

REVIEW ARTICLE

Big Data Analytics and Precision Public Health for Modern Infectious Disease Surveillance

Sulaiman Samaila^{1,*}, Rabia Ammer²¹ Department of Community Medicine, Ahmadu Bello University, Zaria, Kano state study center, Nigeria.² Lincoln University College, Petaling Jaya, Malaysia.***Corresponding Author:** Sulaiman Samaila¹ Department of Community Medicine, Ahmadu Bello University, Zaria, Kano state study center, Nigeria.Email: sulaimansamailadgw@gmail.com**Articleinfo****Received:** 30 March 2026**Accepted:** 12 May 2026**Keywords:** Big Data Analytics; Infectious Disease Surveillance; Artificial Intelligence; Digital Epidemiology.**How to cite this article:** Sulaiman Samaila, Rabia Ammer (2026). Big Data Analytics and Precision Public Health for Modern Infectious Disease Surveillance, 3(3), 11-23.<https://doi.org/10.26524/amr.2026.3.2>**Abstract**

Infectious diseases remains as one of the major public health concerns worldwide. The continuous emergence of new pathogens, rapid pathogen evolution, and environmental changes has increased the population mobility. Meanwhile, the antimicrobial resistance has made disease prevention and control more challenging. Conventional surveillance systems mainly depend on clinical reports and laboratory confirmation. These approaches often delay outbreak detection and public health response. Recent advances in big data analytics have improved the infectious disease surveillance by enabling the analysis of large health related datasets. Information collected from electronic health records, laboratory databases, pathogen genome sequencing, environmental monitoring, mobile health technologies, and digital platforms provides valuable insights into disease transmission and outbreak patterns. Artificial intelligence supports the analysis of these complex datasets and improve outbreak prediction. Precision public health further strengthens disease surveillance by integrating epidemiological, genomic, environmental, and social data. This approach helps to identify the high risk populations and supports targeted intervention strategies. Despite these advantages, several challenges include poor data quality, privacy concerns, algorithmic bias, limited interoperability, and unequal access to digital technologies. Addressing these limitations through standardized data management, ethical governance, and international collaboration will improve surveillance systems and strengthen preparedness against current and emerging infectious disease threats.

1. Introduction

Infectious diseases remains as one of the major challenges to global healthcare. Continuous advances in medicine and vaccination have reduced the burden of many infectious diseases. However, new infectious pathogens continue to emerge in different parts of the world (Bloom & Cadarette, 2019). Antimicrobial resistance has further increased the burden of infectious diseases. Rapid population movement allows pathogens to spread across countries. Urbanization also influences disease transmission. Environmental changes and increased interaction between humans and wildlife create favorable conditions for the emergence of infectious diseases. Recent outbreaks of COVID-19, mpox, avian influenza, Ebola virus disease, and dengue fever have demonstrated the limitations of conventional surveillance systems (Chiu et al.,

2023; Kayembe-Mulumba et al., 2025). Most surveillance programs depend on laboratory confirmation and clinical reporting. These approaches often detect outbreaks only after community transmission has already occurred. Delayed detection reduces the effectiveness of disease control measures and increases healthcare and economic burdens. The rapid development of digital technologies has changed infectious disease surveillance. Large amounts of health related information are generated every day. Electronic health records are one of the major data sources. Laboratory information systems provide valuable clinical information. Pathogen genomic sequencing provides information about pathogen evolution. Mobile health applications and wearable devices generate continuous health related data. Environmental monitoring systems contribute information

about disease spread. Transportation networks provide mobility information. Social media platforms also provide useful signals related to disease activity ([McClymont et al., 2024](#)). These datasets improve the understanding of disease transmission and outbreak patterns. They also support near real time disease surveillance.

Big data analytics has become an important component of modern infectious disease surveillance ([Garattini et al., 2019](#)). Artificial intelligence improves the analysis of complex datasets. Machine learning supports disease prediction and risk assessment. Deep learning identifies hidden patterns within large datasets. Natural language processing extracts useful information from clinical records and online reports. Network analysis helps to identify disease transmission pathways. Cloud computing supports rapid processing of large datasets. These computational approaches improve outbreak detection and public health decision making. During the COVID 19 pandemic, several countries combined mobility data, genomic surveillance, wastewater monitoring, and electronic health records to monitor disease transmission. These studies demonstrated the value of data driven surveillance during public health emergencies ([Nuha et al., 2025](#)). Disease transmission does not occur equally across all populations. Individuals differ in their genetic background, immune status, age, environmental exposure, healthcare accessibility, and socioeconomic conditions. Precision public health was introduced to address these differences. This approach combines epidemiological information with genomic, environmental, demographic, geographic, and social data. The integrated analysis identifies populations with a greater risk of infection. It also supports targeted prevention strategies and efficient healthcare resource allocation. Whole genome sequencing has further improved infectious disease surveillance ([Struelens et al., 2024](#)). It identifies transmission chains and monitors pathogen evolution. It also detects antimicrobial resistance genes and emerging variants. The combination of genomic information with clinical and environmental data provides a better understanding of disease dynamics. Several challenges continue to limit the application of big data analytics in public health. Data quality differs among healthcare systems. Missing information reduces analytical accuracy. Limited interoperability affects data integration. Privacy concerns and algorithmic bias also remain important issues. Many low and middle income countries have limited digital infrastructure and computational resources. Standardized data collection, secure data sharing, transparent analytical methods, and multidisciplinary collaboration will improve future surveillance systems. This review summarizes recent advances in big data analytics and precision public health for infectious disease surveillance. It discusses major data sources and computational approaches used for disease detection and prediction. It also describes recent applications during infectious disease outbreaks. Current

implementation challenges, ethical considerations, and future research directions are also discussed. These developments may contribute to the design of intelligent surveillance systems that support early outbreak detection and improve public health preparedness.

2. Evolution of Infectious Disease Surveillance

The surveillance of infectious diseases has changed considerably over the past several decades. Earlier surveillance systems mainly depended on manual reporting by healthcare professionals. Suspected cases were confirmed through laboratory testing and reported to public health authorities. These systems helped establish national disease notification programs and supported the control of several infectious diseases. However, data collection and analysis required considerable time. In many situations, several days or weeks were required before an outbreak was recognized. Such delays affected the rapid implementation of control measures during infectious disease emergencies. Traditional surveillance systems were mainly based on information collected from hospitals, diagnostic laboratories, outpatient clinics, and community healthcare centers ([Abat et al., 2016](#)). The collected information included patient demographics, clinical findings, laboratory results, and mortality records. These data provided important information about disease patterns. However, conventional surveillance systems had several limitations. Many asymptomatic infections and mild cases were not identified because individuals did not seek medical care. Differences in reporting practices among healthcare facilities also affected the accuracy of disease monitoring.

