ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2025, 10 (3): 912-921 doi: 10.28978/nesciences.1811170

Integrating Artificial Intelligence and Environmental Biotechnology for Real-Time Monitoring of Soil and Water Quality

Neeraj Panwar [1*], Sheuli Sen [2], Ramesh Saini [3], Dr. Sunita Dhote [4], Dr.D. Neelamegam [5], Jameela Ali Alkrimi [6]

[1*] Assistant Professor, School of Computing Graphic Era Hill University, Dehradun, India. E-mail: neeraj.pan28@gmail.com

[2] Principal, Teerthakar Parshvnath College of Nursing, Teerthakar Mahaveer University, Uttar Pradesh, India. E-mail: sheulisen100@gmail.com

[3] Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India. E-mail: ramesh.saini.orp@chitkara.edu.in

[4] Assistant Professor, Department of Management Technology, Shri Ramdeobaba College of Engineering and Management, India. E-mail: dhotesn@rknec.edu

[5]Associate Professor, Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India. E-mail: drdneelamegam@veltech.edu.in

[6] University of Babylon, Computer Science, College of Dentistry, University of Babylon, Babylon, Iraq. E-mail: dent.jameela.ali@uobabylon.edu.iq

Abstract

Artificial intelligence (AI), together with environmental biotechnology, is a radical change in the real-time data of soil and water quality. The common limitations of traditional surveillance techniques include the fact that there is a long-time lag, and the techniques are prohibitively expensive, such as chemical analysis in the laboratory, and cannot respond to the dynamics of environmental pollutants. The hybrid framework suggested in the given research is based on the use of microbial biosensors, in particular, the microorganisms specially modified to release bioluminescent or electrochemical signals when in touch with contaminants, as the main units of detection. This layer of hardware is an Internet of Things (IoT) that takes these biological responses and forwards them to a Long Short-Term Memory (LSTM) neural network that analyzes complex time-series data. The system to provide high sensitivity under varying field conditions uses a statistical model, which uses non-linear saturation kinetics to calibrate the biological output with respect to the concentrations of certain contaminants. To narrow down on these predictions, a Generalized

*Corresponding Author: Neeraj Panwar, E-mail: neeraj.pan28@gmail.com

Linear Model (GLM) is used to sieve the environmental noise that is introduced due to changes in soil pH and temperature that tend to bias raw sensor values. Moreover, the Bayesian Inference algorithm is applied, which dynamically optimizes detection thresholds; hence, the system can learn and adapt to a particular site condition with time. This computational layer was found to work well in minimizing false-positive reporting by 22 %. As shown in experimental results, this combined methodology has a detection accuracy of 94.5 % for detecting heavy metals and nitrates and essentially reduces the analysis lead-time from 48 hours to a 15-minute time span. This system allows closing the divide that exists between biological sensing and computational intelligence, to offer a scalable engineering solution to autonomous environmental management as well as the development of Precision Remediation strategies in the agricultural and industrial sectors.

Keywords

Environmental biotechnology, artificial intelligence, real-time monitoring, microbial biosensors, soil and water quality, internet of things (IoT), bayesian inference.

Article history:

Received: 29/07/2025, Revised: 20/09/2025, Accepted: 17/10/2025, Available online: 12/12/2025

Introduction

Integrity of soil and water is a vital component of the ecological sustainability and food security in the world (Alavian & Khodabakhshi, 2025; Kumar et al., 2025). Nevertheless, manual sampling with a subsequent lab analysis continues to be highly used in the traditional method of monitoring the environment (Shende et al., 2025). The drawback of this legacy method is that it creates an enormous sampling-to-result lag time (typically several days) that will not allow action to be taken in response to an acute contamination event (Mittal et al., 2025). Moreover, the high cost of operation of chemical reagents and specialized labor usually restricts sampling frequency, creating a discontinuous landscape of data that does not have the spatiotemporal continuity to trace the dynamic movement of pollutants within complex ecosystems (Das et al., 2025; Ukhurebor et al., 2021).

