

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

McClymont, H;Lambert, SB;Barr, I;Vardoulakis, S;Bambrick, H;Hu, W

Title:

Internet-based Surveillance Systems and Infectious Diseases Prediction: An Updated Review of the Last 10 Years and Lessons from the COVID-19 Pandemic

Date:

2024-09-01

Citation:

McClymont, H., Lambert, S. B., Barr, I., Vardoulakis, S., Bambrick, H. & Hu, W. (2024). Internet-based Surveillance Systems and Infectious Diseases Prediction: An Updated Review of the Last 10 Years and Lessons from the COVID-19 Pandemic. *Journal of Epidemiology and Global Health*, 14 (3), pp.645-657. <https://doi.org/10.1007/s44197-024-00272-y>.

Persistent Link:

<https://hdl.handle.net/11343/358488>

License:

CC BY



# Internet-based Surveillance Systems and Infectious Diseases Prediction: An Updated Review of the Last 10 Years and Lessons from the COVID-19 Pandemic

Hannah McClymont<sup>1</sup> · Stephen B. Lambert<sup>2,3</sup> · Ian Barr<sup>4,5</sup> · Sotiris Vardoulakis<sup>6,8</sup> · Hilary Bambrick<sup>7</sup> · Wenbiao Hu<sup>1,8</sup>

Received: 4 April 2024 / Accepted: 26 June 2024 / Published online: 14 August 2024  
© The Author(s) 2024

## Abstract

The last decade has seen major advances and growth in internet-based surveillance for infectious diseases through advanced computational capacity, growing adoption of smart devices, increased availability of Artificial Intelligence (AI), alongside environmental pressures including climate and land use change contributing to increased threat and spread of pandemics and emerging infectious diseases. With the increasing burden of infectious diseases and the COVID-19 pandemic, the need for developing novel technologies and integrating internet-based data approaches to improving infectious disease surveillance is greater than ever. In this systematic review, we searched the scientific literature for research on internet-based or digital surveillance for influenza, dengue fever and COVID-19 from 2013 to 2023. We have provided an overview of recent internet-based surveillance research for emerging infectious diseases (EID), describing changes in the digital landscape, with recommendations for future research directed at public health policymakers, healthcare providers, and government health departments to enhance traditional surveillance for detecting, monitoring, reporting, and responding to influenza, dengue, and COVID-19.

**Keywords** Early Warning Systems · Internet-based Surveillance · COVID-19 · Influenza · Dengue · mHealth

✉ Wenbiao Hu  
w2.hu@qut.edu.au

- <sup>1</sup> Ecosystem Change and Population Health (ECAPH) Research Group, School of Public Health and Social Work, Queensland University of Technology (QUT), Brisbane, Australia
- <sup>2</sup> Communicable Diseases Branch, Queensland Health, Brisbane, Australia
- <sup>3</sup> National Centre for Immunisation Research and Surveillance, Sydney Children's Hospitals Network, Westmead, Australia
- <sup>4</sup> WHO Collaborating Centre for Reference and Research on Influenza, The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia
- <sup>5</sup> Department of Microbiology and Immunology, University of Melbourne, Melbourne, Australia
- <sup>6</sup> Health Research Institute, University of Canberra, Canberra, Australia
- <sup>7</sup> National Centre for Epidemiology and Population Health, College of Health and Medicine, The Australian National University, Canberra, Australia
- <sup>8</sup> Healthy Environments and Lives (HEAL) National Research Network, Canberra, Australia

## 1 Introduction

As we move forward managing COVID-19 alongside other communicable diseases, following the World Health Organization declaring the end of the global health emergency in May 2023 [1]. Widespread public health measures have now ended [2] as the immediate threat of COVID-19 has faded following a successful global vaccination effort and the occurrence of less severe variants circulating [3] although the aftermath of the pandemic including loss of life, and economic and societal upheaval endures and will do so for many years to come. COVID-19 is now considered endemic in humans alongside seasonal influenza and respiratory diseases such as respiratory syncytial virus (RSV), human metapneumovirus (HMPV) and others, while the risk of more virulent or vaccine-avoidant variants of SARS-CoV-2 or other viruses causing severe outbreaks in the future remains [4].

In recent years, with the increasing effects of climate change and other environmental and land use changes, the threat of emerging infectious diseases (EIDs) from spillover

events has grown amidst warnings from experts [5]. Over the last decade [6], recent epidemics/pandemics caused by zoonotic spillover events have occurred [7], Avian influenza (H7N9) (2013–17), Swine flu (H1N1) (2009–10) alongside increasing occurrence and range of mosquito-borne diseases including Japanese Encephalitis virus (Australia 2022), Zika virus (2015–2016), and Dengue fever (continuous and ongoing in many countries) [8]. The emergence of the novel coronavirus in December 2019 leading to the COVID-19 pandemic has led to a renewed interest in disease surveillance and outbreak tracking with the need for integrative digital early warning and surveillance systems more pronounced than ever [9].

The COVID-19 pandemic caused significant disruption to global health, with widespread impacts on existing public health measures, vaccination and prevention programs and community surveillance programs for many diseases. In particular, pandemic response measures for the control of COVID-19 including non-pharmaceutical interventions (i.e., hygiene measures, masking, social distancing) [10] and travel restrictions (border closures and movement restrictions) contributed to decreased influenza circulation and other respiratory disease activity during this time [11–13]. These actions have resulted in the global disappearance of Influenza B/Yamagata strain [14], and altered patterns of dengue outbreaks in non-endemic countries [15], even as overall annual incidence and distribution have increased [16].

While research interest in digital surveillance has remained high over the previous decade, particularly for influenza outbreak detection [17], the COVID-19 pandemic was an unprecedented opportunity to employ methods in real-time to detect outbreaks, forecast epidemic growth and tailor effective and locally relevant public health messaging [18, 19]. Internet-based or digital surveillance systems [20], use online data sources to detect digital signals for potential indicators or early signs of infectious disease outbreaks based on online information seeking and trends in user behaviours from a range of social media and search engine sources. Predictive modelling and forecasting use digital signals to estimate the risk of an outbreak, rate of transmission or forecast the spread of disease. By analysing large volumes of online data in real-time, predictive models can be used to identify high-risk clusters, trends and early warning signs often preceding traditional surveillance methods for disease detection (i.e., laboratory-confirmed testing or diagnosis in a healthcare setting) and can provide early warning of outbreaks prior to these health system alerts, and are complementary to event-based electronic surveillance systems such as GPHIN and ProMED [21]. Over the last decade, there have been many changes in the online ecosystem, with emerging social media platforms, changing user

behaviours, and the emergence of AI chatbots integrated into search engines and social media, blurring the lines between sources of information. With the emergence of and widespread implementation of Artificial Intelligence (AI), as described in recent reviews including Brownstein et al. [22], and Macintyre et al. [23], the potential applications in the digital space continue to grow, alongside increased computing capability and signal detection to improve the speed and capacity of existing EWS and surveillance, enabling earlier detection to manage serious epidemics and pandemics in the future [24].

## 1.1 Objectives

In this review, we evaluated studies from the past decade (2013–2024) to capture changing trends in digital surveillance for influenza, COVID-19 and dengue, as representative of broader respiratory and vector-borne diseases with high levels of surveillance. We describe the changes over time in digital surveillance, and forecasting for selected infectious diseases since our previous review [20], the advantages and limitations of using digital surveillance data, and advances in AI and digital technology. Due to the increasing range of social media and search engine data sources and increased integration of multiple data sources the included studies have been grouped by disease of interest rather than data source, these are summarized in Table S2. Finally, we make recommendations for future research into digital surveillance for useful early warning systems (EWS).

## 2 Methods

Using a systematic review approach, we searched PubMed and Scopus databases for peer-reviewed original research publications between July 1, 2013, and March 31, 2024, according to PRISMA 2020 guidelines (See Supplementary Table S1) [25]. Additional relevant publications were identified from references.

