatial variability of parameter relationships and gives room for analyzing this relationships locally, which is made possible by the acceptance of multiple bandwidths, in contrast with the unified, homogeneous bandwidth in the previous regression models. (Mollalo et al., 2020; Yu et al., 2019; Fotheringham et al., 2017).

### 2.3.2.2. Multi-scale geographically weighted regression. MGWR is defined as (Mollalo et al., 2020; Fotheringham et al., 2017):

\[ y_i = \sum_{j=1}^{m} \beta_{bj} x_{ij} + \epsilon_i \quad i = 1, 2, \ldots, n \]

(6)

where \( \beta_{bj} \) denotes the chosen bandwidth with which \( b \)th relationship is calibrated (Fotheringham et al., 2017), while other parameters remain as in Eq. (1). In reality, MGWR is often regarded as a generalized additive model (GAM), which allows for the adoption of backfitting algorithms for its calibration (Buja et al., 1989; Fotheringham et al., 2017; Hastie & Tibshirani, 1986; Mollalo et al., 2020). As a GAM, MGWR is redefined as:

\[ y_i = \sum_{j=1}^{m} \beta_{bj} x_{ij} + \epsilon \]

(7)

where \( \beta_{bj} \) (replacing \( \beta_{bx} \) in equation (3)) denotes the jth additive term and is an applied smoothing function to jth explanatory variable at county \( i \) (Mollalo et al., 2020; Oshan et al., 2019; Fotheringham et al., 2017). In this case, the model calibration will culminate into the establishment of a seterate bandwidth for every \( j \) predictor variables. Differences in spatial scales are represented by different bandwidths. Therefore, by accounting for the influence of variability in spatial relationships, MGWR has a comparative efficacy to integrate local context into parameter relationships in space (Mollalo et al., 2020; Oshan et al., 2019; Fotheringham et al., 2017).
2.4. Analytical procedure

In order to isolate the real predictor variables that influence COVID-19 mortality, we subjected the data to stepwise regression analysis in order to isolate the explanatory variables. In this case, COVID-19 mortality is the dependent variable while other variables (including COVID-19 incidence) were entered into the model as potential predictor variables. The outcome of the initial regression analyses showed that 13 variables could explain COVID-19 mortality in USA. Spatial regression diagnostic analysis showed that % Black, % Hispanic and % Native are spatially correlated. Thus, we generated a composite variable (named vulnerable population (VPP)) from these parameters (i.e., % Black, % Hispanic % Native and % minority). Thereafter, we subjected final set of potential predictor variables to exploratory regression analysis with aim to establish a passing model.

To establish a relationship between the isolated predictor variables and the dependent variable (COVID-19 mortality), we adopted five different models that were proposed by Mallalo et al. (2020). These include three global models (ordinary least squares (OLS), spatial lag model (SLM), spatial error model (SEM)), and two local models (geographically weighted regression (GWR), and multiscale GWR (MGWR)). In this study, we deliberately employ five different regression models for two reasons. First, to assess the comparative strengths and weaknesses of global and local regression models. Second, to compare the outputs of GIS embedded spatial regression algorithms with that of stand-alone algorithms. For a start, we subjected the potential predictor variables to OLS regression in ArcGIS 10.6 environment. Based on the variance inflation factor (VIF) and robust probability (b), the OLS result indicates that a passing model for spatial regression can only be built on case fatality rate (CFR), vulnerable population (VPP), percentage physical inactivity persons (PInactive), percentage accessibility to public spaces for physical exercise (EOpportunity) and poverty. Meaning that, only the five variables passed the test of multi-collinearity, hence, they were taken as the major predictor variables that influenced COVID-19 mortality across US counties during the first eight months of the global pandemic. Also, we subjected all variables to correlation analysis significance based on the available cumulative data for the study period. In the case of COVID-19 mortality, 22% and 40% of the counties were mapped into clusters of hot and cold spots respectively, while the COVID-19 mortality was neither significantly high nor low across 38% of the counties. As at September 16, 2020, 26% and 28% formed the clusters of hot and cold spots of case fatality rate (CFR) while the remaining 46% was characterized by insignificant clusters of CFR. In the same period, 33% and 47% of the counties were characterized by clusters of high and low percentage of vulnerable population (VPP) respectively, while significant cluster was not noticed in the remaining 20% of the counties. As at September 16, 2020, 33% and 26% of the counties were categorized into clusters of high and low percentage of physical inactivity persons respectively, while the remaining 46% of the counties had neither hot nor cold spot in terms of the concentration of physical inactivity persons. For the period of study, 25% and 29% of the counties were mapped into clusters of significantly high and low accessibilities to places for physical activity respectively, while access to places of physical activities in the remaining 46% of the counties was neither significantly high nor low. Results show that hot spots for poverty, case fatality rate, vulnerable population, physical inactivity persons, as well as COVID-19 incidence and mortality are delineated in Western and Southern regions of contiguous USA. In the same vein, the hot spots for accessibilities to places of physical activity are mostly delineated in South with the occurrences of minor clusters in the North (Fig. 2).