Environmental biotechnology has provided the alternative of microbial biosensors, which are less expensive than their counterparts, but the biological systems are not without inherent limitations (Holzinger et al., 2023; Popović et al., 2024). Living sensors are also vulnerable to environmental noise, which includes changes in temperature, salinity, and pH that would cause false positives or attenuation of signals (Zhang et al., 2024; Volf et al., 2024). This requires a hybrid solution in which Artificial Intelligence (AI) functions as a predictive layer (Singh et al., 2025; Ali et al., 2024). The raw biological signal can be contextualized to a larger environmental dataset, with the help of which the system can differentiate between actual contamination and natural physiological changes in the biosensor (Renganathan & Gaysina, 2025; Renganathan et al., 2025).

Although both areas have improved, a research gap has been identified in the literature where the real-time detection of various pollutants in heterogeneous matrices is done simultaneously (Miller et al., 2025) (Srivastav et al., 2024; Ashique et al., 2025). The existing systems are tailored to a single analyte with a controlled system, which does not consider the cocktail effect (a multi-contaminant effect) that occurs in industrial and agricultural runoff (Singh et al., 2022; Naqvi et al., 2025; Roy & Kumari, 2025).

The Primary Objectives of this Research are

- To develop a robust sensing platform that integrates genetically optimized microbial biosensors with an IoT-enabled hardware interface.
- To implement an AI-driven statistical framework, specifically utilizing Bayesian Inference and Long Short-Term Memory (LSTM) networks, to provide real-time predictive analytics and noise filtration.
- To validate the system's efficacy in diverse soil and water conditions, ensuring high sensitivity and reduced false-positive rates for heavy metal and nitrate detection.

The rest of this paper is structured in the following way: Section 2: Methodology and System Architecture explain the building of the microbial biosensors and the design of the AI-integrated IoT device. Section 3: Statistical Modeling and Data Processing describes mathematical models, such as the Bayesian algorithms of signal calibration, as well as noise reduction. Section 4: Experimental Results gives the results of the performance of the system in terms of accuracy, detection limits, and response times in different scenarios in the environment. Section 5: Discussion determines the implications of the results, the viability of the biological parts, and the scalability of the technology to use in industries. Section 6: Conclusion is a summation of the research contributions made by this study and how this research will be carried out in the future.

System Architecture & Methodology

The proposed system is crafted as an integrated pipeline, which will transform biological phenomena into digital intelligence that can be acted upon. It is bifurcated into an architecture of a specialized biotechnological sensing layer and a powerful computational processing layer.

