A conceptional model integrating geographic information systems (GIS) and social media data for disease exposure assessment

Jerry Enoe,[1] Michael Sutherland,[1] Dexter Davis,[1] Bheshem Ramlal,[1] Charisse Griffith-Charles,[1] Keston H. Bhola,[2] Elsai Mati Asefa[3]

[1]Department of Geomatics Engineering and Land Management, The University of the West Indies, St. Augustine, Trinidad and Tobago; [2]Department of Computers and Technology, School of Arts and Science, St George's University, Grenada; [3]School of Environmental Health, College of Health and Medical Sciences, Haramaya University, Harar, Ethiopia

Correspondence: Elsai Mati Asefa, School of Environmental Health, College of Health and Medical Sciences, Haramaya University, Harar, PO Box 235, Ethiopia.
Tel.: +251921832876.
E-mail: Elsai.mati@haramaya.edu.et

Key words: disease map; environmental exposure; GIS; individual mobility patterns; social media.

Contributions: JE, conceptualization, investigation, project administration, resources visualization; JE, KHB, data curation, software; JE, BR, formal analysis, methodology validation; MS, BR, CGC, DD, supervision; JE, EMA, roles/writing – original draft; EMA, MS, BR, CGC, writing – review and editing.

Funding: this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interests: the authors declare that they have no competing interests.

Availability of data and materials: the authors confirm that the data supporting the findings of this study are available within the article.

Acknowledgments: we are grateful to the medical professionals who provided valuable feedback during the development of our model. Their insights and expertise have greatly contributed to the refinement of our work. Furthermore, we would like to extend our gratitude to the reviewers and editors for their constructive comments and suggestions in improving the clarity and overall quality of this work.

Received: 27 December 2023.
Accepted: 28 February 2024.

Geospatial Health 2024; 19:1264
doi:10.4081/gh.2024.1264

Abstract

Although previous studies have acknowledged the potential of geographic information systems (GIS) and social media data (SMD) in assessment of exposure to various environmental risks, none has presented a simple, effective and user-friendly tool. This study introduces a conceptual model that integrates individual mobility patterns extracted from social media, with the geographic footprints of infectious diseases and other environmental agents utilizing GIS. The efficacy of the model was independently evaluated for selected case studies involving lead in the ground; particulate matter in the air; and an infectious, viral disease (COVID-19). A graphical user interface (GUI) was developed as the final output of this study. Overall, the evaluation of the model demonstrated feasibility in successfully extracting individual mobility patterns, identifying potential exposure sites and quantifying the frequency and magnitude of exposure. Importantly, the novelty of the developed model lies not merely in its efficiency in integrating GIS and SMD for exposure assessment, but also in considering the practical requirements of health practitioners. Although the conceptual model, developed together with its associated GUI, presents a promising and practical approach to assessment of the exposure to environmental risks discussed here, its applicability, versatility and efficacy extends beyond the case studies presented in this study.

Introduction

The increase in international and domestic travel can spread disease faster and with a wider geographic distribution than ever before (Schlagenhauf *et al.*, 2015; Otsuki & Nishiura, 2016). Recent outbreaks of diseases such as Chikungunya (Krutikov & Manson, 2016), Zika (Gardner *et al.,* 2018) and coronavirus disease 2019 (COVID-19) (Chinazzi *et al*., 2020) have been directly linked to the global transport network. Additionally, travellers are exposed to infectious diseases that are not endemic to their home countries (Angelo *et al.*, 2017) and an estimated 20-70% of people, who travel to the developing world report illness linked to their travel (Edward, Wilson, & Kain 2002). Indeed, transport geography is becoming an important issue with respect to health (Amer & Bergquist, 2021).

Understanding the influence of location on health is one of the determinants of the epidemiological research framework. The study of health geography (Dummer 2008) can facilitate a spatial understanding of disease propagation, population health, and healthcare service distribution (Photis 2016). As such, efficient diagnosis and treatment of any illness are inherently linked to the speed at which the source of the ailment is identified (Şengül *et al.,* 2013).

Technological advances in geographical information systems (GIS), remote sensing and computer science can increase efficiency of medical diagnosis (Mbunge *et al.*, 2021). In addition, innovations such as mobile geo-enabled sensors with the ability to track an individual's movement can help evaluate the impact of social influences and environmental factors (Kozel & Burnham-Marusich, 2017). Particularly, the widespread use of smartphones and other global positioning system (GPS)-enabled mobile devices has made large volumes of georeferenced data readily available to healthcare researchers (Wu *et al.*, 2014; Tompkins & McCreesh, 2016; Bisanzio *et al*., 2020).

The advent of big data analysis and GIS technologies represent a new era in public health service delivery and problem-solving. Specifically, advances in social networks, mobility, cloud computing and the Internet have made this possible (Alonso *et al.,* 2017). GIS technologies such as geo-visualization (Panteras *et al.*, 2015; Yasobant *et al.*, 2015) and space-time analysis (Briggs 2005) are becoming major tools for analysing big data generated by social media. Social media platforms, such as X (formerly Twitter), Facebook and Instagram have gained considerable significance as valuable data sources for a wide range of applications, including research in the field of public health. These platforms offer access to a diverse array of data types, encompassing real-time streaming content as well as retrievable archival data comprising text, images and videos (Fujita 2017; Liu *et al.*, 2018; Middleton *et al.*, 2018). Leveraging their immense user base, which encompasses millions of individuals across the globe, social media platforms present a unique opportunity to acquire location data shedding light on users' movements and engagements. These platforms have been utilized for purposes such as monitoring disease outbreaks (Heldman *et al.*, 2013; Fung *et al.*, 2016; Aiello *et al.*, 2019;), assessing health behaviours (Allington et al. 2021; Basch, Hillyer, & Jaime 2022) and executing public health interventions (Gunasekeran *et al.*, 2022). Moreover, using GIS tools and methodologies in medicine and public health has been recognized as a valuable approach to improving disease control, monitoring and prevention (Musa *et al.*, 2013; Yasobant *et al.*, 2015; Yousefinaghani *et al.*, 2019). Despite this recognition, the consumption of GIS use within these fields is still relatively low and the literature does not explicitly identify specific tools and methodologies to its application in the assessment of disease exposure.

Despite the considerable potential to support public health initiatives (Aiello *et al.* 2019), current applications in this field require further refinement, particularly in terms of improved integration and validation. For example, real-time mobile phone data can be used to provide mobility pattern information to pre-empt future disease outbreaks and/or retrospective disease exposure integrated with diseases' spatial mapping. Although methods for extracting mobility patterns from social media data (SMD) have been well-established, and studies have demonstrated the feasibility of using SMD to map and predict human movements (Jurdak *et al.,* 2015; Liao & Yeh 2018; Scholz & Jeznik 2020), the extensive use of social media presents a challenge for GIS researchers. On the other hand, limited attention is given to individual-level characteristics, such as mobility patterns and disease exposure, with a focus on refined spatial and temporal resolution (Jurdak *et al.*, 2015; Sinnenberg *et al.,* 2017; Xu, Dredze, & Broniatowski, 2020).

Hence, to fully harness the potential of SMD and effectively integrate it with disease mapping, it is crucial to address the existing challenge and to develop robust spatial data processing techniques. Additionally, the development of new tools and methods should allow for greater use of GIS data by non-GIS experts, such as health practitioners (Li, Wang, & Li, 2015; Yasobant *et al.*, 2015). As a result, these advancements should enable the seamless integration of diverse data sources and bolster the accuracy and timeliness of disease surveillance and response initiatives.

