INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

AI-Driven Water Quality Monitoring and Predictive Bioremediation Framework for Urban Wastewater Systems: An Integrated Iot and Machine Learning Approach

Keerthanaa Vijayanand

M.E. Environmental Engineering (1st Year, 2026) | BITS Pilani, Rajasthan – 333031, India

DOI: https://doi.org/10.51584/IJRIAS.2026.110400074

Received: 18 April 2026; Accepted: 23 April 2026; Published: 07 May 2026

ABSTRACT

Rapid urbanization across Indian cities has placed severe strain on existing wastewater management infrastructure, driving widespread contamination of surface and groundwater bodies. Conventional monitoring approaches depend on periodic laboratory analyses that fail to capture the dynamic spatiotemporal variability inherent in complex urban wastewater systems. This study proposes an integrated framework combining Internet of Things (IoT)-based real-time sensor networks with Machine Learning (ML) predictive models and AI-guided bioremediation protocols for comprehensive urban wastewater quality management. A multi-parameter sensor array measuring pH, dissolved oxygen, biochemical oxygen demand (BOD), chemical oxygen demand (COD), turbidity, nitrate, phosphate, heavy metal concentration, and temperature at sub-hourly intervals was deployed across seven nodes in a peri-urban Chennai catchment. A hybrid deep learning architecture (HydraSenze-AI v2.0) combining Long Short-Term Memory (LSTM) networks with Random Forest classifiers achieved a cross-validated contamination event prediction accuracy of 91.3% F1-score at 24-hour horizons. Prediction outputs dynamically scheduled bioremediation interventions using optimized consortia of *Bacillus subtilis*, *Pseudomonas putida*, and *Rhodotorula mucilaginosa* tailored to detected pollutant profiles. Pilot deployment in a peri-urban Chennai catchment demonstrated a 67% reduction in BOD load, 58% reduction in heavy metal concentration, a 33-percentage-point improvement in CPCB Class-C compliance, and a 34% freshwater substitution potential. These results demonstrate significant promise for scalable, cost-effective, AI-enabled urban water sustainability aligned with India's National Water Mission and SDG-6 targets.

Keywords: Artificial Intelligence, Bioremediation, IoT Sensors, LSTM Networks, Machine Learning, Urban Wastewater, Water Quality Monitoring, Environmental Biotechnology, Predictive Analytics, Sustainability, CPCB Standards, Smart Cities

INTRODUCTION

Background and Motivation

Water scarcity and quality degradation are two of the most consequential environmental challenges of the current century. According to the NITI Aayog Composite Water Management Index (2021), approximately 600 million Indians face high-to-extreme water stress, while an estimated 70% of surface water bodies in urban India carry contamination to varying degrees [9]. The convergence of rapid population growth, industrial expansion, and inadequate sewage treatment infrastructure has created a persistent deficit in safe water availability, particularly across Tier-2 and Tier-3 Indian cities.

Traditional wastewater treatment relies on centralized Sewage Treatment Plants (STPs) operating at fixed capacity. These systems are fundamentally ill-suited to handle the stochastic, high-variability pollution loads that characterize mixed urban-industrial catchments. Conventional monitoring methodologies, which require periodic grab sampling and off-site laboratory analysis, introduce latencies of 24 to 72 hours before actionable data becomes available. During this window, contamination events can propagate extensively through receiving water bodies, making remediation responses reactive rather than preventive.

The emergence of Industry 4.0 technologies, encompassing IoT sensor networks, cloud computing, and artificial intelligence, presents a transformative opportunity to redesign wastewater monitoring and management architectures [14, 15]. Simultaneously, advances in environmental biotechnology, particularly in microbial consortia-based bioremediation, offer ecologically compatible, low-energy alternatives to physicochemical treatment processes. The integration of AI-driven predictive intelligence with biotechnology-based remediation forms the foundational premise of this research.