Despite these limitations, conventional surveillance contributed significantly to public health achievements. Continuous disease monitoring supported the global eradication of smallpox. It also contributed to the reduction of poliomyelitis in many regions. National surveillance programs improved the control of tuberculosis, measles, cholera, and other communicable diseases ([Mwengee et al., 2016](#); [Sparrow et al., 2024](#)). These programs provided essential information for vaccination strategies and disease control activities. However, globalization and increased population movement created new challenges. Infectious pathogens can now spread rapidly across geographical boundaries. This situation increased the importance of early outbreak detection. The development of digital health technologies introduced new possibilities for improving disease surveillance. Electronic health records replaced many paper-based reporting systems. They allowed clinical information to be stored and shared in a standardized digital format. Healthcare data generated during routine patient care became available for epidemiological analysis. Laboratory information systems further improved surveillance by enabling rapid transmission of diagnostic results. These changes reduced reporting delays and

improved the quality of disease information. The expansion of internet-based technologies created additional sources of health information. People increasingly searched online for disease symptoms and shared health-related information through digital platforms (Yang et al., 2015). Although these data were not originally collected for public health purposes, they provided useful information about changing disease patterns. Analysis of online search behavior helped identify early signals of influenza outbreaks and other infectious diseases. This approach contributed to the development of digital epidemiology. Digital epidemiology uses information from online sources to complement traditional surveillance systems (Kostkova et al., 2021). Mobile technologies further expanded the ability to monitor infectious diseases. Smartphones and wearable devices generate continuous health-related information. These technologies can provide data related to symptoms, body temperature, location patterns, and healthcare utilization. Mobile applications developed during infectious disease outbreaks improved communication between health authorities and the public. They supported symptom reporting, exposure notification, vaccination records, and contact tracing activities. During the COVID-19 pandemic, mobile-based surveillance systems were widely used in several countries to monitor disease transmission (Shen et al., 2021). Genomic sequencing has also transformed infectious disease surveillance. Conventional laboratory methods mainly confirm the presence of pathogens. Whole genome sequencing provides additional information about genetic changes, transmission patterns, and antimicrobial resistance. This approach became particularly important during the COVID-19 pandemic. Continuous sequencing of SARS-CoV-2 samples enabled researchers to identify emerging variants and monitor their spread (Gangavarapu et al., 2023a). Similar genomic approaches are now applied for influenza viruses, tuberculosis, bacterial infections, and antimicrobial-resistant organisms.

Environmental surveillance has become another important component of modern disease monitoring. Wastewater-based epidemiology received significant attention during the COVID-19 pandemic. Detection of viral genetic material in wastewater provided early information about community infection trends. This approach allowed public health authorities to monitor disease patterns without testing every individual (Polo et al., 2020). Environmental surveillance is also used for monitoring poliovirus, antimicrobial resistance markers, and other pathogens present in human communities. Geospatial technologies have further improved infectious disease surveillance. Geographic information systems, satellite imaging, and climate monitoring provide information about environmental factors associated with disease transmission (Kamel Boulos & Geraghty, 2020). Temperature, rainfall, humidity, and land conditions influence the spread of several vector-borne diseases. Integrating environmental

information with epidemiological data helps identify high-risk regions. It also supports better planning of disease prevention strategies. The availability of large-scale digital health data has accelerated the use of artificial intelligence in infectious disease surveillance. Traditional statistical methods are useful for analyzing structured datasets. However, they are limited when handling large and complex information generated from multiple sources. Machine learning and deep learning approaches can identify patterns within large datasets. These methods support outbreak prediction, disease risk assessment, transmission analysis, and public health decision-making. The COVID-19 pandemic represented an important stage in the development of integrated surveillance systems. Multiple data sources were combined to monitor disease transmission (Ibrahim et al., 2022; Najafabadi et al., 2015). These included electronic health records, genomic sequencing, laboratory databases, wastewater monitoring, mobility information, and vaccination records. The integration of these sources provided a broader understanding of disease dynamics. However, the pandemic also revealed several challenges. Data privacy concerns, limited data sharing, inconsistent reporting systems, and unequal digital infrastructure affected surveillance efficiency. The next advancement in infectious disease surveillance is represented by precision public health. Traditional surveillance mainly describes disease patterns at the population level. Precision public health focuses on identifying populations with increased disease risk. It integrates epidemiological information with genomic, environmental, behavioral, and socioeconomic data. This approach allows public health authorities to design targeted interventions (Roberts et al., 2024). Vaccination strategies, preventive measures, and healthcare resources can be directed toward populations that require them most. Modern infectious disease surveillance has therefore progressed from passive reporting systems to intelligent data-driven platforms. Digital health technologies, genomic surveillance, environmental monitoring, artificial intelligence, and precision public health have improved the ability to detect and respond to infectious disease threats. Future improvements in data integration, analytical methods, and international collaboration will further strengthen surveillance systems. These developments will be essential for improving preparedness against emerging infectious disease outbreaks.

3. Big Data Sources in Infectious Disease Surveillance

Modern infectious disease surveillance depends on information obtained from multiple sources. The availability of large scale health data has changed the way disease transmission is monitored and predicted. Traditional surveillance systems mainly rely on confirmed clinical cases. However, big data based surveillance combines information from healthcare systems, digital

platforms, environmental monitoring, and population based technologies. These different sources provide valuable information about pathogen spread, disease patterns, and population risk. The integration of multiple datasets improves the ability to identify early disease signals and support outbreak prediction ([Figure 1](#)).

3.1 Electronic Health Records and Clinical Data

Electronic health records (EHRs) represent one of the major sources of clinical information for infectious disease surveillance. The conversion of medical records from paper based systems to digital platforms has increased the availability of patient related data. EHRs contain information about patient characteristics, symptoms, laboratory results, diagnosis, treatment history, hospitalization, and clinical outcomes ([Willis et al., 2019](#)). These data help researchers understand disease severity, healthcare utilization, and patient risk factors. EHR based surveillance can identify unusual patterns within healthcare settings. An increase in patients with similar symptoms may indicate the early stage of an outbreak. Machine learning methods have been applied to EHR datasets to predict disease severity and identify patients at higher risk of complications. During the COVID 19 pandemic, EHR analysis helped identify important risk factors associated with severe infection ([Satterfield et al., 2021](#); [Sheikhtaheri et al., 2024](#)). However, EHR data are mainly collected for patient care rather than disease surveillance. Differences in documentation methods and missing clinical information can affect data quality. Variation between healthcare systems also creates challenges in data integration. Standardized data formats and improved interoperability are required to improve the use of EHRs in infectious disease surveillance.

3.2 Laboratory and Diagnostic Data

Laboratory information provides the foundation for infectious disease detection. Diagnostic laboratories generate large amounts of data through molecular testing, antigen detection, serological analysis, and microbial identification. These data provide direct evidence of pathogen presence and infection trends. The integration of laboratory information systems with public health databases improves real time disease monitoring. During the COVID 19 pandemic, laboratory testing data helped determine infection rates and geographical patterns ([Khan et al., 2023](#)). Microbiology laboratory data also support antimicrobial resistance surveillance by monitoring changes in bacterial susceptibility patterns. Advances in genomic technologies have further expanded the role of laboratory surveillance. Whole genome sequencing provides information about pathogen evolution, transmission routes, and resistance associated genetic changes ([Baker et al., 2018](#)). Combining genomic information with epidemiological data improves outbreak investigation and helps identify infection sources.

