International Journal of Engineering & Tech Development

Volume 2, Issue 9 – September 2026 • E-ISSN: 3067-672X P-ISSN: 3067-669X

AI-Enabled Early Detection of Emerging Health Risks: Integrating Population Data and Biological Signals

Dr. Wouter de Waal

Professor, Leiden University Medical Center (LUMC), Netherlands

Abstract

Emerging health risks may develop through infectious disease transmission, environmental exposure, antimicrobial resistance, changing pathogen characteristics, occupational hazards, or shifts in population vulnerability. Conventional surveillance systems remain essential, but they can be affected by reporting delays, fragmented databases, incomplete geographic coverage, and limited integration of biological and environmental information. Artificial intelligence may strengthen early detection by analysing patterns across clinical reports, laboratory results, population mobility, environmental measurements, digital activity, wearable sensors, and genomic surveillance.This paper examines the potential contribution of AI-enabled multimodal surveillance to the early identification of emerging health risks. A conceptual review is combined with an illustrative simulation in which an early-warning sensitivity index increases from 58 when clinical reports are used alone to 91 after population trends, environmental indicators, biological signals, and genomic surveillance are integrated. These values illustrate a methodological principle and are not clinical surveillance results. The analysis indicates that multimodal systems may detect weak signals earlier, distinguish local anomalies from wider patterns, and support more timely public-health investigation. However, combining multiple data sources can also increase false alerts, surveillance bias, privacy risk, cybersecurity exposure, and dependence on opaque algorithms.

The study argues that AI alerts should initiate structured epidemiological review rather than automatically trigger restrictive public-health interventions. Responsible implementation requires representative data, temporal and geographic validation, uncertainty reporting, transparent alert thresholds, human oversight, data minimization, community accountability, and continuous monitoring for unequal performance. AI can improve public-health preparedness when it supplements laboratory confirmation, clinical expertise, field investigation, and established surveillance infrastructure. Its public value should be evaluated through timeliness, sensitivity, specificity, interpretability, equity, actionability, and the proportionality of any response generated from its predictions.

Keywords: artificial intelligence; early warning; public-health surveillance; population data; biological signals; digital epidemiology; health-risk detection; genomic surveillance

1. Introduction

Public-health surveillance seeks to identify changes in disease, exposure, or population vulnerability before they develop into widespread harm. Timely detection can support epidemiological investigation, laboratory testing, clinical preparedness, targeted communication, and allocation of limited resources. Delayed recognition, by contrast, can allow transmission or exposure to expand before decision-makers understand its scale.

Traditional surveillance draws on clinical diagnoses, mandatory notifications, laboratory results, mortality records, pharmacy data, and structured epidemiological reports. These sources provide essential evidence, but information often becomes available only after a patient develops symptoms, seeks care, receives testing, and enters a reporting system. Populations with limited access to health services may consequently remain underrepresented.

Digital systems have expanded the range of observable health-related signals. Search activity, mobility patterns, emergency calls, wastewater measurements, environmental sensors, wearable devices, genomic sequences, veterinary reports, and anonymized service-use data may reveal changes before they are visible through conventional reporting. Each source is incomplete when considered alone. Search patterns can reflect media attention, mobility data may exclude people without digital access, and wearable information may be concentrated among comparatively affluent users.

Artificial intelligence can analyse relationships across these heterogeneous sources. An algorithm may identify unusual spatial clusters, changes in symptom patterns, correlated environmental conditions, emerging pathogen sequences, or deviations from historical baselines. Machine learning is particularly useful when the signal is distributed across many variables and does not correspond to a single predefined threshold.

The use of AI does not eliminate uncertainty. A statistical anomaly may reflect an outbreak, a reporting change, device malfunction, seasonal variation, migration, media coverage, or a change in health-service access. An alert must therefore be interpreted within epidemiological and social context. This paper evaluates how AI can integrate population and biological information while maintaining scientific validity, privacy, equity, and public accountability.