This study introduces a novel conceptual and practical model that integrates individual mobility patterns extracted from social media, specifically X, with various geo-enabled datasets. These datasets encompass disease outbreaks, endemic diseases and environmental pollution, thereby facilitating the identification of potential individual exposures within both spatial and temporal dimensions. The overarching objective of our study was, therefore, the development and evaluation of a conceptual model that integrates GIS and SMD for disease exposure assessment. This model aims to extend the utilization of GIS technologies and methodologies beyond GIS professionals, enabling medical doctors and public health experts to apply these tools to real-world medical challenges.

Conceptual model development

We focused on the development and evaluation of a conceptual model capable of extracting individual mobility patterns from social media and combining them with disease and environmental agents' geographic footprint using GIS to determine the spatial and temporal relationship between agent and receptor. It is crucial to ascertain the precise needs of medical practitioners involved in disease exposure assessment given that no previous study has provided tools and methodologies considering the specific needs of practitioners and policymakers. As such, a consultation was undertaken with 37 medical doctors to understand the requirements for the model and identify the relevant parameters and variables to be incorporated into the model. The overall feedback received was promising, as it suggests that the conceptual model fulfils a requirement within the medical community, who generally are not experts in advanced GIS. They also identified several diseases, agents and functionalities that the model should include and be able to manipulate, which were the foundation for the selection of case studies presented in this study.

After receiving the feedbacks from medical practitioners, the process of model development was initiated (Figure 1). Here, the proposed conceptual model used several independent datasets (exposure agents and individual mobility patterns extracted from SMD and multiple GIS processes to extract meaningful information from these datasets. This new source of information enhances the capabilities of medical and public health practitioners in identifying the possible locations where agent and receptor (individual/patient) intersect leading to exposure and eventual ailment. Figure 1 gives an illustrative overview of the proposed conceptual model, data types, and GIS processes.

The prevalence and spatiotemporal patterns of the agents of exposure are often represented and modelled through maps and GIS products. These maps and other GIS products typically depict exposure, prevalence and risk gradients that vary over a specific geographic region or globally (De Cola, 2002; Cleckner & Allen, 2014; Leta *et al.*, 2018). With the use of geo-visualisation and other GIS techniques, several diseases and exposure agents have been successfully mapped and risk models created; examples include, but not limited to, malaria (Okami & Kohtake 2016), various vectors (Cleckner & Allen 2014), COVID-19 (Kamel Boulos & Geraghty 2020) and lead (Wallace 2023). These maps and other GIS products are at the core of this conceptual model as they will serve as the base layers for all spatiotemporal queries between the agent and receptor. The role of social media is also vital in this conceptual model since it will be the source of the receptor spatiotemporal patterns. As mentioned above, the ability to utilize SMD as a proxy for human mobility patterns is well established.

Materials and Methods

X data acquisition

Only geo-tagged data from individual X (Twitter) user profiles were downloaded in this study. Other location retrieval methods, such as geoparsing (Middleton *et al.*, 2018), can be used to retrieve location information from social media. Unfortunately, these methods are subjected to more significant errors than actual geotagged tweets due to the ambiguous spelling of placenames (Ye *et al.*, 2016). Data extraction was facilitated through a Python script modified from GitHub.com (https://github.com/goldman88/django_twitter). The script development was done utilizing Spyder 4.1.4 development environment combined with access to X (Twitter)'s application programming interface (API). Figure 2 illustrates the data acquisition process.

The X (Twitter) platform was used for this study because it is one of the most popular social media (Gulnerman *et al.*, 2020; Sloan & Morgan 2015). It was also chosen because of the relative ease of gaining access to its API and the process of obtaining a developer account on the platform compared to other social media platforms (Gulnerman *et al.*, 2020). Mobility data were obtained from randomly selected public X (Twitter) user profiles of three different individuals venturing into the target area (user_1, user_2 and user_3) were selected (with names withheld) from the X (Twitter) tracking website http://onemilliontweetmap.com. Using valid X (Twitter) credentials (consumer key, consumer secret, access token and access token secret) and the python script, all tweets from the selected users were downloaded and stored in an SQLite database connected to ArcGIS Pro (ESRI, Redlands, CA, USA) The first two queries filtered tweets for the last year and tweets containing geographic coordinates (latitude and longitude); see Figure 2.

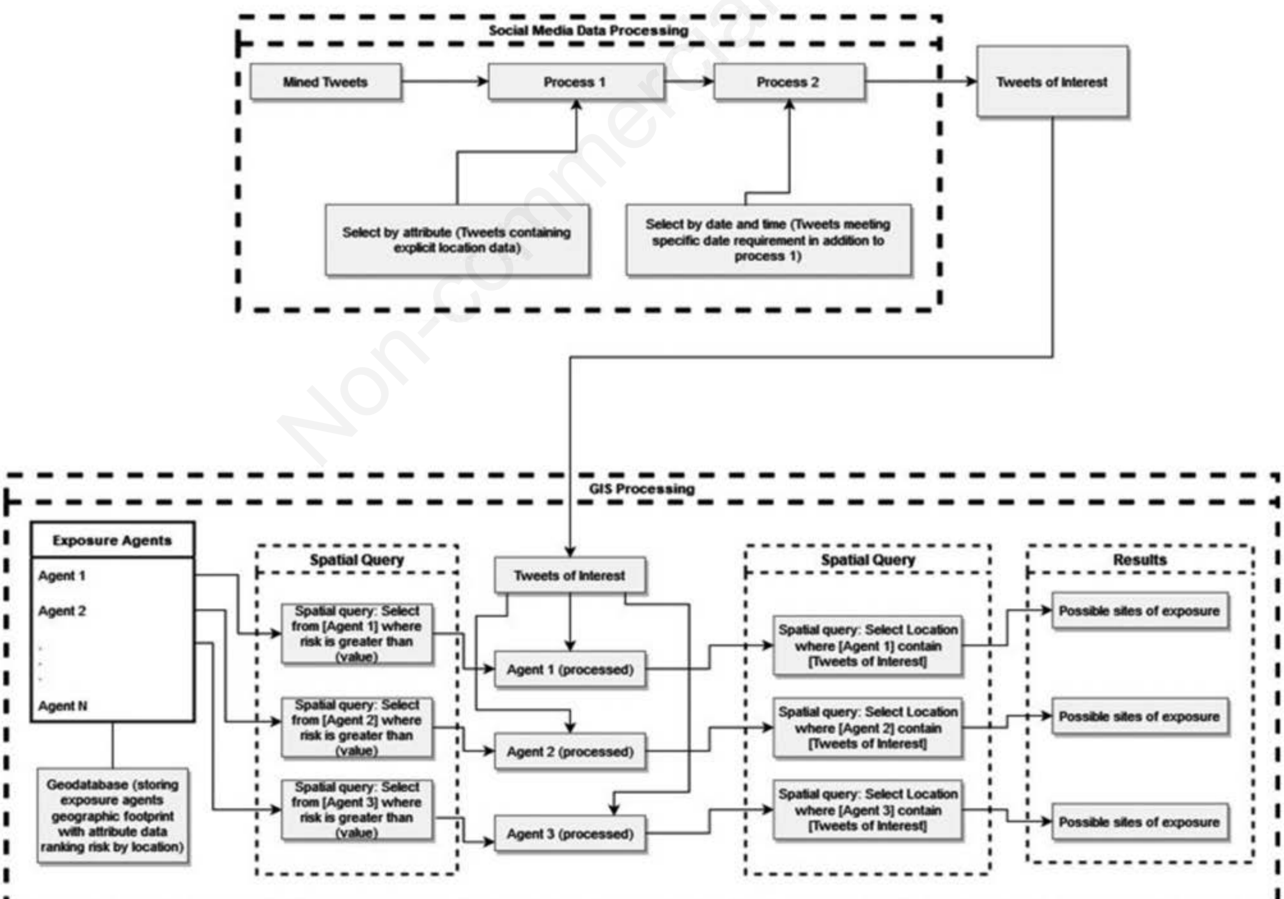

Figure 1. Conceptual design towards mining human mobility data from social media to support exposure assessment.

