Environment and Social Psychology (2025) Volume 10 Issue 11 ISSN: 2424-8975 (O)
doi: 10.59429/esp.v10i11.4109 2424-7979 (P)

RESEARCH ARTICLE

AI and machine learning in environmental monitoring: enhancing legal compliance and public trust

Adel Subhe Abedalkader Abraheem[1], Jameela Khedher Abbas[2], Duha Khalil Ibrahim Ahmed[3], Nazar Habeeb Abbas[4], Baker Mohammed Khalil[5], Petro Ponochovnyi[6*]

[1] *Al-Turath University, Baghdad 10013, Iraq*

[2] *Al-Mansour University College, Baghdad 10067, Iraq*

[3] *Al-Mamoon University College, Baghdad 10012, Iraq*

[4] *Al-Rafidain University College, Baghdad 10064, Iraq*

[5] *Madenat Alelem University College, Baghdad 10006, Iraq*

[6*] *State University of Information and Communication Technologies, Kyiv, 03110, Ukraine*

*** Corresponding author:** Petro Ponochovnyi, petja9186@gmail.com

ABSTRACT

As environmental pollution becomes more complex over the years, finding effective monitoring methods becomes crucial. In real-time monitoring, artificial intelligence (AI) and machine learning (ML) models can be integrated to obtain information about air, water, and soil quality assessment. To improve the accuracy of pollution detection and forecasting, this study proposes a comprehensive framework that integrates IoT-enabled sensor networks, predictive AI models, and statistical validation techniques. The article assesses the relative performance of Gradient Boosting Machines (GBM), Long Short-Term Memory (LSTM) networks, and Transformer-based split networks to predict environmental changes.

The study was conducted across multi-domain urban, suburban, and rural monitoring zones using multimodal datasets derived from IoT sensors, remote sensing streams, and laboratory-validated environmental indicators. Similar integrated AI–IoT ecological monitoring strategies have been highlighted in recent literature as essential for sustainable environmental protection and high-fidelity pollution forecasting. The dataset comprised 216 air samples, 144 water samples, and 96 soil assays collected from three monitoring regions.

Results show that PM2.5 concentrations decreased by 12% ($p < 0.01$), water turbidity declined by 15% ($p < 0.01$), and lead levels in soil were reduced by up to 16.1% in agricultural sites. The GBM model achieved the highest predictive performance with Root Mean Square Error (RMSE) = 2.1 µg/m³, Coefficient of Determination (R^2) = 0.94, and F1-Score = 92.0%, outperforming LSTM and Transformer models.

Beyond technical performance, this study also highlights the legal and societal dimensions of AI-driven monitoring. By improving accuracy and transparency, these systems strengthen regulatory compliance frameworks while fostering public trust in environmental governance. Understanding how citizens and policymakers perceive the reliability of AI-based platforms is essential to ensuring policy acceptance and compliance behavior. This dual perspective—technological and psychological—illustrates that sustainable outcomes depend not only on advanced

ARTICLE INFO
Received: 30 July 2025 | Accepted: 01 November 2025 | Available online: 18 November 2025

CITATION
Abraheem A S A, Abbas J K, Ahmed D K I. AI and Machine Learning in Environmental Monitoring: Enhancing Legal Compliance and Public Trust. *Environment and Social Psychology* 2025; 10(11): 4109. doi:10.59429/esp.v10i11.4109

algorithms but also on social legitimacy and institutional accountability.

Keywords: AI-driven monitoring; machine learning; environmental pollution; IoT sensors; predictive modeling; sustainability; legal compliance; public trust

1. Introduction

Environmental monitoring has been one of the essential pillars for the responsible management of resources while helping to protect ecosystems, conserve biodiversity, and meet regulatory requirements. However, beyond ecological assessment, environmental monitoring is increasingly tied to regulatory compliance and societal trust, as legal systems require evidence-based enforcement and citizens demand transparency in environmental governance [1, 2].

A recent wave of bibliometric and systematic reviews confirms that environmental AI research is rapidly expanding, with significant growth in applications for pollution prediction, compliance automation, ecosystem monitoring, and biodiversity protection [3-5]. The trend highlights a global transition from fragmented environmental datasets toward unified, data-driven oversight ecosystems enabled through AI and machine learning.

As technology advances exponentially, the convergence of artificial intelligence (AI) and machine learning (ML) with environmental monitoring systems offers a paradigm shift that could drive unprecedented levels of efficiency and effectiveness across our environmental assessment frameworks. Recent research confirms that AI enhances not only technical accuracy but also accountability, strengthening legal compliance frameworks and public trust in policy decisions [1, 6, 7]. The last studies further reinforce this shift, emphasizing that AI-enabled environmental analytics are now central to protected area management [8], real-time ocean waste tracking [9], and predictive environmental change modeling [10]. Additionally, AI is increasingly embedded within environmental governance and judicial processes, enabling more transparent, explainable mechanisms for regulatory enforcement [11, 12]. According to recent WHO assessments, air pollution continues to cause more than seven million premature deaths annually, while UNEP reports that over 40% of global population lives in regions where particulate matter consistently exceeds recommended safety thresholds. Additionally, coastal and freshwater systems remain under significant stress, with nearly 60% of monitored water bodies worldwide showing at least one form of chemical or biological contamination. These updated global indicators emphasize the urgent need for scalable, AI-assisted monitoring frameworks capable of supporting real-time environmental governance and early warning systems [8-10].

Traditional approaches to data collection, analysis, and reporting are often labor-intensive and narrowly focused, and falter under the scale and complexity of modern environmental data. With increasing pollution concerns, evolving climate patterns, and rising complexity of regulatory obligations, oversight now requires not only technical monitoring but also frameworks that address compliance behavior and the psychology of risk perception among industries and the public [13, 14].

AI and ML technologies bring significant benefits, such as improved accuracy and the ability to derive actionable insights from large and heterogeneous datasets. In this respect, AI supports more reliable reporting that regulators can integrate into legal procedures, while its transparency fosters greater societal acceptance of enforcement measures [1, 15, 16].

Such architecture enables organizations to distill and help decision-making in time and act accordingly, enabling monitoring of the environment from a reactive process to an accelerative and preventative one [17]. However, decision-making is not only technical; policymakers must weigh legal standards and public perception, where trust in AI-generated evidence plays a decisive role in determining compliance and the acceptance of environmental regulations [18, 19].

Environmental monitoring mainly measures air, water, soil and other natural resources, but air quality monitoring and real measurement of air pollution are becoming increasingly essential. Monitoring programs can also identify emerging threats, evaluate the effects of human activities, and ensure that industries are adhering to environmental safety standards—by establishing baseline conditions and recording changes over time. However, effective monitoring today requires not only ecological assessment but also legal mechanisms and public trust to ensure compliance and enforcement [1, 2]. But the traditional perspectives don't usually manage to keep pace with the steeply evolving environmental realties. Remote sensing, field instruments, and laboratory analyses yield data in such overwhelming volumes that they threaten to outrun traditional data management metadata frameworks for timely, comprehensive processing. Moreover, there are febrile data in the form of satellite images, ground-based sensor data, social media feeds, and citizen science contributions with an intricacy of integration and interpretation of the data [20].

Artificial intelligence and machine learning (AI and ML) technologies, skilled in analyzing large datasets and identifying trends, represent a possible escape hatch for this bind. AI systems are able to quickly analyze streams of sensor data, recognize when changes deviate from normal conditions and even predict future environmental trends based on historical data. In particular, machine learning models may be trained to detect early warning signals around pollution, ecosystem degradation, or natural resource depletion. The result is a high-cadence at-scale application of intelligent automation which is no longer only higher fidelity in the assessment of the environment and threats but also enables real time reaction to these real time emerging threats. Moreover, AI and ML as technologies possess built-in predictive powers to enable regulators, policymakers, and organizations to understand compliance risks in advance and take remedial measures to mitigate breaches before they happen. These proactive measures will minimize environmental impact and be more cost-effective in terms of remediation and enforcement, while still achieving environmental goals[6, 13].Importantly, this dimension connects directly to policy acceptance and public behavior, since environmental rules are only effective if stakeholders perceive them as fair, transparent, and trustworthy [16, 18, 19].

AI and ML is also allowing for a more widespread approach to the issue of environmental monitoring, beyond simply improving efficiency and accuracy. These technologies help to create a holistic view of environmental conditions by aggregating data from a wide variety of sources from remote Internet of Things (IoT) sensors to publicly available climate models. This bird's eye high-level perspective allows stake-holders to appreciate the complex subtleties of the weaving relationships and the relationships that will correlate with ones that they would have previously missed. Machine learning algorithms, for instance, can link industrial emissions to climate variability in regions to health outcomes, information that helps inform policy and target interventions. The applied mathematics we develop in AI and ML can give innovative insights into understanding the complex systems that dictate how we experience the world can lead to optimized sustainable management of our resources [14, 21, 22].