2. Review of Literature

2.1 Digital Epidemiology and Population-Level Signals

Digital epidemiology uses information generated inside and outside health-care systems to study population-health patterns. Salathé and colleagues described how internet searches, social media, mobile devices, and digital communication can contribute to infectious-disease monitoring. These sources may provide rapid signals, but their relationship with actual disease incidence can change over time.

Brownstein, Freifeld, and Madoff demonstrated the value of informal online information for epidemic intelligence. News reports and other web-based sources can complement official surveillance, particularly when verified through structured review. Their work also showed that automation requires human interpretation because duplicated, outdated, or inaccurate information can produce misleading signals.

Pivetta and colleagues compared search-engine and social-media information for epidemic detection. Their findings reinforced the idea that different digital sources have different predictive characteristics. A data stream that performs well for one disease or population may not generalize to another.

Budd and colleagues examined digital technologies used during a major public-health emergency. They emphasized that digital surveillance, mobility analysis, clinical informatics, and contact technologies can support public-health response, while also creating concerns involving effectiveness, privacy, access, and public trust.

2.2 Biological, Environmental, and Genomic Indicators

Biological surveillance includes clinical laboratory results, pathogen sequencing, immune markers, wastewater indicators, and physiological measurements. These signals may provide information about pathogen presence, transmission, or population exposure before routine case counts are complete.

Genomic surveillance can identify mutations, transmission relationships, introductions, and changes in pathogen populations. Its usefulness depends on representative sampling, timely sequencing, consistent metadata, and coordination between laboratories and epidemiological teams. An apparent genomic pattern may be misleading when sampling is concentrated in a small number of locations.

Environmental indicators can support detection of risks linked to air quality, temperature, water contamination, vectors, allergens, or toxic exposure. These observations require geographic linkage and careful attention to latency because an environmental change may precede health outcomes by different intervals.

Wearable devices can provide heart rate, sleep, temperature, oxygen saturation, and activity information. Aggregated physiological changes may help identify population-level anomalies, but device ownership is not evenly distributed. Consumer sensors also differ in accuracy and cannot replace diagnostic testing.

2.3 Artificial Intelligence in Health Surveillance

Machine-learning models can perform anomaly detection, classification, forecasting, spatial risk mapping, and natural-language processing. Anomaly detection is useful when the emerging threat is not yet well defined because the model can identify deviations from an established baseline.

Natural-language processing can extract disease, symptom, location, and timing information from reports. Supervised classifiers can rank signals according to relevance, while time-series systems can estimate whether an observed change exceeds seasonal expectations.

WHO guidance on AI for health emphasizes autonomy, safety, transparency, accountability, inclusiveness, and sustainability. These principles are directly relevant to health surveillance because algorithmic outputs can influence resource allocation, movement restrictions, screening programs, or public communication.

3. Objectives of the Study

3.1 Analytical Objectives

The study aims to examine how clinical, population, environmental, biological, and genomic information can be integrated into an AI-supported early-warning system. The focus is on information complementarity rather than unrestricted data accumulation.

A related objective is to distinguish early-warning sensitivity from confirmed detection. An algorithmic alert represents a signal requiring investigation; it does not independently establish the existence, cause, severity, or geographic extent of a health threat.

3.2 Public-Health Objectives

The second objective is to evaluate how AI-generated alerts can support field investigation, laboratory testing, preparedness, and targeted preventive action. The study considers whether a signal is timely, interpretable, geographically specific, and relevant to a practical response.

The paper also examines the risk that an inaccurate or poorly communicated warning may produce unnecessary anxiety, resource diversion, discrimination, or loss of public trust.

3.3 Governance Objectives

The third objective is to identify the safeguards required for responsible population-health analytics. These include privacy, data minimization, cybersecurity, bias assessment, transparency, community engagement, proportionality, and human oversight.

The study seeks to position AI as a support mechanism operating within accountable public-health institutions rather than as an autonomous decision-maker.

4. Research Methodology

4.1 Research Design

The study uses a conceptual and simulation-based design. It synthesizes literature from epidemiology, public-health informatics, artificial intelligence, environmental health, digital surveillance, and ethics. No patient records, identifiable mobility data, biological samples, or live surveillance information were collected.

