ARTIFICIAL INTELLIGENCE AND DATA SCIENCE IN ENVIRONMENT HEALTH AND SURVEILLANCE: EARLY DETECTION AND RISK-REDUCTION STRATEGIES

Akhere Ibhafidon[*], Ogechi Kate Onyekwere[**], Chinaenye D. Emelike[***], Jennifer Okonkwo[*], Stella Chikadibia Nwachukwu[*]and Tochi Emmanuel Iwuagwu[****]

[*]Department of Health Education, Alvan Ikoku Federal University of Education, Owerri, Imo State

[**]Department of Health Education, Federal College of Education, Eha-Amufu, Enugu State

[***]Department of Computer Science, Alvan Ikoku Federal University of Education, Owerri, Imo State

[****]Department of Human Kinetics and Health Education, University of Nigeria, Nsukka

Abstract

EHS (EHS) is undergoing rapid transformation through the integration of artificial intelligence (AI) and data science. In Africa, particularly Nigeria, rapid urbanization, industrialization, climate variability, and population growth have intensified environmental health threats while exposing limitations in traditional surveillance systems. This review synthesizes current evidence on the application of AI-driven methods in EHS, with emphasis on early detection and risk-reduction strategies. Drawing from studies indexed in Scopus, PubMed, Web of Science, and IEEE Xplore, the review examines methodological approaches including machine learning, deep learning, geospatial artificial intelligence (GeoAI), natural language processing, time-series forecasting, and multimodal data integration. Particular attention is given to the applicability of remote sensing, wearable technologies, health records, and administrative datasets in supporting predictive and preventive environmental health systems. The review further evaluates applications across air quality monitoring, water and soil contamination, vector-borne disease surveillance, and climate-driven hazards. Findings indicate that AI-enabled systems improve anomaly detection, exposure prediction, spatial risk mapping, and decision-support capacity. However, persistent challenges related to data quality, interoperability, explainability, ethical governance, infrastructural inequities, and contextual generalizability constrain implementation in low-resource settings. The article argues that the future of EHS lies not only in technological sophistication but also in transparent methodologies, equitable data infrastructures, and integration of AI outputs into operational public-health workflows. When responsibly implemented, AI and data science can shift EHS from reactive monitoring toward anticipatory prevention and resilient risk management.

Keywords: Artificial intelligence, Data science, Early detection, Risk reduction,Machine learning,GeoAI

Introduction

Environmental Health and Surveillance (EHS) serves as a cornerstone of public-health protection by tracking environmental exposures and identifying emerging hazards that threaten human health. EHS involves the continuous monitoring, assessment, interpretation, and dissemination of environmental and health-related data for prevention and control of environmental risks. In Africa, particularly Nigeria, rapid urbanization, industrialization, climate change, and population growth have increased the complexity of environmental health risks (Tshimula et al., 2024). Traditional surveillance systems, largely dependent on periodic sampling, laboratory analyses, and conventional statistical models, frequently operate reactively and may fail to capture rapidly changing environmental conditions.

Environmental exposures are often diffuse, chronic, and spatially heterogeneous, while associated health outcomes may emerge after prolonged latency periods. Consequently, conventional surveillance approaches are frequently characterized by delayed reporting, limited spatial resolution, fragmented datasets, and weak integration between environmental and health indicators. These limitations reduce the capacity for timely intervention and risk prevention.Recent advances in artificial intelligence (AI) and data science are reshaping this paradigm. Artificial intelligence encompasses computational techniques that enable machines to perform tasks typically associated with human cognition, including pattern recognition, forecasting, prediction, and automated decision-making (Olawade et al., 2024). Machine learning (ML), deep learning (DL), geospatial artificial intelligence (GeoAI), natural language processing (NLP), and predictive analytics increasingly support EHS through the analysis of complex and heterogeneous datasets.

The expansion of remote sensing technologies, sensor networks, Internet-of-Things (IoT) infrastructures, mobile devices, wearable systems, and digital platforms has generated unprecedented volumes of environmental and population-level data. AI-driven analytic systems can integrate these diverse datasets to identify hidden nonlinear relationships, detect anomalies, forecast exposure risks, and generate real-time alerts. Together, these developments are shifting environmental surveillance from reactive monitoring toward predictive and preventive intelligence.AI-enabled systems can forecast pollution episodes, identify contamination events, model vector migration, assess climate-sensitive exposures, and support environmental-health decision-making before hazards escalate into widespread health crises. This capacity is particularly important in low- and middle-income countries (LMICs), where environmental risks are growing rapidly while surveillance resources remain constrained.

This review critically examines the applicability of AI and data science in EHS, with particular emphasis on early detection and risk-reduction strategies. The review synthesizes current developments in data sources, analytic methodologies, surveillance applications, operational frameworks, and implementation challenges. Particular attention is devoted to remote sensing and satellite imagery, wearable and personal monitoring systems, and health and administrative records as critical components of AI-enhanced surveillance ecosystems. The article further explores methodological transparency, ethical

governance, contextual adaptation, and operational integration necessary for sustainable deployment of AI-driven surveillance systems.

