www.envsciarch.com

ENVIRONMENTAL SCIENCE ARCHIVES
ISSN: 2583-5092
Volume V Issue 2, 2026

Received: 2026/07/11
Accepted: 2026/09/02
Published: 2026/09/12

REVIEW

OPEN ACCESS

AI for Surface Water Quality Prediction: A Comprehensive Review

Mohammed Salman* and Mohammed Sharif
Department of Civil Engineering, Jamia Millia Islamia, New Delhi 110025, India
*Correspondence should be addressed to MS (email: mohammed2400650@st.jmi.ac.in)

Abstract
Artificial intelligence for surface water quality prediction has evolved rapidly from classical regression and ensemble learning toward deep sequence, graph-based, physics-informed, and representation-learning approaches, while a largely disconnected systems literature has matured around digital twins, edge intelligence, and federated learning. This critical review synthesizes thirteen primary manuscripts against expanded 2020–2026 literature to examine the relationship between predictive performance and operational deployability. The evidence reveals a persistent bifurcation between an accuracy-maximizing research track and a deployability-maximizing systems track, corroborated independently at classical-modeling, digital-twin, and edge-hardware scales: the corpus's largest single study reports 31,289 samples from 23 stations, whereas a systematic review of 147 water-sector digital-twin studies identifies only 8 cases achieving genuine bidirectional control, and the largest review of environmental federated learning (361 studies) finds only 12 addressing water quality, with predominantly simulated deployments. Ensemble tree methods show the most consistently replicated predictive performance, while near-perfect results reported by simpler models remain constrained by the absence of external or cross-site validation. Explainable AI remains concentrated in feature-attribution methods, convergent SHAP findings require a causal-versus-correlational caveat, and the integration of spatial, physical, and temporal-representational deep-learning strategies remains unexplored. Building on these findings, this review develops a twelve-category Research Gap Matrix and proposes an Integrated Water Intelligence Framework linking prediction, model compression, governance-compliant decision-making, and closed-loop actuation. The review concludes that future progress requires predictive accuracy to be evaluated jointly with validation, explainability, governance, and deployability, and closes with a staged, stakeholder-differentiated roadmap.

Keywords: Surface water quality; Artificial intelligence; Machine learning; Water quality prediction; Explainable AI; Digital twins

Introduction
Surface water quality is a critical component of environmental management, since changes in physicochemical and biological conditions directly affect aquatic ecosystems, water use, and the reliability of freshwater resources. Reliable assessment is challenging because water-quality variables exhibit substantial spatial and temporal variability and are influenced by interacting natural and anthropogenic processes. Conventional monitoring provides observations at discrete locations and times, creating a need for predictive approaches that can extend monitoring information and support timely assessment and management. Recent studies have applied artificial intelligence and machine-learning methods to surface water quality evaluation, monitoring, and assessment, including conventional supervised learning, deep-learning architectures, and explainable AI approaches (Rana et al., 2023; Nallakaruppan et al., 2024). At the same time, emerging work on remote sensing and digital-twin systems is expanding the potential for spatially distributed and operational water-quality intelligence (Deng et al., 2024; Ghorbani Bam et al., 2025; He et al., 2026). However, these developments remain fragmented across predictive-modelling and systems-deployment communities. This fragmentation motivates a critical review that examines not only predictive accuracy but also validation, explainability, data integration, and the practical requirements for deployment.

Surface water quality prediction has no single physically-grounded target: it is a constructed index, the Water Quality Index (WQI), aggregating measurements against reference standards, computed either arithmetically

$$WQI = \Sigma_i w_i q_i \qquad (1)$$

or via entropy weighting, $w_i = (1-e_i)/\Sigma_j(1-e_j)$, where w_i, q_i (Eq. 1) are parameter i's weight/sub-index, and e_i its entropy. Das (2025) is the only study computing both, finding they diverge meaningfully in edge cases. The continuous WQI is typically discretized into a Water Quality Class via five ordinal grades, or six legally defined grades in China (GB3838-2002; Man et al., 2025).

The field's modelling paradigm evolved 2019-2025 through five stages (Fig. 1, upper track): classical regression (Ahmed et al., 2019); ensemble learning with SHAP/LIME (Nallakaruppan et al., 2024; Hridoy et al., 2025); deep sequence/CNN architectures (Rana et al., 2023; Man et al., 2025); graph-relational/physically-constrained models (Wan et al., 2025; Mu et al., 2025); and self-supervised representation learning (Zheng et al., 2025). Meanwhile, systems-deployment literature developed around digital twins, edge intelligence, and federated learning (lower track): most digital twins are not yet closed-loop (Ghorbani Bam et al., 2025).

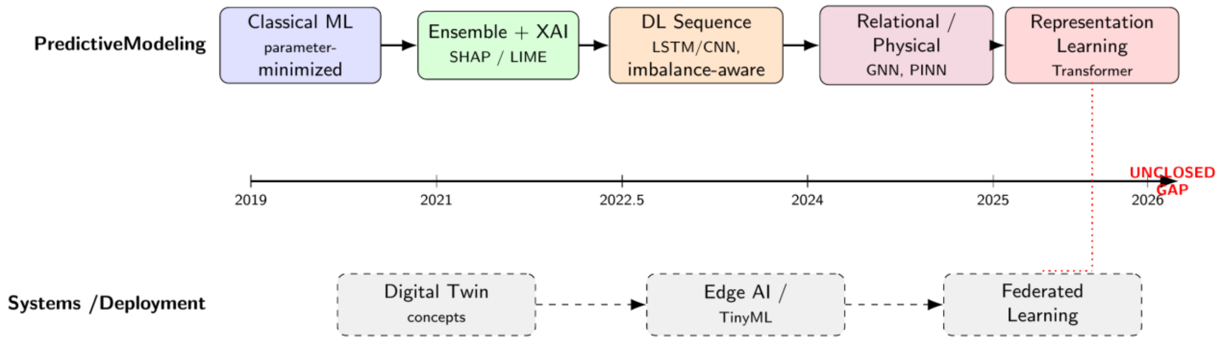

Fig. 1. AI evolution and systems-deployment timeline.

