

<https://doi.org/10.1038/s44259-026-00192-w>

Artificial intelligence for early detection and risk prediction of antimicrobial resistance in aquatic ecosystems



William Calero-Cáceres^{1,2} ✉, Ronan Adler Tavella^{3,4}, Fábio Parra Sellera^{5,6}, Jose Luis Balcazar⁷, Jesus Rodriguez-Manzano^{8,9,10}, Rodrigo Cayô^{4,11}, Nilton Lincopan^{4,12,13}, Ana Cristina Gales^{4,14}, Sergio Schenkman^{4,14}, Zhi Ruan^{15,16} ✉, Eliana Guedes Stehling¹⁷ & João Pedro Rueda Furlan^{4,14,18} ✉

Aquatic environments are key reservoirs and dissemination pathways of antimicrobial resistance (AMR). However, current water-based surveillance remains fragmented and inefficient for the timely detection of emerging threats. Integrating artificial intelligence with embedded metadata provides a powerful pathway to identify novel antimicrobial resistance genes, characterize resistome profiles, and predict AMR dynamics in real-time by combining omics, environmental, and hydrological data into spatiotemporal predictive models. Successful implementation of this framework will require robust governance, ethical safeguards, and capacity building to support predictive AMR monitoring aligned with the One Health approach.

The burden of antimicrobial resistance in aquatic environments

Antimicrobial resistance (AMR) is a major global health challenge associated with millions of deaths and is expected to intensify in the coming decades¹. Growing evidence indicates that aquatic environments are key reservoirs and dissemination routes for antimicrobial-resistant bacteria (ARB) and antimicrobial resistance genes (ARGs)^{2,3}. Wastewater, surface waters, and drinking water systems continuously receive microbial and chemical inputs from diverse human activities, including domestic, healthcare, agriculture, livestock, and industrial sources^{4,5}. These systems function as ecological mixing zones where diverse microbial communities interact with pollutants, nutrients, residual antimicrobials, and mobile genetic elements (MGEs)⁶. This convergence facilitates the persistence, exchange, and spread of AMR determinants across human, animal, plant, and environmental interfaces. In addition to these primary anthropogenic pressures, natural background drivers, such as microbial ecological dynamics, intrinsic gene exchange processes, and environmental stressors, also shape the evolution and dissemination of AMR. In water environments, the synergy between these natural and anthropogenic drivers creates particularly favorable conditions for the emergence and amplification of AMR.

Wastewater treatment plants (WWTPs) highlight the central role of aquatic systems in AMR dissemination. They receive substantial inputs of diverse microbial communities from anthropogenic activities and serve as hotspots of ARB, ARGs, and MGEs. However, even advanced treatment processes do not fully remove clinically relevant AMR determinants⁷.

Studies increasingly highlight the need for environmental surveillance to assess both ARGs abundance and genetic context, their association with MGEs that enable horizontal gene transfer (HGT)⁸. Once discharged into rivers, lakes, irrigation canals, or groundwater, these genetic determinants can disperse through hydrological networks and re-enter human and animal exposure pathways⁹. Drinking water distribution systems may also act as secondary reservoirs due to the persistence of ARB within biofilms¹⁰.

Multiple reports have described the occurrence of clinically important antimicrobial-resistant *Enterobacteriales* in aquatic systems, including species carrying genes encoding for extended-spectrum β -lactamases (ESBLs), carbapenemases, or mobile colistin resistance genes^{10,11}. HGT is a key driver of plasmids, integrative conjugative elements, transposons, and integrons, promoting the rapid exchange of ARGs among diverse microbial taxa. Because many plasmids have broad host ranges, ARGs can traverse phylogenetically distant bacterial groups and eventually reach clinically relevant pathogens^{8,12}. Synergistic interactions among AMR determinants can further strengthen phenotypes, illustrating why aquatic environments, with their high diversity, dense microbial interactions, and strong selective pressures, are fertile grounds for the emergence and evolution of new resistance traits¹³.

Environments heavily influenced by human activity, such as hospital effluents, domestic sewage, food-processing facilities, and pharmaceutical discharges, serve as hotspots of microbial turnover and gene exchange^{14,15}. These settings introduce dense mixtures of pollutants, nutrients, antimicrobials, and diverse microbial communities, accelerating opportunities for HGT and promoting the emergence of novel ARG variants. Some of

A full list of affiliations appears at the end of the paper. ✉e-mail: wr.calero@uta.edu.ec; r_z@zju.edu.cn; joao.furlan@academico.ufpb.br

these locations have been proposed as potential environmental evolution hubs where AMR traits may originate before spreading into clinical settings.

Human and animal exposure to ARB and ARGs from water may occur through recreational activities, consumption of contaminated produce, irrigation of crops, aerosolization near WWTPs, and contact with drinking water systems¹⁶. Because environmental AMR rarely causes immediate disease, it serves as a hidden reservoir of resistance that can translate into clinical impacts when antimicrobial-resistant infections emerge in humans or animals. Consequently, effective surveillance must integrate measurements of ARGs mobility, host associations, and exposure pathways in addition to abundance¹⁷. Accordingly, understanding this burden underscores the need for enhanced and predictive environmental AMR monitoring capable of capturing the true complexity of aquatic systems. Frameworks such as WHO GLASS provide a foundation for developing harmonized approaches to incorporate environmental AMR into global surveillance¹⁸.

Limitations of current antimicrobial resistance surveillance in water

Environmental AMR surveillance in aquatic systems remains limited by substantial technical, methodological, and conceptual challenges. A central limitation is the low sampling typically employed in environmental monitoring programs, resulting in datasets with poor temporal resolution and an incomplete understanding of short-term fluctuations¹⁹. Aquatic environments are highly dynamic, influenced by rainfall, seasonal patterns, hydraulic conditions, industrial discharge pulses, and human activity cycles. As a result, infrequent grab sampling cannot capture the true variability of ARB, ARGs, and MGEs across time^{20,21}. This stands in contrast to clinical surveillance systems, where high-frequency data allow early recognition of emerging threats.

