

A Review on Computational Eco-Dynamics: An Advanced Synthesis of Machine Learning, Subsurface Source Apportionment, and Microbial Profiling across India's Environmental Basins

Ashish Pal¹, Ashish Singh*² and Vaishali³

^{1,3}ICAR-National Bureau of Fish Genetic Resources Canal Ring Road, P.O,
Dilkusha Marg, Lucknow, Uttar Pradesh 226002

²Kerala University of Fisheries and Ocean Studies, Panangad, Kochi, Kerala, 682506 India.

Corresponding Author

Ashish Singh

Email: ashish10899singh@gmail.com



OPEN ACCESS

Keywords

Computational Eco-Dynamics; Machine Learning; Source Apportionment; Microbial Profiling; Environmental Basins

How to cite this article:

Pal, A., Singh, A. and Vaishali. 2026. A Review on Computational Eco-Dynamics: An Advanced Synthesis of Machine Learning, Subsurface Source Apportionment, and Microbial Profiling across India's Environmental Basins. *Vigyan Varta* 7 (10): 34-42.

ABSTRACT

Computational eco-dynamics has emerged as a useful umbrella for integrating hydrochemistry, microbial ecology, source-tracing, and predictive analytics into a single environmental decision framework. In the Indian context, this convergence matters because major river basins and connected aquifers are shaped by dense human settlement, agricultural runoff, industrial discharge, monsoon variability, and rapidly changing land use, all of which create strongly coupled surface-subsurface pollution regimes that cannot be understood through single-discipline methods alone. A growing body of literature now shows that machine learning can improve microbial classification from metagenomic data, identify high-information environmental predictors, and strengthen water-quality assessment and source apportionment when combined with geospatial and hydro chemical evidence. At the same time, microbial profiling studies from the Ganga-Yamuna system

demonstrate that polluted stretches contain distinct taxonomic and functional signatures, including taxa and protein domains associated with biodegradation, metal transport, and xenobiotic transformation.

INTRODUCTION:

Water is both an effective instrument of economic progress and a crucial natural resource for human survival. Among the most important elements of an ideal environment is referred as Water Quality (WQ). Rivers are one of the important sources of fresh water and account for only about 0.3% of the total surface waterbody sources Savari *et al.* (2022). Environmental basins are highly dynamic socio-ecological systems in which hydrological, physicochemical, biological and anthropogenic processes interact across multiple spatial and temporal scales. In India, rapid urbanization, agricultural intensification, industrial discharge and increasing wastewater inputs have altered the ecological functioning of several major river basins. Tiyyasha *et al.* (2020) emphasized that the nonlinear interactions among water-quality parameters make conventional statistical approaches insufficient for comprehensive river-water-quality modelling. Besides, water pollution is also increased by home waste water, municipal solid waste, and industrial wastewater. Eutrophication happens when excess nutrients like nitrates and phosphates cause too much plant and algae growth, reducing dissolved oxygen (DO) and harming aquatic life Das *et al.* (2025). However, due to population expansion and urbanization, there has been a global shortage of freshwater resources in recent years, exposing water supplies to declining quality and quantity Mukhopadhyay *et al.* (2021) The emergence of artificial intelligence and machine-learning techniques has provided new opportunities to overcome these limitations. At the basin scale, however, prediction alone is insufficient because

environmental management also requires identification of pollution sources and their spatial distribution. Choudhury *et al.* (2024) demonstrated the potential of computational and hydrochemical approaches for distinguishing anthropogenic and natural contributions to environmental contamination, emphasizing the importance of source-specific assessment for effective pollution management. Alongside chemical assessment, microbial communities provide an important biological dimension of environmental monitoring. Behera *et al.* (2020) investigated sediment-associated microbial communities of the Ganga and Yamuna and demonstrated the value of metagenomic information for understanding microbial diversity and functional potential under different environmental conditions. Such studies indicate that environmental microbiomes can act as sensitive biological indicators of ecosystem disturbance. This approach has considerable potential for connecting biological fingerprints with pollution pathways and could substantially improve source identification in heterogeneous river-basin environments. Recent advances in explainable artificial intelligence are further changing environmental modelling from a purely predictive discipline toward a more interpretable framework.