No existing review synthesizes these literatures together. Reading the corpus comparatively reveals that papers reporting state-of-the-art accuracy almost never report field deployment, and vice versa, with deploying authors trading sophistication for operational feasibility (Nuangpirom et al., 2025) — this dichotomy is this review's central finding.

A second tension concerns input parameter count: 4–7 parameters is a deployability choice (Ahmed et al., 2019; Al-Adhaileh and Alsaade, 2021), while 9–14 parameters serve regulatory fidelity (Das, 2025; Man et al., 2025). No study controls this tradeoff within one site; it recurs throughout this review.

Methods

This review follows a purposive, targeted critical-review methodology rather than a formal systematic-review protocol. The evidentiary base comprises thirteen primary manuscripts selected for representativeness, expanded through targeted 2020–2026 literature search. Synthesis proceeded thematically, comparing manuscripts within technique clusters to identify consensus and disagreement, not summarizing sequentially. Database names and search strings were not recorded, to avoid implying a more formal protocol than was followed.

Results

Data Ecosystem for AI-Based Water Quality Prediction

All models here are only as reliable as the data driving them: convenience-dataset studies (Ahmed et al., 2019; Nallakaruppan et al., 2024) document sensor methodology in a sentence, while systems literature (Ghorbani Bam et al., 2025) addresses calibration drift and redundancy in detail. Models reporting high accuracy on convenience datasets have never been tested under real sensor-noise conditions.

Man et al. (2025) is the corpus's most rigorous IoT-scale deployment (31,289 samples): six principal components (87.2% variance, KMO/Bartlett-validated) and a severely imbalanced class distribution (Class I:III ≈ 1:41) corrected via SMOTE, lifting Class I recall from 0.677. No other study reports class distribution numerically, so comparable imbalance elsewhere may be masked behind acceptable accuracy. ADASYN and autoencoder-based dimensionality reduction appear in no identified water-quality-specific source. Remote sensing is the single largest structural absence in the primary corpus: no manuscript employs satellite or UAV imagery, despite a mature sub-field evidenced by a 437-study PRISMA review (He et al., 2026), where chlorophyll-a retrieval accounts for 53.3% of publications. Platform selection involves a resolution tradeoff: MODIS offers near-daily revisit at 250–1000 m;

Landsat 30 m at 16 days; Sentinel-2 10–60 m at 5 days with red-edge bands for complex waters (Deng et al., 2024). This division is not incidental: the two sub-literatures rarely integrate, as Fig. 2 illustrates.

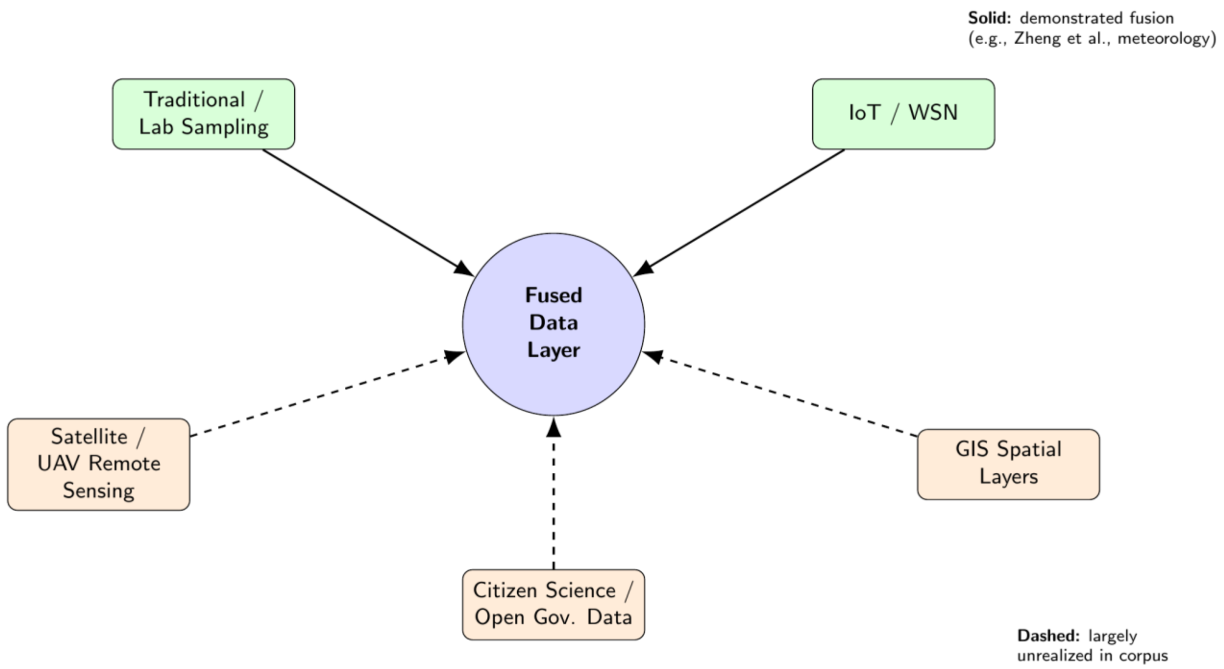

Fig. 2. Data ecosystem for water quality AI

Minimization-for-cost (Ahmed et al., 2019; Al-Adhaileh and Alsaade, 2021) and redundancy-for-trust (Man et al., 2025; Ghorbani Bam et al., 2025) address different problems: a four-sensor system is cheap but cannot detect its own drift; a nine-sensor system can, at higher cost.