Methodological fragmentation further restricts the ability to generate coherent and comparable datasets. Current environmental AMR monitoring relies on culture-based approaches, targeted qPCR assays, and untargeted metagenomic sequencing, each providing only a partial representation of AMR dynamics²². Culture-based methods detect only a fraction of the environmental microbiome; qPCR is highly sensitive but limited to predefined targets; and metagenomics offers broad detection but often lacks sensitivity for low-abundance ARGs^{17,23,24}. The absence of frameworks that integrate these methods means that surveillance remains largely descriptive, with little capacity to predict trends or identify causative relationships. Moreover, delays between sample collection, laboratory analysis, and reporting reduce the utility of surveillance data for timely risk assessment²⁵, where risk refers to upstream environmental signals such as increases in ARG abundance, enhanced association with MGEs, elevated potential for HGT, and greater likelihood of human or animal exposure, rather than direct prediction of infections or clinical outcomes^{8,26}.

Environmental complexity adds further constraints. Aquatic systems encompass heterogeneous and interconnected matrices, including wastewater, surface water, sediments, biofilms, sludge, and irrigation canals, each with unique physicochemical conditions, microbial communities, and selective pressures²⁷. AMR determinants can occur in dissolved DNA, intracellular DNA, or within microbial aggregates, and may shift between these states as they move through the environment. These determinants are transported through hydrological networks, settle into sediments, colonize biofilms, or re-enter human and animal pathways¹⁹. This complexity allows the environment to function not only as a reservoir, but also as a potential biological “reactor” where novel AMR mechanisms may emerge, recombine, and evolve through recombination events²⁸.

Despite extensive evidence showing that anthropogenically impacted ecosystems harbor increased levels of clinically relevant ARGs and ARB, much of the available data remains largely descriptive. Differences in sampling strategies, laboratory workflows, detection thresholds, and reporting formats complicate direct comparisons across studies²⁹. There is no consensus on which indicators best reflect environmental AMR risk, nor agreement on how ARG abundance should be interpreted in terms of

human or ecological health risks, given that environmental AMR risk is fundamentally hazard- and exposure-driven and relies on proxy indicators rather than outcome-based metrics such as infection incidence or disease burden^{8,30}. Consequently, environmental AMR monitoring often generates more unanswered questions than actionable conclusions.

Given these limitations, the integration of artificial intelligence (AI) represents a critical step forward. AI can synthesize fragmented datasets, identify hidden associations, and uncover temporal or spatial trends that traditional methods may fail to detect. By enabling predictive modeling and early recognition of emerging AMR patterns, this approach could transform environmental AMR surveillance from a retrospective and descriptive activity into a proactive system capable of informing timely public health interventions.

Artificial intelligence in environmental antimicrobial resistance surveillance

AI is increasingly reshaping how AMR is monitored in aquatic environments, driven by the recent global expansion of wastewater sequencing programs and advances in deep learning architectures. At its core, AI encompasses a spectrum of computational approaches designed to detect patterns, learn from data, and support decision-making^{31,32}. Among these approaches, machine learning (ML) and its subset, deep learning (DL), have become particularly impactful for environmental genomics because of their ability to extract meaningful structure from high-dimensional, noisy, and heterogeneous datasets³³. ML methods learn statistical associations between features/inputs (e.g., metagenomic features and environmental metadata) and outcomes/outputs (e.g., AMR levels), while DL models use multilayer neural networks to derive high-dimensional representations directly from raw sequences. These methods are complementary, as ML excels at integrating diverse data sources for predictive modeling, whereas DL is particularly effective at identifying novel genetic elements and resolving complex resistome patterns^{34,35}. More recently, large language models (LLMs), an emerging class of DL systems, have expanded the role of AI by assisting with the interpretation and summarization of complex analytical outputs. Together, these tools form an AI toolkit well-suited to the scale and complexity of aquatic AMR surveillance.

ML has already been applied across a broad range of aquatic AMR surveillance tasks and has shown promise in extracting AMR-related patterns that traditional methods often miss. Early studies used ML to predict the persistence of ARB in water systems and to identify dominant sources of ARGs across catchments³⁶. As sequencing efforts scaled up internationally, global wastewater surveillance initiatives began combining resistome profiles with socio-economic and environmental variables to predict ARG abundances in regions without direct measurements, demonstrating how AI can integrate multifactorial drivers into risk assessments³⁷. A recent systematic review of ML in water and wastewater surveillance highlighted this variety of applications, ranging from source attribution and fate modeling to predictions of treatment performance, but also noted that most studies still rely on relatively simple models and descriptive analyses³⁶. This gap highlights the growing interest in more advanced AI approaches, particularly DL, which is now being explored for classifying and predicting resistance profiles in complex environmental samples with greater accuracy and scope³⁸. Notably, these AI methods can integrate high-dimensional data and potentially overcome limitations of traditional culture or PCR-based surveillance, providing a more comprehensive view of environmental resistomes.

One major contribution of AI is improving the identification of ARGs directly from sequencing data. Conventional bioinformatics pipelines rely on aligning sequences to known ARGs, which misses novel or highly diverged ARGs^{35,39}. In contrast, DL models have demonstrated the ability to detect ARGs in metagenomic reads with higher sensitivity³⁵. Modern AI approaches can classify AMR determinants from short reads and contigs, even when homology to reference databases is minimal, and can recognize relationships across antimicrobial classes and mechanisms that traditional pipelines overlook^{39–41}. Newer methods that incorporate pretrained protein

