Magna Scientia Advanced Research and Reviews

eISSN: 2582-9394

Cross Ref DOI: 10.30574/msarr

Journal homepage: https://magnascientiapub.com/journals/msarr/

Magna Scientia
Open access publication

(Review Article)

Check for updates

Data-driven public health surveillance systems improving outbreak prediction, health equity, and policy decision-making effectiveness globally

Omotayo Moni Famodu [1, *], Amarachi Igwilo [2], Adanna Umeano [3], Oluremilekun Oyefolu [4] and oluwamayomikun ifeoluwa soremekun [5]

[1] Microbiologist, Budo Specialist Hospital
[2] Electrocardiogram Technician, National Defense College Clinic, Nigeria
[3] Intern Pharmacist, National Hospital, Abuja, Nigeria
[4] All Saints University, Dominica
[5] Red Cross Blood Drive Program, Chicago, USA.

Magna Scientia Advanced Research and Reviews, 2021, 03(02), 167-179

Publication history: Received on 10 November 2021; revised on 24 December 2021; accepted on 29 December 2021

Article DOI: https://doi.org/10.30574/msarr.2021.3.2.0094

Abstract

Data-driven public health surveillance systems are transforming how outbreaks are detected, monitored, and managed across diverse global contexts. By integrating heterogeneous data streams including electronic health records, laboratory reports, mobility data, environmental indicators, and digital traces these systems enable earlier outbreak prediction and more precise situational awareness than traditional surveillance approaches. Advanced analytics, such as machine learning and geospatial modelling, enhance the capacity to identify emerging transmission patterns, forecast disease spread, and assess population-level risk in near real time. Importantly, when designed with equity-oriented frameworks, data-driven surveillance can illuminate disparities in exposure, access to care, and health outcomes among marginalized and underserved populations, supporting more inclusive public health responses. Beyond technical performance, these systems play a critical role in strengthening policy decision-making effectiveness by providing timely, evidence-based insights to guide resource allocation, intervention targeting, and evaluation of policy impacts. However, their effectiveness depends on governance structures that ensure data quality, interoperability, transparency, and ethical use, particularly in low- and middle-income settings where capacity constraints persist. This paper examines the role of data-driven public health surveillance systems in improving outbreak prediction, advancing health equity, and enhancing policy responsiveness globally, highlighting both their transformative potential and the institutional considerations required for sustainable and trustworthy implementation.

Keywords: Data-driven surveillance; Outbreak prediction; Health equity; Public health policy; Digital epidemiology; Global health systems

1. Introduction

1.1. Global Importance of Public Health Surveillance

Public health surveillance is a core function of health systems, providing the systematic and continuous collection, analysis, interpretation, and dissemination of health-related data for disease prevention and control [1]. Surveillance systems support early detection of outbreaks, monitoring of disease trends, and assessment of population health risks, forming the empirical backbone of public health decision-making at local, national, and global levels [1,2]. In an increasingly interconnected world, characterised by intensified international travel, urbanisation, ecological change, and climate variability, the speed and scale at which infectious diseases can spread have increased substantially,

* Corresponding author: Omotayo Moni Famodu

amplifying the need for timely and sensitive surveillance mechanisms [2,3]. Global health security frameworks, including international reporting obligations, rely heavily on effective surveillance to enable coordinated responses and cross-border information sharing [3]. Beyond infectious disease control, surveillance data are essential for tracking non-communicable diseases, health system performance, and the distribution of health outcomes across populations, reinforcing their central role in long-term public health planning and policy evaluation [2,4].

1.2. Limitations of Traditional Surveillance Systems

Despite their foundational role, traditional public health surveillance systems face well-documented limitations that constrain their effectiveness in contemporary health contexts [4]. Indicator-based surveillance, which depends primarily on routine reporting from healthcare facilities and laboratories, is often affected by reporting delays, under-ascertainment, and incomplete geographic or demographic coverage [4,5]. These constraints can impede early outbreak detection, particularly for rapidly spreading or novel pathogens, where delays of even a few days can significantly affect response effectiveness [5]. Traditional systems also tend to prioritise clinically confirmed cases, limiting their ability to capture mild, asymptomatic, or community-level transmission events occurring outside formal healthcare settings [5,6]. In many low- and middle-income countries, surveillance capacity is further undermined by fragmented data infrastructures, limited laboratory networks, and shortages of trained analytical personnel, resulting in uneven data quality and timeliness [6]. Such structural weaknesses contribute to systematic blind spots, whereby marginalised or underserved populations remain underrepresented in surveillance data, exacerbating existing health inequities and delaying targeted interventions [4,6].

1.3. Emergence of Data-Driven and Digital Surveillance Approaches

Advances in digital technologies, data availability, and computational methods have driven the emergence of data-driven public health surveillance systems designed to address the shortcomings of traditional approaches [7]. These systems integrate heterogeneous data streams, including electronic health records, syndromic surveillance, pharmacy sales, mobility data, environmental indicators, and digital traces from online platforms, to enhance situational awareness and predictive capacity [7,8]. Machine learning, statistical modelling, and geospatial analytics enable the detection of anomalous patterns and spatiotemporal trends that may signal emerging outbreaks earlier than conventional reporting mechanisms [5,7]. Event-based and digital surveillance approaches expand the scope of monitoring beyond formal health system data, improving timeliness and sensitivity while offering contextual insights into population behaviour and exposure risks [8]. However, the growing reliance on complex data infrastructures introduces new challenges related to data bias, interoperability, transparency, and ethical governance, which can influence both the accuracy and legitimacy of surveillance outputs [6–8].

