

Application of Machine Learning and Data Science in Heavy Metal Remediation: Advances, Challenges, and Future Perspectives: A Review

Faisal Latif, Muhammad Bilal, Muhammad Hasnain, Huzaifa, Mahnoor Saeed, Muhammad Ali, Raziya Nadeem

*Department of Chemistry, University of Agriculture, 38040, Faisalabad, Pakistan

Corresponding Author email: faisalbuzdar24@gmail.com

Abstract

Heavy metal contamination of soils and water remains a persistent global environmental and public health challenge, exacerbated by industrialization, mining, and intensive agricultural activities. Although conventional remediation technologies, including chemical precipitation, adsorption, and bioremediation, have proven effective, optimizing them often relies on empirical, site-specific trial-and-error approaches, which limit scalability and increase operational costs. Machine learning (ML) and data science are increasingly proposed as transformative tools for predicting and optimizing remediation performance; however, their practical contribution and field-scale reliability remain insufficiently and critically assessed. The current review evaluates supervised, unsupervised, and deep learning approaches, including regression models, artificial neural networks, support vector machines, ensemble methods, clustering algorithms, and deep neural architectures, for predicting metal-removal efficiency, optimizing operational variables, and Modelling adsorption and remediation kinetics. Beyond reporting predictive performance, the review critically examines how well current ML models provide mechanistic insight, remain transferable across heterogeneous environmental conditions, and outperform conventional statistical and process-based approaches. It focuses on data quality, feature selection, validation strategies, multi-source data integration, and the risks of overfitting and data leakage. The review also assesses ML's potential to advance bioremediation and phytoremediation by resolving complex microbial nutrient contaminant interactions and reducing experimental requirements. It also identifies limited, non-standardized datasets; limited interpretability; inadequate external validation; and poor cross-site generalisability as major barriers to real-world implementation. Future research should therefore prioritize explainable and physics-informed ML, robust uncertainty quantification, IoT-enabled real-time monitoring, and interoperable environmental databases to develop scientifically defensible, scalable, and adaptive remediation decision-support systems.

Keywords: Machine learning, Data science, Heavy metal remediation, Predictive Modelling, Artificial neural networks

1. Introduction

Heavy metals are generally defined as elements possessing a relatively high density (typically greater than 5 g cm^{-3}) that are toxic even at low concentrations. Cadmium (Cd), lead (Pb), chromium (Cr), arsenic (As), and mercury (Hg) are among the most widely studied toxic heavy metals. These elements are used extensively across industry, agriculture, and technology, including battery manufacturing, electroplating, paint pigments, mining and smelting, fossil fuel combustion, agrochemicals, and electronics manufacturing. Their widespread use, combined with inadequate waste disposal practices, has resulted in extensive environmental contamination.

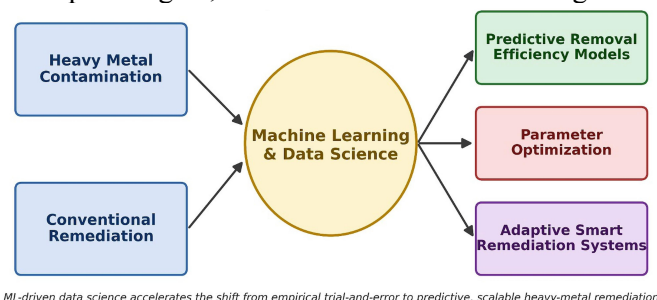
Rapid industrialization and urbanization have increased the environmental burden of heavy metal pollutants. The combined effects of mining, fossil fuel combustion, and agrochemical use are particularly significant because heavy metals are persistent: once released into the environment, they can remain in ecosystems for decades, degrading soil and water quality, disrupting food production, and increasing the risk of cancer and other developmental and immunological disorders in exposed populations. Heavy metals interact with soil matrices, and the limitations of conventional clean-up technologies mean this contamination remains inadequately addressed. Artificial intelligence (AI)-assisted optimization of bioremediation processes has therefore emerged as a promising research direction [1], opening new possibilities for intelligent, technology-enabled remediation.

Predictive Modelling has become an essential tool for assessing immobilization techniques and reducing ecological risk. For example, machine learning models have improved predictions of heavy-metal immobilization rates in soils treated with alkaline materials [2], while predictive frameworks applied to electrokinetic soil remediation have improved transparency of the mechanisms and process variables governing metal removal [3]. Remediation technologies such as adsorption, precipitation, phytoremediation, and chemical fixation remain dependent on empirical experimentation and are time- and resource-intensive. These methods generally lack predictive capability and are constrained by site and soil heterogeneity, variable environmental conditions, and the difficulty and cost of translating laboratory results to field-scale applications. Without uncertainty quantification, traditional models can undermine confidence in remediation decision-making. Coupling machine learning with reliability analysis has been shown to enhance robustness and predictive performance in biochar-based heavy metal removal systems, providing a more systematic framework for remediation design [4]. Data-driven techniques are therefore becoming central to overcoming the limitations of conventional trial-and-error approaches.

The growth of data-driven remediation reflects the broader digital transformation of environmental engineering. Large datasets from laboratory testing, remote monitoring, and field surveys now support computationally efficient machine

learning algorithms that can identify hidden patterns, determine optimal operating parameters, and simulate diverse remediation scenarios. Data-driven frameworks reduce the time and cost of experimentation, making remediation more affordable and scalable, and support a shift from reactive to proactive and prescriptive environmental management.

This review aims to provide an integrated, meta-level synthesis of recent developments in applying machine learning (ML) and data science (DS) to heavy metal remediation, bringing together environmental engineering, computational intelligence, adsorption science, and bioremediation research rather than treating individual technologies in isolation. Through a systematic review of contamination sources, conventional remediation methods, data science workflows, machine learning algorithms, and applied case studies, this review discusses current challenges including data quality, model interpretability, and limited generalisability. It considers emerging directions such as explainable AI, IoT-enabled monitoring, and global environmental databases (Fig. 1). By bringing together these interdisciplinary perspectives, the review aims to connect environmental experimentation with predictive computational Modelling to help researchers, engineers, and decision-makers develop intelligent, sustainable remediation strategies.