The automation of regular tasks, making workflows more efficient, is another major benefit AI and ML can drive. Gathering and processing data using multiple data collection, analysis, and reporting procedures can be tedious and expensive. Human resources towards strategy formation, drafting policy and stakeholders can now be channelized into high-value jobs through the automated processes. It minimizes human error, which includes making data analyses more precise and reliable. As a result, this reliability adds to the credibility of monitoring findings and contributes to increased trust between regulators, industry players and the public [1, 7, 16].

The integration of AI and ML into environmental monitoring systems revolutionizes our capabilities in interpretation, conservation, and protection of natural resources. This technology promises to address long-standing frustrations with data complexity, processing speed and predictive accuracy, such that organizations can meet regulation compliance while gaining better environmental results. This novel approach utilizes the distinctive capabilities of AI and ML toward the establishment of a new paradigm of proactive, science-based environmental monitoring — one that reaches beyond the traditional approaches used to characterize environmental conditions.

1.1. The aim of the article

This article aims to explore how artificial intelligence (AI) and machine learning (ML) are transforming environmental monitoring and compliance from just examining your data to a holistic view. Environmental problems are becoming ever more complex and traditional forms of monitoring can't cope with the massive amounts of data that are needed to present timely and actionable information. Here, we will explain how AI and ML can help us to cover the mentioned gaps, offering more precise, time-efficient, and advance solution of the environmental data.

The article aims to determine how AI and ML can benefit anomaly detection, which can speed up the task of localizing pollution sources, following environmental change and enabling forecasting over risk before it expands. Additionally, it will explore how these technologies help to combine disparate data sets from multiple sources, including satellite imagery, multiplex IoT sensors and high-throughput lab analyses into integrated, real-time monitoring frameworks. In this way, the center makes great strides to showcase benefits of AI-powered data fusion that provides a better insight on the environmental situation and developments.

Besides that, it focuses on exploring how artificial intelligence and machine learning can simplify compliance processes. It also highlights how automation of data analytics and report generation helps reduce human error, lowers operational costs, and enhances regulatory compliance. Moreover, it stresses the psychological and legal aspects of trust in AI-backed supervision, showing how industries and firms remain in accordance with sustainability targets while hedging against penalties and reputational damage.

The article aims to elucidate the revolutionary implications that AI and ML could have for environmental monitoring based on evidence rather than misleading hype. As such, it also presents practical use cases, approaches and advantages in three selected areas with the aim of directing political decision-makers, industry decision-makers and researchers towards the application of these forward-looking technologies. This should lead to more efficient usage of the resources, improved compliance with regulations, and ensure the ecosystem are preserved for our future generations.

1.2. Problem Statement

The size and complexity of the ever-evolving environmental challenges are proving to be severely challenging for traditional monitoring solutions, making resource management and compliance to regulations, very painful and unmanageable. Traditional approaches that rely on manual data collection, geographically-fixed monitoring stations for sensing, and disconnected analytic techniques are often unable to consistently produce the right type of timely, high-fidelity insights. As a result, industries, governments and environmental agencies cannot quickly track pollution sources, predict ecosystem disruptions or forecast future environmental trends. The gap between the generation of environmental data and the robust real-time analysis needed to inform decision-making about the environment threatens the efficacy of that decision-making.

Then consider that the type of data sources — satellite imagery, IoT sensor networks, climate models, field observations, simply adds to the challenge. The most valuable information has a synchronicity problem and isn't aggregated enough to be integrated and analyzed in time. Such fragmentation can lead to inefficiencies, for example, unnecessary duplication of monitoring and delayed responses to new threats, and can dilute overall effectiveness of oversight for the environment.

Another key worry is the increasing regulatory fatigue. Decreasing environmental standards set by governments and international bodies mean that more systems are being developed, thus putting pressure on industry to keep up with the ever-changing rules and regulations. However, the approaches adopted here do not adapt quickly enough to changes. As a result, businesses may be vulnerable to compliance violations that can lead to financial penalties, damage to reputation, and adverse impacts on the environment. Importantly, this creates a dual challenge: technical efficiency of monitoring and psychological acceptance of compliance obligations, since regulation is effective only when stakeholders perceive monitoring technologies as legitimate and trustworthy.

Traditional monitoring methods are resource heavy and prone to human error, making it difficult to maintain a consistent accuracy. It should be stressed that there are numerous situations with high stakes, where informed decisions need to be made based on information as swiftly as possible and in as correctly as possible, making this a challenging situation. These limitations of current methodologies display a need for alternative solutions, specifically that they must be more innovative, effective, and scalable in principle to address these multidimensional issues and increase the efficacy of environmental monitoring and compliance programs.

2. Literature review

The use of artificial intelligence (AI) and machine learning (ML) in environmental monitoring and compliance services has recently gained much traction. Such technologies have gained recognition as solutions to these challenges associated with traditional monitoring approaches that are most often rooted in labor-intensive collection methods and fragmented datasets. Artificial intelligence and machine learning are beginning to disproportionately elevate the processes used to gather, analyze, and utilize environmental data [13, 15]

The current literature includes some attractive research activities making use of AI to either automate or improve anomaly detection. It's must to detect pollution sources, variations in water quality, and alterations in the air and imaging data in reality, Machine learning models have proved their capability to recognize trends and anomalies in massive and intricate datasets. With this capability, organizations can rapidly and efficiently respond to environmental hazards unlike ever before. Consequently, leveraging machine learning techniques within monitoring systems is acting as a significant breakthrough by allowing preemptive action to be taken and have in turn minimizing the occurrence of longer-term environmental impacts [23-25].

Cross-source data integration Another major theme in the literature. Large-scale changes (destruction, population displacement, resource depletion) also put pressure on data in the environmental field, which is now inundated with streams of vastly different data varying in scale, resolution, and frequency, thanks to the proliferation of IoT sensors, drones, and satellite imagery. Data-driven initiatives have been attempted by various organizations to merge these disparate datasets into actionable insights using AI and ML techniques.

Using sophisticated algorithms, practitioners are able to derive more value out of a vast amount of raw data, leading to more accurate and quicker decision-making in environmental assessments[20, 26, 27].

Similarly, predictive modeling of environmental trends is developing rapidly. Studies highlight the capacity of AI to model complex historical patterns and forecast events like pollution spikes or water quality declines, enabling proactive resource allocation and preventing compliance violations [22, 28, 29].

At the same time, there is growing attention to how AI-driven monitoring must operate within legal frameworks and public policy ecosystems. Recent analyses underscore the importance of reviewing technological innovations through the lens of sustainability law to ensure their enforceability[2, 30, 31]. Alongside this, scholars stress the significance of trust and transparency in AI systems, indicating that citizen confidence is essential for the acceptance of environmental policies and oversight mechanisms[17, 32, 33].

Recent work has stressed that explainable and legally aligned AI systems are necessary for strengthening institutional legitimacy, especially in contexts involving environmental risk assessment, protected area governance, and industrial emission oversight [8, 12, 34, 35]. However, despite rapid technological expansion, the literature still identifies key gaps, including insufficient cross-domain fusion (air–water–soil), limited real-time anomaly detection frameworks, and inadequate integration of federated or privacy-preserving architectures for industrial monitoring [36]. Addressing these gaps is essential for advancing next-generation environmental compliance systems.

Building on this, it is increasingly evident that AI and ML serve not only as technological tools but also as facilitators of environmental behavior change and social legitimacy. By reinforcing monitoring accuracy and automating compliance reporting, AI strengthens institutional frameworks and helps alleviate regulatory fatigue. However, its societal impact ultimately hinges on how deeply communities, regulators, and industries perceive its fairness, accountability, and psychological acceptability [7, 37, 38].

3. Materials and methods

3.1. Data collection

Data collection by means of multimodal monitoring framework belonging to various IoT-based sensors, automated sampling stations, and in-house laboratory-based validation techniques. In order to comprehensively capture environmental variations, the strategy for deployment was designed to cover diverse geographical landscapes including urban, suburban, and rural areas. In particular, sensor placement methodology accounted for: historical pollution trends; meteorological data; and regulatory standards in determining optimal locations, while minimizing redundancy. In addition, the spatial design was aligned with regulatory compliance zones to ensure results could inform enforceable environmental standards, linking technical deployment to policy relevance [2, 30].

3.1.1. Air quality monitoring

To monitor the air quality, 50 IoT-enabled air quality sensors were placed in various urban centers with metropolitan traffic, industrial zones, and rural areas to measure PM2 5, PM10, NO_2, and SO_2 levels [37]. A spatiotemporal analysis of pollution patterns and modeling of wind trajectories informed sensor deployment. Through hourly data recording, air quality fluctuations were detected in real time which was wirelessly transmitted using LoRaWAN and 5G networks to a cloud-based AI system for real-time analysis and forecasting [6]. Using Voronoi tessellation, a technique used to reduce spatial redundancy of data obtained from sensors through optimal placement in a study area, was utilized to ensure homogeneity and non-overlapping of regions monitored by tracking changes across study area [20].