3. Results

Between the 1st of January and 16th of September 2020, 6,277,961 incidences of COVID-19 infection were recorded for the contiguous United States. While 97.3% of the infected persons successfully recovered, the remaining 2.7% (170,329) died from COVID-19 related complications. Available data show that the highest incidence (255049) and death (6273) were both recorded for Los Angeles county in the State of California. However, the highest mortality rate was recorded for Hancock county in Georgia State. The highest case fatality rate (33.33) was recorded for Throckmorton in Texas State. In the year 2020, the highest poverty level was recorded for Todd county in South Dakota. The highest percentage of adults that reported no leisure-time physical activity (denoted as PInactive) was recorded for Tattnall county in Georgia while the highest percentage of the population with access to places for physical activity was recorded for Teton county in Wyoming State. The highest percentage of vulnerable population (VPP) (i.e., the aggregate percentage of black, Hispanic, natives and minority) was recorded for Starr county in Texas.

3.1. Hot and cold spot analysis of socio-economic and health variables

In geospatial analysis, hot spot analysis is often employed to map regions of statistically significant clusters of high and low occurrences of events. Hot spot analysis usually delineates three zones within the sphere of occurrence of an event: the hot spot, cold spot and not significant. The hot spot indicates significantly high occurrence and cold spot denotes significantly low occurrence while the not significant connotes neither high nor low occurrence of a particular event. Thus, we employ the GIS-based Optimized Hot Spot (OHS) algorithm to categorize the contiguous US counties to clusters of significantly high, low and not-significant occurrences of the examined variables (Fig. 2). For the study period, COVID-19 incidence was neither significantly high nor low in 94% of the counties while the remaining 6% was characterized by clusters of significantly high COVID-19 incidence in the contiguous USA. It is worth noting that cold spot was not delineated for COVID-19 incidence based on the available cumulative data for the study period. In the case of COVID-19 mortality, 22% and 40% of the counties were mapped into clusters of hot and cold spots respectively, while the COVID-19 mortality was neither significantly high nor low across 38% of the counties. As at September 16, 2020, 26% and 28% formed the clusters of hot and cold spots of case fatality rate (CFR) while the remaining 46% was characterized by insignificant clusters of CFR. In the same period, 33% and 47% of the counties were characterized by clusters of high and low percentage of vulnerable population (VPP) respectively, while significant cluster was not noticed in the remaining 20% of the counties. As at September 16, 2020, 33% and 26% of the counties were categorized into clusters of high and low percentage of physical inactivity persons respectively, while the remaining 46% of the counties had neither hot nor cold spot in terms of the concentration of physical inactivity persons. For the period of study, 25% and 29% of the counties were mapped into clusters of significantly high and low accessibilities to places for physical activity respectively, while access to places of physical activities in the remaining 46% of the counties was neither significantly high nor low. Results show that hot spots for poverty, case fatality rate, vulnerable population, physical inactivity persons, as well as COVID-19 incidence and mortality are delineated in Western and Southern regions of contiguous USA. In the same vein, the hot spots for accessibilities to places of physical activity are mostly delineated in South with the occurrences of minor clusters in the North (Fig. 2).

3.2. Relationship between COVID-19 mortality and the examined explanatory variables

The initial results of the exploratory and stepwise regression analyses reveal that 13 of the 33 potential predictor variables could explain COVID-19 mortality at county level across the contiguous USA. The initial modeling with OLS reveals that only five out of the suggested 13 could be built into a passing model in GIS environment. Thus, we build our spatial regression models on five predictor variables, viz: CFR, VPP, EOpportunity and poverty. The final set of predictor variables have very low variance inflation factors, which is an indicator of acceptable level of multi-collinearity among the predictor variables. It is
worth mentioning that GIS embedded spatial regression algorithms will not run when multi-collinearity exceeds the threshold of 5 (Mollalo et al., 2020; O’brien, 2007). While COVID-19 mortality exhibits positive relationships with CFR, VPP, PInactive and poverty (P = 0.000), it is negatively associated with EOpportunity (P < 0.002) (Table 1). The OLS regression model accounts for 61% of the regression plain (adjusted R²).