Relevant literature was identified through PubMed, Scopus-indexed journals, multidisciplinary databases, and authoritative public-health sources. Search concepts included “AI health surveillance,” “digital epidemiology,” “early outbreak detection,” “biological signals,” “genomic surveillance,” and “population health early warning.”

4.2 Multimodal Surveillance Framework

Five surveillance layers were included in the conceptual model.

Table 1: Data Layers in an AI-Enabled Early-Warning System

Data layer	Representative information	Primary analytical contribution
Clinical reports	Symptoms, diagnoses, admissions and notifications	Establishes direct evidence of health-service demand
Population trends	Mobility, absenteeism, search activity and service use	Identifies changing behavior or symptom-seeking patterns
Environmental signals	Air, water, weather, vectors and wastewater	Provides exposure and community-level indicators
Biological signals	Laboratory results and aggregated physiological changes	Detects biological abnormalities and possible exposure

Data layer	Representative information	Primary analytical contribution
Genomic surveillance	Pathogen sequences and variant patterns	Supports characterization and transmission analysis

Source: Conceptual framework developed for the present study.

Each layer has distinct sampling and measurement limitations. Integration should therefore preserve the provenance, uncertainty, time resolution, and geographic coverage of every input.

4.3 Illustrative Simulation

The simulation assigned an early-warning sensitivity index of 58 to a system using clinical reports alone. The index increased to 68 after population trends were added, 76 after environmental signals, 85 after biological indicators, and 91 after genomic surveillance.

These values are hypothetical. They do not represent sensitivity percentages, diagnostic accuracy, outbreak predictions, or results from an operational health system. The simulation demonstrates how complementary information might improve signal detection under favorable assumptions.

5. Analysis

5.1 Integration of Weak and Heterogeneous Signals

Emerging threats often produce weak signals across several systems before they generate a clear rise in confirmed cases. A small increase in respiratory complaints may be insignificant by itself. When accompanied by unusual absenteeism, wastewater detection, physiological anomalies, and a related genomic signal, the combined evidence may justify investigation.

AI models can assign weights to these observations and evaluate their temporal and geographic alignment. The model should distinguish correlation from confirmation. Multiple data streams may be influenced by the same external event, such as intense media coverage, and therefore cannot always be treated as independent evidence.

The integration process should also account for reporting delays. Clinical admissions, laboratory confirmation, wastewater samples, and genomic sequences become available at different times. A valid system must preserve the chronology of the information to avoid falsely claiming earlier detection through data that were unavailable at the decision point.

5.2 Illustrative Results

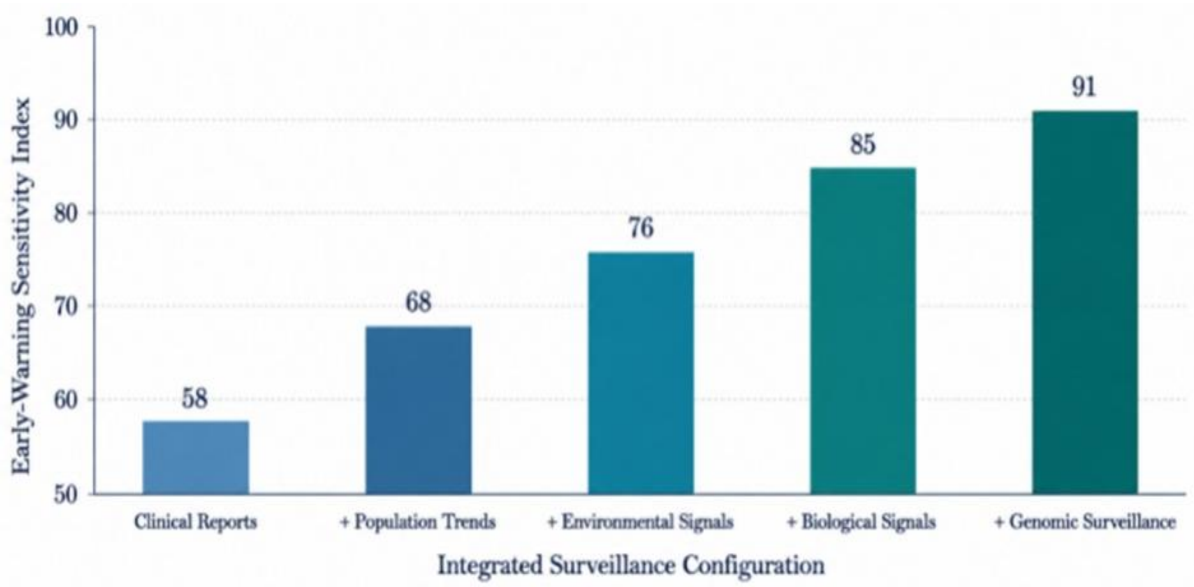
Table 2: Simulated Early-Warning Sensitivity Across Data Configurations