3.3 Pathogen Genomic Data and Molecular Epidemiology

Pathogen genomic surveillance has transformed infectious disease monitoring. Next generation sequencing technologies have increased the ability to generate large genomic datasets at lower cost. These data provide information about pathogen diversity, genetic changes, and evolutionary patterns. During the SARS CoV 2 pandemic, genomic databases allowed researchers to monitor viral variants and evaluate their impact on disease transmission. Similar approaches are now used for influenza surveillance, tuberculosis investigation, and detection of antimicrobial resistant organisms ([Tiwari et al., 2025](#); [Wilkinson et al., 2021](#)). Genomic information combined with clinical and geographical data supports precision epidemiology. Despite its advantages, genomic surveillance requires advanced computational facilities and trained personnel. The analysis and interpretation of large genomic datasets remain important challenges for many healthcare systems.

3.4 Mobile Health Data and Wearable Technologies

Mobile health technologies have created new opportunities for collecting health information directly from populations. Smartphones and wearable devices generate continuous data related to symptoms, body temperature, physical activity, location, and healthcare behavior ([Ng et al., 2018](#)). These data provide additional information that complements traditional surveillance systems. Mobile applications developed during infectious disease outbreaks supported symptom reporting, exposure notification, and public health communication. During the COVID 19 pandemic, digital contact tracing applications were used in several countries to identify possible exposure events ([Nachega et al., 2020](#)). Wearable devices have also been investigated for detecting early physiological changes associated with infection. However, the use of mobile health data requires careful management of privacy and consent issues. Public acceptance depends on secure data handling and transparent policies.

3.5 Social Media and Internet Based Data

Social media and internet platforms have become important sources of digital epidemiological information. People frequently share health concerns, symptoms, and information about disease outbreaks through online platforms. These data can provide early indications of changes in community health patterns. Artificial intelligence and natural language processing methods are used to analyze online information. These approaches can identify disease related discussions and detect unusual increases in health concerns. Internet search patterns have also been explored for monitoring influenza and COVID 19 activity ([Ma et al., 2022](#)). However, online information can be affected by media attention and misinformation. Therefore, digital signals require validation with traditional epidemiological data before being used for public health decisions.

3.6 Environmental and Wastewater Surveillance

Environmental surveillance provides additional information about pathogen circulation in communities. Wastewater based epidemiology has become an important approach for monitoring infectious diseases. Pathogen genetic material released through biological waste can be detected in wastewater samples. During the COVID 19 pandemic, wastewater monitoring provided early indications of increasing infection levels before clinical cases increased (Bogler et al., 2020; Gahlot et al., 2023). This approach is also being explored for poliovirus monitoring, antimicrobial resistance detection, and surveillance of gastrointestinal pathogens. Environmental factors also influence infectious disease transmission. Climate conditions such as temperature, rainfall, and humidity affect the distribution of several vector borne diseases. Combining environmental information with disease data improves prediction models for infections such as malaria and dengue.

3.7 Geospatial and Mobility Data

Geospatial technologies provide important information about the distribution and movement of infectious diseases. Geographic information systems, satellite data, and mobility information help identify patterns of disease spread. Human movement plays an important role in pathogen transmission. Mobility data obtained from transportation systems and mobile networks have been used to evaluate the effects of travel restrictions and social distancing measures (Bogler et al., 2020). Spatial models can identify regions with increased

outbreak risk and support healthcare resource planning. Although mobility data are valuable for surveillance, privacy protection remains essential. Appropriate regulations are required to prevent misuse of personal location information.

3.8 Integration of Multi Source Data for Precision Surveillance

The major advantage of big data surveillance is the ability to combine information from different sources. Clinical data provide information about patients. Genomic data describe pathogen changes. Environmental data explain external risk factors. Mobility data reveal transmission patterns. The integration of these datasets provides a comprehensive understanding of infectious disease dynamics. Artificial intelligence and machine learning approaches help analyze these complex datasets and generate useful information for public health decisions. Combining climate information with vector surveillance can improve outbreak prediction. Linking genomic data with mobility patterns can help identify transmission pathways. Future infectious disease surveillance will depend on secure and effective integration of multiple data sources. Standardized data sharing systems, improved analytical methods, and international collaboration will be important for developing reliable surveillance platforms. These advances will strengthen outbreak prediction, prevention, and response strategies.

Big Data Sources in Infectious Disease Surveillance

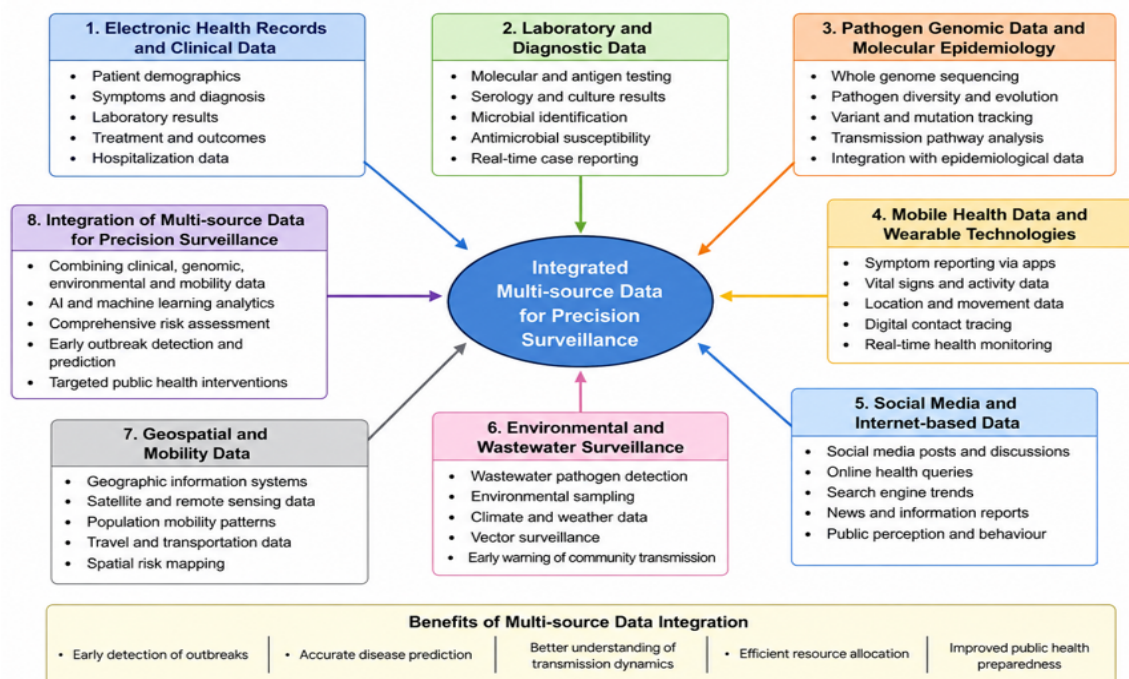


Figure 1. Data ecosystem for precision infectious disease surveillance. This figure categorizes big data inputs into eight domains: clinical records, laboratory diagnostics, pathogen genomics, mobile/wearable sensors, online/social media signals, environmental/wastewater monitors, geospatial/mobility tracking, and their integrative AI-driven fusion. Each domain is represented with its constituent sub-variables. The integrated output enables early outbreak signal detection, real-time risk stratification, and data-guided public health decision-making.

