



## OPEN ACCESS

## EDITED BY

João Valente Cordeiro,  
New University of Lisbon, Portugal

## REVIEWED BY

Raj Setia,  
Punjab Remote Sensing Centre  
(PRSC), India  
Qiyong Liu,  
National Institute for Communicable  
Disease Control and Prevention, China

## \*CORRESPONDENCE

Yuzhou Zhang  
zyzjacky88@gmail.com

## SPECIALTY SECTION

This article was submitted to  
Digital Public Health,  
a section of the journal  
Frontiers in Public Health

RECEIVED 01 June 2022

ACCEPTED 09 November 2022

PUBLISHED 05 December 2022

## CITATION

Sun H, Zhang Y, Gao G and Wu D  
(2022) Internet search data with  
spatiotemporal analysis in infectious  
disease surveillance: Challenges and  
perspectives.  
*Front. Public Health* 10:958835.  
doi: 10.3389/fpubh.2022.958835

## COPYRIGHT

© 2022 Sun, Zhang, Gao and Wu. This  
is an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction  
in other forums is permitted, provided  
the original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Internet search data with spatiotemporal analysis in infectious disease surveillance: Challenges and perspectives

Hua Sun<sup>1</sup>, Yuzhou Zhang<sup>1,2\*</sup>, Guang Gao<sup>1</sup> and Dun Wu<sup>1</sup>

<sup>1</sup>Popsmart Technology (Zhejiang) Co., Ltd, Ningbo, China, <sup>2</sup>College of Computer Science and Technology, Zhejiang University, Hangzhou, China

With the rapid development of the internet, the application of internet search data has been seen as a novel data source to offer timely infectious disease surveillance intelligence. Moreover, the advancements in internet search data, which include rich information at both space and time scales, enable investigators to sufficiently consider the spatiotemporal uncertainty, which can benefit researchers to better monitor infectious diseases and epidemics. In the present study, we present the necessary groundwork and critical appraisal of the use of internet search data and spatiotemporal analysis approaches in infectious disease surveillance by updating the current stage of knowledge on them. The study also provides future directions for researchers to investigate the combination of internet search data with the spatiotemporal analysis in infectious disease surveillance. Internet search data demonstrate a promising potential to offer timely epidemic intelligence, which can be seen as the prerequisite for improving infectious disease surveillance.

## KEYWORDS

internet search data, spatiotemporal analysis, infectious disease, surveillance, prediction

## Introduction

In recent years, it has been realized that internet search data have great potential in infectious disease surveillance, which can be proved by increasing the use of such data to conduct rapid epidemics tracking and surveillance (1). As those data can offer timely surveillance intelligence with a high spatial resolution (2), they could be seen as a novel data source to monitor diseases and epidemics at both space and time scales.

The query “internet search data” in disease surveillance refers to social media data, internet search data, medicine sales data, and online news data (3). Infectious diseases continue to pose public health threats with a large social burden (4), and sometimes, they can cause significant pandemics, such as the coronavirus pandemic. To decrease the effects of infectious diseases on our society, it is critical to improve infectious disease surveillance (5).

The main aim of developing a disease surveillance system was to successfully predict possible diseases and epidemics or even outbreaks *via* their ability of early warning based on the data sources (6). Traditional infectious disease surveillance is based on the passive report system, which collects disease notifications from healthcare organizations. This kind of system is typically accurate but can delay up to 2 weeks from patients' diagnosis to the notifications being compiled into the surveillance system (7). As a result, this lag in the reporting process can post an adverse impact on the capability of the infectious disease surveillance system. Such a system may not offer real-time epidemiological intelligence, which leads to the reduction of the efficiency of infectious disease's quick response (8).

Public health experts consider spatiotemporal uncertainty in several manners. The geocoding process, e.g., using disease surveillance data with unreliable coordinates, could lead to spatial uncertainty. In addition, temporal uncertainty in disease surveillance typically causes a time lag between the occurrence of symptoms and case reporting *via* case identification after medical diagnosis (9). In the disease surveillance field, spatiotemporal uncertainty can present at the stages of data collection and statistical analysis. As a result, considering the spatiotemporal uncertainty can indeed contribute to the improvement of surveillance and even decision-making. The lack of spatiotemporal information in data collection and statistical analysis processes may lead to potential errors (e.g., false alarms) in disease surveillance and weaken the benefits of health actions, which is aimed to reduce the impact of diseases (10, 11).

However, few studies argued the use of internet search data with spatiotemporal analysis approaches in infectious disease surveillance. In this review, we present the challenges and possible future directions for researchers to investigate the combination of internet search data with spatiotemporal analysis in infectious disease surveillance.

## The application of internet search data

### The development of internet-based surveillance

Earlier, internet-based surveillance relied on online news as the main kind of data source, which results from the passive reading of online information for internet users at Web 1.0 (12). However, current studies of internet-based infectious disease surveillance use various internet data sources, including search query data from online search engines and social media such as Twitter (13). This mainly results from the information revolution and the rise of Web 2.0, which triggered the use of the internet as a new tool to actively and frequently seek

health-related information (14). Thus, disease activity can be estimated by collecting and tracking changes in frequencies of related internet searches for key terms (15).