ML-driven data science accelerates the shift from empirical trial-and-error to predictive, scalable heavy-metal remediation

Figure 1: Remediation technologies including ML and DS

1.1. Sources and Characteristics of Heavy Metal Contamination

1.1.1. Natural Sources (Geogenic Processes)

Weathering of rocks and volcanic activity are the main geological processes that release heavy metals into soil and water. Because mineral composition and geological structure vary regionally, background heavy metal concentrations differ substantially between locations. Geogenic inputs generally occur gradually, although some regions show naturally elevated concentrations of arsenic, chromium, or nickel. Establishing baseline concentrations is therefore essential when assessing environmental contamination, since geogenic contributions vary with pH, temperature, and soil texture, and there is growing interest in distinguishing natural background variation from anthropogenic contamination [5]. Comparisons with geological reference values, such as the upper continental crust and shale mean values, together with indices including the geoaccumulation index and enrichment factor, are commonly used to separate lithogenic contributions from human-derived inputs [6]. In granitic and other metal-rich parent-rock terrains, geogenic weathering alone can elevate chromium, nickel, and arsenic concentrations well above conventional guideline values, so multivariate statistical methods such as principal component analysis and cluster analysis are increasingly applied to apportion natural and anthropogenic sources with greater confidence [7]. These source-apportionment approaches provide labeled, source-attributed datasets that later machine learning models use to distinguish geogenic background from pollution signals.

1.1.2. Anthropogenic Sources (Industrial, Mining, Agricultural Activities)

Anthropogenic activity is the dominant driver of heavy metal contamination worldwide. Mining and smelting generate metal-laden waste that accumulates in soils and waterways, while electroplating, battery production, and tannery effluents introduce toxic metals directly into water systems. Agricultural practices, including phosphate fertilizer use and irrigation with wastewater, add further contamination, and urban runoff and landfill leachate contribute to widespread dispersion. Machine learning techniques have been used to predict heavy-metal uptake by activated carbon adsorbents, supporting the evaluation of these materials for environmental applications [8], and AI-enabled resilience assessments further demonstrate the value of intelligent systems for pollution mitigation across diverse ecosystems [9].

1.1.3. Behaviour and Speciation of Heavy Metals in Soil and Water

The environmental fate of heavy metals is governed by their chemical speciation, which in turn determines mobility, bioavailability, and toxicity. Variables such as pH, redox potential, ionic strength, and organic carbon content influence how strongly metal ions bind to soil particles and whether they remain in solution or become immobilized. Because these variables interact in complex ways, interpreting speciation behaviour is analytically challenging. Machine learning frameworks have been applied to model heavy metal fate in coastal ecosystems, capturing spatial and temporal patterns more effectively than conventional deterministic models [10] and improving assessments of contaminant transport and ecological exposure.

1.1.4. Persistence and Bioaccumulation

Heavy metals are among the most persistent and widespread pollutants because they accumulate in sediments and soils and, unlike organic pollutants, do not break down. They enter living organisms through root uptake in terrestrial plants or ingestion in aquatic animals. Bioaccumulation occurs when organisms absorb metals faster than they can eliminate them,

while biomagnification increases metal concentration at successively higher trophic levels, posing risks to both predator species and humans (Table 1). Predictive Modelling of bioaccumulation is therefore important for ecological risk assessment, and intelligent analytical frameworks support predicting long-term effects and developing sustainable management practices [11]. Figure 2 illustrates the pathway of heavy metals from natural and anthropogenic sources through environmental transport, bioaccumulation, and trophic biomagnification.

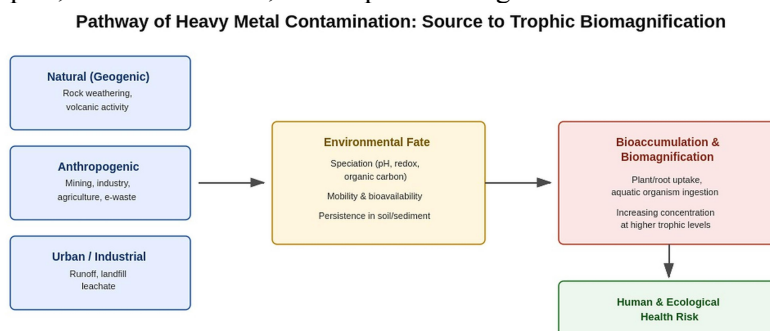


Figure 2: Heavy metal contamination pathway from source to environmental fate and trophic biomagnification[5,11]

Table 1. Major heavy metals, sources, environmental behaviour, and health risks.

Heavy Metal	Major Anthropogenic Sources	Environmental Behaviour	Major Health Risks	Reference
Cadmium (Cd)	Mining, battery manufacturing, phosphate fertilizers	High mobility in acidic soils; bioaccumulative	Kidney damage, bone fragility, carcinogenicity	[5]
Lead (Pb)	Battery recycling, paints, smelting, urban runoff	Strong adsorption to soil particles; persistent	Neurotoxicity, developmental disorders	[12]
Chromium (Cr)	Tannery effluents, electroplating, metallurgy	Speciation-dependent mobility (Cr3+ vs Cr6+)	Carcinogenicity (Cr6+), liver and kidney damage	[11]
Arsenic (As)	Mining, pesticides, groundwater contamination	Highly mobile under reducing conditions	Skin lesions, cancer, cardiovascular disease	[13]
Mercury (Hg)	Fossil fuel combustion, artisanal gold mining	Methylation in aquatic systems; biomagnification	Neurological disorders, fetal toxicity	[14]

1.2. Conventional Metal Remediation Techniques

1.2.1. Physical Remediation Methods

Physical remediation methods for contaminated soils include excavation, thermal desorption, membrane filtration, and soil washing. While these approaches can be effective at localized contamination points, they are typically costly and environmentally disruptive; for instance, excavated soil must be transported for secure landfill disposal, and membrane systems require careful operational control to function efficiently. Recent systematic reviews indicate that artificial intelligence (AI) has improved Modelling of adsorption processes, enhancing predictions of contaminant removal efficiency and optimizing operational parameters [15]. Membrane-based physical treatment is particularly amenable to this integration: artificial neural networks and other data-driven models now forecast fouling behavior and permeate flux decline in membrane systems treating heavy-metal-laden wastewater, allowing operators to schedule cleaning and adjust transmembrane pressure before performance deteriorates rather than relying on fixed maintenance schedules [16]. This coupling of machine learning with physical remediation methods improves treatment-parameter selection and reduces the need for repeated experimentation.

1.2.2. Excavation and off-site disposal.

Excavation, sometimes termed "dig-and-haul," physically removes contaminated soil for secure landfill disposal or ex-situ treatment and is one of the best demonstrated available technologies (BDATs) for small, well-delineated hotspots of heavy-metal contamination. It offers rapid, verifiable removal of the contaminant source. It does not depend on soil chemistry or metal speciation. Still, it is expensive at scale, disrupts soil structure and microbial communities, and merely relocates rather than destroys the contaminant, shifting the liability to the receiving landfill [17,18].

1.2.3. Thermal desorption

Thermal desorption volatilizes metals with appreciable vapour pressure, principally mercury, by heating excavated or in-situ soil to temperatures ranging from around 100 °C to over 600 °C, after which the vapour is captured and condensed for recovery or safe disposal. Field- and pilot-scale trials have reduced residual mercury concentrations to below 1 mg kg⁻¹. The method is fast and can achieve high removal efficiencies for volatile metals. Still, it is energy-intensive, largely ineffective for non-volatile metals such as lead and chromium, and can alter soil structure and organic matter content, sometimes requiring post-treatment amendment before the soil can be reused [18,19].