Exposure agents

Several environmental factors affect human health; these include, but are not limited to, toxic chemicals, air pollution [e.g., particular matter (PM) of 2.5 μm], diseases caused by vector-borne viruses or parasites and poor water quality (bacteria or toxic chemicals); these environmental factors and agents exhibit varying spatial and temporal patterns influenced by natural and anthropogenic factors, such as climate, weather, geology and built infrastructure among others (Matsuyama *et al.*, 2006; Gautam, 2020; Trajer, 2021). For this research, three agents and areas were randomly selected: lead exposure risk in the state of Iowa, COVID-19 risk in Texas, both in the United States (U.S.) and air quality in China. These agents were selected due to their significance in public health, varying modes of exposure and the availability of relevant data sources. For each of them, three specific receptors, namely User_1, User_2 and User_3, were selected to represent individuals with unique characteristics, such as particular travel behaviours, locations or environmental interactions. The geo-tagged data of these individuals was extracted from X (Twitter) and made into GIS layers to show attributes, such as dates and locations. This approach allowed for the integration of receptor-specific data into the analysis, enabling a more comprehensive assessment of the framework's efficacy in identifying exposure sites and informing about risk.

Case study 1. Lead exposure risk in Iowa (U.S.)

Numerous detrimental health effects are associated with lead exposure, including hearing deficit, impaired cognitive function, cardiovascular disease and others (Assi *et al.*, 2016). Annually 800,000 deaths in the U.S. are attributed to cardiovascular diseases; lead has been identified as one of the many recognized contributing risk factors (Lanphear *et al.*, 2018). As a result, lead exposure within the continental U.S. has received significant attention in recent years, resulting in several websites housing various lead-related data, such as risk factors, relative risk factors, populations at risk and geographic exposure variation (Janke 2014). Importantly, the Iowa census tracts were specifically chosen because of their relatively small size and the expected homogeneity of population and demographic characteristics within them.

The risk of exposure to lead was obtained from the database https://www.policymap.com. We investigated the 2010 U.S. national census tract boundaries, where each census tract is assigned a specific lead exposure rating, such as very low, low, medium and high (Figure 3).

Case study 2. Air quality in China

The impact of poor air quality on human health is associated with several health effects, including, but not limited to, premature death due to cardiovascular and pulmonary disease. PM_2 is considered a common indicator of air quality and key constituent of air pollution that can lead to illness. Table 1 highlights the health implications of the different air quality levels according to the U.S. Environmental Protection Agency (EPA) (Janke 2014). Air quality in many parts of China is regarded as serious; 31% of all cities are currently monitored for air quality are considered heavily polluted; moreover, 3/4 of all urban dwellers in China are living under air quality conditions deemed harmful (Hao & Wang 2005).

China has twenty-three provinces, five autonomous regions, four municipalities and two special administrative regions, which are further subdivided into 361 smaller areas (Figure 4). For our case study 2, air quality index (AQI) data including $PM_{2.5}$, SO_3 and NO_2 for January 2021 to December 2021 for all 361 sub-regions within China were obtained from the online database https://aqicn.org/. Each region was assigned an AQI rating based on the 12-month daily average. Though only a one-year average was used, producing an AQI for any specific period is feasible, providing that the data are available.

Case study 3. COVID-19 risk in Texas (U.S.)

Since COVID-19 emerged as a threat to global health (Fauci, Lane, & Redfield, 2020), it has been responsible for close to 7 million deaths worldwide according to the latest recorded figures (https://www.worldometers.info/coronavirus/) and severe economic hardship since March 2019. However, technological advances such as real-time mapping and predictive modelling combined with advances in medicine and traditional public health tools and techniques have played a significant role in reducing the propaga-

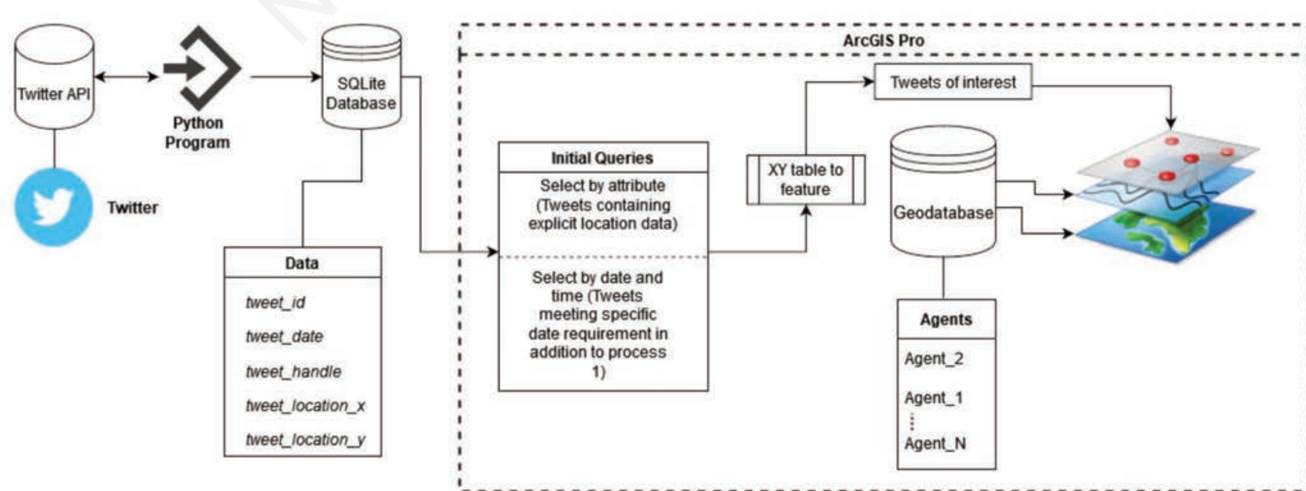

Figure 2. The X (Twitter) data acquisition process.

tion of COVID-19 (Allam & Jones, 2020; Smith *et al.*, 2020). When this article was prepared, COVID-19 cases were still rising in many countries and territories (De Guzman, 2022). Several websites and online databases provide a wide range of COVID-19-related data and interactive tools for analysis by the public and researchers. We used a daily moving average of COVID-19 risk and selected the 7-day time span of January 10-17, 2022 for inclusion in this research. The state of Texas was selected because it was one of the most affected states, ranking second in the number of confirmed cases and deaths in the U.S. Additionally, significant disparities in the number of cases and deaths among different community groups were observed. This state is divided into 254 counties and each county was assigned a daily risk level between 1 and 4, with 4 being the highest risk, as shown in Figure 5.