AI Techniques for Water Quality Prediction

Fig. 3 places the techniques covered in this section into a single taxonomy that covers classical machine learning, deep learning, explainable AI, and systems-level deployment technology, marking those that are demonstrated in, emerging in, or absent from the surveyed literature.

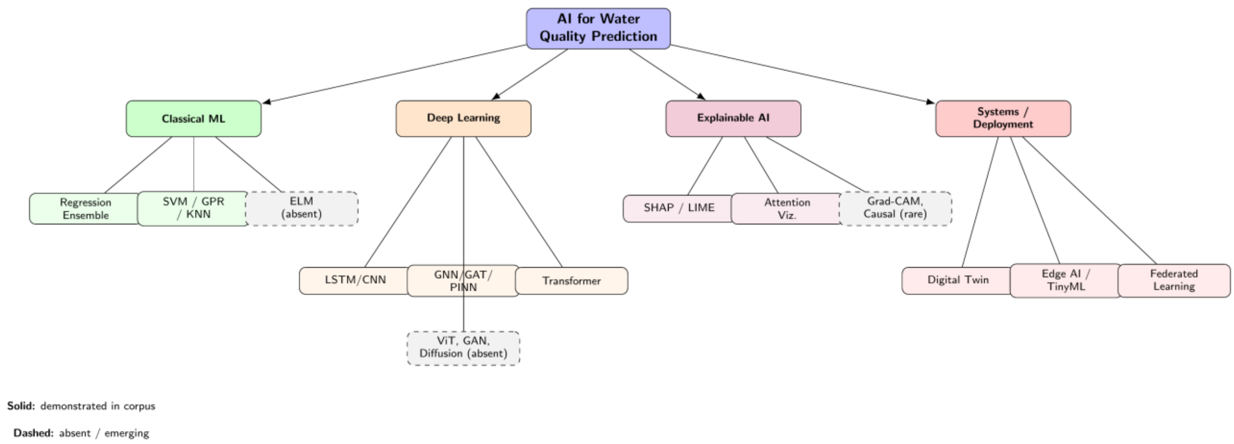

Fig. 3. AI taxonomy for surface water quality prediction

Classical, Ensemble, and Hybrid Learning

Ensemble tree methods constitute this review's most robust, replicated finding: gradient boosting is the best WQI regressor (Ahmed et al., 2019; MAE 1.96); XGBoost achieves R^2=0.973 (Das, 2025); LightGBM/XGBoost reach 99.65% accuracy (Hridoy et al., 2025). Single decision trees underperform ensembled counterparts — Man et al.'s (2025) C4.5 tree reaches only 87.3% accuracy against 94.5%, weakest on the minority classes identified in the data ecosystem section as under-reported. Table 1 summarizes these and the deep-learning results discussed in the AI techniques section, flagging validation status alongside each reported figure.

This consensus should be read alongside a second finding: the corpus's two highest reported accuracy figures belong to comparatively simple, unensembled models — Das (2025) reports MLR at R^2=0.9992; Al-Adhaileh and Alsaade (2021) report a shallow feedforward network at 100% accuracy. Neither has external, cross-site validation, a caution flagged explicitly in Table 1's final column. The most plausible explanation is not that simple models generically rival ensembles, but limited sample diversity and train/test leakage characteristic of single-site environmental data — a methodological caution every subsequent near-perfect accuracy claim in this literature should be read against.

Table 1. Classical/Ensemble/Deep Learning Algorithm Comparison

Algorithm	Best Reported Result	Corpus Source	Validation Status
Gradient Boosting	MAE 1.96 (WQI)	Ahmed et al. (2019)	Single-site, cross-validated
XGBoost	$R^2 = 0.973$	Das (2025)	Held-out test split
LightGBM/XGBoost	99.65% accuracy	Hridoy et al. (2025)	Grid-search tuned
MLR	$R^2 = 0.9992$ (flagged)	Das (2025)	No external validation — caution
FFNN	100% accuracy (flagged)	Al-Adhaileh and Alsaade (2021)	No external validation — caution
LSTM	95% accuracy	Rana et al. (2023)	Single-station time series
ST-GPINN (PINN+GNN)	$R^2 = 98.91\%$	Mu et al. (2025)	Controlled ablation
Transformer (encoder-only)	NSE = 0.80 (149 sites)	Zheng et al. (2025)	Cross-basin, data-scarcity tested

SVM ranks inconsistently across studies: competitive in Palabıyık and Akkan (2024), but below Random Forest in Nallakaruppan et al. (2024). GPR, cited once (Palabıyık and Akkan, 2024), is the only family offering native predictive uncertainty. ANFIS's sole application outperforms neural architectures for WQI regression (Al-Adhaileh and Alsaade, 2021), suiting environmental imprecision.