4. Big Data Analytics Techniques for Infectious Disease Detection and Prediction

The increasing availability of digital health information has created new opportunities for improving infectious disease surveillance. Traditional epidemiological approaches mainly depend on historical disease records and established risk factors. These methods provide important information about disease patterns. However, they have limitations when analyzing large datasets generated from clinical, environmental, genomic, and behavioral sources. Infectious disease transmission involves multiple factors that change continuously over time. Big data analytics provides advanced computational methods to analyze these complex datasets. The application of artificial intelligence, machine learning, deep learning, natural language processing, and network analysis has improved the ability to detect disease signals and predict outbreak patterns (Table 1).

4.1 Artificial Intelligence in Infectious Disease Surveillance

Artificial intelligence (AI) has become an important tool in modern infectious disease surveillance. AI systems can analyze large amounts of information within a short period. These systems identify relationships between different variables that may not be detected using conventional methods. AI applications include outbreak detection, disease prediction, risk assessment, and support for public health decisions. AI models use information obtained from electronic health records, laboratory reports, genomic databases, environmental monitoring systems, and digital platforms. The integration of these data sources allows AI algorithms to identify unusual disease patterns. Changes in healthcare visits, symptom reports, and laboratory-confirmed cases can provide early indications of increasing disease activity. During the COVID-19 pandemic, AI-based approaches were widely used for predicting infection trends and identifying high-risk populations. Machine learning models helped estimate disease severity and hospitalization risk using clinical and demographic information. AI-based genomic analysis also supported the identification of viral mutations and emerging variants (Kaur et al., 2021; Sharma et al., 2026). However, the accuracy of AI systems depends on the quality of available data. Missing information, reporting differences, and unequal healthcare infrastructure can affect model performance. Therefore, AI-based predictions should support public health decisions rather than replace epidemiological knowledge.

4.2 Machine Learning Approaches for Disease Prediction

Machine learning (ML) is one of the most commonly applied approaches in infectious disease surveillance. Unlike traditional statistical methods, ML algorithms can identify patterns directly from large datasets. These models improve their performance by learning from previous data. This ability makes them useful for analyzing continuously

changing infectious disease patterns. Supervised learning methods such as random forest, support vector machines, logistic regression, and gradient boosting have been applied for disease classification and risk prediction (Delpino et al., 2022; Omar et al., 2024). These models have been used to predict hospitalization, mortality, and disease severity. Clinical information and laboratory parameters are commonly used to identify patients at increased risk of severe infection. Unsupervised learning methods provide additional advantages in disease surveillance. These approaches identify hidden patterns without predefined categories. Clustering techniques can group patients with similar clinical features. They can also identify geographical areas with similar transmission characteristics. This approach is useful during emerging outbreaks where limited information is available. Reinforcement learning has recently gained attention for optimizing public health strategies. These models evaluate previous decisions and outcomes to identify effective intervention approaches (Frommeyer et al., 2025). Applications include vaccine distribution, healthcare resource allocation, and outbreak control planning. However, careful evaluation is required because these decisions can influence large populations.

4.3 Deep Learning for Complex Surveillance Data

Deep learning is an advanced form of machine learning that uses artificial neural networks to analyze complex datasets. These models can identify important features directly from raw information. This ability has increased their application in genomic analysis, medical imaging, clinical data interpretation, and disease forecasting. Convolutional neural networks (CNNs) are widely used for infectious disease diagnosis through medical imaging (Narin et al., 2021). These models can identify infection-related patterns in chest radiographs and computed tomography images. Deep learning-based imaging systems have shown potential for rapid screening of respiratory infections. Recurrent neural networks (RNNs) and transformer-based models are used for analyzing time-dependent disease information. Infectious diseases change continuously due to seasonal variation, population behavior, and pathogen evolution. Deep learning models can analyze these changes and predict future disease trends. Long short-term memory networks have been applied for forecasting influenza activity and COVID-19 case patterns. Deep learning also supports pathogen genomic surveillance. Neural network models can identify genetic variations and classify pathogen strains (Zhu et al., 2024). However, these approaches require large datasets and advanced computational resources. The lack of transparency in model decisions remains a major limitation in clinical and public health applications.

4.4 Natural Language Processing for Digital Epidemiology

Natural language processing (NLP) enables computers to analyze human language-based information. It has become

useful for extracting health information from unstructured sources. These sources include scientific publications, electronic health records, social media platforms, and online news reports. A large amount of health information is available in text format. NLP algorithms can identify disease-related terms, symptoms, locations, and outbreak information from these sources. This approach supports early detection of emerging health concerns. During the COVID-19 pandemic, NLP-based systems were used to analyze online discussions and identify changes in public health concerns (Al-Garadi et al., 2022; Fatima et al., 2022). Automated analysis of scientific literature also helped researchers track developments in viral research, treatments, and vaccines. However, digital information can contain misinformation and incomplete descriptions. Therefore, NLP-based findings require confirmation using reliable epidemiological data.

4.5 Predictive Modeling and Disease Forecasting

Predictive modeling is an important application of big data analytics in infectious disease surveillance. These models use previous and current information to estimate future disease trends. Early predictions allow health authorities to prepare preventive strategies before outbreaks become severe. Traditional time-series models have been widely used for disease forecasting. Recent approaches combine statistical methods with machine learning techniques. These models consider infection rates, vaccination coverage, climate conditions, population

movement, and healthcare utilization (Kraemer et al., 2025). Predictive models have been developed for diseases such as influenza, malaria, dengue, tuberculosis, and COVID-19. Environmental factors are particularly important for vector-borne diseases because temperature and rainfall influence pathogen transmission. Mobility patterns are also important for respiratory infections because human interaction affects disease spread. Despite improvements, disease prediction remains challenging. Pathogen evolution, behavioral changes, and policy interventions can influence disease patterns. Continuous model updating and validation are therefore necessary.

4.6 Network Analysis for Disease Transmission

Network analysis provides a framework for understanding infectious disease transmission. It represents individuals or locations as connected networks. These connections help identify transmission pathways and high-risk clusters. Social network analysis has been used to study population interactions and identify individuals who may contribute to disease spread. During COVID-19, network models helped evaluate the effects of travel restrictions and social distancing measures (Le et al., 2022). Network analysis is also valuable in antimicrobial resistance surveillance. It can identify transmission patterns between healthcare facilities and patients. However, incomplete interaction data and privacy concerns remain important challenges.