Agent and receptor

The final product of this conceptual model would be a software application capable of performing the GIS processes, SMD extraction and providing possible sites of exposure in a user-friendly software environment. To understand the potential benefits of this conceptual medical investigation model, a hypothetical scenario would illustrate this best. For example, when a patient presents his/her symptoms, the physician, conducts a series of tests to identify the causative agent and the exposure situation. Figure 6 depicts the graphical user interface (GUI), an application that enables physicians to easily identify potential exposure sites utilizing the patient's SMD. However, it is important to acknowledge that there can be multiple sources of exposure, and it is not feasible to predict exposures as risk factors solely based on the limited data sources incorporated in the current version of the developed tool. Until further enhancements have been made, the GUI developed so far can only monitor some specific threats, such as the agents presented in this study. However, in its current form, the tool's efficacy can be augmented by sending text messages as alert notifications to indi-

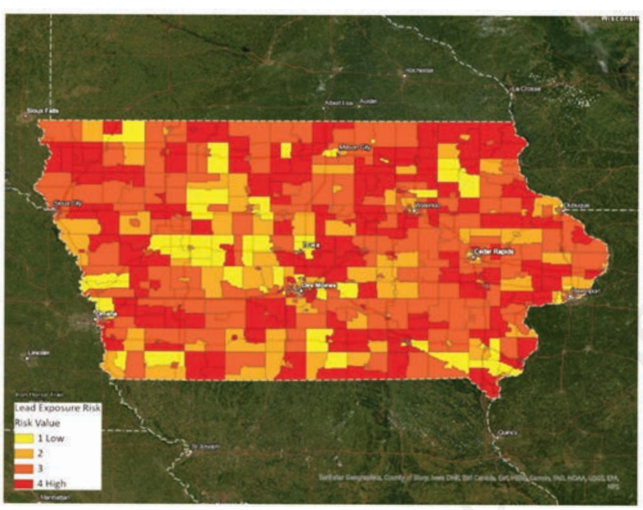

Figure 3. The lead risk in Iowa, U.S. (average records over the years 2015-2019). Source: https://www.policymap.com/

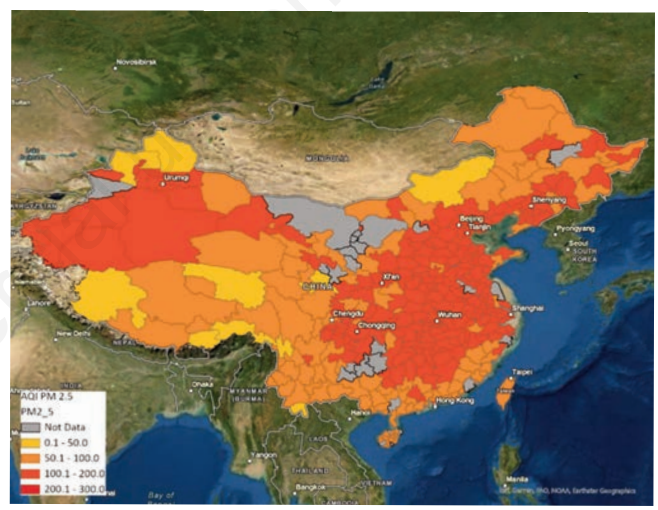

Figure 4. Daily average of air quality with respect to the presence of particulate matter (AQI-$PM_{2.5}$) in China's sub-regions in 2021(January 1-December 31). Source: https://aqicn.org/sources/.

Table 1. Standard air quality indices and association with health implications.

AQI	Level of air pollution	Health implication	Cautionary statement ($PM_{2.5}$)
0 - 50	Good	Air quality is considered satisfactory, with pollution of little or no risk	None
51 -100	Moderate	Air quality is acceptable, with moderate health concern for people who are unusually sensitive to air pollution for certain pollutants	Active children and adults, and people with respiratory diseases, such as asthma, should avoid prolonged outdoor exertion
101-150	Unhealthy for sensitive persons	General public not likely to be affected but sensitive people may experience health effects	Active children and adults, and people with respiratory diseases, such as asthma, should avoid prolonged outdoor exertion
151-200	Unhealthy	Slight health effects for everyone, with of sensitive groups may experience more serious health effects	All, especially active children and people with respiratory disease, such as asthma, should avoid prolonged outdoor exertion
201-300	Very unhealthy	Health warnings, with the entire population likely to be affected	All, especially active children and people with respiratory disease, such as asthma, should limit outdoor exertion
300+	Hazardous	Health alert: everyone may experience serious health effects	All outdoor exertion should be avoided

AQI, air quality index according to U.S. Environmental Protection Agency standard of 2016; $PM_{2.5}$, particular matter of 2.5 μm.

viduals who pass through the identified high-risk locations during outbreak situations. This feature allows for timely communication and awareness to individuals potentially at risk.

Results

A total of 7,108 tweets were downloaded for this study but only 524 tweets contained explicit location data, i.e. latitude and longitude. Figure 7 gives a comparative overview of the total tweets by users_1-3 vs. tweets containing GPS coordinates.

Case study 1 examined the relationship between receptor 1 (X (User_1) and lead exposure. Out of 825 census tracts in the state of Iowa, 302 were identified as high risk during the period January 20 to December 21 in 2021(Figure 8). The mobility patterns extracted from social media indicated that this user traversed locations designated as high-risk 34 times during the year.

Case study 2 focused on air pollution. The mobility pattern extracted from social media of user_2 revealed that the subject traversed a location where air quality was categorized as unhealthy (AQI 121) for sensitive persons based on a daily average from January 13, 2020, to December 29, 2020 (Figure 9). Moreover, 77 days of this period were deemed as very unhealthy, while seven were considered hazardous. Overall, 10 sub-regions had a daily AQI average rated as good, 125 rated as moderate and all other sub-regions in that country classified as unhealthy for sensitive persons.

Case study 3 investigated the possibility of exposure to COVID-19 by user_3 for the period January 10 to 17 in 2022. Here, the social media of user_3 indicated a mobility pattern revealing that the subject had multiple times been in a COVID-19 high-risk area (Figure 10).

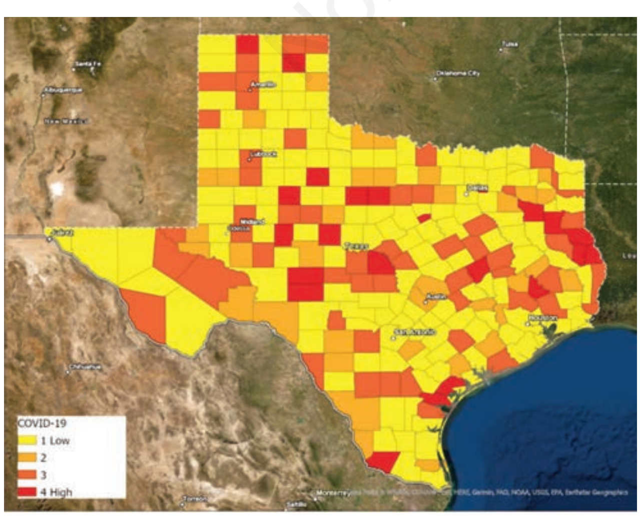

Figure 5. The COVID-19 risk in Texas U.S. in 2022 (January 1017). Source: https://globalepidemics.org/.

