Open Access Library Journal
2026, Volume 13, e14831
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Artificial Intelligence-Driven Approaches to Disease Surveillance and Outbreak Prediction

Enaam Rudwan[1], Heba Elbashir Mohamed Ali[2], Maab Ibrahim[3], Hazim Mohamedelnour[4], Youser Sami Almandalawi[1], Malaz Kamal Mohamed Hassan[5], Leena Moaaz Hashim Mohamedain[6], Leena Saeed[7], Ahmed Rudwan[8]

[1]Obstetrics and Gynaecology, Hamad Medical Corporation, Doha, Qatar
[2]Medicine and Surgery, Ahfad University for Women, Khartoum, Sudan
[3]Medical Relations and Treatment Abroad, Ministry of Public Health, Doha, Qatar
[4]Pharmaceutical Regulatory Compliance, Ministry of Public Health, Doha, Qatar
[5]Medicine and Surgery, The National Ribat University, Khartoum, Sudan
[6]Medical Research Center, Hamad Medical Corporation, Doha, Qatar
[7]Medical Commission, Ministry of Public Health, Doha, Qatar
[8]Cardiology Department, Hamad Medical Corporation, Doha, Qatar
Email: Leenasaeed95@hotmail.com

How to cite this paper: Rudwan, E., Ali, H.E.M., Ibrahim, M., Mohamedelnour, H., Almandalawi, Y.S., Hassan, M.K.M., Mohamedain, L.M.H., Saeed, L. and Rudwan, A. (2026) Artificial Intelligence-Driven Approaches to Disease Surveillance and Outbreak Prediction. *Open Access Library Journal*, **13**: e14831.
https://doi.org/10.4236/oalib.1114831

Received: January 1, 2026
Accepted: January 31, 2026
Published: February 2, 2026

Abstract

Artificial Intelligence (AI) has emerged as a promising approach to enhance disease surveillance and support outbreak prediction in public health. Conventional surveillance systems, while foundational, are often limited by reporting delays, under-detection, and challenges in handling large and complex data streams. Recent advances in AI including machine learning, natural language processing, and deep learning offer new opportunities to address these limitations by enabling automated case detection, syndromic surveillance, real-time anomaly detection, and predictive modeling. This review synthesizes current evidence on AI-driven approaches to disease surveillance and outbreak prediction, focusing on methodological frameworks, data sources, and applications across infectious disease contexts. Key AI-based surveillance strategies, outbreak prediction models, and forecasting techniques are discussed alongside emerging data sources such as electronic health records, environmental data, mobility data, and digital media. The review also highlights challenges related to data quality, interpretability, ethical considerations, and integration with traditional surveillance systems. By summarizing existing knowledge, this review aims to inform future research and support the responsible adoption of AI technologies in public health surveillance and outbreak preparedness.

Subject Areas

Artificial Intelligence

Keywords

Artificial Intelligence, Disease Surveillance, Outbreak Prediction, Traditional Surveillance

1. Introduction

A disease outbreak is defined as the occurrence of disease cases exceeding the level normally expected within a specific community, geographical area, or time period. Such outbreaks are sustained by infectious agents transmitted through direct human-to-human contact, exposure to animal reservoirs or environmental sources, or via insect or animal vectors. Human behavioural factors frequently contribute to the transmission and amplification of these events. Timely detection and reporting of outbreaks are essential to reducing their adverse public health, social, and economic consequences.

Epidemic-prone diseases, including both emerging and re-emerging infections, represent a major threat to public health security and pose substantial risks to social and economic stability. From 1996 to 2023, over 3000 infectious disease outbreaks were documented globally, with the Democratic Republic of Congo, China, and Saudi Arabia reporting the highest frequencies. Influenza was the most reported outbreak, followed by Ebola and Middle East respiratory syndrome-related coronavirus. In 2023, dengue reached unprecedented levels, with approximately five million cases and 5000 deaths worldwide. Fatality rates differed markedly by disease, with the highest mortality observed for Marburg virus (76.86%), followed by haemorrhagic fevers (63.63%) and Ebola (63.00%). These trends highlight the broad geographic spread and variable severity of infectious disease outbreaks, underscoring the need for robust global surveillance systems and targeted public health interventions [1]. Over the past two decades, countries of the Eastern Mediterranean region have experienced a notable increase in the frequency of outbreaks and pandemics caused by emerging and re-emerging diseases, including Alkhurma haemorrhagic fever, chikungunya, cholera, dengue, avian influenza A/H5N1, influenza A/H1N1 2009 pandemic, and Rift Valley fever. These challenges have been further intensified by ongoing acute and chronic humanitarian crises affecting many countries in the region [2].

Public health surveillance is the continuous and systematic collection, analysis, and interpretation of health-related data that are fundamental to the planning, implementation, and evaluation of public health practice, together with the timely dissemination of this information to authorities responsible for disease prevention and control. Its goal is to generate information that supports health-related actions by public health professionals, policymakers, and the public, thereby informing the development and implementation of public health policies and programs. Public health surveillance is used to identify affected individuals and their contacts for appropriate treatment and intervention to detect epidemics, emerging health problems, and changes in health-related behaviours to estimate the magnitude and scope of health conditions, measure temporal trends and characterize disease patterns, monitor changes in infectious and environmental agents, evaluate the effectiveness of public health programs and control measures, and to support hypothesis generation and stimulate further research [3].