Table 1
OLS summary statistics for the COVID-19 mortality predictor variables over contiguous USA.

| Variable     | Coefficient | T-Statistic | P-Value     | Robust_t | Robust_Pr [b] | VIF |
|--------------|-------------|-------------|-------------|-----------|---------------|-----|
| Intercept    | -33.11      | -8.63       | 0.000000*   | -8.37     | 0.000000*     | -  |
| povety       | 0.64        | 6.14        | 0.000000*   | 4.52      | 0.000009*     | 1.54|
| CFR          | 12.37       | 51.27       | 0.000000*   | 16.74     | 0.000000*     | 1.03|
| VPP          | 0.47        | 28.76       | 0.000000*   | 21.32     | 0.000000*     | 1.29|
| PInactive    | 0.76        | 6.88        | 0.000000*   | 5.86      | 0.000000*     | 1.35|
| EOpportunity | -0.08       | -3.75       | 0.000195*   | -3.10     | 0.001979*     | 1.20|

Fig. 2. Cluster maps of some selected variables showing the spatial pattern of hot and cold spots across contiguous USA based on year 2020 datasets. Hotspot represents the cluster of very high parameter value while cold spot denotes a cluster of very low parameter value. Poverty and PInactive hotspots are mainly delineated in the Southern region. CFR hotspots were delineated in the Northeast and South while that of vulnerable population are delineated across the southern axis of the coastal States, extending from California (in the West) to Massachusetts (in the Northeast). EOpportunity cold spots are mainly delineated in the South. COVID-19 incidence hotspots are delineated in the coastal states (in the West), Florida (in the South) as well as in Northeastern States (i.e., Massachusetts, New York, Pennsylvania, New Jersey and Maryland). COVID-19 mortality hotspots are mainly delineated in the southern and northeastern coastal States.
dictor variables could only account for 61% of COVID-19 mortality (Adj. R² = 0.61), which is quite significant. The interpretation is that the performance of the global models is the order of SEM (Adj. R² = 0.68) is slightly higher than that of SLM (Adj. R² = 0.64). Nevertheless, lower value of standard error is recorded for SLM. Despite the significant performances of the global models, the predictor variables could only account for 68% of COVID-19 mortality in the contiguous USA. In this case, 32% of COVID-19 mortality is unaccounted for by the explanatory variables. Thus, the performances of the global regression models could be improved upon if the relationships between the dependent variable and the predictor variables are allowed to vary locally (Mollalo et al., 2020).

To capture the spatial variability of the relationship between dependent and predictor variables at county-level, we implement two local regression models: GWR and MGWR. The local regression models yield an additional 12-13% explanation of the regression plain, which is a significant improvement over the best of the global models (see Table 2). In addition, the average AICc value for the local regression (4074.2) is significantly lower than that of global models (29950.2). On the basis of individual model performance, MGWR has the best goodness of fit with the highest adjusted R² (0.81) value and the lowest AICc value (4031.27). Generally, both the global and local regression models account for substantial percentages of the regression plain. While the efficiency of the global models is the order of SEM (Adj. R² = 0.68) > SLM (Adj. R² = 0.64) > OLS (Adj. R² = 0.61), that of local models is in the order of MGWR (Adj. R² = 0.81) > GWR (Adj. R² = 0.80).

The coefficient maps of GWR and MGWR for the predictor variables are presented in Fig. 3. CFR portrays virtually the same patterns of explanation of the spatial occurrence of COVID-19 mortality across the counties of contiguous USA as modeled by GWR and MGWR. CFR predominantly explains COVID-19 mortality across the counties in Mississippi, Georgia, South Carolina, North Carolina and Florida (Fig. 3a). In contrast, CFR performs poorly in terms of COVID-19 mortality prediction across the West (Particularly in Washington, Idaho, Oregon, Utah, California, and New Mexico), the Midwest (Particularly in Ohio), the South (in West Virginia) and the Northeast (partially in New York, Maine, Vermont, Massachusetts, Pennsylvania and New Hampshire). Vulnerable population is the second most effective predictor variable that give fair explanation to COVID-19 mortality at county-level across contiguous USA. Poverty has almost the same pattern with CFR except that the former only portrays poor performance in the Northeast. Results show that VPP give the most efficient prediction of COVID-19 mortality in Northeastern states (Maine, Vermont, New York, New Hampshire, Massachusetts and New Jersey) as well as some counties around the boundaries of West Virginia, Virginia and Kentucky, in the South. But, VPP portrays poor performance in some states in the South (Alabama, Georgia and Florida) and West (Colorado, Utah, Arizona, New Mexico and Nevada). Both PInactive and EOPportunity have similar pattern with VPP. We notice an obvious disparity in the spatial pattern of coefficient maps of PInactive as modeled by GWR and MGWR. However, there is a considerable agreement between the outputs of ArcGIS-based GWR and that of MGWR 2.2.1.