Deep Learning Architectures

Deep learning representation is uneven: two studies use sequence/convolutional architectures (Rana et al., 2023; Man et al., 2025); one each covers GNN, PINN, and Transformer (Wan et al., 2025; Mu et al., 2025; Zheng et al., 2025). On single-station series, LSTM outperforms ANN (95% vs. 87.5–92.5%); at different scale a simple feedforward network modestly beats CNN/LSTM (Man et al., 2025).

This is not a contradiction but evidence that architectural preference is scale- and structure-dependent: Man et al.'s data is modeled cross-sectionally, leaving LSTM's temporal-memory advantage little native structure to exploit. The corpus's three most advanced studies share no common benchmark (Table 2). Wan et al. (2025) uses plain graph convolution rather than attention-weighted aggregation; the "L-GAT" study (Liu et al., 2025) shows GAT-based aggregation outperforming plain GCN, improving dissolved-oxygen NSE from 0.694 to 0.783. Mu et al.'s (2025) ST-GPINN combines graph structure with an explicit physics-informed loss constraining predictions to the advection-reaction PDE governing chlorine transport,

$$L\_Total = L\_Data + \alpha \cdot L\_PDE, \quad \alpha = 0.3 \text{ (optimal, ablated)} \qquad (2)$$

and its own ablation shows the combined graph-plus-physics model outperforming graph-only or physics-only variants at scale. Zheng et al. (2025) instead encodes structure via cross-basin self-supervised pretraining, sustaining near-full accuracy (NSE 0.80) at 50% training data.

Table 2. Benchmark Datasets Across Corpus and Expanded Literature

Study	Region	N	Target Variable	Public?
Man et al. (2025)	Heilongjiang, China	31,289 samples, 23 stations	6-class WQC (GB3838-2002)	Not stated
Wan et al. (2025)	River network, China	Multi-station	DO, TN	Not stated
Mu et al. (2025)	Water distribution system, China	9–920 junctions	Chlorine residual	Simulated (EPANET)
Zheng et al. (2025)	6 basins, 149 sites, China	2007–2018	COD, DO, NH_3-N, pH	Partially public

These three architectures should be understood as complementary, not competing strategies (Fig. 4). No study combines them; this is this review's most significant single architectural gap. Increasing architectural sophistication has not been matched by reconsideration of whether the WQI/WQC target itself can exploit it. Generative architectures (autoencoders, GANs, diffusion models) and Vision Transformers remain almost entirely unrepresented in water-quality-specific literature — genuine, confirmed absences rather than oversights, stated here plainly rather than papered over with adjacent-domain citations.

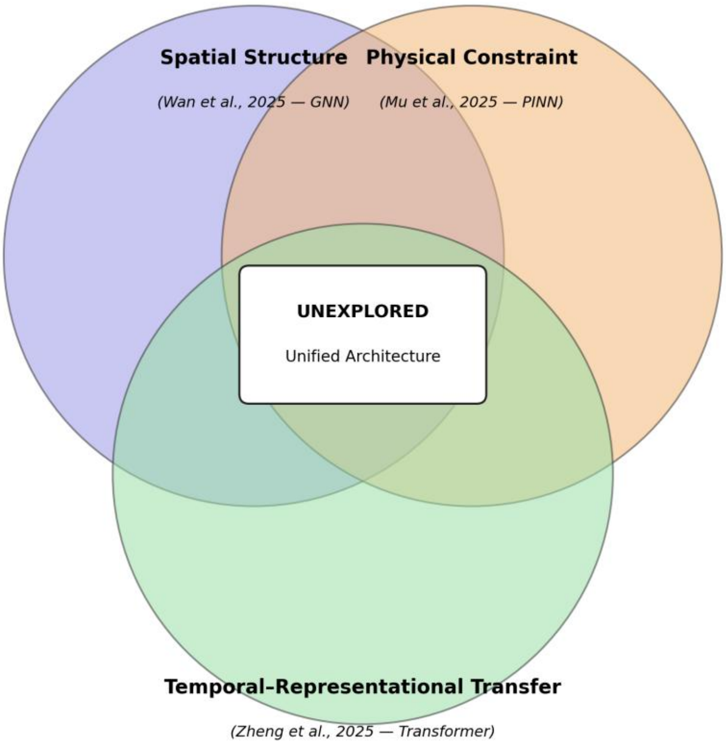

Fig. 4. Complementary deep-learning strategies for water quality prediction

Explainable and Trustworthy AI

Explainability is realized in only three of thirteen manuscripts (Nallakaruppan et al., 2024; Hridoy et al., 2025; Zheng et al., 2025); even there the toolkit is narrow (Fig. 5; Table 3): SHAP, LIME, and one attention-visualization instance exhaust the corpus's methods, with Integrated Gradients, Grad-CAM, PDP/ICE, and counterfactual explanation absent or near-absent. Hridoy et al. (2025) use SHAP as post-hoc auditing atop an optimized ensemble, while Nallakaruppan et al. (2024) select Random Forest specifically for its explainability value, generating SHAP plots jointly with LIME. Both independently designate solids/dissolved-solids as the most important parameter — a substantive convergence rare in this corpus.

