Surveillance of COVID-19 Using Geospatial Data: An Emergency Department Perspective

S. Ashok   Malik Zaka Ullah   Nandakumar Vadivelu   Mohammed Tariqul Islam
Safa Nasereddin   Wajahat Zafar Khan

Department of Emergency Medicine, Rashid Hospital, Dubai Health Authority, Dubai, UAE

Abstract

Background: The outbreak of coronavirus 2019 (COVID-19) which emerged in December 2019 spread rapidly and created a public health emergency. Geospatial records of case data are needed in real time to monitor and anticipate the spread of infection. ***Methods:*** This study aimed to identify the emerging hotspots of COVID-19 using a geographic information system (GIS)-based approach. Data of laboratory-confirmed COVID-19 patients from March 15 to June 12, 2020, who visited the emergency department of a tertiary specialized academic hospital in Dubai were evaluated using ArcGIS Pro 2.5. Spatiotemporal analysis, including optimized hotspot analysis, was performed at the community level. ***Results:*** The cases were spatially concentrated mostly over the inner city of Dubai. Moreover, the optimized hotspot analysis showed statistically significant hotspots ($p < 0.01$) in the north of Dubai. Waxing and waning hotspots were also observed in the southern and central regions of Dubai. Finally, there were nonsustaining hotspots in communities with a very low population density. ***Conclusion:*** This study identified hotspots of COVID-19 using geospatial analysis. It is simple and can be easily reproduced to identify disease outbreaks. In the future, more attention is needed in creating a wider geodatabase and identifying hotspots with more intense transmission intensity.

© 2021 The Author(s)
Published by S. Karger AG, Basel

Introduction

The surveillance of diseases using maps has evolved over time. The use of dot maps by Dr. John Snow in the mid-1800s, which added up to the evidence that cholera is a waterborne disease, is one of the prominent historical events of using maps in health [1]. The spatial analysis of health and disease and their relationship to the geographic environment can provide valuable epidemiological insights [2]. The efficient and prompt collection of community-oriented data by emergency physicians can pave the way for community-responsive medicine and bridge the gap between emergency medicine and public health [3]. Malaria is one of the most common infectious diseases, for which many studies have been conducted through spatial analysis [4]. Moreover, spatiotemporal clustering

Correspondence to:
S. Ashok, aksankar@dha.gov.ae

analysis has demonstrated the diversity and transmission dynamics of H1N1 during the epidemic [5]. The COVID-19 disease which originated in Wuhan (Hubei, China) in December 2019 has spread rapidly to 206 other countries as of September 21, 2021. The availability of real-time strong open data has led to more focus only on aggregated case counts of COVID-19 per geographic location through GIS [6]. The real challenge is obtaining the most accurate data and performing spatiotemporal analysis at finer spatial scales and analyzing them in real time.

Methods

Objective
This retrospective study was carried out to identify hotspots and their progression over a period on the intraurban scale.

Study Settings and Duration
The study period was from March 15, 2020, to June 12, 2020. The study area, the Emirate of Dubai, is part of the United Arab Emirates and has an area of 4,114 km^2.

Study Population
The research covered patients with confirmed COVID-19 according to local national guidelines on visits to the emergency department of the Rashid Hospital, a tertiary specialized academic hospital. The data were obtained through the hospital's health information system. The outcome variable consisted of the location of a patient during a visit to the hospital and was collected from the Hasana Public Health Surveillance Network. The dataset was converted into geospatial data by obtaining the latitude and longitude coordinates of the addresses or the point of interest of patients through the Google maps of Dubai.

Diagnostic Criteria of COVID-19
The diagnostic criteria were according to the definition of national guidelines for Clinical Management and Treatment of COVID-19 of the period [7] and presented with the clinical scenario as follows:
H/O close contact with positive cases – asymptomatic.
H/O close contact with positive cases – symptomatic.
Meeting case definition and stable.
Meeting case definition with moderate and severe symptoms.

Study Participant Selection
Inclusion Criteria
Patients with residential address or point of interest and residing only in Dubai were included.

Exclusion Criteria
Patients with unfilled addresses and those residing in emirates other than Dubai were excluded from the study.

Data Collection and Visualization
The base map of Dubai with its communities which was classified as an administrative division was obtained as a shapefile from the GIS Department Dubai Municipality. The population of Dubai according to its communities was obtained as an Excel file from the Dubai Statistical Center. The GIS software ArcGIS Pro 2.5 developed by the Environmental Systems Research Institute, Redlands, USA, was used for data analysis. A file geodatabase was created in ArcGIS, and the dataset that included the shapefile of Dubai with its communities was added.

The initial visualization of data was done by comparing population density with the number of patients visited among different communities. In all communities, the feature layer population data were merged according to communities. In the symbology of the community layer, data-specific natural break classification was performed and labeled into 5 classes of spatial distribution in population density for easier illustration [8]. By using the geoprocessing tool "Aggregate Points," the total cases in the Dubai base map were aggregated according to each community and classified according to the number of counts. Then, both maps were visualized to determine the patients visited among different population densities.