Remote sensing (RS) and Geographical Information System (GIS) have successfully managed a significant amount of hydrological data in recent decades to create more accurate water susceptibility maps Dash *et al.* (2021). Meanwhile, GIS software is employed to create geographic maps with a variety of

themes, that analyse variations in surface water quality. Furthermore, advancements in RS and GIS technology have facilitated the rapid identification of surface water zones across large regions. A study discussed by, Gugulothu *et al.* (2022), which adopted the joint approach, based on GIS and restricted irrigated WQI, that initiates to evaluate the surface water suitability in the research area of West Pampa Plain, Argentina. Further, utilizing sample points from various locations to create and interpolate a continuous parameter across the value from the observed parameters, and use geostatistical analysts to forecast values for every site in the landscape Romanelli *et al.* (2012)

Why the synthesis matters

The integration of machine learning, source apportionment and microbial profiling is important because environmental pollution cannot be adequately understood through prediction alone. Recent Indian studies show that explainable ML can identify the dominant physicochemical drivers of water quality, while source-apportionment approaches can link contamination with probable anthropogenic sources. Choudhary *et al.* (2025) demonstrated the value of SHAP-based ensemble modelling for interpreting river-water-quality predictions, whereas Karunanidhi *et al.* (2026) integrated hydrochemistry, ML and Monte Carlo simulation to identify major sources of groundwater nitrate contamination. Similarly, combining hydrochemical information with unsupervised learning and explainable AI can reveal spatial water-quality patterns and their environmental drivers. Appukuttan *et al.* (2025) demonstrated this approach in the Ithikkara River Basin, showing how computational models can connect water-quality prediction with hydrochemical processes.

Conceptual architecture of computational eco-dynamics

A practical computational eco-dynamics architecture for Indian basins contains four interacting layers. The first layer is environmental observation: hydrology, hydrochemistry, sediment quality, land use, climate seasonality, and infrastructure information such as drains, barrages, and industrial estates. The second is biological observation through microbial profiling, usually via amplicon sequencing or shotgun metagenomics, supported by taxonomic and functional annotation (Tiyasha *et al.*, 2020; Behera *et al.*, 2020a). The fourth layer is decision support, where outputs are translated into risk maps, priority zones, remediation strategies, and monitoring designs tailored to specific basin segments (Appukuttan *et al.*, 2025; Behera *et al.*, 2020b).

Machine learning in basin diagnostics

Machine learning (ML) has increasingly evolved from a supplementary analytical technique into an important computational layer for environmental and basin-scale diagnostics. In Indian river metagenomics, Chakraborty *et al.* (2023) applied clustering and supervised learning to metagenomic datasets from the Ganga and Yamuna rivers to classify previously unannotated scaffolds. Their results showed that the Random Forest (RF) model achieved superior performance compared with the other evaluated algorithms during 10-fold cross-validation, demonstrating the potential of ML for microbial diversity assessment in complex and high-dimensional riverine metagenomic datasets. Beyond metagenomic classification, ML is increasingly being integrated with hydrochemistry, geospatial variables, and water-quality indices for river-basin assessment. Kushwaha *et al.* (2024) developed hybrid artificial neural network (ANN) models incorporating Random Forest, Support Vector

Machine (SVM), and other algorithms for predicting water-quality conditions in the Bagh River Basin, demonstrating the value of hybrid ML frameworks for environmental monitoring. Recent work has also extended ML from prediction toward source diagnosis and environmental risk assessment. Karunanidhi *et al.* (2026) combined hydrochemical data, machine learning and Monte Carlo simulation to investigate groundwater nitrate contamination in the Suruliyar River Basin, South India.

Microbial profiling as ecological evidence

Microbial profiling has emerged as an important biological layer in river-basin environmental assessment because it captures ecosystem responses that may not be evident from physicochemical parameters alone. Metagenomic analysis of Ganga and Yamuna sediments revealed substantial bacterial and fungal diversity, including 45 bacterial genera with 92 species and 13 fungal genera with 24 species showing potential bioremediation functions. Proteobacteria, Actinobacteria, Firmicutes and Deinococcus–Thermus were among the dominant bacterial groups, while several pollutant-responsive taxa were enriched in contaminated stretches (Behera *et al.*, 2020). Further studies have demonstrated that microbial communities vary with pollution intensity and sediment characteristics. Shukla *et al.* (2022) identified diverse bacterial communities in Ganga sediments and reported functional pathways associated with the degradation of xenobiotic compounds, including benzoate, aminobenzoate and nitrotoluene. Similarly, Behera *et al.* (2022) reported pollutant-degradation genes and enzymes in Ganga sediments from Kanpur, indicating that sediment microbiomes can provide functional evidence of anthropogenic contamination. River confluences represent particularly dynamic microbial environments. Samson *et al.* (2020) showed that fungal communities

around the Ganges–Yamuna confluence exhibited taxonomic and functional patterns associated with anthropogenic pollution and eutrophication. Rajput *et al.* (2022) further demonstrated substantial functional diversity of carbohydrate-active enzymes in the Ganges–Yamuna confluence, indicating that microbial communities can respond to changes in organic matter and environmental conditions across mixing zones.