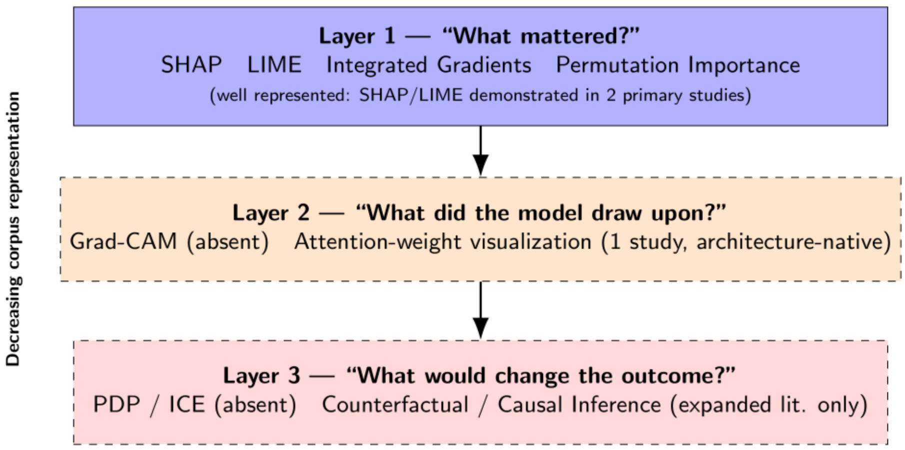

Fig. 5. Three-layer framework for explainable AI

Table 3. XAI Method Comparison

Method	Layer	Demonstrated in WQ Literature?
SHAP	Feature attribution	Yes
LIME	Feature attribution	Yes
Attention-weight visualization	Mechanistic transparency	Yes (native)
Grad-CAM	Mechanistic transparency	No
PDP / ICE	Bridge to causal	No
Counterfactual/Causal Inference	Causal validation	Expanded lit. only

That convergence carries an under-acknowledged caveat: a 2025 causal-ML study of Hong Kong coastal water quality (Zhang et al., 2025) combines SHAP with structural causal modeling, concluding SHAP importance "should be interpreted cautiously, particularly when inferring causality," since consistent model behavior is not evidence a feature is causal, rather than a proxy for an unmeasured cause. This matters for policy: a regulator treating a SHAP-identified parameter as causal may target a symptom, not a cause.

A second asymmetry compounds this: XAI here is applied exclusively to classical or ensemble models, never to the graph, physics-informed, or Transformer architectures discussed in the AI techniques section (except Zheng et al.'s attention visualization). The field's most predictively sophisticated models are, at present, its least explainable. Feature attribution, mechanistic transparency, and causal validation (Fig. 5) are complementary layers of trustworthy explanation, not interchangeable alternatives — and near-total concentration in the first layer limits XAI's practical utility.

$$\phi_i = \Sigma\_\{S \subseteq F \setminus \{i\}\} \left[|S|!(|F|-|S|-1)!/|F|!\right] \cdot \left[f(S \cup \{i\}) - f(S)\right] \qquad (3)$$

where ϕ_i is feature i's Shapley value, F the feature set, S a subset excluding i, and f(S) the model output for S.

Emerging Intelligent Water Systems

Every architecture surveyed in the Results section exists in this corpus as an offline, retrospectively-evaluated artifact. A systematic review of 147 water-sector digital twin studies (2015–2025) (Ghorbani Bam et al., 2025) supplies this review's most consequential finding: only 8 of 147 achieve genuine bidirectional control, the rest functioning as "digital shadows" regardless of modeling sophistication (Fig. 6).

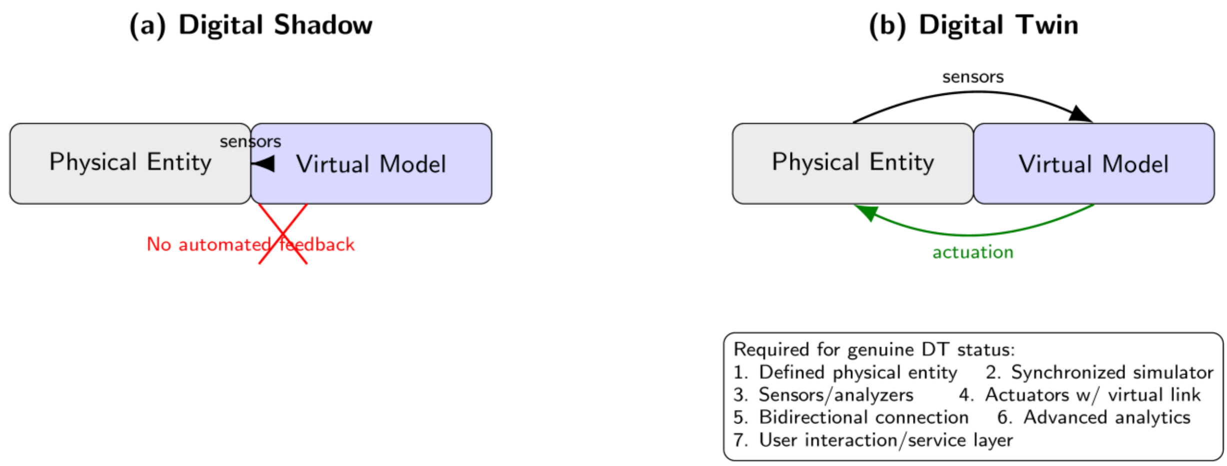

Fig. 6. Digital Shadow versus Digital Twin architecture

Full-scale cases illustrate this (Table 4): Singapore's Changi plant (~1200 SCADA tags) remains advisory-only, as does Eindhoven's, whose forecasts worsen during wet weather (see the data ecosystem section). By contrast, Germany's Cuxhaven plant cuts aeration energy 30% via automated setpoint control alone; China's Lushan is one of two examples combining advisory function with process control. The barrier is institutional, not technical: operators remain legally liable, and safety-critical logic often resides in proprietary code unreplicated in the twin's decision layer.