Data Processing and Analysis
To compare our data for different periods, the 90-day study period was split into 9 selection layers with 10 days each. By using the geoprocessing tool "Select Layer By Date And Time" according to the laboratory-confirmed COVID-19-positive date, the data were split into 9 separate feature layers. Analyzing the pattern toolset is the starting point before conducting an in-depth analysis in spatial statistics. By using the geoprocessing tool "Average Nearest Neighbor" separately for the 9 selection layers, the point pattern of COVID-19 cases was analyzed to determine whether it was random, dispersed, or clustered. It calculates the nearest neighbor index based on the average distance from each feature to its neighboring features. The nearest neighbor ratio R <1 suggests a process toward clustering, and a value >1 indicates that it is dispersed [9]. The null hypothesis of analyzing the pattern toolset is complete spatial randomness. The z-scores and p values returned by the spatial analysis tools tell whether you can reject the null hypothesis that the observed spatial pattern is a consequence of a random process.

Spatial autocorrelation metrics were used to determine the spatial association of variables. It is similar to that of nonspatial statistical methods, except that it is done in the context of location [9]. The Getis-Ord Gi* is the spatial autocorrelation metric used to identify statistically significant clusters with high values (hotspots) and clusters of low values (cold spots) [10]. The "Mapping Clusters toolset" was selected using the geoprocessing pane of ArcGIS Pro. It consists of 2 tools for measuring hotspots: "Hotspot Analysis" and "Optimized Hotspot Analysis." In our study, we used the "Optimized Hotspot Analysis" because it uses the Getis-Ord Gi* in an automated manner to ensure optimal results. It is also able to handle points with no variables attached to it, which supports our data. It is automatically adjusted for spatial dependence and performs multiple testing using the false discovery rate correction method. The optimized hotspot analysis was performed separately for all 9 feature layers. For the analysis, there must be a minimum of 30 patients as data points in each feature layer, and it was one of the supporting points for splitting our data into 10 days each because we had fewer patients in the initial stage. To minimize the impact of a few communities that did not have adequate numbers to produce reliable results, including airports and water bodies, the 44 communities with populations <100 were excluded from the analysis [11]. The optimal distance band was fixed at 1,000 m after

Ashok/Zaka Ullah/Vadivelu/Islam/
Nasereddin/Zafar Khan

Table 1. Average nearest neighbor for all selection layers

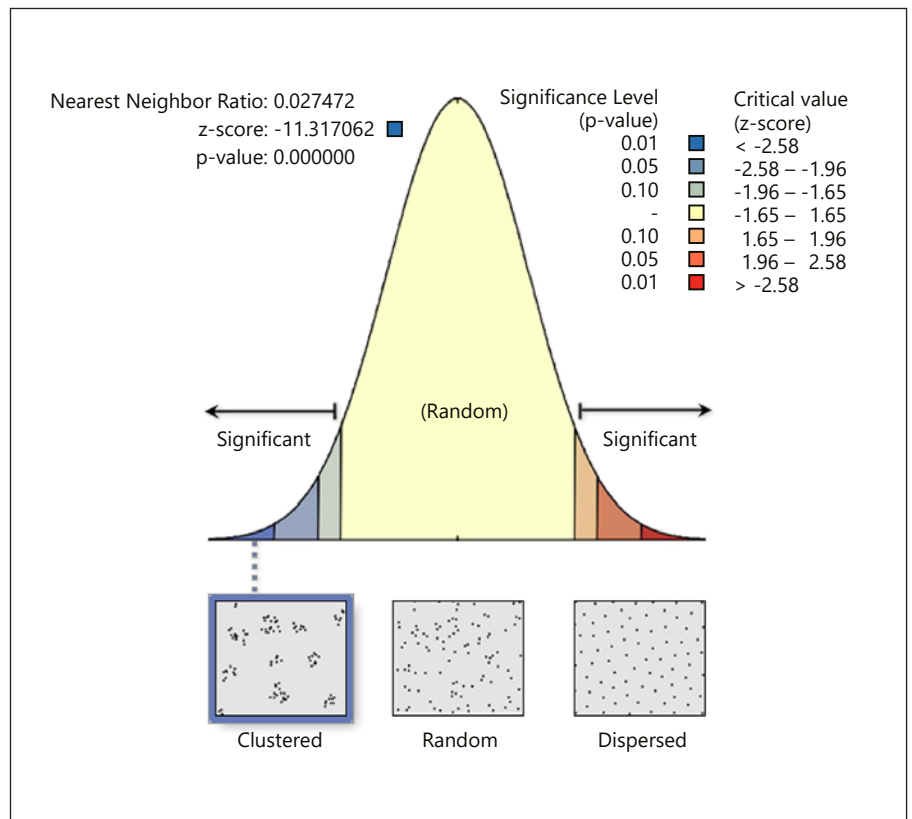| Results	One	Two	Three	Four	Five	Six	Seven	Eight	Nine
Nearest neighbor ratio	0.027	0.014	0.021	0.017	0.030	0.030	0.021	0.023	0.017
z-score	−11.317	−25.636	−24.766	−27.554	−23.388	−23.162	−22.774	−16.809	−18.604
p value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Fig. 1. Average nearest neighbor of selection layer 1. Given the z-score of −11.317062, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

performing multiple analyses using smaller and larger geographic distances [11]. In addition, there were government-imposed restrictions to work from home and closure of shopping malls and postponing major events which had led to more confinement of people in the home [12]. After performing hotspot analysis, a Gi_Bin was created in the output feature class, indicating the presence of statistically significant hotspots or cold spots with related confidence levels. All selection layer hotspots were compared to observe the changes over a defined period.

