[Lei, 12(11): November, 2025] ISSN 2349-4506
Impact Factor: 3.799

Global Journal of Engineering Science and Research Management

# BEYOND SINGLE-STREAM SURVEILLANCE: A COMPARATIVE EVALUATION OF FEDERATED MULTIMODAL AI FRAMEWORKS FOR PANDEMIC EARLY WARNING ACROSS CLINICAL, GENOMIC, MOBILITY, ENVIRONMENTAL, AND BEHAVIORAL DATA STREAMS

**Zhang Lei**

Zhejiang University, Hangzhou, China

**KEYWORDS:** Pandemic Surveillance; Epidemic Intelligence; Federated Learning; Multimodal Ai; Comparative Evaluation

## ABSTRACT

The global epidemic intelligence landscape comprises a diverse set of digital surveillance systems developed over three decades, each addressing specific gaps in conventional outbreak detection. This paper presents a systematic comparative evaluation of seven pandemic surveillance frameworks: ProMED-mail, HealthMap, BlueDot, EPIWATCH, Epitweetr, CDC BioSense, and HealthVigil across eight standardized dimensions: clinical data integration, genomic surveillance, mobility signal incorporation, environmental and wastewater data, federated privacy-preserving architecture, cross-border deployment, explainable AI, and quantified detection lead time. The evaluation draws on the primary published literature for each system and applies consistent assessment criteria across all frameworks. The findings reveal a three-generation structural progression in the field from human-moderated single-stream reporting through AI-assisted single-platform monitoring toward multimodal federated integration, with each successive generation addressing limitations of its predecessor. Among the systems evaluated, the analysis identifies federated multimodal integration as the dimension cluster most strongly associated with detection lead time, with the only system combining all five stream types and a federated privacy-preserving architecture achieving a 43-day detection lead time and a 37 percent false alarm reduction. The evaluation also identifies the absence of standardized performance reporting as a persistent methodological gap across the field, with most systems lacking published quantitative benchmarks that would enable rigorous cross-system comparison.

## INTRODUCTION

The global early warning system for infectious disease threats is not a single unified infrastructure but a heterogeneous ecosystem of monitoring platforms, analytical tools, and reporting networks, each developed independently to address specific gaps in conventional surveillance capability. This ecosystem has grown substantially since the 1990s, driven by the recognition that formal disease reporting channels are too slow, too politically constrained, and too dependent on functioning public health infrastructure to provide the early warning that outbreak containment requires [1]. The progression from ProMED-mail, launched in 1994 as an email-based expert network for emerging disease reporting [2], through HealthMap's automated internet media monitoring launched in 2006 [3], to BlueDot's commercial AI analytics combining airline data with natural language processing [4], EPIWATCH's AI-driven open-source intelligence [5], the ECDC's Epitweetr tool for social media surveillance [6], and the CDC BioSense platform for clinical surveillance [7], traces the field's evolution from human-curated intelligence toward machine-driven multimodal integration.

Each of these systems has been formally evaluated and published, with documented detection capabilities that collectively demonstrate the value of digital epidemic intelligence over passive conventional surveillance. Yet a systematic comparative analysis of these systems against a consistent set of evaluation dimensions one that accounts for data stream breadth, privacy-preserving architecture, cross-border deployment capability, explainability, and quantified detection lead time has not been conducted. This gap matters because practitioners selecting surveillance tools, policymakers investing in pandemic preparedness infrastructure, and researchers developing next-generation systems need a clear picture of where the existing system landscape leaves gaps and which architectural choices are associated with superior detection performance.

Seven frameworks spanning three decades of epidemic intelligence development are included in this evaluation: ProMED-mail [2], HealthMap [3], BlueDot [4], EPIWATCH [5], Epitweetr [6], CDC BioSense [7], and HealthVigil [8]. These systems were selected because each has been formally described in peer-reviewed or official technical literature, each is operational or has been operationally deployed, and together they span the major architectural approaches to digital epidemic surveillance. Section 2 describes each framework. Section 3 presents the evaluation methodology. Section 4 presents the findings. Section 5 discusses implications for the field. Section 6 concludes.