### 2.1 Search Strategy and Selection Criteria

We performed searches with the following terms: “dengue” OR COVID-19 OR “influenza” AND “early warning”, “Google”, “Google Trends”, “internet”, “search engine”, “social media”, “Twitter”, “Facebook”, OR “digital disease detection”, “infodemiology”, “infoveillance”, “real-time disease surveillance”, and “syndromic surveillance”. To be eligible for inclusion, studies needed to be peer-reviewed and describe the use of internet-based data for surveillance, predictive modelling, forecasting or early warning for influenza, COVID-19, or dengue. Studies were excluded

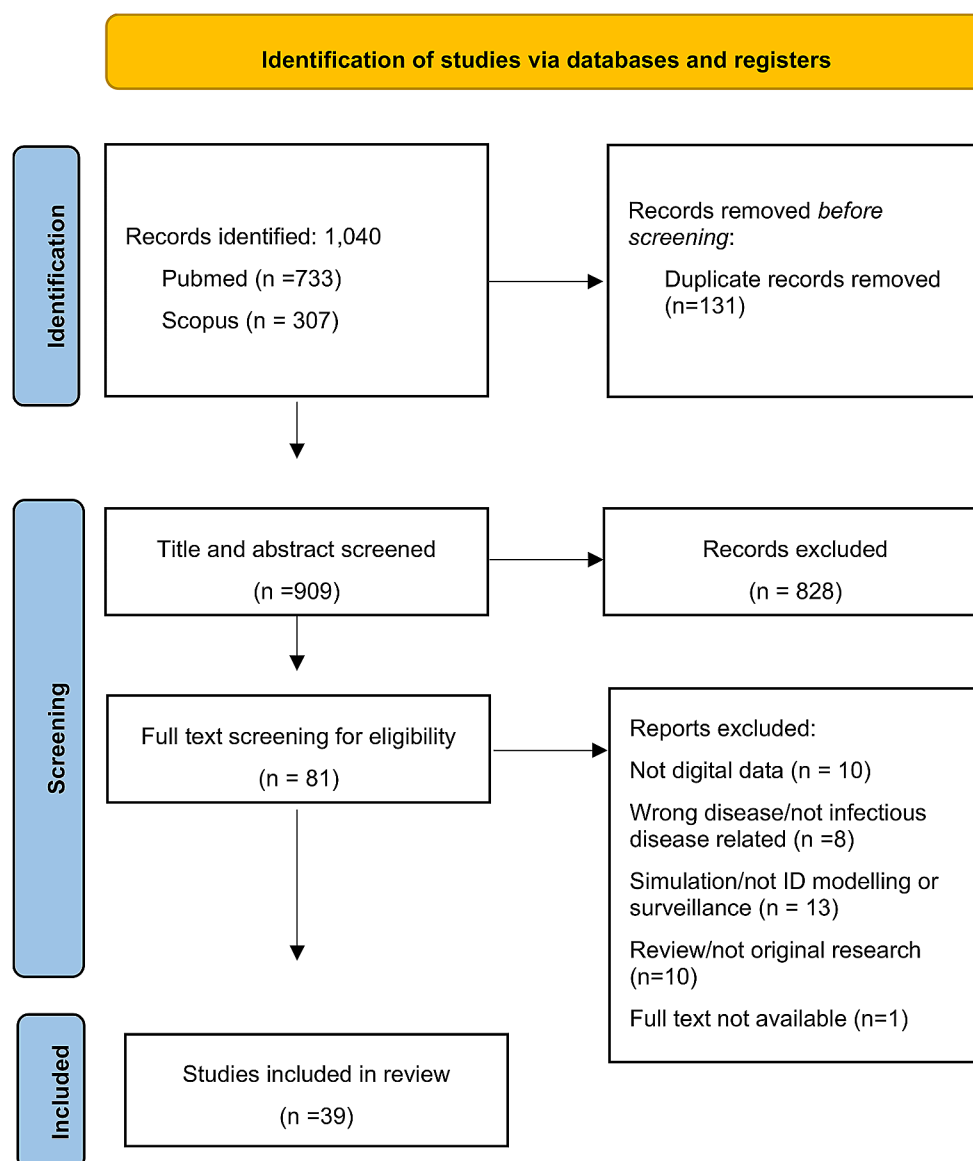
if they were not original research, did not include digital data sources (social media or search trends), or did not discuss influenza, COVID-19, or dengue. Mathematical and computational modelling studies using simulated data were excluded. Due to the evolving nature of AI in its current form and limited access to user data, no studies were published that the authors are aware of at this time. The data for full-text screened articles were extracted and summarized See Supplementary Table S2. Due to varying study designs, methodologies, models, statistical analysis and potential confounders, no meta-analysis was performed.

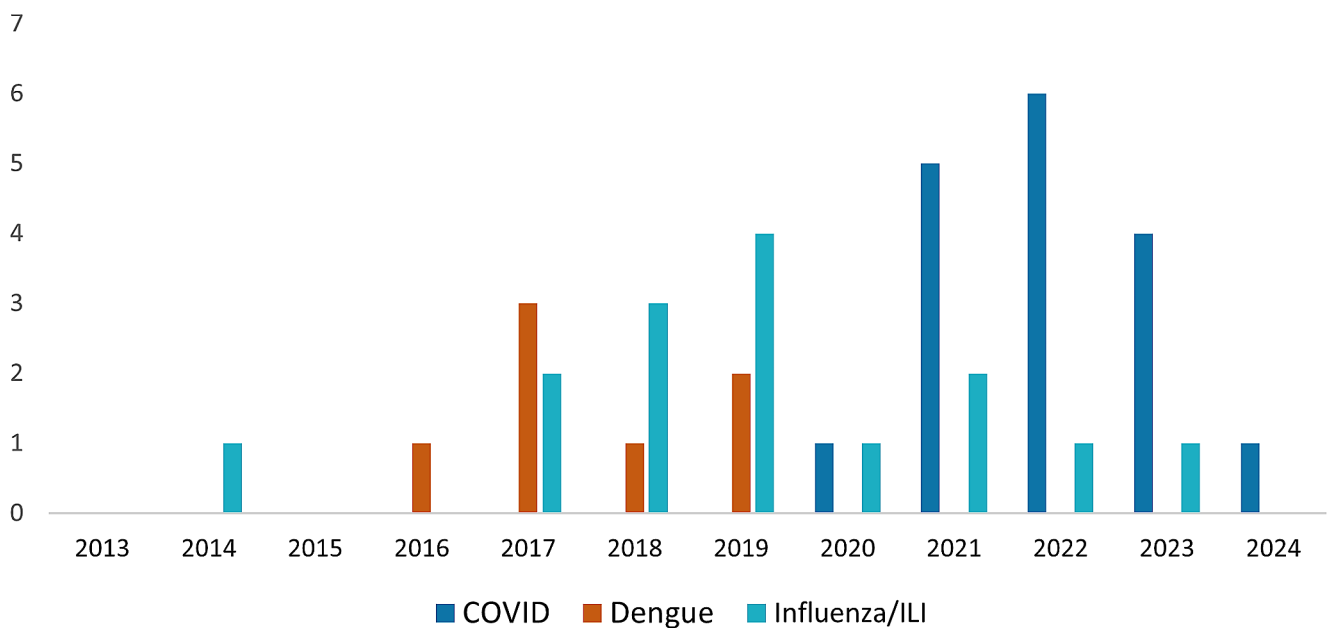
### 3 Results

A total of 1040 studies were identified through the literature search and reference checking, and 131 duplicate records were excluded (Fig. 1). The remaining 909 papers were assessed for eligibility and screened by checking the title and abstract for relevance. Subsequently, 828 papers were excluded, leaving 81 papers for full-text review where references were checked for additional relevant papers, 43 full-text papers were excluded.

Of the 39 selected studies, 17 focused on COVID-19, 15 studies described influenza or influenza-like illnesses (ILI), and 7 studies described dengue fever (Fig. 2). Digital data sources included Google Search Trends and Community Mobility, Apple Mobility, Baidu Search (the main search engine used in China), Bing Search, Wikipedia, and social media including X (formerly known as Twitter), Weibo and

**Fig. 1** Flowchart with article selection





**Fig. 2** Number of selected articles by year of publication and disease of interest from PubMed and Web of Science using digital big data for early warning, surveillance and predictive modelling (Data up to March 2024)

WeChat in China (equivalent to Twitter/X and WhatsApp). Models and methodologies varied ranging from simple correlation (Pearson's or Spearman's Rank Correlation) and Time Series Cross Correlation to spatial and temporal models including Poisson linear regression, generalized linear models, predictive models including SARIMA and ARIMA, LSTM, Prophet and SVR through to machine learning, NLP and neural networks. All these models have varying strengths and limitations, certain models may be more appropriate in specific situations, particularly when considering computational complexity for low resource settings, further discussion on modelling for EWS can be found in Haque et al. [26].