1.4. Research Aims, Scope, and Contribution of the Paper

This paper examines how data-driven public health surveillance systems contribute to improved outbreak prediction, enhanced health equity, and more effective policy decision-making at a global scale [7,8]. It critically analyses the conceptual foundations, data sources, and analytical methods underpinning contemporary surveillance systems, with particular attention to their ability to overcome the structural limitations of traditional models [5,7]. The paper further explores how surveillance design choices influence the visibility of vulnerable populations and the distributional impacts of public health interventions, situating equity as a central analytical lens [4,6]. In addition, it assesses the role of surveillance outputs in shaping policy responses, resource allocation, and governance arrangements across diverse institutional contexts [2,8]. By integrating technical, ethical, and policy perspectives, this study provides a structured framework for understanding both the transformative potential and the practical constraints of data-driven public health surveillance systems globally [7,8].

2. Conceptual foundations of data-driven public health surveillance

2.1. Definition and Evolution of Surveillance Systems

Public health surveillance is traditionally defined as the continuous and systematic collection, analysis, interpretation, and dissemination of health data to support public health action [7]. Early surveillance systems were predominantly indicator-based, relying on routine reporting of clinically diagnosed and laboratory-confirmed cases from healthcare facilities to central authorities [8]. While these systems provided standardised and comparable data over time, their dependence on formal health system interactions limited timeliness and sensitivity, particularly during fast-moving outbreaks [9]. The increasing frequency of emerging infectious diseases, coupled with globalisation and complex transmission dynamics, exposed the inadequacy of purely indicator-based models for early warning and rapid response

[8,10]. As a result, surveillance systems have progressively evolved toward more adaptive, data-driven architectures that prioritise speed, flexibility, and predictive capacity alongside traditional accuracy metrics [9,11].

2.2. From Indicator-Based to Event-Based and Digital Surveillance

Event-based surveillance represents a conceptual shift from structured reporting toward the rapid identification of signals indicative of potential public health threats [10]. These systems incorporate unstructured and semi-structured data sources, such as media reports, community notifications, and digital platforms, to capture early indications of abnormal health events [11]. Digital surveillance further extends this paradigm by leveraging large-scale digital data streams, including search engine queries, social media activity, mobility patterns, and remotely sensed environmental data, to infer disease activity in near real time [12,13]. The integration of event-based and digital approaches enhances surveillance sensitivity and timeliness, particularly in settings where formal reporting systems are weak or delayed [10,14]. However, these approaches also introduce challenges related to signal validity, noise, and representativeness, necessitating robust analytical frameworks to distinguish meaningful patterns from spurious correlations [11–13].

2.3. Core Components of Data-Driven Surveillance Systems

Data-driven public health surveillance systems are underpinned by three interrelated components: diverse data inputs, advanced analytical methods, and feedback mechanisms linking analysis to decision-making [14]. Data inputs typically combine traditional health system data with non-traditional sources, enabling a more comprehensive representation of population health dynamics and contextual risk factors [12,15]. Advanced analytics, including machine learning, statistical inference, and geospatial modelling, are employed to detect anomalies, forecast outbreak trajectories, and assess spatial and temporal risk distributions [13,15]. Crucially, effective surveillance systems incorporate feedback loops that translate analytical outputs into actionable insights for public health authorities, ensuring that predictions inform timely interventions and policy adjustments [14,16]. The conceptual effectiveness of data-driven surveillance therefore depends not only on technical sophistication, but also on institutional capacity, governance structures, and the alignment of analytical outputs with real-world decision processes [15,16].

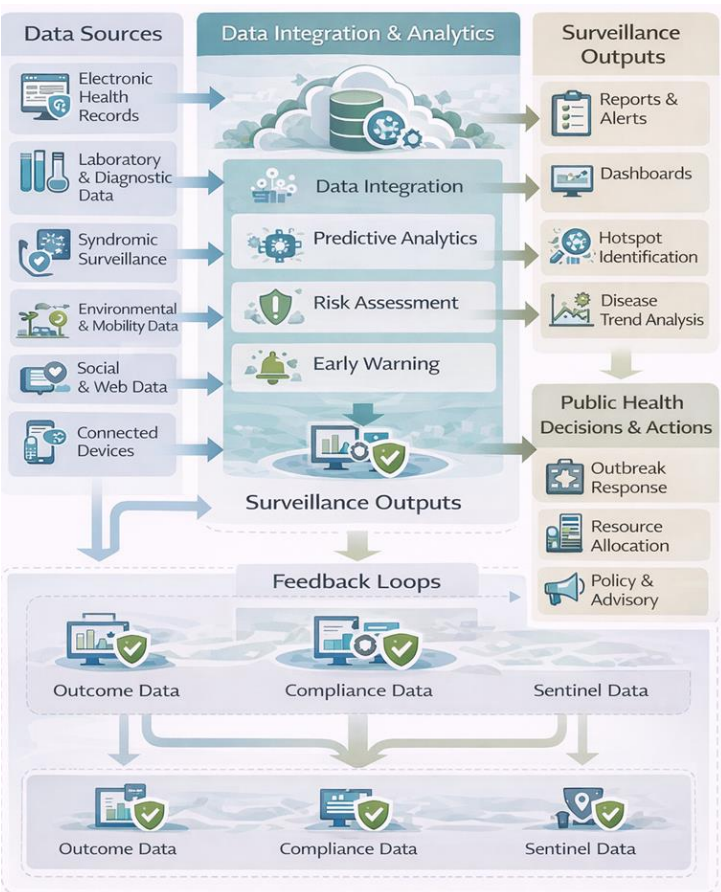

Figure 1 Conceptual framework of data-driven public health surveillance systems

3. Data Sources and Analytical Methods for Outbreak Prediction

3.1. Traditional Health Data Sources

Traditional health data remain a central pillar of outbreak prediction within data-driven surveillance systems, providing clinically validated and standardised information on disease occurrence [15]. Core sources include electronic health records, notifiable disease reports, laboratory-confirmed case data, hospital admissions, and mortality registries, which collectively support longitudinal monitoring of population health trends [15,16]. These data streams offer high specificity and regulatory legitimacy, making them critical for confirming outbreaks and guiding formal public health responses [16]. However, their utility for early outbreak prediction is often constrained by reporting delays, fragmented information systems, and dependence on healthcare-seeking behaviour, which can lead to systematic under-detection during early transmission phases [17]. Despite these limitations, traditional health data continue to serve as an essential reference point for validating signals generated from non-traditional sources and calibrating predictive models [15,17].