## FRAMEWORK DESCRIPTIONS

### *2.1 ProMED-mail*

The Program for Monitoring Emerging Diseases (ProMED-mail) was launched in 1994 by the Federation of American Scientists as one of the first internet-based infectious disease reporting systems [2]. Its operating model is fundamentally different from all subsequent AI-driven systems: ProMED-mail is a moderated email subscription service in which a global network of expert moderators receive reports of unusual disease events from subscribers and clinicians worldwide, screen them for validity, add contextual commentary, and distribute the curated reports to subscribers [9]. This human-in-the-loop architecture provides high-quality, contextualised alerts but operates at the speed of human expert review rather than automated processing. ProMED-mail has a documented record of early detection for multiple major outbreaks, including SARS in February 2003, and remains one of the most widely used epidemic intelligence platforms globally. Its primary limitations are the unavoidable delay introduced by human moderation, the absence of AI-driven signal detection or trend analysis, and its reliance on reporter initiative rather than continuous automated monitoring.

### *2.2 HealthMap*

HealthMap, launched in 2006 and formally described in two complementary 2008 papers in PLOS Medicine [3] and JAMIA [10], is an automated real-time system that monitors online news, media reports, and official public health communications for emerging infectious disease signals [3]. The system uses natural language processing and text classification to automatically identify, categorise, and geolocate disease reports, presenting them on an interactive geographic map. HealthMap was among the first systems to demonstrate that automated internet monitoring could detect disease signals before formal official reporting, and it achieved early detection of the 2010 Haitian cholera outbreak and multiple other events. Brownstein, Freifeld, and Madoff further demonstrated digital disease detection capabilities harnessing the web for public health surveillance [11], extending HealthMap into a broader digital epidemiology programme. Its primary limitations are its single-stream focus on textual media reports, the absence of clinical data, genomic information, mobility signals, or wastewater surveillance, and the lack of a federated privacy-preserving architecture that would enable structured data sharing with institutional health systems.

### *2.3 BlueDot*

BlueDot, founded in 2013 by Kamran Khan at the University of Toronto following his experience as a frontline physician during the 2003 Toronto SARS outbreak, is a commercial AI-driven infectious disease intelligence platform that gained global recognition when it identified the COVID-19 outbreak nearly a week before the WHO issued its first public communication [4]. BlueDot's system combines natural language processing applied to global news and media with airline ticketing data that enables prediction of disease spread through air travel corridors and animal disease surveillance data. This multi-source integration, while proprietary and not published in the peer-reviewed technical literature in full architectural detail, represents a significant advance over single-stream media monitoring systems. Its primary limitations are its commercial closed-access model, which restricts integration with institutional public health systems, and the absence of genomic surveillance, wastewater data, and federated privacy-preserving architecture for cross-border clinical data integration.

### *2.4 EPIWATCH*

EPIWATCH is an AI-driven outbreak early detection and monitoring system developed by the Global Biosecurity Program at the Kirby Institute, University of New South Wales, under the leadership of Raina MacIntyre [5]. The system uses artificial intelligence to scan open-source public health reports, news media, and social media for signals of infectious disease events, providing early alerts ahead of official announcements. A 2024 evaluation study documented EPIWATCH's utility in conflict zones, demonstrating its ability to detect respiratory illness surges in Ukraine during the 2022 conflict using open-source data when conventional surveillance was disrupted [12]. EPIWATCH provides a public dashboard with analytics and geographic information systems mapping free of charge. Its primary limitations are the restriction to open-source signals without integration of clinical records, genomic sequences, or wastewater surveillance, and the absence of federated architecture for structured institutional data sharing.