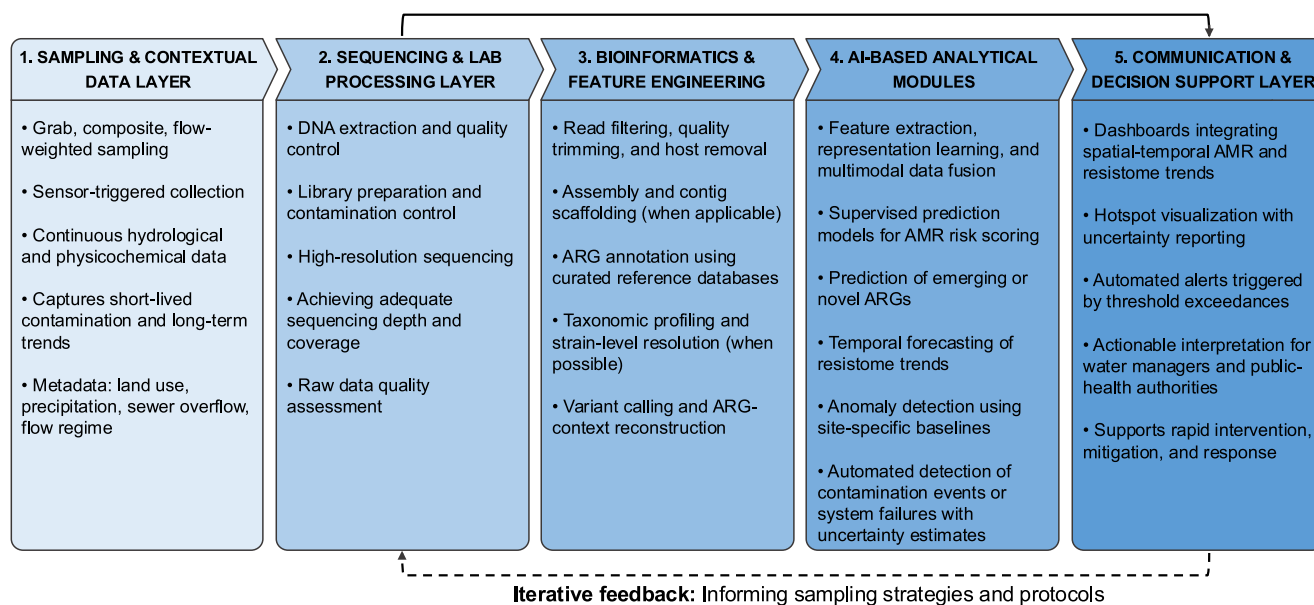


Fig. 1 | Conceptual and integrated AI-driven framework for early AMR detection and risk prediction that combines environmental sampling, laboratory processing, bioinformatics, and AI-based modeling. A feedback loop links model outputs

back to sampling design, enabling iterative refinement of protocols and improving system responsiveness.

representations further enhance detection by capturing higher-order sequence features, enabling the identification of ARGs that would otherwise remain cryptic⁴². Furthermore, ensemble strategies that integrate protein features, genomic context, and evolutionary signals extend this capability by inferring entirely novel ARGs with no detectable similarity to known ones⁴³. The proliferation of such AI tools for ARG discovery demonstrates their promise in extracting richer information from environmental metagenomes than was previously possible with rule-based methods.

AI is not only enhancing gene-level analysis but also complementing existing statistical, epidemiological, environmental, and hydrological approaches by enabling predictive modeling of AMR dynamics in aquatic systems. Data-driven models can learn from historical surveillance data and a multitude of contextual factors to predict where and when AMR hotspots might emerge²⁸. Models that integrate resistome profiles with easily measured environmental parameters have demonstrated the potential to rapidly approximate ARG levels and to reduce the time required for analysis, suggesting that near-real-time surveillance is increasingly feasible³⁸. Importantly, these approaches can also reveal which environmental factors most strongly shape AMR patterns in different regions and therefore provide crucial information on the main drivers of hotspot formation. Building on this progress, researchers are combining disparate datasets such as water quality parameters, antimicrobial consumption/residues, climate and meteorological variables, hydrological data, and land use information into unified predictive frameworks⁴⁴. By training on these multi-layer inputs, AI systems can identify spatiotemporal trends on AMR, detect high-risk locations, and anticipate how resistomes may shift in response to seasonal cycles, human activity, or management interventions. This emerging predictive capacity moves water AMR surveillance beyond static description and toward proactive risk mitigation.

Another emerging application is the use of AI for anomaly detection and early warning in water-based AMR surveillance. Because surveillance data are often collected continuously (for instance, via routine wastewater monitoring), ML algorithms can be deployed to automatically flag unusual patterns that may signal emerging threats (e.g., critical priority clones endemic in some endemic regions emerging in a country)⁴⁵. Unsupervised learning or outlier detection methods can establish baseline expectations for AMR abundances and then detect statistically significant deviations. A

sudden spike in the concentration of a critical ARG in a wastewater treatment plant's effluent could thus trigger an alert for further investigation. In effect, AI can act as an efficient sentinel in the environment, continuously scanning for aberrations in AMR trends that merit rapid public health attention⁴⁶.

Additionally, the maturation of LLMs presents new opportunities to streamline AMR monitoring and decision-making⁴⁷. Although these models do not directly analyze sequencing data, they excel at synthesizing information and generating human-readable output, capabilities that can be harnessed in environmental surveillance workflows. Recent studies have shown that LLM-based systems can extract AMR information from free-text laboratory reports and rapidly organize it for epidemiological use, greatly accelerating the aggregation and interpretation of important findings across sites⁴⁸. LLMs can also assist by generating concise surveillance summaries, explaining complex analytical results to decision-makers, or helping researchers interpret emerging AMR signals^{47,49}. In addition, domain-adapted biomedical models may be used to contextualize novel ARGs or trends by drawing on scientific literature and genomic databases. Although these applications are in their infancy, initial demonstrations indicate that LLMs can assist in interpretation⁴⁸. Integrating their contextual reasoning with the predictive capabilities of ML models may ultimately enhance decision support, enabling systems that not only detect changes in AMR patterns but also interpret their significance and suggest appropriate responses in near real time.