Discussion

The results may seem simplistic for a scientific report. However, the conceptual model proposed in this paper represents only the first step on road to a significantly improved disease diagnosis, a completely new approach, with great implications and potential. The model can help medical practitioners to investigate the source of an ailment more quickly and efficiently by identifying possible exposure sites. By identifying areas where the disease agent and receptor are present at the same time and location, the model can help medical practitioners to identify potential sources

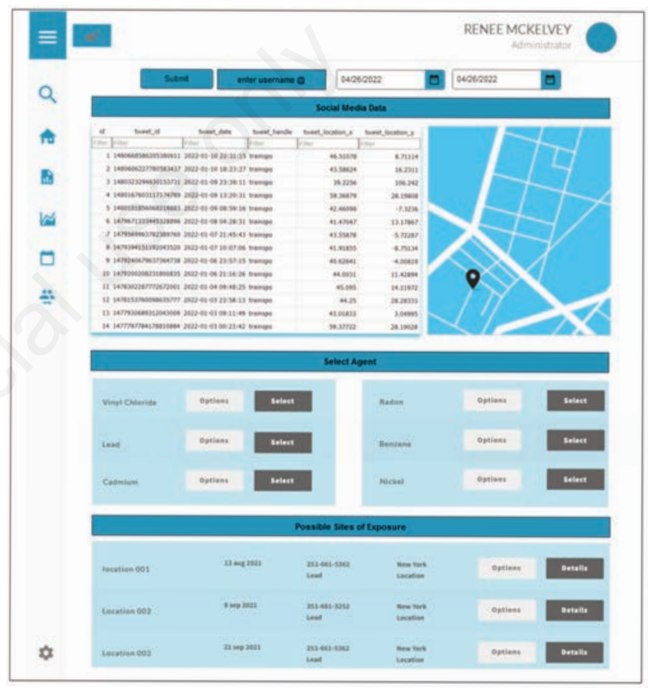

Figure 6. Illustrative mock-up of the graphical user interface application (GUI) showing the main components including Social media data download tab, Agent selection tab and a Results tab with possible exposure sites.

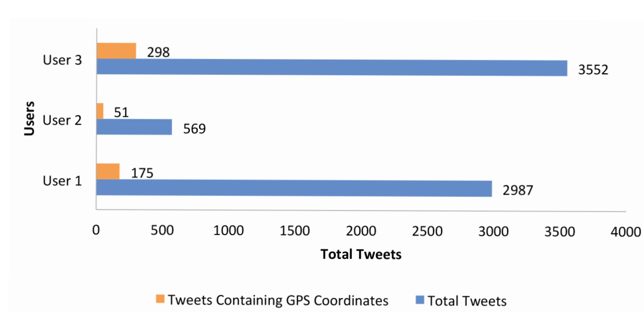

Figure 7. Comparative overview of the total tweets vs. tweets containing GPS coordinates.

of infection, leading to better health outcomes for patients.

In addition to an improved outlook for individual diagnosis, the ramification of the model would also improve public health surveillance. The model can be used as a early warning system (EWS) for potential outbreaks of infectious diseases by identifying and monitoring areas where the disease agent and receptor are present. This would lead to faster responses and more effective disease control measures, ultimately improving public health outcomes. The study also has implications for policy-making. The information generated by the model can help to identify areas of high risk or vulnerability, informing decision-makers where to allocate resources for disease control measures. This information can also contribute to the development of strategies for disease prevention and control. By identifying potential exposure sites, policy-makers can implement measures to reduce the risk of disease transmission and prevent outbreaks.

The X (Twitter) data

The percentage of tweets containing GPS coordinates in this study (7.4%) is slightly above the global average of 1-3% (Zohar 2021). However, the percentage varies by case study; for instance, case study 2, which examined air quality in China and user_2, was the lowest total number of tweets, in particular tweets containing location data. Several factors account for this; principally, X (Twitter) is not one of China's major social media platforms. According to Thomala (2023), the most popular platform in China is WeChat (https://www.wechat.com) followed by Douyin (https://www.douyin.com) (Figure 11). illustrates the absence of

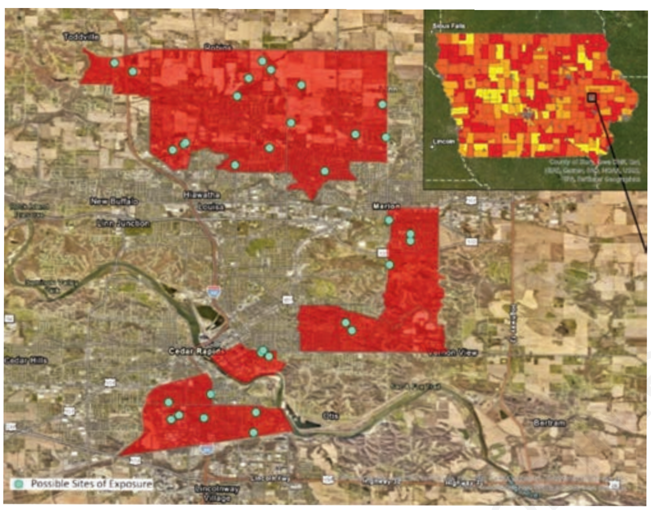

Figure 8. Possible sites of lead exposure for a patient in Iowa, U.S. in 2021 (January 20 to December 21).

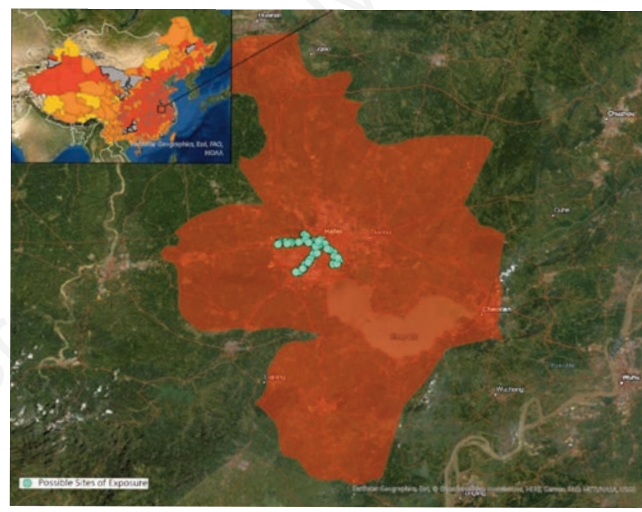

Figure 9. The mobility pattern of a person in China extracted from social media 2020 (January 13 - December 29) indicates high exposure of air pollution.

Figure 10. Mobility extracted from a person in Texas, U.S. in 2022 (January 10 to 17) indicating probable exposures to COVID-19.

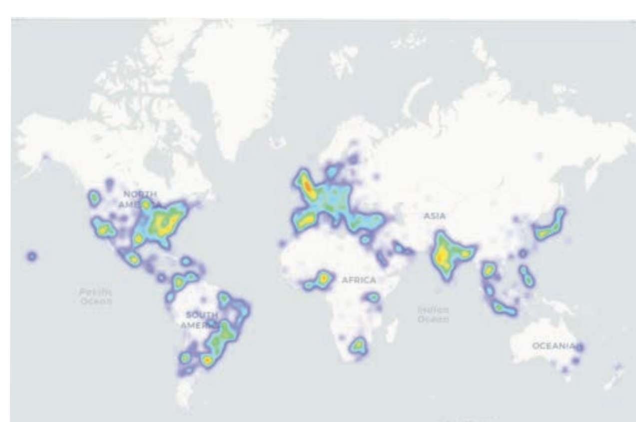

Figure 11. Visible absence of tweeting in China as shown by a 24-hour (August 27, 2020) geo-localized dataset delivered by X (Twitter) streaming public application programming interface. Source: https://onemilliontweetmap.com/

tweets in China compared to their global geographic distribution.