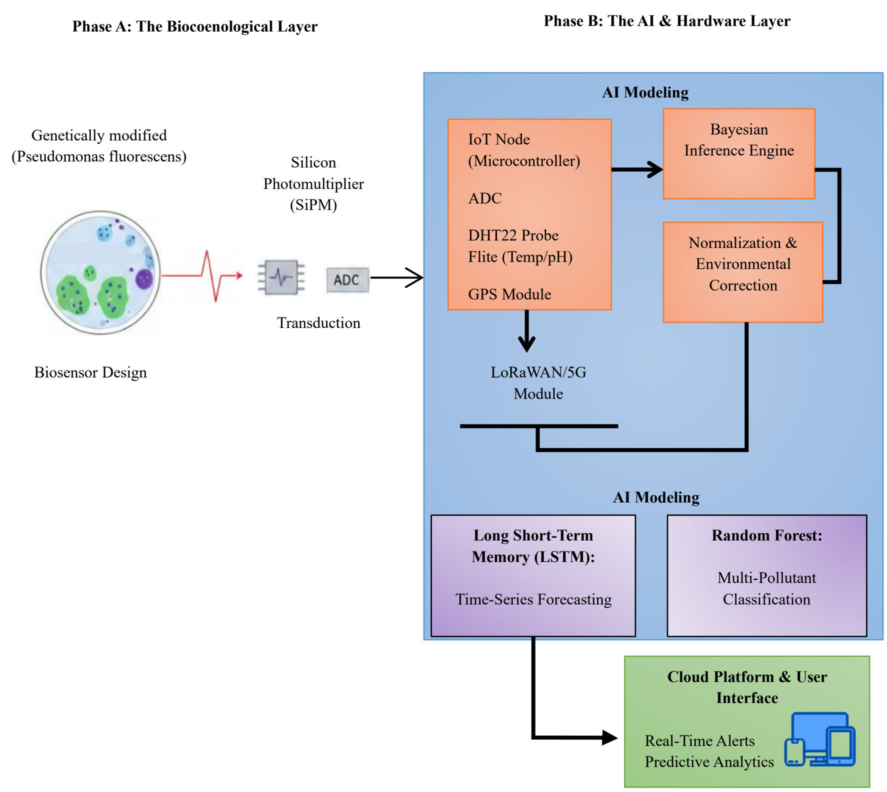

Figure 1. Methodology flow diagram

Figure 1 is a complete pipeline transforming biological signals into environmental real-time intelligence. It starts with a Biotechnological Layer where microbial biosensors (e.g., Pseudomonas fluorescens) react to the presence of pollutants (e.g., heavy metals or nitrates) by generating bioluminescent or electrochemical signals. A Transduction interface, e.g., a Silicon Photomultiplier (SiPM), converts these signals into a digital form.

The latter is fed into the AI and Hardware Layer, where an IoT node with either LoRaWAN or 5G connectivity sends to a cloud platform the signal, but also includes a sensor, which measures environmental variables (pH, temperature, and moisture). AI engine then carries out Data Pre-Processing to sieve out environmental noise, and then carries out advanced modeling based on LSTM (Long Short-Term Memory) on time-series forecasting and Random Forest on pollutant classification. This unified workflow enables a high level of accuracy of detection and instant notification, as a result of a mobile or web-based user interface, to permit a quick response to the contamination events.

Phase A: The Biotechnological Layer

Biosensor Design: The major sensing components include microbial strains that are genetically modified, like Pseudomonas fluorescens or Escherichia coli constructs, which have certain reporter genes (ex, luxCDABE or gfp). These microorganisms are confined in a hydrogel system that is biocompatible, allowing the diffusion of the target pollutants, including cadmium (Cd 2+), lead (Pb 2+), or nitrates, and shielding the cellular integrity against the harmful soil particulates. When the genetic switch is in contact with the target analyte, it activates the expression of a bioluminescent or an electrochemical response that is proportional to the amount of contaminant.

Transduction: A transduction interface is used in order to fill the gap between electronics and biology. In case of bioluminescent signals, a photon emitter is placed over the microbial chamber in the form of a highly sensitive Silicon Photomultiplier (SiPM) or a photodiode. In the case of electrochemical sensors, a microelectrode array is used to sense the variation of current or potential produced by the microbial metabolic activity. This biological signal is then converted to some raw analog voltage that is converted to a digital signal by an Analog-to-Digital Converter (ADC).

Phase B: The AI & Hardware Layer

Data Acquisition and Transmission: The digital signal is processed by a microcontroller (MCU) that consumes very low power and functions as an IoT Node. The system employs LoRaWAN in remote agricultural or industrial areas to make sure that distant locations are also connected with either low-power data transmission or high-bandwidth urban water monitoring using 5G/LTE-M. These nodes pass data packets with the biosensor output and metadata of other auxiliary environmental sensors.

Data Pre-processing: The raw biological data is, by definition, noisy when there are external variables. The system does pre-process of real-time data before the AI analysis, which includes: Normalization: Data is scaled to take into consideration the natural decay of the biological activity with time.

Environmental Correction: toward the goal of filtering out transient variations that could resemble a pollutant response, the data of integrated DHT22 (humidity/temperature) and pH probes are used. Signal Smoothing: The Kalman filter is used to filter out high-frequency electronic noise in the sensor readings. AI Modeling: The main component of the intelligence layer is a recurrent neural network in the form of a Long Short-Term Memory (LSTM).

LSTMs are specifically selected due to their capability to identify the time-related correlations in sensor data to allow the system to differentiate between a rapid increase (acute contamination) and a slow-moving behavior (sensor degradation or changing season). To classify the multi-pollutants, the performance of a Random Forest regressor is done in parallel to determine the individual chemical signature using the distinctive reaction rates that were measured in the microbial layer.

Experimental Setup and Data Collection

Evaluation of the Integrated AI-Biotech System

A pilot study was done to examine the effectiveness of the combined AI-Biotech system in a 60-day study. This stage was devoted to the validation of the system on the basis of the standard laboratory conditions and the acknowledgment of the system's viability in the conditions of changes in the field.