Results

Distribution of COVID-19 Cases

During the study period, there were 1,254 cases with the first COVID-19-positive date, which were successfully geocoded. Communities with a low population density had the highest percentage of patients reported to the hospital for testing (34.2%), followed by communities with moderate density (28.2%) and high density (22.4%). There were only 3 communities with a very high population density, which reported 9.2% of the patients, and the lowest percentage of patients visited were from communities with a population of very low density (5.8%), which comprised mostly the suburbs of Dubai.

Point Pattern Analysis

Average Nearest Neighbor

The ANN tool was used to compare the distribution of cases in all selection layers with different periods over the study area of Dubai to see which is more clustered than

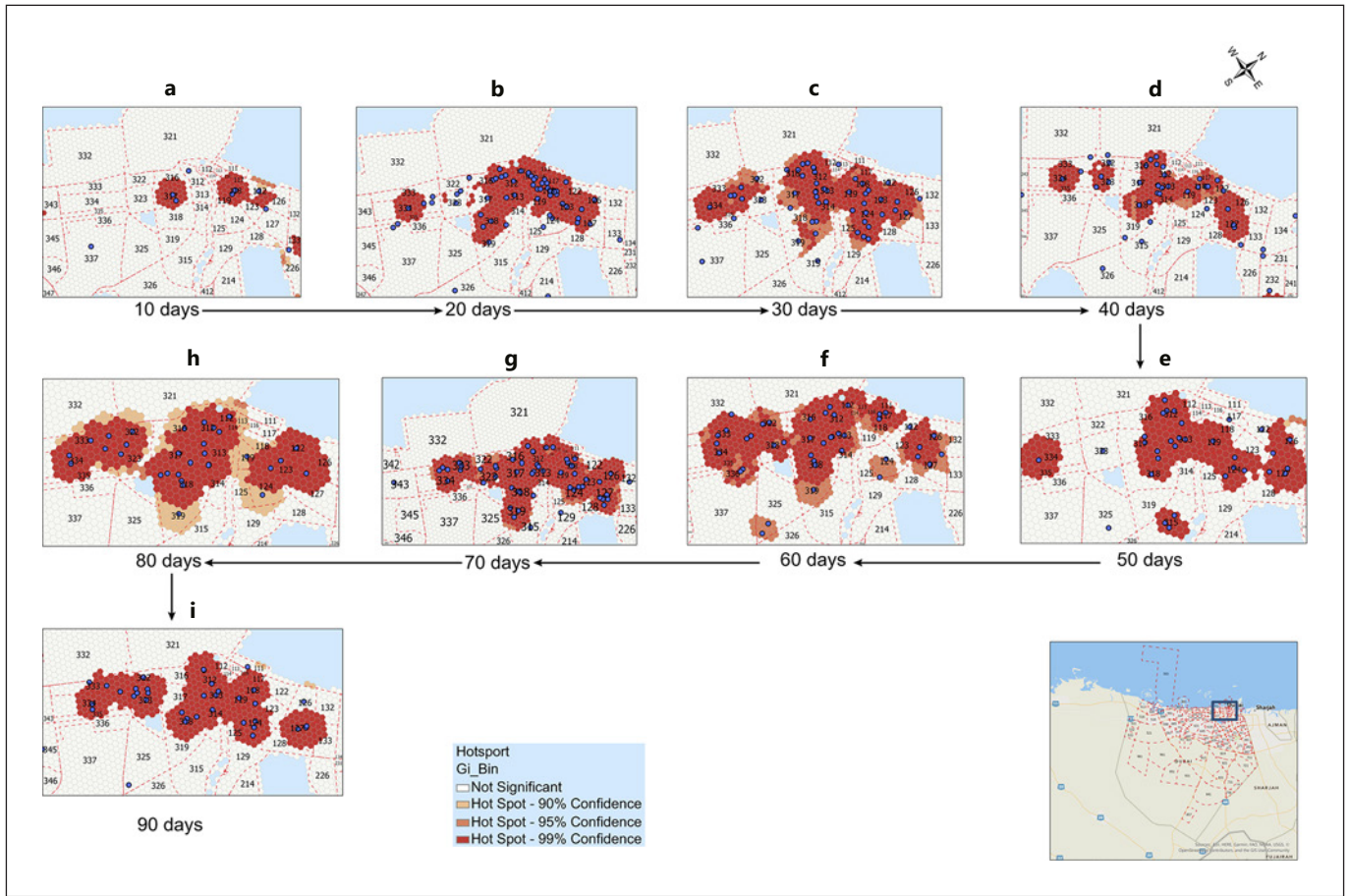**Fig. 2.** Timeline of emergence of hotspots according to selection layers.