The unpopularity of X (Twitter) in China notwithstanding, this does not limit the conceptual model developed in this study, which is developed for any social media platform that allows for data export, including location data. Previous studies have shown that human mobility patterns can be extracted from social media platforms such as Sina Weibo, Tencent (https://www.tencent.com) and QQ (https://im.qq.com) (Kwan, 2018), three of the most widely used social media platforms in China. Conceptually, our model can achieve the same results. However, with the limited proportion of tweets containing geolocation data for user_2, the conceptual model could still identify possible sites of exposure to unhealthy air quality.

Agent and receptor modelling

We have introduced a new and innovative conceptual and practical model that integrates individual mobility patterns extracted from X (Twitter) data with various geo-enabled datasets, including information on disease outbreaks, endemic diseases and environmental pollution, in order to determine potential exposure. While previous research has examined the use of geotagged tweets, limited attention has been given to individual-level characteristics, such as mobility patterns and disease exposure, with a focus on spatial and temporal resolution (Sinnenberg *et al.*, 2017; Xu *et al.*, 2020). This level of detail is crucial for capturing exposure variability within a study population.

As emphasized by Zachlod *et al.* (2022), the volume of user-generated data, coupled with technological advancements and expertise, is continuously expanding. Social media platforms produce a vast quantity of data, a considerable proportion of which may not be immediately applicable for research purposes. Thus, specialized tools and methodologies are required to extract pertinent information from SMD. In public health research, GIS tools can be employed to merge SMD with spatial data to gain insights into the geographical patterns of health-related discussions on social media as well as exposure risk distributions (Ajayakumar *et al.*, 2019; Jing *et al.*, 2023). Consequently, our proposed conceptual model is pertinent in this field.

The proposed conceptual model that combines GIS with individual mobility patterns obtained from SMD and exposure agents' geographic footprints has demonstrated promising results in all three case studies, as discussed below. Case study 1 results demonstrate how the proposed conceptual model can be utilized to identify possible exposure sites to environmental pollutants and provide guidance for medical practitioners in further investigation. The extracted mobility patterns indicated that the user visited locations designated as high risk 34 times during the period. This suggests that the user may have experienced significant exposure to lead, and this can be extrapolated to determine the absolute daily exposure dose for all 34 visits.

Exposure to lead has been known to cause detrimental health effects, especially in vulnerable subgroups, which can result in neurological and developmental issues (WHO 2022). The extent of potential harm from lead exposure can be evaluated based on factors such as exposure duration, frequency, dose and the individual's physiological responses and the severity of lead exposure's adverse effects may vary depending on the level and duration of exposure (Singh *et al.*, 2018; Charkiewicz & Backstrand, 2020). Therefore, examining the mobility patterns and their lead exposure data could offer valuable insights for medical practitioners to explore the connection between exposure and unfavourable health outcomes, aiding in the establishment of a dose-response relationship (DeBord *et al.*, 2016).

The findings of Case study 2 highlight the efficacy of the proposed conceptual model in identifying possible exposure to air pollutants (as measured by the AQI) through individual mobility patterns mined from X (Twitter). The link between exposure to air pollutants and various health effects such as respiratory and cardiovascular diseases is well established (Kelly & Fussell, 2015; Al-Kindi *et al.*, 2020). However, it is important to note that the adverse impacts of air pollutants can vary depending on individual factors, as well as the type and concentration of the pollutant. For instance, frequent exposure to a toxic substance can cause adverse effects over time, high doses of exposure for a short period can lead to immediate adverse effects, while low doses of exposure for a long period can lead to chronic health effects (SCENIHR, SCHER, & SCCS, 2011; Katoto *et al.*, 2021). Thus, obtaining information on the spatial, temporal and specific dosage of air pollutants, as well as individual exposure profile mapped using our conceptual model, is crucial for assessing the potential health implications of exposure to toxic substances.

In Case study 3, we utilized data on a recent global pandemic to examine the likelihood of exposure to COVID-19 using the risk map that we developed. The analysis of mobility patterns extracted from SMD revealed that the individual had frequently visited an area classified as high risk for COVID-19 during the period analysed. This information provides insight into the spread of the disease within the community and the potential sources of infection (Aiello *et al.*, 2019; Allington *et al.*, 2021). For instance, if many people are contracting the virus after visiting a specific facility or location, it suggests that the place may be a transmission hotspot. By analysing exposure patterns, we can also identify individuals who may have been exposed to the virus and take necessary precautions such as testing and quarantine (Singh *et al.*, 2021; WHO 2021). Moreover, this approach can help pinpoint population subgroups or geographic regions that are more susceptible to the infection, enabling public health officials to target interventions effectively.

The case studies presented in this paper demonstrate that our proposed conceptual model has the potential to generate hypotheses regarding the transmission dynamics of infectious diseases and environmental pollutants. Furthermore, the model can guide retrospective studies after an incident occurs. Through the analysis of exposure patterns, researchers can identify variables that contribute to disease transmission, such as demographic factors, behavioural patterns and environmental conditions. By doing so, public health officials can develop focused interventions to limit further spread of the disease (Lin & Wen 2022; Luz & Masoodian 2022). These interventions can include measures such as increased testing, targeted vaccination campaigns and public health messaging designed to encourage behavioural change.

Strengths and limitations of the model

The conventional methods of identifying exposure sites depend on retrospective analysis of data, which may result in delays in identifying and managing health risks. However, the proposed model uses real-time data obtained from social media to identify possible exposure sites, leading to more rapid responses to emerging health risks. Additionally, this model can offer more personalized and targeted assessments of health risks. The proposed conceptual model can expand the use of GIS technologies and methodology to medical doctors and public health professionals beyond GIS professionals. By offering a user-friendly software environment, this proposed application can enable medical professionals to easily query location data against the spatial footprint of exposure agents to identify potential exposure sites. This approach represents a significant stride towards 'democratizing' access to GIS technologies and methodology and empowering medical professionals to use these tools to address real-world medical problems. It is important to acknowledge that the proposed conceptual model is not without limitations, and these limitations must be taken into consideration when utilizing the model. One such limitation is the potential distortion of location data due to the use of virtual private networks (VPNs) and other location-spoofing tools potentially utilized by social media users. Another limitation of the model is the reliance on disease data that may be sparse in certain regions of the world, which makes it difficult to accurately associate mobility patterns with a particular disease footprint. To overcome these limitations, incorporating multiple data sources to validate the location data is recommended. This could involve the use of wearable devices or GPS-based services to cross-reference the location data obtained from social media. Additionally, other methods of acquiring location data, such as text mining, can be used to extract a higher yield of location data from social media. By incorporating such methods in the proposed conceptual model, the accuracy and reliability of the results obtained can potentially be improved.