Beyond monitoring efficiency, this real-time architecture can also be used by regulators for accountability reporting and by communities to increase transparency and trust in policy enforcement [16].This aligns with recent findings showing that next-generation remote-sensing and geospatial workflows now operate at big-data scale, requiring AI-assisted pipelines for efficient extraction of environmental indicators and early detection of pollution dynamics [39].

3.1.2. Water quality monitoring

Water quality was monitored at 10 automated sampling stations established in major water bodies and reservoirs through pH, turbidity, and dissolved oxygen (DO) measurements every 4 h. To identify the optimal sampling locations for high-resolution data collection (avoiding the clustering bias and ensuring that areas were sampled that was prone to contamination from industrial runoff and agricultural activity [2, 7], a Gaussian Mixture Model (GMM) was applied. Watershed turbidity anomalies detected and estimated from in situ sensor data were used to enhance spatial resolution through the integration of satellite-based remote sensing from Sentinel-2 imagery, which can indicate probable contamination sources [26]. A signal filtering algorithm was applied to the collected data to remove noise induced by environmental processes; for instance, sediment resuspension induced by rainfall events [30, 40].

The integration of remote sensing with IoT sensors not only improves scientific accuracy but also provides independent validation sources, which strengthens credibility in legal proceedings and enhances citizen confidence in environmental reporting [14, 40]

3.1.3. Soil quality monitoring

Through the application of a spatiotemporal Kriging Interpolation method, soil quality was monitored at 20 strategically selected sites for enhancing spatial estimation of soil pollution between physical sampling locations. Heavy metals (Pb, Cd) and essential nutrients (N, K) were determined using a combination of in-situ X-ray fluorescence (XRF) spectrometry with laboratory-based Inductively Coupled Plasma Mass Spectrometry (ICP-MS). Soil contamination and nutrient loss were assessed via monthly data collection, facilitating consideration of seasonal variances in soil quality. Time-series decomposition methods were implemented to make sure that long-term trends in the data were accurately identified [38].

Because soil contamination often carries legal and regulatory consequences for land use, the methodology was explicitly designed to align with compliance standards, ensuring that AI-driven results can be applied directly in remediation policies [13, 31].

3.2. Data preprocessing

To ensure data quality, consistency, and reliability, an extensive preprocessing pipeline was implemented, addressing missing values, outliers, and scaling differences across the datasets.

3.2.1. Missing data imputation

Data gaps in air, water, and soil quality records were addressed using a hybrid imputation technique:

- Linear interpolation was applied to short-term gaps (<6 hours) in continuous data streams, such as air and water quality sensor readings, preserving local trends [20].
- Kalman filtering was employed to correct sensor drift and mitigate cumulative measurement errors over extended periods, particularly in soil quality data [38].
- Autoencoder-based reconstruction was used for missing sequences exceeding 6 hours, leveraging deep learning to restore missing patterns while minimizing imputation bias[20].

3.2.2. Outlier detection and removal

Outliers were detected and handled using a two-stage filtering process to ensure that anomalies were correctly classified as either sensor errors or genuine pollution spikes:

1. The 3-standard deviation method was used to identify global outliers that deviated significantly from the historical mean.
2. The Isolation Forest Algorithm was applied to detect localized anomalies, distinguishing between genuine pollution events, such as industrial emissions, sudden algal blooms, and sensor malfunctions [23, 32].

3.2.3. Data normalization

Since environmental variables had differing scales, Min-Max normalization was implemented to standardize feature ranges:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

This transformation improved model training stability and enhanced comparability across datasets, particularly when integrating air, water, and soil parameters into a unified analysis [21].

This preprocessing pipeline also enhances explainability and reproducibility of the models, which are essential for building institutional trust and making AI-based evidence admissible in policy or legal contexts [18, 32].

3.2.4. Feature engineering

Feature engineering, which is a set of processes such extraction and transformation, was utilized in this methodology to increase the predictive power of our machine learning models. Rolling averages with 3-, 6- and 12-hour window were used to capture temporal trends in air pollution levels [33]. Fast Fourier Transform (FFT) was performed on water turbidity data, which helps to identify frequency-domain patterns corresponding to seasonal patterns of contamination trends [24]. The Wavelet Transform decomposition and Savitzky-Golay filtering were applied to minimize fluctuations in long term contamination patterns in soil quality measurements including soil organic carbon (QSOC) and carbon flux (QC) and to filter determined long-term contamination patterns [38].

Importantly, engineered features such as exceedance frequencies and threshold breaches were designed not only for technical robustness but also for direct translation into compliance indicators for regulatory frameworks [28, 29].

3.3. Model development

3.3.1. Machine learning framework

Several machine learning models were trained and optimized for environmental predictions:

- Random Forest (RF): Used as a baseline model due to its robustness in handling mixed-type environmental datasets.
- Gradient Boosting Machines (GBM): Selected for structured regression tasks, as it balances accuracy and computational efficiency [25].
- Long Short-Term Memory (LSTM) Networks: Applied to time-series predictions, particularly for forecasting air pollution trends [22].

- Transformer-Based Models: Integrated air, water, and soil data into a cross-modal framework, capturing complex interdependencies [21].

Model optimization prioritized interpretability in addition to predictive power, ensuring that regulators and policymakers can understand and trust model outputs when they are used for compliance or enforcement [1, 33].

3.3.2. Hyperparameter optimization

Optimizing hyperparameters is one of the important factors that improve predictive performance and generalization of a model. Here, a probabilistic model-based method was used by employing Bayesian Optimization for systematic fine-tuning of hyperparameters of Gradient Boosting Machine (GBM), Long Short-Term Memory (LSTM) networks, and Transformer-based models. Grid and random search techniques were considered but Bayesian Optimization was employed instead, as it features an adaptive exploration-exploitation implementation that quickly homed in on the optimal hyperparameter configuration while minimizing the amount of computation needed.

1. GBM Hyperparameter tuning

The GBM-based model was tuned with respect to learning rate, number of estimators, and maximum tree depth—important parameters to strike a balance between the complexity of the model and generalization performance.

- Learning Rate (0.01–0.1): Determines the size of the steps taken toward minimizing the loss function at each iteration. Smaller values lead to better stability of the model but too small value can make the training time to be very large. While the best value was 0.05 (for fastest convergence without losing predictive accuracy).
- Number of Estimators (100–500): Specifies the number of boosting rounds. A high applies greater representational power for the model but at increased computational costs. The most performing ones had 350 estimators, a good trade-off between overfitting and generalization.
- Max Depth (3–10): Controls single tree complexity. Increased depth helps capture non-linear interactions, but risks overfitting. The best depth for air pollution data was 6, and it was 5 for water and soil datasets to produce good predictions.

The Bayesian Optimization process identified (learning rate: 0.05, estimators: 350, max depth: 6) as the optimal configuration, improving the model's F1-score and reducing mean squared error (MSE) by 8% compared to default hyperparameters.

2. LSTM Hyperparameter tuning

To enhance temporal dependency modeling for time-series predictions in air quality and water quality trends, we optimized Long Short-Term Memory (LSTM) networks.

- Hidden Units (50–200): Specifies the number of neurons in the LSTM layers. While a high number will produce better feature extraction, it will take longer to compute. 128 hidden units was identified as the best configuration to balance model complexity and the time it took to train.
- Dropout Rate (0.1–0.4): Used to reduce overfitting by randomly discarding neurons during training. We found that a dropout rate of 0.2 resulted in the best generalization performance at limiting overfitting while preserving stable predictions.

- Sequence Length (24–72 hours): The total amount of historic data consumed per prediction window. A 48-hour window was best for air pollution forecasting, while water quality trends needed a longer 72-hour window to capture more slowly moving trends.

The optimal LSTM configuration was suggested (128 hidden units, dropout rate: 0.2, sequence length: 48 hours for air pollution, 72 hours for water quality) by Bayesian Optimization. These fine-tuning led to a 10% improvement in the forecasting accuracy and a 15% reduction in the computational overhead relative to the default configuration.

3. Transformer model hyperparameter tuning

The models based on the Transformer architecture were fine-tuned for performing cross-modal learning by merging the datasets on air, water, and soil pollution into a single predictive framework. The tuning process concentrated on attention mechanisms and embedding sizes, which are crucial for processing multi-source environmental data.

- Number of Attention Heads (4–8): Determines how the model attends to different parts of input sequences. More heads allow for richer feature extraction, but too many can cause overfitting and high computational costs. The optimal number was 6 attention heads, balancing expressiveness and efficiency.
- Embedding Dimensions (64–256): Defines the size of input data's vector representations. A higher embedding dimension retains more relationships, but too high values result in redundant information. We measured 128 dimensions as the best fit, as using more was not optimal to represent features nor too heavy on memory.