3.2. Non-Traditional and Digital Data Sources

To enhance timeliness and sensitivity, data-driven surveillance systems increasingly incorporate non-traditional and digital data sources that capture population behaviour and contextual risk factors in near real time [18]. These sources include syndromic surveillance data, pharmacy sales, school and workplace absenteeism records, mobility and transportation data, climate and environmental indicators, and digital traces such as search engine queries and social media activity [18,19]. Digital data can provide early indications of changing health-seeking behaviour, symptom prevalence, or exposure patterns, often preceding formal case confirmation [19]. Mobility and geolocation data, in particular, have proven valuable for modelling disease spread and identifying potential transmission corridors during outbreaks [20]. Nevertheless, non-traditional data sources are subject to biases related to digital access, platform usage, and socio-demographic representation, which can distort outbreak signals if not carefully adjusted [18–20].

3.3. Analytical Methods for Outbreak Prediction

Advanced analytical methods are fundamental to transforming heterogeneous surveillance data into actionable outbreak predictions [21]. Statistical time-series models are commonly used to detect deviations from expected disease baselines, enabling early identification of anomalous patterns [15,21]. Machine learning techniques, including supervised and unsupervised learning algorithms, support complex pattern recognition across high-dimensional data, improving forecasting accuracy and adaptability to evolving epidemiological conditions [21,22]. Geospatial analytics further enhance predictive capacity by incorporating spatial dependence, environmental context, and population density into outbreak models, facilitating hotspot detection and spatial risk stratification [20,22]. While these methods offer substantial predictive gains, their performance depends on data quality, model transparency, and appropriate validation, underscoring the importance of methodological rigor in surveillance design [21,22].

3.4. Strengths, Limitations, and Uncertainty Management

The integration of diverse data sources and analytical techniques strengthens outbreak prediction by improving timeliness, coverage, and contextual awareness [16,18]. Hybrid surveillance models that combine traditional and digital data can offset individual data weaknesses, enhancing robustness and predictive reliability [17,21]. However, uncertainty remains inherent in outbreak prediction due to reporting noise, behavioural variability, and model assumptions, which can affect decision-making if not explicitly communicated [22]. Effective data-driven surveillance systems therefore incorporate uncertainty quantification, sensitivity analysis, and continuous model recalibration to support informed interpretation by policymakers [21,22]. Managing these trade-offs is essential to ensuring that outbreak predictions are both scientifically credible and operationally useful in real-world public health contexts [16,22].

Table 1 Comparative overview of surveillance data sources and predictive value

Data Source Category	Examples	Timeliness	Predictive Value for Outbreak Detection	Key Strengths	Key Limitations
Clinical and Healthcare Data	Electronic health records, hospital admissions, mortality data	Moderate to low	Moderate	Clinically validated, standardised, high specificity, strong regulatory legitimacy	Reporting delays, dependent on healthcare access, under-detection of mild or asymptomatic cases [15–17]
Laboratory Surveillance Data	Confirmatory diagnostic tests, pathogen sequencing	Low	Low to moderate	High diagnostic accuracy, essential for confirmation and strain identification	Significant time lag, limited early warning capability, resource-intensive [15,16]
Syndromic Surveillance Data	Symptom reports, emergency department chief complaints	High	Moderate to high	Early signal detection, less dependent on laboratory confirmation	Lower specificity, potential false positives, requires validation [18,21]
Mobility and Geolocation Data	Mobile phone location data, transport flows	High	High	Captures transmission dynamics, supports spatial forecasting and hotspot analysis	Privacy concerns, access restrictions, population representation bias [18,20,22]
Digital and Online Data	Search queries, social media posts, web reports	Very high	High [context-dependent]	Near real-time availability, early behavioural signals, broad population coverage	Algorithmic bias, data noise, uneven digital access [12,13,19]
Environmental and Climate Data	Temperature, rainfall, air quality, remote sensing	Moderate	Moderate	Supports modelling of vector-borne and climate-sensitive diseases	Indirect relationship with disease outcomes, requires integration with health data [20–22]

4. Improving outbreak prediction and early warning capacity

4.1. Real-Time Detection and Forecasting of Infectious Diseases

Data-driven public health surveillance systems enhance outbreak prediction by enabling near real-time detection of anomalous disease patterns across multiple data streams [21]. Continuous data ingestion and automated analytics allow surveillance platforms to identify deviations from expected baselines earlier than traditional reporting mechanisms, improving lead time for public health response [22]. Real-time forecasting models integrate epidemiological, behavioural, and environmental variables to estimate short- and medium-term disease trajectories under varying transmission scenarios [23]. These predictive capabilities support anticipatory decision-making, allowing authorities to implement targeted interventions before case counts escalate substantially [21,24]. However, real-time systems require robust data governance, computational infrastructure, and skilled analytical capacity to ensure reliability and interpretability of forecasts [22,23].

4.2. Case Studies of Outbreak Prediction Performance

Empirical evaluations of data-driven surveillance systems demonstrate their potential to improve outbreak prediction accuracy and timeliness across diverse contexts [24]. Retrospective analyses of influenza, dengue, and emerging respiratory pathogens show that models incorporating digital and mobility data often detect outbreaks days or weeks earlier than indicator-based systems alone [25]. These gains in timeliness are particularly pronounced in urban settings with high data density and rapid population movement, where traditional surveillance may lag behind transmission dynamics [24,26]. Nevertheless, performance varies considerably across diseases, regions, and data availability, highlighting the context-dependent nature of predictive effectiveness [25,27]. Such variability underscores the need for continuous performance monitoring and context-specific model adaptation within surveillance systems [26,27].

