©IDOSR PUBLICATIONS International Digital Organization for Scientific Research

IDOSR JOURNAL OF APPLIED SCIENCES 11(1):109-116, 2026 ISSN: 2550-7931

https://doi.org/10.59298/IDOSRJAS/2026/111109116 IDOSRJAS11106

Digital Epidemiology: Big Data in Public Health

Nalongo Bina K.

Faculty of Medicine Kampala International University Uganda

ABSTRACT

Digital epidemiology has emerged as a transformative paradigm that integrates big data, computational modelling, and digital platforms into traditional public health surveillance. Driven by rapid advances in technology, data availability, and societal digitalisation, it leverages information generated from non-epidemiological sources such as social media, search engines, electronic health records, and mobility data to detect, monitor, and predict health events in real time. This review explores the foundations, methodological approaches, applications, and ethical considerations of digital epidemiology, highlighting its role in infectious disease monitoring, non-communicable disease (NCD) surveillance, and the assessment of health behaviours and social determinants. Big data technologies support more granular, timely, and wide-ranging insights than conventional epidemiological tools, enabling efficient resource allocation, improved intervention design, and enhanced outbreak preparedness. However, challenges persist regarding data quality, representativeness, interoperability, privacy, equity, and the validity of digital traces. As digital ecosystems grow increasingly complex, robust governance frameworks, methodological innovations, multi-sectoral collaboration, and sustained capacity-building efforts will be vital. This review concludes that digital epidemiology has significant potential to strengthen global public health systems, provided that technological opportunities are matched with ethical safeguards, inclusive policies, and interdisciplinary expertise.

Keywords: Digital Epidemiology, Big Data, Public Health Surveillance, Machine Learning, and Health Informatics.

INTRODUCTION

Epidemiology plays an essential role in controlling disease spread, particularly for infectious pathogens. It monitors distribution, determinants, and health-related phenomena. The field has recently undergone rapid evolution due to significant scientific, technological, and social advances [3]. New data sources and computational models allow epidemiologists to incorporate real-time information on social patterns, connectivity, and human behavior at unprecedented scales into their models. Digital data from the Internet, social media, and mobile applications further complement traditional sources by providing a better understanding of non-communicable diseases (NCDs), health behaviors, risk factors, and complex systems [2]. These extra dimensions also improve resource allocation regarding their spreads, as well as the evaluation of the impact of control measures. To formalise this new era of epidemiology that leverages Digital Data collected for Non-Epidemiological Purposes, the term Digital Epidemiology has been introduced [3]. Early detection of spreading patterns can strengthen monitoring and intervention systems. In addition to governing epidemics, possible extensions cover topics inspired by cross-disciplinary R&D trends that investigate the evolution of climate change, economic crisis, population, and the effect of various monitoring systems [2]. The objective is to sustain effective, robust, replicable, and adaptive solutions for the prevention, management, and mitigation of health crises propelled by any kind of pathogen infestation, integrating also socio-economic and environmental factors that have been neglected. The interest in pursuing R&D in those two strands has significantly heightened since the appearance of SARS-CoV-2, among societal needs yet [5]. Their investigation follows today the systemic sciences acquired in control

theory, covering very wide-ranging yet similar domains [1]. Efforts undertaken at official levels, supplemented at large since the 1990s by academic worlds, notably Geography, predict that challenges to the NCD epidemic are far from over [2].