Methodology

This paper adopted an integrative narrative review design informed by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) principles. The review combines evidence from environmental health, public health, data science, geospatial analytics, and artificial intelligence literature to provide a multidisciplinary synthesis of AI applications in EHS.A structured literature search was conducted across multiple electronic databases, including scopus, PubMed/MEDLINE, Web of Science, IEEE Xplore, Science Direct and Google Scholar.

These databases were selected to ensure comprehensive coverage of environmental health, epidemiology, artificial intelligence, computer science, geospatial analytics, and public-health surveillance literature.Searches were conducted using combinations of Boolean operators and keywords related to artificial intelligence and environmental surveillance. Major search terms included:

“Artificial intelligence” AND “EHS”

“Machine learning” AND “environmental monitoring”

“GeoAI” OR “geospatial artificial intelligence”

“Remote sensing” AND “public health”

“IoT” AND “environmental exposure”

“Deep learning” AND “air pollution prediction”

“Wearable sensors” AND “environmental health”

“Natural language processing” AND “disease surveillance”

“Risk reduction” AND “environmental hazards”

“Climate-driven hazards” AND “AI surveillance”

Reference lists of relevant publications were manually screened to identify additional eligible studies.

The review primarily included peer-reviewed publications published between 2018 and 2025 to capture recent methodological developments and emerging applications of AI-driven environmental surveillance systems.Studies were included if they:

i. Examined applications of AI, machine learning, deep learning, GeoAI, NLP, or data science in EHS.
ii. Focused on environmental exposures including air pollution, water contamination, soil toxicity, climate-related hazards, or vector-borne diseases.

iii. Addressed early detection, prediction, exposure assessment, risk mapping, or intervention support.
iv. Were empirical studies, systematic reviews, scoping reviews, methodological papers, or conceptual frameworks.
v. Were published in English.

Studies were excluded if they:
i. Focused exclusively on clinical AI unrelated to environmental exposures.
ii. Lacked sufficient methodological detail.
iii. Were conference abstracts without full texts.
iv. Addressed unrelated engineering applications without public-health relevance.

Titles and abstracts were independently screened for relevance. Full-text articles meeting eligibility criteria were subsequently reviewed. Duplicate records were removed during the screening process.Approximately 92 articles were initially identified across databases. After removal of duplicates and screening for relevance, 32 studies were retained for full-text review. Following detailed assessment against inclusion criteria, 17 studies formed the core evidence base for the review.The review employed thematic synthesis and critical narrative analysis. Evidence was organized into the following analytical domains:
i. Data sources and integration frameworks.
ii. Analytic approaches in AI-driven surveillance.
iii. Applications in environmental health domains.
iv. Early detection mechanisms.
v. Risk-reduction strategies.
vi. Ethical, technical, and operational challenges.
vii. Future directions and policy implications.
viii. Methodological transparency, conceptual depth, scalability, interpretability, and contextual applicability were emphasized during synthesis.

EHS is the systematic collection, analysis, interpretation, and dissemination of environmental and health data for the prevention and control of disease and environmental risk. Unlike traditional infectious-disease surveillance, environmental surveillance must contend with chronic exposures, multiple interacting determinants, spatial variability, and delayed health effects. Environmental hazards including air pollution, unsafe water, toxic waste exposure, climate-sensitive disease transmission, and industrial contamination contribute substantially to the global burden of disease. According to the World Health Organization (WHO, 2024), ambient (outdoor) air pollution contributes to millions of premature deaths annually. These environmental threats disproportionately affect populations in LMICs, where rapid urbanization and weak environmental governance exacerbate exposure risks.

Conventional environmental surveillance systems frequently suffer from fragmented datasets, delayed reporting mechanisms, inadequate laboratory infrastructure, and limited predictive capacity. In many African settings, surveillance remains heavily

dependent on manual sampling and isolated monitoring systems with poor temporal continuity.Data science and AI introduce opportunities to overcome these limitations through predictive modelling, anomaly detection, automated processing, real-time monitoring, and integration of heterogeneous data streams. The transition from reactive response toward anticipatory surveillance represents one of the most significant conceptual shifts in contemporary environmental health systems.

The effectiveness of AI in EHS depends fundamentally on the availability, quality, interoperability, and representativeness of data. Contemporary AI-driven surveillance systems increasingly rely on interconnected ecosystems of environmental, behavioural, geospatial, climatic, and health-related datasets.

IoT-enabled sensor systems provide continuous environmental monitoring across diverse contexts. These systems collect real-time information on air pollutants, particulate matter, volatile organic compounds, temperature, humidity, water quality parameters, and radiation exposure (Gamit et al., 2024).

The increasing deployment of low-cost environmental sensors has significantly transformed the architecture of EHS systems. Unlike traditional monitoring stations that are expensive and geographically sparse, IoT-enabled systems permit dense monitoring across urban, industrial, rural, and household environments. These sensor ecosystems continuously generate high-frequency environmental data capable of capturing rapidly changing environmental conditions that conventional surveillance approaches may overlook.

AI algorithms process these high-frequency datasets to detect anomalies, identify pollution spikes, and predict hazardous conditions. Sensor networks significantly enhance the temporal resolution of surveillance systems, enabling near-real-time detection and response. Machine-learning systems are particularly useful for recognizing hidden temporal patterns, correcting sensor errors, identifying outliers, and generating automated alerts before environmental exposures escalate into severe public-health threats.

