

Purpose-oriented review of public health surveillance systems: use of surveillance systems and recent advances

Hurmat Ali Shah , Mowafa Househ

To cite: Shah HA, Househ M. Purpose-oriented review of public health surveillance systems: use of surveillance systems and recent advances. *BMJ Public Health* 2024;**2**:e000374. doi:10.1136/bmjph-2023-000374

Received 4 July 2023
Accepted 10 May 2024

ABSTRACT

Public health surveillance systems are an important tool for disease distribution and burden of disease as well as enable efficient distribution of resources to fight a disease. The surveillance systems are used to detect, report, track a disease as well as assess the response to the disease and people's attitudes. This paper provides a framework of review for purpose-oriented categorisation of public health surveillance systems. The framework for review of surveillance systems divides the systems into distribution or monitoring or prediction oriented. While there can be other categorisation based on data sources and data types used, the framework for review in this paper provides a cohesive system which can engulf such categories. The framework of review in this paper is purpose oriented, which categorises the surveillance system according to their stated objectives, which are the most important aspect of any public health surveillance system. This review and the framework of categorisation provide comprehensive details of the surveillance systems in terms of data types used, source of data and purpose of the surveillance system.

INTRODUCTION

Syndromic surveillance comprises monitoring of the outbreak of a disease in its early stages through tools collecting data, which may also include online data.^{1 2} Online data recently have become a rapid source of disease or public health information with the advantage of making early information collected through automated or periodic methodology available through online tools.³ Researchers have also made prediction models based on these online tools and sources. Syndromic surveillance is a type of public health surveillance (PHS),⁴ where surveillance is defined as, 'the ongoing, systematic collection, analysis and interpretation of health data essential to the planning, implementation and evaluation of public health practice, closely integrated with the timely dissemination of (this information) to those who need to

know'.⁵ PHS is essential to modern public health practice. The systems designed for PHS revolve around the tools of data science and data analysis. The aim of surveillance is to monitor the impact of control measures, identify emerging health conditions and ultimately guide action which may impact a significant population. Thus, to take public health action or make a public health policy grounded in social and economic reality such maps and surveillance systems are necessary.

The component of the surveillance system is the infrastructure needed to collect and analyse data while the processes are the data collection as well as data analysis criteria.⁶ While conventional surveillance was based on health data from healthcare organisations such as hospitals, government databases and insurance companies, the web has revolutionised the practice of PHS.⁷ This is not to assert that web-based or online methods of surveillance have replaced the conventional surveillance systems, but that the web has complemented the conventional systems of surveillance.⁸ While conventional PHS is more accurate and based on clinically validated data, the online surveillance systems can be seen to function better as early warning systems as the data collection is in real time. Social media-based surveillance on the other hand can give the public belief about socio-economic associations.⁹

In the literature, there are different ways to classify health surveillance systems. While every classification such as classification based on the source of data, the data format and features has its own benefits, the framework developed in this paper classifies surveillance systems based on the purpose of the map, that is, distribution, deployment for active surveillance and deployment for prediction. PHS systems (PHSS) have been reviewed in literature through different considerations.



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. Published by BMJ.

Hamad Bin Khalifa University, Doha, Qatar

Correspondence to

Dr Mowafa Househ;
Mhouseh@hbku.edu.qa

Systematic reviews of PHSS have also been carried out. While there are reviews done of the academically designed surveillance systems and some reviews also consider larger surveillance systems designed and implemented by governments or on governments' published data,¹⁰⁻¹⁴ there is a need to formulate a framework to see different surveillance systems according to the role they can play. The framework can combine multiple objectives as well as data sources, methodologies and data origins. This paper addresses both the needs.