Foundations of Digital Epidemiology

Epidemiology leverages large Digital Epidemiology integrates aims to obtain information about transmission, spread, and forecasting the evolution of epidemics to the European Centre for Disease Control [2]. Traditional epidemiologies rely on but are subject to substantial retrospective bias. Understanding the timeline of an epidemic is crucial for containing infectious disease threats requires 1 Kostkova et al. highlight that real-time detection of public information monitoring using population mobility data from digital platforms enables tracking pathogens without [3]. Data-and-device device-driven initiations assess intervention policies, estimate transmission, and impact periodically repressing epidemic growth [7]. Quantifying population-scale information diffusion in social networks complements epidemiological models [5]. To properly interpret, participatory surveillance collects diverse, multi-source, fine-grained workflows piloting sensor-based monitoring [4]. Investigating retrospective data does not improve. Combining phenomenological, mechanistic dynamics to track disease propagation facilitates model validation. A balance between scientific advances and technological engagement nurtures transdisciplinary approaches to respond to complex, transboundary threats [6]. Epidemiology is at the crossroads of diverse Digital epidemiology expands the range of phenomena beyond watery border externality feedback loops, metabolic-human ecosystem interactions real-time global-regional connectivities [8]. Combining multiple trajectory patterns facilitates the observation of fractal grain structures, identifies conservation laws spread contagion-like models, and emerges large-scale epidemic simulations, intensiveness delineates observational-theoretical invaluable pandemic governance injection viscosity IMF [9]. Scale-free networks explain ambiguities reconciling abrupt, epidemic-like transitions in progressively free systems. Major outbreaks are habitually characterised by excessive temporal modularity, fuzziness, modelling behavioural socio-physical mobility complex contagion cascades interdependent networks during acute restrictive containment regimes. TransCity mobility dataset bulks quantifiable enable confirmation analytic approaches used kinetic digital surveillance are relevant contemporary-epidemiology design-systemic sustainability enable cross-discipline interpolation pilot non-commercial accessibility connectivity urban complex metropolis stochastic multiplegraph. Public health is intertwined with the military [7]. Public health disruptive scientific revolutions compelled urgency to embrace digitalisation, harnessing the potential of big data discipline-agnostic science for understanding epidemic phenomena, pandemic [9]. Policy decision-making recommending selective, evenly-dispersed population-covering-channels pertinent information dissemination during epidemic communication-oriented healthscape typology govern behaviours adherence-promoting social activity-turning non-epidemic, diffusing scale-free-payoff regain model-fitting rigid accomplish pro-epidemic, wide-velocity diffusion expedite anti-epidemic turnout establish accelerating phase contemporary economic ecosystems[10]. Communication goes beyond spreading epidemic-preventing include mission-centric surveys enhance predictive capability calibrate early-indicators, ascertain timely-phase-emergency-checkpoints injection injecting influence-launch-conditioning-consistency [5]. Dynamic, high-influx large-volume processed throughout inform public knowledge along accompany embed theoretical-flow periodic monitor newly-arisen health-awareness input-file archived stipulating-policy-framework complex contain broader socio-ecology public-policy strategizing culture economics participatory emphasising-dimensional health awareness [7]. Promoting wider participation address myriad digital-resistance phenomenon population subscription-intake formulate spectrum-prescribed on-line pivotal amid thematic insight addressing evasive nature epistemic layers public prioritisation trans-border [8]. Coordination governance scientifically-driven agile formative sectoral-interphase socio-economic assembly socio-health-tech health-data-private-sector collective's cross-country social-resilience engagement enables unprecedented monitoring under-bridge pandemic-agenda data-collective convening watch enabler-sustained [9].

Big Data Technologies in Public Health

Both technology and the availability of microdata are continuously improving, enabling the development of new approaches for collecting and analyzing big data [6]. The first step in conducting digital epidemiology is to collect and integrate data from multiple public platforms and private solutions, such as Google Trends, Twitter, HealthMap, and FitBit. Digital epidemiology relies on several methodological approaches which can be grouped into three main types: [1] qualitative or descriptive epidemiology; [2] mathematical or statistical modeling; and [3] machine learning. The analytics process is therefore highly flexible and customizable, allowing the choice of both software and additional libraries depending on the specific problem. Furthermore, epidemiology is by nature an open and collaborative science. The final component is the implementation of real-time digital surveillance systems [1].