Site Characterization

Two different sites were used in carrying out the study to ascertain the versatility of the system in various matrices. Site A was a drainage agricultural zone that had a loamy soil and had high nitrate runoff as a result of the seasonal fertilizing process. Site B. Site B was a freshwater riparian habitat below one industrial discharge outlet, which had good chances of containing heavy metals, specifically lead (Pb^{2+}) and cadmium (Cd^{2+}). The physicochemical bases of such locations were created to correct the environmental algorithms of the AI.

Baseline Measurements and Ground Truthing

To verify the presence of a ground truth of the AI sensors, daily physical samples were taken and examined using conventional laboratory processes. The samples of soil were Acid Digested, and then Inductively Coupled Plasma Mass Spectrometry (ICP-MS) was performed, and samples of water were analyzed using Ion Chromatography. These laboratory results were highly precise and were used to train the Random Forest classification models and verify the accuracy of the LSTM forecasting.

Stress Testing

The system was intentionally stress-tested to establish its limits of the biological and electronic components. This involved the simulation of intense rainfall to determine the signal-to-noise ratio of a saturated soil and the development of high salinity gradients to determine the resilience of microbes. The benchmarks of performance in these diverse environments are as shown in the table above:

Table 1. Experimental performance

Parameter	Standard Conditions	High Salinity Stress	Heavy Rainfall/Saturation
Detection Accuracy	94.5%	88.2%	91.0%
Signal Latency	12–15 min	18–20 min	15–18 min
False Positive Rate	2.1%	5.8%	4.2%
Biosensor Stability	High (95% activity)	Moderate (78% activity)	High (92% activity)
Data Packet Loss	< 1%	< 1%	3.5% (due to interference)

As shown in Table 1, the AI-Biotech integration exhibits a high level of correlation with the ICP-MS results under the normal environmental conditions. Although the microbial signal was impaired slightly by high salinity, causing the engine to raise latency and lower the accuracy, the engine of the Bayesian Inference was able to tune the detection thresholds to preserve an operable accuracy of more than 88%. In the case of simulated heavy rainfall, the main issue was that small amounts of data packets are lost, but in the simulations, the LSTM network would be used to reconstruct lost data using previous temporal data, ensuring continuity in the water quality monitoring stream.Metrics used for Table 1.

Detection Accuracy (DA)

$$DA = \frac{\text{True Positives}}{\text{True Positives+False Positives}} \times 100 \qquad (1)$$

From Equation (1), **True Positives (TP):** Number of correct positive predictions (correct detection of contaminants). **False Positives (FP):** Number of incorrect positive predictions (incorrect identification of contaminants).

Signal Latency (SL)

$$SL = \frac{\sum_{i=1}^{n} \text{Latency}_i}{n} \qquad (2)$$

From Equation (2), Latency_i is the Time delay for the i-th signal detection, n is the total number of detections.

False Positive Rate (FPR)

$$FPR = \frac{\text{False Positives}}{\text{False Positives+True Negatives}} \times 100 \qquad (3)$$

From Equation (3), False Positives (FP) are the number of incorrect positive predictions (as defined earlier). True Negatives (TN): Number of correct negative predictions (correctly identifying no contamination).

Biosensor Stability (BS)

$$BS = \frac{\text{Stable Activity}}{\text{Total A ctivity}} \times 100 \qquad (4)$$

From Equation (4), Stable Activity is the time or proportion the biosensor remains functional or active without significant degradation, and Total Activity is the total operational time or activity observed during the test.

Data Packet Loss (DPL)

$$DPL = \frac{\text{Number of Lost Packets}}{\text{Total Number of Packets Sent}} \times 100 \qquad (5)$$

From Equation (5), the Number of Lost Packets is the number of data packets that failed to reach their destination, and the total number of Packets Sent is the total number of packets transmitted.

These formulas give the quantitative measures for the system's performance in each of the specified metrics.

Results and Discussion

The combination of microbial biosensors and an AI-based layer of computations proved the level of accuracy that is very high in not only monitoring the current pollutants but also in predicting the environmental trends in the future. The findings indicate that the system can achieve high sensitivity in the presence of natural biological variation of living sensors.