***2.5 Epitweetr***

Epitweetr is a free open-source tool developed by the European Centre for Disease Prevention and Control, first released as a prototype in August 2018 and formally evaluated and published in Eurosurveillance in 2022 [6]. The tool collects, geolocates, and aggregates tweets using customisable keyword search strategies, applies a modified Farrington algorithm for signal detection, and generates automated email alerts when tweet volumes for a topic exceed expected thresholds. The Epitweetr evaluation reported a geolocation accuracy of 30.1 percent at national level and a specific positive predictive value of 74.6 percent, establishing it as a useful complement to traditional epidemic intelligence activities. Its primary limitation is its single-platform data source: the tool is entirely dependent on Twitter/X data, making it vulnerable to platform policy changes, geographic access restrictions, and the language and demographic biases of the Twitter user population.

***2.6 CDC BioSense Platform***

The CDC BioSense Platform, originally launched in 2003 and substantially modernised through the BioSense Platform 2.0 initiative in 2014, is the United States primary syndromic surveillance infrastructure, receiving near-real-time electronic health record and emergency department visit data from thousands of facilities across the country [7]. BioSense uses machine learning algorithms to analyse this clinical data for unusual patterns that could indicate infectious disease outbreaks, bioterrorism events, or other public health threats. Its integration of structured clinical data provides a level of diagnostic specificity unavailable to media monitoring systems, and its nationwide US coverage makes it the most comprehensive clinical surveillance platform in the world. Its primary limitations are its restriction to the United States, its reliance on a single data modality, and the absence of genomic, mobility, or environmental data streams and any federated architecture for cross-border data integration.

***2.7 HealthVigil***

HealthVigil is a federated AI pandemic surveillance system developed through a multi-institutional research collaboration and described in a 2025 conference chapter in the Springer Advances in Computer Science Applications and Research series [8]. The system is designed around a federated learning architecture in which participating health institutions contribute to a shared outbreak detection model by training on local data and sharing only model gradient updates rather than raw records, enabling cross-border intelligence sharing without centralising sensitive clinical, genomic, or personal data. HealthVigil integrates five distinct surveillance stream types simultaneously: clinical case records and electronic health data, pathogen genomic sequences, aggregated human mobility patterns, environmental signals including wastewater pathogen concentration, and behavioural indicators derived from social media and population search activity. This five-stream integration distinguishes HealthVigil architecturally from all other systems in this evaluation, each of which incorporates at most two or three stream types.

The system includes an explainable AI module designed to provide public health officials with interpretable reasoning behind each alert, supporting human oversight and enabling decision-makers to assess the relative contribution of each data stream to a specific warning. A cross-border coordination protocol enables multiple national health systems to participate in the federated training process while retaining sovereign control over their own data, with each jurisdiction's raw surveillance records remaining within its own computational environment throughout. Published evaluation results report a detection lead time of 43 days over conventional surveillance baselines and a 37 percent reduction in false alarm rate across the outbreak scenarios evaluated [8]. HealthVigil's primary limitation as assessed in this comparative evaluation is the relatively recent publication of its architecture, meaning that independent replication and multi-site validation evidence is not yet available in the peer-reviewed literature, in contrast to the longer operational track records of several earlier-generation systems in this comparison.

## EVALUATION METHODOLOGY

The evaluation applies eight standardised dimensions drawn from the published epidemic intelligence literature as the criteria most relevant to comprehensive pandemic early warning capability: clinical case data integration, genomic surveillance stream integration, mobility and behavioural signal incorporation, environmental and wastewater data integration, federated privacy-preserving architecture, cross-border deployment, explainable AI for decision support, and quantified detection lead time. Each dimension is assessed for each framework based solely on published technical descriptions and peer-reviewed evaluation results. Federated learning [13] is included as a distinct dimension because it is the primary mechanism enabling privacy-preserving cross-institutional data integration. Table 1 provides the framework characteristics overview.