Table 4. Digital Twin Case Studies: Deployment Status

Facility	Location	Advisory-only or Closed-loop?	Key Outcome
Changi WRP	Singapore	Advisory-only	5-day forecasting, ~1,200 SCADA tags
Eindhoven WRRF	Netherlands	Advisory-only	Nitrate/MLSS degrade in wet weather
EWE Wasser/Cuxhaven	Germany	Closed-loop (narrow)	30% aeration energy reduction (~1.2M kWh/yr)
Lushan Water Supply	China	Advisory + process control	Automated distribution operation

This repeats at the smallest scale: the only field-tested edge hardware in the corpus is Nuangpirom et al.'s (2025) ESP32 system for dissolved-oxygen/pH control in Thai aquaculture. Of three models evaluated, Random Forest gave the highest offline accuracy (R2>0.80), but Multiple Linear Regression (R2≈0.65–0.72) was deployed, selected for computational efficiency and true offline operability (Table 5); the deployed system maintained safe DO and stabilized pH within roughly an hour. The field's only two rigorously documented deployment successes — Cuxhaven and this ESP32 case — both sacrifice achievable accuracy: the clearest evidence for this review's central bifurcation thesis.

Table 5. Edge/Federated Learning Deployment Comparison

Source	Models Evaluated	Model Actually Deployed	Real-world Validation?
Nuangpirom et al. (2025)	MLR, DTR, RFR	MLR (over higher-accuracy RFR)	Yes
FL water-quality subset (12 studies) (Miller et al., 2025)	LSTM, GRU (federated)	Not specified	Predominantly simulated

Federated learning remains immature here: the largest systematic review of environmental FL (361 studies) (Miller et al., 2025) identifies only 12 addressing water quality, reporting federated LSTM/GRU skill "near-equivalent" to centralized models at 60–80% lower data-transfer. The proposed carbon-aware federated loss remains untested; Green AI here exists only as a proposed framework.

Governance concerns recur independently in the digital-twin and federated-learning literatures (Miller et al., 2025), yet remain almost absent from the predictive-modeling literature, despite being the primary barrier to operational impact. Large language models for water-sector decision support (e.g., Xu et al., 2025) operate only at the communication layer, a separate research community from the Transformer-for-prediction literature.

Discussion

Research Challenges and Future Directions

Consolidating evidence across the Methods, Results, and Discussion sections (Table 6, Fig. 7) surfaces twelve gaps; one is the organizing tension beneath the others. The accuracy/deployability bifurcation recurs independently in three domains: the untested parameter-count tradeoff (Introduction and data ecosystem section), the 8-of-147 digital-twin finding (Emerging Intelligent Water Systems section), and the Random-Forest-to-MLR edge downgrade (Emerging Intelligent Water Systems section) — the largest wedges in Fig. 7. Better benchmarks alone would not resolve deployment unless this tension is addressed.

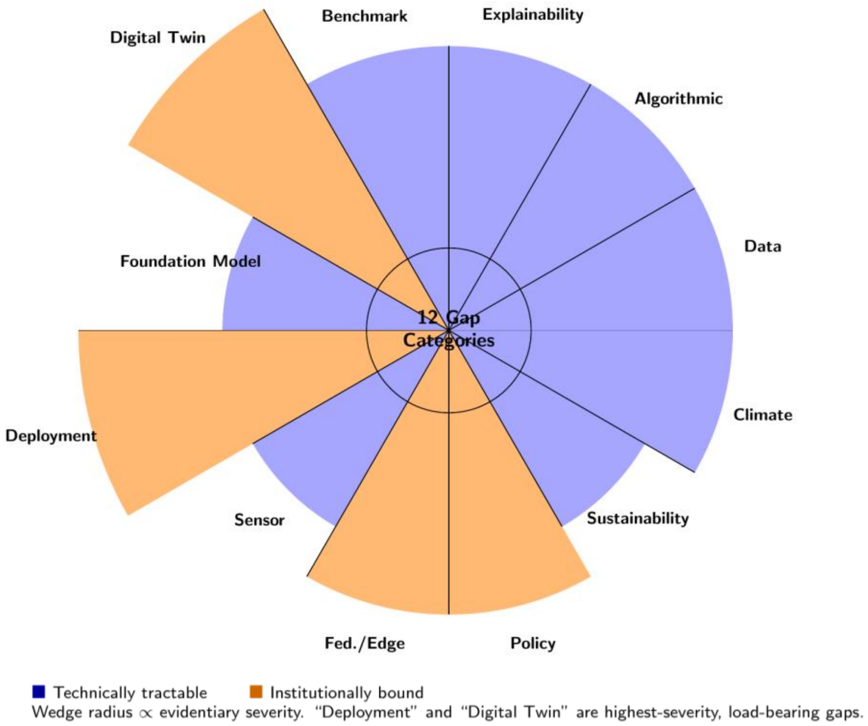

Fig. 7. Research gap framework for water-quality AI

This indicates intervention type: data, algorithmic, benchmark, explainability gaps are tractable through conventional research; digital-twin, deployment, federated/edge, policy gaps are institutional, requiring governance engagement.

