Infection Prevention in Practice 6 (2024) 100382

Available online at www.sciencedirect.com

Infection Prevention in Practice

journal homepage: www.elsevier.com/locate/ipip

Review

Digital epidemiology: harnessing big data for early detection and monitoring of viral outbreaks

Deema Ibrahim Fallatah [a,*,†], **Hafeez Aderinsayo Adekola** [b,†]

[a] *Department of Clinical Laboratory Sciences, Prince Sattam bin Abdulaziz University, Saudi Arabia*
[b] *Department of Molecular Biology and Biotechnology, Nigerian Institute of Medical Research, Nigeria*

ARTICLE INFO

Article history:
Received 16 March 2024
Accepted 13 June 2024
Available online 29 June 2024

Keywords:
Digital epidemiology
Big data
Infectious viral diseases
Epidemics

SUMMARY

Digital epidemiology is the process of investigating the dynamics of disease-related patterns, both social and clinical, as well as the causes of these trends in epidemiology. Digital epidemiology, utilising big data from a variety of digital sources, has emerged as a viable method for early detection and monitoring of viral outbreaks. The present review gives an overview of digital epidemiology, emphasising its importance in the timely detection of infectious disease outbreaks. Researchers may discover and track outbreaks in real time using digital data sources such as search engine queries, social media trends, and digital health records. However, data quality, concerns about privacy, and data interoperability must be addressed to maximise the effectiveness of digital epidemiology. As the global landscape of infectious diseases evolves, integrating digital epidemiology becomes critical to improving pandemic preparedness and response efforts. Integrating digital epidemiology into routine monitoring systems has the potential to improve global health outcomes and save lives in the event of viral outbreaks.

Introduction

Digital epidemiology is an emerging field that uses big data and digital technologies to detect and track viral epidemics [1]. Researchers analyze data from various sources, including social media and electronic health records, to gain insights into the spread and impact of infectious diseases [1]. This approach helps understand the health and disease determinants of human populations, aiding in the rapid understanding of disease spread, risk factors, and intervention impact at the population scale, while mitigating health and economic consequences.

Big data is the vast amount of information collected in various formats, providing potential for analysis and insights [2]. It has five "V" characteristics: volume, variety, velocity, veracity, and value. In digital epidemiology, big data sources include social media, online news, and mobile health applications [3]. Big data analytics are crucial for virus transmission modeling, infection control measures, and emergency response during disease outbreaks. Applications include pandemic prevention, close contact screening, online public opinion monitoring, viral host analysis, and pandemic forecast evaluation [4].

* Corresponding author. Address: Department of Clinical Laboratory Sciences, Prince Sattam bin Abdulaziz University, Saudi Arabia. Tel.: +966532673732.
E-mail address: n.fllatah@psau.edu.sa (D. I. Fallatah).
† Both Authors Contributed equally to the manuscript.

https://doi.org/10.1016/j.infpip.2024.100382

Utilizing digital data sources for early detection

The clinical and research communities have undertaken various efforts to combine and make available data sources for recent viral epidemics, however it is unclear which of these data are reliable for tracking outbreaks in real time [5]. Digital data sources have proven to be reliable and are currently being utilised to enhance public health responses to viral outbreaks because they allow early detection of viral outbreaks [5]. The internet, clinicians' search engines, news reports, crowd-sourced participatory disease surveillance systems, Twitter microblogs, electronic health records, Wikipedia traffic, wearable gadgets, smartphone-connected thermometers, and travel websites are all examples of digital technology [5].

Surveillance of search engines queries

The use of internet-based search tools and social media has opened up intriguing new opportunities for disease surveillance by collecting real-time data and trends for health outcomes [6]. Data obtained via the internet and social media sites is often utilised to supplement existing outpatient, hospital, and laboratory-based monitoring systems [6]. This supplement can be particularly valuable because typical monitoring systems rely on individuals seeking care and so underestimate the entire disease burden due to a lack of representativeness. One example of a supplementary strategy is the use of electronic health records and historical influenza-like sickness data in models based on Google search phrases, which improved accuracy above Google search trend data alone [6].