Traditionally, public health surveillance has been based on manual approaches and epidemiological models derived from mathematical frameworks in epidemiology [4]. These methods rely on the systematic collection and analysis of data from sources such as healthcare facilities, laboratory testing, and case reporting systems. Although effective in many settings, they are constrained by challenges including reporting delays, under-detection of cases, and limited scalability. Moreover, manual approaches have difficulty managing the large and rapidly expanding volumes of data produced in modern, digitally connected environments and lack the capacity to rapidly accommodate confounding and external variables that do not follow predictable or linear patterns [4].

Traditional disease-specific surveillance systems depend on data reported by healthcare facilities and diagnostic laboratories and are based on the collection of structured, predefined information on infectious disease events. However, this indicator-based surveillance approach is often limited by resources, time, and reporting constraints, which can lead to incomplete data, particularly for emerging infectious diseases [5].

AI has the potential to address these challenges by enhancing traditional public health surveillance approaches. AI comprises a set of computational methods, including machine learning, natural language processing, computer vision, and deep learning, developed to support data analysis, prediction, and automation of tasks that have traditionally required human input. Machine learning, as a core component of AI, enables the identification of patterns and generation of predictions from data, while natural language processing facilitates the examination of unstructured textual information from multiple sources. These techniques have demonstrated considerable promise in applications such as disease detection, risk assessment, and outbreak modeling [4].

Technological advancements have increasingly been explored as solutions to the limitations of conventional surveillance systems, with approaches including advanced analytical techniques, AI, the Internet of Things (IoT), remote sensing, and molecular tools that have the potential to strengthen surveillance capacity. These innovations enable real-time monitoring, early detection of anomalies, and enhanced prediction of outbreak risk. Additionally, point-of-care diagnostics, telemedicine, and digital contact tracing tools can substantially improve the timeliness and efficiency of outbreak response efforts [5].

AI and big data have emerged as transformative tools in infectious disease surveillance, enabling the systematic collection and analysis of data related to transmission dynamics, risk factors, and clinical outcomes. These capabilities support the identification of high-risk populations, monitoring of disease spread, and prediction of potential outbreaks, thereby strengthening early intervention efforts and improving the allocation of healthcare resources [5].

The aim of this narrative review is to examine and synthesize existing literature on AI-driven approaches in disease surveillance and outbreak prediction, with particular emphasis on their methodological foundations, applications, and data sources. It also focuses on AI-based methods used in public health surveillance and outbreak prediction, including automated case detection, syndromic surveillance, real-time anomaly detection, and predictive modeling approaches. It covers machine learning-based outbreak prediction models, time-series forecasting methods, and spatial and spatiotemporal analyses. In addition, the review examines the diverse data sources supporting AI-driven surveillance systems, such as electronic health records, environmental and climate data, mobility data, and digital and social media data. The scope also includes a comparison between AI-driven and conventional surveillance approaches, as well as discussion of ethical, legal, and practical challenges associated with the implementation of AI in public health contexts.

2. Artificial Intelligence—Based Disease Detection Methodologies

Automated case detection involves the application of computational methods to identify disease cases using routinely collected electronic health data, with the aim of minimizing reliance on manual case identification.

Studies have shown that automated detection systems using electronic clinical notes can achieve good performance and maintain transferability across healthcare systems in influenza surveillance, supporting their potential application in large-scale surveillance settings [6].

A study by Lake *et al.* [7] examined the application of machine learning methods within an operational syndromic surveillance system. The authors evaluated machine learning as a tool for computer-assisted human decision-making when assessing statistical alarms generated by syndromic surveillance detection algorithms. The study reported that machine learning models could be used to support classification of alarm decision outcomes within the risk assessment process.

Agarwal *et al.* [8] demonstrated the feasibility of real-time anomaly detection in COVID-19 reporting time-series data using a neural network-based framework combined with a binary classification model. The approach was applied to surveillance reporting data from U.S. states, including Pennsylvania and California, and focused on identifying reporting anomalies. The authors discussed several avenues for future work, including extending the method to finer spatial and temporal scales and exploring its potential role in improving forecasting by addressing reporting anomalies.

3. AI-Based Outbreak Prediction and Forecasting

Time-series analysis complements classical epidemiological models by supporting both prediction and forecasting in epidemiological research. Prediction refers to explaining past and current observations based on internal and external influences, whereas forecasting involves estimating possible future values using model predictive ability and hypothesized values of those influences. For straightforward linear models such as Autoregressive Integrated Moving Average (ARIMA), time-series approaches can be easier to implement than classical compartmental models such as SEIR model (Susceptible, Exposed, Infectious, Recovered), although they are limited in forecasting horizon; their primary utility lies in achieving accurate short-term predictions. In its basic form, time-series analysis does not require detailed theoretical knowledge of disease transmission mechanisms and instead estimates relationships directly from observed data, enabling hypothesis testing for factors that may influence epidemic dynamics, including interventions such as school closures and the emergence of viral variants [9]. Applications of time-series analysis include mortality and hospital risk estimation from new cases, seroprevalence studies, assessment of emerging variants, and excess mortality estimation and analysis [9].