By using Bayesian Optimization, each model's hyperparameters were tuned which contributed towards higher prediction accuracy, model generalization and computation/time efficiency. These optimization methods resulted in quicker convergence velocities, lower error rate coefficients, and higher real-time applicability, making an artificial intelligence-driven environmental monitoring system more practical and trustworthy for installation on a large scale.

3.4. Performance validation

3.4.1. Error metrics

Many error metrics were used to assess the accuracy and performance of the machine learning models used in this project to predict environmental pollution. These evaluation metrics are handled in Python code files that compile and then use through separate class objects those to examine various alternatives, such as prediction accuracy, robustness against outliers and detection of pollution events. Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R^2), and F1-Score for pollution event detection are the four main evaluation metrics used in this study.

All these metrics are still handy for model evaluation and provide insights from another analytical perspective. The Root Mean Square Error (RMSE) measures the differences between predicted values and observed ones, giving extra weight to large errors at the whole. MAPE calculates percentage error relative to observed values so its only meaningful use is for prediction across different levels of pollution. R^2 measures the portion of variance that is duplicated by the model (the amount of variation in the data accounted for by the model) and indicates how much of any differences in the level of pollution can be explained by the predictors.

1. Root Mean Squared Error (RMSE)

RMSE is a widely used metric that evaluates the average magnitude of prediction errors, with greater emphasis on larger errors. It is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

Where $y_i$represents the observed pollution level, $\hat{y}_i$ is the predicted pollution level, n is the total number of observations.

The accuracy of PM2 prediction by GBM, LSTM and Transformer models was compared using RMSE. 5, salinity and soil pollution trends. GBM models had the lowest RMSE, confirming their high precision and ability to capture environmental pollution fluctuations.

2. Mean absolute percentage error (MAPE)

MAPE gives a percentage of how much relative error predictions are in, and it is useful for comparing performance across models for variables with different scales (pollutants in our case). It is given by:

$$MAPE = \frac{100}{n}\sum_{i-1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

MAPE evaluated the predictability of pollution levels in datasets covering air, water, and soil quality. MAPE being the lowest for GBM demonstrates the supremacy of GBM in pollution trend prediction. This measure had an added advantage in the context of our final models, they were all LSTM models, meaning that we could interpret it to provide a direct comparison of defined forecasting accuracies across the different pollutants.

3. Coefficient of determination (R²)

R^2, or the coefficient of determination, measures how well the model explains the variability of the observed data. It is computed as:

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} \tag{4}$$

Where y_i represents the observed pollution value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of observed values. R^2 particularly useful in assessing the ability of Transformer-based models to integrate multi-source environmental data for cross-domain learning. GBM achieved the highest R^2 values across all environmental parameters, demonstrating its effectiveness in capturing complex interactions between pollutants. The metric validated that air pollution forecasting models performed better than soil contamination models, due to the more dynamic nature of airborne pollutants.

4. F1-Score for pollution event detection

F1-Score is a classification performance metric that evaluates the model's ability to detect pollution spikes while minimizing false alarms. It is given by:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

Where $Precision$ measures the fraction of correctly identified pollution events out of all predicted pollution events and $Recall$ measures the fraction of actual pollution events that were correctly identified by the model.

By framing reductions in PM2.5, turbidity, and heavy metals in terms of enforceable standards, the validation stage strengthens the link between AI predictions, environmental law, and public trust in governance [22, 25].

3.4.2. Statistical analysis

Statistical tests were conducted to assess the significance of differences in environmental pollution levels across regions and to evaluate the effectiveness of AI-driven monitoring in reducing pollution. The following analyses were performed:

1. Analysis of variance (ANOVA) for regional pollution differences

The one-way Analysis of Variance (ANOVA) test was applied to examine whether the mean pollution levels significantly differed between Region 1 (R1), Region 2 (R2), and Region 3 (R3) for air, water, and soil quality parameters [28]. The null hypothesis (H_0) stated that there were no significant differences in pollution levels among the three regions, while the alternative hypothesis (H_a) suggested at least one region had significantly different pollution levels.

$$H_0: \mu_{R1} = \mu_{R2} = \mu_{R3}$$

$$H_a: At\ least\ one\ mean\ differs \tag{5}$$

For each environmental parameter (X), the test statistic was computed as:

$$F = \frac{Between-group\ variance}{Within-group\ variance} = \frac{\sum_{j-1}^{k} n_j \left(\bar{X}_j - \bar{X}\right)^2 / (k-1)}{\sum_{j-1}^{k} \sum_{j-1}^{n_j} n_j \left(\bar{X}_{i,j} - \bar{X}_j\right)^2 / (N-k)} \tag{6}$$

Where k number of groups (regions), N total number of observations, $\bar{X}_j$ mean of region j, $\bar{X}$ grand mean across all regions $\bar{X}_{i,j}$ individual observations.

Since $p<0.05$, the null hypothesis was rejected, confirming that pollution levels varied significantly across regions, justifying the need for region-specific environmental monitoring interventions.

2. Paired t-Test for PM2.5 reduction post-model implementation

To assess the effectiveness of AI-driven pollution monitoring and prediction, a paired t-test was conducted to compare PM2.5 levels before and after model deployment. The null hypothesis (H_0) stated that there was no significant reduction in PM2.5 levels, while the alternative hypothesis (H_a) suggested a significant decrease.

$$H_0: \mu_{pre} = \mu_{post}$$

$$H_a: \mu_{pre} > \mu_{post} \tag{7}$$

The test statistic was calculated as:

$$t = \frac{\bar{D}}{s_D \sqrt{n}} \tag{8}$$

Where $\bar{D}$ mean difference between pre- and post-model implementation values, s_D standard deviation of the differences, n sample size.

Since $p<0.01$, the null hypothesis was rejected, confirming that the AI-driven monitoring system significantly reduced air pollution levels, likely due to early warning alerts, optimized traffic flow, and enhanced regulatory enforcement.

3. Tukey’s HSD test for soil contamination differences

To determine which regions exhibited significant differences in soil contamination, a Tukey's Honest Significant Difference (HSD) test was conducted following the ANOVA. This test controlled for multiple comparisons by adjusting for family-wise error rate.

The HSD statistic was calculated as:

$$q = \frac{\bar{X}_i - \bar{X}_j}{s_{pooled} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \tag{9}$$

Where $\bar{X}_i, \bar{X}_j$ means of groups i and j, s_{pooled} is pooled standard deviation, n_i, n_j sample sizes of respective groups.

Since $p<0.05$, these results confirmed that soil contamination levels varied significantly based on land use type, supporting targeted remediation strategies and land-use planning policies.

This methodology has developed a comprehensive AI-based paradigm for air, water, and soil quality monitoring, and regulatory compliance. This method serves as a scalable solution for automated environmental assessment by integrating sensor networks, advanced ML models, and real-time anomaly detection. By leveraging multi-source data fusion, ensuring robust preprocessing and predictive analytics, this allows for a proactive approach towards environmental risks with minimum manual oversight possible while increasing forecasting accuracy [1].

4. Results

4.1. Air quality results

The enhanced air quality dataset had resulted in more precise estimates of pollutant concentration across the region, and integrated real time sensor data and predictive analytics. PM2. 5, NO_2, and SO_2 data were provided with higher cadence, enabling a tighter view of pollution hotspots. Results showed that in R1 sites, the NO_2 levels were high and persistent due to heavy traffic and industrial emissions. Most of the highest PM2. 5 levels, which varies considerably according to seasonal meteorological conditions. In rural areas whose pollution levels were less (R3), it was seen that even there were spikes in SO_2 which might be due to the agricultural practices in the area.

Table 1. mean concentrations of air pollutants across monitoring regions.

Region	PM2.5 Mean ± SD (μg/m³)	PM10 Mean ± SD (μg/m³)	NO_2 Mean ± SD (ppb)	SO_2 Mean ± SD (ppb)	Max PM2.5 (μg/m³)	Min PM2.5 (μg/m³)
R1 (Urban)	21.8 ± 4.5	30.2 ± 5.6	18.5 ± 2.3	7.2 ± 1.1	32.1	13.5
R2 (Suburban)	25.4 ± 5.1	33.8 ± 6.2	20.1 ± 3.0	7.5 ± 1.2	36.5	14.2
R3 (Rural)	19.7 ± 3.9	28.5 ± 5.1	16.4 ± 2.1	6.9 ± 1.0	29.8	12.8

The results show PM2. 5 and PM10 concentrations observed in suburban regions (R2) were considerably elevated, measuring at an average of 25.4 ± 5.1 μg/m³ and 33.8 ± 6.2 μg/m³ respectively. That was probably a result of traffic-related emissions and meteorological phenomena that trap fine particulate matter. The maximum NO_2 concentrations, 18.5 ± 2.3 ppb, were present in urban locations (R1); this is most likely due to traffic emissions and industrial emissions, confirming findings of earlier research on urban pollution hotspots. PM2.0 concentrations (R3) were lowest in rural areas. Although levels remained high with median N_2 and SO_2 (30 ppb) with occasional peaks in SO_2, seasonal burning of agricultural waste could be a contributing factor. These differences highlight the importance of pollutant-specific control strategies in different regions.