4.3. Spatial-Temporal Modelling and Hotspot Identification

Spatial-temporal modelling is a critical component of early warning systems, enabling the identification of geographic clusters and transmission hotspots in real time [26]. By integrating geospatial data with temporal trends, these models support granular risk mapping and prioritisation of high-risk areas for intervention [21,27]. Hotspot identification facilitates efficient allocation of limited public health resources, including testing capacity, contact tracing, and preventive measures, thereby improving response efficiency [24,26]. Spatial-temporal analyses also enable assessment of how environmental factors, population density, and mobility patterns interact to shape outbreak dynamics over time [23,28]. Despite these advantages, spatial modelling accuracy depends on data resolution and coverage, which remain uneven across regions, particularly in low-resource settings [26,28].

4.4. Translating Predictive Outputs into Operational Response

The value of outbreak prediction ultimately depends on the effective translation of surveillance outputs into timely and appropriate public health action [22]. Decision-support dashboards, risk thresholds, and alert systems are commonly used to communicate predictive insights to policymakers and operational teams [21,24]. Integrating predictive outputs into established response protocols enhances the likelihood that early warnings result in concrete interventions, such as targeted restrictions, vaccination campaigns, or risk communication strategies [23,27]. However, uncertainties inherent in predictive models can complicate decision-making, particularly when forecasts conflict with political, economic, or social considerations [25,28]. Strengthening institutional trust, transparency, and interpretability of predictive models is therefore essential for maximising the operational impact of data-driven early warning systems [22,28].

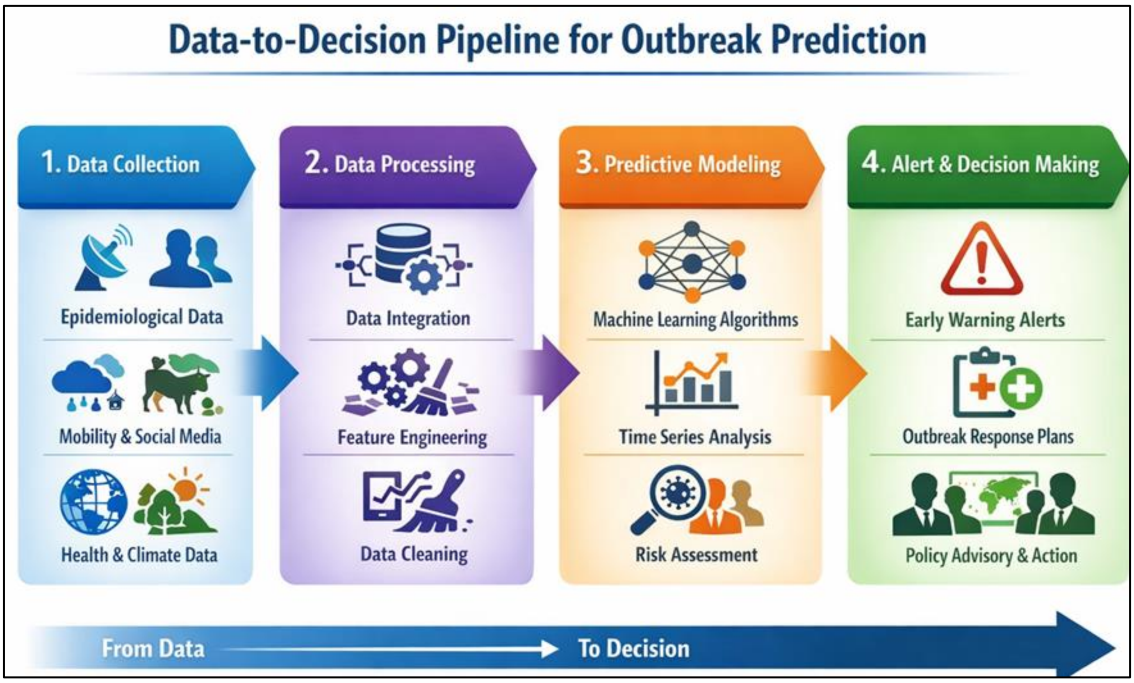

Figure 2 Data-to-decision pipeline for outbreak prediction

5. Advancing health equity through data-driven surveillance

5.1. Identifying Disparities in Exposure, Vulnerability, and Outcomes

Data-driven public health surveillance systems have the potential to improve health equity by systematically identifying disparities in disease exposure, vulnerability, and health outcomes across populations [27]. By integrating demographic, socioeconomic, geographic, and environmental data with epidemiological indicators, surveillance systems can reveal how risks are unevenly distributed among different social groups [28]. Disaggregated surveillance data enable the detection of disproportionate disease burdens affecting marginalised populations, including low-income communities, ethnic minorities, migrants, and informal settlements [27,29]. Such insights are critical for moving beyond aggregate indicators that can obscure inequities and delay targeted public health interventions [28,30]. When effectively implemented, equity-sensitive surveillance supports more precise risk stratification and prioritisation of resources toward populations experiencing the greatest health disadvantages [29,30].

5.2. Surveillance Design for Marginalised and Hard-to-Reach Populations

The extent to which data-driven surveillance advances health equity depends heavily on system design choices and data inclusion strategies [31]. Surveillance systems that rely primarily on formal healthcare data may systematically underrepresent populations with limited access to healthcare services, leading to biased assessments of disease burden [27,31]. Incorporating non-traditional data sources, such as community-based reporting, mobility data, and environmental indicators, can improve coverage of hard-to-reach populations and informal settings [28,32]. Participatory surveillance approaches, which engage communities directly in data generation and validation, further enhance the visibility and relevance of surveillance outputs for marginalised groups [30,32]. However, expanding data collection must be accompanied by safeguards to prevent exploitation, stigmatisation, or unintended harm to vulnerable populations [31,33].