Table 1: Comparative framework for selecting big data analytics approaches in infectious disease surveillance

Analytical approach	Most suitable surveillance situation	Main data requirement	Typical analytical output	Preferred validation approach	Practical
Artificial intelligence	Large surveillance systems requiring automated decision support	Well structured multi source datasets	Risk scores and alerts and decision support indicators	Retrospective testing followed by prospective evaluation	High
Machine learning	Prediction of disease outcomes or population risk	Labelled clinical and epidemiological datasets	Classification or prediction scores	Cross validation and external dataset testing	High
Deep learning	Very large datasets with complex patterns	High volume imaging or sequence or temporal data	Automated feature representation and prediction	Independent test datasets and repeated validation	Moderate to high
Natural language processing	Surveillance based on unstructured text	Clinical notes and online reports and public health documents	Extracted disease signals and event information	Manual annotation and agreement assessment	Moderate

Time series forecasting	Monitoring disease trends over successive time periods	Regularly collected longitudinal data	Short term and long term forecasts	Rolling window validation and forecast error analysis	High
Network analysis	Investigation of connected transmission events	Reliable contact or interaction information	Transmission clusters and network structures	Comparison with epidemiological investigations	Moderate
Multomics integration	Investigation of host pathogen relationships	Genomic and transcriptomic and proteomic datasets	Integrated biological risk profiles	Independent cohort validation	Specialized
Real time analytics	Rapid monitoring during an active outbreak	Continuously updated data streams	Real time alerts and changing risk estimates	Continuous performance monitoring	High
Geospatial analytics	Identification of spatially concentrated disease risk	Geographic and population linked datasets	Risk maps and spatial clusters	Spatial cross validation and field verification	Moderate

5. Precision Public Health Applications in Infectious Disease Control

Precision public health has emerged as an advanced approach for improving infectious disease prevention and control. Traditional public health strategies generally apply similar interventions to entire populations. These strategies include mass vaccination programs, general awareness campaigns, and universal screening approaches. Although these methods have successfully reduced the burden of many infectious diseases, they do not always consider differences between populations. Factors such as genetic background, environmental exposure, healthcare accessibility, and social conditions can influence infection risk. Precision public health addresses these variations by combining large-scale health information with advanced analytical methods (Velmovitsky et al., 2021). This approach helps identify high-risk populations and supports targeted disease control strategies. The success of precision public health depends on the integration of multiple data sources. These include clinical records, pathogen genomic information, environmental factors, demographic characteristics, mobility patterns, and socioeconomic conditions. The analysis of these datasets provides a detailed understanding of disease emergence and transmission. It also allows healthcare authorities to design interventions according to specific population needs rather than applying the same strategy everywhere.

5.1 Precision Surveillance for Early Outbreak Detection

Early detection is one of the major applications of precision public health. Conventional surveillance systems usually identify outbreaks after infected individuals seek medical care. This delay can allow pathogens to

spread within communities before control measures are implemented. Precision surveillance improves early detection by integrating multiple real-time data sources. Electronic health records, laboratory reports, wastewater monitoring, and digital health platforms provide continuous information about disease activity. Artificial intelligence models can analyze these data and identify unusual changes in disease patterns. For example, increases in respiratory symptoms, changes in online health searches, and pathogen detection in wastewater can provide early indications of infection trends. Precision surveillance also improves localized outbreak monitoring (Li et al., 2026). Instead of analyzing disease patterns only at national or regional levels, advanced analytical systems can identify specific communities or geographic areas with increased infection risk. This approach allows targeted interventions and reduces unnecessary restrictions in low-risk populations.

5.2 Genomic-Guided Infectious Disease Control

Pathogen genomic information has become an important component of precision public health. Whole genome sequencing provides detailed information about pathogen evolution and transmission patterns. It also helps identify genetic changes associated with increased transmissibility, immune escape, and antimicrobial resistance. During the COVID-19 pandemic, genomic surveillance played a major role in identifying SARS-CoV-2 variants (Gangavarapu et al., 2023b; Robishaw et al., 2021). Sequencing data helped researchers understand whether changes in infection rates were related to new variants or changes in population behavior. Similar genomic approaches are now used for influenza viruses,

tuberculosis, malaria parasites, and bacterial pathogens. Genomic surveillance is particularly valuable for controlling antimicrobial resistance. Conventional methods often detect resistance after treatment failure occurs. Genomic analysis can identify resistance-associated genes at an early stage. Combining genomic data with healthcare records helps identify transmission pathways within hospitals and communities. This information supports targeted infection control measures.

5.3 Targeted Vaccination Strategies

Precision public health can improve vaccination strategies by identifying populations with increased infection risk. Traditional vaccination programs often depend on age groups or geographic locations. However, infection risk can vary between individuals within the same population. Precision vaccination approaches use information related to disease exposure, immune status, demographic characteristics, and pathogen changes (Soni et al., 2020). These data help identify populations that require priority vaccination during outbreaks. Healthcare workers, elderly individuals, immunocompromised populations, and highly exposed communities can be identified through risk assessment models. Artificial intelligence-based approaches are also being explored to predict vaccine response. These models consider previous infection history, immune characteristics, and clinical conditions. Such strategies may improve vaccine effectiveness, particularly for diseases where immunity decreases over time or pathogen variants influence protection.

5.4 Precision Approaches for Antimicrobial Resistance Surveillance

Antimicrobial resistance is a major global health concern. Resistant pathogens reduce the effectiveness of commonly used treatments and increase disease burden. Traditional surveillance methods often depend on laboratory confirmation after bacterial isolation. This process may delay the detection of emerging resistance patterns. Precision public health integrates genomic surveillance with clinical information and antimicrobial usage data. This approach allows early identification of resistance-associated changes. It also helps identify healthcare facilities and communities with increased resistance risk. Machine learning models have been developed to predict antimicrobial resistance using patient information, previous medication exposure, and microbiological data (Pennisi et al., 2025). These approaches may support better treatment decisions and reduce unnecessary antimicrobial use. The combination of precision surveillance and antimicrobial stewardship can improve control of resistant infections.

5.5 Precision Management of Vector-Borne Diseases

Vector-borne diseases are strongly influenced by environmental conditions. Diseases such as malaria,

dengue fever, chikungunya, and Zika virus infection show seasonal and geographic variation. Uniform control strategies may not be effective in all regions. Precision public health combines climate information, satellite data, vector surveillance, and epidemiological records to predict disease risk. Environmental factors such as temperature, rainfall, and vegetation changes influence vector populations. Including these factors in predictive models improves outbreak forecasting. These approaches help health authorities implement targeted control measures. Vector control activities, diagnostic services, and preventive interventions can be directed toward high-risk areas. For example, malaria control programs use geographic and epidemiological information to prioritize communities requiring preventive measures (López-Monteón et al., 2026).