Real-time sensor data is tested against different AI models to validate its applicability in predicting air pollutants levels. The Gradient Boosting Machine (GBM) model outperforms all other models, while Transformer-based models lock closely behind. The results obtained from the prediction using Random Forests and LSTM would be comparatively good, however both models failed when predicting short-term pollution spikes which attracted a higher prediction error.

Table 2. Model performance for air quality predictions.

Model	RMSE (PM2.5 µg/m³)	MAPE (%)	R^2 Score	F1-Score (Pollution Spike Detection)
Random Forest	2.9	7.8	0.89	88.7
Gradient Boosting Machine (GBM)	2.1	6.4	0.94	92.0
LSTM (Time Series Model)	2.5	7.1	0.91	90.3
Transformer-Based Model	2.3	6.8	0.92	91.1

The GBM model exhibited the lowest RMSE (2.1 µg/m³) and the highest R^2 score (0.94), outperforming all other models in PM2. 5 trends. In summary, boosted ensemble learning methods effectively captured the both long-term pollution trends as well as short-term fluctuations. The LSTM models performed well in general, but were slightly less accurate due to their complex time-series dependency. The competitive performance of the Transformer model also showed potential effectiveness on non-linear relationships of the air pollution dataset with deep-learning approaches. The results also demonstrate that ensemble and deep-learning-based methods are very suitable for real-time air quality prediction.

This measurable reduction is not only a scientific finding but also demonstrates compliance with WHO and national air quality thresholds, making it directly relevant for regulatory enforcement [2, 22]. Furthermore, the high interpretability of GBM outputs strengthens institutional credibility, allowing regulators to justify interventions transparently and thereby enhance public trust in environmental governance [1, 18].

4.2. Water quality results

The processed water quality dataset showed seasonal and geographical variations in pH, turbidity, and dissolved oxygen levels for all monitoring stations. In conclusion, the results revealed increased turbidity levels during periods of heavy rainfall, signifying potential sediment and pollutant runoff. Decreased dissolved oxygen (DO) concentrations during summer months indicate higher microbial activity and lower oxygen solubility at warmer temperatures.

Table 3. Water quality indicators across monitoring stations.

Station	Observed pH ± SD	Predicted pH ± SD	Turbidity (NTU) Observed	Turbidity Predicted	Dissolved Oxygen (mg/L) Observed
S1	7.2 ± 0.2	7.2 ± 0.2	0.5 ± 0.1	0.4 ± 0.1	8.5 ± 0.6
S2	6.8 ± 0.3	6.9 ± 0.3	0.6 ± 0.2	0.5 ± 0.2	7.9 ± 0.5
S3	7.3 ± 0.3	7.3 ± 0.3	0.4 ± 0.1	0.3 ± 0.1	8.1 ± 0.5

Water pH was also constant across all monitoring locations, and there was little variation between the water pH values observed and those predicted by the model, indicating that trends of water pH could be accurately captured using AI-driven models. Across all monitoring stations, pH levels consistently ranged between 6.4 and 7.8, remaining within acceptable ecological thresholds, while dissolved oxygen (DO) values fluctuated between 6.1 and 9.3 mg/L depending on seasonal temperature shifts. These ranges align closely with the predictions generated by the LSTM and GBM models, reinforcing the validity of the combined sensor–AI evaluation pipeline. The maximum turbidity was registered in urban areas (S2) at 0.6 NTU, likely owing to runoffs from infrastructure and human activities. The levels of dissolved oxygen exhibited a seasonal decline, most pronounced at suburban sites (S2), as predicted by the temperature-oxygen solubility relationship. These findings further emphasize the utility of AI models in effectively monitoring and predicting fluctuations in water quality, enabling the development of preemptive strategies for contamination detection and intervention. Comparable studies have documented similar predictive reliability of AI systems for water resource management, noting that multimodal AI–IoT architectures can improve contamination detection speed and reduce decision-making latency for environmental regulators [41, 42]. These findings reinforce the applicability of the present model for policy-relevant water governance.

The ability to cross-validate real-time IoT data with remote sensing enhances transparency and minimizes disputes in legal or policy settings [26, 40]. In addition, the integration of multiple data sources bolsters confidence among local communities that monitoring platforms are unbiased and reliable, thus improving acceptance of compliance obligations [16, 29].

4.3. Soil quality results

The consolidated soil data included an in-depth comparative analysis of heavy metals contamination levels according to different land-use types with special emphasis on lead (Pb) and cadmium (Cd) data. The quality of the soil ranked low but varied by region, as traces were more prominent in industrial and urban locations. The decreasing trend of lead levels for the monitoring period indicates the effectiveness of remediation measures in some areas. Cadmium levels were relatively stable, with increased concentrations in industrial areas because of long-term emissions from industrial activity.

Table 4. Heavy metal concentrations across soil sampling sites.

Site	Lead (mg/kg) Observed	Predicted Lead (mg/kg)	Cadmium (mg/kg) Observed	Predicted Cadmium (mg/kg)	Reduction in Lead Levels (%)
Site 1 (Urban)	10.9 ± 1.2	10.8 ± 1.1	2.1 ± 0.3	2.1 ± 0.2	12.5
Site 2 (Industrial)	12.2 ± 1.3	12.0 ± 1.2	2.5 ± 0.4	2.4 ± 0.3	8.9
Site 3 (Suburban)	9.8 ± 1.1	9.7 ± 1.0	1.9 ± 0.2	1.8 ± 0.2	14.3
Site 4 (Agricultural)	8.5 ± 0.9	8.4 ± 0.8	1.7 ± 0.2	1.7 ± 0.1	16.1

Lead concentration in the soil was found to decrease by 12.5% in urban areas and over 16.1% in agricultural ones, where natural processes of bioremediation could be used to reduce metals. However, the industrial sites (Site 2) showed the least decrease (8.9%), suggesting that the pollution sources are still active and need more strict mitigation measures. Regular monitoring and strict regulations are emphasized by consistently high levels in industrial sites. These results emphasize the value of site-specific remediation efforts based on land-use attributes.

Different machine learning models were compared to the AI-based soil quality predictions to measure their efficiency. Four modelling techniques were compared, and it was found that the GBM model significantly outperformed the other techniques, especially with respect to the prediction of lead levels (the target heavy metal in question).

Table 5. AI model performance for heavy metal contamination predictions.

Model	R^2 Score (Lead)	R^2 Score (Cadmium)	Mean Absolute Error (mg/kg)	Prediction Latency (s)
Random Forest	0.89	0.87	0.3	4.2
Gradient Boosting Machine (GBM)	0.92	0.91	0.2	3.6
LSTM (Time Series Model)	0.90	0.89	0.3	5.8
Transformer-Based Model	0.91	0.90	0.2	6.1

The best R^2 scores for lead and cadmium (0.92 and 0.91, respectively) were produced by the GBM model, indicative of strong predictive performance. Among the models, GBM showed the lowest mean absolute errors (0.2 mg/kg) and was the most reliable model for soil contamination assessments. Although the LSTM and Transformer models performed competitively, the significantly higher prediction latency of 5.8s and 6.1s, respectively, indicates that these methods are less efficient for real-time monitoring These results show that AI can deliver accurate, near-real time predictions of contamination, facilitating a quicker response to soil pollution events.

Because soil contamination carries direct regulatory consequences for agriculture and land use, these reductions demonstrate how AI-enhanced monitoring can support evidence-based liability assessment [13, 31]. The capacity to communicate soil quality improvements in a transparent, data-driven manner also builds public trust, ensuring that remediation policies are not only enforceable but also socially legitimate [7, 37].

4.4. Integrated multi-parameter environmental evaluation

A comprehensive analysis of the environment was carried out by creating a multi-modal pollution prediction framework by combining the AI models from all areas. The performance of these models was further analyzed according to prediction accuracy, precision, recall, and ability to detect environmental anomalies efficiently.

Table 6. Water use efficiency by leadership diversity quartile.

Domain	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Air Quality	94.2	93.0	93.5	93.2
Water Quality	92.7	91.8	92.1	91.9
Soil Quality	93.5	92.4	93.1	92.7
Combined Environmental Model	93.7	92.5	93.1	92.8

The accuracy of the overall integrated model was 93.7%, where air quality prediction (94.2%) was the best. In Weighted Average, precision and recall scores were high across environmental parameters reflecting the ability of the AI system to accurately detect pollution spikes and long-term contamination trends. This needs utilization of different data types — air; water; soil integrated together, and this turned out to be the fulcrum for this analysis facilitating more holistic view which not just enhances predictive ability, but can lead to early warning system.

4.5. Statistical validation of ai-driven environmental monitoring

Statistical tests were conducted to determine the significance of the observed reduction trends in pollution and also to compare environmental quality across regions which further ensured the reliability of the AI models trained.

Table 7. Composite sustainability scores by leadership diversity quartile.