The spatial patterns of local R² values of the local regression models (GWR and MGWR) are presented in Fig. 3b. Local regression models demonstrate optimum prediction strength in the Midwest (in states such as Nebraska, Kansas, Missouri, Arkansas, Minnesota, Iowa, Illinois, Indiana, Ohio and Michigan), the Northeast (in states such as New York, Massachusetts, New Jersey, New Hampshire, Vermont and Maine) and the South (in states such as Kentucky, Tennessee, Georgia, Florida and South Carolina). In contrast, the local regression models put forth poor performances in the West (i.e., in Arizona, New Mexico, Colorado, Wyoming, Idaho and Montana) and the South (particularly in Texas and the border counties of West Virginia, Virginia and North Carolina).

### 3.3. Correlation analyses of variables

The results of the correlation analysis are presented in Table 4. Results show that poverty exhibits positive and significant association with diabetes (0.45), obesity (0.36), percentage of uninsured persons (0.34), mortality (0.34), black race (0.46) and minority (0.45). On the other hand, poverty is inversely related to the percentage of persons that attended college (−0.51) and household income (−0.64). The percentage of persons that attended college exhibits positive and significant relationship with household income (0.76) while it is negatively but significantly associated with poverty, diabetes (−0.45) and obesity (−0.50). The percentage of uninsured persons is positively and significantly associated with Hispanic, minority, poverty, mortality, diabetes, black race and obesity, in descending order of correlation strength. COVID-19 mortality is positively and significantly correlated with black race (0.51), minority (0.48), poverty (0.34), diabetes (0.22), percentage of uninsured persons (0.21) and hispanic (0.16).

### 4. Discussion

In this geospatial study, we attempt to examine the influence of socio-economic and health variables on COVID-19 mortality across the counties of contiguous USA. To achieve this feat, we assemble 33 potential explanatory variables at county-level. In order to isolate the major predictor variables, we subject the variables to stepwise and GIS-based exploratory regression analyses. Preliminary results show that only 13 out of the 33 variables have the potential to explain COVID-19 mortality in the study area. Thereafter, we subject the prospective predictor variables to spatial regression integrity test. We find out that only five variables (i.e. CFR, VPP, PInactive, EOpportunity and poverty) could give meaningful explanation to COVID-19 mortality across the contiguous USA. Thus, we adopt three global and two local regression models to establish an explanatory relationship between the dependent and predictor variables. Case fatality rate (CFR) is a measure of how severe and deadly a disease is. High CFR indicates high probability that an infected person has a slim chance to survive. Vulnerable population (VPP) comprises of Black, Hispanic and Native (i.e., minority people) living in the United States. Among these groups, black race has been the most affected by COVID-19 mortality (Laurencin and Joanne, 2020; Millett et al., 2020). The physical inactivity (PInactive) persons are the individuals who have no access to physical exercises. In the case of COVID-19, higher percentage of PInactive could be an advantage as the spread along this line might be extremely low. On the other hand, physical inactivity persons might suffer more complications when infected by COVID-19, due to the poor state of health that emanate from limited or no physical exercise. In the same vein, accessibility to public spaces for physical exercise could aggravate the spread of infectious disease such as COVID-19. On the other hand, persons that have access to public places for regular physical exercise could have a better chance to survive COVID-19 infection due to physical fitness and good state of health. Poverty usually exposes the humans to hardship and various ill-health conditions. The inability of the poor to afford balanced diet could expose an individual to diseases that could reduce the chance of survival when infected with COVID-19.
Our results provide an insight to the spatial pattern of COVID-19 mortality in the contiguous USA. For instance, COVID-19 mortality is more impactful in the South and Northeast due to high spatial intersection of poverty, VPP and PInactive (see Fig. 2). We observe that COVID-19 mortality determinant variables are more pronounced in the Northeast, Midwest and South. Thus, the prevalence of these determinant variables may increase public health concerns for COVID-19. Our result can be helpful in capturing both social and health determinants of COVID-19 mortality across the three regions. And these patterns reveal an evident spillover effect of the examined socio-economic and health variables across the contiguous USA. The observed spatial pattern of COVID-19 mortality would be a result of consequent spillover effect of the explanatory variables across the three regions (Krisztin et al., 2020; Zhang et al., 2020). Although, mobility and non-pharmaceutical interventions (NPI) could be responsible for these spillover effects. For instance, Kim and Kwan (2021) affirmed a strong association between COVID-19 incidence and mobility in the United States between March and September of 2020. The study further established a strong correlation between restriction policy compliance and COVID-19 incidence within the study period. Thus, non-compliance with COVID-19 intervention policy and guidelines could also result to increased mortality.