Data collection and integration

Data collections can be characterized as active or passive. Active data collection requires people to report their activities using surveys or other methods [2]. With passive data collection, individuals do not have to take any action and hence this data collection method is unobtrusive for users 4. Activity data is often collected through the devices people use in their daily lives, such as mobile phones and computers, and includes information on location, search queries, purchases, etc [8]. A wide range of digital footprints collected for non-epidemiological purposes offers valuable information for studies of physical activity (e.g., geotagged photos, GPS data, and search queries), diet patterns (e.g., shopping lists and food delivery apps), and social interaction (e.g., mobile apps, twitter, and other social media platforms) compared to that collected by active surveys[9]. The COVID-19 pandemic illustrated the potential of data collected through digital services [1]. The integration of digital footprints requires knowledge of the nature of the data geographical domain, period of availability, and inherent behavior that drives the exhibited behavior and the relations between those footprints (between-footprint relation), allowing the design of analytical models considering those relations [9].

Analytics and modeling approaches

Epidemiology has recently advanced by integrating traditional variables with large-scale, real-time social patterns through computational models [7]. The analysis of social interactions is critical for understanding and stopping outbreaks. The availability of Big Data and specific mobile applications enables the validation and refinement of such models with real-world data, facilitating real-time epidemic mapping, resource allocation, and the design of testing and vaccination strategies [1]. Digital epidemiology contributes to epidemic control by informing decision-makers about the introduction and spread of infectious agents, evaluating the impact of population interactions on these processes, and the effectiveness of countermeasures [2]. It also supports the design of operational mechanisms for prevention, monitoring, and the assessment of intervention effects. Epidemics are complex systems that require systemic solutions, and Digital Epidemiology addresses social-ecological, biogeochemical, and climate factors and their interactions [4]. A broad spectrum of data sources, including social media, social networks, and mobile apps, is exploited to understand the evolving and emergent patterns of disease and the social and clinical processes that explain them [7]. Public health surveillance, the collection of data on diseases, their determinants, and intervention measures to inform the responses of public health authorities and early warning systems, has historically relied on limited and, at times, outdated information[6]. Digital epidemiology, enabled by Big Data, new digital tools, and the shared use of private information, presents new opportunities and challenges for monitoring infectious diseases. Epidemiology itself has evolved over the last century, and the new branch offers a wholly different paradigm [2]. Digital epidemiology uses data from tracking devices, social media, and similar sources to detect outbreaks, assess risks, and facilitate rapid responses, such as testing or contact tracing [10]. The extraction of meaningful insights from large, unstructured, and mostly unhuman-readable digital data remains challenging and involves considerable human effort. Nonetheless, the ability of Machine Learning technologies to analyse such datasets, enabling classification, connection grouping, and prediction alongside the growing appreciation of the impact of other diseases and risk factors on epidemic dynamics, confers significant relevance on this discipline [10].

Real-time surveillance systems

Public health surveillance aims to systematically collect, analyze, and interpret data for specific diseases, cases, and interventions [2]. This evidence is essential for detecting outbreaks and informing response measures. Digital epidemiology uses big data and digital technologies to enable the rapid collection and analysis of both established and novel inputs such as social media, online searches, mobility tracking, and weather stations [6]. Despite the challenge of extracting meaningful insights from large volumes of unstructured information, these data improve the understanding of disease emergence and spread [8]. The integration of additional sources such as electronic medical records, genomics, sensor data, and social networks enhances conventional epidemiological analysis. Algorithmic processing and machine learning facilitate collection, aggregation, and decision-making. Digital epidemiological surveillance broadens the diversity of data used in traditional outbreak detection and risk assessment, supporting timely responses to COVID-19 and other public health threats [4].

Applications of Digital Epidemiology

During the initial weeks of the COVID-19 pandemic, the Spanish digital epidemiology research groups. Epidemiology was created to illustrate the potential of using information obtained through non-epidemiological systems to understand the spread of diseases [5]. The group received wide recognition, providing scientific evidence supporting government actions in real time. Materials made publicly available by epidemiology have been downloaded several hundred thousand times [1]. The World Health Organization (WHO) defines epidemiology as "the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to the control of health problems." Digital epidemiology comprises the inclusion of

digital data into any aspect of epidemiology. Any digital information generated by society can be used as a complementary source of evidence to understand, monitor, and control health-related phenomena [2]. Digital epidemiology allows access to health determinants or health-related behaviors that traditional epidemiology lacks, enriching health understanding for a better design and assessment of public health interventions [4].