Metric	ANOVA (p-value)	Paired t-Test (Mean Reduction)	Tukey HSD (Regional Differences, p-value)
PM2.5 (μg/m³)	0.02	12% reduction ($p < 0.01$)	Significant ($p = 0.02$)
NO_2 (ppb)	0.03	9% reduction ($p < 0.05$)	Significant ($p = 0.03$)
SO_2 (ppb)	0.04	8% reduction ($p < 0.05$)	Moderate ($p = 0.06$)
Turbidity (NTU)	0.01	15% reduction ($p < 0.01$)	Significant ($p = 0.01$)
Lead (mg/kg)	0.05	10% reduction ($p < 0.05$)	Significant ($p = 0.02$)

The ANOVA test confirmed that air, water and soil pollution levels varied across regions ($p < 0.05$), which confirmed the significant differences in pollution sources across regions. The paired t-tests showed a 12% decrease of PM2. 5 points and 15% decrease of water turbidity, show that the AI-driven interventions based on drone flyover images significantly impacted the dormitory's environmental quality. By confirming statistically significant regional differences, Tukey's HSD test justified the need for location-specific environmental policies.

The statistical analysis was based on a total of 216 air-quality observations, 144 water-quality samples, and 96 soil assays collected across the three monitoring regions. Effect-size calculations indicated substantial practical significance, with $\eta^2 = 0.41$ for $PM_{2.5}$ reductions, $\eta^2 = 0.37$ for turbidity improvements, and $\eta^2 = 0.33$ for soil lead mitigation. Complementary pairwise comparisons showed medium-to-large effect sizes (Cohen's d ranging from 0.62 to 1.05), confirming that the observed pollutant reductions align not only with statistical significance but with meaningful environmental impacts. These quantitative gains correspond directly to improvements in model accuracy, indicating a strong association between predictive performance and monitored environmental outcomes (recent studies report similar patterns in AI-driven water and soil monitoring [5, 41, 42].

This improves frequency, accuracy, and predictive capabilities of detecting environmental factors, including pollution levels. We reasoned that analyzing these datasets together would allow better interpretation of results, so we integrated them into a unified framework that enables fine-grained anomality discovery to tackle environmental issues with precision. The statistical validation affirmed the effectiveness of AI-driven solutions in realizing concrete pollution reductions and demonstrated that these approaches enhance compliance with environmental policies and strengthen long-term sustainability initiatives.

By demonstrating quantifiable improvements (e.g., 12% PM2.5, 15% turbidity, 16% Pb reduction), the models provide actionable metrics that regulators can use in drafting and enforcing compliance policies [25, 28]. Statistical rigor reinforces the admissibility of AI outputs in both scientific and legal frameworks, while interpretability ensures stakeholder confidence in adopting these technologies [32, 33].

4.6. Policy and social implications

The empirical results of this study demonstrate that AI-driven environmental monitoring can deliver more than technical improvements; they establish a pathway for reshaping compliance and governance practices. The reductions observed in air pollutants, water turbidity, and soil contaminants illustrate how advanced analytics can be directly translated into enforceable standards, providing policymakers with evidence that supports stronger environmental regulation.

At the same time, the interpretability of the models ensures that results are not viewed as opaque outputs of a "black box," but as transparent, auditable evidence that can be used with confidence in both legal and institutional contexts. This quality is crucial in bridging the gap between technological innovation and regulatory acceptance.

From a societal standpoint, the integration of multiple validated data sources fosters public confidence in the credibility of monitoring systems. Trust in these systems is not a peripheral concern but a determining factor in whether communities and industries accept and comply with environmental regulations. Without social legitimacy, even the most accurate technologies risk being underutilized or contested.

The findings highlight that AI and ML are not only instruments of environmental science but also vehicles for strengthening governance. They reinforce accountability, facilitate compliance, and help align institutional frameworks with public expectations. In this sense, the technology becomes both a scientific tool and a social contract mechanism, enabling more sustainable, transparent, and widely accepted environmental management.

5. Discussion

The article demonstrates the potential of AI and machine learning (ML) models to improve environmental monitoring for air, water and soil quality. Thanks to a combination of state-of-the-art IoT sensors, cloud-based AI, and statistical validation methods, we were able to achieve highly accurate real-time pollution detection and predictive analysis. The results indicate that the Gradient Boosting Machines (GBM) were a powerful concurrence for all other ML models used in this experiment with higher accuracy for predicting the pollutants' levels. These findings are consistent with earlier studies that highlighted the promise of AI in environmental change monitoring but also indicate where improvements still are necessary.

Beyond technical accuracy, these outcomes highlight the importance of ensuring that model outputs are perceived as trustworthy by regulators, industries, and communities [1, 18]. Without public acceptance and institutional legitimacy, even the most accurate predictive models may fail to translate into real-world compliance or behavioral change [16].

Importantly, this study represents the potentially the most comprehensive multi-modal integration of environmental parameters in existing literature. Existing works related to individual air-pollutant such as predicting air quality with deep learning[29] or predicting water quality with neural networks [24]. In contrast, this study is the first of its kind to create an air, water and soil data, merged into a single model, from which the resultant can be an all-rounder examination of the trends of pollution in the environment. This integration is key to modeling cross-domain dependencies, for instance the influence of air pollutants on soil content or the relationship between industry emissions and turbidity level in nearby water.

From a governance perspective, such integration has additional implications: multi-domain monitoring creates a more transparent evidence base, which can be used to strengthen environmental legislation and improve accountability mechanisms [2, 31].

Compared to Lowe et al.'s work [26], which surveyed recent AI applications in water treatment and monitoring, this study further builds upon the discussion by proposing a predictive modeling framework that leverages heterogeneous environmental data sources. Most past studies on AI have centered around reactive strategies, for example, AI-aided identification of contaminants in water bodies; our suggested approach involves proactive forecasting which empowers early remediation before pollutants reach dangerous limits. Monitoring is moving from reactive to predictive monitoring, which is a major breakthrough in our environmental management strategies.

Predictive monitoring also reshapes social expectations: when pollution is forecast in advance, policymakers and communities can engage in proactive mitigation, reinforcing a shared responsibility for sustainability [14, 19].

Moreover, this study extends the findings of Wan et al. studied Deep Learning models for predicting water turbidity [24]. Although their findings confirmed that RNNs are consistent in processing time-series variation, this study indicates that GBM models outperform RNNs, in terms of overall efficacy, despite their computational overhead. Furthermore, boosted decision trees perform better than standard neural networks and can be more efficiently deployed in smart real-time environmental monitoring systems. The interpretability of boosted decision trees is also crucial, since interpretable models enhance transparency and improve public confidence in AI-driven decision-making [32, 33].

The use of AI for forecasts of air quality has been well analyzed in literature, especially in works such as published by Subramaniam et al.[29] that reviewed AI-driven forecasting models in the atmospheric system, more specifically, for air pollution. Their results show that the meteorological-based models were limited and that using machine learning can result in the improved accuracy. This study validates those findings: AI-driven models also yielded a 12% drop in PM2. 5 for early warnings and adaptive mitigation strategies. The transformer-based models and LSTM architectures evaluated in this work further highlighted the great efficiency of deep learning when it comes to modeling short-term pollution dynamics, even if their demanding computational requirements prevent real-time implementation. However, successful implementation requires more than model efficiency: policy uptake and social trust determine whether early warnings are acted upon effectively [7, 37]. These findings also correspond with new applications of AI in large-scale forestry management and aquatic biodiversity modeling, where deep learning architectures have been shown to capture complex ecological interactions with unprecedented accuracy [3, 5].

From a soil monitoring view, this study complements a recent review by Chen & Wang[38] examined the potential of heterogeneous sensor networks for soil characterization at large scales. In their research, it was shown that active learning methods increase the efficiency of data gathering, which is in line with what is observed in this study, where the results of a spatiotemporal Kriging interpolation showed that two samples significantly improved predictions of soil contamination. But there is one major limitation: current AI models are not good at predicting trends over longer time scales for soil quality because contamination spreads slowly. While many of the soil degradation processes are slow, causes of air and water pollution show extreme variability over short time scales, making long-term historical datasets necessary to improve model reliability. This limitation also reveals a governance challenge: regulators must balance the short-term variability of pollution with the long-term progression of soil degradation—highlighting the need for policies that integrate both immediate forecasting and long-term resilience planning [13, 25].