This paper presents a holistic approach to both public health and academic surveillance systems under the framework where different categories are organised according to their objectives as explained above. Under the 'data-gathering, deployment and use' framework defined above multiple methodologies or data sources can be seen to combine with the stated objective of the surveillance system to define the category of the surveillance system. Through this framework, the PHSS can be reviewed in a meaningful way such that the overall picture of what has been done and what are the gaps can be pointed out. There is also a need for understanding how the PHS can build towards event detection, which was one of the initial aims of developing health surveillance systems and also how the systems and the data sources can be used for evaluation purposes in public health.

To put it simply, the contributions of this paper are as follows:

1. Provide a purpose-oriented framework to organise health surveillance systems, such that the uses of each system can be clearly identified.
2. Identify the attributes of data and the processes involved in the PHSS which can be used for evaluation and assessment for different purposes in the domain of public health and health information.
3. Discuss the impact of the varied purposes and evaluation attributes on the possibility of event detection.

CLASSIFICATION OF HEALTH SURVEILLANCE SYSTEMS

Proposed framework

According to our survey of the literature available on PHSS and online maps, we can divide the surveillance systems and maps broadly into three categories with their purpose or objective defined broadly. The first category is the distribution of the disease demographically, distributed through different identifiers such as geography, age, gender, income and other such social determinants of health. There are also studies about structural determinants of health or a disease which are surveilled. The second broader category is that of actual surveillance in real time. While this category needs data gathering, some surveillance systems do not report data in real time. We categorise PHSS that report real-time data and cases under this category. While some surveillance systems report real-time data, others go beyond reporting and use the data gathered to predict the spread or distribution of the disease or the public health condition being

surveilled. Machine learning and other artificial intelligence (AI)-related tools are used in this type of system to carry out the prediction. While these three categories are not totally independent, and some have overlaps it is important conceptually to analyse these maps under the broader objective for which they have been defined. The first category can also be seen as the 'data collection' for meaningful interpretation; the second as 'deployment' of the collected data for real-time monitoring while the third category can be seen as 'use' of the surveillance, that is, identifying the future spread of the disease or its distribution so to focus resources to prevent or control damage in the geography or population that is going to be affected in future as per the prediction.

There can also be a division of the PHSS based on methodology, data source and monitored target such as carried by Groseclose and Buckeridge.⁴ For methodology, we can categorise surveillance systems to be using either electronic health records (EHRs), online news, blog and other online content such as videos, and social media networks. Of course, there can be a mix of these three and some surveillance systems use a combination of these sources to complement the use of the primary source. EHRs are clinically validated data, thus they can be considered as the most accurate source for surveillance. But clinically validated data can take time and the surveillance systems based exclusively on this type of data are slow to update. This can be complemented by online news sources. But news sources are also used as the single source of data in PHSS with the aim to focus or zoom in more on a demography or population that is identified to have a particular disease or public health concern. These areas or sections of the population can further be analysed through clinically validated data to guide a public health policy decision. The third source of data can be online social networks.

Social media has become a huge and important source of data. It is immediate and can be the first source of data. Social media has especially been used to monitor disease outbreaks, surveil pandemic and to monitor health issues that require intimate input such as mental health issues. While the accuracy and quality of social media data may not be as good as an indicator as the clinically validated data it is personal, diverse and can give multiple insights such as the correlation of a disease or a public health issue with socioeconomic indicators and social determinants of health as well as the public belief about a public health issue.⁹

This purpose-oriented classification can also be seen as an important contribution to evaluating these systems, along with the data sources and processes used, in the domain of public health. Although the data sources do overlap, these categories can guide a public health and informatics practitioner into the specific characteristics which can be evaluated. Assessment of attributes of the systems can be helpful in identifying what impacts these have in gathering data, monitoring of diseases and prediction, hence the overall surveillance. Another

contribution of this classification or the result of this classification can be the delineation of which systems can lead to event detection. Which approaches, process and data can be important for event detection can also be known from this purpose-oriented classification.

Methodology

We wanted this paper to present how accurately the health surveillance systems can be classified to further help in investigating the PHSS. We did not intend this paper to be a systematic review, therefore, we used a narrative review approach to give the classification. It is very hard to focus on one area and methods used as this is a paper about divergent areas of health surveillance literature. If this is made a systematic review of a topic, then the purpose of the paper which is to have a larger classification will not be met.