Moreover, an important component of epidemiology is Spatial analysis, with its origins tracing back to early work linking geographic features to disease occurrence. The development of Geographic Information Systems (GIS) in the 1990s substantially expanded the scope and application of spatial analysis in public health by enabling spatial analyses and visualization techniques [10]. GIS technology allows visualization of maps and assessment of spatial variability using dynamic information, including temporal information that changes over short periods [11]. These technological advances have encouraged the use of spatiotemporal methodologies that incorporate both time and place in epidemiological research [12].

On the other hand, a comparative study evaluating machine learning approaches for COVID-19 outbreak prediction examined the performance of Multilayer Perceptron (MLP) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models in modeling outbreak time-series data. The study reported that both machine learning models demonstrated higher generalization ability for long-term prediction compared with standard epidemiological approaches. Given the complexity of COVID-19 dynamics and variability across countries, the authors highlighted machine learning as a promising tool for outbreak time-series modeling, while emphasizing that the findings represent an initial benchmarking framework for future research [13].

Zhang *et al.* [14] developed a machine learning-based universal outbreak risk prediction tool intended to predict outbreak risk across multiple diseases and countries. The study used outbreak data from 43 diseases in 206 countries to build a universal risk prediction system. The system employed five machine learning models (Neural Network, XGBoost, Logistic Boost, Random Forest, and Kernel support vector machine (SVM)) that voted together to generate ensemble predictions. Using economic, cultural, social, and epidemiological factors, the model achieved approximately 80% - 90% accuracy, and performance was evaluated using three datasets designed to reflect different realistic situations. The authors reported that the approach provided outbreak risk assessment that is not limited by country borders or disease type and may facilitate rapid response, government decision-making, and international cooperation.

4. Data Sources Supporting AI-Driven Surveillance and Prediction

4.1. Electronic Health Records and Clinical Data

AI has become a central approach in the analysis of Electronic Health Records (EHRs) for public health applications. Machine learning algorithms enable the analysis of large-scale EHRs data to identify patterns and support disease trend prediction. Natural language processing methods further enhance EHRs analysis by extracting relevant information from unstructured clinical text, such as physicians' notes, allowing a more comprehensive understanding of patient health status. In addition, deep learning techniques are increasingly applied to evaluate complex EHRs datasets and generate accurate predictions, including forecasts of patient outcomes such as hospital readmissions. These approaches have potential public health applications, including informing policy decisions and identifying population groups that may benefit from targeted interventions, such as vaccination strategies for vulnerable populations [15].

The integration of AI into EHRs analysis supports improved identification of disease patterns, personalized treatment planning, and early outbreak detection, thereby enabling more informed clinical and public health decision-making. However, the use of AI for EHRs-based public health surveillance faces several challenges. Prediction accuracy depends heavily on the quality, completeness, and standardization of EHRs data, while data heterogeneity including unstructured text, images, and time-series information adds complexity to model development. Data privacy and security are critical concerns, requiring compliance with regulatory frameworks such as the Health Insurance Portability and Accountability Act. Despite these challenges, AI-driven analysis of EHRs holds significant potential to enhance the effectiveness and precision of public health data analysis and contribute to improved health outcomes at both individual and population levels [15].

4.2. Environmental and Climate Data

Environmental data have been incorporated as inputs in machine learning-based infectious disease surveillance systems. In a recent study by Oh *et al.* [16] on food-borne disease surveillance, environmental information was included alongside wastewater microbiome and crowdsourced data within a data-driven machine learning modeling framework. The environmental data consisted of meteorological variables, specifically wind-chill temperature, humidity, and sunshine rate, which were treated as environmental predictors in the model. These variables were integrated as input features during model development and were used in combination with other data streams to support surveillance of foodborne infectious diseases. The study demonstrates the feasibility of incorporating environmental and climate-related variables into machine learning-based surveillance models without isolating their effects from other contributing data sources.

4.3. Mobility and Population Movement Data

Mobility data have been incorporated into machine learning-based frameworks to examine their association with infectious disease dynamics. In a hybrid machine learning study evaluating the COVID-19 pandemic, aggregated population mobility data were integrated with demographic and environmental variables to generate predictions of county-level COVID-19 case trends. The mobility data represented temporal changes in population movement patterns and were analysed in relation to different COVID-19 control phases, including periods of lockdown and reopening. These mobility indicators were used as input features within a machine learning hybrid prediction framework, and the predicted case trends were evaluated against reported COVID-19 case data to examine associations between mobility patterns and variations in case counts over time [17].