Beyond technical monitoring, IoT infrastructures also support decentralized and community-based surveillance. In resource-constrained settings, portable sensors integrated with cloud-based analytics can provide low-cost alternatives for environmental-risk assessment. This is particularly relevant in many African countries where centralized environmental monitoring systems remain inadequate. Nevertheless, challenges involving calibration drift, inconsistent maintenance, network instability, and uneven geographical coverage continue to affect reliability and scalability of sensor-driven surveillance systems. However, sensor-based systems also introduce methodological challenges. Low-cost sensors may suffer calibration drift, inconsistent sensitivity, environmental interference, and incomplete coverage. AI-based calibration and data-imputation models increasingly address these limitations through automated correction mechanisms and predictive reconstruction.

Remote Sensing and Satellite Imagery

Remote sensing and satellite imagery constitute one of the most transformative developments in AI-driven environmental surveillance. Earth-observation technologies

provide continuous large-scale information on land use, vegetation cover, atmospheric composition, urban expansion, thermal anomalies, hydrological conditions, and environmental degradation (Iyer et al., 2025). The applicability of remote sensing in EHS has expanded considerably due to improvements in satellite resolution, cloud computing, and geospatial analytics. Satellite platforms now provide environmental observations at multiple spatial and temporal scales, allowing surveillance systems to monitor environmental change continuously across vast geographic regions. This capability is particularly valuable in regions where ground-based monitoring infrastructure is weak or absent.

Remote sensing technologies support indirect measurement of environmental-health indicators such as aerosol concentrations, land-surface temperature, vegetation density, water quality proxies, and urban heat-island intensity. When integrated with AI algorithms, these environmental indicators can be transformed into predictive models for disease risk, pollution exposure, ecological degradation, and climate-sensitive hazards.AI-driven analysis of satellite imagery supports: air-quality estimation, detection of wildfire smoke and dust events, flood and drought monitoring, heat-island assessment, vegetation and ecosystem surveillance, vector-habitat identification, water-body contamination mapping, urban environmental risk assessment among others.

Deep-learning architectures, particularly convolutional neural networks (CNNs), are increasingly used for image classification, environmental segmentation, land-cover analysis, and spatio-temporal prediction. These models identify subtle environmental changes that may indicate emerging hazards.GeoAI approaches combine satellite imagery with meteorological, demographic, mobility, and epidemiological data to create dynamic exposure maps and predictive risk models. According to Adefemi et al. (2023), Iyer et al. (2025), and VoPham et al. (2018), GeoAI enables scalable geospatial exposure assessment within large population-health databases.

Importantly, remote sensing also strengthens environmental-health equity research by identifying populations disproportionately exposed to environmental risks. Satellite-derived exposure models can reveal spatial inequalities in pollution burden, access to green infrastructure, flood vulnerability, and climate-related hazards. Such insights are increasingly valuable for environmental justice research and targeted public-health intervention.

Nevertheless, methodological concerns remain. Satellite imagery may vary in temporal frequency, spectral resolution, and atmospheric sensitivity. Environmental exposures inferred from satellite proxies may not perfectly reflect individual-level exposures. Additionally, many remote-sensing models remain computationally intensive and difficult to operationalize in resource-limited contexts. Despite these constraints, remote sensing remains indispensable for environmental surveillance in regions lacking extensive ground-monitoring infrastructure. In African contexts, satellite-supported AI systems may substantially improve environmental-health intelligence where conventional monitoring networks are sparse.

Satellite-derived environmental indicators are increasingly integrated into machine-learning frameworks to improve exposure assessment and risk prediction.

Traditional environmental monitoring frequently relied on sparse ground stations with limited spatial representativeness. In contrast, remote sensing enables scalable and geographically extensive surveillance.

Wearable and Personal Monitoring Technologies

Wearable devices and personal monitoring systems represent an emerging frontier in environmental exposure science. Portable biosensors, smart watches, mobile phones, and wearable environmental monitors enable individualized assessment of exposure patterns across time and space. Traditional environmental surveillance systems often estimate exposure at the community or regional level, assuming that populations within the same geographic area experience similar environmental conditions. However, exposure patterns are highly individualized and influenced by mobility, occupation, housing conditions, behavioural practices, and daily activities. Wearable technologies address this limitation by generating individualized, continuous, and context-sensitive exposure data.

Unlike conventional community-level surveillance, wearable technologies capture dynamic and personalized exposure trajectories. These systems monitor variables including particulate exposure, noise pollution, heat stress, ultraviolet radiation, physical activity, heart rate variability, respiratory parameters, and geographic mobility. AI algorithms process wearable-generated data to identify behavioural-environmental interactions, predict exposure vulnerabilities, and assess cumulative health risks. Machine-learning systems can detect patterns linking environmental conditions to physiological responses in real time. This capacity is particularly important for vulnerable populations such as children, older adults, outdoor workers, and individuals with chronic respiratory or cardiovascular diseases.

The applicability of wearable systems in EHS includes: personalized exposure assessment, identification of vulnerable populations, behavioural-environmental modeling, occupational exposure monitoring, climate and heat-stress surveillance, urban mobility and pollution mapping among others.