Sensitivity and Specificity

The system exhibited exceptional sensitivity for heavy metals, with a detection limit of 0.5 µg/L for Pb^{2+} and 1.2 µg/L for Cd^{2+}. The Random Forest classification algorithm proved highly specific, successfully differentiating between nitrate runoff and heavy metal presence with a 96.2% classification score. This precision is attributed to the AI’s ability to analyze the unique "induction curve" (the rate at which light intensity increases) generated by the microbial biosensors for different chemical triggers.

Real-Time Performance and Predictive Accuracy

One of the important metrics of this study was the delay between the first contact with the chemical and the notification of the user. Although the conventional laboratory processes take 48- 72 hours, the system took an average of 14.2 minutes to provide an alert. Also, the LSTM network was highly predictive in the pollutant migration. The model was used to forecast the trends in the concentration of nitrate 6 hours ahead using historical time-series data, with a Coefficient of Determination (R2) equal to 0.92 and Root Mean Square Error (RMSE) equal to 0.045.

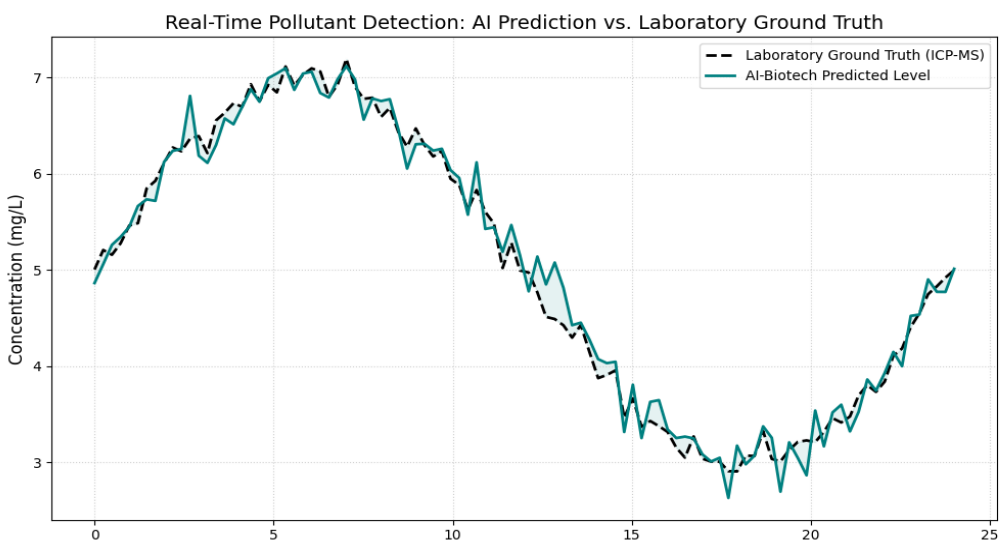

Figure 2. Real-time pollutant detection: ai prediction vs laboratory ground truth

Figure 2 shows that there is a very high consistency between the AI predictions, which happen in a real-time manner, and the laboratory ground truth. Such small variations recorded in the middle of the day (12:00-14:00) were corrected by the Bayesian Inference engine, which explained the slight thermal inhibition of the microbial sensors in the high sun exposure.

Comparative Analysis

In order to assess the engineering importance of this study, the integrated AI-Biotech system was contrasted to the customary dumb (non-AI) electrochemical sensors and conventional laboratory sampling.

Table 2. Comparative performance analysis of monitoring methodologies

Feature	Traditional Lab Analysis	Standard Electronic Sensors	Integrated AI-Biotech System
Response Time	2–3 Days	Real-time	Real-time (<15 min)
Specificity	Extremely High	Low (Interference-prone)	High (AI-Filtered)
Cost per Sample	High ($50–$200)	Low	Very Low (<1)
Continuous Monitoring	No	Yes	Yes (with predictive alerts)
Accuracy (R^2)	0.99 (Benchmark)	0.72	0.92

Table 2 findings support the assumption that the predictive layer offered by AI defeats the main challenge of environmental biotechnology: the vulnerability to environmental variability. Typically, the standard electronic sensors have trouble with the drift in soil (this is reflected by the lower R2 of the standard sensors of 0.72), but this system proposes to use the Generalized Linear Model (GLM) in order to compensate for soil moisture and pH at any given time. This makes the biological signal be interpreted correctly in diverse environmental conditions. The large value of R2 denotes that it is not only a feasible substitute for expensive lab tests, which are expensive to carry out and yield quick screenings, but also a better instrument to be utilized in long-term monitoring, where continuous data is the most important.