The study also provides bearing challenges and constraints that need to be solved in further studies despite promising outcomes. One area that is a limitation is data variability between geographic regions and its effect on generalizability of AI models. Although predictions achieved high accuracy within the places of study, calibrating the models to variable climatic or industrial situations would likely need to be performed in another place. Fan et al. [14] raised similar concerns in study of how AI models can adapt to applications in environmental sustainability. The study stated that AI models trained on a certain dataset tend to fail when encountering unseen environments, hence a push to develop adaptive algorithms. In this respect, generalizability is not only a technical problem but also a social one: stakeholders are less likely to trust monitoring systems if they appear to fail outside controlled conditions [28]. Recent research similarly notes that environmental AI adoption depends on public confidence in the fairness, transparency, and accuracy of algorithmic tools, particularly among youth and environmentally active communities [43]. Trust-building mechanisms, including explainable AI interfaces and transparent IoT data-sharing frameworks, are increasingly recognized as central to enhancing compliance behavior and encouraging pro-environmental decision-making. Building adaptive models that can demonstrate robustness across contexts is therefore essential for maintaining legitimacy [16].

Another restriction concerns sensor reliability and missing data issues especially in monitoring of water and soil quality. Although the Kalman filter and reconstruction techniques based on autoencoders were employed to address the problem of missing data, sensor failures and communication problems remain to be solved. As Reis et al. [18], AI-driven applications for data science have inherent challenges when it comes to hardware degradation and signal interference in sensor networks, resulting in systematic biases. Addressing this will need self-correcting AI models that are capable of auto-regulation to overcome sensor failures, and missing data in real-time. Reliability also has a psychological dimension: when communities perceive frequent breakdowns or inconsistencies in monitoring data, their confidence in enforcement and compliance declines [16]. Future AI systems must therefore address both technical resilience and perceived trustworthiness [1]. Other studies warn that AI adoption for natural resource and environmental safety management requires explicit legal frameworks to avoid compliance disputes, ensure admissibility of algorithmic evidence in judicial settings, and maintain procedural fairness in environmental litigation [11, 34].

Findings are also in line with the work of Korra & Sadhana [19], in which Internet of Things sensors, cloud AI and satellite broadband are integrated for monitoring environmental elements in real time. Although their study highlighted the relevant applications of the smart city infrastructures in the realm of pollution management, this study builds on their framework by proving the efficacy of the predictive AI models not only in urban contexts but also exploring their applicability in rural and industrial contexts. However, there is still a lag in data transmission, particularly in areas with little to no connectivity. Further work needs to focus on optimizing the edge architecture to reduce the latency in AI-based environmental monitoring. This is consistent with 2025 evidence showing that federated and distributed learning architectures can help mitigate digital inequities by allowing local monitoring systems to function even under limited connectivity, thus maintaining environmental safety across heterogeneous regions [36]. Connectivity gaps also raise equity issues: if only well-connected regions benefit from AI-based oversight, disparities in environmental protection may widen, reducing societal trust in governance mechanisms [2].

Additionally, the outcomes underscore the importance of policy formulation and regulatory adjustment to ensure that AI-enabled environmental tracking translates into meaningful benefits. Literature such as Adefemi et al.[16] argues that for effective protection of the environment, AI-centric pollution tracking systems should be integrated into government policy frameworks. This study illustrates that even if AI models generate more precise forecasts, the effectiveness of these systems is ultimately determined by their uptake by policy makers, industries and municipal authorities. Thus, the social legitimacy of AI is as important as its technical capacity: adoption depends on whether stakeholders trust the system and perceive it as aligned with their rights, duties, and collective environmental goals [7, 33].

Although this study focused on predictive modeling and data-driven insights, the integration of AI with autonomous intervention systems can be explored in future research. Recent developments in AI-enabled robotic remediation and automated pollution control systems indicate that AI has the potential not just to monitor but to actively reduce environmental pollution [31]. The next step toward creating self-sustaining frameworks for environmental protection will be to investigate how machine learning can be integrated with autonomous remediation systems. Future research should also examine the psychological and institutional readiness for such interventions, since their success will rely not only on technical performance but also on acceptance by regulators, industries, and affected communities [14, 37].

Evidence from literature supports this conclusion, emphasizing that integrated AI–IoT–remote sensing architectures are becoming foundational elements of global sustainability strategies, climate-change mitigation initiatives, and environmental protection systems. These developments highlight the growing importance of data-driven, legally aligned, and socially trusted AI infrastructures for environmental governance.

The article illustrates how transformative AI and machine learning can be to environmental monitoring, delivering accurate predictions of pollution levels, near-real-time anomaly detection, and the ability to integrate cross-domain datasets. As indeed this comparison with other literature shows that research in the field of AI-based environmental insight has experienced a noticeable professionalization; however, generalizability of data, reliability of sensors, and adaptability of policy all remain major challenges. By focusing on scalable AI frameworks, adaptive learning techniques, and autonomous intervention systems, future studies will offer transformative potential that converts AI-driven environmental monitoring into actionable and sustainable solutions. The broader impact of this work lies not only in advancing environmental science but also in showing how AI can strengthen compliance regimes, foster societal trust, and contribute to sustainable governance [1, 2, 16]. This dual contribution: scientific precision and social legitimacy—marks the key to making AI-driven monitoring a cornerstone of environmental psychology and policy [18, 31].

The strong alignment between predicted and observed pollutant patterns demonstrates that high-accuracy AI systems can directly support evidence-based regulatory interventions. When deployed within environmental agencies, such predictive stability enables earlier issuance of compliance warnings, more targeted inspections, and scientifically defensible policy decisions. This outcome echoes broader findings that AI-supported governance frameworks enhance institutional transparency and increase the public's willingness to trust environmental information generated by automated systems (recent governance-oriented environmental studies support similar conclusions [11, 12, 43].

6. Conclusions

Environmental monitoring has also been aided with the use of AI and machine learning techniques that have proven to significantly invest in improving the accuracy, efficiency and prediction of pollution in the air, water and soil. A key take-away from the study was the successful generation of an AI-powered multi-modal environmental detection framework by utilizing the real-time data from IoT sensors and state-of-the-art machine learning methods to predict pollution trends and discover anomalies. Enabling such predictive models would facilitate these proactive instead of reactive solutions. The research is enlightening because it shows the role of Artificial Intelligence (AI)-based approaches like gradient boosting machines, long short-term memory networks, and transformer-based architectures towards improved forecasting accuracy and pollution detection that enable timely and data-driven environmental management decisions.

The findings provide compelling evidence that AI-enable environmental monitoring enhances pollution assessment across diverse geophysical contexts, closely capturing geographical variation in pollution and pollution dynamics. Notably, the advanced feature engineering techniques and statistical validation methodology adopted contributes to the robustness of the AI models, making it capable of identifying breakthrough environmental hazards at unprecedented levels of sensitivity. Based on these results, it is further highlighted in the study that cross-domain correlations could be introduced into a combined AI model by integrating multi-dimensional environmental parameters into the same network, which could facilitate a profound comprehension of that pollution sources, interactions and long-term trends. And integrating air, water and soil data into a unified predictive framework is a significant advance in terms of environmental analytics, providing informed insights not available through conventional monitoring techniques.

In spite of the numerous advantages of AI-based monitoring, there still some challenges and limitations that should be addressed in order to improve AI monitoring system in long term. One of the crucial things is variability in data due to different environmental circumstances, which might restrict the generalization of AI models in other regions. Sensor networks are commonly pioneers in real-time data collection and therefore the integrity of AI-based predictions could be compromised due to hardware failure, data voids, and transmission delays. Besides, computationally expensive deep-learning based models may also pose scalability issues in resource-poor settings with poor digitally enabled infrastructure or cloud computing capabilities. These challenges highlight the importance of the need for flexible learning strategies, refined sensor calibration methods and robust computing platforms to facilitate continuous and accurate monitoring of the environment.

This study further enhances this concept by making the models transferable using the transfer learning and reinforcement learning approaches, which now that is an area of research for AI models to learn from cross environmental conditions. Moreover, AI technologies with built-in autonomy concerning intervention could also provide the protection of ecosystems, enabling a form of automation of pollution control, instant remediation, and optimal regulatory enforcement. Research investigating the role of AI for climate adaptation strategies and disaster response would enable a better understanding of the contribution of AI-powered environmental monitoring tools towards broader goals of sustainable global development.

The study highlights the potential of AI and machine learning to change the landscape of environmental science and transform how pollution is assessed and managed. As artificial intelligence technologies evolve, their adoption within the context of environmental policy, smart city infrastructure, and industrial regulation will be critical to realize their full promise. AI-powered environmental monitoring has the potential to enhance sustainability and environmental management by building on existing challenges and monitoring innovations, and contributing to greater data-driven decisions and ultimately a healthier global ecosystem space.

The study underscores that technical innovation alone is not sufficient: for AI to deliver lasting value, it must be embedded in governance systems that emphasize compliance, accountability, and fairness. The social acceptance of these technologies—rooted in transparency, trust, and perceived legitimacy will determine whether they become widely adopted or resisted. By aligning predictive power with psychological and societal dimensions, AI-driven monitoring can evolve from being a scientific tool to serving as a foundation for collective environmental stewardship. This dual capacity, advancing precision while reinforcing trust, marks its most transformative potential within environmental psychology and policy.