6. Challenges and Ethical Considerations in Big Data-Driven Infectious Disease Surveillance

Although big data analytics has improved infectious disease surveillance, several challenges limit its effective implementation. The success of these systems depends mainly on the quality and reliability of the data used for analysis. Modern surveillance platforms collect information from different sources such as electronic health records, laboratory databases, genomic sequencing systems, mobile applications, environmental monitoring, and digital platforms. However, these datasets often differ in their accuracy, format, and reporting methods. Electronic health records may contain incomplete clinical information. Laboratory surveillance may fail to identify infections in regions with limited diagnostic facilities. Digital sources such as social media data may represent public behavior rather than actual disease occurrence. These limitations can influence the accuracy of analytical models and may affect public health decisions. Therefore, improving data standardization and establishing reliable data management systems are essential for effective surveillance. The integration of multiple data sources remains another important limitation. Health records, genomic information, environmental data, and mobility patterns are usually collected and maintained by different organizations. The absence of common data formats and limited communication between systems can reduce the efficiency of data sharing. Developing interoperable platforms will help connect healthcare institutions, research centers, and public health agencies. Such integration is necessary for generating comprehensive models that can accurately represent disease transmission patterns.

The use of large-scale health data also raises important concerns related to privacy and security. Infectious disease surveillance systems may involve sensitive information including medical records, genetic information, geographic location, and personal behavior. Although these data provide valuable insights into disease spread, improper handling may compromise individual privacy. Secure storage systems, data encryption, controlled access, and clear governance policies

are required to ensure responsible use of health information. Ethical issues associated with artificial intelligence based surveillance also require careful consideration. Machine learning models depend on the quality of training datasets. If these datasets contain population bias or incomplete information, the resulting predictions may not accurately represent all communities. For example, populations with limited healthcare access may appear to have lower infection rates because of reduced testing rather than lower disease risk. Similarly, digital surveillance methods based on smartphones or internet access may exclude individuals living in areas with poor digital infrastructure. Such limitations may increase existing health inequalities. The digital divide remains a major barrier, particularly in low and middle income countries. Limited access to advanced healthcare infrastructure, computational resources, and trained personnel can restrict the adoption of big data based surveillance systems. Expanding digital health infrastructure and strengthening technical capacity are important steps for achieving equitable implementation. Without these improvements, advanced surveillance technologies may benefit only regions with strong technological resources. Another challenge is the lack of transparency in complex artificial intelligence models. Deep learning systems can provide accurate predictions, but their decision-making processes are often difficult to understand. This limitation may reduce confidence among healthcare professionals and policymakers. The development of explainable artificial intelligence approaches can improve transparency and support responsible decision-making. Future infectious disease surveillance requires a balanced approach that combines technological advancement with ethical responsibility. Improving data quality, protecting privacy, reducing algorithmic bias, and strengthening international collaboration will be essential for developing reliable surveillance systems. Cooperation between epidemiologists, healthcare professionals, data scientists, policymakers, and communities will help overcome current limitations. Addressing these challenges will allow big data analytics and precision public health to support early outbreak detection, targeted interventions, and better preparedness for future infectious disease threats.

7. Conclusion

Big data analytics and precision public health have changed the approach of infectious disease surveillance. These technologies allow earlier identification of outbreaks and improve the prediction of disease trends. The use of clinical records, pathogen genomic information, environmental data, and digital health platforms provides valuable information about disease transmission and population risk factors. However, several limitations still affect the implementation of data-driven surveillance systems. Issues related to data accuracy, privacy protection, system integration, and unequal access to

digital technologies need to be addressed. Improving data management strategies and developing ethical frameworks will be important for the successful application of these approaches. Future surveillance systems will depend on the effective combination of multiple data sources with advanced analytical methods. Artificial intelligence and machine learning approaches can further improve outbreak monitoring and support timely public health decisions. The integration of these technologies with conventional epidemiological methods may help develop more reliable surveillance platforms and improve global preparedness for future infectious disease challenges.

Declarations

Ethics approval statement

Not applicable

Consent to participate

Not applicable

Consent to publish

Not applicable

Data Availability Statement

The data are available from the corresponding author upon reasonable request

Competing Interests

The authors declare that they have no conflict of interest

Funding

Not Applicable

Author contribution

Conceptualization, Data curation, Investigation, Formal analysis, Writing, review, and editing: S.S & R.A. All authors have read and agreed to the published version of the manuscript

Acknowledgements

Not Applicable

References

1. Abat, C., Chaudet, H., Rolain, J.-M., Colson, P., & Raoult, D. (2016). Traditional and syndromic surveillance of infectious diseases and pathogens. *International Journal of Infectious Diseases*, 48, 22–28. <https://doi.org/10.1016/j.ijid.2016.04.021>
2. Al-Garadi, M. A., Yang, Y.-C., & Sarker, A. (2022). The Role of Natural Language Processing during the COVID-19 Pandemic: Health Applications, Opportunities, and Challenges. *Healthcare*, 10(11), 2270. <https://doi.org/10.3390/healthcare10112270>
3. Baker, S., Thomson, N., Weill, F.-X., & Holt, K. E. (2018). Genomic insights into the emergence and spread of antimicrobial-