Ashok/Zaka Ullah/Vadivelu/Islam/
Nasereddin/Zafar Khan
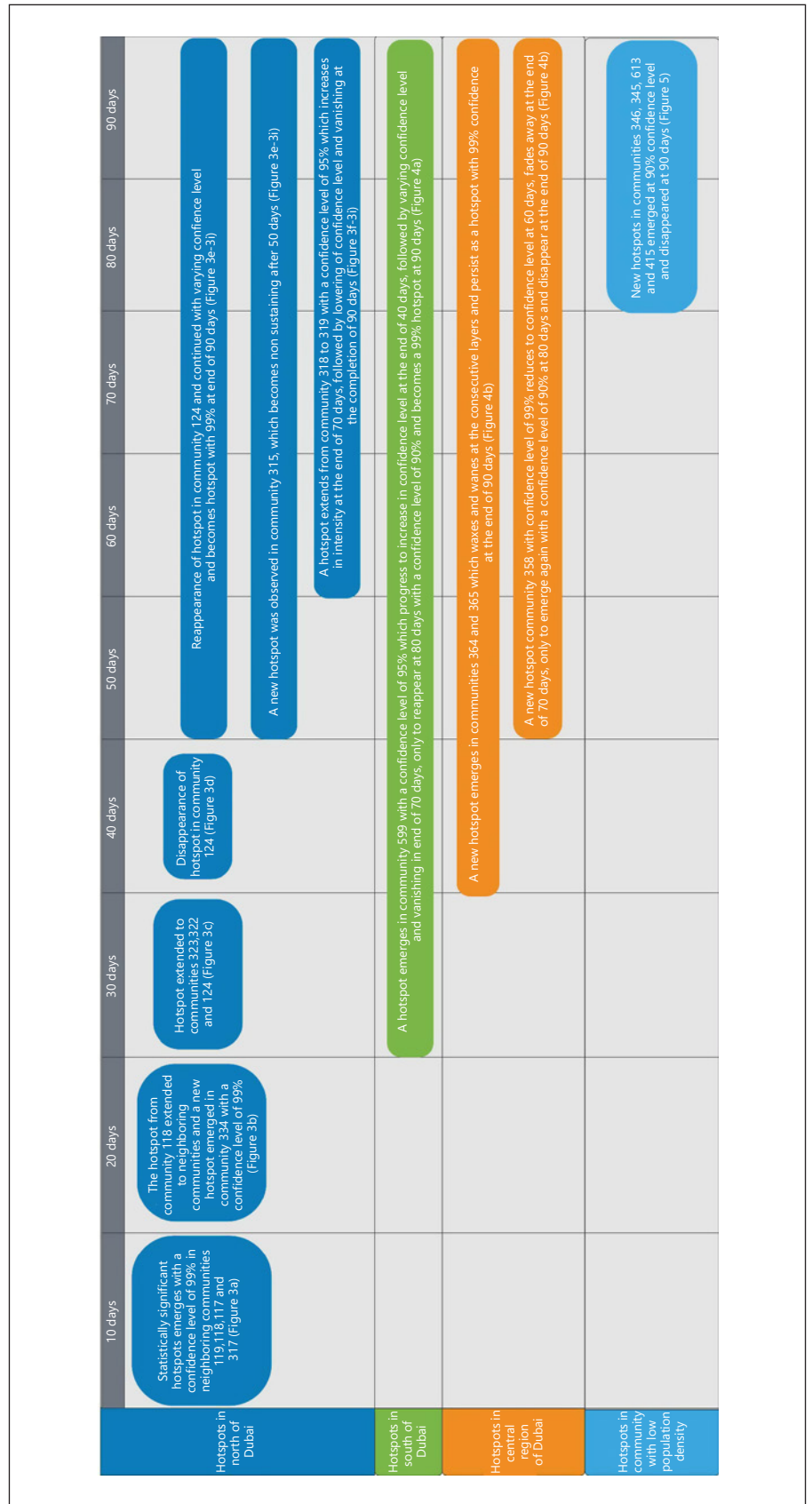
Fig. 3. Hotspots in north Dubai communities.

the other. The z-score of all selection layers was < -2.58, and the p value was <0.0001, as shown in Table 1. The nearest neighbor ratio and z-score of selection layer 4 showed that it was more clustered than the others. Selection layer 1, which includes the data of the first 10 days, showed the least clustering compared to the other layers (Fig. 1).

Optimized Hotspot Analysis

The hotspot analysis showed statistically significant clusters across the timeline (shown in Fig. 2–5). There were no cold spots found in our analysis.

Discussion

Tobler [13] invoked in his first law of geographics that everything is related to everything else, but things that are near are more related than those that are distant. In terms