Infectious Disease Monitoring

Public health surveillance involves the systematic collection of population-level data on infectious diseases, cases, interventions, and microbial agents to guide public health responses and provide early warning of potential outbreaks [2]. Surveillance also incorporates analysis of relevant media coverage. In a digital public health surveillance context, epidemiological data from digital sources can be integrated into the reporting systems designed to monitor traditional indicators, thereby supplementing and enhancing existing surveillance mechanisms [3]. A global digital public health surveillance system can facilitate the rapid detection of emerging transnational threats, as demonstrated during the COVID-19 pandemic [9]. Worldwide genomic sequencing of SARS-CoV-2, data on cross-border mobility, and digital surveillance of the pathogens were integrated into international analysis to better inform the Bali Process dialogue on security and humanitarian response, which addressed transnational crime, people smuggling, and human trafficked victims [8]. The rapid identification of the virus in December 2019, followed by the timely release of gene sequence data and accelerated large-scale genomic surveillance efforts, contributed to an effective public health response to COVID-19[10].

Non-Communicable Diseases and Risk Factors

Dozens of online sources allow tracking of non-communicable diseases (NCDs) and their risk factors [3]. These include smoking tobacco (e.g., mentions in social media, maps of shops selling tobacco), drinking alcohol (e.g., social media messages, locations of liquor stores), unhealthy diets, lack of physical activity, and exposure to harmful chemicals [4]. Internet searches and social media posts reveal widespread concern or specific questions about NCDs and their risk factors [2]. NCDs account for nearly three-quarters of global mortality [4]. The WHO defines an NCD as a disease that lasts for long periods of time and is not spread through the environment. NCDs develop slowly, are often preventable, and respond poorly to treatment, resulting in life-long management [1]. NCD risk factors include unhealthy diets, physical inactivity, tobacco use, alcohol abuse, air pollution, and stress.

Health Behavior and Social Determinants

Health behavior and social determinants, including tobacco and alcohol use, diet, sedentary behavior, and body weight, have a profound societal impact and are pivotal risk factors for non-communicable diseases (NCDs) [1]. Currently, rapid advances in technology and data availability raise novel opportunities for tracking, monitoring, and understanding these factors [3]. Digital data from varied sources, including short messaging services (SMS), telemetry devices, search, social media, and data applications, inform exposure levels with spatial resolution, immediacy, or frequency unattainable by traditional surveys [3]. These data allow the study of physical activity, tobacco use, obesity, stress, water quality, nutrition, food safety, and other themes with rich observational data across regions, times, and socioeconomic contexts [4].

Ethical, Legal, and Social Implications

Digital epidemiology combines rigorous evidence with critical appraisal of data sources, focusing on how big data informs public health decisions, surveillance, and policy [4]. Rapid proliferation of data and technological advances raises a plethora of ethical, legal, and social implications in digital epidemiology that vary across contexts and regions [5]. Key considerations include privacy and consent, data governance and accountability, and equity and inclusion [6]. The use of massive amounts of mobile metadata in various contexts covering behaviour, media preferences, economic indices, and more demands a thorough examination of the methods and underlying assumptions before extracting conclusions on public health or environmental variables [3]. Previously, ambient-to-user, especially private, large signalling traces have gained accessibility for public health, yet poorly annotated aggregated analytical formats remain available [4]. Restoration of better annotations along meta-stamped criteria or presentation of data at a higher-level secure user privacy is critical. That significantly continues to rise even if large segments were removed during processing, situated connectivity, speed distributions, and crucially, beyond all individual interactions remain accessible [8]. Apparent high-impact events simply may not emerge as higher connectivity and reactivity occur at larger scales than they do at personal size [5]. Rapid proliferation of data and technological advances raises a plethora of ethical, legal, and social implications in digital epidemiology that vary across contexts and regions. Key considerations include privacy and consent, data governance and accountability, and equity and inclusion [8].