Digital surveillance employing search query data has played a crucial role in several viral outbreaks. One example is the creation of Google flu trends (GFT) in 2008 to track influenza incidence in Canada [6]. GFT was a publicly available tool for communicating anticipated influenza incidence based on Google search query volumes. GFT intended to predict influenza incidence one week in advance and was reliable from 2007 to 2009, but it failed to predict the 2009 A/H1N1 pandemic. This failure resulted in an update that was able to predict the pandemic in retrospect; however, the most recent version overestimated the peak intensity of influenza incidence in subsequent years of 2012 and 2013. Because of this failure, the public website was withdrawn [6]. The CDC's FluSight yearly competition is another example of digital surveillance through search queries. Competitors employed digital data sources such as search queries, social media, and other internet-based data to make the most accurate weekly regional influenza predictions. Flusight used influenza modelling with digital data to generate usable objectives for predicting influenza onsets, peaks, and incidence up to four weeks in advance. Because of the success of the FluSight Models, competition for other viral diseases such as dengue, Ebola, and Chikungunya has emerged (Fig. 1). During the COVID-19 Pandemic, Digital Surveillance data were gathered from Daily Google Trends and daily Baidu Attention Index by searching for the keyword "Coronavirus" and collected corresponding data [7]. Digital surveillance of internet search engines found that search phrases for shortness of breath, anosmia, dysgeusia and ageusia, headache, chest pain, and sneezing were strongly associated with both new daily verified cases and COVID-19 deaths [8]. GT COVID-19 (search term) and GT coronavirus (virus) searches came 12 days before real-world confirmed cases [8]. Searches for symptoms of diarrhoea, fever, shortness of breath, cough, nasal obstruction, and rhinorrhea all had a negative lag of more than one week compared to new daily cases, whereas searches for anosmia and dysgeusia peaked globally and in China with positive lags of 5 days and 6 weeks, respectively, corresponding to widespread media coverage of these symptoms in COVID-19 [8].

Monitoring social media platforms for real time information

Social media data has also played an important role in digital surveillance activities, with Twitter being the most widely used [6]. It is clear that social media users are increasingly

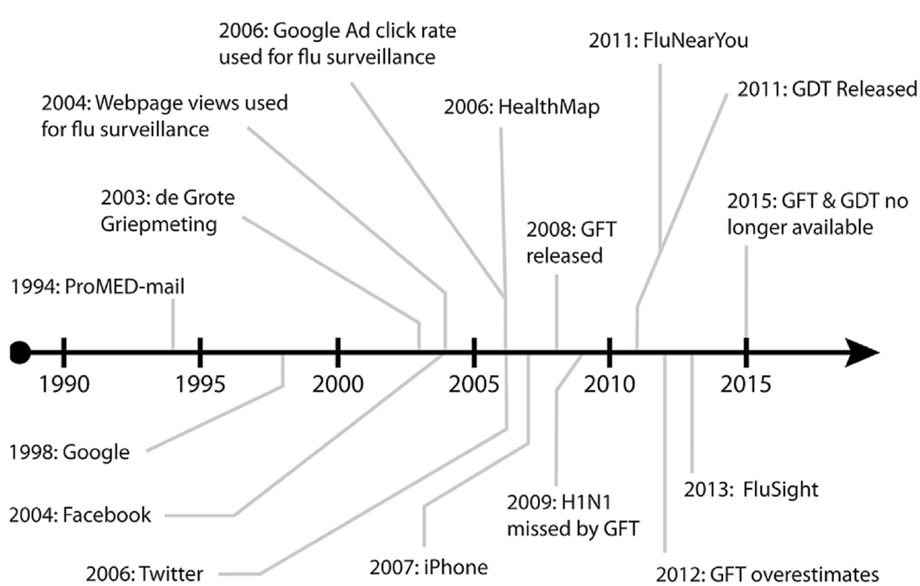

Figure 1. Major events in digital public health surveillance. Abbreviations: GFT: Google Flu Trends, GDT: Google Dengue Trends. Source [6].