It would be possible to apply Preferred Reporting Items for Systematic Reviews and Meta-Analyses if our aim and topic were very specific or narrow in the field of PHS, for example, health surveillance systems review for influenza, or for leishmaniasis or COVID-19, or public health systems which use a particular methodology or type of data such as social media data, news sources or clinical data, etc. We would have found limited papers on a specific area and selected a plausible number of papers to be reviewed systematically. On the search, we got 995 articles from Google Scholar Search. We need a criterion to trim the number down and have a manageable number of papers to review. This paper does not consider any specific surveillance system; therefore, we selected the health surveillance systems which could fit under the category.

The classification framework is based on an extensive review of literature. We found out that the PHSS can have a purpose and it would add to the literature on health surveillance to have a purpose-oriented classification. It must be mentioned that the systems under each category are not exhaustive. There can be other systems under the same category. We wanted the systems given here to be indicative and to present how our classification framework can be effective and helpful.

This study is a framework paper of already published studies and during this study, no patient data were collected. Thus, no patient was sought or taken.

PURPOSE-ORIENTED CLASSIFICATION OF PHSS

Distribution of disease

The distribution maps represent a spatially refined assessment of a particular public health issue or a disease. These provide a starting point for various public health interventions in terms of developing strategies for control and assessing disease burden. Every surveillance system has data that can be used to represent the distribution, but some surveillance systems and health maps have the explicit function of showing disease distribution. Thus, these systems collect, refine and analyse the data

primarily to represent distribution. Moreover, such maps can either show a representation of global distribution of the disease or it can be focused on a geographic region.

A publicly available public health map on www.healthmap.org developed by Friefield *et al*¹⁰ and Brownstein *et al*¹¹ is a web-based tool which also has other resources available. The map is developed through online content such as news reports, blogs, alerts and other online tools to give a distribution of 87 disease categories in 89 countries. The map was constructed in Friefield *et al*¹⁰ by analysing 778 online reports about disease outbreaks. The map version in¹¹ provides real-time disease outbreak but the major purpose of the map as it is publicly available is a disease distribution system.

The global distribution of diseases can also be academically designed ventures with the ability to apply different academic tools to analyse and visualise the global burden of the disease. One such example is the global distribution of maps of the leishmaniasis diseases carried out by Pigott *et al*.¹² To understand the global distribution of the disease information from various sources such as published literature, online reports, strain archives and genetic data from GenBank was aggregated. The result was detailed maps of the distribution of the disease with the estimate that around 1.7 billion people live in areas where they are at the potential risk of leishmaniasis. Insights such as affected population, level of risk and the distribution of risk for earlier intervention and control can be gained from surveillance system with distribution as their primary focus. Another such example is the yellow fever distribution map developed by Shearer *et al*¹³ for worldwide infection risk zones. Geographical records were analysed to find 5×5 km regions across all the risk zones. The regression model used also took into consideration environmental, biological, vaccination coverage and spatial disease variability. The vaccination data estimate found out that in the risk zone the vaccination averts between 94336 and 118500 cases of yellow fever annually.

Global distribution maps are not limited to diseases only and they are used also for estimating factors and causes associated with a particular disease. Messina *et al*¹⁴ carried out a global environment suitability study for Zika virus. Through a specie distribution modelling, it was shown that tropical and subtropical regions globally have suitable environmental conditions for the spread of the virus. Specie distribution models for finding out the niche environment for developing vectorborne diseases are also carried out for other diseases too. Examples of these include dengue,¹⁵ leishmaniasis¹² and Crimean-Congo Haemorrhagic Fever.¹⁴ Distribution maps are also used to measure attitude towards a public health issue. Twitter activity and sentiment analysis were used to find out public attitude towards immunisation and awareness about vaccination campaigns.¹⁶ Twitter and Facebook were analysed to find out the attitude of the people and public leaders while analysis of news showed a different set of actors. This varying preference of use

of social media and topic-wise selection of social media gives insights into the demographic that prefers one social media type over the other. Similar studies about public attitude towards a health-related issue can engulf studies such as finding attitude of people towards the use of a pharmaceutical product or recording adverse events associated with a medication.¹⁷