### 3.1 Disease Specific Internet-Based Surveillance

#### 3.1.1 Influenza and ILI Surveillance

This review identified research using search queries for surveillance and early warning for influenza/ILI from Mexico [27], the United States [28, 29], Australia [30], Hong Kong [31], South Africa [32], and a multi-country study [33]. These studies reported Google search results with climate and weather variables included in models enhanced predictive accuracy and emerging outbreaks and seasonal variations for ILIs and respiratory diseases were detected weeks earlier compared with conventional surveillance [34]. Studies from China utilised Baidu search, Guo et al. proposed a surveillance framework based on significant keywords [35], Chen et al. [36] used seasonal SARIMA models with

Baidu search ( $\beta=0.008$ ,  $p<0.001$ ) and Weibo ( $\beta=0.002$ ,  $p=0.036$ ) for search terms “H7N9”, “avian influenza”, and “live poultry” to explore the association with weekly laboratory-confirmed H7N9 cases. Yang et al. [37] developed deep learning prediction models for ILI, reporting that integrating both climate factors and search trends enhanced model accuracy. Additionally, Google and Baidu search queries were used to forecast seasonal influenza outbreaks cross-hemisphere for the United States, United Kingdom, and China using Australian Influenza surveillance data. The resulting SARIMA models demonstrated high correlation coefficients (China=0.96, the US=0.97, the UK=0.96,  $p<0.01$ ) and low Maximum Absolute Percent Error (MAPE) values (China=16.76, the US=96.97, the UK=125.42), significantly improving predictive accuracy over case-only models [38].

Beyond search queries, text mining of symptom keywords on social media enabled near real-time syndromic surveillance for flu/unwell in Australia [39], identifying changes in frequency counts compared to public health notifications. Natural Language Processing was used to detect avian influenza notifications, by processing Tweets, 75% of official outbreak notifications (i.e. farm records, outbreaks and individual cases) were identified from the sample dataset, and a third of these detections were identified earlier than official notifications [40]. Using Twitter/X geolocation data, Nagar et al [41]. used tweet vector maps to identify clusters of ILI in New York providing insight into spatio-temporal patterns of ILI. Another study [42] found there was a strong temporal association with flu-related Tweet

activity for healthcare seeking behaviour preceding official reports of influenza cases by up to one month, and identified hotspots related to public spaces.

### 3.1.2 Dengue Fever (DF) Surveillance

While there were fewer studies on internet-based surveillance for DF compared with Influenza and COVID-19, the magnitude of dengue outbreaks significantly increased in 2023, compared with the previous periods from 2018 to 2022. This may potentially be attributable to public health interventions during the pandemic response impacting transmission and testing availability and research focus on COVID-19 over this time [43]. These studies covered China, Brazil, the Philippines, and Indonesia. In China, researchers used Baidu Search trends for dengue to determine thresholds to detect outbreaks. Weekly search indexes showed a positive correlation with incidence rates, and a lagged moving average of 1–3 weeks greater than 99.3, indicated there was an 89.28% chance of an outbreak in Guangzhou. In Zhongshan, weekly BSI at 1–5 weeks was over 68.1 with the chance of an outbreak increased by 100% [44]. In Guangdong province, Guo and co-authors developed a forecasting model using Baidu search queries and weather factors, with support vector regression (SVR) model consistently demonstrating lower prediction error rates [45].

Liu et al. reported increased predictive accuracy of models using weather, Baidu search and demographics in Guangzhou city [46]. Li et al. [47] found that dengue-related searches at a lag of one week were positively correlated with DF occurrence. The model including search indexes had greater predictive capability (ICC:0.94, RMSE:59.86) compared to the model without search data (ICC:0.72, RMSE:203.29). Ho et al. explored temporal and spatial patterns of dengue incidence in Manila, Philippines, using Google Dengue Trends (GDT). Weekly values of DF incidence were moderately associated ( $r=0.405$ ) with weekly GDT values, while spatial analysis was not significant ( $r=0.223$ ,  $p=0.283$ ) [48]. In Indonesia, Husnayain et al. reported a significant correlation with Google search terms for dengue symptom, dengue and dbd (dengue abbreviation) showing the highest correlation one week preceding;  $r=0.937$ ,  $0.931$  and  $0.921$  respectively ( $p\leq 0.05$ ) [49].

Finally, Marques-Toledo and co-authors explored multiple digital data sources for estimating and forecasting DF at the city level in Brazil. They accessed Google Search, X and Wikipedia to explore real-time health-seeking behaviour for DF. The authors reported a positive linear association with Tweets ( $r=0.87$ ,  $p<0.001$ ), Google Trends ( $r=0.92$ ,  $p<0.001$ ), and Wikipedia ( $r=0.71$ ,  $p<0.01$ ). Tweets selected by city were used to develop nowcasting and forecasting models, demonstrating temporally association with

DF up to 8 weeks in advance, with the strongest association at lag week 1 [50].

### 3.1.3 COVID-19 Pandemic

From the initial detection of COVID-19 in December 2019 to the present, the novel pandemic situation received significant attention and resulted in extensive epidemiological research. Early research concentrated on the initial outbreak and the first wave in China, using search queries (Baidu) and social media/microblogs (Weibo, WeChat and Douyin/TikTok) to detect early signals of the emerging pandemic as people searched for the latest news and updates on escalating outbreaks [51].

Studies using Baidu search data found search terms associated with COVID-19 valuable for early outbreak warning. Tu et al. reported an average “search to confirmed interval” of 19.8 days, with optimal time lags for search queries at 0–4 days [52]. Li et al. reported significant lags at 4–7 days preceding conventional surveillance identification, with early warning signs up to 20 days earlier than lockdown policy implementation [53]. Integrating multiple data sources enhanced predictive accuracy and early warning capabilities. Gong et al. compared search interest and microblogs for daily new cases, new deaths and outbreak severity, revealing advanced trends between lags of 3–16 days for both Baidu and Weibo [54]. Baidu and Weibo showed a significant positive correlation with cases and deaths, with Baidu search having a stronger correlation.

Weibo keyword trends for symptoms and diagnosis were useful for detecting early signals of COVID-19. Guo et al. used Weibo data to improve the predictive accuracy of early epidemic models and Shen et al. used social media data to accurately identify early digital disease signals [55, 56]. Li et al. combined diverse internet data, including online news articles, microblogs and search trends to improve model forecasting accuracy providing early warning signals for outbreaks 2 to 6 days in advance [57]. Gao et al. utilized search queries and video-based social media Douyin keyword trends to predict asymptomatic or undetected transmission [58], capturing data from younger subpopulations.

As COVID-19 spread globally, international travel restrictions and public health measures were implemented following the WHO pandemic declaration [59]. The number of studies using digital signals grew quickly, with an English study using Bing search to detect early warning of COVID-19 [60], and X data to explore symptom keywords for early detection of COVID-19 in Europe and globally [61, 62]. During this time, there was an increasing use of multisource digital data, combining search trends and social media data. Twitter/X data proved useful for providing early warning of outbreaks in studies from the United States and



Canada [63–65]. Symptom-based keyword searches for both Google search and X were temporally correlated, with Google searches for “cough,” “runny nose,” and “anosmia” correlated with COVID-19 incidence and peaked 9, 11, and 3 days earlier than the incidence peak, respectively. This improved the predictive accuracy of LSTM forecasting models ( $MSE = 124.78$ ,  $R^2 = 0.88$ ) [66]. In California, Habibdoust *et al* used search volumes for “Fever,” “COVID Testing,” “Signs of COVID,” “COVID Treatment,” and “Shortness of Breath” to predict daily incidence comparing GMDH-type neural network and LSTM models over three time periods [67], where models with queries improved predictive accuracy by as much as 22.6%, 21%, and 37.3% in NRMSE across the different study periods.