5.3. Bias, Data Gaps, and Representation Challenges

Despite their promise, data-driven surveillance systems are vulnerable to multiple forms of bias that can undermine equity objectives [33]. Digital data sources often reflect existing inequalities in technology access, literacy, and platform usage, resulting in systematic underrepresentation of certain demographic groups [29,33]. Algorithmic bias can further amplify these disparities if predictive models are trained on incomplete or unrepresentative datasets [31,34]. Data gaps related to informal housing, undocumented populations, and fragile settings limit the ability of surveillance systems to capture the full spectrum of health risks [28,34]. Addressing these challenges requires deliberate strategies for bias detection, data triangulation, and continuous model evaluation to ensure equitable surveillance performance [30,31].

5.4. Ethical Considerations and Equity-Oriented Metrics

Ethical governance is central to aligning data-driven surveillance with health equity goals, particularly when sensitive population data are involved [32]. Issues related to privacy, consent, data ownership, and secondary use are especially salient for marginalised communities that may already experience heightened surveillance or discrimination [33]. Equity-oriented surveillance frameworks emphasise transparency, accountability, and proportionality in data use, ensuring that surveillance activities demonstrably serve public health interests [31,34]. The development of explicit equity metrics, such as differential detection rates or intervention coverage across population groups, supports systematic evaluation of whether surveillance systems are reducing or reinforcing disparities [27,30]. Embedding ethical and equity considerations throughout the surveillance lifecycle is therefore essential to realising the inclusive potential of data-driven public health surveillance systems [32,34].

Table 2 Equity Risks and Mitigation Strategies in Data-Driven Public Health Surveillance

Equity Risk	Description of Risk	Implications for Surveillance and Policy	Mitigation Strategies
Underrepresentation of Marginalised Populations	Surveillance data rely heavily on formal healthcare systems or digital platforms that underserved groups may not access	Disease burden among low-income, rural, migrant, or informal populations may be underestimated, delaying targeted interventions	Integrate community-based reporting, participatory surveillance, and non-traditional data sources; strengthen primary care data capture

Digital Divide and Access Bias	Unequal access to internet-enabled devices and digital services skews digital and mobility data	Predictive models disproportionately reflect behaviours of digitally connected populations	Apply weighting and adjustment techniques; combine digital data with traditional and environmental data
Algorithmic and Model Bias	Predictive models trained on incomplete or biased datasets reinforce existing inequities	Resource allocation and policy responses may systematically disadvantage certain groups	Conduct bias audits, use diverse training datasets, and apply fairness-aware modelling approaches
Data Gaps in Informal or Fragile Settings	Limited data availability in informal settlements, conflict zones, or low-resource regions	Surveillance blind spots reduce early warning capacity and equity of response	Invest in local data infrastructure, remote sensing, and partnerships with local organisations
Privacy and Stigmatisation Risks	Use of granular demographic or location data may expose vulnerable communities to stigma or discrimination	Loss of trust and reduced participation in surveillance systems	Implement strong data governance, anonymisation, proportional data use, and community engagement
Lack of Equity-Focused Metrics	Surveillance performance assessed using aggregate indicators only	Improvements in overall outcomes may mask widening disparities	Incorporate disaggregated indicators and equity-sensitive performance metrics into surveillance evaluation
Limited Community Trust and Engagement	Historical mistrust of authorities affects willingness to share data	Reduced data quality and surveillance effectiveness	Ensure transparency, clear communication of public health benefits, and participatory governance mechanisms

6. Policy decision-making and governance implications

6.1. Evidence-Informed Policy Formulation and Resource Allocation

Data-driven public health surveillance systems play a critical role in strengthening evidence-informed policy formulation by providing timely, granular, and actionable intelligence to decision-makers [33]. Predictive analytics and real-time surveillance outputs enable policymakers to anticipate outbreak trajectories and allocate resources proactively rather than reactively [34]. Surveillance-informed policies support more efficient distribution of limited public health resources, including testing capacity, healthcare personnel, vaccines, and non-pharmaceutical interventions [35]. By aligning interventions with spatial and demographic risk profiles, data-driven systems enhance policy precision and effectiveness while minimising unnecessary social and economic disruption [33,36]. However, the policy value of surveillance data depends on decision-makers' capacity to interpret probabilistic outputs and integrate uncertainty into planning processes [34,35].

6.2. Surveillance Data in Emergency and Routine Policy Contexts

The role of surveillance data differs substantially between emergency response and routine public health governance, requiring adaptable policy frameworks [36]. During acute outbreaks, real-time surveillance supports rapid decision-making under conditions of uncertainty, enabling authorities to trigger emergency powers, deploy surge capacity, and implement containment measures [37]. In routine contexts, surveillance data inform long-term policy development, health system strengthening, and prevention strategies by identifying persistent vulnerabilities and structural risk factors [33,38]. Balancing short-term emergency responsiveness with long-term policy learning remains a central governance challenge, particularly when political pressures prioritise immediate outcomes over sustained system investment [36,39]. Effective integration of surveillance into both contexts requires institutional mechanisms that support continuity, learning, and policy feedback loops [37,38].

6.3. Governance, Data Sharing, and Institutional Coordination

Robust governance structures are essential to ensure that data-driven surveillance systems operate effectively, ethically, and sustainably [39]. Surveillance systems often span multiple institutions, including health ministries, data agencies, research organisations, and private-sector data holders, necessitating clear coordination and accountability mechanisms [34,39]. Data sharing frameworks that promote interoperability and standardisation enhance the timeliness and comparability of surveillance outputs across jurisdictions [35,40]. At the same time, governance arrangements must address legal and ethical concerns related to privacy, data protection, and cross-border information exchange to maintain public trust [33,37]. Weak governance and fragmented institutional responsibilities can undermine the translation of surveillance insights into coherent policy action [38,40].