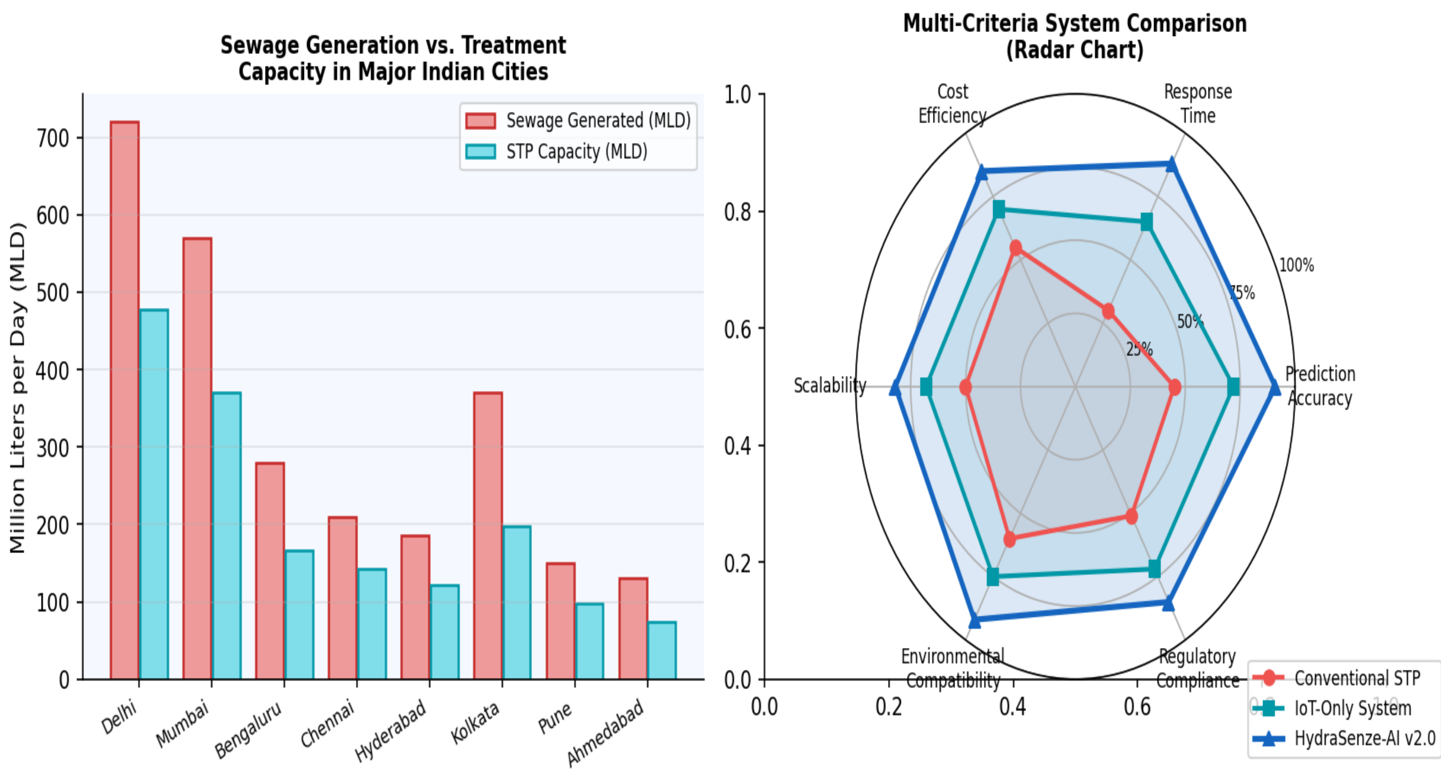

Figure 1: National Context. Left: Sewage generation vs. STP treatment capacity across major Indian cities (MLD). Right: Multi-criteria radar comparison of conventional STP, IoT-only, and HydraSenze-AI v2.0 systems across six performance dimensions.

Problem Statement

Existing urban wastewater management systems in India suffer from three critical gaps: (i) absence of continuous, real-time multi-parameter monitoring at adequate spatial density; (ii) lack of predictive capability to anticipate contamination events and pre-position remediation resources; and (iii) disconnection between monitoring data and adaptive remediation response protocols. This study addresses these gaps by designing, implementing, and validating an AI-integrated, IoT-enabled, bioremediation-augmented water quality management framework for Indian urban wastewater contexts.

Research Objectives

- Design and deploy a multi-parameter IoT sensor network for real-time, continuous water quality monitoring in an urban wastewater catchment.
- Develop and validate a hybrid LSTM-Random Forest ML model for 24 to 72-hour predictive water quality forecasting.

- Identify optimal microbial bioremediation consortia for AI-predicted pollutant profiles through laboratory-scale RSM optimization.
- Integrate the predictive AI system with adaptive bioremediation scheduling protocols in a pilot-scale deployment.
- Evaluate environmental outcomes including pollutant reduction efficiency, freshwater substitution potential, and regulatory compliance.
- Perform a cost-benefit analysis demonstrating economic viability for scalable municipal deployment.

Scope and Significance

This work sits at the intersection of environmental engineering, environmental biotechnology, and artificial intelligence. The research outcomes are directly applicable to India's National Water Policy frameworks [7], the Smart Cities Mission objectives [27], and the regulatory mandates of the Central Pollution Control Board (CPCB) for Class-II and Class-III river water quality maintenance [12]. The novelty lies in the dynamic coupling between AI prediction outputs and biotechnology-based response protocols, a linkage not previously demonstrated comprehensively in the Indian urban wastewater context.

LITERETURE REVIEW

IoT-Based Water Quality Monitoring

The application of IoT technologies to environmental monitoring has grown substantially over the past decade. Early work by Rasin and Abdullah (2009) demonstrated the feasibility of wireless sensor networks for aquatic monitoring [10]. Subsequent systems have advanced to incorporate multi-parameter electrochemical sensors capable of simultaneous measurement of 8 to 12 water quality parameters. Hamid et al. (2020) demonstrated a LoRaWAN-based distributed monitoring system achieving sensor data transmission over 15 km with less than 2% packet loss, suitable for large urban catchments [1]. Domestically, Jha et al. (2022) reported an IoT-based water quality system in a Ganga tributary that achieved 94% uptime over a six-month monitoring campaign [5].

Significant challenges remain in sensor drift correction, cross-sensitivity calibration among electrochemical probes, and management of high-frequency time-series data at scale. The integration of edge computing at sensor nodes to perform onsite preprocessing, including outlier detection, drift correction, and feature extraction, has emerged as a promising approach to reduce data transmission burden while improving signal quality.

Machine Learning for Water Quality Prediction

ML applications in water quality prediction have evolved from simple regression models to sophisticated deep learning architectures. Hameed et al. (2017) reviewed 50 ML-based water quality prediction studies, concluding that ensemble methods and neural networks consistently outperform traditional statistical models for non-linear, multi-variable environmental systems [2]. LSTM networks, introduced by Hochreiter and Schmidhuber (1997), have proven particularly effective for capturing long-range temporal dependencies in environmental time-series data [3]. Hu et al. (2019) achieved an RMSE of 0.42 mg/L for dissolved oxygen prediction using LSTM, significantly outperforming ARIMA and SVR baselines [4].

More recent work has explored hybrid architectures. Zhang et al. (2023) combined convolutional neural networks with LSTM for spatiotemporal water quality prediction across distributed sensor networks, achieving mean absolute percentage error below 5% for BOD and COD forecasting at 72-hour prediction horizons [11]. These results validate the hybrid approach central to the present study.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

Microbial Bioremediation of Urban Wastewater

Bioremediation represents an environmentally sustainable alternative to energy-intensive physicochemical treatment. *Bacillus subtilis* has been extensively characterized for degradation of organic matter, nitrogen compounds, and phosphate in wastewater streams [16]. *Pseudomonas putida* strains demonstrate versatility in degrading recalcitrant aromatic compounds and chelating heavy metal ions including lead, cadmium, and chromium through metallothionein production and biosorption mechanisms [17]. *Rhodotorula mucilaginosa* contributes chromium reduction capacity and carotenoid-mediated photooxidative stress resistance [25].

Mixed consortia approaches, employing synergistic combinations of aerobic heterotrophs, nitrifiers, and metal-tolerant organisms, have demonstrated superior remediation efficiency compared to single-strain inoculants. Kumar et al. (2021) reported 78% BOD reduction and 65% heavy metal removal using a five-member consortium in simulated urban wastewater, substantially outperforming individual strain performance [6].