Privacy and Consent

The collection and use of personal data pose significant risks to individual privacy [1]. Governmental, corporate, and academic entities often use consent form documents that fall short of complying with statutory requirements

for informed consent [8]. Public health officials committed to the protection of public health must develop modified forms of consent that enable the conduct of certain public health etiological research activities while respecting individuals' right to privacy and obtaining mutually acceptable informed consent [6]. Advances in wireless technology, digital access, and the Internet have profoundly altered the quantity, richness, and availability of everyday societal digital data, including data from consumer use of mobile devices, web browsing, social media, search engines, and e-commerce [7]. Such data have the potential to enhance understanding of the human condition by providing unprecedented access to health-related consumer interest, social determinants of health, and non-communicable disease (NCD) risk factor development and dissemination at massive uninterrupted societal scales, thereby facilitating population health analysis and modelling rather than relying on traditional survey-based methods of health inference from samples that may be poorly suited to urban environments and the data themselves [7].

Data Governance and Governance Models

Data governance, the structure of authority and decision-making that determines how data is collected, stored, shared, and used, is essential to digital epidemiology [6]. The vast and diverse data sources leveraged in digital epidemiology contribute to understanding social patterns, enrich real-time data monitoring, enable integration of multi-source information, and enhance analysis of disease spread. Digital epidemiology applies established models and mechanisms of data governance for efficient epidemiological surveillance and supports operational planning for prevention, mitigation, and monitoring of epidemics [1]. Data governance frameworks must balance flexible, innovative data use with individuals' rights, public health objectives, and epidemiological utility. Stakeholders recognize that, while the existing governance landscape embraces the spirit of the Open Data movement, improvements and adaptation remain necessary [9]. Consent remains fundamental to data sharing. Modern data governance frameworks increasingly accommodate broad dissemination of data derived from public sources, anonymized files, and government activities [8]. Public health stakeholders draw on multiple governance frameworks and disciplines, including public health, epidemiology, privacy, open data, health informatics, research, and social science, to address challenges and enhance planning and response activities [3].

Equity and Inclusion

Social media and other mobile applications provide another avenue for data collection and analysis, although the intention behind conducting such analyses is often unclear [5]. When using crowd-sourced data from the public for scientific research, the data should be collected ethically, with either implicit or explicit consent obtained. The use of Search Engine Query volume data from Google has documented increases in breast, prostate, and other cancers in southern England that are more accurate than those provided by the Office for National Statistics. Social media is being used for other health-related areas such as obesity, vaccination, e-mental health, and infectious diseases [3]. Even though they are rich sources for researching infectious diseases, several ethical issues persist. The appropriate balance that limits government intervention without compromising citizen safety, social media privacy, usage of data without authorization, and evidence of information is still needed [5]. To ensure that personal privacy is maintained, electronic health data should never be used, and personal health literacy (PHL) levels or longitudinal health status during social media communication should be excluded when conducting health-related data analyses. The inequity and exclusion standpoint warrants greater attention from researchers to ensure that economically, socially, and politically disadvantaged populations are not precluded [6]. Studies have indicated low-level attention to HAI and HPV among African American marginalised and economically disadvantaged youth, leading to reduced exposure to health information. Journals increasingly solicit papers on health promotion relevant to neglected cultures, communities of colour, and other groups receiving less than their fair share of attention [5]. Access to information can be observed in many urban poor communities. Most engage in public discussions concerning the availability of certain health facilities. Even with the presence of the Internet and mobile phones in urban poor communities, limited access to health-related materials such as magazines, leaflets, pamphlets, and newspapers is being experienced [7]. The urban poor in high-GDP countries observe that health matters are not a priority topic in their communities, which is reflected in the scarcity of health brokerage and third-party health solutions on the Internet and mobile phone platforms [4]. Even with the use of social media platforms such as Facebook, Blogger, and Twitter by urban poor communities, health information is still conspicuously neglected, diminishing the likelihood of locating candid health discussions [6].