Table 6. Comprehensive Research Gap Matrix

Category	Specific Gap	Severity	Tractability
Data	No cross-site validation; remote sensing disconnect	High	Technical
Algorithmic	No unified spatial+physical+representational model	High	Technical
Explainability	XAI absent for deep/graph models; no causal validation	High	Technical
Benchmark	No shared datasets across GNN/PINN/Transformer work	High	Technical
Digital Twin	Only 8/147 studies achieve bidirectional control	Highest	Institutional
Foundation Model	LLM/prediction communities non-overlapping	Medium	Technical
Deployment	Accuracy/deployability bifurcation (load-bearing)	Highest	Institutional
Sensor	Cost-vs-trustworthiness unresolved	Medium	Technical
Federated/Edge	Zero real-world FL deployments	High	Institutional
Policy/Governance	Governance absent from predictive literature	High	Institutional
Sustainability	Carbon-aware learning untested	Medium	Technical
Climate	Non-stationarity unaddressed in corpus	High	Technical

In response, this review proposes an Integrated Water Intelligence Framework (Fig. 8): a prediction layer combining the spatial, physical, and representational strategies discussed in the AI techniques section rather than treating them as exclusive; a compression/distillation stage evaluated against edge-deployment constraints early; and a governance-compliant decision layer gated by constraint logic. This is a proposed contribution requiring future validation, not a definitive finding.

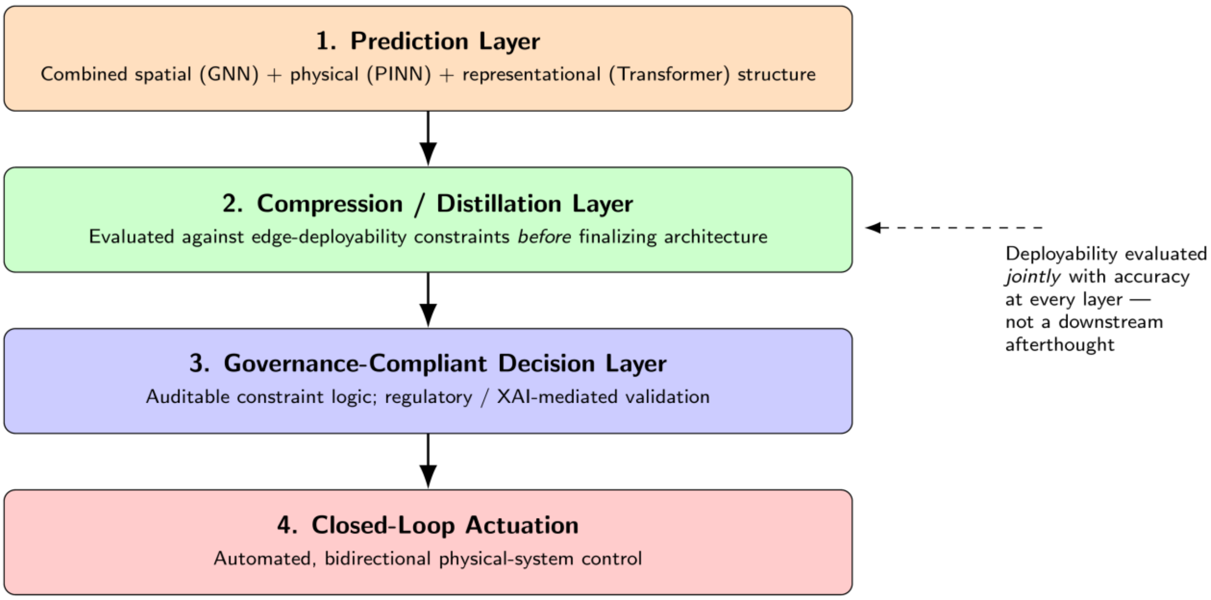

Fig. 8. Integrated Water Intelligence Framework

A staged roadmap follows (Table 7). Near term: parameter-count ablation studies; shared GNN/PINN/Transformer benchmarks; XAI extension to deep architectures; bounded-automation digital-twin pilots. Mid term: a unified architecture combining the three strategies discussed in the AI techniques section; non-simulated federated-learning deployments; carbon-aware learning validation; LLM-based decision-support integration. Long term: utility-scale Framework deployment, with governance pathways for validation standards.

Table 7. Roadmap Priorities by Stakeholder

Stakeholder	Near-Term (2026–2028)	Long-Term (2032+)
Researchers	Deployability-aware architecture design	Unified Integrated Water Intelligence Framework validation
Utilities	Narrow, bounded automation (Cuxhaven model)	Full closed-loop systems integration
Regulators	AI-model validation pilot frameworks	Standardized certification pathways

Conclusions

Reading this literature comparatively rather than domain-by-domain surfaces a finding no siloed review has established: the field has bifurcated into an accuracy-maximizing research track and a deployability-maximizing systems track, with no study occupying both. This finding, formalized as a Research Gap Matrix and addressed through an Integrated Water Intelligence Framework, is this review's central contribution toward a field capable of more than it has been permitted to deploy. Three findings qualify the field's progress. First, the highest single-site accuracy figures lack external validation and remain a methodological caution, not a benchmark. Second, explainability remains concentrated in feature-attribution methods, leaving sophisticated models unexplained and convergent SHAP findings vulnerable to being misread as causal. Third, governance is concentrated in the systems literature and nearly absent from predictive literature, despite being the primary barrier to deployment.