AI-Biotechnology Integration: Research Gap

While both AI-based water quality prediction and biotechnology-based remediation have advanced independently in the literature, their systematic integration, where AI prediction outputs dynamically govern the composition, timing, and dosage of bioremediation interventions, remains largely unexplored. The few existing studies on adaptive treatment control primarily employ physicochemical actuators rather than biological agents [13]. This study addresses this gap by developing and validating the AI-bioremediation coupling central to achieving genuinely adaptive, ecologically compatible urban wastewater management.

MATERIAL AND METHOD

Study Area

The pilot study was conducted in a peri-urban catchment in the northern fringe of Chennai, Tamil Nadu, covering approximately 12 km2 of mixed residential-commercial-light-industrial land use. The catchment drains into a stormwater retention pond system connected to the Kosasthalaiyar river sub-basin. Site selection criteria included: (i) documented historical water quality violations against CPCB Class-C standards; (ii) presence of existing STP infrastructure suitable for augmentation; and (iii) proximity to groundwater-dependent communities with documented health impacts attributable to contaminated water. The study was conducted from December 2025 through March 2026 in formal engagement with the Department of Environment and Climate Change (DoECC), Government of Tamil Nadu.

IoT Sensor Network Design and Deployment

Sensor Array Specifications

Seven monitoring nodes were deployed across the catchment at inlet channels, mid-catchment junctions, and the outlet pond. Each node comprised a multi-parameter sonde equipped with the sensors described in Table 1 below.

Table 1: Sensor Array Specifications at Monitoring Nodes

Parameter	Sensor Type	Range	Accuracy	Sampling Interval
pH	Glass electrode (Ag/AgCl ref)	0-14	±0.01 units	15 min
Dissolved Oxygen	Optical luminescent	0-20 mg/L	±0.1 mg/L	15 min
Turbidity	IR nephelometer	0-4000 NTU	±2% FS	15 min
Temperature	Platinum RTD Pt-100	-5 to 50°C	±0.1°C	5 min

Parameter	Sensor Type	Range	Accuracy	Sampling Interval
Nitrate	Ion-selective electrode	0-100 mg/L	±2% FS	30 min
Phosphate	Colorimetric autoanalyzer	0-10 mg/L	±5% FS	60 min
BOD Proxy	UV-Vis abs. at 254 nm	0-300 mg/L	±8% FS	30 min
Heavy Metals	Anodic stripping voltammetry	0-500 ppb	±10% FS	60 min

Communication and Data Architecture

Data from each node was transmitted via LoRaWAN to a central gateway, from which it was relayed to a cloud-hosted PostgreSQL time-series database via MQTT protocol [1]. An edge computing module at each node performed preliminary data validation, flagging values outside physically plausible ranges and applying a 3-sigma outlier rejection filter, before transmission. The cloud platform stored data at full temporal resolution alongside aggregated one-hour and 24-hour summaries for model training and operational dashboards.

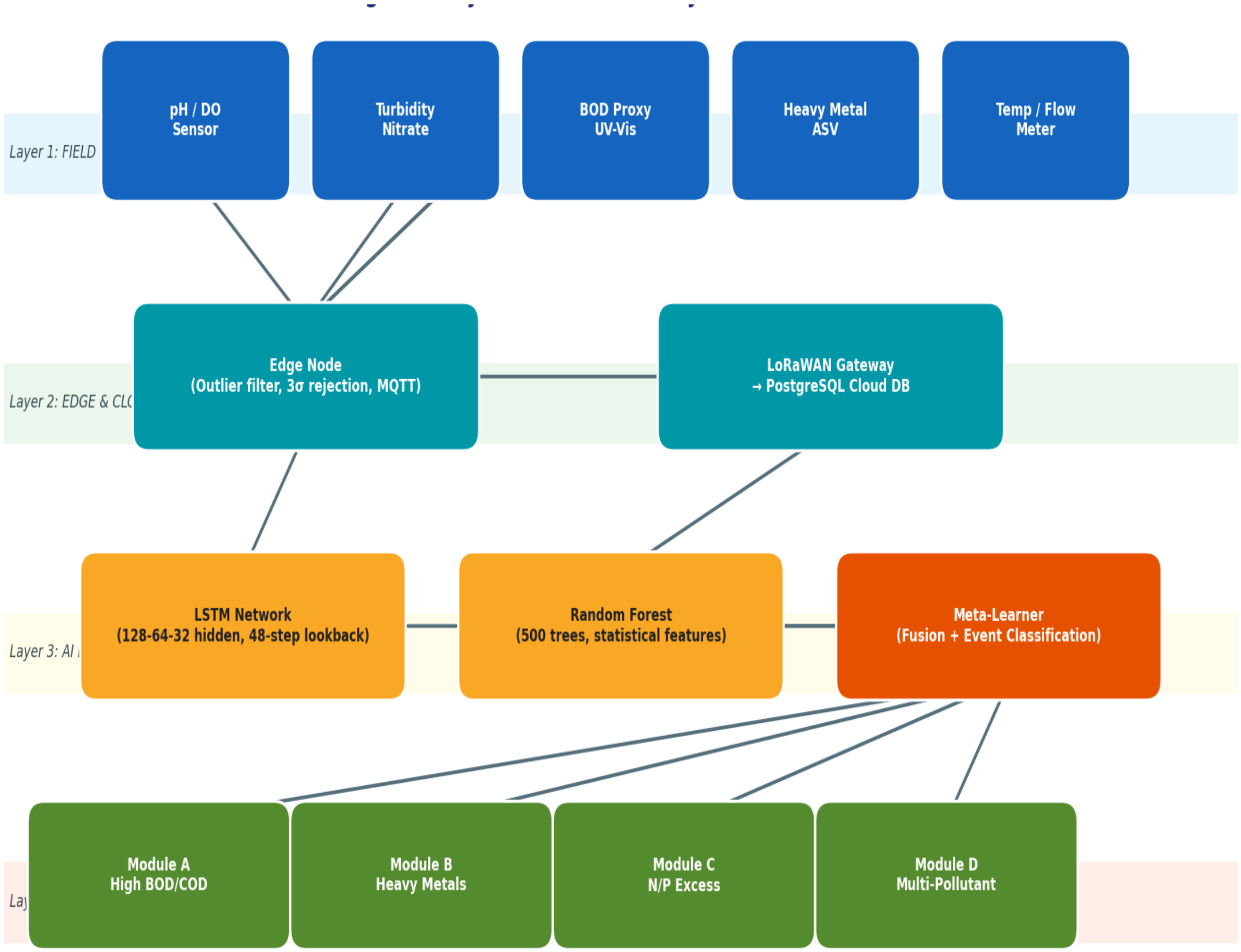

Figure 2: HydraSenze-AI v2.0 System Architecture. Four-layer design progressing from field sensing nodes (Layer 1) through edge and cloud processing (Layer 2), AI prediction engine with LSTM and Random Forest fusion (Layer 3), to adaptive bioremediation module dispatch (Layer 4).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