The results of the optimized hot spot analyses indicate that substantial percentage of the COVID-19 infected persons are liable to die due to poor socio-economic and health conditions. We observe that relatively high poverty level and percentages of vulnerable persons and people with no access to physical activities, as well as the lowest access to public spaces are found in the South, Northeast and Midwest. The significant contribution of CFR to the regression plain indicates that early stage COVID-19 mortality is more of a function of its fatality than socio-economic and health conditions. However, available data show that socio-economic and health conditions are also potent determinant of COVID-19 mortality across the counties of contiguous USA. In particular, the vulnerable population (i.e., black race, hispanic, minority and native) has suffered more COVID-19 deaths in USA within the study period. Similar findings have earlier been reported by Petroni et al. (2020), Maroko et al. (2020) and Saffray et al. (2020). To be specific, Pan et al. (2020) and Saffray et al. (2020) have reported positive and significant relationship between COVID-19 epidemic and percentage Black, Asian and Minority Ethnic (BAME) individuals in the contiguous USA.

Though the results of all the regression models are generally optimal, the most reliable results are recorded for the local models. Perhaps, the main strength of the local regression model is its ability to model the variability inherent in the interaction among spatial phenomena (Sun et al., 2020). And the ability of MGWR to capture the variation of differential scale of the spatial interaction makes it the most suitable local regression model for examining the relationship between COVID-19 epidemic and ecological variables (Mollalo et al., 2020). However, it is important to note that a generalized prediction model could only be built on global regression model. Thus, the legacy of global regression models cannot be undone by the efficacy of local regression models.

Our results reflect slight difference in the spatial patterns of the coefficient of determination of GIS-embedded GWR and that of the stand-alone MGWR algorithm. Furthermore, MGWR algorithm does not have the capacity to present its results in form of maps. In this case, results must be exported to a GIS environment before MGWR results could be visualized in form of map. Therefore, embedding the MGWR algorithm in a standard GIS software would reduce user stress and also give room for result comparison. Moreover, the results of the regression analyses show that the prominent influencing factors of COVID-19 mortality vary significantly across the regions of USA. And this underscores the importance of spatial context in modelling the outbreak of infectious diseases.

All the outputs of this study reveal an interesting pattern that reflects cross-border spatial autocorrelation of events (also referred to as spillover effect). For instance, we notice two spatial regimes in COVID-19 mortality: first, a situation whereby mortality epi-centers emerge within some states and spread towards the boundaries. Second, a situation whereby the epi-centers emerge within the border counties and spread into the core counties of the neighboring states. This observation underscores the comparative efficacy of local regression models over their global counterparts (Sun et al., 2020)