Recent metagenomic evidence has further identified microbial enzymes associated with plastic and xenobiotic degradation in Ganga sediments. Rout *et al.* (2024) reported differences in microbial communities between Devprayag and Prayagraj sediments and identified genes encoding enzymes potentially involved in the degradation of plastics and several xenobiotic compounds. These findings indicate that microbial communities reflect both contamination pressure and the potential for natural biotransformation processes. Another metagenomic study of the Ganges–Yamuna confluence at Prayagraj showed that the Yamuna had a transient influence on taxonomic and functional microbial patterns downstream, with particularly notable effects on xenobiotic degradation, harmful-substance resistance, and antibiotic-resistance related functions. This finding matters because it demonstrates how confluences can create short-lived but ecologically meaningful perturbations, which static sampling may miss. In computational terms, it supports event- and location-aware modeling rather than reliance on broad annual averages. (Samson *et al.*, 2019)

Indian basin Evidence Base

The current evidence base for India is promising but uneven. The strongest evidence comes from the Ganga–Yamuna system, while recent studies are expanding microbial and computational approaches to other major basins. Regar *et al.* (2024) conducted a

comparative metagenomic assessment of the Ganga, Narmada, Cauvery and Gomti rivers and demonstrated distinct microbial signatures associated with anthropogenic contamination and metal stress. This provides important evidence that microbial community patterns can be compared across geographically and environmentally contrasting Indian river systems. The Brahmaputra Basin is another emerging area for microbial evidence. Sharma *et al.* (2024) used shotgun metagenomics to characterize Brahmaputra River sediments and identified 31 bacterial phyla, 1,016 genera and 3,630 species, together with functional genes and antimicrobial-resistance determinants. Such datasets demonstrate that microbial profiling can extend computational eco-dynamics beyond the better-studied Ganga system and provide biological evidence for assessing pollution, ecological functioning and resistance pressure.



Figure 1: Illustrates the review’s central argument: environmental diagnosis improves when hydrological, chemical, spatial, and microbial evidence are fused into a common inference layer. This integrated workflow is consistent with recent basin studies that combine environmental monitoring, metagenomic analysis, and predictive modeling for classification or source diagnosis.

Table 1: Comparative role of major analytical components

Analytical component	Main purpose	Typical data inputs	Strengths	Main limitations
Machine learning	Prediction, classification, variable ranking, anomaly detection	Water chemistry, sediments, land use, sequencing features	Captures nonlinear patterns and high-dimensional relationships	Risk of overfitting, weak transferability, limited interpretability in some models
Source apportionment	Quantify contributions from likely pollutant sources	Tracers, ions, nutrients, GIS covariates, discharge context	Supports management action by linking contamination to likely origins	Sensitive to assumptions, incomplete source libraries, and mixing effects
Microbial profiling	Detect ecological response and transformation potential science	Amplicon reads, shotgun metagenomes, functional annotations	Reveals biological stress response and biodegradation capacity	Taxonomy does not guarantee function, and databases remain incomplete
Integrated computational eco-dynamics	Connect source, transport, and ecological response in one framework	Combined hydrochemical, spatial, and microbial layers	Most useful for basin management and adaptive monitoring	Requires harmonized sampling, higher computational capacity and careful validation

Methodological strengths

The major strength of computational eco-dynamics is triangulation, where hydrochemistry describes contaminant conditions, machine learning identifies nonlinear relationships, and microbial profiling provides biological evidence of ecosystem response. Combining these evidence streams can provide a more reliable interpretation than relying on a single monitoring approach (Behera *et al.*, 2020; Dhapre *et al.*, 2025). A second strength is the ability of machine learning to capture complex and nonlinear water-quality relationships. In the Godavari River Basin, Satish *et al.* (2024) demonstrated that a stacking ANN framework combined with geospatial variables and land-use/land-cover information could improve stream-water-quality prediction.

Recent Ganga studies further demonstrate the predictive value of computational approaches. Das *et al.* (2025) used multiple machine-learning models, including Random Forest and Boost, to predict nitrate-N concentrations in the lower Ganga using hydro chemical and climatic variables. The study illustrates how historical monitoring data can be transformed into predictive information for ecosystem-health assessment. Another methodological advantage is spatial integration. GIS can connect monitoring observations with land use, drainage networks, pollution sources and hydrological characteristics, allowing machine-learning predictions to be mapped across larger areas. Recent Mahanadi Basin studies have combined GIS, water-quality indices and machine learning for pollution assessment and source identification, demonstrating the potential for spatially explicit environmental diagnosis (Das, 2025).