Several famous internet-based infectious disease surveillance systems have been successfully built using non-structured, event-triggered, internet search data. The Public Health Agency of Canada developed The Global Public Health Intelligence Network (GPHIN) to assist public health agencies, as well as the World Health Organization (WHO) Global Outbreak Alert and Response Network to detect infectious disease outbreaks using retrieved online information, such as online news. The network has first displayed its great ability during the severe acute respiratory syndrome (SARS) outbreak in 2003, with 2-months earlier reports for SARS than the official one by WHO (16).

Moreover, in the age of "Web 2.0," the technologies of the proliferation of Really Simple Syndication (RSS) and Asynchronous JavaScript and XML enable researchers to develop more interactive infectious disease surveillance, such as HealthMap (17). This powerful internet search data-based surveillance used a wide range of internet external feeds, such as online news to collect valuable disease-related information, and then visualized the critical information, such as disease type, date, and location to the public as an early warning.

## The internet search data type

A variety of internet search data can be applied for infectious disease surveillance. Generally, the applied categories include internet search metrics (The volume of internet search activity) and mined social media data (The volume of social media posts). Additionally, the combination of the above internet data source with other data sources also has a great potential for surveillance, such as self-diagnosis questionnaires online, medication sales data, and school absenteeism data.

As a new tool, internet search data relies on the basis that the population group who have a great possibility of infections will actively search related information online about their health conditions. Thus, disease epidemic patterns can be tracked by watching the dynamic of search volume in related internet search activities for certain internet search queries. Internet search data enables investigators to discover disease patterns from timely intelligence at a larger spatial scale (18). As internet use worldwide is currently dominated by various search engines by country, reviewed studies used the dominating internet search engine data by study settings. In our review, several studies using Google (19, 20), Baidu (21–23), and other search term data (24–27) have been performed worldwide to successfully detect infectious disease events.

Social media communication is an increasingly utilized platform to monitor personal health information and contents, which is the main advancement of it compared to other internet

data sources (28). Moreover, the interest in using social media data to track infectious diseases is increasing because of the timely data generated by internet users on the platform (29). Thus, social media data are a perfect source for detecting disease in the early stage because of their timeliness characteristic. This characteristic also enables health authorities to contact the public in the early phase of disease outbreak detection (30). We identified several original, exploratory studies on infectious diseases targeting social media users between 1 January 2000 and 30 June 2017. Both Twitter and other blogs claimed to be seen as valuable social media data sources in infectious disease surveillance (31–34).

## The data processing of internet search data

Through the common data collecting and processing steps of our reviewed studies (Figure 1), first, all data related to infectious diseases were collected from the internet. The studies that used “internet search data” as a variable collected the search volume from search engine websites and that used “social media data” as a data source that collected diseases-related contents through their application programming interface (API). In this stage, most of the included articles in this systematic review collected their data using key terms within specific time periods and locations. However, for social media data, not all data collected are associated with the specified diseases (35). Textual analysis of data was needed to identify disease-related and non-disease-related data to detect and track disease events. Thus, the second stage involves efficient social media data filtering and classification. Machine learning approaches are commonly performed in reviewed studies to classify whether the collected social media data are relevant to disease events (36). The final stage is to evaluate the predictive accuracy and time efficiency of internet-based surveillance compared to conventional surveillance.

## The spatiotemporal internet search data with analysis approaches

Traditional data analysis approaches often generate correlations with biases under the assumption that the independent variables have no autocorrelation at both spatial and temporal scales (37). However, such autocorrelation is very common in the real world.

The main advancement of internet search data is including rich spatiotemporal information, with uncertainty contents at both space and time scales. Internet search data generated from users' Internet Protocol Address can accurately reveal the users' locations. For example, internet search volume data (e.g., Google Trends) aggregated the internet search activities