The results of the regression models reveal that COVID-19 mortality is more related to socio-economic and health conditions in the North-east, South and Midwest. Earlier, studies have attributed the impacts of COVID-19 to the socio-economic disadvantages and inequalities arising from the pandemic itself (Ahmed et al., 2020; Mollalo et al., 2020). We observe that the severity of the impacts of COVID-19 is a function of cumulative inappropriate health and socio-economic conditions of the people. Recently, Kim and Kwan (2021) affirm strong relationship between COVID-19 incidence and political inclination in the United States. Our results show that spatial pattern of COVID-19 mortality does not correlate with political inclination. Rather, we find that the spatial pattern of COVID-19 mortality correlates well with poverty, PInactive, CFR and VPP in the United States. Just as earlier studies (e.g., Kim & Kwan, 2021; Maroko et al., 2020; Petroni et al., 2020; Saffray et al., 2020) have attributed COVID-19 incidence to poverty, our results show that COVID-19 mortality is highly related to poverty in the United States. While Kim and Kwan (2021) attributes COVID-19 incidence among the poor persons to high mobility of essential workers, our study shows that COVID-19 mortality is influenced by the socioeconomic status and health conditions of the infected persons in the United States. For instance, poverty ridden persons may be prone to underlying health conditions such as diabetes, obesity and upper respiratory tract infections. These may be as a result of eating habits or inappropriate access to health care. Such underlying health conditions in turn expose the sick individuals to complications when infected with COVID-19. For example, Gupta et al. (2021) observed that many of the COVID-19 inpatients in a black American dominated community had one or more comorbidities such as hypertension, diabetes and chronic kidney disease. Thus, the observed high COVID-19 mortality in Northeast is strongly connected with poverty and the consequent inappropriate health conditions. Hence, investing in the improvement of people’s standard of living is another way of boosting the immunity of a country against unforeseeable outbreak of infectious disease such as COVID-19. We observe that COVID-19 mortality does not correlate with COVID-19 mortality.

### Table 3

| Variable | Coefficient | Std. error | Z-score | P-value |
|----------|-------------|------------|---------|---------|
| Intercept | -30.35      | 15.71      | -15.71  | 0.00000 |
| poverty  | 0.57        | 0.35       | 10.10   | 0.01055 |
| CFR      | 11.44       | 12.31      | 0.94    | 0.00000 |
| VPP      | 0.32        | 0.48       | 0.02    | 0.00000 |
| PInactive| 0.71        | 0.15       | 0.02    | 0.00000 |
| E_Oppotunity | 0.10    | 0.05       | 0.02    | 0.00000 |
| Rho      | -0.45       | 0.89       | 0.02    | 0.00000 |

Note: All the outputs of this study reveal an interesting pattern that reflects cross-border spatial autocorrelation of events (also referred to as spillover effect). For instance, we notice two spatial regimes in COVID-19 mortality: first, a situation whereby mortality epi-centers emerge within some states and spread towards the boundaries. Second, a situation whereby the epi-centers emerge within the border counties and spread into the core counties of the neighboring states. This observation underscores the comparative efficacy of local regression models over their global counterparts (Sun et al., 2020).
incidence in the United States during the study period and up till date (as at 31st of December 2021), as indicated by USA Facts (2022) data on COVID-19 pandemic. However, the general expectation is that the best way to reduce COVID-19 mortality is to reduce its incidence rate. But thus far, human experience through the ongoing pandemic suggests that it might be better to explore the options of developing resistance against COVID-19 rather than fighting its spread.

Correlation analysis results show that poor persons are susceptible to disease conditions while the educated persons are not likely to be poor in the United States. Our results link household income to education status while the percentage of uninsured persons higher among the black race and the minority. COVID-19 mortality is linked with black race, minority, poverty, diabetes and percentage of uninsured persons. Earlier, Saffary et al. (2020) and Zhang et al. (2020) have reported that COVID-19 incidence and mortality exhibit significant relationship with socioeconomic factors across the counties in the United States. We observe that education is the key to improve socio-economic and health conditions. Therefore, we suggest that more efforts should be put in

**Fig. 3a.** The influence of causal fatality rate (CFR), percentage of black, minority and native (vulnerable population), and percentage of adults that report no leisure-time physical activity (PInactive) on COVID-19 mortality variability across the contiguous USA, as determined by GWR and MGWR coefficients. *Parameter prediction strength is directly proportional to coefficient value.*
Fig. 3b. The influence of percentage of the population with access to places for physical activity (EOpportunity) and poverty on COVID-19 mortality variability; and the local adjusted $R^2$ values for counties as determined by GWR and MGWR across the contiguous USA. Parameter prediction strength is directly proportional to adjusted $R^2$ value.