Conclusions

The proposed conceptual model, which integrates SMD and GIS technologies, presents a promising and practical approach to enhancing conventional disease exposure assessments. Significant strides can be made in addressing practical medical issues by leveraging the model. It has the potential to revolutionize public health investigations by providing real-time, dynamic information on individual behaviour and mobility patterns, facilitating the identification of exposure sites for infectious diseases and environmental agents. Despite the limitations mentioned, the importance of this type of innovative tools will only continue to grow in the face of various public health challenges. Integrating multiple data sources and emerging technological advancements, such as wearable devices and text mining, could potentially enhance the model further, paving the way for a transformative impact on public health and disease investigation.

References

Aiello AE, Renson A, Zivich PN, 2019. Social media- and internet-based disease surveillance for public health. Annu Rev Public Health 41:101-18.

Ajayakumar J, Curtis AJ, Curtis J, 2019. Addressing the data guardian and geospatial scientist collaborator dilemma: how to share health records for spatial analysis while maintaining patient confidentiality. Int J Health Geogr 18:30.

Al-Kindi SG, Brook RD, Biswal S, Rajagopalan S, 2020. Environmental determinants of cardiovascular disease: lessons learned from air pollution. Nat Rev Cardiol 17:656–72.

Allam Z, Jones DS, 2020. On the Coronavirus (Covid-19) outbreak and the smart city network: universal data sharing standards coupled with Artificial Intelligence (Ai) to benefit urban health monitoring and management. Healthcare (Basel) 8:46.

Allington D, Duffy B, Wessely S, Dhavan N, Rubin J, 2021. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 51:1763–69.

Alonso SG, de la Torre Díez I, Rodrigues JJPC, Hamrioui S, López-Coronado M, 2017. A Systematic Review of Techniques and Sources of Big Data in the Healthcare Sector. J Med Syst 41:183.

Amer S, & Bergquist R. 2021. Transport geography: Implications for public health. Geospatial Health 16:1009

Angelo KM, Kozarsky PE, Ryan ET, Chen LH, Sotir MJ, 2017. What proportion of international travellers acquire a travel-related illness? A review of the literature. J Travel Med 24:10.1093/jtm/tax046.

Assi MA, Hezmee MNM, Haron AW, Sabri MYM, Rajion MA, 2016. The detrimental effects of lead on human and animal health. Vet World 9:660–71.

Basch CH, Hillyer GC, Jaime C, 2022. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 34:367–69.

Bisanzio D, Kraemer MUG, Bogoch II, Brewer T, Brownstein JS, Reithinger R, 2020. Use of twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale. Geospatial Health 15:882

Briggs D, 2005. The Role of GIS: coping with space (and time) in air pollution exposure assessment. J Toxicol Environ Heal Part A 68:1243–61.

Charkiewicz AE, Backstrand JR, 2020. Lead toxicity and pollution in Poland. Int J Environ Res Public Health 17:4385.

Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Pastore Y Piontti A, Mu K, Rossi L, Sun K, Viboud C, Xiong X, Yu H, Halloran ME, Longini IM Jr, Vespignani A, 2020. The effect of travel restrictions on the spread of the 2019 novel Coronavirus (COVID-19) Outbreak. Science 368:395–400.

Cleckner H, Allen TR, 2014. Dasymetric Mapping and Spatial Modeling of Mosquito Vector Exposure, Chesapeake, Virginia, USA. ISPRS Int J Geo-Information 3:891–913.

De Cola L, 2002. Spatial forecasting of disease risk and uncertainty. Cartogr Geogr Inf Sci 29:363–80.

De Guzman C, 2022. Asia has kept COVID-19 at bay for 2 years. Omicron could change that. Retrieved September 29, 2022. Available from: https://time.com/6139851/asia-omicron-covid-surge/

DeBord DG, Carreón T, Lentz TJ, Middendorf PJ, Hoover MD, Schulte PA, 2016. Use of the 'exposome' in the practice of epidemiology: a primer on -omic technologies. Am J Epidemiol 184:302–14.

Dummer T, 2008. Health geography: supporting public health policy and planning. C Can Med Assoc J 178:1177–80.

Edward R, Wilson M, Kain K, 2002. Illness after international travel. N Engl J Med 347:1984.

Fauci AS, Lane HC, Redfield RR, 2020. Covid-19 — navigating the uncharted. N Engl J Med 382:1268–69.

Fujita H, 2017. Information extraction and visualization from Twitter considering spatial structure. Cartogr Int J Geogr Inf Geovisualization 52:178–93.