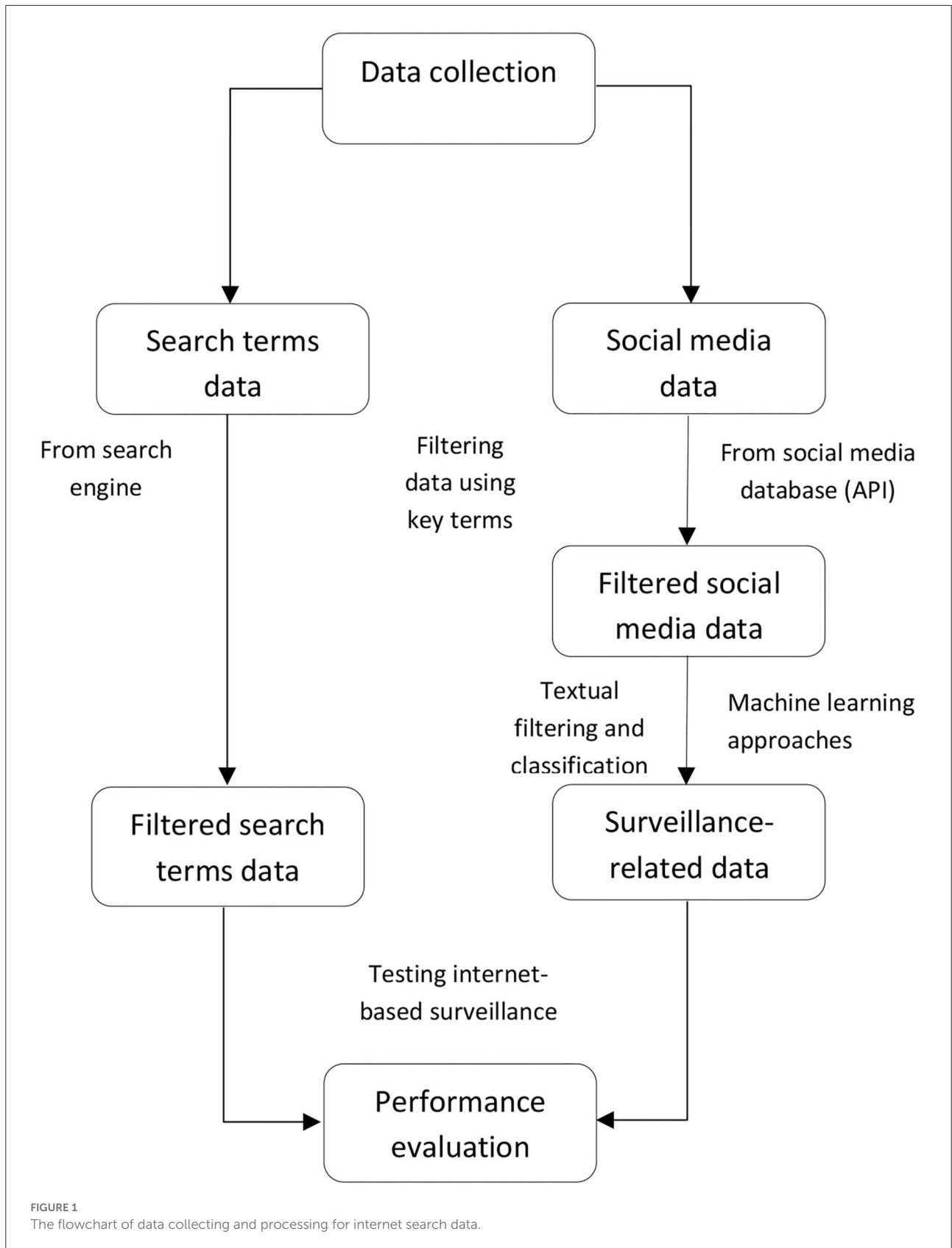
in a certain area with search terms. The geographically tagged tweets indicated the specific locations where the health issues occurred. This can avoid the spatial uncertainty between the real locations of the users and the geocoded address using coordinates, which has been widely used in traditional disease surveillance (38). Moreover, internet search data are produced in real time, including personal health issues, symptoms, and so on. Thus, using internet search data in disease surveillance can limit the time lag in the disease reporting process (39).

Spatiotemporal analysis in the domain of disease surveillance refers to analyzing the surveillance data with geospatial attributes in a time series (40). Elliott et al. identified several spatiotemporal analysis approaches in epidemiology, including disease mapping, disease cluster identification, and correlation analysis (41). Disease mapping is used to show the geographic location of events or attributes, which is useful for communicating trends or averages in an area. Moreover, disease cluster identification can ensure an unusually large aggregation of a relatively uncommon medical condition or event within a particular geographical location or period, which can be seen as an essential prerequisite for identifying an outbreak. Furthermore, the term “correlation analysis in disease surveillance” refers to measuring the strength of the association using statistical methods between event occurrence and potential risk factors, such as environmental factors and sociodemographic factors.

The advances in spatiotemporal analysis approaches can compensate for the residual variability as the spatial variation in data analysis processing, which may lead to a decrease in the effects of potential errors (42). It is critical to simultaneously include the spatiotemporal components and spatiotemporal uncertainty variables in the surveillance (43). Spatiotemporal analysis approaches have additional advances compared to the analysis methods that purely applied spatial or temporal analysis approaches, which is due to the dynamics in spatial patterns over time and temporal patterns at different spatial units. Overall, the spatiotemporal analysis approaches enable simultaneous data analysis at both space and time, as well as the investigation of any unusual spatiotemporal patterns (44).

## The spatiotemporal visualization in surveillance

As both internet search data and infectious disease surveillance data have rich information in space and time, an increasing number of studies used spatiotemporal visualization tools to map such variables. These tools enable researchers to show the spatiotemporal distribution of the route of disease transmission combined with the related internet search data. For instance, the HealthMap allows the demonstration of the map of currently active infectious diseases. Furthermore, the



map contains the links for further latest information about the diseases, which were retrieved from the internet (45).

## The spatiotemporal clustering in surveillance

The detection of disease clusters plays a critical role in surveillance, which can help health organizations to identify relatively high-risk areas. Mackey and colleagues investigated the clusters of Tweets with COVID-19-related symptoms or experiences from March 3, 2020 to March 20, 2020. The results indicated that the regions with a larger number of population-normalized COVID-19 confirmed case exhibited more tweets with COVID-19 associated symptoms or experiences (46). Chowell et al. applied online news data and health bulletins to discover the clustering of Ebola virus disease (EVD) cases from January 2014 to January 2016. In the study, there was a high correlation coefficient (Spearman  $\rho = 0.86$ ;  $P < 0.001$ ) between the monthly clusters number retrieved from online news and the officially reported number of EVD cases using traditional surveillance (47).