- resistant bacterial pathogens. *Science*, 360(6390), 733–738. <https://doi.org/10.1126/science.aar3777>
4. Bloom, D. E., & Cadarette, D. (2019). Infectious Disease Threats in the Twenty-First Century: Strengthening the Global Response. *Frontiers in Immunology*, 10. <https://doi.org/10.3389/fimmu.2019.00549>
 5. Bogler, A., Packman, A., Furman, A., Gross, A., Kushmaro, A., Ronen, A., Dagot, C., Hill, C., Vaizel-Ohayon, D., Morgenroth, E., Bertuzzo, E., Wells, G., Kiperwas, H. R., Horn, H., Negev, I., Zucker, I., Bar-Or, I., Moran-Gilad, J., Balcazar, J. L., ... Bar-Zeev, E. (2020). Rethinking wastewater risks and monitoring in light of the COVID-19 pandemic. *Nature Sustainability*, 3(12), 981–990. <https://doi.org/10.1038/s41893-020-00605-2>
 6. Chiu, K. H.-Y., Sridhar, S., & Yuen, K.-Y. (2023). Preparation for the next pandemic: challenges in strengthening surveillance. *Emerging Microbes & Infections*, 12(2). <https://doi.org/10.1080/22221751.2023.2240441>
 7. Delpino, F. M., Costa, Â. K., Farias, S. R., Chiavegatto Filho, A. D. P., Arcêncio, R. A., & Nunes, B. P. (2022). Machine learning for predicting chronic diseases: a systematic review. *Public Health*, 205, 14–25. <https://doi.org/10.1016/j.puhe.2022.01.007>
 8. Fatima, R., Samad Shaikh, N., Riaz, A., Ahmad, S., El-Affendi, M. A., Alyamani, K. A. Z., Nabeel, M., Ali Khan, J., Yasin, A., & Latif, R. M. A. (2022). A Natural Language Processing (NLP) Evaluation on COVID-19 Rumour Dataset Using Deep Learning Techniques. *Computational Intelligence and Neuroscience*, 2022, 1–17. <https://doi.org/10.1155/2022/6561622>
 9. Frommeyer, T. C., Gilbert, M. M., Fursmidt, R. M., Park, Y., Khouzam, J. P., Brittain, G. V., Frommeyer, D. P., Bett, E. S., & Bihl, T. J. (2025). Reinforcement Learning and Its Clinical Applications Within Healthcare: A Systematic Review of Precision Medicine and Dynamic Treatment Regimes. *Healthcare*, 13(14), 1752. <https://doi.org/10.3390/healthcare13141752>
 10. Gahlot, P., Alley, K. D., Arora, S., Das, S., Nag, A., & Tyagi, V. K. (2023). Wastewater surveillance could serve as a pandemic early warning system for COVID - 19 and beyond. *WIREs Water*, 10(4). <https://doi.org/10.1002/wat2.1650>
 11. Gangavarapu, K., Latif, A. A., Mullen, J. L., Alkuzweny, M., Hufbauer, E., Tsueng, G., Haag, E., Zeller, M., Aceves, C. M., Zaiets, K., Cano, M., Zhou, X., Qian, Z., Sattler, R., Matteson, N. L., Levy, J. I., Lee, R. T. C., Freitas, L., Maurer-Stroh, S., ... Hughes, L. D. (2023). Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Nature Methods*, 20(4), 512–522. <https://doi.org/10.1038/s41592-023-01769-3>
 12. Garattini, C., Raffle, J., Aisyah, D. N., Sartain, F., & Kozlakidis, Z. (2019). Big Data Analytics, Infectious Diseases and Associated Ethical Impacts. *Philosophy & Technology*, 32(1), 69–85. <https://doi.org/10.1007/s13347-017-0278-y>
 13. Ibrahim, D. A., Zebari, D. A., Mohammed, H. J., & Mohammed, M. A. (2022). Effective hybrid deep learning model for COVID -19 patterns identification using CT images. *Expert Systems*, 39(10). <https://doi.org/10.1111/exsy.13010>
 14. Kamel Boulos, M. N., & Geraghty, E. M. (2020). Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal of Health Geographics*, 19(1), 8. <https://doi.org/10.1186/s12942-020-00202-8>
 15. Kaur, I., Behl, T., Aleya, L., Rahman, H., Kumar, A., Arora, S., & Bulbul, I. J. (2021). Artificial intelligence as a fundamental tool in management of infectious diseases and its current implementation in COVID-19 pandemic. *Environmental Science and Pollution Research*, 28(30), 40515–40532. <https://doi.org/10.1007/s11356-021-13823-8>
 16. Kayembe-Mulumba, B., N'gattia, A. K., & Belizaire, M. R. D. (2025). One Health, Many Gaps: Rethinking Epidemic Intelligence in Resource-Limited Settings to Prepare for the Global Threat of Disease X. *Microorganisms*, 13(11), 2615. <https://doi.org/10.3390/microorganisms13112615>
 17. Khan, M. M., Odoi, A., & Odoi, E. W. (2023). Geographic disparities in COVID-19 testing and outcomes in Florida. *BMC Public Health*, 23(1), 79. <https://doi.org/10.1186/s12889-022-14450-9>
 18. Kostkova, P., Saigí-Rubió, F., Eguia, H., Borbolla, D., Verschuuren, M., Hamilton, C., Azzopardi-Muscat, N., & Novillo-Ortiz, D. (2021). Data and Digital Solutions to Support Surveillance Strategies in the Context of the COVID-19 Pandemic. *Frontiers in Digital Health*, 3. <https://doi.org/10.3389/fdgh.2021.707902>
 19. Kraemer, M. U. G., Tsui, J. L.-H., Chang, S. Y., Lytras, S., Khurana, M. P., Vanderslott, S., Bajaj, S., Scheidwasser, N., Curran-Sebastian, J. L., Semenova, E., Zhang, M., Unwin, H. J. T., Watson, O. J., Mills, C., Dasgupta, A., Ferretti, L., Scarpino, S. V., Koua, E., Morgan, O., ... Bhatt, S. (2025). Artificial intelligence for modelling infectious disease epidemics. *Nature*, 638(8051), 623–635. <https://doi.org/10.1038/s41586-024-08564-w>
 20. Le, T.-M., Raynal, L., Talbot, O., Hambridge, H., Drovandi, C., Mira, A., Mengersen, K., & Onnela, J.-P. (2022). Framework for assessing and easing global COVID-19 travel restrictions. *Scientific Reports*, 12(1), 6985. <https://doi.org/10.1038/s41598-022-10678-y>
 21. Li, J.-H., Tseng, Y.-J., Chen, S.-H., & Chen, K.-F. (2026). Artificial intelligence in infection surveillance: Data integration, applications and future directions. *Biomedical Journal*, 49(2), 100929. <https://doi.org/10.1016/j.bj.2025.100929>
 22. López-Monteón, A., Ramos-López, M., Ramos-López, E., & Ramos-Ligonio, A. (2026). Vector-Borne Disease Control in the Anthropocene: Integrating Biotechnology, Genomic Surveillance, and Environmental Management. *Current Tropical Medicine Reports*, 13(1), 15. <https://doi.org/10.1007/s40475-026-00374-6>
 23. Ma, S., Sun, Y., & Yang, S. (2022). Using Internet Search Data to Forecast COVID-19 Trends: A Systematic Review. *Analytics*, 1(2), 210–227. <https://doi.org/10.3390/analytics1020014>
 24. McClymont, H., Lambert, S. B., Barr, I., Vardoulakis, S., Bambrick, H., & Hu, W. (2024). Internet-based Surveillance Systems and Infectious Diseases Prediction: An Updated Review of the Last 10 Years and Lessons from the COVID-19 Pandemic. *Journal of*