1.2.4. Soil washing and soil flushing.

Soil washing uses water, acids, chelating agents such as EDTA, or surfactants to desorb metals from soil particles into a liquid washing solution, exploiting differences in particle size, density, and surface chemistry to concentrate the contamination into a smaller volume for further treatment. Removal efficiencies exceeding 90% have been reported for cadmium, lead, and zinc under optimized washing conditions. The main limitations are the generation of a metal-laden spent washing solution that also requires treatment, potential loss of soil fertility from removing fine particles and organic matter, and reduced effectiveness in fine-textured or heterogeneous soils [17,18].

1.2.5. Membrane filtration.

Membrane filtration passes metal-laden water or leachate through microfiltration, ultrafiltration, nanofiltration, or reverse osmosis membranes, which reject metal ions or metal-bearing particles and complexes based on size and charge; nanofiltration and reverse osmosis routinely achieve rejection rates above 90–99% for dissolved heavy metals [20]. Membrane filtration is compact, modular, and produces a relatively clean permeate suitable for reuse. Still, it suffers from membrane fouling, high energy demand for high-pressure processes, and a need for pre-treatment to protect membranes from scaling.

1.3. Chemical Treatment Methods

Chemical remediation methods, including precipitation, coagulation-flocculation, ion exchange, and chemical stabilization, are widely used to treat heavy metal contamination. Each process requires precise control of chemical dosing and environmental conditions, and excessive chemical use can generate secondary waste streams; the complexity of chemical equilibria also complicates predicting treatment outcomes. Integrating machine learning with capacitive deionization systems has substantially improved prediction accuracy and operational efficiency for heavy metal removal [21], enabling adaptive control of chemical processes that reduces inefficiency and environmental risk.

1.3.1. Chemical precipitation.

Chemical precipitation converts soluble metal ions into insoluble solids, typically metal hydroxides, sulfides, or carbonates, by raising pH or adding a precipitating reagent, after which the resulting solids are separated by sedimentation or filtration [22]. It is one of the most widely applied industrial methods because of its operational simplicity and low reagent cost. It can achieve removal efficiencies above 95% for many divalent metals. However, it generates large volumes of metal-bearing sludge that require further dewatering and disposal, and its performance is highly sensitive to pH and competing complexing agents.

1.3.2. Coagulation-flocculation.

Coagulation destabilizes colloidal and fine particulate metal species by neutralizing their surface charge, typically using aluminum or iron salts. At the same time, flocculation aggregates the destabilized particles into larger flocs, often with the aid of polymeric flocculants such as polyaluminium chloride, so they can be removed by sedimentation [22]. The process is effective as a pre-treatment or polishing step ahead of filtration and is well suited to high-turbidity effluents. Still, it produces substantial secondary sludge, requires ongoing reagent dosing, and is less effective for metals present in truly dissolved rather than colloidal form.

1.3.3. Ion exchange.

Ion exchange removes dissolved metal cations by exchanging them for less harmful ions, commonly sodium or hydrogen, held on synthetic resins or natural zeolites; because the resin can be regenerated and reused, the process is capable of near-complete removal of metals such as Cd(II), Cr(III), Cu(II), Ni(II), and Zn(II) even at low influent concentrations [22]. Ion exchange is therefore particularly valuable for polishing dilute effluents to meet strict discharge limits, but resin costs, sensitivity to fouling by organic matter and competing ions, and the need to manage the resulting regenerant brine limit its use in large or highly variable waste streams.

1.3.4. Chemical stabilisation (immobilisation)

Chemical stabilization reduces the mobility and bioavailability of heavy metals in situ by amending soil with materials such as lime, phosphate compounds, biochar, or industrial by-products including coal fly ash, which promote precipitation, adsorption, complexation, and ion-exchange reactions that bind metals into less soluble forms without physically removing them [23]. The approach is comparatively low-cost, minimally disruptive to soil structure, and applicable over large areas. Still, it does not reduce the total metal inventory in the soil, so its long-term effectiveness depends on maintaining stable geochemical conditions such as pH, and periodic monitoring is required to confirm that immobilized metals do not re-mobilize as conditions change.

1.4. Biological Approaches (Bioremediation and Phytoremediation)

Microbial and phytoremediation approaches remove or immobilize heavy metals through biological mechanisms, using microorganisms or plants to improve soil, sediment, or water quality. Microbial metabolism can convert heavy metals into less toxic or non-toxic forms. At the same time, phytoremediation relies on plants' ability to take up and stabilize metals from soil, offering a cost-effective and environmentally sustainable remediation pathway. However, these methods are strongly dependent on environmental conditions such as soil type, nutrient availability, and climate, all of which affect treatment performance. Predictive Modelling is increasingly used to optimize biological remediation systems and

improve scalability [24]. Figure 3 summarises the integration of machine learning with bioremediation and phytoremediation processes for predictive Modelling and optimization of microbial- and plant-based heavy metal removal.

Machine Learning Integration in Bioremediation and Phytoremediation

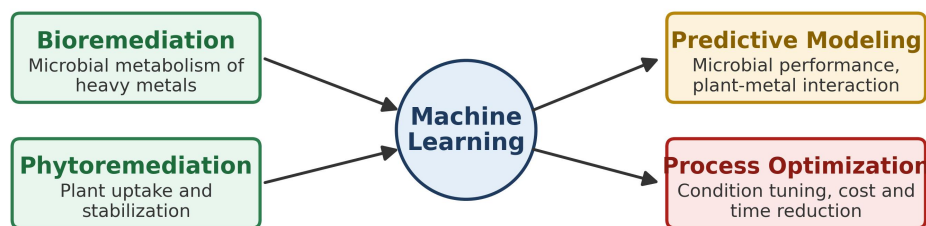


Figure 3: Integration of machine learning in bioremediation and phytoremediation processes for heavy metal removal [29,24]

1.5. Limitations of Traditional Trial-and-Error Approaches

Conventional remediation design relies primarily on incremental, trial-and-error experimentation, which is costly in both time and resources and typically fails to capture the complex interactions between variables. Scaling laboratory results to field conditions introduces substantial uncertainty, and the absence of predictive frameworks slows decision-making. Coupling data-driven Modelling with conventional remediation methods can improve efficiency, reduce costs, and enhance environmental outcomes. Table 2 compares the principal conventional heavy metal remediation techniques, their advantages, and their limitations.

Table 2. Comparison of conventional heavy metal remediation techniques.