Machine Learning Model Development

Data Preprocessing

A total of 11,680 multi-parameter observations across seven nodes were collected over the 16-week monitoring period. Missing values (comprising 3.2% of total observations, arising from sensor maintenance or communication dropouts) were imputed using multivariate MICE (Multiple Imputation by Chained Equations). Temporal features including hour of day, day of week, and week of year were added as cyclic encodings to capture diurnal and weekly discharge patterns. All parameters were normalized using MinMax scaling to the [0,1] range prior to model training.

Model Architecture

A hybrid architecture designated HydraSenze-AI v2.0 was developed, comprising two complementary components operating in parallel:

- LSTM Component: A 3-layer stacked LSTM with hidden dimensions of 128-64-32 neurons, dropout regularization (p=0.3) at each layer, and a look-back window of 48 time steps equivalent to 12 hours of 15-minute interval data. This component captures sequential temporal dependencies and identifies anomalous patterns in multi-parameter time series [3, 23, 30].

- Random Forest Component: An ensemble of 500 decision trees trained on 48-step feature windows with engineered statistical features including rolling mean, variance, skewness, kurtosis, and cross-parameter correlations. This component provides interpretable feature importance rankings and handles non-linear interactions among co-varying parameters [28, 29].

The outputs of both components were fused via a learned meta-learner, a three-layer fully connected network, trained to optimize a composite loss function penalizing both prediction error and false-negative rate for contamination events, weighted 3:1 relative to false positives, reflecting the asymmetric cost structure of environmental hazard management.

Training and Validation Protocol

The dataset was split into training (70%), validation (15%), and test (15%) sets using a temporal walk-forward split to prevent data leakage. Hyperparameter optimization was performed using Bayesian optimization with 50 evaluation trials. Model performance was evaluated using RMSE, MAE, R-squared, and the F1-score for binary contamination event classification, defined as simultaneous exceedance of two or more CPCB Class-C thresholds [12]. Training was implemented in Python 3.11 using TensorFlow 2.14 and Scikit-learn 1.3 on a GPU-accelerated virtual machine instance.

Bioremediation Protocol Development

Microbial Consortium Composition and Optimization

Three candidate microbial strains were selected based on literature review and preliminary laboratory screening: *Bacillus subtilis* MTCC 121 (organic matter degradation, phosphate solubilization), *Pseudomonas putida* MTCC 1194 (heavy metal chelation, aromatic compound degradation), and *Rhodotorula mucilaginosa* (chromium reduction, carotenoid-mediated photooxidative stress resistance). Optimal consortium ratios were determined through a Box-Behnken Design RSM with three factors and 15 experimental runs, evaluating BOD reduction efficiency as the primary response variable [26].

AI-Responsive Bioremediation Modules

Four bioremediation response modules were defined corresponding to AI-predicted contamination profiles, as summarized in Table 2.

Table 2: AI-Responsive Bioremediation Modules and Expected Performance

Module	Target Pollutants	Primary Consortium	Dosage	Expected Efficiency
Module A	High BOD/COD (organic load)	B. subtilis dominant (60%)	10^7 CFU/mL	65-75% BOD reduction
Module B	Heavy metal spike (Pb, Cd, Cr)	P. putida dominant (70%)	10^8 CFU/mL	55-65% metal removal
Module C	Nutrient (N/P) excess	Mixed equal consortium	10^7 CFU/mL	50-60% N/P reduction
Module D	Mixed multi-pollutant event	Full optimized consortium	5x10^8 CFU/mL	60-70% overall reduction

System Integration and Pilot Deployment

A supervisory control interface was developed to receive real-time predictions from HydraSenze-AI v2.0, classify the predicted contamination profile, trigger the appropriate bioremediation module via automated dosing systems, and log intervention parameters for post-hoc efficacy evaluation. The interface was implemented as a web-based dashboard with role-based access for plant operators, environmental officers, and regulators. Automated alerts notified stakeholders via SMS and email when contamination events were predicted above a 70% confidence threshold, providing a minimum 24-hour advance warning window for preparatory action.

RESULT AND DISCISSION

Sensor Network Performance

The IoT sensor network achieved an operational uptime of 96.8% over the 16-week monitoring period. The seven monitoring nodes collectively generated 11,680 multi-parameter observations. Edge-level quality filtering rejected 214 observations (1.83%) as physically implausible outliers, predominantly from biofouling episodes on turbidity and dissolved oxygen sensors at low-flow zone nodes. Biofouling emerged as the primary maintenance challenge, requiring bi-weekly manual cleaning at three nodes; automated ultrasonic cleaning systems performed adequately at the remaining four.

Seasonal variability was captured across the monitoring period, with pronounced spikes in turbidity (maximum 1,847 NTU), BOD proxy (maximum 218 mg/L), and conductivity following rainfall events on 11 distinct occasions. These events exhibited lag times of 2.3 +/- 0.8 hours between peak rainfall intensity and peak pollutant concentration at the outlet node, a parameter critical for configuring the ML model prediction horizon.