Conflict of interest

The authors declare no conflict of interest

References

1. Hacker P. Article: Sustainable AI Regulation. Common Market Law Review. 2024.
2. Qi Y. The Impact of Artificial Intelligence on Environmental Protection. Highlights in Science, Engineering and Technology. 2024;96:152-6.
3. Miller T, Michoński G, Durlik I, Kozlovska P, Biczak P. Artificial Intelligence in Aquatic Biodiversity Research: A PRISMA-Based Systematic Review. Biology. 2025;14.
4. Okafor C, Otunomo F, Nnadi V, Nzekwe C, Nwoye A, Ajaero C. Artificial intelligence in environmental research: bibliometric, text mining and content analysis. Discov Artif Intell. 2025;5:124.
5. Wang T, Zuo Y, Manda T, Hwarari D, Yang L. Harnessing Artificial Intelligence, Machine Learning and Deep Learning for Sustainable Forestry Management and Conservation: Transformative Potential and Future Perspectives. Plants. 2025;14.
6. Popescu SM, Mansoor S, Wani OA, Kumar SS, Sharma V, Sharma A, et al. Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management. Frontiers in Environmental Science. 2024;12.
7. Alprol AE, Mansour AT, Ibrahim ME, Ashour M. Artificial Intelligence Technologies Revolutionizing Wastewater Treatment: Current Trends and Future Prospective. Water [Internet]. 2024; 16(2).
8. Atalay A, Perkumienė D, Safaa L, Škėma M, Aleinikovas M. Artificial Intelligence Technologies as Smart Solutions for Sustainable Protected Areas Management. Sustainability. 2025.
9. Adeoba M, Pandelani T, Ngwangwa H, Masebe T. The Role of Artificial Intelligence in Sustainable Ocean Waste Tracking and Management: A Bibliometric Analysis. Sustainability. 2025.
10. Anand A. Machine Learning and Artificial Intelligence in Environmental Change Prediction. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 2025.

11. G A, D S. Greening the justice system: assessing the legality, feasibility, and potential of artificial intelligence in advancing environmental sustainability within the Indian judiciary. Frontiers in Political Science. 2025.
12. Zhou C. Artificial Intelligence and Urban Air Quality: The Role of Government and Public Environmental Attention. Sustainability. 2025.
13. Miller T, Cembrowska-Lech D, Kisiel A, Krzemińska A, Kozlovska P, Jawor M, et al. HARNESSING AI FOR ENVIRONMENTAL RESILIENCE: MITIGATING HEAVY METAL POLLUTION AND ADVANCING SUSTAINABLE PRACTICES IN DIVERSE SPHERES. Grail of Science. 2023(26):151-6.
14. Fan Z, Yan Z, Wen S. Deep Learning and Artificial Intelligence in Sustainability: A Review of SDGs, Renewable Energy, and Environmental Health. Sustainability. 2023.
15. Emerging Technology Integration - Artificial Intelligence (AI) and Machine Learning (ML) for Predictive Analysis for Safety and Toxicity Assessment in Environmental Toxicology. International Journal of Scientific Research and Management (IJSRM). 2024;12(05):1182-95.
16. Adefemi A, Ukpoju EA, Adekoya O, Abatan A, Adegbite AO. Artificial intelligence in environmental health and public safety: A comprehensive review of USA strategies. World Journal of Advanced Research and Reviews. 2023.
17. Giannopoulos M, Tsagkatakis G, Tsakalides P. Higher-Order Convolutional Neural Networks for Essential Climate Variables Forecasting. Remote Sensing [Internet]. 2024; 16(11).
18. Reis T, Dumberger L, Bruchhaus S, Krause T, Schreyer V, Bornschlegl MX, et al. AI-Based User Empowerment for Empirical Social Research. Big Data and Cognitive Computing [Internet]. 2024; 8(2).
19. Korra C, & Sadhana, A. Integrating IoT Sensors, Cloud AI, and Satellite Broadband for Enhanced ESG Governance in Smart Cities: A Tripartite Approach to Real-Time Environmental Monitoring. International Journal of Innovative Science and Research Technology (IJISRT). 2024;9(3).
20. Decorte T, Mortier S, Lembrechts JJ, Meysman FJR, Latré S, Mannens E, et al. Missing Value Imputation of Wireless Sensor Data for Environmental Monitoring. Sensors [Internet]. 2024; 24(8).
21. Khan M. Advances in Architectures for Deep Learning: A Thorough Examination of Present Trends
22. Journal of Artificial Intelligence General science (JAIGS) 2024.
23. Kumar R, Goel R, Sidana N, Sharma A, ghai S, Singh T, et al. Enhancing climate forecasting with AI: Current state and future prospect [version 1; peer review: 1 approved]. F1000Research. 2024;13(1094).
24. Noh S-H, Moon HJ. Anomaly Detection Based on LSTM Learning in IoT-Based Dormitory for Indoor Environment Control. Buildings [Internet]. 2023; 13(11).
25. Wan S, Yeh M-L, Ma H-L, Chou T-Y. The Robust Study of Deep Learning Recursive Neural Network for Predicting of Turbidity of Water. Water [Internet]. 2022; 14(5).
26. Si MaW, Brett M. and Du, Ke, . Long-Term Evaluation of Machine Learning Based Methods for Air Emission Monitoring. SSRN Electronic Journal. 2023.
27. Lowe M, Qin R, Mao X. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. Water [Internet]. 2022; 14(9).
28. Hino M, Benami E, Brooks N. Machine learning for environmental monitoring. Nature Sustainability. 2018;1(10):583-8.
29. Ameh B. Digital tools and AI: Using technology to monitor carbon emissions and waste at each stage of the supply chain, enabling real-time adjustments for sustainability improvements. International Journal of Science and Research Archive. 2024.
30. Subramaniam S, Raju N, Ganesan A, Rajavel N, Chenniappan M, Prakash C, et al. Artificial Intelligence Technologies for Forecasting Air Pollution and Human Health: A Narrative Review. Sustainability [Internet]. 2022; 14(16).
31. Hernan G, Dubel AK, Caselle JE, Kushner DJ, Miller RJ, Reed DC, et al. Measuring the Efficiency of Alternative Biodiversity Monitoring Sampling Strategies. Frontiers in Marine Science. 2022;9.
32. Kumar D, Bassill NP, Ghosh S. Analyzing recent trends in deep-learning approaches: a review on urban environmental hazards and disaster studies for monitoring, management, and mitigation toward sustainability. International Journal on Smart Sensing and Intelligent Systems. 2024;17(1).
33. Chapman A, Lauro L, Missier P, Torlone R. Supporting Better Insights of Data Science Pipelines with Fine-grained Provenance. ACM Trans Database Syst. 2024;49(2):Article 6.
34. Satish K, Krishna Kishor T, Sandhyarani G, Er. Aman S, Prof. Sangeet V, Shalu J. Leveraging AI and Machine Learning to Optimize Retail Operations and Enhance. Darpan International Research Analysis. 2024;12(3):1037-69.
35. Bredikhina V, Batsman Y, Tolkushcha K. Features of the use of artificial intelligence technologies in the management of natural resources and environmental safety. Uzhhorod National University Herald Series: Law. 2025.

36. Xu Z, Shi X, Shu W, Xin Y, Zan X, Si Z, et al. Machine learning classifiers to detect data pattern change of continuous emission monitoring system: A typical chemical industrial park as an example. Environment international. 2025;201:109594.
37. Ramadan M, Ali M, Khoo SY, Alkhedher M. SecureIoT-FL: A Federated Learning Framework for Privacy-Preserving Real-Time Environmental Monitoring in Industrial IoT Applications. Alexandria Engineering Journal. 2025.
38. Ab. Rahman E, Hamzah FM, Latif MT, Azid A. Forecasting PM2.5 in Malaysia Using a Hybrid Model. Aerosol and Air Quality Research. 2023;23(9):230006.
39. Chen H, Wang J. Active Learning for Efficient Soil Monitoring in Large Terrain with Heterogeneous Sensor Network. Sensors [Internet]. 2023; 23(5).
40. Dritsas E, Trigka M. Remote Sensing and Geospatial Analysis in the Big Data Era: A Survey. Remote Sensing. 2025.
41. Hino M, Benami E, Brooks N. Machine learning for environmental monitoring. Nature Sustainability. 2018;1:583-8.
42. Ponnuru A, Madhuri J, Saravanan S, Vijayakumar T, Manimegalai V, Das A, et al. Data-Driven Approaches to Water Quality Monitoring: Leveraging AI, Machine Learning, and Management Strategies for Environmental Protection. Journal of Neonatal Surgery. 2025.
43. Sheik AG, Kumar A, Srungavarapu CS, Azari M, Ambati S, Bux F, et al. Insights into the application of explainable artificial intelligence for biological wastewater treatment plants: Updates and perspectives. Eng Appl Artif Intell. 2025;144:110132.
44. Popa R-G, Chenic A. Artificial Intelligence, Consumer Trust and the Promotion of Pro-Environmental Behavior Among Youth. Sustainability. 2025.