sharing information, and researchers are more interested in analysing social media activity for public health purposes. Twitter is a model platform for digital monitoring, and its use entails content identification, either by keyword search or natural language processing, to detect tweets related to current epidemics [6]. Epidemic levels are then estimated as a function of tweet frequency. However, another advantage of Twitter data is the availability of geolocated tweets, which may be utilised to model disease propagation as a function of human geographic movement, potentially improving accuracy [9]. While Twitter is by far the most popular medium for digital monitoring, others have been employed as well. For example, Chaudhary and Naaz (2017) accessed digital health data for various diseases from Practo's Facebook pages, the National Portal of India, and the Integrated Disease Surveillance Programme (IDSP) [10]. Gittelman *et al*. used Facebook "likes" as a predictor of mortality, illnesses, and lifestyle behaviours [11]. YouTube is another prominent platform for analysing videos on numerous health topics, including Ebola virus disease [12].

Analysis of digital health records and wearable devices

Digital health record analysis is uniquely positioned to improve the detection and management of infectious viral diseases [13]. Both medical gadgets and wearables can be repurposed to detect emerging trends that indicate disease outbreaks [13]. Fitbit devices, for example, have been utilised to help inform timely and accurate models of influenza patterns at the population level [14]. The COVID-19 pandemic has stimulated innovation in digital health devices, particularly in locations under heavy lockdown. These advances centred on telemedicine, smartphone apps, and website creation [15]. For example, the IBM Watson Health Explorys was used to collect electronic health record data from 360 hospitals and 317,000 providers across 50 states in the United States to investigate the link between COVID-19 and dementia [16]. Neurological and psychological consequences of COVID-19 survivors were examined using TriNetX electronic health record networks that included over 81 million individuals. The electronic health records were able to provide evidence for significant neurological and mental morbidity among COVID-19 survivors [17]. Aside from COVID-19, electronic health records have been utilised to study other viral epidemics. Wu *et al*. (2020) investigated the interactions between rhinoviruses and influenza A viruses. The Epic Systems electronic medical record system was utilised to analyse data from all testing modalities and age groups. Using the records, the study was able to compare observed and expected co-detections [18].

Wearable technologies can help healthcare professionals prevent and control viral outbreaks by tracking devices, sharing information, and raising awareness to reduce the risk of viral outbreaks. People now wear a variety of wearable devices, including fitness trackers, smart glasses, smart rings, smart shoes, and smart contact lenses [19]. Researchers have presented wearable sensors and demonstrated great accuracy in identifying patients in the prodromal phase as well as monitoring patients' symptoms, such as respiration rate, heart rate, temperature, and so on [19]. Zhu *et al*. (2020) examined large-scale wearable device data to estimate the COVID-19 pandemic trend using four geographically segmented models. The results of the study revealed that prediction models can be used to predict outbreaks and advocated for a health surveillance system based on wearable device data [20]. Radin *et al*. (2020) also acquired de-identified sensor data from hundreds of thousands of individuals living in five states of the United States who wore a Fitbit wearable device for at least 60 days to research influenza-like illness (ILI). The study found that Fitbit data considerably improved ILI predictions in all five states studied. Access to Fitbit and other comparable data could improve real-time and geographically refined influenza surveillance, as well as give critical information for implementing timely outbreak response measures to prevent further transmission of influenza infections during outbreaks [14].

Analytical methods for big data in epidemiology

Many nations use early disease monitoring and detection methods, such as BioSense (USA), CDPAC (Canada), SAMSS, AIHW (Australia), SentiWeb (France), and so on [21]. The adoption of intelligent data analytics tools to extract information from massive amounts of digital data is critical. Data analytics researchers are positioned to make significant advances in meeting the demand for efficient, sensitive, and cost-effective disease prevention solutions [21]. Although, data analytics, geographic information systems, machine learning, and data mining have enabled early disease detection and treatment [21]. There is significant potential for data analytics applications, particularly in the healthcare industry, to avert future outbreaks [21].