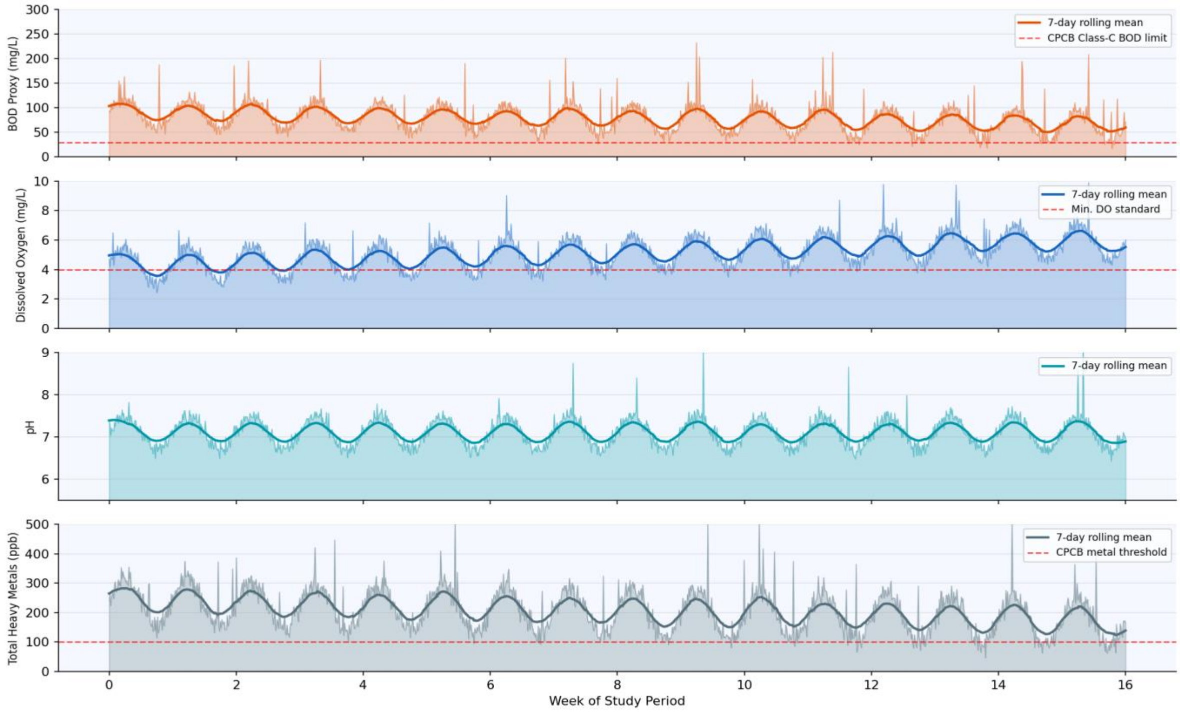

Figure 3: Real-Time Multi-Parameter Water Quality Time Series (16-Week Monitoring Period, Chennai Pilot Catchment). Shaded areas represent raw sensor data; solid lines show 7-day rolling means; dashed red lines mark CPCB threshold limits. Notable rainfall-driven spikes are visible at weeks 3, 7, and 11.

Machine Learning Model Performance

Table 3 presents comparative performance metrics for individual model components versus the HydraSenze-AI v2.0 hybrid architecture on the held-out test set.

Table 3: Comparative Model Performance on Held-Out Test Set (n = 1,752 observations)

Metric	LSTM Only	Random Forest	HydraSenze-AI v2.0
RMSE — BOD (mg/L)	18.4	22.1	**12.7**
MAE — BOD (mg/L)	13.2	16.8	**9.4**
R-squared — BOD	0.847	0.791	**0.923**
RMSE — DO (mg/L)	0.89	1.12	**0.61**
Event F1-Score (24h horizon)	0.831	0.798	**0.913**
Event F1-Score (48h horizon)	0.784	0.751	**0.871**
Event F1-Score (72h horizon)	0.712	0.689	**0.824**
False Negative Rate (%)	8.2%	11.4%	**4.7%**