Fung IC, Duke CH, Finch KC, Snook KR, Tseng PL, Hernandez

AC, Gambhir M, Fu KW, Tse ZTH, 2016. Ebola virus disease and social media: a systematic review. Am J Infect Control 44:1660–71.
Gardner LM, Bóta A, Gangavarapu K, Kraemer MUG, Grubaugh ND, 2018. Inferring the Risk Factors behind the Geographical Spread and Transmission of Zika in the Americas. PLoS Negl Trop Dis 12:e0006194.
Gautam S, 2020. The influence of COVID-19 on air quality in India: a boon or inutile. Bull Environ Contam Toxicol 104:724–26.
Gulnerman AG, Karaman H, Pekaslan D, Bilgi S, 2020. Citizens' spatial footprint on Twitter—anomaly, trend and bias investigation in Istanbul. ISPRS Int J Geoinf 9:222.
Gunasekeran DV, Chew A, Chandrasekar EK, Rajendram P, Kandarpa V, Rajendram M, Chia A, Smith H, Leong CK, 2022. The impact and applications of social media platforms for public health responses before and during the COVID-19 pandemic: systematic literature review. J Med Internet Res 24:e33680.
Hao J, Wang L, 2005. Improving urban air quality in China: Beijing case study. J Air Waste Manage Assoc 55:1298–305.
Heldman AB, Schindelar J, Weaver JB, 2013. Social media engagement and public health communication: implications for public health organizations being truly 'social.' Public Health Rev 35:1–18.
Janke K, 2014. Air pollution, avoidance behaviour and children's respiratory health: evidence from England. J Health Econ 38:23–42.
Jing F, Li Z, Qiao S, Zhang J, Olatosi B, Li X, 2023. Using geospatial social media data for infectious disease studies: a systematic review. Int J Digit Earth 16:130–57.
Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D, 2015. Understanding human mobility from Twitter. PLoS One 10:e0131469.
Kamel Boulos MN, Geraghty EM, 2020. 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 19:8.
Katoto PDMC, Brand AS, Bakan B, Obadia PM, Kuhangana C, Kayembe-Kitenge T, Kitenge JP, Nkulu CBL, Vanoirbeek J, Nawrot TS, Hoet P, Nemery B, 2021. Acute and chronic exposure to air pollution in relation with incidence, prevalence, severity and mortality of COVID-19: a rapid systematic review. Environ Heal 20:41.
Kelly FJ, Fussell JC, 2015. Air pollution and public health: emerging hazards and improved understanding of risk. Environ Geochem Health 37:631–49.
Kozel TR, Burnham-Marusich AR, 2017. Crossm diseases: past, present, and future. J Clin Microbiol 55:2313–20.
Krutikov M, Manson J, 2016. Chikungunya virus infection: an update on joint manifestations and management. Rambam Maimonides Med J 7:e0033.
Kwan M, 2018. Human mobility, spatiotemporal context, and environmental health: recent advances in approaches and methods. Int J Environ Res Public Health 15(308).
Lanphear BP, Rauch S, Auinger P, Allen RW, Hornung RW, 2018. Low-level lead exposure and mortality in US adults: a population-based cohort study. Lancet Public Heal 3:e177–84.
Leta S, Beyene TJ, De Clercq EM, Amenu K, Kraemer MUG, Revie CW, 2018. global risk mapping for major diseases transmitted by Aedes Aegypti and Aedes Albopictus. Int J Infect Dis 67:25–35.
Li D, Wang S, Li D, 2015. Spatial data mining. Berlin, Heidelberg: Springer Berlin Heidelberg, 308 pp.
Liao Y, Yeh S, 2018. Predictability in human mobility based on geographical-boundary-free and long-time social media data. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, IEEE 2068-2073 pp.
Lin CH, Wen TH, 2022. How spatial epidemiology helps understand infectious human disease transmission. Tropical Medicine and Infectious Disease 7:164.
Liu Z, Zhou X, Shi W, Zhang A, 2018. Towards detecting social events by mining geographical patterns with VGI data. ISPRS Int J Geoinf 7:481.
Luz S, Masoodian M, 2022. Exploring environmental and geographical factors influencing the spread of infectious diseases with interactive maps. Sustainability 14:9990.
Matsuyama A, Yasuda Y, Yasutake A, Xiaojie L, Pin J, Li L, Mei L, Yumin A, Liya Q, 2006. etailed pollution map of an area highly contaminated by mercury containing wastewater from an organic chemical factory in People's Republic of China. Bull Environ Contam Toxicol 77:82–87.
Mbunge E, Akinnuwesi B, Fashoto SG, Metfula AS, Mashwama P, 2021. A critical review of emerging technologies for tackling COVID-19 pandemic. Hum Behav Emerg Technol 3:25–39.
Middleton SE, Kordopatis-Zilos G, Papadopoulos S, Kompatsiaris Y, 2018. Location extraction from social media. ACM Trans Inf Syst 36:1–27.
Musa GJ, Chiang PH, Sylk T, Bavley R, Keating W, Lakew B, Tsou HC, Hoven CW, 2013. Use of GIS mapping as a public health tool - from cholera to cancer. Heal Serv Insights 6:111–16.
Okami S, Kohtake N, 2016. Fine-scale mapping by spatial risk distribution modeling for regional malaria endemicity and its implications under the low-to-moderate transmission setting in western cambodia. PLoS One 11:e0158737.
Otsuki S, Nishiura H, 2016. Reduced risk of importing ebola virus disease because of travel restrictions in 2014: a retrospective epidemiological modeling study. PLoS One 11:e0163418.
Panteras G, Wise S, Lu X, Croitoru A, Crooks A, Stefanidis A, 2015. Triangulating social multimedia content for event localization using Flickr and Twitter. Trans GIS 19:694–715.
Photis Y, 2016. Disease and health care geographies: mapping trends and patterns in a GIS. Heal Sci J 10:1–8.
SCENIHR, SCHER, and SCCS. 2011. Toxicity and Assessment of Chemical Mixtures. European Union; pp. 1–50.
Schlagenhauf P, Weld L, Goorhuis A, Gautret A, Weber R, von Sonnenburg F, Lopez-Vélez R, Jensenius M, Cramer J, Field V, Odolini S, Gkrania-Klotsas E, Chappuis F, Malvy D, van Genderen PJJ, Mockenhaupt F, Jauréguiberry S, Smith C, Beeching NJ, Ursing J, Rapp C, Parola P, Grobusch MP, EuroTravNet. 2015. Travel-Associated Infection Presenting in Europe (2008-12): An Analysis of EuroTravNet Longitudinal, Surveillance Data, and Evaluation of the Effect of the Pre-Travel Consultation. Lancet Infect Dis 15:55–64.
Scholz J, Jeznik J, 2020. Evaluating geo-tagged twitter data to analyze tourist flows in Styria, Austria. ISPRS Int J Geo-Information 9:doi: 10.3390/ijgi9110681.
Şengül AT, Bilgin Büyükkarabacak Y, Durgun Yetim T, Pirzirenli MG, Çelik B, Başoğlu A, 2013. Early diagnosis saves lives in

esophageal perforations. Turkish J Med Sci 43;939945.
Singh G, Singh V, Wang ZX, Voisin G, Lefebvre F, Navenot JM, Evans B, Verma M, Anderson DW, Schneider JS, 2018. Effects of developmental lead exposure on the hippocampal methylome: influences of sex and timing and level of exposure. Toxicol Lett 290:63–72.
Singh PK, Nandi S, Ghafoor KZ, Ghosh U, Rawat DB, 2021. Preventing COVID-19 spread using information and communication technology. IEEE Consum Electron Mag 10:18–27.
Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM, 2017. Twitter as a tool for health research: a systematic review. Am J Public Health 107:e1–8.
Sloan L, Morgan J, 2015. Who tweets with their location? understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. PLoS One 10:e0142209.
Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, Caffery LJ, 2020. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 26:309–13.
Thomala L, 2023. China: most popular social media platforms 2022. Statista Inc. Retrieved February 19, 2024. Available from: https://www.statista.com/statistics/250546/leading-social-network-sites-in-china/
Tompkins AM, McCreesh N, 2016. Migration statistics relevant for malaria transmission in Senegal derived from mobile phone data and used in an agent-based migration model. Geospatial Health 11:408
Trajer A, 2021. Aedes Aegypti in the mediterranean container ports at the time of climate change: a time bomb on the mosquito vector map of Europe. Heliyon 7:e07981.
Wallace E, 2023. Lead exposure risk in your neighborhood | PolicyMap. Retrieved May 25, 2023. Available from: https://www.policymap.com/blog/lead-exposure-risk-in-your-neighborhood
WHO, World Health Organization, 2021. Contact tracing in the context of COVID-19. WHO guidelines: contact tracing in the context of COVID-19 2019 (May, 10):1–7.
WHO, World Health Organization, 2022. Lead poisoning fact sheet. 4:230. Available from: https://www.who.int/news-room/fact-sheets/detail/lead-poisoning-and-health#:~:text=At high levels of exposure,intellectual disability and behavioural disorders
Wu L, Zhi Y, Sui Z, Liu Y, 2014. Intra-urban human mobility and activity transition: evidence from social media check-in data. PLoS One 9:e97010.
Xu P, Dredze M, Broniatowski DA, 2020. The twitter social mobility index: measuring social distancing practices with geolocated tweets. J Med Internet Res 22:e21499.
Yasobant S, Vora KS, Hughes C, Upadhyay A, Mavalankar DV, 2015. Geovisualization: A Newer GIS Technology for Implementation Research in Health, J Geogr Inf Syst 07:20–28.
Ye X, Li S, Yang X, Qin C, 2016. Use of social media for the detection and analysis of infectious diseases in China. ISPRS Int J Geoinf 5:156.
Yousefinaghani S, Dara R, Poljak Z, Bernardo TM, Sharif S, 2019. The assessment of Twitter's potential for outbreak detection: avian influenza case study. Sci Rep 9:18147.
Zachlod C, Samuel O, Ochsner A, Werthmüller S, 2022. Analytics of social media data – state of characteristics and application. J Bus Res 144:1064–76.
Zohar M, 2021. Geolocating Tweets via spatial inspection of information inferred from Tweet Meta-Fields. Int J Appl Earth Obs Geoinf 105:102593.