4.4. Digital and Social Media Data

Digital and social media data have been used as alternative data sources for infectious disease surveillance. In a review of social media-based influenza surveillance, user-generated textual data from digital platforms, particularly Twitter, were analysed using machine learning and text-mining techniques. The reviewed studies utilized social media posts and user-generated messages to identify influenza-related signals and monitor influenza activity over time. Machine learning classifiers were applied to process unstructured textual data and classify content relevant to influenza-related activity. The review also describes multiple challenges in using social media data for influenza detection and prediction, including restrictions on data collection, large data volume, informal language and spelling variations, heterogeneity of users, sampling bias, limitations in dataset consistency, lack of accurate user location information, difficulties in defining proxy populations, the presence of spam accounts, and challenges in evaluation due to limited ground-truth data beyond CDC Illness-Like Influenza Surveillance Network (ILINet) [18]. Although this approach is considered reliable, it is often constrained by high cost and delayed reporting. Consequently, numerous studies have focused on developing real-time analytical methods to monitor Influenza-Like Illness (ILI) using alternative data sources. Social networking platforms, particularly social media streams such as Twitter, offer timely and large-scale data that can be leveraged to estimate influenza activity within populations. Analysis of these data has shown potential for anticipating trends in influenza transmission and providing early warning signals ahead of traditional surveillance systems [18].

5. AI System Implementation in Infectious Disease Surveillance

AI has become a pivotal component of contemporary infectious disease surveillance, offering advanced analytical capabilities that address the structural limitations of traditional epidemiological systems. Conventional surveillance methods often rely on delayed case reporting, fragmented data sources, and retrospective analysis, which can impede timely public health responses. In contrast, AI-driven systems leverage machine learning, deep learning, and Natural Language Processing (NLP) techniques to analyze large volumes of heterogeneous data, including electronic health records, laboratory results, environmental indicators, population mobility data, and unstructured online information such as news reports and social media content [5] [19] [20]. This capacity to integrate diverse datasets enables earlier detection of anomalous disease patterns and supports near-real-time situational awareness, which is critical for effective outbreak prevention and control.

Empirical evidence demonstrates that AI-enhanced surveillance systems substantially improve early warning and predictive capabilities. Time-series forecasting models, ensemble learning approaches, and recurrent neural networks have been shown to outperform traditional statistical methods in predicting disease incidence and temporal trends [19] [21]. Furthermore, NLP-based event detection systems facilitate automated monitoring of multilingual media and digital sources, allowing for the rapid identification of emerging health threats before formal notification through official surveillance channels [22]. Systematic reviews consistently report that AI applications enhance the sensitivity and timeliness of outbreak detection, particularly when combined with non-traditional data sources such as wastewater surveillance and climate data [5] [20].

AI-based surveillance systems operate across multiple spatial scales, with important distinctions between regional and global implementations. Regional surveillance systems typically focus on localized datasets, including hospital admissions, diagnostic test results, and environmental sensor data, enabling high-resolution detection of outbreaks and targeted public health interventions [19] [21]. These systems are particularly valuable for supporting local decision-making and optimizing resource allocation. In contrast, global AI surveillance platforms aggregate data from multiple countries and regions, incorporating international travel patterns, global media reports, and cross-border epidemiological information to identify transnational disease threats and assess pandemic risk [6]. While global systems provide broader situational awareness, they face challenges related to data heterogeneity, reporting bias, integration difficulties, language diversity, and interoperability across national health systems [5] [20] (**Table 1**).

Beyond general outbreak detection, AI has demonstrated significant value in the surveillance of specific infectious disease categories, particularly vector-borne and viral infections. Vector-borne diseases such as malaria, dengue, Zika, chikungunya, and West Nile virus are strongly influenced by environmental and climatic factors. AI-based models for forecasting disease risk and mapping its geographic distribution use a variety of data sources to enhance prediction accuracy. These sources include healthcare records, online search and behavioral data, environmental and climate information, geographic and human mobility patterns, as well as social media, news reports, and genomic data from pathogens. Additional inputs from environmental sensors and wearable devices can provide further context on conditions that affect disease spread. By combining these diverse datasets, AI systems can better anticipate areas at risk and support timely public health actions [23].

AI models have been widely applied to integrate satellite imagery, meteorological variables, land-use data, and vector ecology information to predict disease risk and geographic distribution [24]. These approaches enable spatial risk mapping and seasonal forecasting, allowing public health authorities to implement proactive vector control strategies in high-risk areas. Studies indicate that AI-based models outperform conventional regression techniques in capturing complex, nonlinear relationships that characterize vector-borne disease transmission dynamics [24] [25]. AI applications have also played a crucial role in the surveillance of viral infections, including influenza, COVID-19, Ebola, and other emerging viral pathogens. In this context, AI-driven systems utilize time-series analysis, mobility data, and NLP techniques to detect early signals of viral spread and estimate transmission dynamics [20] [22]. During the COVID-19 pandemic, AI-powered global surveillance platforms successfully identified abnormal patterns of viral activity through the analysis of digital media and travel data prior to widespread official reporting, underscoring their potential for early pandemic intelligence [22]. Deep learning models have further been used to forecast viral incidence, evaluate the impact of public health interventions, and support strategic planning at both regional and global levels [21].