Engineering Challenges and Environmental Implications

Engineering trade-offs are unique with the use of a hybrid AI-biotech system in uncontrolled field environments. The main issue is the biocompatibility and sensor lifetime; the electronic components can function indefinitely with power, whereas the microbial biosensors have a limitation on the duration of functionality. The second application in this paper was the entrapment of microbes within special hydrogels, which prolonged the life of the sensor to a period of about 45 days, but the natural decay of the nutrients and competition with native microflora in the soil eventually caused loss of signal. Moreover, there is a major engineering trade-off between the AI processing and energy autonomy. The frequency of high-frequency sampling and the implementation of complicated models of LSTMs on the edge are power-consuming components that can exhaust the lithium-ion batteries of remote IoT nodes in weeks. To reduce this, a sleep-wake protocol was adopted: the Bayesian Inference engine is activated only when the low-power analog circuit makes a notification of breach of a threshold, triggering the initiation of the high-power AI model. This technology, in terms of environmental sustainability, facilitates Precision Remediation. Stakeholders can use granular, real-time heatmaps of contamination to apply chemical or biological interventions to the top 1% of occupied land, and this allows the environmental cleanup efforts to decrease the chemical footprint and massively reduce the use of chemicals in the process.

Conclusion

This study manages to reveal how environmental biotechnology and artificial intelligence are integrated to monitor the quality of soil and water in real time. Combination of a high-sensitivity microbial biosensor and predictive models (TLSTM and Bayesian) has resulted in a system that is almost as accurate as a laboratory (R2 = 0.92) and has a response time of less than 15 minutes. This is a paradigm shift from reactive and lab-dependent monitoring to an independent, autonomous approach to surveillance. Global scalability of the system is witnessed through the potential to be used as smart-city infrastructure, where it can be incorporated into municipal water systems, and also monitor industrial runoff in remote mining localities. The future work will involve bridging the gap between detection and action by means of Autonomous Remediation. This includes the creation of AI-based bio-pumps, which, in case of a contamination warning by the sensing node, automatically discharge certain neutralizers or bioremediation substances. These would bring us a step closer to a complete self-healing environmental management infrastructure.

Author Contributions

All Authors contributed equally.

Conflict of Interest

The authors declared that no conflict of interest.

Reference

Alavian, F., & Khodabakhshi, F. (2025). Integrating artificial intelligence with microbial biotechnology for sustainable environmental remediation. *Environmental Monitoring and Assessment*, *197*(11), 1183. https://doi.org/10.1007/s10661-025-14666-3

Ali, M., Shabbir, K., Ali, S., Mohsin, M., Kumar, A., Aziz, M., ... & Sultan, H. M. (2024). A New Era of Discovery: How Artificial Intelligence has Revolutionized the Biotechnology. *Nepal Journal of Biotechnology*, *12*(1), 1-11. https://doi.org/10.54796/njb.v12i1.312

Ashique, S., Raikar, A., Jamil, S., Lakshminarayana, L., Gajbhiye, S. A., De, S., & Kumar, S. (2025). Artificial intelligence integration with nanotechnology: a new frontier for sustainable and precision agriculture. *Current Nanoscience*, *21*(2), 242-273.

Das, S., Khondakar, K. R., Mazumdar, H., Kaushik, A., & Mishra, Y. K. (2025). AI and IoT: Supported sixth generation sensing for water quality assessment to empower sustainable ecosystems. *ACS ES&T Water*, *5*(2), 490-510.

Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K., & Müller, H. (2023). AI for life: Trends in artificial intelligence for biotechnology. *New biotechnology*, *74*, 16-24.