Real-time Health-Risk Alerts

Wearables also bridge a major methodological gap between ecological exposure estimates and individual-level environmental experience. By integrating wearable data with geospatial and atmospheric datasets, AI systems can generate more precise exposure models.Beyond surveillance, wearable systems may contribute to preventive public health by enabling personalized risk communication and adaptive behavioural interventions. For instance, AI-enabled mobile applications can warn individuals about hazardous pollution levels, unsafe temperatures, or excessive ultraviolet exposure based on real-time environmental conditions and physiological indicators.

However, substantial ethical and technical concerns accompany wearable surveillance. Issues of privacy, informed consent, data ownership, digital exclusion, and algorithmic bias remain unresolved. Wearable systems also generate extremely large volumes of continuous data, requiring robust computational infrastructure and secure

storage mechanisms. In LMIC contexts, cost, battery limitations, internet access, and unequal digital literacy may constrain adoption. Nevertheless, the integration of wearable systems into environmental-health surveillance presents significant opportunities for precision public health and context-sensitive risk prevention.

Health and Administrative Records

Health and administrative records provide critical linkage between environmental exposures and population-level health outcomes. Hospital admissions, toxicology reports, mortality databases, disease registries, insurance claims, emergency department visits, and laboratory records collectively support epidemiological analysis and environmental-risk attribution.

The integration of health and administrative datasets into AI-driven surveillance systems has substantially expanded the capacity of environmental epidemiology. These records enable researchers and public-health authorities to examine long-term relationships between environmental exposures and disease outcomes across large populations and extended time periods.AI systems increasingly integrate environmental datasets with health records to identify exposure-disease relationships, predict health outcomes, and support decision-making. Machine-learning models can detect complex nonlinear associations between environmental variables and disease incidence patterns. Such systems are capable of identifying subtle exposure-response relationships that may be difficult to detect using traditional statistical approaches. Applications include: predicting respiratory admissions during pollution episodes, linking heat exposure to cardiovascular mortality, identifying climate-sensitive disease outbreaks, monitoring environmental toxicities, forecasting healthcare demand during environmental crises, evaluating intervention effectiveness among others.

Administrative records also support longitudinal surveillance, enabling retrospective analysis of chronic exposure impacts over time. In addition, integration of environmental and administrative records supports population stratification and vulnerability assessment by linking exposure conditions with age, socio-economic status, geographic location, and pre-existing health conditions.Importantly, health-record analytics strengthen evidence-based policymaking by enabling evaluation of environmental interventions and public-health responses. AI-driven analysis of hospitalization trends, mortality records, and exposure histories can help determine whether pollution-control measures, climate-adaptation policies, or environmental regulations are achieving measurable health benefits.

Nevertheless, methodological transparency is essential when integrating health and environmental datasets. Many administrative databases suffer missing data, coding inconsistencies, reporting bias, and temporal delays. Ethical concerns surrounding confidentiality, geolocation tracking, and sensitive health information further complicate implementation.

Unstructured and Digital Data Streams

Digital platforms including social media, online news, search-engine queries, and citizen-reporting systems increasingly contribute to environmental surveillance ecosystems. Natural language processing (NLP) techniques enable automated extraction of information related to environmental events, disease outbreaks, public sentiment, odour complaints, contamination concerns, and hazard perception. Methuku et al. (2025) discussed automation of disease surveillance through NLP and AI, emphasizing its role in public-health intelligence. Although unstructured digital data may contain misinformation and demographic biases, these sources can provide valuable early-warning signals before conventional reporting systems detect emerging hazards.

Analytic Approaches in AI-Driven EHS

Artificial intelligence and data-science methods constitute the computational backbone of modern EHS. These methods transform complex, heterogeneous, and high-dimensional datasets into actionable environmental-health intelligence.

Machine Learning and Deep Learning

Machine-learning algorithms including random forests, support vector machines, gradient boosting, and neural networks are widely used for environmental prediction, classification, clustering, and exposure modelling.

Deep learning, a subset of machine learning, employs multilayered neural architectures capable of modelling complex nonlinear relationships within large datasets. Deep-learning systems are particularly valuable for image analysis, remote sensing, environmental forecasting, and spatio-temporal modelling. Deep learning for image or remote-sensing data: e.g., convolutional nets are designed to classify satellite imagery or detect land cover change relevant to vector spread or pollutant sources (Zheng et al., 2025).

Convolutional neural networks (CNNs) support environmental image classification and land-use detection, while recurrent neural networks (RNNs) and long short-term memory (LSTM) architectures facilitate time-series forecasting for pollution and climate variables. Banjara et al. (2025) demonstrated that deep-learning methods can reconstruct incomplete water-quality time series, thereby improving continuity and predictive accuracy.The applicability of machine learning in environmental surveillance includes: air-quality forecasting, contamination prediction, exposure modeling, climate-risk forecasting, disease-vector prediction, environmental anomaly detection, dynamic risk mapping, automated environmental classification among others. Despite high predictive performance, many deep-learning systems remain difficult to interpret. The “black-box” nature of these algorithms presents challenges for accountability, transparency, and public trust.