## The application of spatiotemporal models in surveillance

The high resolution of internet search data at both space and time has great promise to enhance disease surveillance by developing adaptive spatiotemporal models at different levels (national, state, and local government) of public health authorities (48).

Generous et al. applied internet search data on Wikipedia to forecast location of the disease in 14 countries with spatiotemporal linear models. Overall, the models successfully estimated the disease activities at a variety of time scales (49). Ma and Yang attempted to use Google Trends data to predict COVID-19 patterns in the United States at both national and state levels *via* regularized linear model, which incorporated a cross-state, cross-region spatiotemporal framework. The proposed model performed well in the predictions up to 4 weeks ahead (50). Zhang et al. developed seasonal auto-regressive integrated moving average (SARIMA) models in different local regions to discover the relationships between seasonal influenza epidemics and Google Trends data with identified key terms. The spatiotemporal contents were considered through the different parameters in the models by region, which can better fit the spatial heterogeneity (51). Li et al. developed generalized additive models (GAM) to predict dengue fever using Baidu Index data at the city level. The results indicated that internet search data (Baidu Index) promoted the forecasting performance at different time scales, compared to the model not using Baidu Index data (23).

The feature of internet search data provides a great opportunity to develop spatiotemporal models in surveillance at a finer spatial scale and time series. This enables public health authorities to better understand disease risks, especially in areas where traditional disease surveillance is poor (52). Furthermore, the flexible spatiotemporal modeling enables internet search data to generate dynamic surveillance in near real time (e.g., disease mapping and risk mapping) (53–56).

## The challenges in internet search data and spatiotemporal analysis

### The access to internet

As the internet search data is mainly based on the search activity online, it is crucial to consider the ability of internet access. A previous study reported that the majority of internet users were located in developing areas, such as Asia (53.7%), South America (10.2%), and Africa (10.1%) (57). As a result, these regions can be seen as the ideal places to collect internet search data. These regions may have great opportunities to develop infectious disease surveillance using internet search data.

### The internet search behavior

The overall performance of the models that purely used internet search data in infectious disease surveillance is widely noticed in previous research. However, such models may be subject to bias and come with potential errors. For instance, internet search data was applied to forecast the number of influenza peaking cases, but the number was two-fold higher than the number reported by the CDC (58). The accuracy of using internet search data in surveillance can be varied by internet search behaviors and media-driven bias. The widespread media reports may lead to many internet search activities by internet users who were not ill (59). Google Flu Trends (GFT) pioneered the internet search data-based flu surveillance in the world. GFT kept watching any changes in internet search behaviors to update its predictive models annually, which contribute to the goodness-of-fit to the reference flu surveillance data (60).

### The development of finer spatiotemporal analysis resolution

The main aim of developing infectious disease surveillance was to timely collect disease-related intelligence, which can contribute to decrease the impact of epidemics on the vulnerable population (32). Internet search data source is in an ideal location to conduct quick surveillance and monitoring as it

can timely reflect epidemic patterns at the defined spatial units, where internet access is available (35). Internet search data usually included geographic information. Search engines usually provide search volume data at the state or even lower level, and social media data usually include users' geographic locations. This nature can identify high-risk areas of infectious diseases by determining the areas with high volumes of internet data generation.

Google Flu Trends have made some great progress in the finer spatiotemporal resolution analysis, which offers city-level or finer spatial resolution internet search data in influenza surveillance (61). Although a finer spatiotemporal resolution in internet search data is limited by the capacity of data aggregation and internet search volume, the rapidly increasing use of internet search as a health knowledge tool could lead to the development of a better spatiotemporal analysis using a finer resolution internet search data in space and time (62).