In summary, AI offers a powerful set of approaches that can transform environmental AMR surveillance from a reactive, data-sparse endeavor to a proactive, data-driven system. By applying AI systems, AMR profiles in complex water microbiomes can be classified with higher precision, ARGs can be detected with greater sensitivity, and high-risk hotspots can be forecast before they escalate. These advances improve both the speed and the resolution of surveillance, enabling earlier warnings of emerging threats. The examples above illustrate how integrating diverse data sources through AI can yield actionable ideas that were previously unattainable^{32,49}. Realizing this potential will require interdisciplinary collaboration, high-quality data inputs, and careful validation of models. However, without sufficiently large and evenly distributed environmental datasets, AI systems risk overfitting to local sampling conditions and reinforcing existing biases, undermining their predictive reliability. Even so, the trajectory is clear: intelligent algorithms

can help identify worrisome resistance trends in water before they spread, guiding interventions to protect public health. Harnessing AI in environmental AMR monitoring is an essential step toward more predictive and preventive management of AMR in the One Health context.

A conceptual AI-driven framework for early detection and risk prediction

Developing an AI-enabled framework for AMR surveillance in aquatic environments requires the coordinated integration of sampling, sequencing, bioinformatics, and analytical layers into a single workflow that transforms raw environmental data into actionable public health intelligence^{32,50,51}. The conceptual structure proposed here illustrates how these components interact to support both early detection of emerging AMR signals and forward-looking risk prediction in water systems, representing one of the first attempts to formalize a truly predictive, end-to-end framework for aquatic AMR surveillance (Fig. 1). Importantly, this framework is designed to support tiered and adaptive sampling strategies, prioritizing longitudinal surveillance at sentinel sites and targeted sampling intensification during periods or locations of elevated risk, rather than relying on uniformly high-frequency sampling across all settings.

The workflow begins with systematic and, when possible, automated water sampling. Grab samples are valuable for detecting short-lived contamination events, whereas composite or flow-weighted sampling provides a more representative view of daily or weekly resistome fluctuations. Increasingly, sensor-augmented samplers allow high-frequency collection triggered by water level, flow, or precipitation events, ensuring that critical hydrological dynamics are captured. These systems often operate alongside in situ sensors that record temperature, turbidity, conductivity, dissolved oxygen, and flow rate, enabling continuous contextualization of the microbiological data⁵².

The second layer consists of laboratory processing and sequencing. Shotgun metagenomics remains the most comprehensive strategy for resistome monitoring, offering unbiased detection of known and candidate ARGs across a broad range of taxa. Hybrid and long-read sequencing resolves gene-host associations, plasmid structures, and MGEs, providing insight into the potential for dissemination⁵³. Targeted approaches, such as multiplexed amplicon sequencing, can complement these methods when surveillance programs require high throughput or prioritization of specific resistance markers⁵².

Downstream, bioinformatic workflows convert sequencing data into structured features compatible with AI analysis. These steps include read filtering, removal of host contamination, assembly where appropriate, and annotation against curated ARG and MGEs databases^{52,53}. This process generates a multidimensional feature space capturing ARG abundance, diversity, mobility indicators, taxonomic context, and environmental metadata. The performance and reliability of this framework can be evaluated through retrospective analyses of historical datasets, prospective pilot deployment at sentinel sites, and perturbation-based validation using natural or operational disturbances, allowing assessment of predictive accuracy, robustness, and real-world utility.

AI-based analytical modules constitute the core of the framework, complementing established surveillance and analytical practices. Feature extraction components integrate resistome and environmental characteristics, providing the inputs for supervised and unsupervised modeling. Risk assessment models estimate the likelihood of elevated AMR levels or the emergence of clinically relevant resistance mechanisms at a given location. Temporal forecasting algorithms predict future trends in ARG abundance, enabling authorities to anticipate periods of increased risk. In parallel, anomaly-detection systems establish baselines for each monitoring site and flag deviations that may indicate contamination events, infrastructure failures, or localized outbreaks. Together, these analytical layers convert complex, noisy data into interpretable signals.

The final segment of the framework emphasizes communication and decision support. Interactive dashboards visualize AMR trends across space and time, highlight hotspots, provide uncertainty estimates, and issue

automated alerts when predefined thresholds are exceeded. These interfaces facilitate rapid interpretation by public health authorities, water managers, and policymakers, supporting timely interventions^{32,46,47}. This workflow is designed to operate iteratively with model outputs informing adjustments in sampling strategies, laboratory protocols, or analytical thresholds, progressively strengthening system performance.

It should be acknowledged that this framework is primarily conceptualized for settings with established laboratory infrastructure, sustained funding, and access to advanced sequencing technologies, which are more common in high-income countries. In low- and middle-income countries, implementation may be constrained by limited resources, fragmented surveillance networks, and reduced analytical capacity. Nevertheless, partial adoption of individual components, such as targeted sequencing, periodic sampling, or simplified metadata collection, can still generate valuable evidence on AMR in aquatic environments. Even fragmented datasets contribute to baseline knowledge, situational awareness, and progressive capacity building, reinforcing that limited data are preferable to the absence of surveillance altogether.

Data requirements, standardization, and technical challenges

AI-enabled AMR surveillance in aquatic environments demands coherent, high-quality data ecosystems that capture microbial, environmental, and ecological complexity. High-resolution sequencing data must be generated using robust laboratory protocols to ensure accurate detection of both dominant ARGs and low-abundance or emerging AMR determinants. Inadequate sequencing depth and coverage, poor read quality, batch effects, or contamination can distort microbial community structure and impair model performance, particularly when AI is used to identify early signals of change. AI cannot compensate for inadequate sampling design, insufficient spatial or temporal coverage, or fundamental inconsistencies in laboratory and analytical methods, its performance is inherently constrained by the quality and representativeness of the underlying data.

The predictive power of machine learning systems depends not only on genomic data but also on standardized, context-rich metadata describing temperature, flow rate, rainfall, pollution pressure, hydrological dynamics, and land-use characteristics. As highlighted in recent AMR surveillance perspectives, metadata harmonization is essential for moving from descriptive environmental analyses to actionable, One Health-aligned early-warning systems^{37,54,55}. Within this framework, environmental AMR risk is interpreted as a composite and trend-based signal rather than as the exhaustive detection of all rare ARGs in individual samples, making shotgun metagenomics, particularly when combined with longitudinal sampling and targeted validation of priority markers, a suitable foundation for risk-oriented surveillance.