Machine learning also provides an opportunity to improve monitoring efficiency. Rather than sampling every location with equal intensity, predictive models can identify high-risk zones

and guide where additional field measurements are most valuable. This is particularly important in India, where large river basins and limited monitoring resources make uniformly dense sampling difficult. The methodological value extends to groundwater systems as well. Dhapre *et al.* (2025) highlighted the increasing application of machine learning for groundwater monitoring, including prediction, anomaly detection and contamination assessment. Integrating these approaches with hydrochemistry and GIS can help bridge the gap between individual sampling points and basin-scale groundwater management. An additional strength is the growing use of explainable and ensemble machine learning. Kundu *et al.* (2025) demonstrated the application of explainable artificial intelligence for assessing water-quality criteria across nine major rivers, while ensemble approaches can improve prediction by combining the strengths of several algorithms. This improves not only predictive performance but also the interpretability required for environmental decision-making.

Implications for remediation and governance

The practical value of this review lies in environmental governance. If pollution sources can be estimated more confidently and microbial response patterns can indicate where natural attenuation is active or failing, agencies can design more selective interventions rather than relying on generalized cleanup approaches. In a basin such as the Ganga, this could mean distinguishing stretches where sewage interception is the priority from stretches where sediment legacy contamination, industrial effluents, or groundwater inflows are more important controls. (Das, 2025; Dhapre *et al.*, 2025; Karunanidhi *et al.*, 2026) The approach also supports adaptive monitoring. Machine learning can be used to identify the smallest variable set that still predicts

ecological condition well, allowing agencies to reserve high-cost omics analysis for decision-critical nodes while maintaining cheaper hydrochemical surveillance elsewhere. Over time, this can produce a more efficient and scientifically grounded monitoring architecture for India's river basins and aquifer systems.

CONCLUSION

Computational eco-dynamics offers a credible next step for environmental basin science in India because it treats contamination as a dynamic interaction among sources, transport pathways, and ecological response rather than as a static exceedance problem. Current evidence from the Ganga-Yamuna system and emerging groundwater studies shows that machine learning, source apportionment, and microbial profiling already have complementary strengths, and that their integration can improve both diagnosis and management relevance. The field is not yet mature. Basin coverage is still fragmented, seasonal datasets remain limited, and microbial function is often inferred more confidently than validated. Even so, the direction is clear: India's environmental basins are ready for a new generation of computational, biologically informed, and policy-linked assessment systems that can move from reactive monitoring toward predictive stewardship.

REFERENCES

- Appukuttan, A., Aju, C. D., Reghunath, R., *et al.* (2025). Exploring hydrochemical drivers of drinking water quality in a tropical river basin using self-organizing maps and explainable AI. *Water Research*, 284, 123884.
- Behera, B. K., Patra, B., Chakraborty, H. J., Sahu, P., Rout, A. K., Sarkar, D. J., Parida, P. K., Raman, R. K., Rao, A. R., Rai, A., Das, B. K., Jena, J., & Mohapatra, T. (2020). Metagenome analysis from the sediment of river Ganga and Yamuna: In search of beneficial microbiome. *PLOS ONE*, 15(10), e0239594.
- Choudhury, N., Sahu, T. K., Rao, A. R., Rout, A. K., & Behera, B. K. (2023). An improved machine learning-based approach to assess the microbial diversity in major North Indian River ecosystems. *Genes*, 14(5), 1082.
- Das, A. (2025). An optimized approach for predicting water quality features and a performance evaluation for mapping surface water potential zones based on Discriminant Analysis (DA), Geographical Information System (GIS) and Machine Learning (ML) models in Baitarani River Basin, Odisha. *Desalination and Water Treatment*, 321, 101039.
- Das, A. (2025a). An optimization-based framework for water quality assessment and pollution source apportionment employing GIS and machine learning techniques for smart surface water governance. *Discover Environment*, 3, 117.
- Das, A. (2025b). Pollution source apportionment and application of machine learning approaches in surface water suitability for irrigation based on hydrochemical analysis. *Green Technology, Resilience, and Sustainability*, 5, 4.
- Das, B. K., Paul, S., Mandal, B., Gogoi, P., Paul, L., Saha, A., Johnson, C., Das, A., Ray, A., Roy, S., & Das Gupta, S. (2025). Integrating machine learning models for optimizing ecosystem health assessments through prediction of nitrate-N concentrations in the lower stretch of Ganga River, India.