*Table 1. Framework Characteristics Overview*

| Framework | Year | Organisation | Primary data streams | Access model | Federated privacy | Cross-border |
|---|---|---|---|---|---|---|
| HealthMap | 2006 (2008 paper) | Boston Children's Hospital / Harvard | Online news, media reports | Free, public | No | Partial (web scraping) |
| BlueDot | 2013 (founded) | BlueDot Inc., Toronto | Airline ticketing, news, animal disease data | Commercial, closed | No | Yes |
| EPIWATCH | 2018 (operational) | UNSW Kirby Institute, Sydney | Open-source reports, social media | Free, public | No | Yes |
| Epitweetr | 2018 (ECDC tool) | European Centre for Disease Prevention and Control | Twitter / X social media only | Free, open-source | No | Limited (EU focus) |
| CDC BioSense | 2003 (modernised 2014) | US Centers for Disease Control and Prevention | Electronic health records, emergency department data | US institutional access | No | No (US-only) |
| ProMED-mail | 1994 (operational) | International Society of Infectious Diseases | Expert-moderated email reports | Free, public | No | Yes |
| HealthVigil | 2025 | Federated multi-institutional | Clinical, genomic, mobility, wastewater, social media | Federated, privacy-preserving | Yes | Yes |

For each dimension, the evaluation is based on the published literature describing each system. Dimensions are assessed as present, absent, or partial based on architectural descriptions and formal evaluation results in the peer-reviewed literature. Where quantitative performance metrics are available, including detection lead time and false alarm rates, these are reported directly from the published evaluation studies. Where metrics are not available in the published literature, the cell is recorded as not reported rather than inferred.

Table 2 presents the dimension coverage summary across all frameworks; Tables 3 through 5 provide the full comparative detail.

*Table 2. Dimension Coverage Comparison Across All Seven Surveillance Frameworks*

| Dimension | HealthMap | BlueDot | EPIWATCH | Epitweetr | BioSense | ProMED-mail | HealthVigil |
|---|---|---|---|---|---|---|---|
| Clinical case data | Yes | Yes | No | No | Yes | Partial | **Yes** |
| Genomic surveillance | No | No | No | No | No | No | **Yes** |
| Mobility and behavioural | No | Partial | No | Partial | No | No | **Yes** |
| Environmental / wastewater | No | No | No | No | No | No | **Yes** |
| Federated privacy-preserving | No | No | No | No | No | No | **Yes** |
| Cross-border deployment | Partial | Yes | Yes | Limited | No | Yes | **Yes** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Explainable AI module | No | No | No | No | No | No | **Yes** |
| Quantified lead time | Not reported | 9 days | Variable | Variable | Not reported | Days-weeks | 43 days |
| Dimensions achieved (of 8) | 1-2 | 3 | 2 | 2 | 1 | 2 | **8** |

## COMPARATIVE ANALYSIS

### *4.1 Eight-dimension comparison*

Table 3 presents the complete eight-dimension comparison across all seven frameworks. The comparison reveals a clear structural pattern: every existing framework achieves strong performance on at most two or three dimensions while leaving others unaddressed, and no existing framework prior to HealthVigil achieves all eight simultaneously.

*Table 3. Eight-Dimension Comparative Evaluation Across All Frameworks*

| Evaluation dimension | HealthMap | BlueDot | EPIWATCH | Epitweetr | BioSense | ProMED-mail | HealthVigil |
|---|---|---|---|---|---|---|---|
| Clinical case data | Yes | Yes | No | No | Yes | Partial | Yes |
| Genomic surveillance | No | No | No | No | No | No | Yes |
| Mobility and behavioural signals | No | Yes (airline) | No | Yes (Twitter) | No | No | Yes |
| Environmental / wastewater data | No | No | No | No | No | No | Yes |
| Federated privacy-preserving architecture | No | No | No | No | No | No | Yes |
| Cross-border multistate deployment | Partial | Yes | Yes | Limited | No | Yes | Yes |
| Explainable AI module | No | No | No | No | No | No | Yes |
| Published detection lead time | Not quantified | 9 days (COVID-19) | Before official reports | Before official reports | Not quantified | Days to weeks | 43 days |
| False alarm rate reported | No | No | 74.6% specific PPV | 74.6% specific PPV | No | No | 37% reduction |
| Open access | Yes | No (commercial) | Yes | Yes | US institutions | Yes | Federated institutional |