of health, there is always a strong relationship between place and infectious disease transmission [14]. In the current study, we analyzed the data from COVID-19 patients who presented to the emergency department using geospatial data for rapid identification of hotspots. So far, similar research has been conducted in infectious diseases using different spatiotemporal techniques [5, 15, 16]. The population of Dubai is around 3.355 million people spread across 226 communities. The classification of population density and associated aggregate case counts of each community is shown in Figure 6. The total number of CO-VID-19 cases reported was mostly concentrated in 116 communities, which were mainly in the center of Dubai. The patient visits varied across communities with different population density, and the least patients visited were from communities with very low population density which was mostly on the outskirts of Dubai. A study on contact rate scaling with population density among infectious diseases showed an initial sublinear density-dependent pattern and
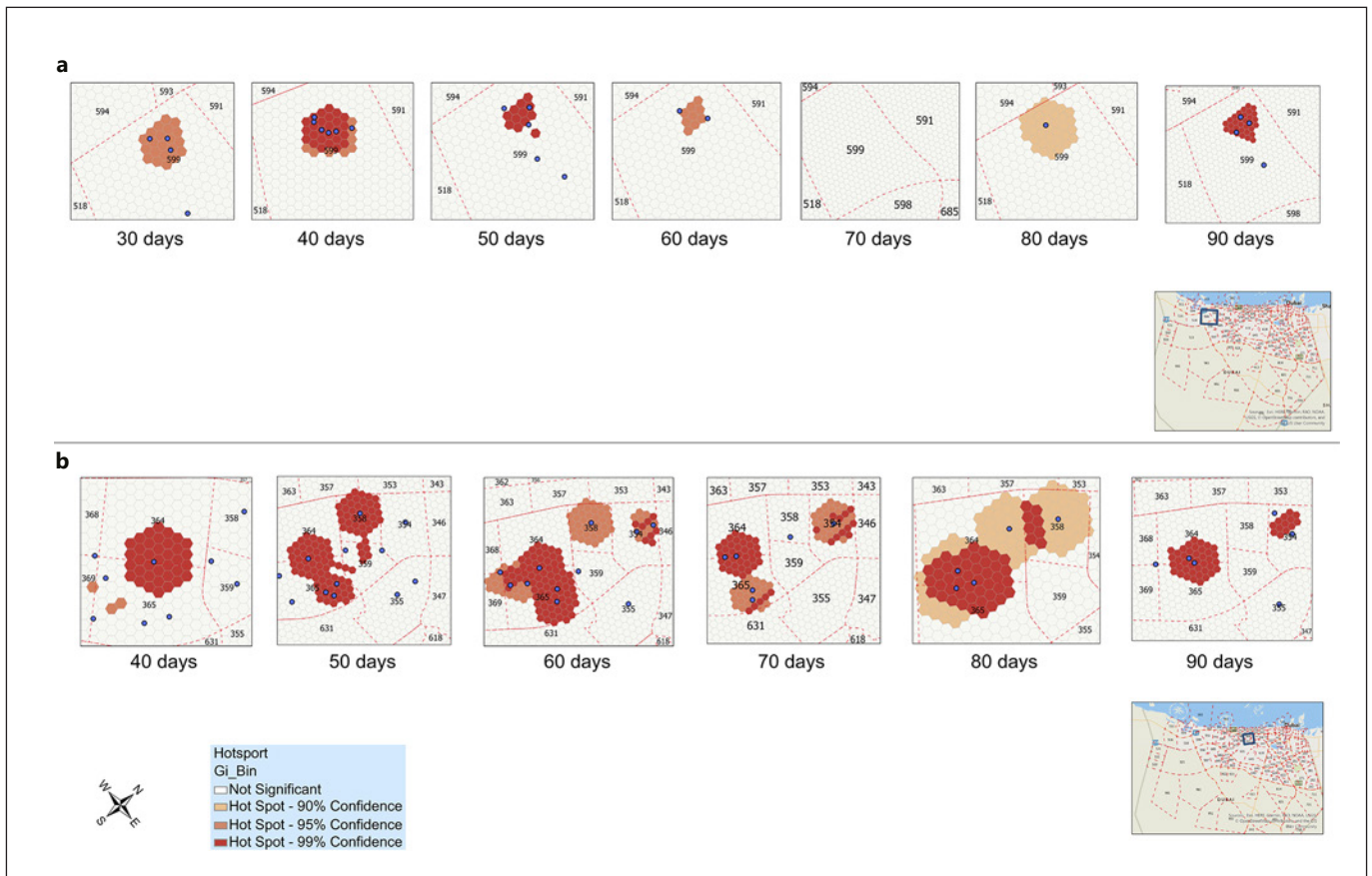
Fig. 4. Hotspots in communities in south and central regions of Dubai.

transitioning to a frequency-dependent pattern independent of population density [17]. In the present COVID-19 pandemic, there is conflicting evidence to the association of population density and disease transmission [18–20]. Analysis of the spread pattern of the cases showed all selection layers where statistically significant clusters were formed, which became more intense at the end of 40 days. Given the *z*-score of < −2.58 and a *p* value of <0.0001 in all selection layers, there is <1% likelihood for the observed clustered pattern to be due to random chance. The null hypothesis of complete spatial randomness is rejected. As the study area is fixed for all selection layers and the size of the study area is large, its influence on the analysis must be kept in mind. However, the nearest neighbor analysis remains to be the most useful approach for tracing the clustering in disease outbreaks [9].