Table 4
Correlation matrix of selected socio-economic and health variables over contiguous USA.

|       | Black | hispanic | minority | college | hhincome | poverty | diabetes | obesity | PCTUI | Mortality |
|-------|-------|----------|----------|---------|----------|---------|----------|---------|-------|-----------|
| black | 1.00  |          |          |         |          |         |          |         |       |           |
| hispanic | −0.12 | 1.00     |          |         |          |         |          |         |       |           |
| minority | 0.61  | 0.62     | 1.00     |         |          |         |          |         |       |           |
| college | −0.15 |          | −0.10    | 1.00    |          |         |          |         |       |           |
| hhincome | −0.20 | 0.10     |          | 0.76    | 1.00     |         |          |         |       |           |
| poverty | 0.46  | 0.10     | 0.45     | −0.51   | −0.64    | 1.00     |         |         |       |           |
| diabetes | 0.37  | −0.11    | 0.20     | −0.45   | −0.39    | 0.45    | 1.00     |         |       |           |
| obesity | 0.31  | −0.20    | 0.10     | −0.50   | −0.42    | 0.36    | 0.51     | 1.00    |       |           |
| PCTUI  | 0.17  | 0.44     | 0.47     | −0.26   | −0.30    | 0.34    | 0.18     | 0.10    | 1.00  |           |
| Mortality | 0.51  | 0.76     | 0.48     | −0.14   | −0.10    | 0.34    | 0.22     | 0.15    | 0.21  | 1.00     |

All the associations are significant at alpha = 0.01, but the underline r values are relatively high.
place to encourage the Americans to attain the highest possible status of education. Though our study did not consider the quantity and quality of health-care providers as explanatory variables, the evaluation of the influence of such variables on COVID-19 mortality would be highly revealing. In fact, Bueraus et al. (2020) and Mollalo et al. (2020) had earlier highlighted the influence of quantity and quality of frontline health workers on COVID-19 incidence. In the same vein, researchers had earlier speculated that demographic characteristics could be influencing factors on COVID-19 pandemic (Bayne et al., 2020; Mollalo et al., 2020). Our results show that the percentage of people above of 64 years did not have significant influence on COVID-19 mortality in the contiguous USA.

Though Mollalo et al. (2020) reported that environmental factors do not have any influence on COVID-19 pandemic in USA, our results indicate that some of the examined socio-economic and health variables are indirectly influenced by climate. For instance, we observe that the vulnerable population (comprising of black race, Hispanic, minority and the native) concentrate mostly in the Southern USA (Texas, Florida, Louisiana) where the climate is relatively warmer. In this case, we observe that population density is influenced by environmental factors, particularly climate. On the other hand, the significantly high population density in the Northeast (i.e., Massachusetts, New York, New Jersey) could be attributed to the fact that New York city is a major gateway to the United States and it is a sanctuary city that accommodate numerous illegal migrants. These high-density areas coincide with COVID-19 mortality hot spots. Thus, climate has indirect influence on COVID-19 pandemic in the USA. Moreover, Ma et al. (2020) has earlier established a direct relationship between daily COVID-19 mortality and diurnal temperature in mainland China. Elsewhere in Latin America and Caribbean region, Bolano-Ortiz et al. (2020) has also established that COVID-19 incidence exhibits a significant association with mean and minimum temperature values as well as air quality, while COVID-19 mortality is found to have significant association with humidity, rainfall and wind speed.

The major challenge of this study has to do with the limitation inherent in the details of the available data. Mollalo et al. (2020) had earlier emphasized the setback associated with the coarse spatial granularity of the available COVID-19 data. Furthermore, we observed that more influencing factors could be identified, if the data were directly linked to the infected persons. The availability of such data would pave the way for the objective evaluation of the influence of some variables (such as underlying health conditions: obesity, diabetes and upper respiratory tract infections) on COVID-19 mortality. Gupta et al. (2020), Zheng et al. (2020) and Mollalo et al. (2020) had earlier identified the above underlying health conditions as potential factors that have the capacity to aggravate COVID-19 mortality. Perhaps the most influencing but the most difficult to capture is the influence of behavioral factors on COVID-19 mortality. For instance, mortality could be influenced by the willingness of the people to comply with rules and regulations regarding COVID-19 pandemic. Also, addiction to certain behaviors or lifestyles could expose some individuals or group of people to infections. Also, Mollalo et al. (2020) highlighted the possible influence of the dichotomy in enforcing COVID-19 guidelines among the states. Therefore, there are still more to learn about the factors influencing COVID-19 pandemic.

The modified/added sentences are colored red and highlighted in yellow.

The major challenge of this study has to do with the limitation inherent in the details of the available data. Mollalo et al. (2020) had earlier emphasized the setback associated with the coarse spatial granularity of the available COVID-19 data. Furthermore, we observed that more influencing factors could be identified, if the data were directly linked to the infected persons. The availability of such data would pave the way for the objective evaluation of the influence of some variables (such as underlying health conditions: obesity, diabetes and upper respiratory tract infections) on COVID-19 mortality. Gupta et al. (2020), Zheng et al. (2020) and Mollalo et al. (2020) had earlier identified the above underlying health conditions as potential factors that have the capacity to aggravate COVID-19 mortality.