Machine learning and data mining techniques

Machine learning (ML) approaches, which use mathematical models and statistical background, are the most essential data processing tools for detecting outbreaks in their early stages [22]. Machine learning techniques can be used to analyse and interpret medical data in order to anticipate illnesses [22]. Former epidemics (such as Ebola, cholera, swine fever, H1N1 influenza, dengue fever, Zika, and oyster norovirus) were modelled using machine learning approaches. During these previous epidemics, machine learning techniques were frequently used to aid healthcare practitioners and authorities in taking better disease-related measures. For example, Sandhu *et al*. (2016) suggested an ML model that uses GPS technology, cloud computing capacity, and google maps to depict possibly infected patients while also providing an alternative route for uninfected users, potentially reducing the spread [23]. The model achieves a classification accuracy of 80% while rerouting away from contaminated individuals. Using ML classification techniques, healthcare providers can better diagnose COVID-19. Furthermore, with the goal of discovering a treatment for the virus, ML algorithms have been used for medication discovery/repurposing and even vaccine development. Thus, machine learning techniques allow us to process labelled and unlabelled datasets using supervised and unsupervised approaches, respectively. While supervised learning methods such as Naïve Bayes (NB), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), and Artificial Neural

Networks (ANN) can predict the emergence of previous outbreaks, unsupervised learning methods such as principal component and cluster analysis can also detect new outbreaks [22].

Since the threat of pandemics has elevated public health concerns, researchers have used data mining techniques to uncover hidden knowledge. Data mining is a semi-automated process of extracting relevant, previously unknown patterns from enormous databases [24]. The rising popularity of data mining may be traced back to the ease of collecting data and storage, which has resulted in large databases containing a plethora of data that traditional methods of analysis struggle to convert into valuable knowledge. Meaningful patterns, in particular, are frequently concealed and unexpected, suggesting that hypothesis-driven approaches may fail to detect them [24]. During the COVID-19 pandemic in 2020, Jiaming Pei and the World Health Organisation (WHO) developed a Bayesian algorithm to give a theoretical and feasible pandemic control method [25]. However, in the Muhammad *et al*. (2020) study, data mining models were created to predict COVID-19 infected patients' recovery using an epidemiological dataset of COVID-19 patients from South Korea. To develop the models, the decision tree, support vector machine, naive Bayes, logistic regression, random forest, and K-nearest neighbour algorithms were applied directly to the dataset in Python. The study showed that a model constructed using the decision tree data mining technique is more efficient at predicting the chance of recovery of infected individuals from the COVID-19 pandemic, with an accuracy of 98.5% [26].

Natural language processing for text analysis

Natural language processing (NLP), also known as text mining, enables automated processing and analysis of unstructured texts, such as extracting relevant information and presenting it in a structured format suitable for computational analysis, or applying transformations such as summarization or translation to make the text more digestible for human readers [27]. NLP allows for faster and bigger scale studies than manual analysis, and it has long been recognised as a method of reducing information overload in biological research [27]. In the last decade, NLP methods have seen tremendous growth with successful applications in a variety of applications, including literature-based discovery (LBD), facilitating the analysis of high-throughput (gene expression/genome-wide association) data, and pharmacovigilance, among many others [27].

Information extraction is a critical component of natural language processing. This extraction via text analysis can be classified into four types [27]. NLP can be used in biomedical literature; digital health records; social media and; news stories, However, because of the various aims of these source types, the text found in each has significantly distinct features [27]. Both clinical notes and social media, for example, have significantly greater rates of spelling and grammatical errors than published literature or news items [27]. The content of clinical notes might differ greatly throughout health care systems and even individual hospitals [27]. There are also considerable variances between abstracts and full-text articles in published literature, which drives differences in the NLP algorithms used. Despite these challenges, it is clear that natural language processing (NLP) can be used to fulfil many of the information demands that have become urgent as a result of viral outbreaks [27].