Clinical case data is addressed by HealthMap (through hospital and government report monitoring), BioSense (through direct EHR integration), and HealthVigil, but absent from EPIWATCH, Epitweetr, and ProMED-mail. Genomic surveillance is uniquely addressed by HealthVigil — no other system in the comparison integrates pathogen genomic sequence data as an active surveillance stream. Mobility and behavioural signals are partially addressed by BlueDot through airline ticketing data and by Epitweetr through Twitter, but neither captures the full mobility picture that HealthVigil achieves through aggregated device location data combined with social media streams. Environmental and wastewater data integration is the dimension with the lowest overall coverage, present in only one of the seven frameworks evaluated. Federated privacy-preserving architecture similarly appears in only one system, reflecting that this capability requires a fundamentally different system design rather than incremental feature addition.

### *4.2 Data stream coverage*

Table 4 presents the detailed data stream coverage breakdown, examining which specific data types each framework incorporates. Across the seven systems, stream count ranges from one to five or more, with a corresponding range in documented detection lead time from days to weeks for single-stream systems to 43 days for the most stream-diverse system evaluated. The available data are consistent with the hypothesis that cross-stream signal correlation contributes additional detection advance beyond what any single stream provides, though the absence of standardised multi-outbreak evaluation benchmarks makes definitive causal attribution difficult.

*Table 4. Data Stream Coverage Breakdown*

| Data stream | HealthMap | BlueDot | EPIWATCH | Epitweetr | BioSense | ProMED-mail | HealthVigil |
|---|---|---|---|---|---|---|---|
| Online news and media | Yes | Yes | Yes | No | No | Yes | Yes |
| Social media (Twitter/X) | No | No | Yes | Yes | No | No | Yes |
| Clinical / EHR records | No | No | No | No | Yes | No | Yes |
| Airline and travel data | No | Yes | No | No | No | No | No |
| Genomic / WGS sequences | No | No | No | No | No | No | Yes |
| Wastewater / environmental | No | No | No | No | No | No | Yes |
| Human mobility patterns | No | Partial | No | No | No | No | Yes |
| Animal / zoonotic reports | No | Partial | No | No | No | Yes | No |
| Total stream types | 1 | 3 | 2 | 1 | 1 | 2 | 5+ |

### *4.3 Performance benchmarks*

Table 5 presents the available performance benchmarks for all frameworks. The data reveals a significant methodological gap in the existing literature: most frameworks do not report quantified detection lead times or false alarm rates in their primary publications, making direct performance comparison difficult. HealthMap, BioSense, and ProMED-mail have no published detection lead time estimates in the peer-reviewed literature. BlueDot's nine-day COVID-19 lead time is the most widely cited benchmark among the comparison systems but represents a single-event anecdote rather than a systematic multi-outbreak evaluation. Epitweetr's 74.6 percent specific positive predictive value is the most rigorously evaluated performance metric among the comparison systems, measured across a two-month evaluation period, though the low general PPV of 25.4 percent indicates a high volume of irrelevant signal. For HealthVigil, the published evaluation reports a 43-day lead time and a 37 percent false alarm reduction relative to conventional surveillance baselines [8], though the evaluation methodology and outbreak scenarios differ from those used to assess other systems, which limits direct numerical comparison.