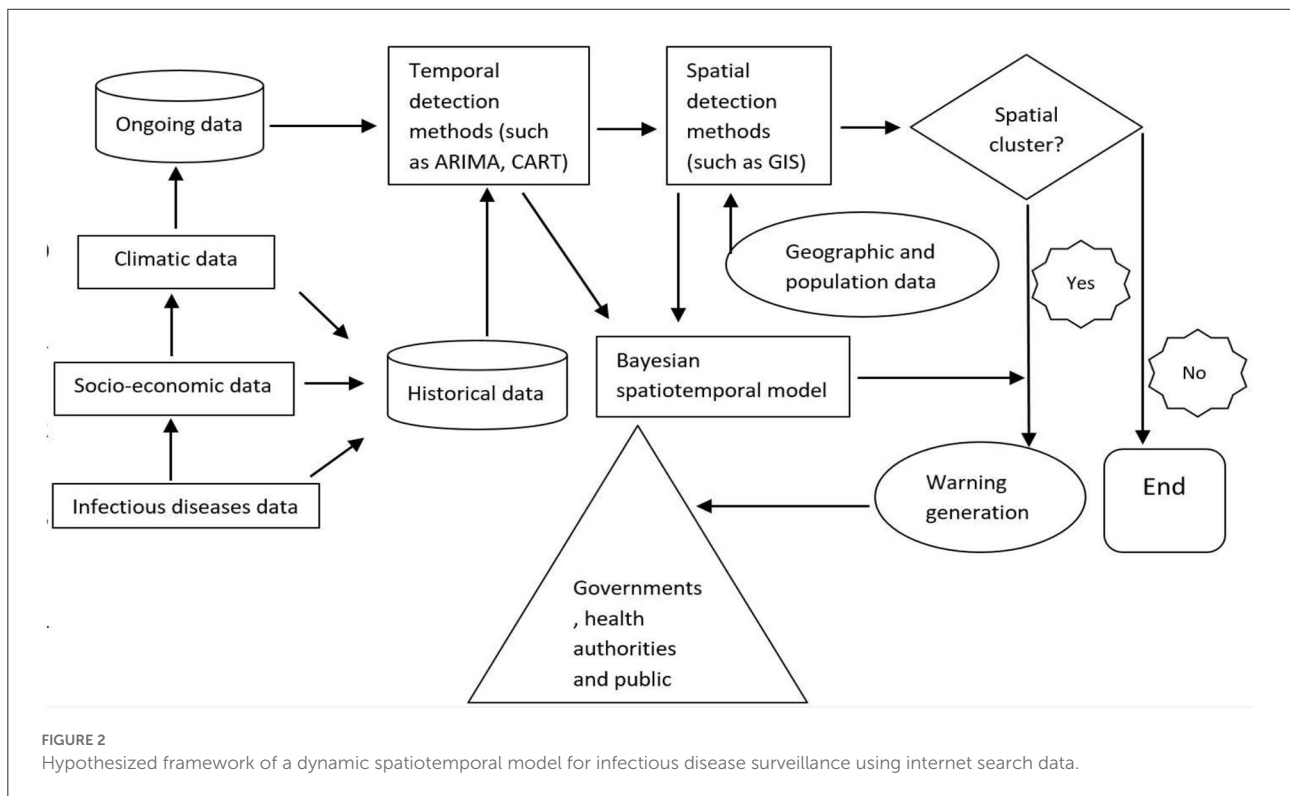
### Developing integrated surveillance using internet search data with spatiotemporal analysis

The surveillance system purely based on internet search data may generate noise value and failed predictive results. First, the accuracy of internet-based surveillance may be impacted by the levels of internet access (63). Second, it is acknowledged that

there are different internet-seeking behaviors, self-reporting, and media-driven bias between different sectors of the community (64). Previous studies reported that media bias can adversely impact internet-based surveillance systems (65). Third, the absence of involving other risk factors for infectious diseases, such as climatic and socio-economic factors, may contribute to the noise or failed prediction of infectious disease events. Finally, exploring spatiotemporal clustering plays a crucial role in surveillance. This enables health authorities to trigger disease intervention in high-risk windows to reduce the burden and impact of infectious diseases.

The successful surveillance systems for infectious diseases are affected by the complex conditions and the transmission patterns of the disease, and social-environmental variables (66). Previous studies indicated that infectious diseases have strong climatic and social-environmental patterns (67, 68). The involvement of climatic and social-environmental factors could improve the predictive performance of epidemiological models (69–71). The prediction results can contribute to the decision-making of certain control measures and surveillance, such as the allocation of healthcare resources, social distance control, vaccination plan, and health education.

Thus, a dynamic, integrated surveillance system using big data has the potential for timely and specifically detecting infectious disease events and reducing the potential errors introduced by factors such as fear-based searching. We designed a flowchart of a dynamic spatiotemporal model for infectious



disease surveillance using big data, which provides possible research directions for future study (Figure 2).

## Conclusion

Internet search data hold the potential as a free, easily accessible data source to access large community fraction of health-related data to reflect disease activity and generate timely disease information by targeting people in the early phase of the disease process (72). Ongoing evaluation, validation, and verification of internet search data-based surveillance with epidemiological and clinical data by users, developers, and agencies will greatly improve the utilization of this new surveillance approach for infectious disease detection and tracking. This study provides the necessary groundwork and critical appraisal of the use of internet search data and spatiotemporal analysis approaches. This study also provides future directions to researchers to investigate the combination of internet search data with spatiotemporal analysis in a wider range of infectious disease surveillance in more regions worldwide.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

HS and YZ contributed to the conceptualization, methodology, and writing—review and editing. GG and DW contributed to the data curation and funding acquisition.

## References

- Zhang Y, Milinovich G, Xu Z, Bambrick H, Mengersen K, Tong S, et al. Monitoring pertussis infections using internet search queries. *Sci Rep.* (2017) 7:10437. doi: 10.1038/s41598-017-11195-z
- Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature.* (2008) 451:990–3. doi: 10.1038/nature06536
- Wilson K, Brownstein JS. Early detection of disease outbreaks using the internet. *Can Med Assoc J.* (2009) 180:829–31. doi: 10.1503/cmaj.1090215
- Vlieg WL, Fanoy EB, van Asten L, Liu X, Yang J, Pilot E, et al. Comparing national infectious disease surveillance systems: China and the Netherlands. *BMC Public Health.* (2017) 17:415. doi: 10.1186/s12889-017-4319-3
- Garattini C, Raffle J, Aisyah DN, Sartain F, Kozlakidis Z. Big data analytics, infectious diseases and associated ethical impacts. *Philos Technol.* (2019) 32:69–85. doi: 10.1007/s13347-017-0278-y
- Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. *J Med Ethics.* (2020) 46:427–31. doi: 10.1136/medethics-2020-106314

All authors read and agreed to the published version of the manuscript, participated in data collection, preliminary analysis, early drafting of the manuscript, made substantive contributions to the development and revision of the manuscript, contributed to the article, and approved the submitted version.