6.4. Global Disparities in Surveillance Capacity and Policy Impact

Significant disparities in surveillance capacity persist between high-income and low- and middle-income countries, shaping the global distribution of policy effectiveness [36]. Resource constraints, limited digital infrastructure, and shortages of skilled personnel reduce the ability of many health systems to fully leverage data-driven surveillance for policy decision-making [38,39]. These inequities can delay outbreak detection and response in vulnerable regions, increasing the risk of regional and global disease spread [34,40]. International cooperation, capacity-building initiatives, and equitable data-sharing partnerships are therefore essential to strengthening global surveillance and policy resilience [35,37]. Addressing these structural disparities is a prerequisite for ensuring that the policy benefits of data-driven public health surveillance systems are realised equitably across global contexts [36,40].

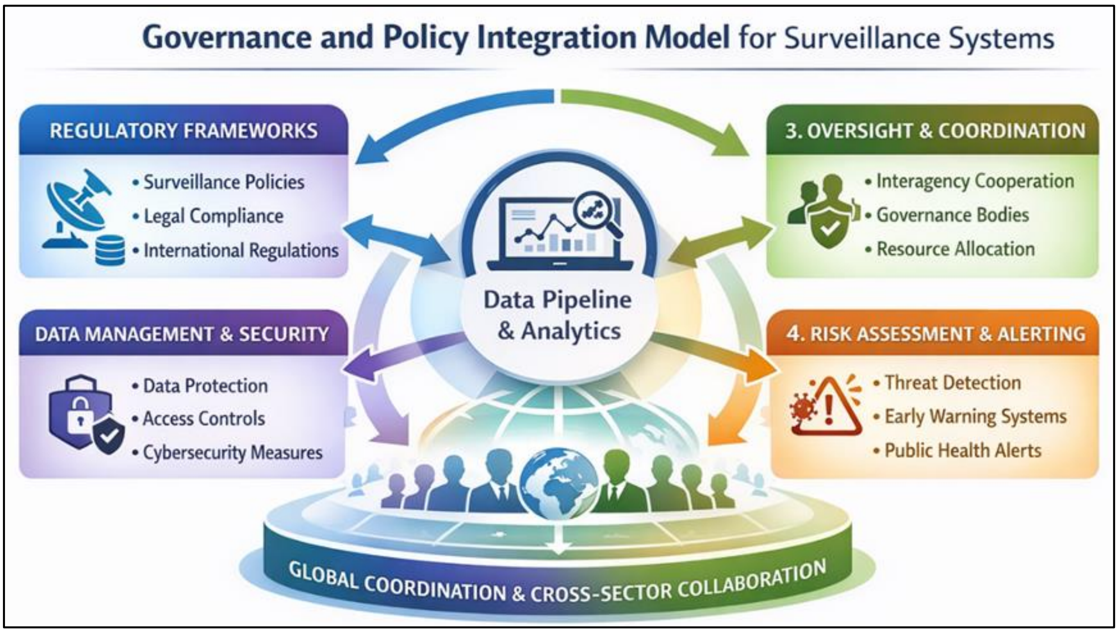

Figure 3 Governance and policy integration model for surveillance systems

7. Implementation challenges and systemic constraints

7.1. Data Quality, Interoperability, and Infrastructure Limitations

The effectiveness of data-driven public health surveillance systems is fundamentally shaped by the quality, completeness, and timeliness of underlying data inputs [32]. Inconsistent reporting standards, missing data, and variable diagnostic capacity can compromise the reliability of surveillance outputs and undermine outbreak prediction accuracy [33,34]. Interoperability challenges arising from fragmented health information systems, incompatible data formats, and siloed institutional databases further limit the integration of diverse data streams [35,36]. In many settings, particularly low- and middle-income countries, inadequate digital infrastructure, limited internet connectivity, and insufficient computational capacity constrain the scalability of advanced surveillance platforms [37,38]. Addressing these structural limitations requires sustained investment in health information systems, standardisation frameworks, and technical capacity-building [39,40].

7.2. Legal, Privacy, and Public Trust Barriers

Legal and ethical considerations represent significant constraints on the implementation of data-driven surveillance systems, particularly when sensitive personal or location-based data are involved [32,41]. Variability in data protection laws, consent requirements, and regulatory oversight across jurisdictions complicates data sharing and cross-border surveillance collaboration [33,42]. Public concerns regarding surveillance overreach, misuse of data, and potential discrimination can erode trust and reduce willingness to participate in data-driven initiatives [41,43]. Transparent governance, clear communication of public health benefits, and proportional data use are therefore essential to maintaining legitimacy and social licence for surveillance activities [32,44]. Failure to adequately address privacy and trust concerns can result in resistance, legal challenges, and reduced effectiveness of surveillance systems [42,45].

7.3. Capacity Constraints in Low- and Middle-Income Countries

Capacity constraints remain a major barrier to the equitable implementation of data-driven surveillance systems globally [34,37]. Shortages of trained epidemiologists, data scientists, and informatics specialists limit the ability of health authorities to develop, maintain, and interpret advanced surveillance models [38,39]. Financial constraints and competing health priorities further restrict investment in digital surveillance infrastructure and long-term system maintenance [36,40]. Dependence on external funding and technical assistance can undermine sustainability and local ownership of surveillance systems in resource-constrained settings [37,41]. Strengthening national capacity through workforce development, institutional strengthening, and locally relevant system design is therefore critical to sustainable surveillance implementation [39,45].

7.4. Sustainability and Scalability Considerations

Ensuring the sustainability and scalability of data-driven surveillance systems requires alignment between technological innovation, institutional capacity, and long-term funding mechanisms [40]. Pilot surveillance initiatives often demonstrate technical feasibility but fail to scale due to lack of integration with existing health systems and policy processes [36,42]. Sustainable systems must be designed to adapt to evolving data sources, emerging health threats, and changing governance environments without excessive complexity or cost [38,43]. Continuous evaluation, stakeholder engagement, and iterative system refinement support resilience and long-term relevance of surveillance platforms [41,44]. Without deliberate attention to sustainability and scalability, data-driven surveillance systems risk remaining fragmented, underutilised, or dependent on short-term crisis-driven investments [40,45].