In our study from Victoria, Australia, we used Google Mobility as a proxy for population movement and non-pharmaceutical interventions on COVID-19 transmission, integrating search trends and weather factors to forecast epidemic growth [68]. The multivariable weather and mobility model demonstrated the highest predictive accuracy ( $R^2 = 0.948$ ,  $RMSE = 137.57$ ,  $MAPE = 21.26$ ) compared to cases only ( $R^2 = 0.942$ ,  $RMSE = 141.59$ ,  $MAPE = 23.19$ ). Finally, Kogan *et al.* developed an integrated EWS for detecting ILI globally, monitoring COVID-19 activity using multiple digital sources including Google search trends, Apple Mobility, Twitter/X API with ILINet (CDC sentinel system) and UpToDate physician search trends and smart thermometer data and found digital proxies for COVID-19 preceded detection through normal clinical surveillance [69].

## 4 Discussion

### 4.1 Brief Overview of EWS

Within the broader scheme of global outbreak detection and pandemic preparedness, timely disease detection and notifications are limited by aging infrastructure and reduced public health funding leading to downgrading or discontinuation of existing early warning surveillance systems. In the lead-up to the detection of atypical pneumonia and official reports to the WHO in December 2019–January 2020, the US dissolved the National Security Council (NSC) Directorate for Global Health Security and Biodefense in May 2018, responsible for monitoring global health risk and coordinating government response [70]. As the oldest digital surveillance system, ProMED-mail has provided reliable event-based surveillance for over 25 years [71], and successfully detected the first alert of COVID-19 in December 2019. Despite playing an important role in global public health surveillance, as with many pandemic preparedness

tools, ProMED-mail suffers from funding shortages reducing the availability and capacity for detecting future outbreaks. Another well-known digital surveillance system, the Global Public Health Intelligence Network (GPHIN) [72], an event-based EWS developed by the Canadian Government to collect global and multilingual media reports to create alerts in the wake of SARS. Having successfully detected early signs of outbreaks including the Middle East Respiratory Syndrome (MERS), influenza pandemics, and Zika virus, leading up to the COVID-19 detection, GPHIN alerts were increasingly limited by outdated system capabilities and downgraded reporting and response.

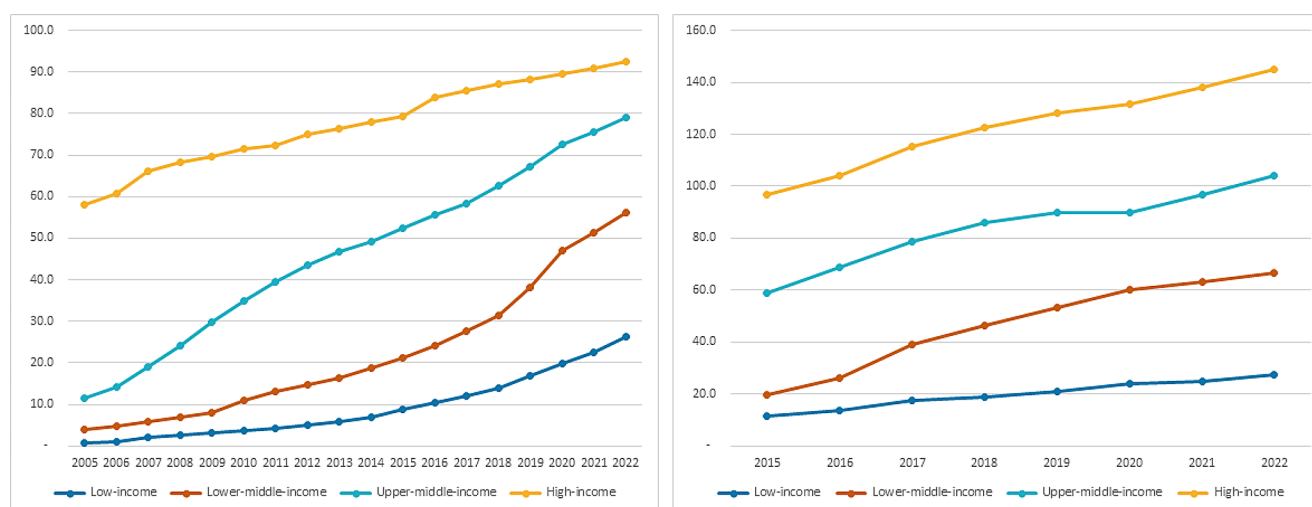
### 4.2 Emerging Trends in Digital Surveillance from the COVID-19 Pandemic

In the last decade, research consistently supports the use of internet-based data for infectious disease surveillance, with continued research using these methods, though more recent studies often utilised multi-source digital data *i.e.*, both search queries and social media and an expanded range of sources. This approach offers advantages in capturing information from difficult-to-reach sources beyond the medical system, including individuals seeking testing, accessing health resources, or being hospitalized. Digital surveillance excels at detecting early signals from recently exposed individuals, those with milder disease states, and younger demographics with limited access to regular healthcare or a lower inclination to seek care. Digital data provides near real-time availability, facilitates discussions about symptoms or keywords with geotagging functionality, and links online behaviours to unique user accounts and online networks, potentially reflecting real-world connections like family, friends, and co-workers.

Over the last decade, there have been many technological advances affecting internet usage and digital health care seeking. Levels of internet access in the home have increased alongside significant growth in smartphone usage since 2013, numbers of which are higher in advanced economies compared to regions with higher rates of poverty (see Fig. 3). Along with the proliferation of smart devices or the Internet of Things (IoT), GPS-enabled and WIFI-connected wearables and mobile devices are widespread and able to collect biometric, audio and location data in real-time potentially able to diagnose IDs in pre-symptomatic infected individuals through biosignals or distinctive COVID-19 or pneumonia coughing [73].

#### 4.2.1 Changing Digital Landscape and Data Availability

Along with the continuous 24-hour news cycle and inundation of health information over a range of platforms; the



**Fig. 3** Individuals using the Internet per 100 population (Left) Active mobile-broadband subscriptions per 100 population (Right) (Data not available before 2015) by global development status. (Data Source: <https://datahub.itu.int/>). (Source: ITU)

way users consume and interact with media for healthcare-seeking and news has shifted over the last decade. Digital usage trends since the early 2020 pandemic lockdowns contributed to changing online behaviours and increased digital communication. While social media is a useful medium for targeted public health messaging where users are increasingly accessing news and health advice, information from official sources appears alongside content containing misinformation and disinformation [74]. With the emergence of AI, unlimited access by LLM for training data has led to restrictions on data from many of the previously freely available data sources. Over time the scope and availability of aggregate data sets have changed, these changes will significantly impact digital research, affecting applications in short- and long-term trend analysis for pandemic detection, disaster responses during social unrest, and extreme weather events [75].

While the potential applications for AI are evolving and changing rapidly, recent changes to integrate AI chatbots into search engines and social media i.e., Microsoft Bing and Google search engines powered by AI chatbots and Meta AI and X-AI on social media platforms are shaping online user behaviour [76], meaning search results and social media algorithms are trained on user behaviour aiming to provide faster, more accurate and relevant answers based on user profiles and behaviour across the online ecosystem [77]. Using AI trained on user-provided information across the digital landscape to answer user queries rather than directing traffic to websites will contribute to changes in the consumption of social media as a news source and information seeking, where users are increasingly searching and accessing news and information from social media sources, i.e., health and government information releases and media coverage in real time, meaning the distinction

between digital signals from search engines and posting about symptoms on social media are increasingly blurring.

#### 4.2.2 Integrating AI Technology and Advances in Digital Data

Emerging technological advances are accelerating hand in hand with changing usage trends, with potential uses in future research and applications in healthcare quickly evolving [78]. The growth of AI and machine learning algorithms and increased computing capability allow for processing more complex and larger data sets, faster data mining, and improved capacity for predictive models. With increased capability to deal with high volumes of text-based communications including surveillance reports and news coverage for identifying and classifying infectious disease signals and detecting epidemics [79].

Evaluating the usefulness of digital data sources is essential where some sources may contain a greater amount of noise and, positive signals can overwhelm the capacity of a system to recognise and respond to events in real-time. Ensuring data fidelity where data is captured accurately, with precision and timeliness, is essential [80]. This can be achieved through a range of methods including real-time monitoring and constant sampling to improve the detection of data signals and show complex patterns and changes over time can improve signal detection accuracy and are less prone to noise and artifacts in the data [81]. Methodological approaches including multisource data and improved machine learning and neural networks can improve data fidelity, alongside real-world reporting and test results to validate and fine-tune models for best results, and is an important consideration in future research.