Perhaps the most influencing but the most difficult to capture is the influence of behavioral factors on COVID-19 mortality. For instance, mortality could be influenced by the willingness of the people to comply with rules and regulations regarding COVID-19 pandemic. Also, addiction to certain behaviors or lifestyles could expose some individuals or group of people to infections. Also, Mollalo et al. (2020) highlighted the possible influence of the dichotomy in enforcing COVID-19 guidelines among the states. Therefore, there are still more to learn about the factors influencing COVID-19 pandemic.

5. Conclusions

In this study, we employ both global and local regression models to evaluate the relationship between COVID-19 mortality and a set of socio-economic and health variables. Our main intention is to evaluate the cumulative influence of socio-economic and health conditions on COVID-19 mortality in contiguous USA. Besides, we also evaluated the comparative efficacy of the local regression models in the spatial modelling of causal relationships between a dependent variable and explanatory variables. Our results reflect the fatality of COVID-19 across the contiguous USA. The modeled spatial pattern of COVID-19 mortality showed that the survival chance of COVID-19 infected individuals partly depends on their socio-economic and health conditions. The results of the regression analyses further confirmed the comparative efficiency of local regression models over their global counterparts. Nevertheless, all the adopted models performed optimally in this study. The overall results showed that SEM had the best performance among the global models while MGWR had the best statistics and the highest goodness-of-fit among all the adopted regression models. However, some disparities are evident in the spatial patterns of the outputs of other models and that of MGWR. Thus, the comparative assessment of the efficacy of the adopted models would better be done if all the regression model algorithms were implemented on the same platform. We conclude that investing in the living standard of the people is a formidable way to reduce the susceptibility of a population to killer infectious diseases such as COVID-19.

Author’s statement

The authors are pleased with the comments and suggestions of the reviewers. Therefore, we have carefully revised the manuscript accordingly.

- We have resolved all issues raised by the reviewers.
- The paper has been formatted in line with the guidelines of the Journal.
- The modified-added sentences are colored red and highlighted in yellow.
- The revision does not alter the focus of the study in any way.

We appreciate the efforts of the editors and reviewers.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

The authors are grateful to Emory University, Robert Wood Johnson Foundation and University of Wisconsin for making COVID-19 and socio-economic and health data freely available.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeog.2022.102671.

References

Abdelrahman, E., El Emr, F., El Komaity, M., & Swart, J. H. (2013). The effect of a high density area on the incidence of COVID-19 in South Africa. PLoS One, 8(8), e72437. https://doi.org/10.1371/journal.pone.0072437
Abdolahri, E., Champredon, D., Langley, J. M., Galvani, A. P., & Moghadas, S. M. (2020). Temporal estimates of case-fatality rate for COVID-19 outbreaks in Canada and the United States. Canadian Medical Association Journal, 192(25), E666–E670. https://doi.org/10.1503/cmaj.200711
Adekunle, I. A., Onanuga, A. T., Akinola, O. O., & Ogungbade, O. W. (2015). Modelling spatial variations of coronavirus disease (COVID-19) in Africa. The Science of the Total Environment, 542, 388–398. https://doi.org/10.1016/j.scitotenv.2015.10.066
Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., & Xia, L. (2020). Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. Radiology, 296(2), E2–E40. https://doi.org/10.1148/radiol.2020200642
Anbrug, R. W. (1996). Population, health and environmental health environment. Disease and Health Care Planning, 1(1), 1–7.
Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section. Papers in Regional Science, 92(1), 3–17.
Bayne, E., Norris, C., & Edward, T. E. (2020). A primer on emergency occupational licensing reforms for combating COVID-19. SSRN Electron. J. https://doi.org/10.2139/ssrn.3502340
Ahmed, F., Ahmed, N., Pissarides, C., & Stiglitz, J (2020). Why inequality could spread COVID-19. Lancet Public Health, 5(5). doi:10.1016/S2468-2667(20)30085-2.

Desjardin, M.R., Hohl, A., & Delmelle, E.M (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. Journal of Applied Geography, 118(102202). http://doi.org/10.1016/j.apgeog.2020.102202.

O’brien, R.M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality and Quantity, 41(5), 673–690. doi:https://doi.org/10.1007/s11135-006-9018-6.