8. Conclusion

Data-driven public health surveillance systems have emerged as a critical advancement in the detection, monitoring, and management of health threats in an increasingly complex and interconnected world. By integrating diverse data sources with advanced analytical methods, these systems enhance the timeliness and precision of outbreak prediction, enabling public health authorities to move from reactive responses toward anticipatory and preventive action. This shift has significant implications for improving situational awareness, optimising resource allocation, and strengthening the overall effectiveness of public health interventions.

Beyond technical performance, data-driven surveillance holds important potential for advancing health equity by making disparities in exposure, vulnerability, and outcomes more visible within public health decision-making processes. When surveillance systems are designed to account for data gaps, bias, and representation, they can support more inclusive and proportionate policy responses that prioritise populations at greatest risk. However, realising this potential requires deliberate attention to ethical governance, transparency, and institutional trust, particularly where sensitive or non-traditional data sources are involved.

The effectiveness of data-driven surveillance ultimately depends on its integration into policy and governance structures that can translate analytical insights into timely and meaningful action. Persistent challenges related to data quality, interoperability, capacity constraints, and sustainability must be addressed to ensure long-term impact, especially in resource-limited settings. Strengthening global cooperation, investing in institutional capacity, and aligning technological innovation with public health values are essential steps toward building resilient surveillance systems capable of supporting equitable and effective responses to future public health challenges.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclsoed.

References

[1] Unanah OV, Yunana AP, Adefolaju I. Community-based vaccine advocacy in rural Sub-Saharan Africa: combating misinformation through culturally tailored campaigns. *Int J Comput Appl Technol Res*. 2021;10(12):371–385. ISSN: 2319-8656.

[2] Reeves JJ, Pageler NM, Wick EC, Melton GB, Tan YH, Clay BJ, Longhurst CA. The clinical information systems response to the COVID-19 pandemic. Yearbook of medical informatics. 2021 Aug;30(01):105-25.

[3] KOMI LS, CHIANUMBA EC, YEBOAH A, FORKUO DO, MUSTAPHA AY. Advances in community-led digital health strategies for expanding access in rural and underserved populations. Iconic Res Eng J. 2021 Sep;5(3):299-317.

[4] Carlson CJ, Farrell MJ, Grange Z, Han BA, Mollentze N, Phelan AL, Rasmussen AL, Albery GF, Bett B, Brett-Major DM, Cohen LE. The future of zoonotic risk prediction. Philosophical Transactions of the Royal Society B: Biological Sciences. 2021 Nov 8;376(1837).

[5] Zafar A. BIG DATA ANALYTICS FOR PUBLIC POLICY: A MULTIDISCIPLINARY APPROACH TO DECISION-MAKING. Multidisciplinary Research in Computing Information Systems. 2021 Jan 4;1(1):11-21.

[6] Araz OM, Choi TM, Olson DL, Salman FS. Role of analytics for operational risk management in the era of big data. Decision Sciences. 2020 Dec;51(6):1320-46.

[7] Uzor D. Systems-level modeling of antimicrobial utilization patterns to optimize stewardship interventions and suppress emerging resistance pathways. GSC Adv Res Rev. 2021;9(3):203–218. doi:10.30574/gscarr.2021.9.3.0310.

[8] Nwenekama Charles-Udeh. Leveraging financial innovation and stakeholder alignment to execute high-impact growth strategies across diverse market environments. Int J Res Finance Manage 2019;2(2):138-146. DOI: 10.33545/26175754.2019.v2.i2a.617

[9] Omitoyin G, Moshood M. Evaluating digital transformation strategies in global supply chain management through strategic project leadership and measurable economic impact. *Magna Scientia Adv Res Rev*. 2021;3(2):122–138. doi:10.30574/msarr.2021.3.2.0092. Available from: https://doi.org/10.30574/msarr.2021.3.2.0092

[10] Udeh NC. *Building sustainable SME banking strategies that expand market access, boost client retention, and support economic inclusion*. International Journal of Financial Management and Economics. 2018;1(1):126-135. doi:10.33545/26179210.2018.v1.i1.674.

[11] Eze Dan-Ekeh. DEVELOPING ENTERPRISE-SCALE MARKET EXPANSION STRATEGIES COMBINING TECHNICAL PROBLEM-SOLVING AND EXECUTIVE-LEVEL NEGOTIATIONS TO SECURE TRANSFORMATIVE INTERNATIONAL ENERGY PARTNERSHIPS. International Journal Of Engineering Technology Research & Management (IJETRM). 2018Dec21;02(12):165–77.

[12] Anthony P, Dada SA. Data-driven optimization of pharmacy operations and patient access through interoperable digital systems. Int J Multidiscip Res Growth Eval. 2020;1(2):229-44.

[13] Kolawole Oloke. Architecting autonomous financial decision engines through federated learning and hybrid cloud frameworks. Int J Appl Res 2019;5(6):500-510. DOI: 10.22271/allresearch.2019.v5.i6d.13166

[14] Eze Dan-Ekeh. Engineering high-value commercialization frameworks integrating technical innovation with strategic sales leadership to drive multimillion-dollar growth in global energy markets. World J Adv Res Rev. 2019;4(2):256-268. doi:10.30574/wjarr.2019.4.2.0152

[15] World Health Organization. Public health surveillance. Geneva: WHO; 2014.

[16] Heymann DL, Rodier GR. Global surveillance, national surveillance, and SARS. Emerg Infect Dis. 2004;10[2]:173–175.

[17] Morse SS. Global infectious disease surveillance and health security. JAMA. 2012;307[4]:372–373.

[18] Thacker SB, Berkelman RL. Public health surveillance in the United States. Epidemiol Rev. 1988;10:164–190.

[19] Paramasivan A. Big Data to Better Care: The Role of AI in Predictive Modelling for Healthcare Management. International Journal of Innovative Research and Creative Technology. 2020;6(3):1-9.