## Funding

This research was supported by the Major Scientific and Technological Projects in Ningbo (2021Z050) and the Science and Technology Project of Zhejiang Provincial Department of Natural Resources (2020-16).

## Acknowledgments

The authors wish to acknowledge the support of Major Scientific and Technological Projects in Ningbo.

## Conflict of interest

HS, YZ, GG, and DW were employed by Popsmart Technology (Zhejiang) Co., Ltd.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Gilbert GL, Degeling C, Johnson J. Communicable disease surveillance ethics in the age of big data and new technology. *Asian Bioethics Rev.* (2019) 11:173–87. doi: 10.1007/s41649-019-00087-1
- Zhang Q. Data science approaches to infectious disease surveillance. *Philos Trans R Soc A.* (2022) 380:20210115. doi: 10.1098/rsta.2021.0115
- Delmelle EM, Desjardins MR, Jung P, Owusu C, Lan Y, Hohl A, et al. Uncertainty in geospatial health: challenges and opportunities ahead. *Ann Epidemiol.* (2022) 65:15–30. doi: 10.1016/j.annepidem.2021.10.002
- Griffith DA. Uncertainty and context in geography and giscience: reflections on spatial autocorrelation, spatial sampling, and health data. *Ann Am Assoc Geogr.* (2018) 108:1499–505. doi: 10.1080/24694452.2017.1416282
- Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. *Ann Epidemiol.* (2017) 27:1–9. doi: 10.1016/j.annepidem.2016.12.001
- Al-Garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: a systematic review. *J Biomed Inform.* (2016) 62:1–11. doi: 10.1016/j.jbi.2016.05.005