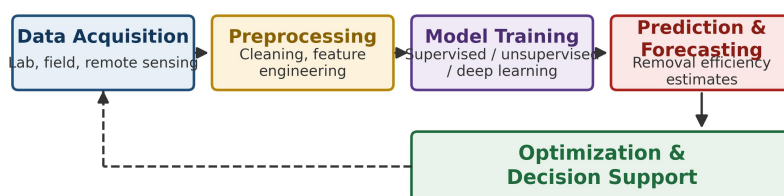
Technique Type	Examples	Advantages	Limitations	Reference
Physical Methods	Excavation, soil washing, membrane filtration	Rapid contaminant removal	Expensive, energy-intensive, secondary waste	[15]
Chemical Methods	Precipitation, ion exchange, stabilisation	High removal efficiency	Chemical sludge formation, reagent cost	[21]
Biological Methods	Bioremediation, phytoremediation	Eco-friendly, cost-effective	Slow process, sensitive to conditions	[24]
Electrokinetic Methods	Electrokinetic soil remediation	Effective for fine-grained soils	High energy demand, scaling challenges	[3]

1.6. Introduction to Data Science in Environmental Remediation

1.6.1. Concept of Data-Driven Environmental Engineering

Data-driven methods are transforming environmental engineering by shifting decision-making from observation-based to prediction-based approaches. Machine learning algorithms can identify patterns in large environmental datasets that traditional statistical methods may miss, improving understanding of nonlinear relationships between operational parameters and remediation performance. The application of computational techniques in environmental science continues to expand the range of analytical methods available for optimizing physical remediation systems [25], supporting the development of predictive frameworks to select effective remediation strategies. Figure 3 illustrates a general machine learning framework for heavy metal remediation, encompassing data acquisition, Modelling, prediction, and optimization.

General Machine Learning Framework for Heavy Metal Remediation



Dashed line: feedback loop refines operating parameters and re-informs model training

Figure 3: Machine learning framework for heavy metal remediation, highlighting data acquisition, Modelling, prediction, and optimization [25]

This shift supports a transition from reactive remediation, where action is taken only after contamination occurs, to proactive, sustainability-oriented environmental management. Table 3 summarises the main machine learning techniques used to remove lead and other heavy metals from soil and water, along with their use in assessing and mitigating environmental impact.

Table 3. Machine learning algorithms applied in heavy metal remediation.

Algorithm Category	Common Algorithms	Applications in Remediation	Strengths	Limitations	Reference
Supervised Learning	Linear Regression, SVM, Random Forest	Prediction of removal efficiency, adsorption capacity	High predictive accuracy	Requires labelled datasets	[36]
Artificial Neural Networks	ANN, Deep Neural Networks	Modelling nonlinear adsorption kinetics and optimization	Handles complex nonlinear relationships	Limited interpretability	[34]
Ensemble Methods	Gradient Boosting, XGBoost	Process optimization and robustness enhancement	Improved generalization	Computational complexity	[35]
Unsupervised Learning	K-means, Hierarchical Clustering	Contamination hotspot detection	No labeled data required	Limited predictive capability	[39]

1.6.1. Role of Big Data in Remediation Planning

Environmental monitoring systems, laboratory experiments, and remote sensing technologies generate large volumes of structured and unstructured data that substantially exceed the scale of traditional environmental datasets. Managing and interpreting such data requires advanced analytical tools that can convert large, complex datasets into actionable insights. Big data analytics combined with machine learning algorithms improves pattern recognition, forecasting accuracy, and risk assessment for contaminated sites [26], enabling stakeholders to evaluate treatment efficiency, simulate project scenarios, and develop optimal operational strategies. As a result, big data integration supports a more transparent, scalable, and agile approach to performance-based remediation projects.

1.6.2. Integration of Experimental and Monitoring Data

Combining controlled laboratory data with real-time monitoring data improves predictive-model accuracy: laboratory data provide controlled insight into adsorption behavior, while monitoring data capture performance under field conditions. Integrating these datasets lets machine learning models generalize predictions across different environments. For example, combining environmental magnetism with machine learning has enabled rapid diagnosis of heavy metal contamination in lake sediments using multi-source data [27], demonstrating the value of integrated data for pollution source detection. Such integrative frameworks improve spatial mapping of contamination, strengthen risk assessment, and support earlier identification of at-risk areas.

1.6.3. Digital Transformation in Environmental Sciences

Artificial intelligence and digital technologies are modernizing environmental science through automation, smart sensors, cloud computing, and advanced analytics, enabling continuous environmental monitoring and predictive maintenance of remediation systems. Digital transformation has strengthened collaboration among researchers, policymakers, and engineers through improved data sharing and real-time decision-making, and AI-driven tools are supporting more resource-efficient remediation planning as digital innovation continues to reshape global contamination management [28].

1.7. Data Science Workflows in Metal Remediation

1.7.1. Data Acquisition and Preprocessing

Laboratory experiments, pilot-scale remediation studies, remote sensing observations, industrial discharge records, and ecological monitoring programs are the principal sources of environmental datasets. Reliable machine learning Modelling depends on the accuracy and completeness of these data. Well-structured environmental data support developing and refining models of microbial-assisted remediation processes [29]. High-resolution datasets are critical for effective model training and validation; historical contamination records improve the ability to predict future occurrences, and careful documentation of experimental conditions strengthens the reliability of resulting datasets.

1.7.2. Laboratory and Field Monitoring Data

Controlled laboratory conditions generate datasets suitable for characterizing adsorption efficiency and kinetic and equilibrium behavior. In contrast, field monitoring captures real-time, in situ variation from climate change, soil heterogeneity, and hydrological variability. Using machine learning to predict bioaccumulation and histopathological effects in aquatic organisms illustrates the value of long-term monitoring data for ecological risk prediction [30]. Combining laboratory and field datasets improves model generalisability, while well-designed monitoring frameworks better represent natural environmental variability.

1.7.3. Data Cleaning and Normalization

Raw environmental data are often inconsistent, contain outliers, and are recorded on different measurement scales. Data cleaning involves removing erroneous entries, correcting measurement errors, and standardizing units, while normalization improves numerical stability and algorithm performance [31]. Commonly applied scaling procedures include min-max normalization, which rescales variables to a fixed range such as zero to one, and z-score standardization, which centers variables on a mean of zero with unit variance; the appropriate choice depends on the sensitivity of the

downstream algorithm to feature magnitude, with distance-based and gradient-based methods such as k-nearest neighbors, support vector machines, and neural networks being particularly sensitive to unscaled inputs. Effective preprocessing prevents predictive models from being disproportionately influenced by extreme values and improves comparability across experimental conditions. Recent methodological guidance for environmental machine learning further emphasizes that preprocessing decisions, including how outliers and duplicate records are treated and how variables from heterogeneous field, laboratory, and remote-sensing sources are harmonized, should be documented transparently, since inconsistent preprocessing pipelines are a recognized source of poor reproducibility and inflated performance estimates across environmental prediction studies.