Geospatial Artificial Intelligence (GeoAI)

GeoAI integrates geographic information systems (GIS), spatial statistics, remote sensing, and machine learning to support environmental-health analysis. Traditional exposure

assessments frequently relied on static or coarse environmental proxies. GeoAI now enables scalable, high-resolution exposure assessment within large population-health databases (Adefemi et al., 2023; Iyer et al., 2025; VoPham et al., 2018).GeoAI methodologies support: Spatial hotspot identification, Environmental vulnerability mapping, Vector-habitat modeling, Climate-sensitive disease prediction, Urban environmental analysis, Environmental injustice assessment, Spatial intervention targeting. By integrating spatial, demographic, climatic, and epidemiological variables, GeoAI supports highly contextualized environmental-health surveillance systems. Importantly, GeoAI contributes not only to hazard detection but also to environmental-health equity analysis by identifying populations disproportionately exposed to environmental risks.

Natural Language Processing (NLP)

Natural language processing (NLP) enables extraction and interpretation of information from unstructured textual datasets. The NLP approaches are increasingly used to analyze: Social-media posts, Environmental complaints, News reports, Public-health bulletins, Toxicological reports, Community risk perceptions among others. Methuku et al. (2025) emphasized the role of NLP in automating disease surveillance and public-health intelligence.

The NLP systems may detect environmental-health concerns before official systems recognize emerging hazards. For example, clusters of social-media complaints regarding unusual odours, respiratory irritation, or water discoloration may indicate contamination events. However, NLP-derived surveillance data are susceptible to misinformation, demographic bias, linguistic variability, and inconsistent geolocation.

Time-Series Forecasting and Anomaly Detection

Environmental surveillance increasingly depends on predictive time-series modelling and anomaly detection. Machine-learning systems trained on historical environmental datasets can forecast future exposure conditions and identify deviations from expected environmental patterns. Applications include forecasting pm2.5 levels, predicting heat waves, anticipating harmful algal blooms, detecting water-quality anomalies, predicting vector outbreaks, monitoring industrial emissions among others. These methods support anticipatory interventions and operational preparedness.

Explainable AI (XAI)

Explainable AI seeks to improve transparency and interpretability in AI-driven decision systems.McGovern et al. (2021) identified the “black-box” nature of AI systems as a major implementation barrier in environmental monitoring. The XAI approaches attempt to clarify: Why predictions are generated, Which variables drive model outputs, how uncertainty influences predictions, whether models exhibit bias. Explainability is particularly important in environmental-health contexts where policy decisions may affect vulnerable populations.

Edge Computing and Real-Time Analytics

Edge computing processes environmental data near the point of collection rather than relying exclusively on centralized cloud infrastructure. This approach reduces latency, enhances responsiveness, and supports real-time hazard detection. Edge analytics are particularly valuable for IoT-enabled environmental surveillance systems operating in resource-constrained settings.

Domains and Applications of AI in EHS

i. Air Quality Monitoring and Prediction

Air-quality surveillance represents one of the most advanced domains of AI-driven environmental monitoring. AI systems integrate sensor data, satellite imagery, meteorological variables, and traffic information to predict pollutant concentrations and identify pollution hotspots. Gryech et al. (2024) conducted a systematic review demonstrating the effectiveness of ML and IoT systems in outdoor air-pollution monitoring and forecasting. AI-driven air-quality surveillance enables: real-time pollution forecasting, exposure hotspot detection, urban environmental planning, community-level warnings, health-risk communication, evaluation of emission-control strategies among others .These systems provide substantially finer spatial and temporal resolution than conventional fixed-monitoring stations.

ii. Water Quality and Waterborne Hazard Surveillance

AI-enhanced water surveillance integrates IoT systems, satellite imagery, sensor networks, and predictive analytics. Deep-learning algorithms support detection of contamination anomalies, prediction of pollutant transport, and reconstruction of incomplete datasets (Banjara et al., 2025; Liang et al., 2025). Applications include: monitoring heavy metals, predicting microbial contamination, flood-related contamination surveillance, harmful algal bloom detection, waterborne disease prevention among others.

iii. Soil and Chemical Hazard Surveillance

AI systems increasingly support soil contamination assessment and chemical-risk modelling. Guo et al. (2025) demonstrated that machine-learning models can prioritize contaminated sites and optimize remediation strategies. Remote sensing combined with AI supports: toxicity mapping, industrial contamination assessment, agricultural chemical monitoring, exposure-risk modeling, land restoration prioritization among others.

iv. Vector-Borne and Climate-Driven Hazards

Climate variability is altering vector ecology, disease transmission patterns, and environmental exposures. GeoAI and ecological modelling support prediction of vector migration, habitat suitability, and climate-sensitive disease risks (Iyer et al., 2025). Applications include: malaria risk forecasting, dengue hotspot prediction, flood-risk assessment, heat-wave surveillance, wildfire smoke modeling, climate-sensitive exposure mapping among others. These approaches support proactive preparedness and targeted intervention.

v. Personal and Community-Level Exposure Assessment
Wearables and mobile sensors allow individualized exposure assessment across dynamic environmental contexts. AI systems aggregate these datasets to generate: exposure heatmaps, mobility-linked pollution estimates, occupational exposure profiles, physiological stress indicators, behavioural-environmental interaction models among others. These systems advance precision public health by integrating environmental, behavioural, and physiological dimensions of exposure.