Curated reference databases, such as CARD, ResFinder, and MEGARes, remain foundational to ARG detection and annotation^{56–58}. However, inconsistencies in nomenclature, lack of prediction of resistance mutations, gene clustering, update frequency, and annotation criteria can lead to divergent outputs across analytic pipelines. Harmonized ontologies, cross-database compatibility standards, and transparent versioning, combined with benchmarking using environmental mock communities, would substantially improve reproducibility and strengthen the reliability of AI-based inference. Expanding these repositories to include environmental-specific resistomes (i.e., structural AMR determinants associated with environmental bacterial species), together with integration of mobile genetic element databases to better track HGT potential, will further enhance model capacity to capture novel or niche AMR dynamics in water systems.

Data scarcity, imbalance, and integration pose additional challenges. While global sewage metagenomics cohorts are expanding, many regions, particularly in low- and middle-income countries, still lack longitudinal datasets that capture temporal and spatial variability in aquatic resistomes. ARG distributions in natural environments are often heavily skewed, with rare but clinically important determinants posing

difficulties for model training and further reinforcing the predictive challenge. Environmental noise driven by seasonal variation, storm events, industrial discharge, and wastewater fluctuations introduces additional complexity, influencing microbial turnover and gene mobilization. AI models should therefore incorporate uncertainty quantification, domain adaptation techniques, and environmental covariates to avoid overfitting and improve generalizability.

Reproducibility and transparency remain central to credible AI-assisted environmental AMR surveillance. Differences in DNA extraction kits, sequencing approaches, and bioinformatic tools can introduce artefacts that propagate into AI predictions, undermining cross-study comparability. Adoption of FAIR principles (Findable, Accessible, Interoperable, and Reusable) along with open-source pipelines, standardized reporting checklists, and containerized workflows will support robust model validation and international interoperability^{59,60}. As emphasized in recent perspectives on environmental genomics, coordinated global data infrastructures that integrate wastewater, clinical, agricultural, and ecological datasets are necessary to enable predictive, One Health-oriented AMR intelligence networks^{54,61}. Developing such infrastructures will allow AI systems not only to detect AMR patterns earlier but also to generate actionable risk forecasts that support targeted interventions.

Ethical, operational, and regulatory considerations

The deployment of AI-driven AMR surveillance in aquatic environments introduces important ethical, operational, and regulatory considerations that must be addressed to ensure responsible, equitable, and sustainable implementation^{62,63}. First, ethical concerns primarily relate to data governance, privacy, and equity. Although environmental metagenomic datasets generally do not contain identifiable human genomic information, wastewater and aquatic surveillance can inadvertently capture host-associated microbial signals that correlate with community-level health patterns or anthropogenic chemical signatures⁶⁴. Transparent data governance frameworks are therefore needed to define what constitutes sensitive information, determine who controls data access, and establish safeguards that prevent misuse, stigmatization of communities, or unintended inferences about population behavior⁶⁵. Furthermore, because AI models learn from available datasets, there is a risk that long-standing global inequities in environmental monitoring, particularly underrepresentation of low-resource regions, may propagate into AI predictions, thereby reinforcing data-driven biases⁶⁶. Ensuring equitable participation in surveillance networks and fair distribution of benefits is essential for building trust and achieving global AMR preparedness.

Operational challenges also shape the feasibility of AI-enabled aquatic AMR surveillance. Developing high-resolution datasets requires sustained investments in sampling infrastructure, sequencing capacity, bioinformatics expertise, and cloud-based computational resources. Many water utilities and environmental agencies lack in-house capacity to manage large-scale metagenomic datasets, let alone deploy advanced machine-learning workflows. This gap highlights the need for capacity building that includes not only technical training but also the standardization of AI pipelines, ensuring consistent implementation and reducing heterogeneity across monitoring programs⁶⁷. Capacity-building programs, regional sequencing hubs, standardized training modules, and user-friendly analytic platforms are therefore needed to reduce technical barriers. The maintainability of AI systems poses another challenge: models must be retrained and recalibrated as environmental conditions, microbial communities, and AMR profiles evolve. Without continuous data inflow, version control, and performance auditing, prediction accuracy may degrade over time⁶⁸. Integrating AI workflows into existing water-quality monitoring operations require clear standard operating procedures, interoperability with laboratory information systems, and robust mechanisms for cross-sector communication among public health authorities, wastewater managers, and environmental regulators.

Regulatory considerations represent a third pillar for responsible deployment. AMR surveillance in aquatic systems operates across

fragmented regulatory domains encompassing environmental protection, water safety, public health, and data governance⁶⁹. Clear regulatory guidance is necessary to standardize sampling frequencies, sequencing thresholds, metadata reporting requirements, and quality control procedures to ensure that AI-derived outputs are interpretable and comparable across jurisdictions. At the same time, regulatory bodies will need to set evidence standards for when AI-generated alerts can trigger public-health action, and how uncertainty should be communicated to decision-makers. The integration of AI models into environmental policy frameworks also raises the question of accountability, particularly when risk forecasts inform high-stakes decisions such as issuing advisories, modifying wastewater treatment operations, or regulating industrial discharge. Clarifying the legal accountability of model-driven decisions, as well as establishing formal validation and performance requirements, will be essential to ensure transparent, defensible, and trusted use of AI in environmental governance⁷⁰. Developing transparent audit trails, algorithmic explainability tools, and external validation procedures will be critical for ensuring regulatory confidence in AI-assisted risk prediction. Finally, international harmonization is needed to align environmental AMR surveillance with global One Health strategies, enabling data sharing, cross-border early warnings, and coordinated responses to emerging ARB threats⁶⁶.