Sentiment analysis and topic modelling are effective NLP text analysis tools for understanding public attitudes and discussions about viral outbreaks. Sentiment analysis is the process of determining if a digital text's emotional tone is positive, negative, or neutral. Sentiment analysis can be used to assess how individuals feel about viral epidemics by examining social media posts, news articles, or public forums. Social media, in particular, has played an important role in evaluating public opinions, with researchers claiming that many social media data points could have been quickly managed if experts had studied social media data [28]. Topic modelling is also an NLP approach used in information retrieval to infer hidden topics in a collection of documents, providing an automatic method for organising, understanding, and summarising vast volumes of textual material. For viral epidemics, topic modelling can assist in identifying significant themes or subjects mentioned in online chats [29]. It can reveal common subjects such as symptoms, prevention strategies, treatment options, conspiracy theories, and public responses. This can help health officials target specific locations for communication and intervention [29]. Combining sentiment analysis and topic modelling leads to a more comprehensive understanding. For example, Garcia and Berton (2021) utilised topic modelling and sentiment analysis to examine English and Portuguese tweets about COVID-19 from Brazil and the United States between April and August 2021. Over 6 million tweets in English and Portuguese were compared and debated to determine the efficacy of subject recognition and sentiment analysis in each language. The study discovered that seven of the ten COVID-19 topics were the same in both languages, although negative emotions predominated in almost all of the topics investigated. It was argued that governments and authorities may counterbalance sentiments through strategic health communication [30]. Praveen and Ittamalla (2021) also examined the general people's attitude towards COVID-19 crises using sentiment analysis and topic modelling, gathering over 400,000 tweets to trace changes in public sentiment towards the COVID-19 outbreak. The findings demonstrated that the general public's attitude shifted from reasonably neutral at the start of the epidemic to increasingly negative as the pandemic became severe [31].

Geographic information system (GIS) and spatial analysis

Geographic information systems (GIS) are technologies that enable the collection, display, and analysis of spatial data. The GIS has been used to process health data, analyse spatial distribution, map illness prediction, conduct surveillance, and manage epidemics [32]. GIS have been widely utilised to create epidemiological maps of viral epidemics. Disease risk mapping, identifying coronavirus-sensitive areas, and disease clustering are all examples of location-based analysis that can aid in disease prevention and management [32]. Spatial analysis, a critical component of GIS, involves investigating patterns, correlations, and trends in spatial data to gain useful insights [33]. Spatial analysis can be performed using a variety of methodologies, including statistics and geographical information systems [33].

Mapping and visualisation are critical in monitoring and responding to viral epidemics. GIS technology combines epidemiological data, demographic information, and spatial features to generate dynamic maps depicting the distribution and concentration of infectious diseases [34]. These maps provide a visual picture of the impacted areas, assisting health professionals, policymakers, and the general public in understanding the geographical patterns and possible hotspots of viral epidemics [34]. Over the last decade, several researchers have used mapping and visualisation to analyse viral epidemics. Dey *et al*. (2020) used a visual exploratory data analysis approach to investigate the COVID-19 epidemic. The study compiled and analysed epidemiological outbreak information on COVID-19 using numerous free datasets on 2019-nCoV provided by Johns Hopkins University, World Health Organisation, Chinese Centre for Disease Control and Prevention, and National Health Commission. The exploratory data analysis with visualisations revealed the number of different cases reported (confirmed, deceased, and recovered) in various provinces of China and outside of China [35]. Yalcin (2022) also used cartograms to map the worldwide spatio-temporal dynamics of the COVID-19 epidemic across the first 150 days of the pandemic. Using daily country-level data from WHO, the study was able to depict the pandemic's distribution and spatial patterns, and recommended cartograms as a useful visualisation technique [36]. Mast *et al*. (2021) used real-world data and GIS visualisation of rotavirus data to assess whether immunisation improved health outcomes. Visualisation revealed increased vaccination coverage rates at all geographic levels over time; this study informs tactics for monitoring the impact of SARS-CoV-2 vaccinations [37].