Challenges and Limitations

Public health practitioners need comprehensive and representative datasets to enable timely, accurate analyses of the spread and impact of COVID-19[1]. Yet many readily available datasets derive from social, mobile, and online platforms often used by a concentration of the population, raising serious concerns about their generalizability. When presented with the opportunity to apply a series of publicly available, digital traces to map the spatiotemporal dynamics of the spread of COVID-19, researchers encountered challenges previously observed in

the scientific community [6]. Platforms like Google, Instagram, and Twitter often lack sufficient geographic tagging or present usage data on a national, or even global, scale. Periodic monitoring of Twitter and Wikipedia also requires significant time and processing power, ultimately making these datasets difficult for the broader community to exploit [2].

Data Quality and Representativeness

Digital epidemiology encompasses a variety of data sources, which can restrict their applicability to specific geographical areas or population groups [4]. Modeling combinations of these data streams demands careful consideration of their distinct collecting environments and longitudinal nature [3]. Moreover, the frequency of the data itself has high variability, which suggests data devaluation over time [1]. High temporal resolutions may quickly render the information obsolete, while low frequencies can limit their ability to detect rapid system changes [5]. These characteristics have implications for hygiene, non-communicable diseases, and social determinants of health. Nevertheless, the exploitation of diverse data streams potentially fosters a more comprehensive public health knowledge base in a progressively interconnected world [9].

Bias and Validity

Digital epidemiology must be examined critically; data alone do not produce knowledge. A key challenge closing the gap between Internet-digital talk and the epidemiology that properly informs public health practice lies in the sensitivity and validity of the observations, coupled with issues of context [5]. Data aggregation by corporations easily loses context, leading to significant overestimation of effects [7]. The major outcome from internet-based activity it's very novelty, its interaction with a myriad of variables such as media activity and public transport, and the complexity of the data and the relevant modelling approaches required for countries as opposed to cities or regions pose serious hurdles to formulating reliable measurements [3]. In the distribution of epidemic events, the measurement approaches available strongly affect validity: the choice of a variable or activity, the spatial zones considered, and even the individual adjusting under registration are paramount [6].

Interoperability and standards

The internal mechanisms of interoperability begin with standardization in situations of low to medium data integration, aiming to provide the minimum conditions that different datasets must fulfill to allow a subsequent combination or comparison through, for example, intensity mapping [6]. Moreover, methods emerge that seek to enhance data integration possibilities, for instance, by developing ontologies or dictionaries that bolster the capacity to link observations of bio-ecological phenomena to the activities of individuals recorded in social data, even when no individual-level link can be established [3]. The vast amount of information available on the internet now permits the development of aggregation measures, which, however, concerning health observations, possess a low degree of representativeness of the monitored population; an overestimation of effects is frequent when the baseline average of health observations has insufficient articulation, especially when only the mean is considered [1]. Symbolically, the same activity carried out at the same time when people commute conveys a denser informational flow for measurement aggregation than activity also carried out daily but infrequently for each individual, such as buying groceries or occasionally receiving a health-detecting vaccination [9].

Future Directions

Innovations in statistical methodologies, machine learning approaches, and graphical data representation provide new tools to evaluate public health information [5]. The traditional epidemiological cycle of data collection is no longer linear, as systems that host multi-source data now enable real-time analysis and modeling [3]. Public health entities increasingly view policies related to social determinants of health, non-communicable diseases, and health promotion in conjunction with public health approaches associated with infectious disease epidemiology [3]. On the public health side, integration of experimental, monitoring, and modelling components allows domain expertise to be integrated at multiple scales [8]. Initiatives to train a new generation of public health workers in data science, machine learning, and systems science will enable enhanced capability to use and act on large, disparate datasets for a broader range of public health issues [1].