How users are interacting with AI chatbots, as an alternative to internet searches for self-diagnosis, may have the potential to transform healthcare, by capturing the healthcare-seeking actions of internet users to detect early warning signals [82]. Applications include the use of chatbots trained on medical research for answering consumer health questions as natural language alternatives to conventional keyword-based search methods for healthcare seeking [83, 84], as mHealth (mobile health) applications with virtual educator interface for providing health information [85], diagnostic purposes and assessing illness severity.

### 4.3 Developing Integrative Multisource Early Warning Systems

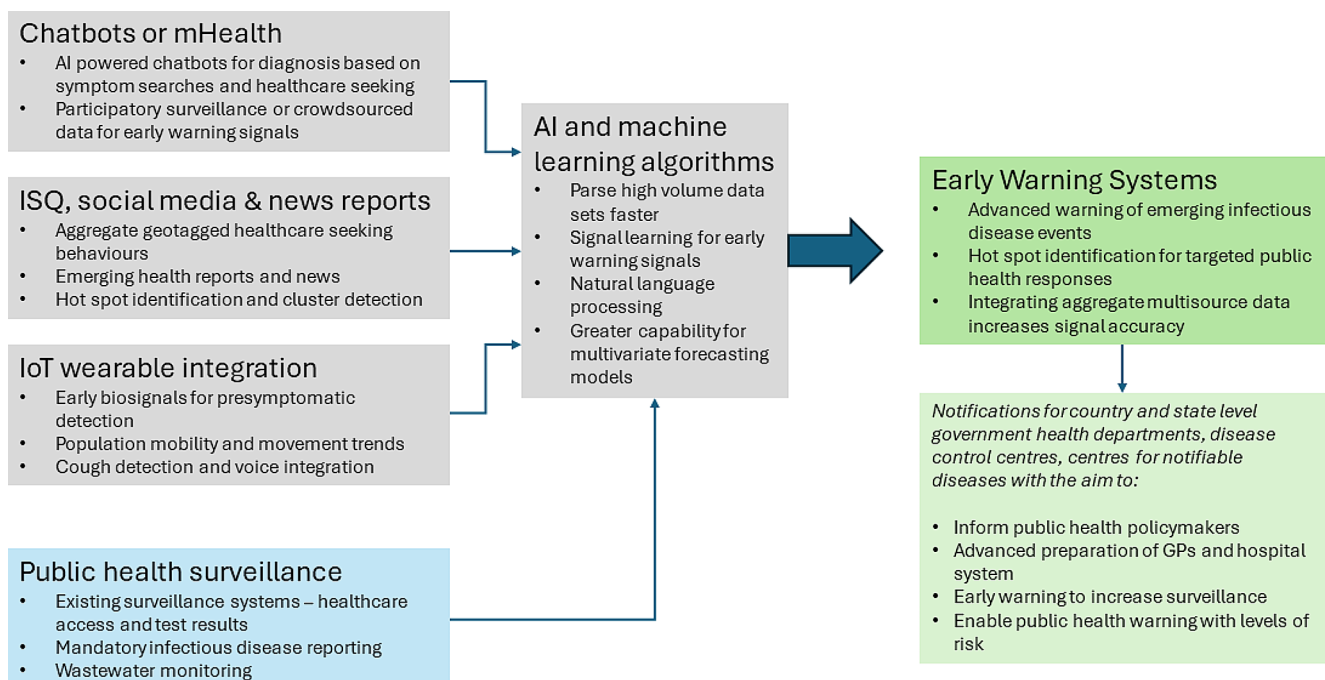
Developing integrated multisource EWS holds significant potential for improving the early detection of disease signals and identifying emerging outbreaks, particularly for climate and weather-sensitive vector-borne or respiratory diseases. This is particularly important in regions experiencing increasing climate stress, at greater risk of spillover events, and where the population may experience high levels of poverty or limited access to healthcare infrastructure. Our previous research has emphasised the usefulness of incorporating socio-environmental factors including weather with internet-based data across a range of diseases [86, 87]. Additionally, utilizing Geographical Information Systems (GIS) at high spatial resolution for climate-sensitive

or vector-borne diseases [88], has the potential to enhance predictive modelling and early warning for outbreaks [89].

Using existing surveillance and wastewater monitoring, mHealth applications with AI chatbots, internet-based data and socio-environmental factors can be useful for the timely detection and notification of emerging outbreaks, identifying spatiotemporal variations in outbreaks or hot spots or high-risk clusters. An EWS would facilitate timely public health notifications to the public and assist health staff and policymakers with broader applications in detecting mass gatherings and emergencies (See Fig. 4). Some related examples of this include EpiWatch [79], an AI-based system providing early warnings of epidemics on a global scale based on open data, and HealthMap (<https://www.health-map.org>), using natural language processing and Bayesian machine-learning classification trained to identify relevant information from digital data and PROMED-mail [71] alerts to identify infectious disease outbreaks.

### 4.4 Advantages and Limitations of Digital Surveillance

Digital data for infectious disease surveillance remains an important area of research. This approach offers advantages in capturing information from hard-to-reach sources outside the medical system i.e., people seeking testing, accessing health resources or hospitalised. Digital surveillance is useful for detecting signals from recently exposed individuals, those experiencing milder disease states, and younger



**Fig. 4** Integrating multisource internet-based data with existing surveillance methods using AI and machine learning for improving early warning of infectious diseases

demographics with limited access to regular healthcare or less likely to seek care. Digital data has near real-time availability, enables discussion of symptoms or keywords with geotagging functionality, and connects online behaviours to unique user accounts and networks, potentially representing real-world connections with family, friends, and co-workers [90]. However, the digital divide among social media users based on age may lead to under-detection for disease surveillance purposes.

While internet-based surveillance holds significant potential for EWS, there are limitations to consider. Although internet access has improved, coverage remains limited in LMIC, reducing early signal detections for infectious disease surveillance, particularly in rural or remote areas [91]. Post-COVID changes in laboratory diagnostics and respiratory disease surveillance may impact the future utility of these methods. Spatial resolution is limited [92], with limited availability of city-level data. Data sets are often discontinued or monetised, measurement indexes may have accuracy issues over time, potentially impacting research reproducibility across various geospatial contexts [93].

As Internet usage has increased, so has the availability of digital data. However, interpreting signals and predictive values with appropriate sensitivity and specificity, and distinguishing true signals from false positives or noise, remains a significant challenge. Moreover, as internet usage has increased and data linkage raises ethical concerns about accessing and using personal health information [94, 95] with potential risks for user safety and privacy, care must be taken to ensure patient and individual privacy and personal data are protected. Finally, the current generation of AI chatbots have a tendency to ‘hallucinate,’ [96] providing incorrect or nonsensical responses which is particularly dangerous in healthcare scenarios.

#### 4.5 Future Research

Future research must focus on a better understanding of digital health signals and digital surveillance, the growing use of AI, and how these are changing the online environment in terms of healthcare information-seeking and collation of web-available data. Finally, an informative and useful EWS based on an integrated framework including conventional surveillance data, crowdsourced surveillance, and a wide range of internet-based surveillance sources to build the capacity of existing surveillance for the future requires global cooperation for information and resource sharing.

In summary, for improving infectious disease early warning surveillance we recommend:

1. Development of novel dynamic GIS-based spatiotemporal models to link infectious diseases with internet-based data and social-environmental data.
2. Integration of internet-based models with social-environmental data to produce infectious disease surveillance systems able to better identify vulnerable/susceptible communities over space and time.
3. Developing innovative mHealth applications with a chatbot and virtual educator interfaces.
4. Expand the knowledge of big data utility in infectious disease early warning systems and development of functional infectious disease early warning systems, minimising noise and ensuring real signals are identified early and accurately without overwhelming existing system capacity.