Conclusions

Aquatic environments play a central, often underrecognized role in the emergence, evolution, and dissemination of AMR. However, current water-based surveillance remains fragmented, temporally sparse, and methodologically inconsistent, limiting our ability to detect emerging threats and translate environmental findings into actionable public health insights. AI offers a transformative opportunity to overcome these limitations. By integrating high-dimensional sequencing data with environmental, hydrological, and socioecological metadata, AI systems can identify novel ARGs, characterize complex resistome profiles, predict AMR trends, and detect anomalies far earlier than conventional surveillance approaches. The proposed AI-driven framework illustrates a pathway toward real-time environmental intelligence capable of informing targeted, evidence-based interventions. Ethical, operational, and regulatory safeguards must evolve in parallel to ensure the responsible deployment of AI in aquatic AMR surveillance. Investments in capacity building, governance structures, and cross-sector coordination will be essential for embedding these technologies into routine monitoring systems. Altogether, advancing AI-enabled AMR surveillance in aquatic systems represents not merely a technical improvement but a critical step toward predictive, preventive, and truly One Health-aligned management of AMR. By bridging environmental monitoring with data-driven risk assessment, AI-based frameworks offer a pathway to transform complex environmental patterns into interpretable metrics capable of informing public health decision-making. Ultimately, integrating AI into aquatic AMR surveillance has the potential to redefine how environmental data are converted into timely, actionable insights that support coordinated responses across environmental, clinical, and policy sectors.

Data availability

No datasets were generated or analysed during the current study.

Received: 16 December 2025; Accepted: 17 February 2026;

Published online: 26 March 2026

References

1. Naghavi, M. et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet* **404**, 1199–1226 (2024).
2. Grenni, P. Antimicrobial resistance in rivers: a review of the genes detected and new challenges. *Environ. Toxicol. Chem.* **41**, 687–714 (2022).

3. European Environment Agency. Antimicrobial resistance in European surface waters: a developing area. European Environment Agency (EEA). <https://www.eea.europa.eu/en/analysis/publications/antimicrobial-resistance-in-european-surface-waters-a-developing-area> (2023).
4. Sivalingam, P. et al. Anthropogenic pollution may enhance natural transformation in water, favouring the spread of antibiotic resistance genes. *J. Hazard. Mater.* **475**, 134885 (2024).
5. Cocker, D. et al. Healthcare as a driver, reservoir and amplifier of antimicrobial resistance: opportunities for interventions. *Nat. Rev. Microbiol.* **22**, 636–649 (2024).
6. Zhang, L. et al. Antibiotic resistance genes and mobile genetic elements in different rivers: The link with antibiotics, microbial communities, and human activities. *Sci. Total Environ.* **919**, 170788 (2024).
7. Pazda, M. et al. Antibiotic resistance genes identified in wastewater treatment plant systems – A review. *Sci. Total Environ.* **697**, 134023 (2019).
8. Klümper, U. et al. Towards the integration of antibiotic resistance gene mobility into environmental surveillance and risk assessment. *npj Antimicrob. Resist.* **3**, 81 (2025).
9. Pinamonti, D. et al. Water-mediated dissemination and detection of antibiotic resistance across livestock, agri-food, and aquaculture systems. *Micromachines* **16**, 934 (2025).
10. Calero-Cáceres, W. et al. Genomic insights of mcr-1 harboring *Escherichia coli* by geographical region and a One-Health perspective. *Front. Microbiol.* **13**, 1032753 (2023).
11. Furlan, J. P. R. et al. Emergence of carbapenem-resistant *Klebsiella pneumoniae* species complex from agrifood systems: detection of ST6326 co-producing KPC-2 and NDM-1. *J. Sci. Food Agric.* **104**, 7347–7354 (2024).
12. Tao, S. et al. The spread of antibiotic resistance genes: in vivo model. *Can. J. Infect. Dis. Med. Microbiol.* **2022**, 3348695 (2022).
13. Singer, A. C. et al. Review of antimicrobial resistance in the environment and its relevance to environmental regulators. *Front. Microbiol.* **7**, 1728 (2016).
14. Kunhikannan, S. et al. Environmental hotspots for antibiotic resistance genes. *MicrobiologyOpen* **10**, e1197 (2021).
15. Salazar, C. et al. Human microbiota drives hospital-associated antimicrobial resistance dissemination in the urban environment and mirrors patient case rates. *Microbiome* **10**, 208 (2022).
16. Singh, A. K., Kaur, R., Verma, S. & Singh, S. Antimicrobials and antibiotic resistance genes in water bodies: pollution, risk, and control. *Front. Environ. Sci.* **10**, 830861 (2022).
17. Bengtsson-Palme, J. et al. Towards monitoring of antimicrobial resistance in the environment: for what reasons, how to implement it, and what are the data needs? *Environ. Int.* **178**, 108089 (2023).
18. World Health Organization. Global Antimicrobial Resistance and Use Surveillance System (GLASS). World Health Organization. Available at: <https://www.who.int/initiatives/glass> (2024).
19. Beltrán de Heredia, I., Alkorta, I., Garbisu, C. & Ruiz-Romera, E. A practical framework for environmental antibiotic resistance monitoring in freshwater ecosystems. *Antibiotics* **14**, 840 (2025).
20. Barrantes-Jiménez, K. et al. Influence of seasonality and pollution on the presence of antibiotic resistance genes and potentially pathogenic bacteria in a tropical urban river. *Antibiotics* **14**, 798 (2025).
21. Calero-Cáceres, W., Méndez, J., Martín-Díaz, J. & Muniesa, M. The occurrence of antibiotic resistance genes in a Mediterranean river and their persistence in the riverbed sediment. *Environ. Pollut.* **229**, 927–936 (2017).
22. Sousa-Carmo, R. R. & Lincopan, N. Importance of wastewater-based and environmental surveillance for monitoring resistance to next-generation antibiotics. *Future Microbiol.* **20**, 849–854 (2025).
23. Taylor, W., Bohm, K., Dyet, K., Weaver, L. & Pattis, I. Comparative analysis of qPCR and metagenomics for detecting antimicrobial resistance in wastewater: a case study. *BMC Res. Notes* **18**, 5 (2025).
24. Vitorino, L. C. & Bessa, L. A. Microbial Diversity: the Gap between the estimated and the known. *Diversity* **10**, 46 (2018).
25. Cajas-Corrales, N., Vallejo-Espin, D. & Calero-Cáceres, W. A case for global standardisation of genomics and wastewater-based epidemiology. *Lancet Microbe* **6**, 101092 (2025).
26. Hamilton, K. A. et al. Next Generation Quantitative Microbial Risk Assessment (QMRA): bigger, better, faster. *Environ. Sci. Technol. Lett.* **12**, 1471–1480 (2025).
27. Larsson, D. G. J. & Flach, C. F. Antibiotic resistance in the environment. *Nat. Rev. Microbiol.* **20**, 257–269 (2021).
28. Zambrano, M. M. Interplay between antimicrobial resistance and global environmental change. *Annu. Rev. Genet.* **57**, 275–296 (2023).
29. Calero-Cáceres, W., de Almeida Kumlien, A. C. M. & Balcázar, J. L. Advances and challenges in assessing antimicrobial resistance in environmental settings. *Curr. Opin. Environ. Sci. Health* **41**, 100571 (2024).
30. Collineau, L. et al. Integrating whole-genome sequencing data into quantitative risk assessment of foodborne antimicrobial resistance: a review of opportunities and challenges. *Front. Microbiol.* **10**, 1107 (2019).
31. Lv, J. et al. A review of artificial intelligence applications for antimicrobial resistance. *Biosaf. Health* **3**, 22–31 (2021).
32. Arnold, A. et al. How AI can help us beat AMR. *npj Antimicrob. Resist.* **3**, 18 (2025).
33. Ali, T. et al. Artificial intelligence for antimicrobial resistance prediction: challenges and opportunities towards practical implementation. *Antibiotics* **12**, 523 (2023).
34. Cesaro, A. et al. Challenges and applications of artificial intelligence in infectious diseases and antimicrobial resistance. *npj antimicrob. Resist.* **3**, 2 (2025).
35. Olatunji, I. et al. Artificial intelligence tools for the identification of antibiotic resistance genes. *Front. Microbiol.* **15**, 1437602 (2024).
36. Foroughi, M. et al. Application of machine learning for antibiotic resistance in water and wastewater: a systematic review. *Chemosphere* **358**, 142223 (2024).
37. Hendriksen, R. S. et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **10**, 1124 (2019).
38. Miglietta, L. et al. Artificial intelligence and infectious disease diagnostics: state of the art and future perspectives. *Lancet Infect. Dis.* **S1473-3099**(25), 00354–8 (2025).
39. Arango-Argoty, G. et al. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 23 (2018).
40. Pei, Y. et al. ARGNet: using deep neural networks for robust identification and classification of antibiotic resistance genes from sequences. *Microbiome* **12**, 84 (2024).
41. Borelli, T. C. & Paschoal, A. R. da Silva RR. DeepSEA: an alignment-free explainable approach to annotate antimicrobial resistance proteins. *BMC Bioinform.* **26**, 224 (2025).
42. Wu, J. et al. PLM-ARG: antibiotic resistance gene identification using a pretrained protein language model. *Bioinformatics* **39**, btad690 (2023).
43. Rannon, E. et al. DRAMMA: a multifaceted machine learning approach for novel antimicrobial resistance gene detection in metagenomic data. *Microbiome* **13**, 67 (2025).
44. Jiang, P. et al. A rapid approach with machine learning for quantifying the relative burden of antimicrobial resistance in natural aquatic environments. *Water Res.* **262**, 122079 (2024).
45. Emon, M. I. et al. CIWARS: a web server for antibiotic resistance surveillance using longitudinal metagenomic data. *J. Mol. Biol.* **437**, 169159 (2025).