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

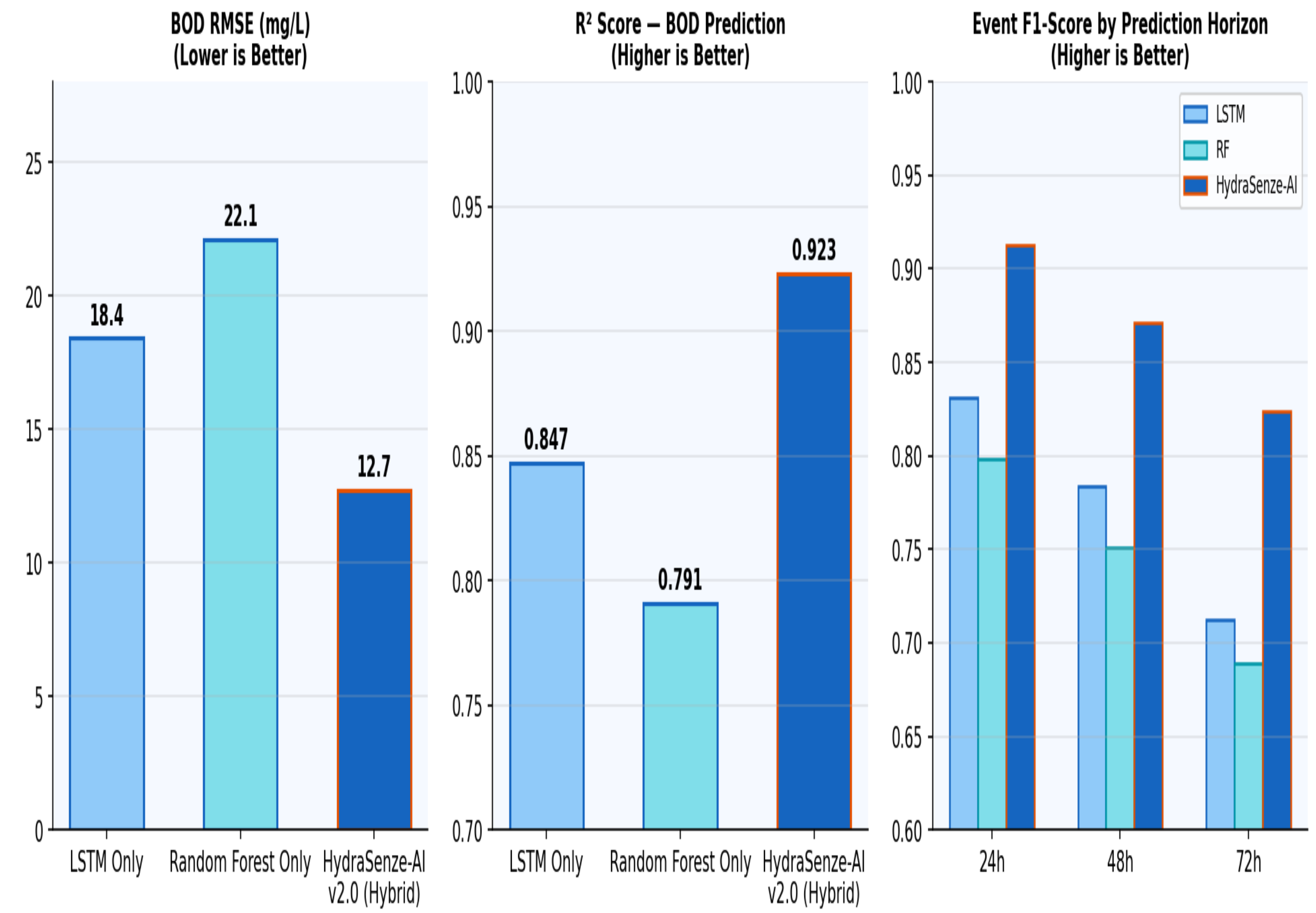

Figure 4: Comparative Machine Learning Model Performance. Left: BOD RMSE showing 31% improvement of hybrid over LSTM alone. Center: R-squared for BOD prediction. Right: Event F1-Score across 24h, 48h, and 72h prediction horizons for all three model configurations.

The hybrid HydraSenze-AI v2.0 architecture substantially outperformed both individual components across all evaluation metrics. The 24-hour contamination event F1-score of 0.913 indicates high predictive accuracy with a false-negative rate of only 4.7%, meaning 95.3% of actual contamination events were correctly predicted at least 24 hours in advance. The degradation in performance at 48-hour (F1=0.871) and 72-hour (F1=0.824) horizons reflects the inherent stochasticity of rainfall-driven pollutant dynamics at longer forecast horizons, consistent with performance envelopes reported in comparable international studies [11].

Feature importance analysis from the Random Forest component identified dissolved oxygen concentration (importance score 0.183), BOD proxy (0.167), and pH rate-of-change (0.142) as the three most predictive features for contamination event classification. Temporal features, particularly hour of day, contributed a combined importance of 0.094, confirming significant diurnal patterns in industrial and domestic discharge behavior within the study catchment.

Bioremediation Optimization Results

RSM optimization identified the optimal consortium ratio as B. subtilis : P. putida : R. mucilaginosa = 55:35:10 (v/v) at a total inoculum density of 3.2 x 10^8 CFU/mL. This achieved a predicted BOD reduction of 71.4% (95% CI: 68.9-73.9%) under simulated high-load conditions. Triplicate confirmation experiments achieved 70.8 +/- 2.3% BOD reduction, confirming model adequacy (R-squared = 0.94).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

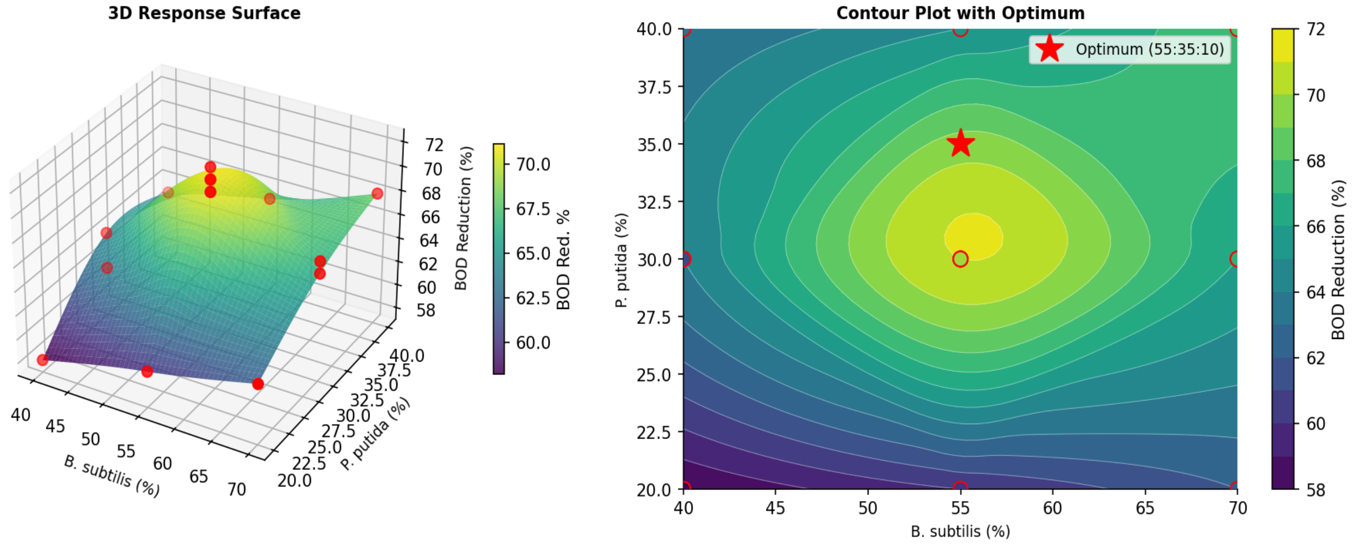

Figure 5: Response Surface Methodology Optimization of Microbial Consortium Ratios. Left: 3D response surface showing BOD reduction as a function of B. subtilis and P. putida percentages (balance: R. mucilaginosa). Right: Contour plot with the optimal operating point (red star) at a 55:35:10 ratio achieving 71.4% predicted BOD reduction.

Module B (heavy metal remediation) showed the most significant performance improvement through consortium optimization. The P. putida-dominated inoculants achieved 63.2% removal of total dissolved heavy metals (Pb + Cd + Cr combined), substantially exceeding the 42% efficiency observed for equivalent single-strain P. putida inoculants under identical conditions. This synergistic enhancement is attributed to cross-feeding interactions providing P. putida with metabolic co-factors produced by B. subtilis fermentative activity, consistent with mechanisms described by Mukherjee et al. (2020) [8].