## 5 Conclusion

With increasing internet access globally, the world is more connected digitally than ever before as the availability and range of digital data sources available for disease surveillance have grown. Our increased interconnectedness and globalisation as COVID-19 swept through the world underlined the importance of consistent global surveillance, for the detection and reporting of such events. The next 10 years will be a challenging period, and it remains to be seen if we can indeed achieve the holy grail of being able to forecast existing or newly emerging diseases with accuracy and early enough to allow measures to be enacted that will mitigate or halt these outbreaks.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s44197-024-00272-y>.

**Acknowledgements** HM is supported by the Australian Government RTP Scholarship. SV and WH acknowledge support from the HEAL Network from the National Health and Medical Research Council Special Initiative in Human Health and Environmental Change (Grant no. 2008937) and National Foundation for Australia-China Relations (Grant no. 220011).

**Author Contributions** HM: Literature search and extraction, Writing—Original Draft, Review and Editing, Figure Preparation. SL IB HB SV: Writing—Review and Editing. WH: Conceptualisation, Methodology, Writing—Review & Editing, Supervision.

**Funding** This research did not receive any specific funding.

**Data Availability** No datasets were generated or analysed during the current study.

## Declarations

**Ethics Approval and Consent to Participate** N/A.

**Consent for Publication** N/A.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Wise J. Covid-19: WHO declares end of global health emergency. *BMJ*. 2023;381:1041. <https://doi.org/10.1136/bmj.p1041>.
2. Murray CJL. COVID-19 will continue but the end of the pandemic is near. *Lancet*. 2022;399(10323):417–9. [https://doi.org/10.1016/S0140-6736\(22\)00100-3](https://doi.org/10.1016/S0140-6736(22)00100-3).
3. Wrenn JO, Pakala SB, Vestal G, Shilts MH, Brown HM, Bowen SM, et al. COVID-19 severity from Omicron and Delta SARS-CoV-2 variants. *Influenza Other Respir Viruses*. 2022;16(5):832–6. <https://doi.org/10.1111/irv.12982>.
4. El-Sadr WM, Vasan A, El-Mohandes A. Facing the New Covid-19 reality. *N Engl J Med*. 2023;388(5):385–7. <https://doi.org/10.1056/NEJMp2213920>.
5. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. *Nat Rev Microbiol*. 2022;20(4):193–205. <https://doi.org/10.1038/s41579-021-00639-z>.
6. Morens DM, Fauci AS. Emerging Pandemic diseases: how we got to COVID-19. *Cell*. 2020;182(5):1077–92. <https://doi.org/10.1016/j.cell.2020.08.021>.
7. Carlson CJ, Albery GF, Merow C, Trisos CH, Zipfel CM, Eskew EA, et al. Climate change increases cross-species viral transmission risk. *Nature*. 2022;607(7919):555–62. <https://doi.org/10.1038/s41586-022-04788-w>.
8. Rocklöv J, Dubrow R. Climate change: an enduring challenge for vector-borne disease prevention and control. *Nat Immunol*. 2020;21(5):479–83. <https://doi.org/10.1038/s41590-020-0648-y>.
9. McClymont H, Bambrick H, Si X, Vardoulakis S, Hu W. Future perspectives of emerging infectious diseases control: a One Health approach. *One Health*. 2022;14:100371. <https://doi.org/10.1016/j.onehlt.2022.100371>.
10. Cot C, Cacciapaglia G, Sannino F. Mining Google and Apple mobility data: temporal anatomy for COVID-19 social distancing. *Sci Rep*. 2021;11(1):4150. <https://doi.org/10.1038/s41598-021-83441-4>.
11. Sullivan SG, Carlson S, Cheng AC, Chilver MB, Dwyer DE, Irwin M, et al. Where has all the influenza gone? The impact of COVID-19 on the circulation of influenza and other respiratory viruses, Australia, March to September 2020. *Euro Surveill*. 2020;25(47):2001847. <https://doi.org/10.2807/1560-7917.ES.2020.25.47.2001847>.
12. Zipfel CM, Colizza V, Bansal S. The missing season: the impacts of the COVID-19 pandemic on influenza. *Vaccine*. 2021;39(28):3645–8. <https://doi.org/10.1016/j.vaccine.2021.05.049>.
13. El-Heneidy A, Ware RS, Robson JM, Cherian SG, Lambert SB, Grimwood K. Respiratory virus detection during the COVID-19 pandemic in Queensland, Australia. *Aust N Z J Public Health*. 2022;46(1):10–5. <https://doi.org/10.1111/1753-6405.13168>.
14. Koutsakos M, Wheatley AK, Laurie K, Kent SJ, Rockman S. Influenza lineage extinction during the COVID-19 pandemic? *Nat Rev Microbiol*. 2021;19(12):741–2. <https://doi.org/10.1038/s41579-021-00642-4>.
15. Lu X, Bambrick H, Pongsumpun P, Dhewantara PW, Toan DTT, Hu W. Dengue outbreaks in the COVID-19 era: Alarm raised for Asia. *PLoS Negl Trop Dis*. 2021;15(10):e0009778. <https://doi.org/10.1371/journal.pntd.0009778>.
16. Ebi KL, Nealon J. Dengue in a changing climate. *Environ Res*. 2016;151:115–23. <https://doi.org/10.1016/j.envres.2016.07.026>.
17. Simonsen L, Gog JR, Olson D, Viboud C. Infectious Disease Surveillance in the Big Data era: towards faster and locally relevant Systems. *J Infect Dis*. 2016;214(suppl4):S380–5. <https://doi.org/10.1093/infdis/jiw376>.
18. Morgan OW, Abdelmalik P, Perez-Gutierrez E, Fall IS, Kato M, Hamblion E, et al. How better pandemic and epidemic intelligence will prepare the world for future threats. *Nat Med*. 2022;28(8):1526–8. <https://doi.org/10.1038/s41591-022-01900-5>.
19. Morgan OW, Aguilera X, Ammon A, Amuasi J, Fall IS, Frieden T, et al. Disease surveillance for the COVID-19 era: time for bold changes. *Lancet*. 2021;397(10292):2317–9. [https://doi.org/10.1016/S0140-6736\(21\)01096-5](https://doi.org/10.1016/S0140-6736(21)01096-5).
20. Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. *Lancet Infect Dis*. 2014;14(2):160–8. [https://doi.org/10.1016/S1473-3099\(13\)70244-5](https://doi.org/10.1016/S1473-3099(13)70244-5).
21. Aiello AE, Renson A, Zivich PN. Social media- and Internet-Based Disease Surveillance for Public Health. *Annu Rev Public Health*. 2020;41(1):101–18. <https://doi.org/10.1146/annurev-publhealth-040119-094402>.
22. Brownstein JS, Rader B, Astley CM, Tian H. Advances in Artificial Intelligence for Infectious-Disease Surveillance. *N Engl J Med*. 2023;388(17):1597–607. <https://doi.org/10.1056/NEJMr2119215>.
23. MacIntyre CR, Lim S, Quigley A. Preventing the next pandemic: use of artificial intelligence for epidemic monitoring and alerts. *Cell Rep Med*. 2022;3(12):100867. <https://doi.org/10.1016/j.xcrm.2022.100867>.
24. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. *Lancet Digit Health*. 2020;2(8):e435–40. [https://doi.org/10.1016/S2589-7500\(20\)30142-4](https://doi.org/10.1016/S2589-7500(20)30142-4).
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. <https://doi.org/10.1136/bmj.n71>.
26. Haque S, Mengersen K, Barr I, Wang L, Yang W, Vardoulakis S, et al. Towards development of functional climate-driven early warning systems for climate-sensitive infectious diseases: statistical models and recommendations. *Environ Res*. 2024;249:118568. <https://doi.org/10.1016/j.envres.2024.118568>.
27. Gonzalez-Bandala DA, Cuevas-Tello JC, Noyola DE, Comas-Garcia A, Garcia-Sepulveda CA. Computational forecasting methodology for Acute Respiratory Infectious Disease Dynamics.