46. Kasse, G. E. et al. Leveraging artificial intelligence for one health: opportunities and challenges in tackling antimicrobial resistance—scoping review. *One Health Outlook* **7**, 51 (2025).
47. Maity, S. & Saikia, M. J. Large language models in healthcare and medical applications: a review. *Bioengineering* **12**, 631 (2025).
48. Chou, A. et al. Extracting antibiotic susceptibility from free-text microbiology reports using natural language processing. *Infect. Control Hosp. Epidemiol.* **46**, 1–3 (2025).
49. Chen, C. et al. Using large language models to assist antimicrobial resistance policy development: Integrating the environment into health protection planning. *Environ. Sci. Technol.* **59**, 1243–1252 (2025).
50. Waldo, W. J. et al. Mitigating antimicrobial resistance by innovative solutions in AI (MARISA): a modified James Lind Alliance analysis. *npj Antimicrob. Resist.* **3**, 75 (2025).
51. Liguori, K. et al. Antimicrobial resistance monitoring of water environments: a framework for standardized methods and quality control. *Environ. Sci. Technol.* **56**, 9149–9160 (2022).
52. Ottesen, A. et al. Advancing antimicrobial resistance monitoring in surface waters with metagenomic and quasimetagenomic methods. *PLoS Water* **1**, e0000067 (2022).
53. Davis, B. C. et al. Recommendations for the use of metagenomics for routine monitoring of antibiotic resistance in wastewater and impacted aquatic environments. *Crit. Rev. Env. Sci. Tec.* **53**, 1731–1756 (2023).
54. Larsson, D. G. J. et al. Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. *Environ. Int.* **117**, 132–138 (2018).
55. Singh, S. et al. Advancing AMR surveillance: confluence of one health and big data integration: converging one health and big data for AMR. *Ecohealth* **22**, 403–414 (2025).
56. Alcock, B. P. et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, D517–D525 (2020).
57. Bortolaia, V. et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**, 3491–3500 (2020).
58. Doster, E. et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res.* **48**, D561–D569 (2020).
59. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
60. McDermott, M. B. A. et al. Reproducibility in machine learning for health research: still a ways to go. *Sci. Transl. Med.* **13**, eabb1655 (2021).
61. Balcázar, J. L. Wastewater-based epidemiology as a complementary tool for antimicrobial resistance surveillance: overcoming barriers to integration. *Bioessays* **47**, e70034 (2025).
62. Rodríguez-Manzano, J. et al. Innovative diagnostic technologies: navigating regulatory frameworks through advances, challenges, and future prospects. *Lancet Digit. Health* **6**, e934–e943 (2024).
63. Pennisi, F. et al. The role of artificial intelligence and machine learning models in antimicrobial stewardship in public health: a narrative review. *Antibiotics* **14**, 134 (2025).
64. Port, J. A. et al. Metagenomic frameworks for monitoring antibiotic resistance in aquatic environments. *Environ. Health Perspect.* **122**, 222–8 (2014).
65. Tiffin, N. et al. How to use relevant data for maximal benefit with minimal risk: digital health data governance to protect vulnerable populations in low-income and middle-income countries. *BMJ Glob. Health* **4**, e001395 (2019).
66. Peykani, P. et al. Large language models: a structured taxonomy and review of challenges, limitations, solutions, and future directions. *Appl. Sci.* **15**, 8103 (2025).
67. Van Den Bossche, T. et al. The microbiologist's guide to metaproteomics. *Imeta* **4**, e70031 (2025).
68. Khan, T. et al. TwinEco: a unified framework for dynamic data-driven digital twins in ecology. *Ecol. Inform.* **91**, 103407 (2025).
69. Craddock, H. A. et al. The challenge of reducing antimicrobial resistance (AMR) across the one health landscape: Diverse perspectives on AMR risks and their mitigation in sinks, drains, and wastewater. *Sci. Total Environ.* **992**, 179935 (2025).
70. Zhao, T. et al. Leveraging immersive digital twins and AI-driven decision support systems for sustainable water reserves management: a conceptual framework. *Sustainability* **17**, 3754 (2025).