1.7.4. Handling Missing and Noisy Data

Environmental datasets frequently contain missing values due to sensor malfunction or sampling constraints. Imputation and noise-reduction techniques improve dataset completeness and reliability, and advanced machine learning methods can accommodate incomplete data through probabilistic Modelling. Recommended practice is to distinguish data missing completely at random from data missing systematically, since the latter can bias model outcomes if imputed naively; multiple imputation and model-based approaches such as random-forest imputation are generally preferred over simple mean substitution for environmental datasets with structured missingness. Addressing these data quality issues strengthens predictive robustness and reduces uncertainty.

1.7.5. Feature Engineering and Selection

Identifying relevant environmental variables and selecting appropriate predictors are central to effective feature engineering and directly influence remediation outcomes. Well-chosen predictor variables improve both model interpretability and computational efficiency. A clear relationship has been observed between feature quality and prediction accuracy in machine learning models used to evaluate chitosan-based flocculants for heavy metal removal [32]. Effective feature engineering therefore improves model accuracy while reducing computational cost.

1.7.6. Data Integration and Management

Integrating diverse datasets, such as chemical analyses, hydrological measurements, and biological indicators, provides a more detailed basis for prediction. Multi-source data fusion supports a comprehensive assessment of contamination and a more holistic approach to remediation planning. Nevertheless, the harmonization of variables acquired at differing spatial resolutions, sampling frequencies, and levels of instrumental precision across field, laboratory, and remote-sensing platforms remains a significant methodological challenge that must be addressed to ensure the reliable integration of such heterogeneous datasets

1.7.7. Real -Time Monitoring Systems

Smart sensor networks enable continuous data collection to support adaptive machine learning models, improving response speed and operational efficiency in real-time monitoring systems. Coupling artificial intelligence agents with Internet of Things sensor infrastructure allows water-quality and other environmental parameters to be streamed, processed, and acted upon with minimal delay, enabling remediation systems to move from periodic manual sampling toward continuous, autonomous surveillance of contamination levels.

1.7.8. Database Development for Remediation Projects

Structured environmental databases support long-term sustainability and knowledge sharing, giving stakeholders greater transparency through a centralized repository. A recent global soil study illustrates the value of this approach at scale: by compiling 9,489 heavy-metal fraction measurements from 219 published sources across 56 countries and pairing them with open geospatial soil-property databases such as SoilGrids, researchers were able to train an XGBoost classification model that maps mobility hotspots worldwide and, using mercury as a case study, identified regions exposing an estimated 15.1 million people and over 100 million hectares of cropland to high-mobility contamination [33]. Such consolidated, publicly accessible repositories illustrate how database development directly extends the geographic reach and reliability of predictive remediation models. Integrating real-time monitoring data with machine learning enables automated control and dynamic optimization of remediation processes as programs scale. Figure 5 illustrates a smart remediation system integrating machine learning, real-time data analysis, and automated control for soil and water heavy metal removal.

Smart Remediation System: Real-Time Monitoring and Adaptive Control

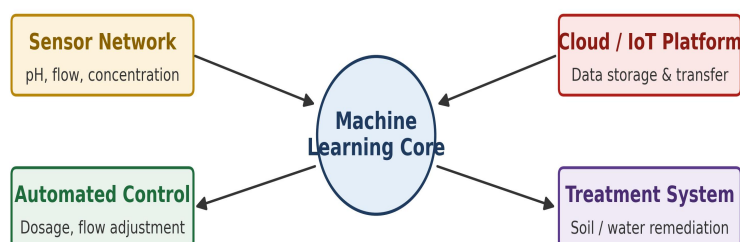


Figure 5: Smart remediation system integrating machine learning, real-time data analysis, and automated control for soil and water heavy metal removal [28,53]

1.8. Machine Learning Models for Metal Remediation

1.8.1. Supervised Learning Models

Supervised machine learning algorithms use labeled data to predict heavy-metal removal performance and optimize operating parameters. Linear regression models provide an initial predictive benchmark, while artificial neural networks capture nonlinear relationships more effectively. Support vector machines and ensemble methods, such as random forests and gradient boosting, further improve predictive reliability. A review of machine learning applications in biochar evaluation found ensemble algorithms particularly effective for Modelling complex adsorption systems [34,35], and random forest models have accurately and stably predicted heavy metal removal efficiency in zeolite-embedded systems, outperforming traditional empirical correlations by capturing multidimensional relationships among variables [36].

1.8.2. Unsupervised Learning Models

Unsupervised learning methods, including K-means and hierarchical clustering, identify patterns in unlabelled environmental data, while principal component analysis reduces dimensionality while preserving data variance. These methods support the classification and identification of contamination sources and provide an exploratory way to analyze large, complex datasets without prior knowledge of outcomes, improving the detection of contamination hotspots and informing targeted remediation planning. In a large-scale field study of cultivated soils, unsupervised source-apportionment techniques combined with random forest Modelling were used alongside 577 soil samples to identify the dominant natural and anthropogenic drivers of heavy-metal accumulation and to map the spatial correlation between ecological risk and environmental factors, demonstrating that clustering-based approaches remain a practical complement to supervised prediction when the objective is source discovery rather than direct forecasting [37].

1.8.3. Deep Learning Approaches

Deep neural networks and hybrid machine learning systems provide advanced tools for Modelling highly nonlinear, multidimensional environmental data. Convolutional neural networks (CNNs) support the analysis of spatially distributed contaminant data and remote sensing imagery, and graph neural network architectures that explicitly encode spatial dependency between sampling locations have been shown to outperform conventional spatial-interpolation and standard neural network baselines when mapping soil heavy-metal distributions across an industrial river basin [38]. Analyses of large-scale remediation case studies demonstrate the scalability of diverse machine learning models for environmental decision-making [39], and deep learning methods that combine multiple data types and formats support the development of more effective decision-making tools for large-scale environmental management.

1.9. Applications of Machine Learning in Metal Remediation

Machine learning has been applied across a broad range of tasks in heavy metal remediation, from predicting removal efficiency to selecting remediation agents and characterizing contaminated sites. This section discusses these applications in more detail, with reference to representative studies from the reviewed literature.

1.9.1. Prediction of Heavy Metal Removal Efficiency

Accurately anticipating the effectiveness of a given remediation method allows engineers to plan treatment systems with greater confidence and to avoid costly overdesign or underperformance. Machine learning techniques applied to predicting heavy metal removal from liquid streams using polymer inclusion membranes (PIMs) have been shown to outperform traditional regression-based analyses in predictive accuracy [40]. Similarly, ensemble and neural network models trained on adsorption isotherm and breakthrough data have been used to forecast removal efficiency across a range of adsorbents, including biochar, zeolites, and activated carbon, under varying influent concentrations and operating conditions. By reducing dependence on exhaustive experimental trials, predictive Modelling shortens development timelines and lowers the resource burden of process optimization.