13. Johnson HA, Wagner MM, Hogan WR, Chapman WW, Olszewski RT, Dowling JN, et al., editors. *Analysis of Web Access Logs for Surveillance of Influenza*. Amsterdam: Medinfo (2004).
14. Andrianou XD, Pronk A, Galea KS, Stierum R, Loh M, Riccardo F, et al. Exposome-based public health interventions for infectious diseases in urban settings. *Environ Int*. (2021) 146:106246. doi: 10.1016/j.envint.2020.106246
15. Zimmerman DM, Mitchell SL, Wolf TM, Deere JR, Noheri JB, Takahashi E, et al. Great ape health watch: enhancing surveillance for emerging infectious diseases in great apes. *Am J Primatol*. (2022) 84:e23379. doi: 10.1002/ajp.23379
16. Mykhalovskiy E, Weir L. The global public health intelligence network and early warning outbreak detection: a Canadian contribution to global public health. *Can J Public Health*. (2006) 97:42–4. doi: 10.1007/BF03405213
17. Najebullah K, Liebig J, Darbro J, Jurdak R, Pains D. Timely surveillance and temporal calibration of disease response against human infectious diseases. *PLoS ONE*. (2021) 16:e0258332. doi: 10.1371/journal.pone.0258332
18. Goodman LB, Whittaker GR. Public health surveillance of infectious diseases: beyond point mutations. *Lancet Microbe*. (2021) 2:e53–4. doi: 10.1016/S2666-5247(21)00003-3
19. Jain D, Nair K, Jain V. *Factors Affecting GDP (Manufacturing, Services, Industry): An Indian Perspective*. Pune: Social Science Electronic Publishing (2015).
20. Husnayain A, Foad A, Su EC-Y. Applications of google search trends for risk communication in infectious disease management: a case study of the COVID-19 outbreak in Taiwan. *Int J Infect Dis*. (2020) 95:221–3. doi: 10.1016/j.ijid.2020.03.021
21. Han L, Wang Y, Hu K, Tang Z, Song X. The therapeutic efficacy of Huashi Baidu formula combined with antiviral drugs in the treatment of COVID-19: a protocol for systematic review and meta-analysis. *Medicine*. (2020) 99:e22715. doi: 10.1097/MD.00000000000022715
22. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from Baidu. *PLoS ONE*. (2013) 8:e64323. doi: 10.1371/journal.pone.0064323
23. Li Z, Liu T, Zhu G, Lin H, Zhang Y, He J, et al. Dengue Baidu search index data can improve the prediction of local dengue epidemic: a case study in Guangzhou, China. *PLoS Negl Trop Dis*. (2017) 11:e0005354. doi: 10.1371/journal.pntd.0005354
24. Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. *PLoS ONE*. (2009) 4:e4378. doi: 10.1371/journal.pone.0004378
25. Santillana M, Nsoesie EO, Mekaru SR, Scales D, Brownstein JS. Using clinicians' search query data to monitor influenza epidemics. *Clin Infect Dis*. (2014) 59:1446. doi: 10.1093/cid/ciu647
26. Seo D-W, Jo M-W, Sohn CH, Shin S-Y, Lee J, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. *J Med Internet Res*. (2014) 16:e289. doi: 10.2196/jmir.3680
27. Thorner AR, Cao B, Jiang T, Warner AJ, Bonis PA, editors. *Correlation between UpToDate Searches and Reported Cases of Middle East Respiratory Syndrome During Outbreaks in Saudi Arabia*. New York, NY: Oxford University Press (2016). doi: 10.1093/ofid/ofw043
28. Gianfredi V, Provenzano S, Santangelo OE. What can internet users' behaviours reveal about the mental health impacts of the COVID-19 pandemic? A systematic review. *Public Health*. (2021) 198:44–52. doi: 10.1016/j.puhe.2021.06.024
29. Huang W, Cao B, Yang G, Luo N, Chao N. Turn to the internet first? Using online medical behavioral data to forecast COVID-19 epidemic trend. *Inf Process Manag*. (2021) 58:102486. doi: 10.1016/j.ipm.2020.102486
30. Xu Q, Su Z, Zhang K, Yu S. Fast containment of infectious diseases with e-healthcare mobile social internet of things. *IEEE Internet Things J*. (2021) 8:16473–85. doi: 10.1109/JIOT.2021.3062288
31. Memos VA, Minopoulos G, Stergiou KD, Psannis KE. Internet-of-Things-Enabled infrastructure against infectious diseases. *IEEE Internet Things Mag*. (2021) 4:20–5. doi: 10.1109/IOTM.0001.2100023
32. Meraj M, Singh SP, Johri P, Quasim MT. Detection and prediction of infectious diseases using IoT sensors: a review. *Smart Comp*. (2021) 56–61. doi: 10.1201/9781003167488-8
33. Maxwell S, Grupac M. Virtual care technologies, wearable health monitoring sensors, and internet of medical things-based smart disease surveillance systems in the diagnosis and treatment of COVID-19 patients. *Am J Med Res*. (2021) 8:118–31. doi: 10.22381/ajmr8220219
34. Riley A, Nica E. Internet of things-based smart healthcare systems and wireless biomedical sensing devices in monitoring, detection, and prevention of COVID-19. *Am J Med Res*. (2021) 8:51–64. doi: 10.22381/ajmr8220214
35. Batrinenko AV, Denisova S, Lisovskii D, Orlov S, Soshnikov S. The internet search engines as an additional tool in public health research in the context of disease outbreaks. *Int J Health Govern*. (2022) 27:194–207. doi: 10.1108/IJHG-09-2021-0094
36. Arima Y, Kanou K, Arashiro T, Ko YK, Otani K, Tsuchihashi Y, et al. Epidemiology of coronavirus disease 2019 in Japan: descriptive findings and lessons learned through surveillance during the first three waves. *JMA J*. (2021) 4:198–206. doi: 10.31662/jmaj.2021-0043
37. Anggraeni W, Yuniarno EM, Rachmadi RF, Purnomo PMH. A Sparse Representation of Social Media, Internet Query, and Surveillance Data to Forecast Dengue Case Number using Hybrid Decomposition-Bidirectional LSTM. *Int J Intell Eng Syst*. (2021) 14:209–25. doi: 10.22266/ijies2021.1031.20
38. Siffel C, Strickland MJ, Gardner BR, Kirby RS, Correa A. Role of geographic information systems in birth defects surveillance and research. *Birth Defects Res Part A Clin Mol Teratol*. (2006) 76:825–33. doi: 10.1002/bdra.20325
39. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. *J Infect Dis*. (2016) 214:S375–9. doi: 10.1093/infdis/jiw400
40. Dong Z, Guo C, editors. A literature review of spatio-temporal data analysis. *J Phys Conf Ser*. (2021) 1792:012056. doi: 10.1088/1742-6596/1792/1/012056
41. Elliot P, Wakefield JC, Best NG, Briggs DJ. *Spatial Epidemiology: Methods and Applications*. New York, NY: Oxford University Press (2000). doi: 10.1093/acprof:oso/9780198515326.001.0001
42. Javadi M, Khan IH. Internet of things (IoT) enabled healthcare helps to take the challenges of COVID-19 pandemic. *J Oral Biol Craniofacial Res*. (2021) 11:209–14. doi: 10.1016/j.jobcr.2021.01.015
43. Garrett R, Young SD. Digital public health surveillance tools for alcohol use and HIV risk behaviors. *AIDS Behav*. (2021) 25:333–8. doi: 10.1007/s10461-021-03221-z
44. Lami F, Amiri M, Majeed Y, Barr KM, Nsour MA, Khader YS. Real-time surveillance of infectious diseases, injuries, and chronic conditions during the 2018 Iraq Arba'een mass gathering. *Health Secur*. (2021) 19:280–7. doi: 10.1089/hs.2020.0074
45. Kopsco HL, Krell R, Connally NP, Mather T. Identifying trusted sources of lyme disease prevention information among internet users. *Prim Care*. 9:8–2. doi: 10.2196/preprints.37871
46. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: retrospective big data infoveillance study. *JMIR Public Health Surveill*. (2020) 6:e19509. doi: 10.2196/19509
47. Chowell G, Cleaton JM, Viboud C. Elucidating transmission patterns from internet reports: Ebola and Middle East respiratory syndrome as case studies. *J Infect Dis*. (2016) 214:S421–6. doi: 10.1093/infdis/jiw356
48. Lee EC, Asher JM, Goldlust S, Kraemer JD, Lawson AB, Bansal S. Mind the scales: harnessing spatial big data for infectious disease surveillance and inference. *J Infect Dis*. (2016) 214:S409–13. doi: 10.1093/infdis/jiw344
49. Generous N, Fairchild G, Deshpande A, Del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with wikipedia. *PLoS Comput Biol*. (2014) 10:e1003892. doi: 10.1371/journal.pcbi.1003892
50. Ma S, Yang S. Covid-19 forecasts using internet search information in the united states. *Sci Rep*. (2022) 12:11539. doi: 10.1038/s41598-022-15478-y
51. Zhang Y, Bambrick H, Mengersen K, Tong S, Hu W. Using google trends and ambient temperature to predict seasonal influenza outbreaks. *Environ Int*. (2018) 117:284–91. doi: 10.1016/j.envint.2018.05.016
52. Shaman J, Karspeck A, Yang W, Tamerius J, Lipsitch M. Real-time influenza forecasts during the 2012–2013 season. *Nat Commun*. (2013) 4:2873. doi: 10.1038/ncomms3837
53. Galić Z, Mešković E, Osmanović D. Distributed processing of big mobility data as spatio-temporal data streams. *Geoinformatica*. (2017) 21:263–91. doi: 10.1007/s10707-016-0264-z
54. Gomide J, Veloso A, Meira W Jr, Almeida V, Benevenuto F, Ferraz F, et al., editors. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In: *Proceedings of the 3rd International Web Science Conference*. Koblenz (2011). doi: 10.1145/2527031.2527049
55. Gao S, Mioc D, Anton F, Yi X, Coleman DJ. Online GIS services for mapping and sharing disease information. *Int J Health Geogr*. (2008) 7:8. doi: 10.1186/1476-072X-7-8
56. Hassan Zadeh A, Zolbanin HM, Sharda R, Delen D. Social media for nowcasting flu activity: spatio-temporal big data analysis. *Inform Syst Front*. (2019) 21:743–60. doi: 10.1007/s10796-018-9893-0