Methodological innovations

Digital epidemiology is revolutionizing traditional epidemiography by employing computational models that integrate unconventional yet readily accessible data sources [9]. This permits analysis of expansive social systems on unprecedented temporal and spatial scales, yielding enhanced insights into virus dissemination and the overarching socio-ecological dynamics that underpin epidemiological events [8]. Technologies such as mobile applications and electronic health records enable validation and calibration of simulation models against empirical observations, while the mass dissemination of big data actively guides public health measures [1]. Methodological innovations encompass real-time analyses using non-epidemiological data [4], automatic detection of pandemics such as COVID-19 on social media platforms [3], and digital tools that monitor the spread of misinformation and adjust public health responses accordingly.

Policy integration and public health decision making

Digital Epidemiology relies on an evidence-centric approach combined with a critical assessment of data sources. Public health decisions, surveillance, and policy formation are influenced by urbanization, environmental variables, and disease spread-related requests [6]. Physical mobility is pivotal in pathogen dissemination during epidemics, whereas the Internet has become a major vector for entire populations [1]. Data and models describing mobility and online interactions facilitate the mapping of human behavior and offer real-time insight into population density and movement, with the potential for inferring the onset and track of outbreaks, establishing indicators of exposure to contagion, and following the macroscopic evolution of epidemics. Additionally, through a consistent integration of social data into models, models act as observatories for monitoring risk factors, behavioral changes, and socio-epidemiological, socio-ecological, and socio-political systems [3]. Digital Epidemiology creates direct links among early-warning capabilities, identification of the conditions enhancing transmissibility, and policy interventions [7]. These connections, foundational to disruption and mitigation strategies, offer frameworks for approximation, priority determination, and investment in real-time assessments matched to specific outbreaks. Direct modeling of the impact of interventions informs decision-making relative to specific conditions [8]. High-fidelity multi-scale multi-physics models enable exploration of an ensemble of interventions based on institutional responses and an understanding of behaviors, highlighting the need for investment in detailed models even when screening approaches may suffice [10-12]. The lessons learned during the COVID-19 pandemic underline the importance of aligning resources and capabilities with epidemiological, societal, and economic needs to maximize the societal benefits from public health investments.

Education and capacity building

The global health landscape is undergoing a profound transformation driven by rapid advances in data generation, storage capacities, and computational power [4]. One of today's leading priorities is to harness the potential of these data. Worldwide, organizations sustain efforts to monitor health data streams, aiming to forge a comprehensive overview of health conditions, disease evolution, and risk determinants. In this context, education and capacity building are crucial [9]. Commitment to these priorities is evident in recent programs integrating big data training into global health education, acknowledging the escalating use of analytics in health surveillance. Big data is defined as data generated from multiple sources at high velocity, volume, and variety. Even simple documents may encompass thousands of records, a challenge for traditional approaches. Capacity building will enhance basic analytics and design for a wider audience. Despite limitations related to sample size and composition, results echo findings worldwide [13-15].

CONCLUSION

Digital epidemiology represents a major shift from traditional epidemiological practice, offering unprecedented access to real-time, population-level insights derived from vast digital ecosystems. By integrating data from mobile devices, social media, search engines, sensors, and electronic health systems with advanced analytical tools, including machine learning and computational modeling, digital epidemiology enhances the capacity to detect outbreaks early, understand transmission dynamics, and evaluate public health interventions more effectively. The review highlights the breadth of its applications from infectious disease monitoring and genomic surveillance during events such as COVID-19, to the analysis of non-communicable disease risk factors and social determinants of health. These developments have fundamentally expanded the scope of public health practice, enabling predictive, participatory, and adaptive approaches that align with contemporary societal behaviours and technological realities. Yet, the field faces significant challenges. Data quality, representativeness, and validity remain central concerns, as digital traces may exclude marginalised or low-connectivity populations, thereby reinforcing existing inequities. Ethical issues surrounding privacy, consent, surveillance, and data ownership demand rigorous governance structures that balance innovation with the protection of individual rights. Interoperability barriers and the absence of universal standards further constrain the effective integration of diverse datasets. Looking ahead, digital epidemiology will only achieve its full potential if supported by methodological innovation, strong governance frameworks, cross-sector collaboration, and investment in education and capacity building. Training a new generation of public health practitioners equipped with data science and informatics skills is essential. With appropriate safeguards and interdisciplinary engagement, digital epidemiology can serve as a cornerstone of resilient, equitable, and responsive public health systems capable of addressing both current and future global health challenges.