Monitoring of disease

While surveillance and monitoring can mean the same thing, in the context of this paper, surveillance is a broader term which includes public health maps and systems that are meant to find out the distribution of disease and make predictions in the future based on the collected data. For the purpose of this review, we classify those surveillance systems and health maps under monitoring which use real-time data to update the system, either demographic-wise and geographical distribution-wise or risk-level wise. Data collection and geographical or demographic distribution are included in this category, but the collected data are a continuous process with various methods used to sift and verify the data. Over the last few years, web-based and social media-based maps have been developed under this category which regularly gets updated. The real-time data can also come from medically validated sources such as EHRs, databases of diseases and so on, but not necessarily.

The surveillance systems under this classification category are often larger systems initiated by governments or large global public health organisations. The real-time web-based surveillance systems were meant to strengthen global disease surveillance systems. The first system to be developed through such an approach was the Programme for Monitoring Emerging Diseases (ProMED-Mail) established in 1994.¹⁸ It was chartered by the Federation of American Scientists with the aim to disseminate information to a wide audience in real time. After this WHO created an efficient infrastructure called Global Outbreak Alert Response Network (GOARN) which builds capacities in partnered networks to coordinate response to global disease outbreaks. From the initial news-based monitoring surveillance systems such as ProMED-Mail, the advent of social media has led to adoption of other real-time social media-based monitoring systems. These are often adopted at the national level such as the Generating Epidemiological Trends from Web Logs Like is being officially accepted by the Swedish government.¹⁹ It has been used as a complementary tool for daily surveillance by epidemiologists.

Under this classification category of surveillance systems used for monitoring, there are some which do not update automatically but the data collected are first analysed and vetted by a human expert. ProMED-Mail and GOARN²⁰ are such systems. There also are collaborations among governments and public health organisations such as WHO in developing PHSS. One example is Global Public Health Intelligence Network²¹ which is a collaboration between Health Canada and WHO for

early warning of potential public health threats which also includes chemical, biological radiological and nuclear induced public health threats. Some other social media-based surveillance systems add more functionality or alertness to previously implemented PHSS such as EpiSPIDER²² extracts emerging infectious disease information from ProMED, combines it with CIA Factbook, extracts location using natural language processing and then posts it on Google Maps. Google Trends is another tool widely used as news aggregator which gives topic-wise mention and frequency. It has been used in tracing epidemics, disease outbreak and distribution of diseases.

Because of COVID-19, there has been a surge in surveillance maps specifically for monitoring cases of COVID-19. Almost every government has one type or the other surveillance system in place to monitor the cases and number of COVID-19. While it will be out of the scope of this paper to mention even the most efficient and informative surveillance systems for COVID-19, the global dashboard developed by John Hopkins University Center for Systems Science Engineering is a comprehensive one.²³ For modelling the outbreak and spread of COVID-19, a stochastic metapopulation epidemic simulation tool is used to simulate global outbreak dynamics. The raw data for the simulation tool are also available while the user interface is an interactive GUI. Similarly, WHO also has a comprehensive global dashboard with interactive user interface to surveil the COVID-19 outbreak and report various data associated with the epidemic.²⁴ There have been other national-level surveillance maps for various purposes available. These maps provide access to the national-level health data and the maps are regularly updated, for example, the health map of the Australian health department.²⁵

There are some limitations to the web-based and social media-based surveillance systems. The biggest issue is internet penetration and asymmetrical global access. The surveillance systems based on social media or web-based online news content are skewed towards developed or developing countries while major portions of the world where the access to internet is not as pervasive as in the developed world may be left out from the data collection process. The other problem is that of reliability of data which is self-reported or comes from news sources which may also not be as reliable as clinically validated data. Another issue with automated surveillance technologies is that of analysis of the language used. The machine learning algorithms used for sentiment analysis or reporting may not contain the nuances of language such as cultural tones, language shifts and colloquiums. These language barriers may affect the accuracy of detecting a disease outbreak or reporting of the disease.