Surveillance configuration	Sensitivity index	Incremental gain	Cumulative gain
Clinical reports	58	—	—
Clinical reports and population trends	68	10	10
Addition of environmental signals	76	8	18
Addition of biological signals	85	9	27
Addition of genomic surveillance	91	6	33

Source: Author-developed illustrative simulation; values are not clinical surveillance results.
The simulated index increases with each data layer. The largest conceptual gain occurs when information outside delayed clinical reporting is incorporated. The smaller final increment reflects the illustrative assumption that some evidence overlaps with information already captured by biological surveillance.

The pattern should not be interpreted to mean that every additional dataset improves performance. Poor-quality or biased information can reduce accuracy, increase false alerts, and obscure genuine threats.

Graph 1: Illustrative Effect of Data Integration on Early-Warning Sensitivity



Supporting data: Clinical reports = 58; population trends = 68; environmental signals = 76; biological signals = 85; genomic surveillance = 91.
Source: Author-developed illustrative simulation; values are not clinical surveillance results.
Interpretation: The graph demonstrates progressive improvement within the simulated system as complementary surveillance layers are integrated. Population and environmental data may provide earlier contextual signals, while biological and genomic evidence can improve specificity and characterization.

Key Finding: Multimodal integration may improve early-warning capability, but operational benefit depends on data quality, timely access, independent validation, and a clear investigation protocol.

5.3 Alert Generation and Public-Health Response

A useful alert should identify the location, time period, contributing signals, uncertainty, and reason for deviation from the baseline. A single unexplained risk score provides inadequate support for public-health action.

Alert thresholds should reflect the seriousness of the potential threat and the cost of investigation. A low threshold may identify more threats but generate excessive false alarms. A high threshold may reduce unnecessary investigations while delaying genuine warnings.

AI output should lead to a tiered response. Initial alerts may trigger data-quality review and comparison with known events. Persistent signals can lead to laboratory testing, local clinical review, field investigation, or enhanced sampling. Restrictive measures require substantially stronger evidence than an algorithmic anomaly.

6. Challenges

6.1 Data Quality, Missingness, and Concept Drift

Surveillance data are affected by changes in testing, health-seeking behavior, reporting rules, technology use, and service availability. A model may interpret these operational changes as evidence of a new health threat.

Concept drift occurs when relationships learned from historical data change. Seasonal patterns, pathogen characteristics, population immunity, communication behavior, and clinical practice may evolve. Models require continuous monitoring and recalibration.

Missing data should be represented explicitly. Absence of reporting is not evidence of absence of risk, particularly in regions with limited laboratory or digital infrastructure.

6.2 False Alerts and Unequal Detection

High sensitivity can produce frequent false positives. Unnecessary alerts may divert laboratory capacity, exhaust surveillance teams, and weaken trust. Evaluation must report sensitivity together with specificity, positive predictive value, alert frequency, and investigation burden.

Detection performance may vary geographically. Digitally connected populations generate more data and may therefore appear more visible to the system. Communities with limited access to health care, laboratories, internet services, or wearable technology may remain systematically underdetected.