REFERENCES

1. Pastor-Escuredo D. Digital Epidemiology: A review. arXiv preprint arXiv:2104.03611. 2021 Apr 8.
2. Kostkova P, Saigí-Rubió F, Eguia H, Borbolla D, Verschuuren M, Hamilton C, Azzopardi-Muscat N, Novillo-Ortiz D. Data and digital solutions to support surveillance strategies in the context of the COVID-19 pandemic. Frontiers in digital health. 2021 Aug 6;3:707902.

3. Ugwu CN, Ugwu OP, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI, Uti DE. Sustainable development goals (SDGs) and resilient healthcare systems: Addressing medicine and public health challenges in conflict zones. Medicine. 2025 Feb 14;104(7):e41535.
4. Velasco E. Disease detection, epidemiology and outbreak response: the digital future of public health practice. Life sciences, society and policy. 2018 Apr 1;14(1):7.
5. Ugwu OP, Ogenyi FC, Ugwu CN, Basajja M, Okon MB. Mitochondrial stress bridge: Could muscle-derived extracellular vesicles be the missing link between sarcopenia, insulin resistance, and chemotherapy-induced cardiotoxicity?. Biomedicine & Pharmacotherapy. 2025 Dec 1;193:118814.
6. Park HA, Jung H, On J, Park SK, Kang H. Digital epidemiology: use of digital data collected for non-epidemiological purposes in epidemiological studies. Healthcare informatics research. 2018 Oct 31;24(4):253-62.
7. Paul-Chima UO, Nneoma UC, Bulhan S. Metabolic immunobridge: Could adipose-derived extracellular vesicles be the missing link between obesity, autoimmunity, and drug-induced hepatotoxicity?. Medical Hypotheses. 2025 Sep 28:111776.
8. Velasco E. Disease detection, epidemiology and outbreak response: the digital future of public health practice. Life sciences, society and policy. 2018 Apr 1;14(1):7.
9. Goldstein ND, Sarwate AD. Privacy, security, and the public health researcher in the era of electronic health record research. Online journal of public health informatics. 2016 Dec 28;8(3):e207.
10. Paul-Chima UO, Nnaemeka UM, Nneoma UC. Could dysbiosis of urban air microbiota be an overlooked contributor to pediatric asthma and neurodevelopmental disorders?. Medical Hypotheses. 2025 Sep 12:111758.
11. Grande D, Luna Marti X, Merchant RM, Asch DA, Dolan A, Sharma M, Cannuscio CC. Consumer views on health applications of consumer digital data and health privacy among US adults: qualitative interview study. Journal of Medical Internet Research. 2021 Jun 9;23(6):e29395.
12. Ugwu OP, Okon MB, Alum EU, Ugwu CN, Anyanwu EG, Mariam B, Ogenyi FC, Eze VH, Anyanwu CN, Ezeonwumelu JO, Egba SI. Unveiling the therapeutic potential of the gut microbiota–brain axis: Novel insights and clinical applications in neurological disorders. Medicine. 2025 Jul 25;104(30):e43542.
13. Holmes JH. Privacy, security, and patient engagement: the changing health data governance landscape. eGEMs. 2016 Mar 31;4(2):1261.
14. Leonelli S, Tempini N. Where health and environment meet: the use of invariant parameters in big data analysis. Synthese. 2021 May;198(Suppl 10):2485-504.
15. Olayinka O, Kekeh M, Sheth-Chandra M, Akpinar-Elci M. Big data knowledge in global health education. Annals of Global Health. 2017 May 1;83(3-4):676-81.

CITE AS: Nalongo Bina K. (2026). Digital Epidemiology: Big Data in Public Health. IDOSR JOURNAL OF APPLIED SCIENCES 11(1):109-116. https://doi.org/10.59298/IDOSRJAS/2026/111109116