A novel contribution of our research is the spatiotemporal analysis of COVID-19 at the community level in urban areas, which becomes a critical aspect in understanding the spatial distribution and cluster areas of infectious diseases [21]. Higgs and colleagues [22] proved that using exact geographic coordinates as the spatial base instead of census tract centroids of TB patients was able to identify localized clusters of smaller sizes earlier in time which adds value to our study, wherein we observed the emergence of hotspots using the precise location of patients. In May 2020, Rossman et al. [23] reported through population survey and integrating data with Israeli MOH that a higher prevalence of symptomatic patients in the same neighborhood of confirmed COVID-19 patients demonstrated the potential of detecting disease clusters at higher geographical resolution. In our analysis, we were able to observe the initial hotspots originating in closely connected communities (e.g., communities 119, 118, and 117) with a mixed-density population in the north of Dubai. The intensity of hotspots persisted and was confined to these historic neighborhoods, suggesting that crowded spaces play a more important role in the spread of CO-VID-19 [24]. The confinement of hotspots in this region may be the result of strict lockdown initiated in the dis-
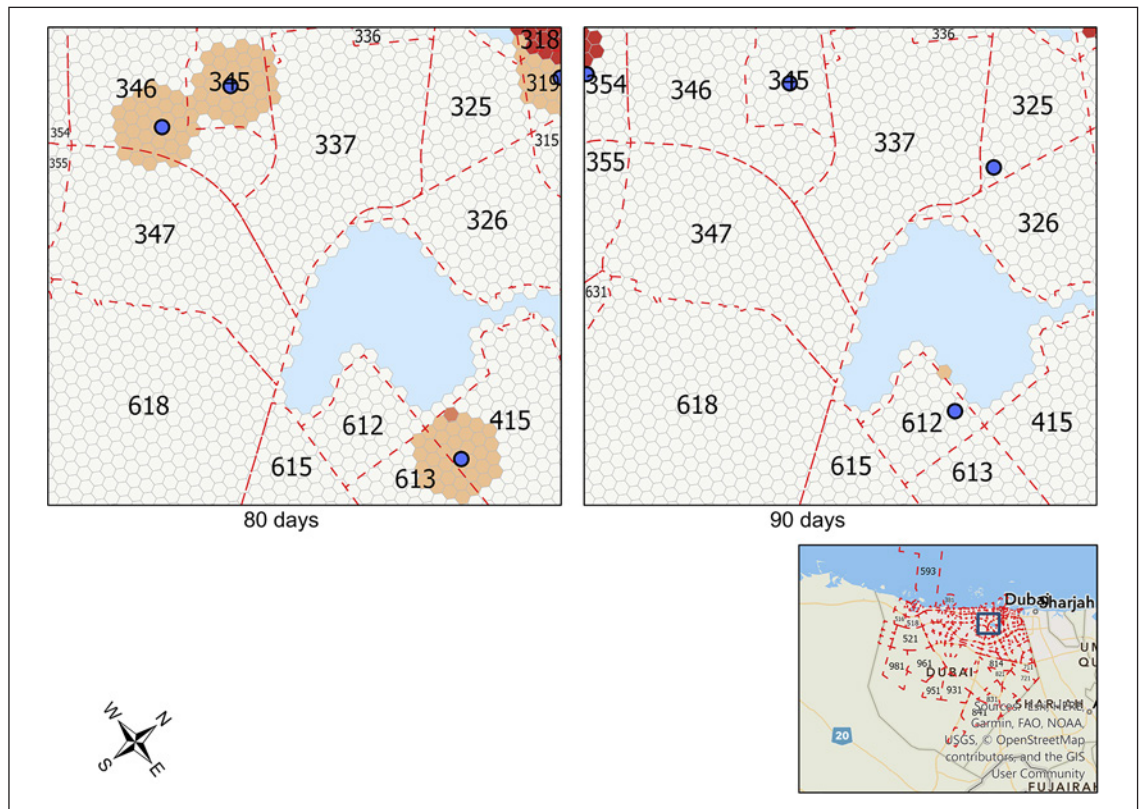
Ashok/Zaka Ullah/Vadivelu/Islam/
Nasereddin/Zafar Khan

Fig. 5. Hotspots in communities with very low population density.

tricts to the north of Dubai, which started on March 31, 2020. The hotspot analysis showed that the hotspots are mostly concentrated in densely populated communities (e.g., communities 599, 365, 117, and 118). In the current COVID-19 pandemic, recent research comparing basic reproductive number (Ro) with population density suggested that there are increased interactions between susceptible and infectious individuals in densely populated communities [25]. The temporal trends of emerging clusters in communities 364 and 358 by the end of 40 days and 50 days followed by being active for the next 40 days with varying confidence levels can help us understand the spatial patterns of emerging infectious diseases [26]. Desjardins and colleagues [27] utilized prospective space-time scan statistics using case data provided by John Hopkins University in detecting active and emergent clusters of COVID-19 at the county level in the USA. In our study, we were able to observe nonsustaining hotspots in communities 346, 345, 613, and 415, which are communities with very low population density. A similar study by Malvin et al. [28] demonstrated the identification of urban-rural clusters of malaria hotspots and compared it with a population density map using the Optimized Hotspot Analysis through ArcGIS. To the best of our knowledge, this is one of the few studies in which we were able to identify acute illness clusters of COVID-19 at the finer spatial scale using data from health care delivery systems. Technological advancements have made map-based dashboards that move COVID-19 information faster, but the real benefit will be through the collection of data at a fine spatial resolution [29]. The challenges lie ahead in the integration of heterogeneous data and creating unified storage systems that can lead to the rapid construction of GIS during emergencies [30]. The creation of a technology-driven spatial decision support system through the emergency department can help us move faster than the disease and toward early containment.