Table 1. Comparison of regional and global AI-based surveillance systems.

Dimension	Regional AI Surveillance Systems	Global AI Surveillance Systems
Geographic scope	Subnational or national	Multinational or global
Data sources	Local clinical records, laboratories, and environmental sensors	Global media, international reports, travel and mobility data
Primary objective	Early detection of localized outbreaks	Identification of emerging global threats and pandemics
Analytical focus	High-resolution spatial and temporal modeling	Broad pattern recognition across regions
Strengths	Timely local response and targeted interventions	Global situational awareness and preparedness
Key challenges	Limited scalability	Data heterogeneity and interoperability
Examples	AI-assisted urban disease monitoring systems [19] [21]	Global digital surveillance platforms such as HealthMap [22]

Collectively, these applications highlight AI's capacity to shift infectious disease surveillance from a reactive to a proactive paradigm. By enabling predictive risk assessment rather than retrospective analysis, AI supports earlier intervention, targeted prevention strategies, and more efficient use of public health resources [5] [21]. Nevertheless, challenges remain regarding data quality, model interpretability, ethical governance, and equitable representation of vulnerable populations, particularly in low- and middle-income countries where infectious disease burdens are highest [5] [20]. Addressing these challenges will be essential to ensuring that AI-driven surveillance systems are transparent, reliable, and globally inclusive.

6. Comparison of AI-Based and Traditional Surveillance and Prediction Methods

AI-based surveillance and prediction methods represent a substantial progression beyond traditional infectious disease surveillance approaches, which typically depend on passive case reporting, periodic aggregation of clinical data, and conventional statistical analyses. Traditional methods, including syndromic surveillance and compartmental or regression-based models, are frequently limited by reporting delays, incomplete case capture, and restricted ability to represent complex and nonlinear transmission dynamics [26]. In contrast, AI-driven surveillance systems employ machine learning and deep learning algorithms that can process large-scale, high-dimensional data in near real time, enabling earlier outbreak detection and more precise forecasting [20]. Unlike conventional prediction models that rely predominantly on historical case counts, AI methods integrate diverse data streams such as population mobility, environmental and climatic variables, genomic data, and unstructured digital information thereby identifying subtle patterns and early signals of disease emergence [22]. Moreover, AI-based models exhibit greater adaptability, as they can be continuously updated and retrained in response to evolving epidemiological conditions, whereas traditional models often require manual recalibration and are less responsive to rapid changes in transmission dynamics [20]. Nevertheless, AI approaches are most effective when used in conjunction with traditional surveillance systems, which provide validated laboratory and clinical confirmation essential for public health decision-making [26]. This evidence supports the adoption of hybrid surveillance frameworks that combine the robustness of conventional epidemiology with the predictive power of AI to enhance accuracy, timeliness, and preparedness in infectious disease control [22] (**Table 2**).

Table 2. Comparison of AI-based and traditional infectious disease surveillance methods.

Dimension	Traditional Surveillance & Prediction Methods	AI-Based Surveillance & Prediction Methods
Data sources	Clinical reports, laboratory confirmations, historical case counts	Multisource data (clinical, mobility, environmental, genomic, digital media)
Timeliness	Often delayed due to reporting and aggregation processes	Near real-time analysis and early signal detection

Continued

Analytical approach	Statistical models, rule-based systems, linear regression	Machine learning, deep learning, ensemble models
Ability to model complexity	Limited capacity to capture nonlinear interactions	Strong ability to model complex and nonlinear patterns
Adaptability	Manual model updates and recalibration	Continuous learning and dynamic model updating
Prediction accuracy	Moderate, dependent on historical trends	Generally higher, particularly for short-term forecasting
Interpretability	High transparency and ease of interpretation	Variable; may require explainable AI techniques
Role in public health	Provides validated case confirmation and official statistics	Enhances early warning, forecasting, and decision-making support
Limitations	Reporting delays, underreporting, and limited predictive scope	Data quality dependence, potential bias, interpretability challenges

7. Ethical, Legal, and Privacy Considerations in AI-Based Disease Surveillance and Outbreak Prediction

The implementation of artificial intelligence in infectious disease surveillance and outbreak prediction raises significant ethical, legal, and privacy considerations that must be carefully addressed. Ethically, AI systems process large volumes of sensitive health and behavioral data, including clinical records, mobility patterns, and digital traces, which may affect individuals or communities without their explicit consent [20] [27]. Algorithmic bias is another concern; models trained on incomplete or unrepresentative datasets may generate inequitable predictions, disproportionately affecting vulnerable populations [28]. From a legal perspective, AI surveillance must comply with data protection regulations such as the General Data Protection Regulation (GDPR) or national health privacy laws, including rules on data storage, cross-border sharing, and secondary use of information for predictive modelling [27]. Privacy challenges are particularly pronounced in outbreak prediction, where real-time monitoring and integration of multiple data sources could inadvertently expose personally identifiable information. Effective mitigation requires robust technical safeguards, including encryption, anonymization, and privacy-preserving data architectures, alongside governance mechanisms such as ethical review boards, transparent decision-making processes, and public engagement [20] [28]. Overall, while AI can greatly enhance early detection and predictive capacity in disease surveillance, careful attention to ethical principles, legal compliance, and privacy protection is essential to ensure responsible and equitable application of these technologies [20] [27] [28].