Resource allocation based solely on data volume can intensify existing inequality. Models must incorporate measures of surveillance coverage and should not equate limited data with low risk.

6.3 Privacy, Cybersecurity, and Public Trust

Population surveillance may include highly sensitive health, mobility, behavioral, and biological information. Even deidentified datasets can present reidentification risks when linked across multiple sources.

Data minimization should guide collection. Public-health agencies should use only the information necessary for a legitimate surveillance purpose, restrict secondary use, apply retention limits, and implement strong access controls.

Trust depends on transparency about what information is collected, why it is used, how decisions are reviewed, and when surveillance will end. Emergency conditions should not automatically justify permanent expansion of monitoring.

7. Discussion

7.1 From Prediction to Actionable Intelligence

The value of an early-warning system does not lie solely in predicting an event. Its output must be understandable and linked to a feasible response. An alert that cannot be investigated because of limited laboratory capacity may have little immediate value.

Evaluation should therefore consider warning time, geographical precision, stability, actionability, and response burden. A modest improvement in prediction may be valuable if it provides enough time to arrange testing or prepare health facilities.

Conversely, a statistically accurate forecast may have limited practical value when it is too broad, arrives after conventional detection, or lacks information needed for intervention.

7.2 Human–AI Collaboration

Epidemiologists, laboratory specialists, clinicians, environmental scientists, and local health workers contribute knowledge that an algorithm cannot fully reproduce. They can identify reporting artifacts, interpret unusual clusters, and determine whether a signal is biologically plausible.

AI can support these teams by processing large datasets, identifying deviations, and ranking signals for review. Human investigators should retain authority over confirmation, communication, and intervention.

Local knowledge is particularly important. Community practitioners may recognize festivals, migrations, environmental events, or health-service changes that explain an apparent anomaly.

7.3 Responsible Governance

AI surveillance should operate through predefined governance structures. These should specify data access, validation, alert review, escalation, communication, and accountability.

Independent audit is necessary to assess performance across regions and demographic groups. Public reporting should distinguish simulations, retrospective models, prospective evaluations, and operational deployments.

WHO’s ethical principles support a governance model centred on autonomy, well-being, transparency, accountability, inclusiveness, and sustainability. These principles should guide both technical design and institutional use.

8. Conclusion

AI-enabled surveillance can integrate clinical, population, environmental, biological, and genomic information to support earlier recognition of emerging health risks. Multimodal analysis can reveal relationships and patterns that may remain hidden when information is stored in isolated databases. By combining diverse data sources, surveillance systems may improve situational awareness and help identify unusual changes across populations or environments. However, effective integration depends on data quality, interoperability, standardized formats, appropriate analytical methods, and continuous validation to ensure that detected signals remain meaningful and actionable.

The simulation presented in this paper illustrates a potential improvement in early-warning sensitivity when complementary data layers are progressively incorporated. However, the simulated findings should not be interpreted as evidence of clinical or operational effectiveness. Real-world performance requires prospective evaluation using predefined outcomes, transparent performance criteria, and independent comparisons with established surveillance systems. Such evaluation should also examine sensitivity, specificity, timeliness, false-alert rates, and practical usefulness. Only rigorous validation can determine whether AI provides meaningful improvements over conventional surveillance approaches.

Responsible implementation of AI-enabled surveillance requires careful attention to false alerts, unequal data coverage, privacy, cybersecurity, interpretability, and institutional capacity. Excessive or inaccurate alerts may increase workload and reduce trust, while incomplete datasets may cause important populations or communities to remain underrepresented. Strong governance should therefore establish clear standards for data protection, model monitoring, accountability, and human oversight. Collecting larger quantities of information alone does not guarantee better surveillance. Data must be reliable, relevant, timely, representative, and directly connected to an accountable public-health response.

AI should strengthen rather than displace public-health expertise and essential surveillance infrastructure. Its most defensible role is to identify potentially important signals that can be examined through structured investigation and professional judgment. Laboratory confirmation, epidemiological reasoning, local knowledge, community participation, and proportionate public-health action must remain central to decision-making. Human experts should evaluate AI-generated alerts within their broader contextual understanding and determine appropriate responses. This approach can support innovation while maintaining scientific accountability, institutional responsibility, transparency, and public trust throughout the surveillance process.