[20] Mohapatra S, Satpathy S, Paul D. Data-driven symptom analysis and location prediction model for clinical health data processing and knowledgebase development for COVID-19. InApplications of artificial intelligence in COVID-19 2021 Sep 30 (pp. 99-117). Singapore: Springer Singapore.

[21] Kringos D, Carinci F, Barbazza E, Bos V, Gilmore K, Groene O, Gulácsi L, Ivankovic D, Jansen T, Johnsen SP, De Lusignan S. Managing COVID-19 within and across health systems: why we need performance intelligence to coordinate a global response. Health research policy and systems. 2020 Jul 14;18(1):80.

[22] World Health Organization. Early detection, assessment and response to acute public health events. Geneva: WHO; 2014.

[23] Uzor D. Behavioral economics-informed frameworks increasing sustained adherence to infection prevention protocols within complex hospital workflows. International Journal of Research in Medical Science. 2020;2(2):22-33. doi:10.33545/26648733.2020.v2.i2a.186.

[24] Dankwa-Mullan I, Scheufele EL, Matheny ME, Quintana Y, Chapman WW, Jackson G, South BR. A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle. Journal of Health Care for the Poor and Underserved. 2021;32(2):300-17.

[25] Gado P, Oparah OS, Ezeh FE, Gbaraba SV, Adeleke AS, Omotayo O. Framework for Developing Data-Driven Nutrition Interventions Targeting High-Risk Low-Income Communities Nationwide. Framework. 2020 Jul;1(3).

[26] Burger M. The risk to population health equity posed by automated decision systems: A narrative review. arXiv preprint arXiv:2001.06615. 2020 Jan 18.

[27] Du M, Sai A, Kong N. A data-driven optimization approach for multi-period resource allocation in cholera outbreak control. European Journal of Operational Research. 2021 Jun 16;291(3):1106-16.

[28] Uddoh J, Ajiga D, Okare BP, Aduloju TD. Next-Generation Business Intelligence Systems for Streamlining Decision Cycles in Government Health Infrastructure. Journal of Frontiers in Multidisciplinary Research. 2021 Jan;2(1):303-11.

[29] O 'Neil S, Taylor S, Sivasankaran A. Data equity to advance health and health equity in low-and middle-income countries: a scoping review. Digital Health. 2021 Dec;7:20552076211061922.

[30] Brakefield WS, Ammar N, Olusanya OA, Shaban-Nejad A. An urban population health observatory system to support COVID-19 pandemic preparedness, response, and management: design and development study. JMIR Public Health and Surveillance. 2021 Jun 16;7(6):e28269.

[31] Naudé W, Vinuesa R. Data, global development, and COVID-19: Lessons and consequences. WIDER Working Paper; 2020.

[32] Majeed A, Hwang SO. Data-driven analytics leveraging artificial intelligence in the era of COVID-19: an insightful review of recent developments. Symmetry. 2021 Dec 23;14(1):16.

[33] Yang H, Zhang S, Liu R, Krall A, Wang Y, Ventura M, Deflitch C. Epidemic informatics and control: A review from system informatics to epidemic response and risk management in public health. InINFORMS International Conference on Service Science 2020 Dec 19 (pp. 1-58). Cham: Springer International Publishing.

[34] Schwalbe N, Wahl B. Artificial intelligence and the future of global health. The Lancet. 2020 May 16;395(10236):1579-86.

[35] Bertsimas D, Boussioux L, Cory-Wright R, Delarue A, Digalakis V, Jacquillat A, Kitane DL, Lukin G, Li M, Mingardi L, Nohadani O. From predictions to prescriptions: A data-driven response to COVID-19. Health care management science. 2021 Jun;24(2):253-72.

[36] Tariq N. THE USE OF BIG DATA IN PREDICTING AND MANAGING DISEASE OUTBREAKS. International journal of applied biology and forensic. 2020 Oct 31;4(4):278-94.

[37] Yu S, Qing Q, Zhang C, Shehzad A, Oatley G, Xia F. Data-driven decision-making in COVID-19 response: A survey. IEEE Transactions on Computational Social Systems. 2021 May 28;8(4):1016-29.

[38] Adiga A, Chen J, Marathe M, Mortveit H, Venkatramanan S, Vullikanti A. Data-driven modeling for different stages of pandemic response. Journal of the Indian Institute of Science. 2020 Oct;100(4):901-15.

[39] Subbhuraam V, Olatinwo I. Predictive analytics in public health surveillance. InPredictive Analytics in Healthcare, Volume 1: Transforming the future of medicine 2021 Dec 1 (pp. 2-1). Bristol, UK: IOP Publishing.

[40] Bitetto A, Cerchiello P, Mertzanis C. A data-driven approach to measuring epidemiological susceptibility risk around the world. Scientific Reports. 2021 Dec 15;11(1):24037.

[41] Johnson WG. Using precision public health to manage climate change: opportunities, challenges, and health justice. The Journal of Law, Medicine & Ethics. 2020 Dec;48(4):681-93.

[42] Yang H, Zhang S, Liu R, Krall A, Wang Y, Ventura M, Deflitch C. Epidemic informatics and control: A holistic approach from system informatics to epidemic response and risk management in public health. InAI and Analytics for Public Health-Proceedings of the 2020 INFORMS International Conference on Service Science 2021 (pp. 1-46). Berlin/Heidelberg, Germany: Springer.

[43] Chianumba EC, Ikhalea NU, Mustapha AY, Forkuo AY, Osamika DA. A conceptual framework for leveraging big data and AI in enhancing healthcare delivery and public health policy. IRE Journals. 2021 Dec;5(6):303-10.

[44] Alamo T, Reina DG, Millán P. Data-driven methods to monitor, model, forecast and control covid-19 pandemic: Leveraging data science, epidemiology and control theory. arXiv preprint arXiv:2006.01731. 2020 Jun 1.

[45] Asokan GV, Mohammed MY. Harnessing big data to strengthen evidence-informed precise public health response. InBig data in psychiatry# x0026; neurology 2021 Jan 1 (pp. 325-337). Academic Press.