Prediction of disease

Prediction about the disease outbreak in risk zones, identifying the risk zones and tracing the trajectory of a disease is the third purpose of the surveillance maps. Monitoring the disease is part of the prediction as the

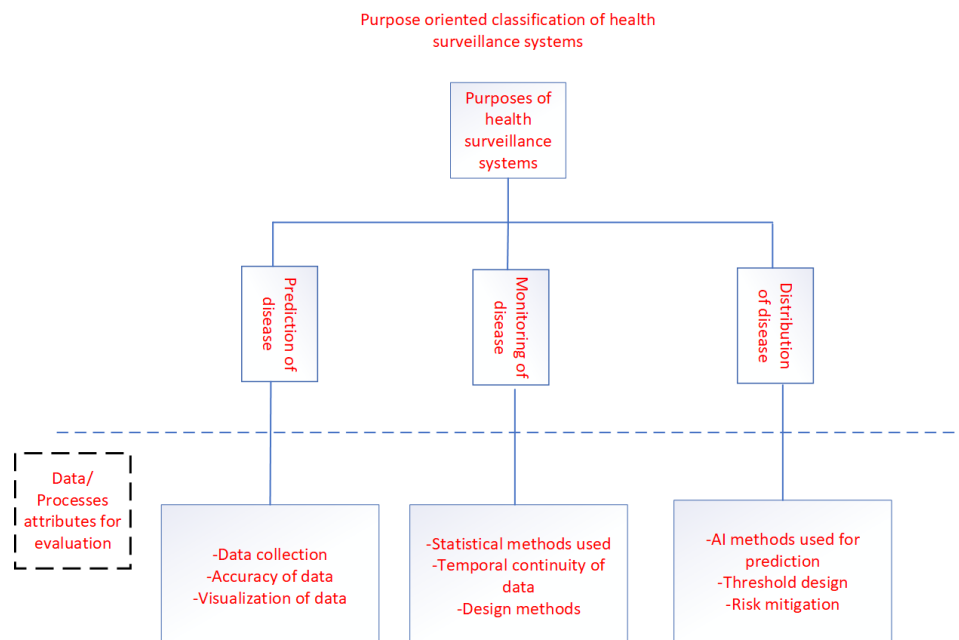


Figure 1 Representation of classification categories.

data collected is used for prediction through different tools. The introduction of AI and machine learning tools into PHS has given the surveillance systems the ability to accurately follow the disease and enable policy-makers to take pre-emptive action. Other than prediction AI is also used to collect the data and analyse the data. AI provides modelling tools that can assess the pattern of disease transmission and spread and can also assess public attitude and responses towards the disease. The predictions of AI are context-based through quantification of variables, responses and other factors in the interacting environment.²⁶

The evolution of an epidemic or a disease in space, time and particular demography is a complex process involving a degree of uncertainty and non-linearity. The application of aggregated statistics and linear interactions is thus limited in predicting the pattern of transmission or outbreak of a disease in a risk zone is limited. AI tools are also used to know the distribution of complex diseases such as AI tools were employed to simulate the global distribution of mosquito-borne infections.²⁷ The risk of dengue transmission was predicted through the use of random forest, an AI tool, in Singapore by taking into consideration the dengue, environmental, entomological and population data. Similarly, deep learning was used to predict the risk of Zika virus outbreak in Americas.²⁸

Prediction using non-linear, unstructured and heterogeneous sources of data is also carried out over the years through surveillance systems. Thapen *et al* combined Twitter data with news sources to predict the outbreak detection.²⁹ The use of Twitter and social media data for outbreak prediction can be challenging as making semantic sense of the tweets and social media posts are informal and often incomplete. Luo *et al*³⁰ proposed a long-short-term RNN structure to classify tweets

containing infection-related information and showed that the model outperformed conventional prediction systems. However, as noted previously, prediction-oriented surveillance systems can have problems such as unequal availability of data in terms of geospatial distribution, the heterogeneity of users and language barriers.