Early Detection Strategies Through AI-Enabled Surveillance
The central value of AI-driven surveillance lies in its capacity for early detection and anticipatory response. AI-based early detection systems integrate heterogeneous environmental and health datasets to identify hazards before they escalate into significant public health events. For instance, AI-driven Internet of Things (IoT) systems have been employed to monitor and predict indoor air pollution in rural African households, providing real-time data and predictive analytics to mitigate health risks (Ooko & Rweyemamu, 2024). Other strategies adopted for early detection include:

i. Baseline modelling and anomaly detection: Using historical sensor data, satellite imagery and health outcome data, ML models can learn expected patterns and flag deviations.
ii. Forecasting future exposures or risk: Time series forecasting (e.g., pollutant levels, vector habitat suitability) allows pre-emptive warnings a few hours/days ahead.
iii. Spatial hot spot identification: GeoAI methods can identify which geographic zones are at increased risk, based on dynamic variables (e.g., land cover change, meteorology, demographics) and thereby detect new or emerging risk zones. (VoPham et al., 2018)
iv. Use of unstructured digital data (Digital Epidemiology and NLP): NLP models on social-media, news, sensor logs or citizen reports may detect early signals of environmental-health events (e.g., smell complaints, water discolouration) before formal monitoring captures them.
v. Data fusion and multi-source integration: Combining environmental, health, mobility, and socio-demographic data increase the signal strength and detection accuracy.
vi. Real-time streaming and alert systems: IoT networks providing real-time or near-real-time data allow rapid detection and alerts; AI models reduce lag substantially compared to traditional systems. (Gamit et al., 2024).
vii. Social-media mining and textual analysis provide early indicators of environmental-health concerns.

These approaches collectively support earlier intervention, improved preparedness, and prevention of exposure escalation.

Risk-Reduction Strategies Enabled by AI and Data Science

Detection alone is insufficient unless linked to actionable interventions. AI and data science therefore support environmental-health risk reduction through predictive decision systems.

Vulnerability Mapping and Risk Stratification

GeoAI enables integration of exposure, demographic, and socio-economic datasets to identify vulnerable populations and high-risk communities.

i. Resource Optimization

Machine-learning systems support strategic allocation of sensors, monitoring campaigns, vector-control measures, and mitigation resources.

ii. Scenario Modelling and Preparedness Planning

Predictive models simulate future exposure scenarios under climate change, urbanization, and environmental degradation.

iii. Intervention Evaluation

AI systems assess effectiveness of environmental interventions and refine surveillance models through continuous feedback loops.

iv. Behavioural and Policy Support

Environmental-health analytics support evidence-based urban planning, zoning policies, green infrastructure development, and public advisories.

These capabilities shift environmental-health systems from reactive crisis management toward proactive prevention.

v. Ensuring Environmental Sustainability: Consider the energy and resource costs of AI systems as part of environmental responsibility (Samuel & Lucassen, 2022).

Benefits of AI and Data Science in EHS

The adoption of AI and data science offers several major advantages:

i. Early detection of environmental hazards: AI enables timely identification of anomalies or hazardous trends.Enhanced precision: Integration of multimodal data provides finer spatial and temporal insights.
ii. Timeliness: The sooner a hazard is detected, the more time there is for intervention and reducing exposures and health impacts.
iii. Enhanced preparedness and intervention targeting: Early detection allows focusing resources on high-risk areas/populations rather than blanket responses.
iv. Prevention of escalation: By detecting anomalies early, systems can help prevent a small exposure event from evolving into a large scale health impact.

v. Improved efficiency: Data-driven early detection reduces resource wastage (e.g., monitoring in the wrong place/time) and enhances resource allocation.
vi. Better evaluation and feedback through dashboards and visualizations: Early detection systems enable continuous monitoring and adjustment of strategies based on evolving data.
vii. Efficiency: Automated analysis reduces labour demands and accelerates decision cycles.
viii. Complex pattern recognition: AI can detect nonlinear and multi-factorial relationships among environmental variables.
ix. Improved communication: AI-powered visualizations and dashboards enhance stakeholder understanding and engagement.

Improved spatial and temporal precision.

Enhanced predictive capacity.

Automated processing of large datasets.

Real-time environmental monitoring.

Improved allocation of surveillance resources.

Identification of complex nonlinear relationships.

Together, these capabilities foster a paradigm shift from reaction to anticipation; from responding to hazards after exposure to predictive and preventive surveillance systems.(preventing them through foresight).