1.9.2. Optimization of Operational Parameters

Machine learning algorithms are increasingly used to identify optimal operating conditions, such as solution pH, temperature, adsorbent dosage, and contact time, for a given remediation system. Coupling predictive models with sensitivity analysis allows researchers to quantify the relative influence of each operational variable on removal performance, which supports targeted process control rather than uniform adjustment of all parameters. This approach has been applied to capacitive deionization systems, where machine learning-guided optimization improved both removal efficiency and energy use, and to electrokinetic soil remediation, where interpretable models clarified which process variables most strongly influenced metal removal. Optimizing operational parameters this way reduces resource consumption, minimizes secondary waste generation, and improves the overall efficiency of remediation systems.

1.9.3. Selection of Suitable Remediation Agents

Selecting an appropriate remediation agent, whether an adsorbent material, chemical reagent, or biological agent, traditionally requires extensive comparative laboratory testing. Machine learning-based comparative Modelling lets researchers screen candidate materials computationally before committing to laboratory validation, narrowing the experimental search space. Feature-based models trained on the physicochemical properties of biochar, for example, have been used to predict adsorption performance across different biochar formulations, supporting the selection of materials best suited to specific metal contaminants and site conditions [41]. This screening capability is particularly valuable in

review-level assessments, where a wide range of candidate materials must be compared across heterogeneous experimental studies.

1.9.4. Modelling Adsorption and Kinetic Behaviour

Adsorption isotherms (e.g., Langmuir and Freundlich models) and kinetic models (e.g., pseudo-first-order and pseudo-second-order) have traditionally been fitted using linear or nonlinear regression. Machine learning approaches extend this capability by capturing nonlinear, multi-variable interactions among metal ions, adsorbent surface properties, and solution chemistry that classical isotherm and kinetic models cannot represent. A review of machine learning applications for simulating heavy metal adsorption in soil and water systems highlights these methods' ability to model complex interactions between metals and matrix materials more flexibly than conventional kinetic frameworks [42]. This capability is particularly valuable for scaling laboratory-derived kinetic parameters to field-scale or multi-contaminant systems, where classical models often perform poorly.

1.9.5. Risk Assessment and Site Characterization

Predictive machine learning frameworks support contamination mapping, exposure assessment, and site characterization by integrating spatial, chemical, and hydrological data. For example, machine learning-based mapping of soil adsorption capacity at regional and global scales has been used to estimate spatial variation in heavy metal retention, supporting risk assessment across large and data-sparse geographic areas [43]. Such frameworks help regulators and site managers prioritize remediation efforts and allocate resources to the highest-risk locations, improving decision-making efficiency in large-scale environmental management programs.

1.10. Machine Learning in Bioremediation and Phytoremediation

1.10.1 Prediction of Microbial Performance

Machine learning models have been used to gain insight into the interactions between heavy metals, biochar, and microbial communities [44], improving predictions of microbial removal effectiveness and supporting the optimization of microbial remediation procedures, including the selection of microbial consortia and the tuning of nutrient and environmental conditions that maximize degradation or immobilization performance.

1.10.2. Plant–Metal Interaction Modelling

Plant-metal interaction Modelling addresses how plants take up, translocate, and stabilize heavy metals within their tissues, and how these processes can be predicted and optimized computationally. Because plant uptake depends on many interacting variables, including root architecture, rhizosphere microbial activity, soil pH, and metal speciation, this process is well suited to machine learning approaches that can capture nonlinear, multi-variable relationships that classical dose-response models cannot represent. Predictive models of plant uptake typically use soil physicochemical properties (pH, organic matter content, cation exchange capacity), metal speciation, and plant physiological traits as input features to estimate bioconcentration and translocation factors. Regression-based and ensemble models have been applied to predict metal accumulation in plant tissues from soil metal concentrations and soil properties, allowing researchers to estimate phytoextraction potential without exhaustive pot- or field-trial testing. Random forest and gradient boosting algorithms, in particular, can rank the relative importance of soil and plant variables, helping identify which factors most strongly limit or enhance uptake in a given system.

Machine learning has also been applied to model rhizosphere and root-soil interactions, including how root exudates and associated microbial communities influence metal bioavailability. Because plant-growth-promoting rhizobacteria and mycorrhizal fungi can alter metal solubility and plant stress tolerance, coupling microbial and plant uptake data within a single predictive framework allows researchers to model bioremediation and phytoremediation as an integrated biological system rather than as separate processes. This integrated Modelling approach helps identify plant-microbe combinations that maximize stabilization or extraction efficiency for a given contaminant and soil type.

Another application of plant-metal interaction Modelling is identifying hyperaccumulator or tolerant plant species suited to specific contamination profiles. By training classification models on datasets of plant tolerance and accumulation traits across multiple heavy metals, researchers can predict candidate species for phytoremediation trials at a given contaminated site, substantially narrowing the pool of species requiring field validation and accelerating the deployment of phytoremediation programs in previously uncharacterized soils.

1.10.3. Optimization of Phytoremediation Conditions

Ensemble machine learning models trained on biochar physicochemical properties have improved the accuracy of predicting heavy metal adsorption performance across different biochar formulations, supporting the optimization of phytoremediation conditions in ways that enhance both sustainability and cost-effectiveness. More directly, machine learning models trained on more than 20 interacting variables spanning plant characteristics, soil physicochemical properties, and metal speciation have been used to model phytoremediation performance in heavy-metal-contaminated soils, enabling the identification of the critical factors that govern uptake and offering a data-driven basis for selecting plant species and adjusting soil amendments before field trials begin [45].

1.10.4. Reduction of Experimental Cost and Time

By reducing dependence on extensive laboratory and field trials, predictive Modelling shortens the experimental timelines associated with bioremediation and phytoremediation research, allowing researchers to prioritize the most

promising experimental conditions identified computationally before committing to costly field-scale validation, an efficiency gain that has been demonstrated across microbial, algal, and plant-based remediation systems where machine-learning-guided screening substantially narrows the experimental search space.

Implications for Remediation Strategies

1.10.5. Shift from Empirical to Predictive Frameworks

Advances in machine learning for water treatment reflect a broader shift toward predictive, data-driven remediation planning [46], enabling more proactive management of environmental systems.

1.10.6. Identification of Critical Controlling Factors

AI-based optimization of river bioremediation has enabled identification of the dominant environmental variables governing pollutant removal [47]. Understanding these controlling factors supports more targeted and effective remediation system design.

1.10.7. Resource-Efficient Remediation

Integrating automation with predictive tools supports the development of adaptive, resource-efficient remediation systems that reduce secondary pollution and can adjust to varying scales of contamination, from localized industrial spills to diffuse agricultural pollution, echoing evidence that AI-guided process control lowers reagent and energy consumption in wastewater heavy-metal treatment while maintaining removal performance [48].