FluSight task is hosted by the US Centers for Disease Control and Prevention which carries seasonal influenza forecasting at the national and regional levels using weighted influenza-like illness (wILI) data. There are other tools developed which are AI based to carry out prediction based on wILI data such as the framework developed by Adhikari *et al*.³¹ There is presymptomatic surveillance where disease outbreaks which are novel and cannot be placed in the current categories are predicted using different machine learning tools such as are used in.^{32 33} Modelling disease transmission is also carried out by prediction-oriented surveillance systems. Scarpino and Petri³⁴ used dynamic approaches such as permutation entropy, Markov chain simulations and epidemic simulations. Machine learning methods were used by Tripathi *et al*³⁵ to predict controllability of disease on complex networks.

Prediction of outbreaks or pattern of transmission of disease involves non-linear modelling and simulation. Hence, it is a non-trivial task. The costs associated with a false positive or false negative are also high as it may result in waste of resources or negligence. Prediction of surveillance systems and health maps will benefit more from advanced AI tools such as deep learning in data collection, making the data operationalisable and analysing the data as well as modelling the disease pattern of transmission.

Table 1 Classification based on purpose design of health surveillance systems

Category	Main themes	Systems considered
Distribution of disease	Data are collected and employed to understand the distribution of disease. The data collection process can be evaluated.	10–17
Monitoring of disease	The state of diseases, especially infectious disease, is monitored. Statistical methods and threshold design, etc can be evaluated.	18–25
Prediction of disease	Systems designed to predict events, that is, occurrence of disease. Methods design and risk mitigation can be evaluated.	26–29 31–35

DISCUSSION

Figure 1 gives a representation of the classification presented here as well as how these systems can be used in evaluation. PHSS was first presented as a fast way to event detection such as discussed by Buehler *et al.*³⁶ The ability of the PHSS to gather information and present the distribution and spread of disease was seen as a route to that purpose. However, these systems are now seen as having primary goal of monitoring and situational awareness with event detection, such as disease outbreak as secondary advantages. However, the classification framework proposed in this paper can be used for two other major concerns in health surveillance and public health. The first is what attributes of data and processes in these categories can be used for evaluation and assessment. While second whether these attributes can be used for event detection.

In the use of these classification categories in evaluation and assessment, the purposes for which the systems are deployed or whether are there significant contributions can also lead to specific attributes for evaluation. For the ‘distribution of disease’ category, various attributes of data as well as the process of data collection can be assessed. The spread, diversity, accuracy and different approaches to data collection can be assessed through their use in these systems. While data collection and the accuracy of data will be the most significant aspect of all systems, as these systems are about the presentation and collection of data, the processes used to collect the data as well as how the data are presented and visualised can be assessed from these systems.

For the second classification category ‘monitoring of disease’, the ability of these systems to formulate algorithms and design methods to monitor diseases can be assessed. Statistical methods used as processes to monitor diseases can be evaluated for their rigour and accuracy. The methods employed also draw prior knowledge from the dataset, therefore, the way the data are preprocessed, or the assumptions drawn from the data can be subject to assessment. Another evaluation aspect in these systems can be the temporal continuity of data, such as how the data are integrated and how the interpretation of historical data is carried out. This aspect of designing methods for historical data integration can be a valuable aspect of the systems under this category.