There were some limitations to this study. The data obtained are from a single center, and this may not provide a complete picture. It is also arguable whether the results can be generalized for all COVID-19 cases in Dubai. However, it is also likely that our total number of patients was representative of the total COVID-19 patients during the same study period. Moreover, we selected patients from a
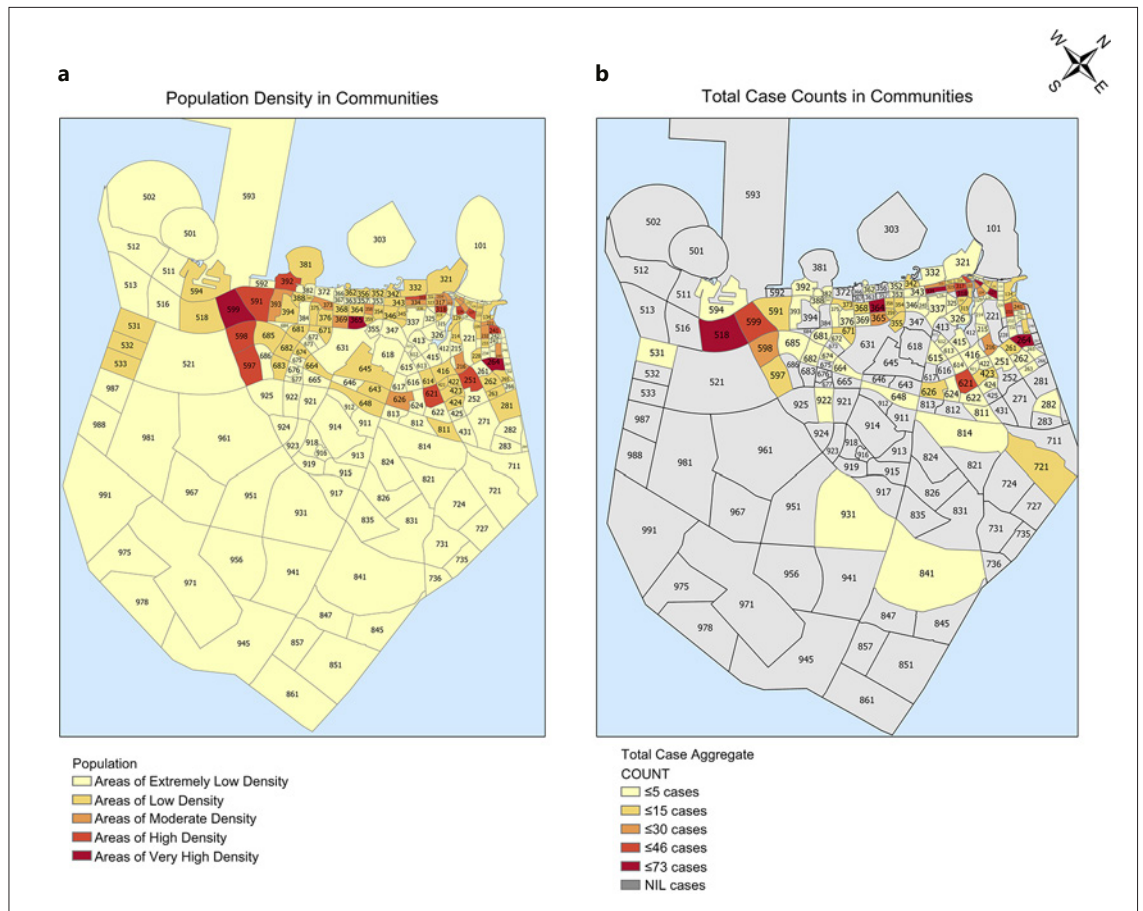
Fig. 6. Population with case count aggregates according to community.

single center because their data integrity was better. And, there were few datasets with relative location geocodes such as point of interest which could be avoided in the future by having absolute geocodes with recommended address submission format. Finally, our patient data points did not include other variables such as the severity of illness, socioeconomic factors, and environmental factors attached to it, which could better explain the spatial variability of disease incidence.

Conclusion

To our knowledge, this study was the first in Dubai to provide information regarding the geospatial analysis of COVID-19 cases. This method is simple and can be easily reproduced to identify disease outbreaks using big data. More refined absolute location data using advanced geo-addressing systems like Makani, a Dubai municipality application, can enrich the detection of potential disease threats. Future work could focus on creating a wider geodatabase by classifying disease severity and identifying hotspots with more intense transmission intensity.

Statement of Ethics

The research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki and was approved by the Dubai Scientific Research Ethics Committee, Dubai

Ashok/Zaka Ullah/Vadivelu/Islam/
Nasereddin/Zafar Khan

Health Authority (Reference DSREC-06/2020_14, 15 June 2020). The patients' deidentified data used in this research were carried out retrospectively. The need to obtain a written consent rather than the general patient's consent to treat and use their data for research was waived by the Dubai Scientific Research Ethics Committee.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

The authors did not receive any funding.

Author Contributions

A.K.S., M.Z.U., N.V., M.T.I., S.N., and W.Z.K. contributed to conceptualization; A.K.S., M.Z.U., N.V., M.T.I., S.N., and W.Z.K. contributed to data structuring; A.K.S. contributed to software; A.K.S., M.Z.U., and N.V. contributed to formal analysis; M.T.I., S.N., and W.Z.K. contributed to supervision; A.K.S. and M.Z.U. contributed to visualization; A.K.S., M.Z.U., and N.V. contributed to writing – original draft; A.K.S., M.Z.U., N.V., M.T.I., S.N., and W.Z.K. contributed to writing – review and editing.