1.11. Challenges and Limitations

1.11.1. Data Availability and Quality Issues

Predicting soil adsorption capacity at regional and global scales requires standardized and extensive datasets, which remain limited. Current mapping efforts highlight the need for improved data collection protocols and more rigorous validation of environmental datasets.

1.11.2. Model Interpretability

The limited transparency of complex machine learning models, particularly ensemble methods and deep neural networks, remains a barrier to regulatory approval and to building trust among stakeholders and end users, particularly where remediation decisions carry significant financial or public health implications. Explainable artificial intelligence (XAI) techniques are increasingly used to address this barrier: SHapley Additive exPlanations (SHAP) and related feature-attribution methods quantify the contribution of individual input variables, such as pH, adsorbent surface area, or contact time, to a model's predicted removal efficiency, converting an otherwise opaque prediction into a mechanistically interpretable output that can be checked against known physicochemical behaviour [49]. Applying SHAP-based interpretability to adsorption-based heavy metal remediation has, for example, allowed researchers to confirm that predicted removal efficiencies are driven by physically plausible variables rather than dataset artifacts. However, reviewers caution that SHAP importance rankings can be sensitive to feature correlation and dataset composition and should therefore be interpreted alongside, rather than as a replacement for, mechanistic understanding. Embedding such interpretability methods into remediation Modelling pipelines is increasingly viewed as a precondition for regulatory and stakeholder acceptance of machine-learning-based decision support.

1.11.3. Overfitting and Generalization Problems

Many reviewed studies report limited generalisability when models trained in one geographic or environmental context are applied elsewhere, underscoring the need to establish model robustness across diverse geographical settings and environmental conditions.

1.11.4. Computational Constraints

High-dimensional, multi-source environmental datasets demand substantial computational resources, which can limit access to advanced machine learning methods for smaller research groups or resource-constrained settings. This concern is echoed in recent benchmarking of boosting algorithms for environmental and energy sensor analytics, where the computational cost of repeated retraining under realistic deployment conditions, including calibration drift and temporal shift, was identified as a major barrier to sustained real-world use rather than a one-off training expense [50]. More broadly, training and running complex AI models demands substantial computational infrastructure that smaller organizations and lower-resource settings often lack, a disparity that risks concentrating advanced remediation-Modelling capability among well-resourced institutions unless lighter-weight or shared computational approaches are developed [51].

1.11.5. Lack of Standardized Environmental Datasets

The absence of standardized environmental datasets and measurement protocols constrains the transferability and reproducibility of machine learning models internationally. Efforts to consolidate heterogeneous, literature-derived measurements into unified, publicly documented datasets, such as a recently compiled global soil heavy-metal dataset spanning 9,489 field observations from 56 countries, illustrate both the feasibility and the labour intensity of building genuinely standardized resources, and highlight the continued reliance on freely available geospatial layers such as SoilGrids to fill gaps where field measurements remain sparse. Broader adoption of shared data standards, harmonized sampling protocols, and open repositories would substantially improve the cross-study comparability that heavy-metal

remediation research currently lacks. Figure 6 summarises the principal challenges facing machine learning applications in heavy metal remediation, including data quality, model interpretability, overfitting, and computational constraints.

Key Challenges in Machine Learning Applications for Heavy Metal Remediation

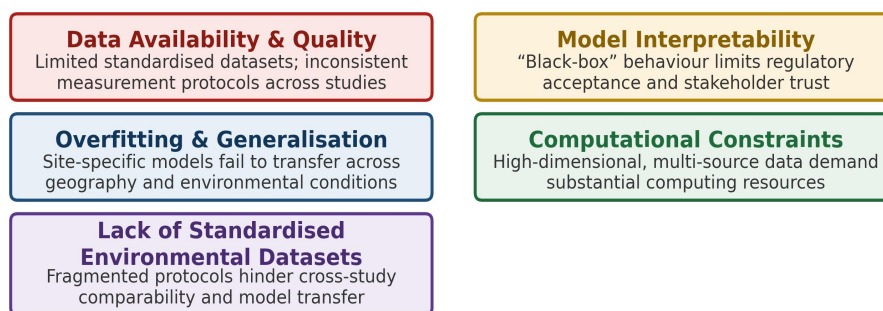


Figure 6: Challenges in machine learning applications for heavy metal remediation[48,43]

1.12. Future Perspectives

1.12.1. Integration with IoT and Smart Sensors

Artificial intelligence is playing an increasingly important role in wastewater treatment and environmental protection [52], and the growing integration of Internet of Things (IoT) technologies enables organizations to acquire real-time environmental data at unprecedented scale.

1.12.2. Real-Time Adaptive Remediation Systems

Real-time adaptive remediation systems represent an emerging paradigm in environmental management, in which sensors, machine learning algorithms, and automated control systems operate together to provide adaptive control over treatment processes. Algorithms such as artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), and support vector machines (SVM) have been used to predict operating conditions, including pH, adsorbent dosage, contact time, and temperature, that maximize removal efficiency in adsorption-based treatment systems. Coupling predictive models with real-time data, such as influent concentration, flow rate, and physicochemical parameters, allows treatment systems to adjust operational parameters to sustain optimal removal performance dynamically. AI-based automated systems of this kind reduce the need for manual intervention, improve treatment reliability, and enhance resource efficiency [53].

1.12.3. Explainable Artificial Intelligence (XAI)

Explainable artificial intelligence (XAI) techniques are expected to play a growing role in improving model transparency and building stakeholder confidence in machine learning-based remediation decisions. Machine learning can transform metal remediation from a largely experimental, trial-and-error process into a data-driven, predictive, and optimized technology. Integrating it with environmental engineering, nanotechnology, biotechnology, and XAI can support the development of more efficient, cost-effective, and sustainable strategies for remediating heavy-metal-contaminated environments. An increasingly important development is integrating Explainable Artificial Intelligence (XAI) with ML-based remediation models. While conventional ML models can provide highly accurate predictions, their decision-making processes may be difficult to interpret. XAI techniques can identify which variables, such as pH, surface functional groups, initial metal concentration, or contact time, most influence remediation efficiency. It improves scientific understanding, model transparency, and confidence in applying ML predictions to real environmental systems. An increasingly important development is integrating Explainable Artificial Intelligence (XAI) with ML-based remediation models. While conventional ML models can provide highly accurate predictions, their decision-making processes may be difficult to interpret. SHAP (Shapley Additive exPlanations) is an XAI method based on Shapley values from cooperative game theory that explains individual predictions by quantifying each input feature's contribution to the predicted outcome. In heavy-metal remediation, SHAP can determine how variables such as pH, surface functional groups, initial metal concentration, adsorbent dosage, and contact time contribute to the predicted removal efficiency. Positive and negative SHAP values indicate whether a variable increases or decreases the model prediction, respectively, providing insight into the factors controlling remediation performance. This approach improves scientific understanding, model transparency, and confidence in applying ML predictions to real environmental systems [54].