The systems in the third category ‘prediction of disease’ can be the most important to event detection. In fact, they can be seen as systems whose primary purpose is to detect

events. The systems under the second category do also contribute to event detection. In the evaluation context, the methods and algorithms designed for prediction and event detection can be the most important attributes of the processes to be assessed. As the systems under this category focus on prediction rather than situational awareness as the end goal, the accuracies of these predictions can also be assessed. The false prediction and the strategies associated with such scenarios can be assessed for accuracy of the systems. Another aspect which can be assessed is how capable these systems are of playing event simulation based on data.

From a general evaluation standpoint, the data sources and their stability as well as availability can be assessed in all these systems. Many of the systems presented here use social media data for monitoring and prediction, thus event detection, however, the social media data cannot always be available. This is other than the fact that the same data cannot be as trustworthy as clinically validated data or data gathered through surveys. How the disruption in data gathering, for example, the recent changes in the data availability from X and Reddit should also be evaluated into the development of these systems. How a system that is dependent on social media data or web data and others which use more slow and traditional forms of data collection fare are aspects to be evaluated.

Evaluation of the usability of these systems is also a combined aspect of all the classification categories. As these systems are user-facing and the results and their purpose are for situational awareness as a significant purpose, evaluating how these systems can interact with users in terms of visualisation, appeal, ease of access and operation is an important aspect. The web-based systems have a higher reach and are designed to keep an average end-user in mind rather than a specialist. Recently, evaluation of these PHSS has attracted attention from the research community but the recommendations of these evaluations need to be incorporated in future design of these systems.³⁷ Table 1 gives the themes and one point of each category which can be evaluated.

CONCLUSION

Information-based tracking of a disease outbreak and tracking of a disease as well as its burden of distribution are important aspects of a PHSS. It is important to assess the purpose of the surveillance system beforehand. If the

purpose is to know the exact burden of disease, then the best data type employed should be clinically validated data. However, other sources of data such as web-based content and social media activity of users can be important to gain an overall picture of the distribution of disease or possible outbreak of a disease. The possible risk zones can then be focused on and the accurate picture can be found through clinically validated data. While clinically validated data provide the best accuracy, it is slow to report. The gap here is filled by web-based and social media data. The framework of categorisation of PHSS developed in this paper can take into consideration multiple data types, reporting methods and accuracy levels based on the purpose for which the system is employed. Further research directions can be based on providing a comprehensive review of the use of AI and machine learning tools in predictions of surveillance systems.

Contributors HAS designed the study, carried out and wrote the study while MH supervised the study.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval As this study is a review of already published papers and involves no experiments, no ethical approval was sought or taken from any ethical approval body.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Hurmat Ali Shah <http://orcid.org/0000-0002-5691-4819>