Challenges and Limitations

Despite significant promise, AI-driven environmental surveillance faces major methodological, ethical, and operational constraints.

i. Data quality and representativeness: The effectiveness of AI models is contingent upon the quality and availability of data. Incomplete, biased, or inaccurate data can lead to unreliable predictions and undermine public trust in AI-driven systems. Sensor malfunctions and regional data disparities compromise reliability (Guo et al., 2025).
ii. Transparency and interpretability: AI models, particularly deep learning algorithms, are often criticized for their "black-box" nature, making it challenging to interpret their decision-making processes Deep-learning models can lack explainability, complicating trust and accountability (McGovern et al., 2021). Also, many AI/ML models are “black boxes”, raising issues of trust among public-health practitioners
iii. Ethical, privacy and governance issues: The deployment of AI in EHS raises ethical and privacy issues, including data security, informed consent, and potential misuse of information. Data-driven monitoring must protect personal information and avoid reinforcing environmental inequities (Iyer et al., 2025). Privacy concerns arise when integrating personal health and geolocation data, necessitating strong data protection frameworks. In addition, use of sensor networks, health records, social-media data raises privacy concerns; algorithmic bias could disadvantage vulnerable populations.

iv. Technical and infrastructural constraints/gaps: Implementing AI systems requires computing capacity, trained personnel, and maintenance to avoid model drift (Gamit et al., 2024). Infrastructure constraints, including limited high-speed internet and computing resources, may hinder large-scale deployment in rural areas.
v. Model generalizability and scalability: Context-specific training may limit transferability across regions and climates (Gryech et al., 2024). A model trained in one region may not work well elsewhere due to differing environmental, socioeconomic or demographic factors. (VoPham et al., 2018)
vi. Actionability: Early detection is useful only if it triggers timely, effective interventions, systems must be embedded in public health or environmental health workflows. Predictive outputs must be integrated into public health decision frameworks to translate insights into real-world impact (Mendes et al., 2025).
vii. Heterogeneous data integration: Combining data from sensors, satellites, health records, social media requires harmonisation of formats, spatial/temporal alignment, and dealing with missingness.
viii. Real time processing constraints: To be effective, early detection systems must process data rapidly and reliably; this demands robust infrastructure, especially in resource-constrained settings.

Operational Integration Predictive systems only generate impact when integrated into public-health workflows and decision-support systems (Mendes et al., 2025)
Resource Inequities

Many LMICs face limited access to surveillance infrastructure, computational capacity, and technical training.

Conclusion

Artificial intelligence and data science are reshaping EHS by enabling predictive, scalable, and increasingly real-time monitoring systems. Through integration of remote sensing, IoT infrastructures, wearable technologies, administrative records, and advanced analytics, AI-driven surveillance systems enhance early detection, exposure assessment, and risk forecasting across multiple environmental-health domains. This review demonstrates that machine learning, deep learning, GeoAI, NLP, and predictive analytics are increasingly central to environmental health intelligence. Applications now extend across air-quality prediction, water contamination surveillance, vector-borne disease forecasting, climate-risk modelling, and individualized exposure assessment. However, technological advancement alone is insufficient. Persistent challenges involving data quality, interpretability, privacy, infrastructural inequities, and contextual transferability continue to constrain operational implementation, particularly in resource-limited settings.

The future of EHS therefore depends on responsible implementation frameworks emphasizing transparency, interoperability, ethical governance, explainable AI, and equitable access to digital infrastructure. Cross-disciplinary collaboration between

environmental scientists, public health practitioners, data scientists, policymakers, and communities will be essential. When responsibly implemented, AI and data science can shift EHS from reactive response toward anticipatory prevention, thereby strengthening resilience, reducing environmental risk, and improving population health outcomes.Based on the reviewed evidence and identified gaps, the following recommendations are proposed:

Build Integrated Environmental Data Ecosystems

Expand deployment of sensor networks (air, water, soil, climate, wearable exposures) remote-sensing infrastructures, and interoperable surveillance systems, particularly in underserved or data scarce regions. This is best done by developing robust data pipelines that integrate multi modal data (sensor time-series, satellite/remote sensing imagery, health outcome records, socioeconomic data) with spatial and temporal context:

i. Promote Explainable and Ethical AI

Prioritize explainable AI frameworks, transparency audits, algorithmic fairness, and privacy protection mechanisms.

ii. Strengthen Capacity Building

Develop interdisciplinary training programmes linking environmental health, epidemiology, geospatial science, and data analytics.

iii. Enhance Operational Integration

Integrate AI outputs directly into environmental-health workflows, emergency-response systems, and public health decision support platforms.

iv. Improve Contextual Adaptation

Develop transfer-learning and context-sensitive modelling approaches suitable for African and other LMIC environments.

v. Expand Community Engagement

Engage local communities in surveillance design, implementation, interpretation, and risk communication.

vi. Standardize Governance and Interoperability

Develop national and international standards for environmental health data sharing, metadata harmonization, and AI governance.

vii Prioritize Impact Evaluation

Future research should move beyond predictive accuracy toward evaluation of real-world health outcomes, intervention effectiveness, and equity impacts.

viii Encourage Cross-Sector Collaboration

Governments, universities, technology providers, and environmental health agencies should collaborate in developing scalable and context-appropriate AI systems.

ix Promote Sustainable and Inclusive AI Systems

Embedding AI into decision support by developing operational early-warning and visualization tools that integrate directly into agency workflows.

Cross-disciplinary collaboration between public health professionals, data scientists, and policymakers will be critical for sustainable surveillance solutions in Nigeria and other African countries.

Governments and health agencies should foster partnerships with universities and technology providers to co-develop AI tools tailored to local contexts.
Public engagement and transparency are necessary to build trust in AI-based surveillance systems and encourage community participation in risk reduction initiatives.