- Epidemiology and Global Health, 14(3), 645–657. <https://doi.org/10.1007/s44197-024-00272-y>
25. Mwengee, W., Okeibunor, J., Poy, A., Shaba, K., Mbulu Kinuani, L., Minkoulou, E., Yahaya, A., Gaturuku, P., Landoh, D. E., Nsubuga, P., Salla, M., Mihigo, R., & Mkanda, P. (2016). Polio Eradication Initiative: Contribution to improved communicable diseases surveillance in WHO African region. *Vaccine*, 34(43), 5170–5174. <https://doi.org/10.1016/j.vaccine.2016.05.060>
 26. Nachega, J. B., Leisegang, R., Kallay, O., Mills, E. J., Zumla, A., & Lester, R. T. (2020). Mobile Health Technology for Enhancing the COVID-19 Response in Africa: A Potential Game Changer? *The American Journal of Tropical Medicine and Hygiene*, 103(1), 3–5. <https://doi.org/10.4269/ajtmh.20-0506>
 27. Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1. <https://doi.org/10.1186/s40537-014-0007-7>
 28. Narin, A., Kaya, C., & Pamuk, Z. (2021). Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, 24(3), 1207–1220. <https://doi.org/10.1007/s10044-021-00984-y>
 29. Ng, Y. C., Alexander, S., & Frith, K. H. (2018). Integration of Mobile Health Applications in Health Information Technology Initiatives. *CIN: Computers, Informatics, Nursing*, 36(5), 209–213. <https://doi.org/10.1097/CIN.0000000000000445>
 30. Nuha, N., Ali Pitchay, S., Ab Halim, A. H., Sahbudin, M. A. Bin, & Sahbudin, I. (2025). Beyond the outbreak: a review of big data analytics in proactive infectious disease prevention for risk mitigation for COVID-19. *Journal of Big Data*, 12(1), 185. <https://doi.org/10.1186/s40537-025-01245-z>
 31. Omar, E. D., Mat, H., Abd Karim, A. Z., Sanaudi, R., Ibrahim, F., Omar, M. A., Ismail, M. Z. H., Jayaraj, V., & Goh, B. L. (2024). Comparative Analysis of Logistic Regression, Gradient Boosted Trees, SVM, and Random Forest Algorithms for Prediction of Acute Kidney Injury Requiring Dialysis After Cardiac Surgery. *International Journal of Nephrology and Renovascular Disease*, Volume 17, 197–204. <https://doi.org/10.2147/IJNRD.S461028>
 32. Pennisi, F., Pinto, A., Ricciardi, G. E., Signorelli, C., & Gianfredi, V. (2025). The Role of Artificial Intelligence and Machine Learning Models in Antimicrobial Stewardship in Public Health: A Narrative Review. *Antibiotics*, 14(2), 134. <https://doi.org/10.3390/antibiotics14020134>
 33. Polo, D., Quintela-Balaja, M., Corbishley, A., Jones, D. L., Singer, A. C., Graham, D. W., & Romalde, J. L. (2020). Making waves: Wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. *Water Research*, 186, 116404. <https://doi.org/10.1016/j.watres.2020.116404>
 34. Roberts, M. C., Holt, K. E., Del Fiore, G., Baccarelli, A. A., & Allen, C. G. (2024). Precision public health in the era of genomics and big data. *Nature Medicine*, 30(7), 1865–1873. <https://doi.org/10.1038/s41591-024-03098-0>
 35. Robishaw, J. D., Alter, S. M., Solano, J. J., Shih, R. D., DeMets, D. L., Maki, D. G., & Hennekens, C. H. (2021). Genomic surveillance to combat COVID-19: challenges and opportunities. *The Lancet Microbe*, 2(9), e481–e484. [https://doi.org/10.1016/S2666-5247\(21\)00121-X](https://doi.org/10.1016/S2666-5247(21)00121-X)
 36. Satterfield, B. A., Dikilitas, O., & Kullo, I. J. (2021). Leveraging the Electronic Health Record to Address the COVID-19 Pandemic. *Mayo Clinic Proceedings*, 96(6), 1592–1608. <https://doi.org/10.1016/j.mayocp.2021.04.008>
 37. Sharma, A., Kumar, A., Dhanka, S., Maini, S., Rautela, K., Kant, S., Kaur, A., & Palta, A. (2026). Artificial Intelligence in Epidemic Modeling and Pandemic Preparedness: A Comprehensive Review. *Archives of Computational Methods in Engineering*, 33(5), 7085–7115. <https://doi.org/10.1007/s11831-026-10525-7>
 38. Sheikhtaheri, A., Tabatabaee Jabali, S. M., Bitaraf, E., TehraniYazdi, A., & Kabir, A. (2024). A near real-time electronic health record-based COVID-19 surveillance system: An experience from a developing country. *Health Information Management Journal*, 53(2), 145–154. <https://doi.org/10.1177/18333583221104213>
 39. Shen, Y.-T., Chen, L., Yue, W.-W., & Xu, H.-X. (2021). Digital Technology-Based Telemedicine for the COVID-19 Pandemic. *Frontiers in Medicine*, 8. <https://doi.org/10.3389/fmed.2021.646506>
 40. Soni, D., Van Haren, S. D., Idoko, O. T., Evans, J. T., Diray-Arce, J., Dowling, D. J., & Levy, O. (2020). Towards Precision Vaccines: Lessons From the Second International Precision Vaccines Conference. *Frontiers in Immunology*, 11. <https://doi.org/10.3389/fimmu.2020.590373>
 41. Sparrow, A., Smith-Torino, M., Shamamba, S., Chirakarhula, B., Lwaboshi, M., Benn, C., & Chumakov, K. (2024). A Risk Management Approach to Global Pandemics of Infectious Disease and Anti-Microbial Resistance. *Tropical Medicine and Infectious Disease*, 9(11), 280. <https://doi.org/10.3390/tropicalmed9110280>
 42. Struelens, M. J., Ludden, C., Werner, G., Sintchenko, V., Jokelainen, P., & Ip, M. (2024). Real-time genomic surveillance for enhanced control of infectious diseases and antimicrobial resistance. *Frontiers in Science*, 2. <https://doi.org/10.3389/fsci.2024.1298248>
 43. Tiwari, S., Dhakal, T., Kim, B.-J., Jang, G. S., & Oh, Y. (2025). Genomics in Epidemiology and Disease Surveillance: An Exploratory Analysis. *Life*, 15(12), 1848. <https://doi.org/10.3390/life15121848>
 44. Velmovitsky, P. E., Bevilacqua, T., Alencar, P., Cowan, D., & Morita, P. P. (2021). Convergence of Precision Medicine and Public Health Into Precision Public Health: Toward a Big Data Perspective. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.561873>
 45. Wilkinson, E., Giovanetti, M., Tegally, H., San, J. E., Lessells, R., Cuadros, D., Martin, D. P., Rasmussen, D. A., Zekri, A.-R. N.,

- Sangare, A. K., Ouedraogo, A.-S., Sesay, A. K., Priscilla, A., Kemi, A.-S., Olubusuyi, A. M., Oluwapelumi, A. O. O., Hammami, A., Amuri, A. A., Sayed, A., ... de Oliveira, T. (2021). A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science*, 374(6566), 423–431. <https://doi.org/10.1126/science.abj4336>
46. Willis, S. J., Cocoros, N. M., Randall, L. M., Ochoa, A. M., Haney, G., Hsu, K. K., DeMaria, A., & Klompas, M. (2019). Electronic Health Record Use in Public Health Infectious Disease Surveillance, USA, 2018–2019. *Current Infectious Disease Reports*, 21(10), 32. <https://doi.org/10.1007/s11908-019-0694-5>
47. Yang, J.-J., Li, J., Mulder, J., Wang, Y., Chen, S., Wu, H., Wang, Q., & Pan, H. (2015). Emerging information technologies for enhanced healthcare. *Computers in Industry*, 69, 3–11. <https://doi.org/10.1016/j.compind.2015.01.012>
48. Zhu, R., Pan, W., Liu, J., & Shang, J. (2024). Epileptic seizure prediction via multidimensional transformer and recurrent neural network fusion. *Journal of Translational Medicine*, 22(1), 895. <https://doi.org/10.1186/s12967-024-05678-7>