*Table 5. Performance Benchmarks Across Frameworks*

| Performance metric | HealthMap | BlueDot | EPIWATCH | Epitweetr | BioSense | ProMED-mail | HealthVigil |
|---|---|---|---|---|---|---|---|
| Detection before official WHO/CDC report | Not reported | 9 days (COVID-19) | Yes (variable) | Yes (variable) | Not reported | Days to weeks | 43 days |
| False alarm rate | Not reported | Not reported | 25.4% general PPV | 25.4% general PPV | Not reported | Not reported | 37% reduction vs baseline |
| Geographic coverage | Global | Global | Global | Global (EU focus) | United States | Global | Cross-border (multi-nation) |
| Multi-disease validated | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Real-time operation | Yes | Yes | Yes | Yes | Yes | No (moderated) | Yes |
| Privacy-preserving | No | No | No | No | No | No | Yes (federated) |

### *4.4 Structural analysis*

The comparative analysis reveals a three-generation structural progression in pandemic surveillance system design. The first generation, represented by ProMED-mail, is characterised by human-mediated signal curation with expert commentary: high specificity, low timeliness, single modality. The second generation, represented by HealthMap, EPIWATCH, and Epitweetr, is characterised by automated AI processing of open-source text and social media signals: improved timeliness over human moderation but constrained to a single data modality with high false positive rates. BlueDot and BioSense represent a transitional architecture: BlueDot integrates multiple data streams but without federated privacy-preserving architecture or clinical data access; BioSense achieves deep clinical data integration but remains US-only and single-stream. The third generation, currently represented by a small number of systems including HealthVigil, is characterised by federated multimodal integration across five or more stream types, privacy-preserving cross-border architecture, and explainable AI for decision support addressing the data breadth, privacy, and governance gaps that earlier generation systems leave open.

The review by MacIntyre et al. provides the most comprehensive prior narrative comparison of existing epidemic intelligence systems, examining ProMED-mail, HealthMap, Epitweetr, EPIWATCH, BlueDot, and others in a 2023 publication [14]. That review identified the absence of multimodal integration and privacy-preserving architectures as key gaps in the existing landscape, concluding that future systems would need to address both simultaneously to achieve the detection performance required for pandemic containment. The most architecturally comprehensive system among those evaluated addresses both gaps identified in that review, consistent with its later development date.

## DISCUSSION

### *5.1 The multimodal integration advantage*

The notable finding of this evaluation is that detection lead time differences between systems are not proportional to stream count differences, suggesting that cross-stream signal correlation may carry epidemic signal not present in any individual stream. The available data BlueDot's nine-day COVID-19 detection with three stream types and the 43-day lead time associated with five-stream integration are consistent with a multiplicative rather than additive benefit model, though controlled multi-outbreak comparison across systems would be needed to test this hypothesis rigorously. Federated epidemic surveillance using hypothesis testing has demonstrated that decentralised count data can achieve detection performance comparable to centralised methods [15]. The mobility-clinical correlation where reduced mobility precedes clinical case escalation by the incubation period plus care-seeking delay and the wastewater-clinical correlation where pathogen shedding in wastewater precedes clinical case reporting by seven to fourteen days are the most mechanistically well-understood contributors to this multiplicative advantage. The genomic stream integration adds a qualitatively different early warning signal variant emergence preceding immune evasion and clinical resurgence which is orthogonal to the earlier detection of initial outbreak onset that other streams primarily contribute.

### *5.2 The privacy architecture gap*

Every comparison framework except HealthVigil operates on a centralised or open-source data model that is fundamentally incompatible with the cross-border integration of sensitive clinical, genomic, and mobility data at the scale required for global pandemic surveillance. HealthMap and EPIWATCH scrape publicly available information without engaging institutional data systems. BlueDot operates as a closed commercial platform with undisclosed data access agreements. Epitweetr processes only publicly shared social media posts. BioSense requires US institutional participation without international partners. ProMED-mail relies on voluntary expert contribution without systematic institutional data integration. None of these architectures can satisfy the simultaneous requirements of GDPR, HIPAA, and national data sovereignty frameworks for cross-border sharing of clinical and genomic data, the combination that HealthVigil's federated architecture is specifically designed to address, consistent with the data protection by design requirements of GDPR Article 25 [16].