References

Adefemi, A., Ukpoju, E. A., Adekoya, O., Abatan, A., & Adegbite, A. O. (2023).Artificial intelligence in environmental health and public safety: A comprehensive review of USA strategies. *World Journal of Advanced Research and Reviews, 20*(3), 1420–1434. https://doi.org/10.30574/wjarr.2023.20.3.2591

Banjara, I. P., Poudel, S., Pariyar, K., Upreti, D., Zafeirakou, A., & Paudel, S. R. (2025). AI for enhanced water quality data imputation: A deep learning perspective. *Environmental Science: Advances, 4*, 817–823. https://doi.org/10.1039/D4VA00367E

Gamit, V., Vasava, M. J., & Patel, A. S. (2024). AI-Powered environmental surveillance: Enhancing air and water quality monitoring through real-time predictive analytics.*ShodhKosh: Journal of Visual and Performing Arts, 5*(6), 3157–3162. https://doi.org/10.29121/shodhkosh.v5.i6.2024.6169

Gryech, I., Assad, C., Ghogho, M., &Kobbane, A. (2024). *Applications of machine learning and IoT for outdoor air pollution Monitoring and prediction: A systematic literature review*. arXiv:2401.01788[cs] https://arxiv.org/abs/2401.01788

Guo, Z., Huong, L., Yan, J., Zhang, H., Tia, X., Li, M., & Li, H. (2025). Artificial intelligence technology in environmental research and health: Development and prospects. *Environmental International, 203*, 109788. https://doi.org/10.1016/j.envint.2025.109788

Iyer, H. S., Karasaki, S., Yi, L., Hiswen, Y., James, P., &Vopham, T. (2025). Harnessing geospatial artificial intelligence (GeoAI) for environmental epidemiology: A narrative review. *CurrentEnvironmental Health Reports, 12*(34). https://doi.org/10.1007 /s40572-025-00497-4

Liang, Y., Ding, F., Liu, L., Yin, F., Hao, M., Kang, T., Zhao, C., Wang, Z., & Jang, D. (2025). Monitoring water quality parameters in urban rivers using multi-source

data and machine learning approach. *Journal of Hydrology, 648*, 132394. https://doi.org/10.1016/j.jhdrol.2024.132394

McGovern, A., Ebert-Uphoff, I., Gagne, D. J. II, & Bostrom, A. (2021). *The need for ethical, responsible, and trustworthy artificial intelligence for environmental sciences.* arXiv:2112.08453 [cs]. https://arxiv.org/abs/2112.08453

Mendes, V. S., Mendes, B. M. F., Moura, R. P., Lourenço, I. W., Oliveira, M. F. A., Ng, K. L., & Pinto, C. S. (2025). Harnessing artificial intelligence for enhanced public health surveillance: A narrative review. *Frontiers in Public Health, 13*, 1601151. https://doi.org/10.3389/fpubh.2025.1601151

Methuku, V. (2025). NLP and AI for public health intelligence: Automating disease surveillance from unstructured data ICCK. *Transactions on Emerging Topics in Artificial Intelligence,2*(1), 43-56.https://doi.org/10.62762/TETAI.2025.222799

Olawade, D. B., Wada, O. Z., Ige, A. O., Egbewole, B. I., Olojo, A., & Oladapo, B. I. (2024). Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions. Hygiene and Environmental Health Advances, 100114. https://doi.org/10.1016/j.heha.2024.100114

Ooko, S. O., & Rweyemamu, E. (2024). Monitoring and predicting African rural household air pollution using Internet of Things and artificial intelligence. *Pan-African Journal of Health and Environmental Science, 3(*1), 59–73. https://doi.org/10.56893/ajhes2024v03i01.06

Samuel, G., & Lucassen, A. M. (2022). The environmental sustainability of data-driven health research: A scoping review. *Digital Health*. https://doi.org/10.1177/2055207622111129

Tshimula, J. M., Kalengayi, M., Makenga, D., Lilonge, D., Asumani, M., Madiya, D., Kalonji, E. N., Kanda, H., Galekwa, R. M., Kumbu, J., Mikese, H., Tshimula, G., Muabila, J. T., Mayemba, C. N., Nkashama, D. K., Kalala, K., Ataky, S., Basele, T. W., Didier, M. M., Kasereka, S. K., Dialufuma, M. V., Kumwita, G. I. W., Muyuku, L., Kimpesa, J., Muteba, D., Abedi, A. A., Ntobo, L. M., Bundutidi, G. M., Mashinda, D. K., Mpinga, E. K., & Kasoro, N. M. (2024). Artificial intelligence for public health surveillance in Africa: Applications and opportunities. *arXiv:2408.02575v1 [cs].* https://doi.org/10.48550/arXiv.2408.02575

VoPham, T., Hart, J. E., Laden, F., & Chiang, Y.-Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): Potential applications for environmental epidemiology. *Environmental Health, 17*, 40. https://doi.org/10.1186/s12940-018-0386-x

World Health Organization. (2024). *Ambient (outdoor) air pollution*. https://www.who.int/westernpacific/newsroom/factsheets/detail/ambient%28outdoor%29-air-quality-and-health

Zheng, W., Li, K., & Liu, X. (2025). GeoAI for land use observations, analysis, and forecasting. *Land, 14*(10), 2058. https://doi.org/10.3390/land14102058