8. Challenges and Limitations of AI-Driven Approaches

AI has shown substantial potential in improving disease surveillance and outbreak prediction by enabling rapid analysis of large, heterogeneous datasets to detect early warning signals and forecast epidemic trends. However, the real-world application of AI-driven approaches faces several critical challenges and limitations. One major constraint is data quality and availability, as surveillance data are often incomplete, delayed, biased, or inconsistently reported, particularly in Low- and Middle-Income Countries (LMICs), which undermines model accuracy and generalizability [29]. Additionally, many advanced AI models operate as black-box systems, offering limited interpretability and making it difficult for public health authorities to understand, validate, or trust model outputs for decision-making [20]. To address this limitation, literature has proposed the use of explainable AI techniques, such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), which provide post-hoc insights into feature contributions and improve the interpretability of complex models [30] [31].

Algorithmic bias represents another significant concern, as models trained on non-representative datasets may exacerbate existing health inequities and yield unreliable predictions for marginalized populations [32]. Ethical and legal challenges, including data privacy, informed consent, and governance, further complicate the integration of AI into surveillance systems, especially when sensitive personal, genomic, or mobility data are involved [28] [33]. Furthermore, technical and infrastructural limitations, such as lack of interoperability between health information systems, insufficient computational resources, and absence of standardized data frameworks, hinder scalable and real-time implementation of AI solutions [29] [34].

Finally, excessive reliance on automated predictions without adequate human oversight and multidisciplinary collaboration may result in false alarms or misinterpretation of results, emphasizing the need for cautious validation and integration with traditional epidemiological expertise [20].

9. Future Directions and Research Opportunities

AI is poised to play an increasingly influential role in the future of disease surveillance and outbreak management by addressing long-standing limitations in traditional public health monitoring systems. Conventional surveillance approaches often depend on delayed reporting, fragmented data sources, and manual analysis, which can impede timely outbreak detection. Future research should therefore focus on embedding AI technologies within national surveillance systems in ways that enhance, rather than replace, established epidemiological practices. This includes the development of interoperable AI platforms capable of integrating diverse data streams such as clinical records, laboratory results, environmental indicators, and population mobility data. Ensuring compatibility with national health information infrastructures and public health workflows will be essential for sustainable and effective implementation.

Another critical direction for research is the advancement of explainable artificial intelligence for outbreak prediction. While machine learning and deep learning models can capture complex, non-linear patterns associated with disease transmission, their opaque nature limits their utility in public health decision-making. Explainable AI approaches offer a pathway to improve transparency by revealing how specific variables contribute to predictions and alerts. Future work should aim to design explanation methods that are meaningful to epidemiologists and policymakers, allowing them to assess model reliability, validate predictions against domain knowledge, and justify intervention strategies. Integrating explainability with causal reasoning can further strengthen confidence in AI-generated insights and reduce the risk of misleading conclusions.

The deployment of AI for disease surveillance in LMICs presents distinct research opportunities and challenges. These settings often experience higher vulnerability to infectious disease outbreaks while simultaneously facing constraints related to data quality, technical infrastructure, and human resources. Future research should prioritize the development of AI models that are resilient to incomplete or irregular data and that can function effectively in resource-limited environments. Lightweight algorithms, mobile-based systems, and context-specific model adaptation are particularly relevant for LMICs. Equally important is the emphasis on local capacity development, ensuring that AI systems are governed, maintained, and adapted by domestic institutions to support long-term public health resilience.

Policy and regulatory considerations will be central to the responsible expansion of AI-driven disease surveillance. The use of large-scale health and non-health data raises concerns related to privacy, data ownership, and potential misuse. Research is needed to inform policy frameworks that define ethical standards, accountability mechanisms, and regulatory oversight for AI applications in public health. This includes establishing guidelines for data sharing, algorithmic transparency, and continuous performance evaluation after deployment. Clear and adaptive regulatory structures can help balance innovation with public trust, ensuring that AI technologies contribute positively to disease surveillance without exacerbating existing inequities or ethical risks.

Overall, future research on AI in disease surveillance should adopt an interdisciplinary and context-sensitive approach that aligns technological innovation with public health priorities. By focusing on system integration, explainability, equity in low-resource settings, and robust policy frameworks, AI can evolve into a trustworthy and effective component of outbreak preparedness and response at both national and global levels.