References

Ahmed U, Mumtaz R, Anwar H, Shah AA, Irfan R and García-Nieto J (2019) Efficient water quality prediction using supervised machine learning. Water 11(11):2210.

Al-Adhaileh MH and Alsaade FW (2021) Modelling and prediction of water quality by using artificial intelligence. Sustainability 13(8):4259.

Ambika K, Shanmugapriya P, Alghamdi AG and Alsolai H (2026) Monitoring coastal water quality using remote sensing and machine learning. Ocean Science Journal 61:22.

Ashiquzzaman A, Kim J and Kim C (2025) Variational autoencoder-regulated data augmentation for precise estimation of organic pollutants in water with deep learning neural network. Applied Artificial Intelligence 39(1).

Das S (2025) Comparative water quality index modelling using arithmetic and entropy weighting with machine learning in Lakhanpur Block, Odisha. Desalination and Water Treatment 324:101548.

Deng Y, Zhang Y, Pan D, et al. (2024) Review of recent advances in remote sensing and machine learning methods for lake water quality management. Remote Sensing 16(22):4196.

Ghorbani Bam P, Rezaei N, Roubanis A, et al. (2025) Digital twin applications in the water sector: a review. Water 17(20):2957.

He Q, Wang D, Ji F, Zhu L, Li R, Tian T, Zhang Q, Tao Y and He M (2026) Machine learning-driven multi-source remote sensing for surface water quality retrieval: progress and prospects. Water 18(14):1744.

Hridoy RC, et al. (2025) Machine learning framework with SHAP-based explainability for aquaculture water quality classification and regression. Science of the Total Environment 1008:180965.

Ibrahim A, Alkarbi N, Alsaadi M, Almazrouei A, Alshamsi S and Hamouda MA (2026) Satellite-based machine learning models for chlorophyll-a and TSS retrieval in Abu Dhabi's coastal waters. Frontiers in Marine Science 13:1787597.

Liu Y, Zheng H and Zhao J (2025) Enhanced water quality prediction by LSTM and graph attention network (L-GAT): an analytical study of the Pearl River Basin. Water Research X 28:100383.

Man G, Yun Q, Qilin Z, et al. (2025) Surface water quality classification and prediction model based on multiple machine learning algorithms. Scientific Reports 15:39907.

Miller T, Durlik I, Kostecka E, et al. (2025) Federated learning for environmental monitoring: a review of applications, challenges, and future directions. Applied Sciences 15(23):12685.

Mu T, Duan F, Ning B, et al. (2025) ST-GPINN: a spatio-temporal graph physics-informed neural network for enhanced water quality prediction in water distribution systems. npj Clean Water 8:74.

Nallakaruppan MK, Gangadevi E, Lawanya Shri M, Balusamy B, Bhattacharya S and Selvarajan S (2024) Reliable water quality prediction and parametric analysis using explainable AI models. Scientific Reports 14:7520.

Nuangpirom P, Pitjamit S, Jaikampan V, et al. (2025) Machine learning on low-cost edge devices for real-time water quality prediction in tilapia aquaculture. Sensors 25(19):6159.

Palabıyık EK and Akkan Ö (2024) Comparative modelling of water quality using ANN, SVM, GPR and ensemble methods for Aksu Creek, Türkiye. Environment, Development and Sustainability.

Rana R, Kalia A, Boora A, Alfaisal FM, Alharbi RS, Berwal P, Alam S, Khan MA and Qamar O (2023) Artificial intelligence for surface water quality evaluation, monitoring and assessment. Water 15(22):3919.

Ravindran JS, Sermet Y, Mount J, Vald G, Shrestha S, Cwiertny D and Demir I (2025) Application of large language models in developing conversational agents for water quality education, communication, and operations. Water Practice and Technology 20(10):2094–2109.

Wan H, Xiang L, Cai Y, et al. (2025) Temporal and spatial feature extraction using graph neural networks for multi-point water quality prediction in river network areas. Water Research 281:123561.

Xu B, Wu G, Li Z, Xu G, Zeng H, Tong R and Ng HY (2025) Towards domain-adapted large language models for water and wastewater management: methods, datasets and benchmarking. npj Clean Water 8:82.

Zhang H, Li Y, Li M, Wei S and Zhang H (2025) Explainable and causal machine learning to investigate the spatiotemporal dynamics patterns of coastal water quality in Hong Kong. Water Research 290:125026.

Zheng Y, Zhang X, Zhou Y, et al. (2025) Deep representation learning enables cross-basin water quality prediction under data-scarce conditions. npj Clean Water 8:33.

Author Contributions

Mohammed Salman: conceptualization, literature analysis, synthesis, and manuscript preparation. Mohammed Sharif: supervision, critical review, and manuscript evaluation. Both authors conceived the concept, wrote and approved the manuscript.

Acknowledgements

The authors gratefully acknowledge the Department of Civil Engineering, Jamia Millia Islamia, New Delhi, for providing the academic and research environment supporting this work. The authors sincerely acknowledge Prof. Mohammed Sharif for his guidance, supervision, and constructive support throughout the development of this review.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Competing interest

The authors declare no competing interests.

Ethics approval

Not applicable.

Citation: Salman M and Sharif M (2026) AI for Surface Water Quality Prediction: A Comprehensive Review. Environmental Science Archives 5(2): 485-494.