- Environmental Science and Pollution Research, 32, 4670–4689.
- Dash, S., & Kalamdhad, A. S. (2021). Discussion on the existing methodology of entropy-weights in water quality indexing and proposal for a modification of the expected conflicts. *Environmental science and pollution research*, 28(38), 53983-54001.
- Dhapre, M., Jadhav, S., Das, D., Khan, J., Kim, Y., Chiao, S., & Danielson, T. (2025). A systematic review of machine learning in groundwater monitoring. *Environmental Modelling & Software*, 192, 106549.
- Gugulothu, S., Subba Rao, N., Das, R., Duvva, L. K., & Dhakate, R. (2022). Judging the sources of inferior groundwater quality and health risk problems through intake of groundwater nitrate and fluoride from a rural part of Telangana, India. *Environmental Science and Pollution Research*, 29(32), 49070-49091.
- Karunanidhi, D., Raj, M. R. H., Subramani, T., & Roy, P. D. (2026). Source apportionment, drinking water quality prediction and health risk appraisal of groundwater nitrate using hydrochemistry, machine learning and Monte Carlo simulation: A case study from the Suruliyar River basin, South India. *Environmental Geochemistry and Health*, 48(3), 138.
- Kundu, S., Datta, P., Pal, P., Ghosh, K., Das, A., & Das, B. K. (2025). Unveiling the hidden connections: Using explainable artificial intelligence to assess water quality criteria in nine giant rivers. *Journal of Cleaner Production*, 492, 144861.
- Mukhopadhyay, R., Sarkar, B., Jat, H. S., Sharma, P. C., & Bolan, N. S. (2021). Soil salinity under climate change: Challenges for sustainable agriculture and food security. *Journal of environmental management*, 280, 111736.
- Rajput, V., Samson, R., Yadav, R., Dastager, S. G., Khairnar, K., & Dharne, M. (2022). Metagenomic mining of Indian river confluence reveal functional microbial community with lignocellulolytic potential. *3 Biotech*, 12, 132.
- Romanelli, A., Lima, M. L., Quiroz Londoño, O. M., Martínez, D. E., & Massone, H. E. (2012). A GIS-based assessment of groundwater suitability for irrigation purposes in flat areas of the Wet Pampa Plain, Argentina. *Environmental management*, 50(3), 490-503.
- Rout, A. K., Dixit, S., Tripathy, P. S., Rout, S. S., Parida, S. N., Parida, P. K., Sarkar, D. J., Das, B. K., Singh, A. K., & Behera, B. K. (2024). Metagenomic landscape of sediments of river Ganga reveals microbial diversity, potential plastic and xenobiotic degradation enzymes. *Journal of Hazardous Materials*, 471, 134377.
- Rout, A. K., Tripathy, P. S., Dixit, S., Behera, D. U., Behera, B., Das, B. K., & Behera, B. K. (2023). Unveiling the microbiome landscape: A metagenomic study of bacterial diversity, antibiotic resistance, and virulence factors in the sediments of the River Ganga, India. *Antibiotics*, 12(12), 1735.
- Samson, R., Rajput, V., Shah, M., Yadav, R., Sarode, P., Dastager, S. G., Dharne, M. S., & Khairnar, K. (2020). Deciphering taxonomic and functional diversity of fungi as potential bioindicators within

- confluence stretch of Ganges and Yamuna Rivers, impacted by anthropogenic activities. *Chemosphere*, 252, 126507.
- Satish, N., Anmala, J., Rajitha, K., & Varma, M. R. R. (2024). A stacking ANN ensemble model of ML models for stream water quality prediction of Godavari River Basin, India. *Ecological Informatics*, 80, 102500.
- Savari, M., Yazdanpanah, M., & Rouzaneh, D. (2022). Factors affecting the implementation of soil conservation practices among Iranian farmers. *Scientific Reports*, 12(1), 8396. *Sci Rep*. 2022; 12:8396.
- Sharma, N., Das, B. K., Bhattacharjya, B. K., Chaudhari, A., Behera, B. K., Kumar, A. P., & Chakraborty, H. J. (2024). Metagenomic insights into microbial community, functional annotation, and antibiotic resistance genes in Himalayan Brahmaputra River sediment, India. *Frontiers in Microbiology*, 15, 1426463.
- Shukla, *et al.* (2022). Ganga River sediments of India predominate with aerobic and chemo-heterotrophic bacteria majorly engaged in the degradation of xenobiotic compounds. *Environmental Science and Pollution Research*.
- Tiyasha, T., Tung, T. M., & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, 585, 124670.