10. Conclusion

AI-driven approaches have the potential to substantially enhance traditional public health surveillance by enabling real-time analysis, early outbreak detection, and improved prediction using diverse and large-scale data sources. Compared with conventional indicator-based systems, AI methods offer greater flexibility, scalability, and capacity to capture complex and non-linear disease patterns. However, challenges related to data quality, ethical governance, transparency, and integration with existing surveillance infrastructure remain significant. Overall, AI should be viewed as a complementary tool that strengthens, rather than replaces, established surveillance systems, with careful implementation and continued evaluation essential to maximizing its public health impact.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Liu, Q., Liu, M., Liang, W., Li, X., Jing, W., Chen, Z., *et al.* (2025) Global Distribution and Health Impact of Infectious Disease Outbreaks, 1996-2023: A Worldwide Retrospective Analysis of World Health Organization Emergency Event Reports. *Journal of Global Health*, **15**, Article No. 04151. https://doi.org/10.7189/jogh.15.04151

[2] World Health Organization Regional Office for the Eastern Mediterranean (2025) Disease Outbreaks. WHO Eastern Mediterranean Region. https://www.emro.who.int/health-topics/disease-outbreaks/

[3] Centers for Disease Control and Prevention (2021) Introduction to Public Health Surveillance. U.S. Department of Health and Human Services. https://www.cdc.gov/training-publichealth101/media/pdfs/introduction-to-surveillance.pdf

[4] Mendes, V.I.S., Mendes, B.M.F., Moura, R.P., Lourenço, I.M., Oliveira, M.F.A., Ng, K.L., *et al.* (2025) Harnessing Artificial Intelligence for Enhanced Public Health Surveillance: A Narrative Review. *Frontiers in Public Health*, **13**, Article ID: 1601151. https://doi.org/10.3389/fpubh.2025.1601151

[5] Idahor, C.O., Esomu, E.O., Ogbonna, N., Momoh, Z., Ogbeide, O.A., Ikhu-Omoregbe, O., *et al.* (2025) Infectious Disease Surveillance in the Era of Big Data and AI: Opportunities and Pitfalls. *Cureus*, **17**, e93929. https://doi.org/10.7759/cureus.93929

[6] Ye, Y., Wagner, M.M., Cooper, G.F., Ferraro, J.P., Su, H., Gesteland, P.H., *et al.* (2017) A Study of the Transferability of Influenza Case Detection Systems between Two Large Healthcare Systems. *PLOS ONE*, **12**, e0174970. https://doi.org/10.1371/journal.pone.0174970

[7] Lake, I.R., Colón-González, F.J., Barker, G.C., Morbey, R.A., Smith, G.E. and Elliot, A.J. (2019) Machine Learning to Refine Decision Making within a Syndromic Surveillance Service. *BMC Public Health*, **19**, Article No. 559. https://doi.org/10.1186/s12889-019-6916-9

[8] Agarwal, P., Aluru, P. and Prakash, B.A. (2022) Real-Time Anomaly Detection in Epidemic Data Streams. OpenReview. https://share.google/0UNSev1ErdYgBxWeW

[9] Tomov, L., Chervenkov, L., Miteva, D.G., Batselova, H. and Velikova, T. (2023) Applications of Time Series Analysis in Epidemiology: Literature Review and Our Experience during COVID-19 Pandemic. *World Journal of Clinical Cases*, **11**, 6974-6983. https://doi.org/10.12998/wjcc.v11.i29.6974

[10] Meliker, J.R. and Sloan, C.D. (2011) Spatio-Temporal Epidemiology: Principles and Opportunities. *Spatial and Spatio-Temporal Epidemiology*, **2**, 1-9. https://doi.org/10.1016/j.sste.2010.10.001

[11] Pfeiffer, D.U. and Stevens, K.B. (2015) Spatial and Temporal Epidemiological Analysis in the Big Data Era. *Preventive Veterinary Medicine*, **122**, 213-220. https://doi.org/10.1016/j.prevetmed.2015.05.012

[12] Musa, G.J., Chiang, P., Sylk, T., Bavley, R., Keating, W., Lakew, B., *et al.* (2013) Use of GIS Mapping as a Public Health Tool—From Cholera to Cancer. *Health Services Insights*, **6**, 111-116. https://doi.org/10.4137/hsi.s10471

[13] Ardabili, S., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A., Reuter, U., *et al.* (2020) COVID-19 Outbreak Prediction with Machine Learning. *Algorithms*, **13**, Article No. 249. https://doi.org/10.3390/a13100249

[14] Zhang, T., Rabhi, F., Chen, X., Paik, H. and MacIntyre, C.R. (2024) A Machine Learning-Based Universal Outbreak Risk Prediction Tool. *Computers in Biology and Medicine*, **169**, Article ID: 107876. https://doi.org/10.1016/j.compbiomed.2023.107876

[15] Olawade, D.B., Wada, O.J., David-Olawade, A.C., Kunonga, E., Abaire, O. and Ling, J. (2023) Using Artificial Intelligence to Improve Public Health: A Narrative Review. *Frontiers in Public Health*, **11**, Article ID: 1196397. https://doi.org/10.3389/fpubh.2023.1196397