57. Internet World Stats. *Internet Users in the World by Regions-2017 Q2*. Internet World Stats. (2017). Available online at: <http://www.internetworldstats.com/stats.htm>
58. Lazer D, Kennedy R, King G, Vespignani A. The parable of google flu: traps in big data analysis. *Science*. (2014) 343:1203–5. doi: 10.1126/science.1248506
59. Butler D. When google got flu wrong. *Nature*. (2013) 494:155. doi: 10.1038/494155a
60. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS ONE*. (2011) 6:e23610. doi: 10.1371/journal.pone.0023610
61. Hamalaw SA, Bayati AH, Babakir-Mina M, Benvenuto D, Fabris S, Guarino M, et al. Assessment of core and support functions of the communicable disease surveillance system in the Kurdistan Region of Iraq. *J Med Virol*. (2022) 94:469–79. doi: 10.1002/jmv.27288
62. Kariithi H, Ferreira H, Welch C, Ateya L, Apopo A, Zoller R, et al. Surveillance and genetic characterization of virulent newcastle disease virus subgenotype V. 3 in indigenous chickens from backyard poultry farms and live bird markets in Kenya. *Viruses*. (2021) 13:103. doi: 10.3390/v13010103
63. Lu T, Reis BY. Internet search patterns reveal clinical course of COVID-19 disease progression and pandemic spread across 32 countries. *NPJ Digit Med*. (2021) 4:22. doi: 10.1038/s41746-021-00396-6
64. Seo Y, Choi K-H, Lee G. Characterization and trend of co-infection with neisseria gonorrhoeae and chlamydia trachomatis from the Korean national infectious diseases surveillance database. *World J Mens Health*. (2021) 39:107. doi: 10.5534/wjmh.190116
65. Beebeejaun K, Elston J, Oliver I, Ihueze A, Ukenedo C, Aruna O, et al. Evaluation of national event-based surveillance, Nigeria, 2016–2018. *Emerg Infect Dis*. (2021) 27:694. doi: 10.3201/eid2703.200141
66. Racloz V, Ramsey R, Tong S, Hu W. Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. *PLoS Negl Trop Dis*. (2012) 6:e1648. doi: 10.1371/journal.pntd.0001648
67. McMichael AJ. Environmental and social influences on emerging infectious diseases: past, present and future. *Philos Trans R Soc Lond B Biol Sci*. (2004) 359:1049–58. doi: 10.1098/rstb.2004.1480
68. Epstein PR. Climate change and emerging infectious diseases. *Microbes Infect*. (2001) 3:747–54. doi: 10.1016/S1286-4579(01)01429-0
69. Racloz V, Griot C, Stärk K. Sentinel surveillance systems with special focus on vector-borne diseases. *Anim Health Res Rev*. (2006) 7:71–9. doi: 10.1017/S1466252307001120
70. Eisen L, Eisen RJ. Using geographic information systems and decision support systems for the prediction, prevention, and control of vector-borne diseases. *Annu Rev Entomol*. (2011) 56:41–61. doi: 10.1146/annurev-ento-120709-144847
71. Cordeiro R, Donalizio MR, Andrade VR, Mafra AC, Nucci LB, Brown JC, et al. Spatial distribution of the risk of dengue fever in southeast Brazil, 2006–2007. *BMC Public Health*. (2011) 11:355. doi: 10.1186/1471-2458-11-355
72. Milinovich GJ, Avril SM, Clements AC, Brownstein JS, Tong S, Hu W. Using internet search queries for infectious disease surveillance: screening diseases for suitability. *BMC Infect Dis*. (2014) 14:690. doi: 10.1186/s12879-014-0690-1