References

1. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., ... Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine, 1*, Article 18. https://doi.org/10.1038/s41746-018-0029-1
2. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*, 115–118. https://doi.org/10.1038/nature21056
3. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics, 19*(6), 1236–1246. https://doi.org/10.1093/bib/bbx044
4. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA, 319*(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391
5. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7

6. Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology, 69*(21), 2657–2664. https://doi.org/10.1016/j.jacc.2017.03.571
7. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. *IEEE Journal of Biomedical and Health Informatics, 22*(5), 1589–1604. https://doi.org/10.1109/JBHI.2017.2767063
8. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine, 25*, 24–29. https://doi.org/10.1038/s41591-018-0316-z
9. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs, 33*(7), 1123–1131. https://doi.org/10.1377/hlthaff.2014.0041
10. Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. *Journal of the American Medical Informatics Association, 24*(1), 198–208. https://doi.org/10.1093/jamia/ocw042
11. Ruiz, V. M., Goldsmith, M. P., Shi, L., Simpao, A. F., Gálvez, J. A., Naim, M. Y., Nadkarni, V., Gaynor, J. W., & Tsui, F. R. (2022). Early prediction of clinical deterioration using data-driven machine-learning modeling of electronic health records. *The Journal of Thoracic and Cardiovascular Surgery, 164*(1), 211–222.e3. https://doi.org/10.1016/j.jtcvs.2021.10.060
12. Garriga, R., Mas, J., Abraha, S., Nolan, J., Harrison, O., Tadros, G., & Matic, A. (2022). Machine learning model to predict mental health crises from electronic health records. *Nature Medicine, 28*, 1240–1248. https://doi.org/10.1038/s41591-022-01811-5
13. Sadasivuni, S., Saha, M., Bhatia, N., Banerjee, I., & Sanyal, A. (2022). Fusion of fully integrated analog machine learning classifier with electronic medical records for real-time prediction of sepsis onset. *Scientific Reports, 12*, Article 5711. https://doi.org/10.1038/s41598-022-09712-w
14. Arya, S. S., Dias, S. B., Jelinek, H. F., Hadjileontiadis, L. J., & Pappa, A.-M. (2023). The convergence of traditional and digital biomarkers through AI-assisted biosensing: A new era in translational diagnostics? *Biosensors and Bioelectronics, 235*, 115387. https://doi.org/10.1016/j.bios.2023.115387
15. Chen, S., Yu, J., Chamouni, S., Wang, Y., & Li, Y. (2024). Integrating machine learning and artificial intelligence in life-course epidemiology: Pathways to innovative public health solutions. *BMC Medicine, 22*, Article 354. https://doi.org/10.1186/s12916-024-03566-x
16. Lee, Y. J., Park, C., Kim, H., Cho, S. J., & Yeo, W.-H. (2024). Artificial intelligence on biomedical signals: Technologies, applications, and future directions. *Med-X, 2*, Article 25. https://doi.org/10.1007/s44258-024-00043-1
17. Ford, E., Milne, R., & Curlewis, K. (2024). Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia. *WIREs Data Mining and Knowledge Discovery, 14*(2), e1492. https://doi.org/10.1002/widm.1492
18. Garriga, R., et al. (2024). AI-based epidemic and pandemic early warning systems: A systematic scoping review. *Health Informatics Journal*. https://doi.org/10.1177/14604582241275844

19. Hurst, W., Tekinerdogan, B., Alskaif, T., Boddy, A., & Shone, N. (2022). Securing electronic health records against insider threats: A supervised machine learning approach. *Smart Health, 26*, 100354. https://doi.org/10.1016/j.smhl.2022.100354
20. Alotaibi, E., & Nassif, N. (2024). Artificial intelligence in environmental monitoring: In-depth analysis. *Discover Artificial Intelligence, 4*, Article 84. https://doi.org/10.1007/s44163-024-00198-1