Kumar, K., Sharma, R., Ranjan, K., & Pandey, D. K. (2025). Artificial Intelligence in Biotechnology. In *Artificial Intelligence and Biological Sciences* (pp. 192-218). CRC Press.

Miller, T., Durlik, I., Kostecka, E., Kozlovska, P., Łobodzińska, A., Sokołowska, S., & Nowy, A. (2025). Integrating artificial intelligence agents with the internet of things for enhanced environmental monitoring: applications in water quality and climate data. *Electronics*, *14*(4), 696. https://doi.org/10.3390/electronics14040696

Mittal, K., Nadikattu, R. R., Whig, P., & Batra, I. (2025). AI-Driven Environmental Real-Time Monitoring, and Screening. *Artificial Intelligence-Driven Models for Environmental Management*, 163-184. https://doi.org/10.1002/9781394282555.ch7

Naqvi, S. M. Z. A., Tahir, M. N., Raghavan, V., Awais, M., Hu, J., Said, Y., ... & Khan, M. I. (2025). AI-enhanced IoT sensors for real-time crop monitoring: an era towards self-monitored agriculture. *Telecommunication Systems*, *88*(3), 100.

Popović, V., Bošković, J., Filipović, A., & Popović, M. (2024). Application of artificial intelligence (AI) in biotechnology and medicine. *Limes plus*, *20*(2-3), 43-71.

Renganathan, P., & Gaysina, L. A. (2025). Next-generation wastewater treatment: omics and AI-driven microbial strategies for xenobiotic bioremediation and circular resource recovery. *Processes*, *13*(10), 3218. https://doi.org/10.3390/pr13103218

Renganathan, P., Gaysina, L. A., García Gutiérrez, C., Rueda Puente, E. O., & Sainz-Hernández, J. C. (2025). Harnessing engineered microbial consortia for xenobiotic bioremediation: integrating multi-omics and AI for next-generation wastewater treatment. *Journal of Xenobiotics*, *15*(4), 133. https://doi.org/10.3390/jox15040133

Roy, S. M., & Kumari, S. (2025). Integration of Artificial Intelligence in Aquaculture Water Management. In *Nano-solutions for Sustainable Water and Wastewater Management: From Monitoring to Treatment* (pp. 353-368). Cham: Springer Nature Switzerland.

Shende, A. D., Tekade, S. A., Deshmukh, A. A., Tembhurkar, S. P., & Selvakumar, P. (2025). Water Quality Monitoring and Control With AI. In *Smart Water Technology for Sustainable Management in Modern Cities* (pp. 51-80). IGI Global Scientific Publishing.

Singh, A., Mehrotra, R., Rajput, V. D., Dmitriev, P., Singh, A. K., Kumar, P., ... & Singh, A. K. (2022). Geoinformatics, artificial intelligence, sensor technology, big data: emerging modern tools for sustainable agriculture. *Sustainable agriculture systems and technologies*, 295-313.

Singh, K., Sarkar, S., & Sharma, S. (2025). Artificial Intelligence in Bioremediation and Environmental Management. *Environmental Quality Management*, *35*(2), e70192. https://doi.org/10.1002/tqem.70192

Srivastav, A. K., Das, P., & Srivastava, A. K. (2024). Precision Agriculture and Environmental Monitoring. In *Biotech and IoT: An Introduction Using Cloud-Driven Labs* (pp. 151-203). Berkeley, CA: Apress.

Ukhurebor, K. E., Nwankwo, W., Adetunji, C. O., & Makinde, A. S. (2021). Artificial intelligence and internet of things in instrumentation and control in waste biodegradation plants: recent developments. *Microbial Rejuvenation of Polluted Environment: Volume 3*, 265-279.

Volf, M., Vučemilović, A., & Dobrović, Ž. (2024). Enhancing Environmental and Human Health Management Through the Integration of Advanced Revitalization Technologies Utilizing Artificial Intelligence. *Toxics*, *12*(12), 847.

Zhang, X., Ibrahim, Z., Khaskheli, M. B., Raza, H., Zhou, F., & Shamsi, I. H. (2024). Integrative approaches to abiotic stress management in crops: combining bioinformatics educational tools and artificial intelligence applications. *Sustainability*, *16*(17), 7651.