[16] Oh, S., Byeon, H. and Wijaya, J. (2024) Machine Learning Surveillance of Foodborne Infectious Diseases Using Wastewater Microbiome, Crowdsourced, and Environmental Data. *Water Research*, **265**, Article ID: 122282. https://doi.org/10.1016/j.watres.2024.122282

[17] Kuo, C. and Fu, J.S. (2021) Evaluating the Impact of Mobility on COVID-19 Pandemic with Machine Learning Hybrid Predictions. *Science of The Total Environment*, **758**, Article ID: 144151. https://doi.org/10.1016/j.scitotenv.2020.144151

[18] Alessa, A. and Faezipour, M. (2018) A Review of Influenza Detection and Prediction through Social Networking Sites. *Theoretical Biology and Medical Modelling*, **15**, 2. https://doi.org/10.1186/s12976-017-0074-5

[19] Alwakeel, M.M. (2025) AI-Assisted Real-Time Monitoring of Infectious Diseases in Urban Areas. *Mathematics*, **13**, Article No. 1911. https://doi.org/10.3390/math13121911

[20] Villanueva-Miranda, I., Xiao, G. and Xie, Y. (2025) Artificial Intelligence in Early Warning Systems for Infectious Disease Surveillance: A Systematic Review. *Frontiers in Public Health*, **13**, Article ID: 1609615. https://doi.org/10.3389/fpubh.2025.1609615

[21] Kareem, S.A. and Quazi, F. (2024) Cloud Based AI Solutions for Public Health Surveillance and Crisis Management. *International Journal of Global Innovations and Solutions*. https://ijgis.pubpub.org/pub/n4rt24ur

[22] MacIntyre, C.R., Chen, X., Kunasekaran, M., Quigley, A., Lim, S., Stone, H., *et al.* (2023) Artificial Intelligence in Public Health: The Potential of Epidemic Early Warning Systems. *Journal of International Medical Research*, **51**, 1-18. https://doi.org/10.1177/03000605231159335

[23] Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., *et al.* (2013) The Global Distribution and Burden of Dengue. *Nature*, **496**, 504-507. https://doi.org/10.1038/nature12060

[24] Kraemer, M.U.G., Reiner, R.C., Brady, O.J., Messina, J.P., Gilbert, M., Pigott, D.M., *et al.* (2019) Past and Future Spread of the Arbovirus Vectors Aedes Aegypti and Aedes Albopictus. *Nature Microbiology*, **4**, 854-863. https://doi.org/10.1038/s41564-019-0376-y

[25] Ong, S.Q., Isawasan, P., Ngesom, A.M.M., Shahar, H., Lasim, A.M. and Nair, G. (2023) Predicting Dengue Transmission Rates by Comparing Different Machine Learning Models with Vector Indices and Meteorological Data. *Scientific Reports*, **13**, Article No. 19129. https://doi.org/10.1038/s41598-023-46342-2

[26] World Health Organization (2014) Early Detection, Assessment and Response to Acute

Public Health Events: Implementation of Early Warning and Response with a Focus on Event-Based Surveillance. World Health Organization. https://www.who.int/publications/i/item/WHO-HSE-GCR-LYO-2014.4

[27] World Health Organization (2021) Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. World Health Organization. https://www.who.int/publications/i/item/9789240029200

[28] Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016) The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, **3**. https://doi.org/10.1177/2053951716679679

[29] Dhanda, S.S., Panwar, D., Lin, C., Sharma, T.K., Rastogi, D., Bindewari, S., *et al*. (2025) Advancement in Public Health through Machine Learning: A Narrative Review of Opportunities and Ethical Considerations. *Journal of Big Data*, **12**, Article No. 154. https://doi.org/10.1186/s40537-025-01201-x

[30] Lundberg, S.M. and Lee, S.-I. (2017) A Unified Approach to Interpreting Model Predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., *et al*., Eds., *Advances in Neural Information Processing Systems*, Vol. 30, Curran Associates, Inc., 4765-4774. https://papers.nips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[31] Ribeiro, M.T., Singh, S. and Guestrin, C. (2016) "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *Proceedings of the* 22*nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 1135-1144. https://doi.org/10.1145/2939672.2939778

[32] Parker, M. (2022) Ethical Challenges in the Use of AI for Infectious Disease Epidemiology: A Double-Edged Sword. Ethox Centre, University of Oxford. https://www.ethox.ox.ac.uk/blog/ethical-challenges-in-the-use-of-ai-for-infectious-disease-epidemiology-a-double-edged-sword/

[33] Simonsen, L., Gog, J.R., Olson, D. and Viboud, C. (2016) Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. *Journal of Infectious Diseases*, **214**, S380-S385. https://doi.org/10.1093/infdis/jiw376

[34] Chadwick, F. (2025) AI for Disease Surveillance in the Modern Era: Early Detection and Rapid Response. *Global Journal of Medical and Biomedical Case Reports*, **6**, 22-30. https://biomedicalcasereports.com/article/view/ai-for-disease-surveillance-in-the-modern-era-early-detection-and-rapid-response