- Int J Environ Res Public Health. 2020;17(12):1–20. <https://doi.org/10.3390/ijerph17124540>.
28. Zimmer C, Leuba SI, Yaesoubi R, Cohen T. Use of daily internet search query data improves real-time projections of influenza epidemics. *J R Soc Interface*. 2018;15(147):20180220. <https://doi.org/10.1098/rsif.2018.0220>.
  29. Yang S, Ning S, Kou SC. Use internet search data to accurately track state level influenza epidemics. *Sci Rep*. 2021;11(1):4023. <https://doi.org/10.1038/s41598-021-83084-5>.
  30. 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. <https://doi.org/10.1016/j.envint.2018.05.016>.
  31. Fan B, Peng J, Guo H, Gu H, Xu K, Wu T. Accurate forecasting of Emergency Department Arrivals with Internet Search Index and Machine Learning models: Model Development and performance evaluation. *JMIR Med Inf*. 2022;10(7):e34504. <https://doi.org/10.2196/34504>.
  32. Olukanmi SO, Nelwamondo FV, Nwulu NI. Utilizing Google Search Data with Deep Learning, Machine Learning and Time Series modeling to Forecast Influenza-Like illnesses in South Africa. *IEEE Access*. 2021;9:126822–36. <https://doi.org/10.1109/access.2021.3110972>.
  33. Zhou X, Yang F, Feng Y, Li Q, Tang F, Hu S, et al. A spatial-temporal method to detect global influenza epidemics using heterogeneous data collected from the Internet. *IEEE/ACM Trans Comput Biol Bioinform*. 2018;15(3):802–12. <https://doi.org/10.1109/TCBB.2017.2690631>.
  34. Zhang Y, Bambrick H, Mengersen K, Tong S, Hu W. Using internet-based query and climate data to predict climate-sensitive infectious disease risks: a systematic review of epidemiological evidence. *Int J Biometeorol*. 2021;65(12):2203–14. <https://doi.org/10.1007/s00484-021-02155-4>.
  35. Guo P, Zhang J, Wang L, Yang S, Luo G, Deng C, et al. Monitoring seasonal influenza epidemics by using internet search data with an ensemble penalized regression model. *Sci Rep*. 2017;7(1):46469. <https://doi.org/10.1038/srep46469>.
  36. Chen Y, Zhang Y, Xu Z, Wang X, Lu J, Hu W. Avian influenza A (H7N9) and related internet search query data in China. *Sci Rep*. 2019;9(1):10434. <https://doi.org/10.1038/s41598-019-46898-y>.
  37. Yang L, Li G, Yang J, Zhang T, Du J, Liu T, et al. Deep-learning model for Influenza Prediction from Multisource Heterogeneous Data in a megacity: Model Development and evaluation. *J Med Internet Res*. 2023;25:e44238. <https://doi.org/10.2196/44238>.
  38. Zhang Y, Yakob L, Bonsall MB, Hu W. Predicting seasonal influenza epidemics using cross-hemisphere influenza surveillance data and local internet query data. *Sci Rep*. 2019;9(1):3262. <https://doi.org/10.1038/s41598-019-39871-2>.
  39. Sparks RS, Robinson B, Power R, Cameron M, Woolford S. An investigation into social media syndromic monitoring. *Commun Stat - Simul Comput*. 2016;46(8):5901–23. <https://doi.org/10.1080/03610918.2016.1186182>.
  40. Yousefinaghani S, Dara R, Poljak Z, Bernardo TM, Sharif S. The Assessment of Twitter's potential for outbreak detection: Avian Influenza Case Study. *Sci Rep*. 2019;9(1):18147. <https://doi.org/10.1038/s41598-019-54388-4>.
  41. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *J Med Internet Res*. 2014;16(10):e236. <https://doi.org/10.2196/jmir.3416>.
  42. 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(4):743–60. <https://doi.org/10.1007/s10796-018-9893-0>.
  43. World Health Organization. Dengue - the Region of the Americas. <https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON475> (2023). Accessed.
  44. Liu K, Wang T, Yang Z, Huang X, Milinovich GJ, Lu Y, et al. Using Baidu Search Index to Predict Dengue Outbreak in China. *Sci Rep*. 2016;6(1):38040. <https://doi.org/10.1038/srep38040>.
  45. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: a case study in China. *PLoS Negl Trop Dis*. 2017;11(10):e0005973. <https://doi.org/10.1371/journal.pntd.0005973>.
  46. Liu D, Guo S, Zou M, Chen C, Deng F, Xie Z, et al. A dengue fever predicting model based on Baidu search index data and climate data in South China. *PLoS ONE*. 2019;14(12):e0226841. <https://doi.org/10.1371/journal.pone.0226841>.
  47. 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(3):e0005354. <https://doi.org/10.1371/journal.pntd.0005354>.
  48. Ho HT, Carvajal TM, Bautista JR, Capistrano JDR, Viacrusis KM, Hernandez LFT, et al. Using Google trends to examine the spatio-temporal incidence and behavioral patterns of Dengue Disease: a Case Study in Metropolitan Manila, Philippines. *Trop Med Infect Dis*. 2018;3(4). <https://doi.org/10.3390/tropicalmed3040118>.
  49. Husnayain A, Fuad A, Lazuardi L. Correlation between Google trends on dengue fever and national surveillance report in Indonesia. *Glob Health Action*. 2019;12(1):1552652. <https://doi.org/10.1080/16549716.2018.1552652>.
  50. Marques-Toledo CA, Degener CM, Vinhal L, Coelho G, Meira W, Codeco CT, et al. Dengue prediction by the web: tweets are a useful tool for estimating and forecasting Dengue at country and city level. *PLoS Negl Trop Dis*. 2017;11(7):e0005729. <https://doi.org/10.1371/journal.pntd.0005729>.
  51. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. *Lancet Digit Health*. 2021;3(3):e175–94. [https://doi.org/10.1016/S2589-7500\(20\)30315-0](https://doi.org/10.1016/S2589-7500(20)30315-0).
  52. Tu B, Wei L, Jia Y, Qian J. Using Baidu search values to monitor and predict the confirmed cases of COVID-19 in China: - evidence from Baidu index. *BMC Infect Dis*. 2021;21(1):98. <https://doi.org/10.1186/s12879-020-05740-x>.
  53. Li K, Shi J, Liu X, Ward MP, Wang Z, Liu R, et al. Early warning signals for Omicron outbreaks in China: a retrospective study. *J Med Virol*. 2023;95(1):e28341. <https://doi.org/10.1002/jmv.28341>.
  54. Gong X, Hou M, Han Y, Liang H, Guo R. Application of the internet platform in Monitoring Chinese Public attention to the outbreak of COVID-19. *Front Public Health*. 2021;9:755530. <https://doi.org/10.3389/fpubh.2021.755530>.
  55. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using reports of symptoms and diagnoses on Social Media to Predict COVID-19 Case counts in Mainland China: Observational Infoveillance Study. *J Med Internet Res*. 2020;22(5):e19421. <https://doi.org/10.2196/19421>.
  56. Guo S, Fang F, Zhou T, Zhang W, Guo Q, Zeng R, et al. Improving Google Flu trends for COVID-19 estimates using Weibo posts. *Data Sci Manage*. 2021;3:13–21. <https://doi.org/10.1016/j.dsm.2021.07.001>.
  57. Li J, Huang W, Sia CL, Chen Z, Wu T, Wang Q. Enhancing COVID-19 epidemic forecasting accuracy by combining real-time and historical data from multiple Internet-based sources: analysis of social media data, online news articles, and search queries. *JMIR Public Health Surveill*. 2022;8(6):e35266. <https://doi.org/10.2196/35266>.
  58. Gao C, Zhang R, Chen X, Yao T, Song Q, Ye W, et al. Integrating internet multisource big data to predict the occurrence and