REFERENCES

- Al-Garadi MA, Khan MS, Varathan KD, *et al*. Using online social networks to track a pandemic: a systematic review. *J Biomed Inform* 2016;62:1–11.
- Velardi P, Stilo G, Tozzi AE, *et al*. Twitter mining for fine-grained syndromic surveillance. *Artif Intell Med* 2014;61:153–63.
- Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009;11:e11.
- Groseclose SL, Buckeridge DL. Public health surveillance systems: recent advances in their use and evaluation. *Annu Rev Public Health* 2017;38:57–79.
- Thacker SB, Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev* 1988;10:164–90.
- McNabb S, Surdo AM, Redmond A, *et al*. Applying a new conceptual framework to evaluate tuberculosis surveillance and action performance and measure the costs, Hillsborough County, Florida, 2002. *Ann Epidemiol* 2004;14:640–5.
- Centola D. The spread of behavior in an online social network experiment. *Science* 2010;329:1194–7.
- Ginsberg J, Mohebbi MH, Patel RS, *et al*. Detecting influenza epidemics using search engine query data. *Nature* 2009;457:1012–4.
- Choi J, Cho Y, Shim E, *et al*. Web-based infectious disease surveillance systems and public health perspectives: a systematic review. *BMC Public Health* 2016;16:1238.
- Freifeld CC, Mandl KD, Reis BY, *et al*. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 2008;15:150–7.
- Brownstein JS, Freifeld CC, Reis BY, *et al*. Surveillance Sans Frontiers: internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med* 2008;5:e151.
- Pigott DM, Bhatt S, Golding N, *et al*. Global distribution maps of the leishmaniasis. *Elife* 2014;3:e02851.
- Shearer FM, Longbottom J, Browne AJ, *et al*. Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet Glob Health* 2018;6:e270–8.
- Messina JP, Kraemer MU, Brady OJ, *et al*. Mapping global environmental suitability for Zika virus. *Elife* 2016;5:e15272.
- Bhatt S, Gething PW, Brady OJ, *et al*. The global distribution and burden of dengue. *Nature* 2013;496:504–7.
- Himelboim I, Xiao X, Lee DKL, *et al*. A social networks approach to understanding vaccine conversations on Twitter: network clusters, sentiment, and certainty in HPV social networks. *Health Commun* 2020;35:607–15.
- Freifeld CC, Brownstein JS, Menone CM, *et al*. Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Saf* 2014;37:343–50.
- Madoff LC, Woodall JP. The internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. *Arch Med Res* 2005;36:724–30.
- Hulth A, Rydevik G. GET WELL: an automated surveillance system for gaining new epidemiological knowledge. *BMC Public Health* 2011;11:252.
- 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.
- Global Public Health Intelligence Network. Available: https://gphin.canada.ca/cepr/aboutgphin-rmispnref.jsp?language=en_CA
- Programmableweb. Available: <https://www.programmableweb.com/category/mapping>
- Squarespace. Available: <https://publichealthmaps.org/coronavirus#Dashboards>
- WHO. Available: <https://covid19.who.int/>
- Healthmap. Available: <https://healthmap.com.au/>
- Zeng D, Cao Z, Neill DB. Artificial intelligence-enabled public health surveillance—from local detection to global epidemic monitoring and control. In: *Artificial intelligence in medicine*. Academic Press, 2021: 437–53.
- Ding F, Fu J, Jiang D, *et al*. Mapping the spatial distribution of *Aedes aegypti* and *Aedes albopictus*. *Acta Tropica* 2018;178:155–62.
- Akhtar M, Kraemer MUG, Gardner LM. A dynamic neural network model for predicting risk of Zika in real-time. *bioRxiv* 2019:466581.
- Thapen N, Simmie D, Hankin C, *et al*. DEFENDER: detecting and forecasting epidemics using novel data-analytics for enhanced response. *PLoS One* 2016;11:e0155417.
- Luo L, Wang Y, Liu H. COVID-19 personal health mention detection from Tweets using dual Convolutional neural network. *Expert Syst Appl* 2022;200:117139.
- Adhikari B, Xu X, Ramakrishnan N, *et al*. Epideep. Paper presented at the proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining - KDD '19; Anchorage AK USA, July 25, 2019. 10.1145/3292500.3330917 Available: <https://dl.acm.org/doi/proceedings/10.1145/3292500>
- Lall R, Levin-Rector A, Mathes R, *et al*. Detecting unanticipated increases in emergency department chief complaint keywords. *Online J Public Health Inform* 2014;6:e93.
- Walsh A, Hamby T, St. John TL. Identifying clusters of rare and novel words in emergency department chief complaints. *Online J Public Health Inform* 2014;6:e146.
- Scarpino SV, Petri G. On the predictability of infectious disease outbreaks. *Nat Commun* 2019;10:898.
- Tripathi R, Reza A, Garg D. Prediction of the disease controllability in a complex network using machine learning algorithms. *arXiv [Preprint]* 2019.
- Buehler JW, Hopkins RS, Overhage JM, *et al*. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recomm Rep* 2004;53:11.
- Shah HA, Househ M, Schneider J, *et al*. Studying usability of public health surveillance maps through framework based heuristic evaluation. *Comput Methods Programs Biomed Update* 2024;5:100143.