Challenges and considerations in digital epidemiology

One of the most significant issues in digital epidemiology is ensuring that data obtained from multiple digital sources is accurate, complete, and reliable. Data that cannot be duplicated due to concerns about its quality and dependability is a waste [38]. To ensure the quality and integrity of epidemiological data, comprehensive data validation, cleaning, and standardisation processes must be developed [38]. The use of digital data for epidemiological research also presents serious privacy and ethical concerns about the acquisition, storage, and dissemination of personal health information. Protecting people's privacy and data confidentiality are critical factors in digital epidemiology. To protect people's privacy, researchers must follow strict ethical norms and regulations, such as obtaining informed consent, anonymizing data, and implementing strong data security measures [39]. Digital epidemiology is based on the integration of data from various sources, such as electronic health records, wearable devices, environmental sensors, and social media platforms. However, these data sources frequently use multiple formats, standards, and protocols, posing obstacles for data interoperability and integration [40]. To overcome these issues involves the creation of standardised data formats, interoperable systems, and data-sharing agreements to promote smooth data exchange and integration across diverse platforms and sources [40].

Conclusion

Digital epidemiology is at the cutting edge of public health surveillance, providing unprecedented prospects for early detection and monitoring of viral epidemics. Digital epidemiology allows for real-time monitoring of disease dynamics and rapid responses to emerging threats by using the power of big data from a variety of digital sources such as social media, internet searches, mobile health applications, and wearable devices. The value of digital epidemiology in improving early detection and monitoring of viral epidemics cannot be overemphasised. Digital epidemiologists can sift through massive volumes of data using modern analytics and machine learning algorithms to spot outbreak signals before they spread to a larger population. Furthermore, digital epidemiology provides proactive surveillance in remote or resource-constrained areas where traditional surveillance methods may be insufficient. As the worldwide landscape of infectious illnesses evolves, integrating digital epidemiology becomes critical to improving pandemic preparedness and response efforts. Public health authorities may keep ahead of emerging threats, reduce disease spread, and better allocate resources by leveraging big data analytics, predictive modelling, and real-time monitoring systems. Incorporating digital epidemiology into routine surveillance systems has the potential to improve global health outcomes while also saving lives during viral epidemics.

Acknowledgements

None Declared by Authors.

Sources of funding

This study was supported by funding from Prince Sattam Bin Abdulaziz University under Project Number PSAU/2024/R/1445.

Ethics statement

Not Relevant.

Conflict of interest statement

None declared.

References

[1] Pastor-Escuredo D, Olmos L. A Survey on Digital Epidemiology: Networks and Models. Med Res Arch 2022;10. https://doi.org/10.18103/mra.v10i10.3175.

[2] Pham Q, Nguyen D, Huynh T, Hwang W, Pathirana P. Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts. IEEE Access 2020;8:130820–39. https://doi.org/10.1109/ACCESS.2020.3009328.

[3] Wu J, Wang J, Nicholas S, Maitland E, Fan Q. Application of Big Data Technology for COVID-19 Prevention and Control in China: Lessons and Recommendations. J Med Internet Res 2020;22:e21980. https://doi.org/10.2196/21980.

[4] Corsi A, de Souza FF, Pagani RN, Kovaleski JL. Big data analytics as a tool for fighting pandemics: a systematic review of literature.

J Ambient Intell Hum Comput 2021;12:9163–80. https://doi.org/10.1007/s12652-020-02617-4.

[5] Kogan N, Clemente L, Liautaud P, Kaashoek J, Link N, Nguyen A, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. Sci Adv 2021;7. https://doi.org/10.1126/sciadv.abd6989.

[6] Aiello AE, Renson A, Zivich PN. Social Media– and Internet-Based Disease Surveillance for Public Health. Annu Rev Publ Health 2020;41:101–18. https://doi.org/10.1146/annurev-publhealth-040119-094402.