- development of COVID-19 cryptic transmission. *NPJ Digit Med*. 2022;5(1):161. <https://doi.org/10.1038/s41746-022-00704-8>.
59. Singh S, McNab C, Olson RM, Bristol N, Nolan C, Bergstrom E, et al. How an outbreak became a pandemic: a chronological analysis of crucial junctures and international obligations in the early months of the COVID-19 pandemic. *Lancet*. 2021;398(10316):2109–24. [https://doi.org/10.1016/S0140-6736\(21\)01897-3](https://doi.org/10.1016/S0140-6736(21)01897-3).
  60. Yom-Tov E, Lampos V, Inns T, Cox IJ, Edelstein M. Providing early indication of regional anomalies in COVID-19 case counts in England using search engine queries. *Sci Rep*. 2022;12(1):2373. <https://doi.org/10.1038/s41598-022-06340-2>.
  61. Loprete M, Panzarasa P, Puliga M, Riccaboni M. Early warnings of COVID-19 outbreaks across Europe from social media. *Sci Rep*. 2021;11(1):2147. <https://doi.org/10.1038/s41598-021-81333-1>.
  62. Didi Y, Walha A, Ben Halima M, Wali A. COVID-19 outbreak forecasting based on Vaccine Rates and tweets classification. *Comput Intell Neurosci*. 2022;2022:4535541. <https://doi.org/10.1155/2022/4535541>.
  63. Yousefinaghani S, Dara R, Mubareka S, Sharif S. Prediction of COVID-19 waves using social media and Google search: a case study of the US and Canada. *Front Public Health*. 2021;9:656635. <https://doi.org/10.3389/fpubh.2021.656635>.
  64. Feng Y, Shah C. Unifying telescope and microscope: a multi-lens framework with open data for modeling emerging events. *Inf Process Manag*. 2022;59(2):102811. <https://doi.org/10.1016/j.ipm.2021.102811>.
  65. Stoleran LM, Clemente L, Poirier C, Parag KV, Majumder A, Masyn S, et al. Using digital traces to build prospective and real-time county-level early warning systems to anticipate COVID-19 outbreaks in the United States. *Sci Adv*. 2023;9(3):eabq0199. <https://doi.org/10.1126/sciadv.abq0199>.
  66. Yang Y, Tsao SF, Basri MA, Chen HH, Butt ZA. Digital Disease Surveillance for emerging infectious diseases: an early warning system using the internet and Social Media Data for COVID-19 forecasting in Canada. *Stud Health Technol Inf*. 2023;302:861–5. <https://doi.org/10.3233/SHTI230290>.
  67. Habibdoust A, Seifaddini M, Tatar M, Araz OM, Wilson FA. Predicting COVID-19 new cases in California with Google Trends data and a machine learning approach. *Inf Health Soc Care*. 2024;1–17. <https://doi.org/10.1080/17538157.2024.2315246>.
  68. McClymont H, Si X, Hu W. Using weather factors and Google data to predict COVID-19 transmission in Melbourne, Australia: a time-series predictive model. *Heliyon*. 2023;9(3):e13782. <https://doi.org/10.1016/j.heliyon.2023.e13782>.
  69. Kogan NE, Clemente L, Liautaud P, Kaashoek J, Link NB, Nguyen AT, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Sci Adv*. 2021;7(10). <https://doi.org/10.1126/sciadv.abd6989>.
  70. Environmental Data & Governance Initiative E. An embattled Landscape Series, Part 2a: Coronavirus and the three-year Trump Quest. to Slash Science at the CDC; 2020.
  71. Carrion M, Madoff LC. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. *Int Health*. 2017;9(3):177–83. <https://doi.org/10.1093/inthealth/ihx014>.
  72. Wark W. Building a better global health security early-warning system post-COVID: the view from Canada. *Int Journal: Canada's J Global Policy Anal*. 2021;76(1):55–67. <https://doi.org/10.1177/0020702020985227>.
  73. Kelly JT, Campbell KL, Gong E, Scuffham P. The internet of things: impact and implications for Health Care Delivery. *J Med Internet Res*. 2020;22(11):e20135. <https://doi.org/10.2196/20135>.
  74. Basch CH, Meleo-Erwin Z, Fera J, Jaime C, Basch CE. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. *Hum Vaccin Immunother*. 2021;17(8):2373–7. <https://doi.org/10.1080/21645515.2021.1894896>.
  75. Finazzi F. Replacing discontinued big tech mobility reports: a penetration-based analysis. *Sci Rep*. 2023;13(1):935. <https://doi.org/10.1038/s41598-023-28137-7>.
  76. Hirvonen N, Jylhä V, Lao Y, Larsson S. Artificial intelligence in the information ecosystem: Affordances for everyday information seeking. *J Association Inform Sci and Technology.n/a(n/a)*. <https://doi.org/10.1002/asi.24860>.
  77. Lindemann NF. Chatbots, search engines, and the sealing of knowledges. *AI Soc*. 2024. <https://doi.org/10.1007/s00146-024-01944-w>.
  78. Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, et al. Scientific discovery in the age of artificial intelligence. *Nature*. 2023;620(7972):47–60. <https://doi.org/10.1038/s41586-023-06221-2>.
  79. MacIntyre CR, Chen X, Kunasekaran M, Quigley A, Lim S, Stone H, et al. Artificial intelligence in public health: the potential of epidemic early warning systems. *J Int Med Res*. 2023;51(3):03000605231159335. <https://doi.org/10.1177/03000605231159335>.
  80. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *Npj Digit Med*. 2022;5(1):2. <https://doi.org/10.1038/s41746-021-00549-7>.
  81. Liang W, Tadesse GA, Ho D, Fei-Fei L, Zaharia M, Zhang C, et al. Advances, challenges and opportunities in creating data for trustworthy AI. *Nat Mach Intell*. 2022;4(8):669–77. <https://doi.org/10.1038/s42256-022-00516-1>.
  82. Will ChatGPT transform healthcare?. *Nat Med*. 2023;29(3):505–6. <https://doi.org/10.1038/s41591-023-02289-5>.
  83. De Angelis L, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health*. 2023;11:1166120. <https://doi.org/10.3389/fpubh.2023.1166120>.
  84. Amiri P, Karahanna E. Chatbot use cases in the Covid-19 public health response. *J Am Med Inf Assoc*. 2022;29(5):1000–10. <https://doi.org/10.1093/jamia/ocac014>.
  85. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI Chatbot for Medicine. *N Engl J Med*. 2023;388(13):1233–9. <https://doi.org/10.1056/NEJMs2214184>.
  86. Milinovich GJ, Magalhaes RJ, Hu W. Role of big data in the early detection of Ebola and other emerging infectious diseases. *Lancet Glob Health*. 2015;3(1):e20–1. [https://doi.org/10.1016/S2214-109X\(14\)70356-0](https://doi.org/10.1016/S2214-109X(14)70356-0).
  87. Lui CW, Wang Z, Wang N, Milinovich G, Ding H, Mengersen K, et al. A call for better understanding of social media in surveillance and management of noncommunicable diseases. *Health Res Policy Syst*. 2021;19(1):18. <https://doi.org/10.1186/s12961-021-00683-4>.
  88. 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(1):690. <https://doi.org/10.1186/s12879-014-0690-1>.
  89. Rohat F, Milinovich GJ, Avril SM, Le Cao KA, Tong S, Hu W. Disease surveillance based on internet-based linear models: an Australian case study of previously unmodeled infection diseases. *Sci Rep*. 2016;6(1):38522. <https://doi.org/10.1038/srep38522>.
  90. Wilson AE, Lehmann CU, Saleh SN, Hanna J, Medford RJ. Social media: a new tool for outbreak surveillance. *Antimicrob Steward Healthc Epidemiol*. 2021;1(1):e50. <https://doi.org/10.1017/ash.2021.225>.

91. Makri A. Bridging the digital divide in health care. *Lancet Digit Health*. 2019;1(5):e204–5. [https://doi.org/10.1016/s2589-7500\(19\)30111-6](https://doi.org/10.1016/s2589-7500(19)30111-6).
92. 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(suppl\_4):S409–13. <https://doi.org/10.1093/infdis/jiw344>.
93. Kedron P, Li W, Fotheringham S, Goodchild M. Reproducibility and replicability: opportunities and challenges for geospatial research. *Int J Geogr Inf Sci*. 2020;35(3):427–45. <https://doi.org/10.1080/13658816.2020.1802032>.
94. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics*. 2021;22(1):122. <https://doi.org/10.1186/s12910-021-00687-3>.
95. Hussein R, Griffin AC, Pichon A, Oldenburg J. A guiding framework for creating a comprehensive strategy for mHealth data sharing, privacy, and governance in low- and middle-income countries (LMICs). *J Am Med Inf Assoc*. 2023;30(4):787–94. <https://doi.org/10.1093/jamia/ocac198>.
96. Egli A, ChatGPT. GPT-4, and other large Language models: the Next Revolution for Clinical Microbiology? *Clin Infect Dis*. 2023;77(9):1322–8. <https://doi.org/10.1093/cid/ciad407>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.