Data Availability Statement

The data that support the findings of this study are not publicly available as they contain information that could compromise the privacy of research participants but are available upon reasonable request from the corresponding author who can be reached at aksankar@dha.gov.ae.

References

1 Snow J. On the mode of communication of cholera. 1855. Salud Publica Mex. 1991;33(2):194–201.

2 Mullner RM, Chung K, Croke KG, Mensah EK. Geographic information systems in public health and medicine. J Med Syst. 2004;28(3):215–21.

3 Garrison HG, Runyan CW, Tintinalli JE, Barber CW, Bordley WC, Hargarten SW, et al. Emergency department surveillance: an examination of issues and a proposal for a national strategy. Ann Emerg Med. 1994;24(5):849–56.

4 Lyseen AK, Nøhr C, Sørensen EM, Gudes O, Geraghty EM, Shaw NT, et al. A review and framework for categorizing current research and development in health related geographical information systems (GIS) studies. Yearb Med Inform. 2014;9(1):110–24.

5 Lee SS, Wong NS. The clustering and transmission dynamics of pandemic influenza a (H1N1) 2009 cases in Hong Kong. J Infect. 2011;63(4):274–80.

6 Xu B, Gutierrez B, Mekaru S, Sewalk K, Goodwin L, Loskill A, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. Sci Data. 2020;7(1):106.

7 Available from: https://services.dha.gov.ae/sheryan/wps/portal/home/search-circulars.

8 Campbell J, Shin M. Essentials of geographic information systems. Saylor Foundation; 2011.

9 Grekousis G. Spatial analysis methods and practice: describe–explore–explain through GIS. Cambridge University Press; 2020.

10 Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Perspect Spatial Data Analysis. 1992;24(3):189–206.

11 Lee DC, Jiang Q, Tabaei BP, Elbel B, Koziatek CA, Konty KJ, et al. Using indirect measures to identify geographic hotspots of poor glycemic control: cross-sectional comparisons with an A1C registry. Diabetes Care. 2018;41(7):1438–47.

12 Available from: https://u.ae/en/information-and-services/justice-safety-and-the-law/handling-the-covid-19-outbreak/government-efforts-to-contain-covid-19.

13 Tobler WR. A computer movie simulating urban growth in the Detroit Region. Econ Geogr. 1970;46:234–40.

14 Mennis J, Yoo EE. Geographic information science and the analysis of place and health. Trans GIS. 2018;22(3):842–54.

15 Al-Kindi KM, Alkharusi A, Alshukaili D, Al Nasiri N, Al-Awadhi T, Charabi Y, et al. Spatiotemporal assessment of COVID-19 spread over Oman using GIS techniques. Earth Syst Environ. 2020;4:797–811.

16 Tewara MA, Mbah-Fongkimeh PN, Dayimu A, Kang F, Xue F. Small-area spatial statistical analysis of malaria clusters and hotspots in Cameroon; 2000-2015. BMC Infect Dis. 2018;18(1):636.

17 Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. Math Biosci. 2013;244(2):125–34.

18 Carozzi F. Urban density and COVID-19. Bonn, Germany: IZA: Institute of Labor Economics; 2020. http: //www.econis.eu/PPNSET?PPN=1724848917. Accessed 2021 Jan 10.

19 Hamidi S, Ewing R, Sabouri S. Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: early evidence from 1,165 metropolitan counties in the United States. Health Place. 2020;64:102378.

20 Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. J Travel Med. 2020;27(3):taaa038.

21 Zhu B, Fu Y, Liu J, Mao Y. Spatial distribution of 12 class B notifiable infectious diseases in China: a retrospective study. PLos One. 2018;13(4):e0195568.

22 Higgs BW, Mohtashemi M, Grinsdale J, Kawamura LM. Early detection of tuberculosis outbreaks among the San Francisco Homeless: trade-offs between spatial resolution and temporal scale. PLoS One. 2007;2(12):e1284.

23 Rossman H, Keshet A, Shilo S, Gavrieli A, Bauman T, Cohen O, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. Nat Med. 2020;26(5):634–8.

24 Almagro M, Orane-Hutchinsonb A. The determinants of the differential exposure to COVID-19 in New York City and their evolution over time. J Urban Econ. 2020:103293.

25 Sy KTL, White LF, Nichols BE. Population density and basic reproductive number of COVID-19 across United States counties. PLoS One. 2021;16(4):e0249271.

26 Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990–3.

27 Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl Geogr. 2020;118:102202.

28 Tewara MA, Mbah-Fongkimeh PN, Dayimu A, Kang F, Xue F. Small-area spatial statistical analysis of malaria clusters and hotspots in Cameroon;2000-2015. BMC Infect Dis. 2018;18(1):636.

29 Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr. 2020;19(1):8.

30 Zhou C, Su F, Pei T, Zhang A, Du Y, Luo B, et al. COVID-19: challenges to GIS with big data. Geogr Sustain. 2020;1(1):77–87.