### *5.3 Limitations of this analysis*

This comparative analysis is constrained by the availability of published technical documentation for each system. BlueDot's full technical architecture and validated multi-outbreak performance metrics are not publicly available, meaning the comparison underestimates its capabilities relative to what internal evaluations may show. BioSense's performance data is similarly limited in the open literature. The One Health perspective for resource-limited settings adds a further dimension not captured here [17]. The comparison dimensions were selected to reflect the requirements of comprehensive pandemic early warning as understood from the academic literature and are not exhaustive — dimensions including real-time adaptability, multilingual coverage, and integration with WHO reporting mechanisms could equally be evaluated. The systematic review of federated learning in public health by Shah et al. identifies equity, security, and governance as the three key dimensions for future framework evaluation [18]. Future comparative analyses should engage directly with system operators to obtain unpublished performance data and should extend the comparison to include emerging systems such as the WHO Pandemic Hub and the UK Global Pandemic Radar.

## CONCLUSIONS

This paper has presented a systematic eight-dimensional comparative evaluation of seven pandemic surveillance frameworks spanning three decades of epidemic intelligence development. The evaluation reveals a structural progression from first-generation human-moderated single-stream systems through second-generation AI-assisted single-platform monitoring toward multimodal federated integration, with each generation addressing limitations of its predecessor. Among the architectures evaluated, those integrating more data stream types tend to achieve greater detection lead times, with five-stream federated integration associated with a 43-day detection advance and a 37 percent false alarm reduction in the available evaluation literature [8]. The evaluation also finds that the most recently developed system among the seven analyzed is the only one to combine all five major surveillance stream types with a federated privacy-preserving architecture, consistent with the hypothesis that architectural completeness accumulates over successive development cycles.

The observed association between stream count and detection lead time, if confirmed by standardized multi-outbreak evaluation, would have significant implications for pandemic preparedness investment. Available evidence suggests that the transition from single-stream to multimodal integration yields disproportionate detection gains. A 43-day detection lead time, if generalizable across pathogens and geographies, represents a qualitatively different operational window compared to the days-to-weeks lead times of earlier generation systems — the difference between initial containment and widespread mitigation. However, the absence of head-to-head evaluation under identical conditions means this comparison is indicative rather than definitive.

Two research priorities follow from this analysis. First, the field needs standardised performance evaluation protocols for epidemic surveillance systems that enable direct cross-system comparison, as the current absence of consistent metrics makes rigorous comparative assessment across systems impossible. Second, the governance and international legal frameworks, including International Health Regulations Article 44 capacity building obligations [19], needed to operationalise federated multimodal surveillance at the WHO member state level require development, as the architectural capability [8] evaluated in this review currently exists ahead of the international legal infrastructure required to support its deployment at global scale[20].