[7] Xie T, Tan T, Li J. An Extensive Search Trends-Based Analysis of Public Attention on Social Media in the Early Outbreak of COVID-19 in China. Risk Manag Healthc Pol 2020;13:1353–64. https://doi.org/10.2147/RMHP.S257473.

[8] Higgins T, Wu AW, Sharma D, Illing E, Rubel K, Ting J. Correlations of Online Search Engine Trends With Coronavirus Disease (COVID-19) Incidence: Infodemiology Study. JMIR Public Heal Surveill 2020;6:e19702. https://doi.org/10.2196/19702.

[9] Rocklöv J, Tozan Y, Ramadona A, Sewe M, Sudre B, Garrido J, et al. Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017. Emerg Infect Dis 2019;25:1041–9. https://doi.org/10.3201/eid2506.180138.

[10] Chaudhary S, Naaz S. Use of big data in computational epidemiology for public health surveillance. In: 2017 international conference on computing and communication technologies for smart nation (IC3TSN). IEEE; 2017. p. 150–5.

[11] Gittelman S, Lange V, Gotway Crawford C, Okoro C, Lieb E, Dhingra S, et al. A New Source of Data for Public Health Surveillance: Facebook Likes. J Med Internet Res 2015;17:e98. https://doi.org/10.2196/jmir.3970.

[12] Basch CH, Basch CE, Ruggles KV, Hammond R. Coverage of the Ebola Virus Disease Epidemic on YouTube. Disaster Med Public Health Prep 2015;9:531–5. https://doi.org/10.1017/dmp.2015.77.

[13] Alwashmi M. The Use of Digital Health in the Detection and Management of COVID-19. Int J Environ Res Publ Health 2020;17:2906. https://doi.org/10.3390/ijerph17082906.

[14] Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. Lancet Digit Heal 2020;2:e85–93.

[15] Fagherazzi G, Goetzinger C, Rashid M, Aguayo G, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020;22:e19284. https://doi.org/10.2196/19284.

[16] Wang Q, Davis PB, Gurney ME, Xu R. COVID-19 and dementia: Analyses of risk, disparity, and outcomes from electronic health records in the US. Alzheimer's Dementia 2021;17:1297–306. https://doi.org/10.1002/alz.12296.

[17] Taquet M, Geddes J, Husain M, Luciano S, Harrison P. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. Lancet Psychiatr 2021;8:416–27. https://doi.org/10.1016/S2215-0366(21)00084-5.

[18] Wu A, Mihaylova VT, Landry ML, Foxman EF. Interference between rhinovirus and influenza A virus: a clinical data analysis and experimental infection study. The Lancet Microbe 2020;1:e254–62. https://doi.org/10.1016/S2666-5247(20)30114-2.

[19] Channa A, Popescu N, Skibinska J, Burget R. The Rise of Wearable Devices during the COVID-19 Pandemic: A Systematic Review. Sensors 2021;21:5787. https://doi.org/10.3390/s21175787.

[20] Zhu G, Li J, Meng Z, Yu Y, Li Y, Tang X, et al. Learning from Large-Scale Wearable Device Data for Predicting the Epidemic Trend of COVID-19. Discret Dyn Nat Soc 2020;2020:1–8. https://doi.org/10.1155/2020/6152041.

[21] Razzak MI, Imran M, Xu G. Big data analytics for preventive medicine. Neural Comput Appl 2020;32:4417–51. https://doi.org/10.1007/s00521-019-04095-y.

[22] Şenol A, Canbay Y, Mahmut K. Trends in Outbreak Detection in Early Stage by Using Machine Learning Approaches. Bilişim Teknol Derg 2021;14:355–66.

[23] Sandhu R, Sood SK, Kaur G. An intelligent system for predicting and preventing MERS-CoV infection outbreak. J Supercomput 2016;72:3033–56. https://doi.org/10.1007/s11227-015-1474-0.

[24] Thakur M, Olafsson S, Lee J-S, Hurburgh CR. Data mining for recognizing patterns in foodborne disease outbreaks. J Food Eng 2010;97:213–27. https://doi.org/10.1016/j.jfoodeng.2009.10.012.