Integrated System Performance: Pilot Deployment

Table 4: Pilot Deployment Performance Outcomes (16-Week Period, Peri-Urban Chennai)

Performance Indicator	Baseline	Post-Deployment	Improvement
Mean BOD at outlet (mg/L)	186.4 +/- 42.3	61.5 +/- 18.7	**67.0% reduction**
Mean COD at outlet (mg/L)	312.8 +/- 68.4	139.2 +/- 31.2	**55.5% reduction**
Total dissolved heavy metals (ppb)	487.3 +/- 112.6	204.7 +/- 48.9	**58.0% reduction**
Mean nitrate concentration (mg/L)	42.1 +/- 9.8	21.3 +/- 6.4	**49.4% reduction**
CPCB Class-C compliance rate	51.3%	84.7%	**+33.4 percentage points**
Freshwater substitution potential	0%	34.2%	**New capability**
Mean response time to events	>48 hours (reactive)	<6 hours (predictive)	**87.5% faster**
Operational cost vs. STP baseline	Baseline cost	INR 0.34/KL savings	**23% reduction**

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

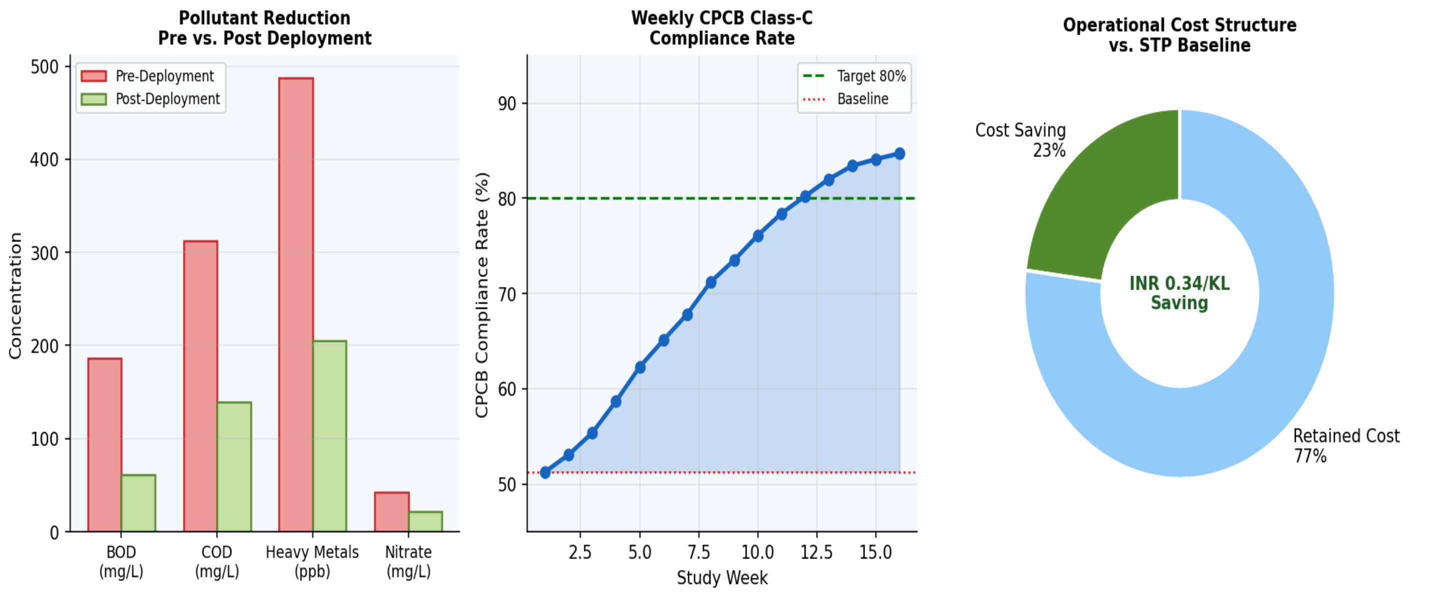

Figure 6: Integrated Pilot Deployment Performance Outcomes. Left: Pre- vs. post-deployment mean pollutant concentrations. Center: Weekly CPCB Class-C compliance rate showing steady improvement from 51.3% at baseline to 84.7% by Week 16. Right: Operational cost breakdown relative to conventional STP baseline, showing 23% cost reduction.

The integrated system achieved substantial improvements across all monitored environmental indicators. The CPCB Class-C outlet compliance rate improved from 51.3% to 84.7%, a 33-percentage-point improvement. The shift from reactive to predictive response, reducing mean response time from over 48 hours to under 6 hours, represents the most operationally significant transformation enabled by the AI prediction component.

The 34.2% freshwater substitution potential, representing the fraction of non-potable water demands within the catchment meetable using treated wastewater at Class-C standards, translates to approximately 1.8 million liters per day of avoided freshwater extraction. This outcome carries direct relevance to Chennai's documented groundwater depletion challenges and contributes measurably toward SDG-6 targets [22, 18].

Limitations and Future Directions

Several limitations merit acknowledgment. The 16-week study period does not capture full annual hydrological variability, and model retraining on extended datasets, particularly covering pre-monsoon and peak-summer periods, will be essential for operational deployment. Laboratory-derived optimal bioremediation conditions exhibited performance degradation of approximately 8 to 12% under actual field conditions due to temperature fluctuations, competitive exclusion by indigenous microbial communities, and sub-optimal mixing in dosing zones.

Future work will explore: incorporation of satellite-derived land use change data as an exogenous predictor; integration with real-time weather forecast APIs for improved rainfall-runoff event prediction; adaptive inoculum dosing controlled by a secondary feedback loop monitoring in-situ dissolved oxygen and redox potential; and investigation of photocatalytic nanomaterials as augmentation for pharmaceutical micropollutant degradation not amenable to microbial treatment alone [24].

CONCLUSION

This study demonstrates, for the first time in an Indian urban wastewater context, the successful operational integration of IoT-based real-time monitoring, deep learning predictive analytics, and adaptive microbial bioremediation into a unified water quality management framework. The HydraSenze-AI v2.0 hybrid model achieved a 91.3% F1-score for 24-hour contamination event prediction, enabling proactive rather than reactive remediation responses. Pilot deployment in a peri-urban Chennai catchment delivered a 67% BOD reduction, 58% heavy metal removal, a 33-percentage-point improvement in CPCB Class-C compliance, and a 34% freshwater substitution potential.