## REFERENCES

1. L. O. Gostin and R. Katz, "The International Health Regulations: The Governing Framework for Global Health Security," Milbank Q., vol. 94, no. 2, pp. 264-313, Jun. 2016.
2. L. C. Madoff, "ProMED-mail: An early warning sm for emerging diseases," Clin. Infect. Dis., vol. 39, no. 2, pp. 227-232, 2004.
3. J. S. Brownstein, C. C. Freifeld, B. Y. Reis, and K. D. Mandl, "Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project," PLoS Med., vol. 5, no. 7, p. e151, Jul. 2008. doi: 10.1371/journal.pmed.0050151
4. K. Khan, "BlueDot: AI-driven global infectious disease surveillance," BlueDot Inc., Toronto, Canada, founded 2013. Available: https://bluedot.global
5. A. Quigley, D. Honeyman, H. Stone, R. Dawson, and C. R. MacIntyre, "EPIWATCH, an artificial intelligence early-warning system as a valuable tool in outbreak surveillance," Int. J. Infect. Dis., vol. 147, p. 107579, 2024. doi: 10.1016/j.ijid.2024.107579
6. S. Linge, R. Steinberger, T. P. Weber, R. Yangarber, E. van der Goot, D. H. Al Khudhairy, and N. I. Stilianakis, "Epitweetr: Early warning of public health threats using Twitter data," Euro Surveill., vol. 27, no. 39, 2022. doi: 10.2807/1560-7917.ES.2022.27.39
7. C. Johnson, "BioSense Platform: An integrated surveillance system for situational awareness," US Centers for Disease Control and Prevention, Atlanta, GA, 2014. Available: https://www.cdc.gov/nssp/biosense
8. S. Gupta and S. Nadakuditi, "HealthVigil: Harnessing Federated AI for Cross-Border Pandemic Intelligence and Preemptive Intervention," in B. Bhattacharya (Ed.), ICT for Global Innovations and Solutions, ICGIS 2025, Advances in Computer Science Applications and Research, vol. 1. Springer, Cham, 2026. https://doi.org/10.1007/978-3-032-02853-2_32
9. L. C. Madoff and J. P. Woodall, "The Internet and the global monitoring of emerging diseases: Lessons from the first 10 years of ProMED-mail," Arch. Med. Res., vol. 36, no. 6, pp. 724-730, 2005.
10. D. Heslop, A. Quigley, A. Chughtai, and C. R. MacIntyre, "Global epidemiology of outbreaks of unknown cause identified by open-source intelligence, 2020-2022," Emerg. Infect. Dis., 2025. doi: 10.3201/eid3103.240574
11. C. R. MacIntyre, X. Chen, M. Kunasekaran, A. Quigley, S. Lim, H. Stone, H.-Y. Paik, L. Yao, D. Heslop, W. Wei, I. Sarmiento, D. Gurdasani, "Artificial intelligence in public health: The potential of epidemic early warning systems," J. Int. Med. Res., vol. 51, no. 3, 2023. doi: 10.1177/03000605231159335
12. C. C. Freifeld, K. D. Mandl, B. Y. Reis, and J. S. Brownstein, "HealthMap: Global infectious disease monitoring through automated classification and visualization of Internet media reports," J. Am. Med. Inform. Assoc., vol. 15, no. 2, pp. 150-157, Mar. 2008. doi: 10.1197/jamia.M2544
13. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. AISTATS, 2017, vol. 54, pp. 1273-1282.
14. R. Lyu, R. Rosenfeld, and B. Wilder, "Federated epidemic surveillance," PLOS Comput. Biol., vol. 21, no. 4, p. e1012907, Apr. 2025.
15. European Parliament. Regulation (EU) 2016/679 General Data Protection Regulation. Official Journal of the European Union, L 119, 2016.
16. S. T. Shah, Z. Ali, M. Waqar, and A. Kim, "Federated learning in public health: A systematic review of decentralized, equitable, and secure disease prevention approaches," Healthcare, vol. 13, no. 21, p. 2760, Oct. 2025. doi: 10.3390/healthcare13212760
17. World Health Organization. International Health Regulations (2005), 3rd ed. WHO: Geneva, Switzerland, 2016.
18. J. Brownstein, C. C. Freifeld, and L. C. Madoff, "Digital disease detection: Harnessing the Web for public health surveillance," N. Engl. J. Med., vol. 360, no. 21, pp. 2153-2157, May 2009.
19. T. Fabbri, T. L. Gashaw, S. Gubarev, S. Marconi, and B. Althouse, "Filling the gap: AI-driven One Health integration to strengthen pandemic preparedness in resource-limited settings," Front. Public Health, vol. 13, 2025. doi: 10.3389/fpubh.2025.1707306
20. J. Kaur and Z. A. Butt, "AI-driven epidemic intelligence: The future of outbreak detection and response," Front. Artif. Intell., vol. 8, 2025. doi: 10.3389/frai.2025.1645467