Acknowledgements

This study was supported by the São Paulo Research Foundation (FAPESP) (grant numbers 21/10599-3 and 23/12947-4), the Ecuadorian Corporation for the Development of Research and Academia (CEDIA) (Project UTA number PEFICIAL15), the National Natural Science Foundation of China (82472335), "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province (2024C03217), and the Key Program of the Zhejiang Medical and Health Science and Technology Project (WKJ-ZJ-2506). R.A.T. is a postdoctoral fellow at the São Paulo Research Foundation (FAPESP) (grant number 24/02579-0). F.P.S., R.C., N.L., A.C.G., S.S., and E.G.S. are research fellows at the National Council for Scientific and Technological Development (CNPq) (305590/2025-1, 307915/2022-0, 314336/2021-4, 305081/2025-0, 303788/2020-8, and 304905/2022-4, respectively). J.P.R.F. thanks the São Paulo Research Foundation (FAPESP) (grant number 23/16216-4) for the postdoctoral fellowship. W.C.C. expresses his gratitude to Johanna Mora for her support and for being a source of inspiration in the development of this perspective.

Author contributions

W.C.C.: Conceptualization, writing—original draft, writing—review & editing, visualization. R.A.T.: writing—original draft, writing—review & editing, visualization. F.P.S.: Writing—review & editing, visualization. J.R.M.: writing—review & editing, visualization. J.L.B.: Writing—review & editing, visualization. R.C.: writing—review & editing, visualization. N.L.: writing—review & editing, visualization. A.C.G.: Writing—review & editing, visualization. S.S.: Writing—review & editing, visualization. Z.R.: Writing—original draft, writing—review & editing, visualization. E.G.S.: Writing—original draft, writing—review & editing, visualization. J.P.R.F.: Conceptualization, project administration, writing—original draft, writing—review & editing, visualization.

Competing interests

William Calero-Cáceres is a member of the Editorial Board of *npj Antimicrobials and Resistance*. William Calero-Cáceres was not involved in the journal's review of, or decisions related to, this manuscript.

Additional information

Correspondence and requests for materials should be addressed to William Calero-Cáceres, Zhi Ruan or João Pedro Rueda Furlan.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

¹Biotechnology Program, Department of Food and Biotechnology Science and Engineering, Universidad Técnica de Ambato, Ambato, Ecuador. ²UTA-RAM-One Health, Group for Universal Advance in Bioscience, Department of Food and Biotechnology Science and Engineering, Universidad Técnica de Ambato, Ambato, Ecuador. ³Institute of Environmental, Chemical and Pharmaceutical Sciences, Federal University of São Paulo, Diadema, Brazil. ⁴Antimicrobial Resistance Institute of São Paulo (ARIES), São Paulo, Brazil. ⁵School of Veterinary Medicine, Metropolitan University of Santos, Santos, Brazil. ⁶Department of Internal Medicine, School of Veterinary Medicine and Animal Science, University of São Paulo, São Paulo, Brazil. ⁷Catalan Institute for Water Research (ICRA-CERCA), Girona, Spain. ⁸Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, UK. ⁹Centres for Antimicrobial Optimisation Network, Imperial College London, London, UK. ¹⁰The Fleming Initiative, Imperial College London and Imperial College Healthcare NHS Trust, London, UK. ¹¹Laboratory of Environmental Antimicrobial Resistance (LEARN), Department of Biological Sciences (DCB), Institute of Environmental, Chemical and Pharmaceutical Sciences (ICAQF), Federal University of São Paulo (UNIFESP), Diadema, Brazil. ¹²Department of Microbiology, Instituto de Ciências Biomédicas, University of São Paulo, São Paulo, Brazil. ¹³Department of Clinical Analysis, Faculty of Pharmacy, University of São Paulo, São Paulo, Brazil. ¹⁴Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, Brazil. ¹⁵Department of Clinical Laboratory, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China. ¹⁶Key Laboratory of Precision Medicine in Diagnosis and Monitoring Research of Zhejiang Province, Hangzhou, Zhejiang, China. ¹⁷Department of Clinical Analyses, Toxicology and Food Science, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo, Brazil. ¹⁸Department of Pharmaceutical Sciences, Health Sciences Center, Federal University of Paraíba, João Pessoa, Paraíba, Brazil. ✉e-mail: wr.calero@uta.edu.ec; r_z@zju.edu.cn; joao.furlan@academico.ufpb.br