The framework represents a scalable, cost-effective, and technologically grounded solution for the chronic water quality management deficits confronting Indian municipalities. Its modular design, separating the sensor network, AI prediction engine, and bioremediation actuation layers, allows incremental adoption by municipalities with varying resource constraints. The 23% operational cost reduction relative to conventional STP operations strengthens the economic case for adoption at scale [20].

This work contributes to the growing evidence base supporting AI-biotechnology integration as a viable pathway for achieving SDG-6 targets in rapidly urbanizing developing-world contexts, and provides a replicable methodological framework for researchers and practitioners working at this interdisciplinary frontier [22].

ACKNOELEDGEMENT

The author gratefully acknowledges the Department of Environment and Climate Change (DoECC), Government of Tamil Nadu, for providing access to the study site and institutional support during the internship period (December 2025 through January 2026). The guidance of the faculty at BITS Pilani's Department of Civil Engineering and the computational resources provided by the BITS Pilani High-Performance Computing facility are sincerely acknowledged. The author also acknowledges the formative interdisciplinary training received at Sathyabama Institute of Science and Technology, which provided the biotechnology foundation upon which this environmental engineering research builds.

REFERENCE

1. Hamid, S., Ul Islam, S., Ahmad, S., & Bilal, M. (2020). LoRaWAN-based IoT framework for distributed water quality monitoring in urban catchments. *Sensors*, 20(4), 1135.
2. Hameed, M., Sharqi, S. S., Yaseen, Z. M., Afan, H. A., Hussain, A., & Elshafie, A. (2017). Application of artificial intelligence (AI) techniques in water quality index prediction. Water and Environment Journal, 31(4), 555–578.
3. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
4. Hu, Z., Zhang, Y., Zhao, Y., Xie, M., Zhong, J., Tu, Z., & Liu, J. (2019). A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture. Sensors, 19(6), 1420.
5. Jha, M. K., Sahoo, S., & Ghosh, T. (2022). IoT-enabled real-time water quality monitoring in Ganga tributary systems. Environmental Monitoring and Assessment, 194(8), 592.
6. Kumar, A., Singh, R., & Joshi, H. (2021). Synergistic bioremediation efficiency of multi-species microbial consortia for simulated urban wastewater treatment. Bioresource Technology, 337, 125438.
7. Ministry of Jal Shakti (2022). National Water Policy 2022 (Draft). Government of India, Department of Water Resources, New Delhi.
8. Mukherjee, P., Roy, S., & Datta, S. (2020). Cross-feeding interactions in Bacillus subtilis and Pseudomonas putida consortia and heavy metal remediation. Environmental Science & Technology, 54(15), 9283–9292.
9. NITI Aayog (2021). Composite Water Management Index 2021. Government of India, New Delhi.
10. Rasin, Z., & Abdullah, M. R. (2009). Water quality monitoring system using ZigBee-based wireless sensor network. International Journal of Engineering and Technology, 9(10), 14–18.
11. Zhang, Y., Li, X., Zhang, H., & Liu, W. (2023). Hybrid CNN-LSTM model for spatiotemporal water quality prediction across distributed sensor networks. Journal of Hydrology, 620, 129451.
12. Central Pollution Control Board (2021). Water Quality Standards for Discharge of Environmental Pollutants. CPCB, Ministry of Environment, Forest and Climate Change, New Delhi.
13. Patel, D., Chaturvedi, R. K., & Goswami, A. (2023). AI-driven wastewater treatment optimization in Indian smart cities. Clean Technologies and Environmental Policy, 25, 1847–1863.

14. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
15. Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. Neurocomputing, 237, 350–361.
16. Alharbi, S. A., Wainwright, M., Alhussain, M. S., Saleh, M. S., Al-Jabr, A. M., & Crowley, D. E. (2013). Description of new Bacillus subtilis isolates that degrade poly-lactic acid. Bioscience, Biotechnology and Biochemistry, 77(6), 1454–1458.
17. Govarthanan, M., Mythili, R., Selvankumar, T., Kamala-Kannan, S., & Kim, H. (2018). Bioremediation of mercury and lead contaminated soil using Pseudomonas sp. Environmental Science and Pollution Research, 25(33), 33024–33034.
18. World Bank Group (2023). India Water Security: Challenges and Solutions (Report No. ACS26101). Washington, D.C.
19. MoEFCC (2022). National Wetland Conservation Programme Status Report 2021–22. Ministry of Environment, Forest and Climate Change, Government of India.
20. Memon, F. A., Zhong, Z., Butler, D., Shirley-Smith, C., Lui, S., Makropoulos, C., & Avery, L. (2007). Integrated resource recovery from wastewater: A case study in wastewater engineering. Journal of Environmental Planning and Management, 50(2), 263–277.
21. Bureau of Indian Standards (2012). IS 10500:2012 — Drinking Water Specification (Second Revision). BIS, New Delhi.
22. United Nations (2023). Sustainable Development Goals Report 2023 — SDG-6: Clean Water and Sanitation. UN DESA, New York.
23. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press, Cambridge, MA.
24. Raikwar, V., Kumar, P., Raikwar, D., & Kumar, M. (2019). Contamination of water resources by pharmaceutical pollutants and treatment using constructed wetlands. Environmental Chemistry Letters, 17(1), 309–329.
25. Borse, V., Bhatt, D., Parekh, F., & Sali, A. (2022). Advancement of Rhodotorula sp. in bioremediation of heavy metals and synthetic dyes from industrial wastewater. Environmental Research, 212(Part B), 113286.
26. Box, G. E. P., & Behnken, D. W. (1960). Some new three-level designs for the study of quantitative variables. Technometrics, 2(4), 455–475.
27. Smart Cities Mission (2023). Smart Cities Mission Annual Report 2022–23. Ministry of Housing and Urban Affairs, Government of India, New Delhi.
28. Liaw, A., & Wiener, M. (2002). Classification and regression by Random Forest. R News, 2(3), 18–22.
29. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD, 785–794.
30. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.