[25] Pei J. Big data mining in the control of epidemic. Basic Clin Pharmacol Toxicol 2020:429–30.

[26] Muhammad LJ, Islam MM, Usman SS, Ayon SI. Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery. SN Comput Sci 2020;1:206. https://doi.org/10.1007/s42979-020-00216-w.

[27] Chen Q, Leaman R, Allot A, Luo L, Wei C, Yan S, et al. Artificial Intelligence in Action: Addressing the COVID-19 Pandemic with Natural Language Processing. Annu Rev Biomed Data Sci 2021;4:313–39. https://doi.org/10.1146/annurev-biodatasci-021821-061045.

[28] Singh R, Singh R, Bhatia A. Sentiment analysis using Machine Learning technique to predict outbreaks and epidemics. Int J Adv Sci Res 2018;3:19–24.

[29] Dai J, Lyu F, Yu L, He Y. Temporal and Emotional Variations in People's Perceptions of Mass Epidemic Infectious Disease After the COVID-19 Pandemic Using Influenza A as an Example: Topic Modeling and Sentiment Analysis Based on Weibo Data. J Med Internet Res 2023;25:e49300. https://doi.org/10.2196/49300.

[30] Garcia K, Berton L. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Appl Soft Comput 2021;101:107057. https://doi.org/10.1016/j.asoc.2020.107057.

[31] Praveen S, Rajesh I. An analysis of attitude of general public toward COVID-19 crises – sentimental analysis and a topic modeling study. Inf Discov Deliv 2021;49:240–9. https://doi.org/10.1108/IDD-08-2020-0097.

[32] Razavi-Termeh SV, Sadeghi-Niaraki A, Choi S-M. Coronavirus disease vulnerability map using a geographic information system (GIS) from 16 April to 16 May 2020. Phys Chem Earth, Parts A/B/C 2022;126:103043. https://doi.org/10.1016/j.pce.2021.103043.

[33] Venkatramanan S, Prasanna MV, Chung SY. GIS and geostatistical techniques for groundwater science. Elsevier; 2019.

[34] Nayak P, Pai J, Singla N, Somayaji K, Kalra D. Geographic information systems in spatial epidemiology: Unveiling new horizons in dental public health. J Int Soc Prev Community Dent 2021;11:125. https://doi.org/10.4103/jispcd.JISPCD_413_20.

[35] Dey SK, Rahman MM, Siddiqi UR, Howlader A. Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. J Med Virol 2020;92:632–8. https://doi.org/10.1002/jmv.25743.

[36] Yalcin M. Mapping the global spatio-temporal dynamics of COVID-19 outbreak using cartograms during the first 150 days of the pandemic. Geocarto Int 2022;37:3791–800. https://doi.org/10.1080/10106049.2020.1844310.

[37] Mast T, Heyman D, Dasbach E, Roberts C, Goveia M, Finelli L. Planning for monitoring the introduction and effectiveness of new vaccines using real-word data and geospatial visualization: An example using rotavirus vaccines with potential application to SARS-CoV-2. Vaccine 2021;X 7:100084. https://doi.org/10.1016/j.jvacx.2021.100084.

[38] Quiroga Gutierrez A, Lindegger D, Taji Heravi A, Stojanov T, Sykora M, Elayan S, et al. Reproducibility and Scientific Integrity of Big Data Research in Urban Public Health and Digital Epidemiology: A Call to Action. Int J Environ Res Publ Health 2023;20:1473. https://doi.org/10.3390/ijerph20021473.
[39] Kreuter F, Haas G, Keusch F, Bähr S, Trappmann M. Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent. Soc Sci Comput Rev 2020;38:533–49. https://doi.org/10.1177/0894439318816389.
[40] Ndlovu K, Scott RE, Mars M. Interoperability opportunities and challenges in linking mhealth applications and eRecord systems: Botswana as an exemplar. BMC Med Inf Decis Making 2021;21:246. https://doi.org/10.1186/s12911-021-01606-7.