1.12.4. Policy and Regulatory Integration

Regulatory frameworks that align AI-driven environmental management with existing environmental standards remain underdeveloped and represent an important direction for future policy research. A comparative analysis of global AI regulations found that existing environmental governance structures were largely developed before AI-based monitoring and decision-support tools became widespread, leaving significant gaps around algorithmic accountability, data privacy, and transparency in automated environmental decision-making, and recommended adaptive governance frameworks built around transparency, inclusiveness, and cross-institutional accountability as AI becomes further embedded in remediation planning [55].

1.12.5. Development of Global Environmental Databases

The development of standardized, internationally shared environmental databases will require coordinated efforts among research institutions and regulatory bodies across countries. Recent large-scale efforts demonstrate what this coordination can achieve: by systematically compiling heavy-metal fraction data from 219 published sources spanning 56 countries and combining it with open global soil-property layers, researchers have produced globally applicable predictive maps of contamination mobility that would be impossible to generate from any single national dataset, underscoring the scientific payoff of continued investment in shared, harmonized environmental data infrastructure. Future research should prioritize explainable AI, the development of global environmental databases, integration of real-time monitoring, alignment of regulatory policy, and the application of advanced deep learning techniques, as summarised in Figure 7.

Future Research Directions in ML-Driven Heavy Metal Remediation

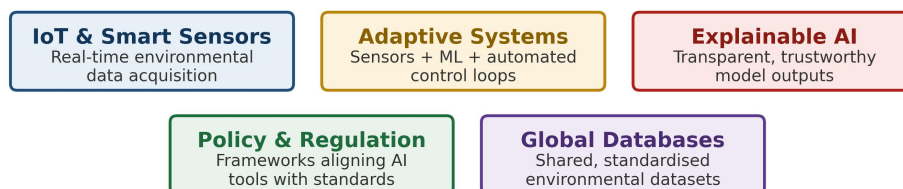


Figure 7: Future research directions in ML-driven heavy metal remediation [28,53]

2.0. Conclusion

This review shows that machine learning and data science are reshaping heavy metal remediation by shifting from empirical, trial-and-error methods to predictive, data-driven decision-making. Supervised, unsupervised, and deep learning approaches have consistently demonstrated value in predicting removal efficiency, optimizing operational parameters, and Modelling adsorption and kinetic behavior across chemical, physical, and biological remediation systems. Integrating machine learning with bioremediation and phytoremediation, particularly through plant-metal interaction Modelling and microbial performance prediction, offers a promising pathway toward more sustainable, cost-effective, and scalable remediation strategies. However, practical deployment remains constrained by limited, non-standardized datasets; limited model interpretability; overfitting when models transfer across geographic and environmental contexts; and substantial computational demands. Addressing these barriers will require coordinated efforts to develop standardized environmental databases, advance explainable AI methods that build regulatory and stakeholder trust, and integrate real-time IoT-enabled monitoring with adaptive, automated treatment systems. As these challenges are addressed, the continued convergence of computational intelligence and environmental science is expected to support remediation strategies that are more sustainable, resilient, and responsive to real-world environmental variability.

Acknowledgment

The authors gratefully acknowledge the Department of Chemistry, University of Agriculture, Faisalabad, for institutional support during the preparation of this review.

Author's Contribution

Faisal Latif conceived the review and drafted the initial manuscript. Muhammad Bilal and Mahnoor Saeed conducted the literature synthesis and prepared the tables and figures. Muhammad Hasnain contributed substantially to the major revision of the manuscript, including incorporating reviewers' comments, updating the literature and references, and improving relevant scientific sections. Muhammad Ali and Huzaifa contributed to revising the manuscript, checking the literature and references, and improving relevant sections. Raziya Nadeem supervised the study, critically revised the manuscript for intellectual content, and approved the final version for submission. All authors read and approved the final manuscript.

Conflict of Interest

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this review.

Funding Resource

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Hao, L., Zhang, D., Zhou, H., Wang, Z., Zhang, J., Zhao, Z., & Li, M. (2025). Machine learning applications in prediction and optimization of heavy metal bioremediation: A review. *Journal of Environmental Chemical Engineering*, 13(5), 118845. <https://doi.org/10.1016/j.jece.2025.118845>
2. Ju, C., Xu, X., Wang, Q., Park, J., Meng, L., Ruan, Z., & Zhou, L. (2025). Application of machine learning methods to predict the immobilization rate of heavy metal contaminated soils by alkaline solid waste. *Journal of Environmental Chemical Engineering*, 13(2), 116064. <https://doi.org/10.1016/j.jece.2025.116064>

3. Barkhordari, M. S., Zhou, N., Li, K., & Qi, C. (2024). Interpretable machine learning for predicting heavy metal removal efficiency in electrokinetic soil remediation. *Journal of Environmental Chemical Engineering*, 12(6), 114330. <https://doi.org/10.1016/j.jece.2024.114330>
4. Barkhordari, M. S., & Qi, C. (2025). Integrating machine learning and reliability analysis: A novel approach to predicting heavy metal removal efficiency using biochar. *Ecotoxicology and Environmental Safety*, 299, 118381. <https://doi.org/10.1016/j.ecoenv.2025.118381>
5. Jadaa, W., & Mohammed, H. K. (2023). Heavy metals – Definition, natural and anthropogenic sources of releasing into ecosystems, toxicity, and removal methods – An overview study. *Journal of Ecological Engineering*, 24(6), 249–271. <https://doi.org/10.12911/22998993/162955>
6. More, S., & Dhakate, R. (2025). Geogenic and anthropogenic sources of heavy metals in soil: An ecological and health risk assessment in the granitic terrain of South India. *CATENA*, 254, Article 108960. <https://doi.org/10.1016/j.catena.2025.108960>
7. Soltani-Gerdefaramarzi, S., Ghasemi, M., & Ghanbarian, B. (2021). Geogenic and anthropogenic sources identification and ecological risk assessment of heavy metals in the urban soil of Yazd, central Iran. *PLOS ONE*, 16(11), e0260418. <https://doi.org/10.1371/journal.pone.0260418>
8. El Shafie, M., Mubarak, M. F., Nasr, M., Shaltout, A., & El Shahawy, A. (2025). Application of machine learning in predicting heavy metal uptake by activated carbon adsorbents. *Advanced Sciences and Technology Journal*, 2(3), 1–20. <https://doi.org/10.21608/astj.2025.397523.1080>
9. Miller, T., Cembrowska-Lech, D., Kisiel, A., Krzemińska, A., Kozłowska, P., Jawor, M., Kołodziejczak, M., & Durlík, I. (2023). Harnessing AI for environmental resilience: Mitigating heavy metal pollution and advancing sustainable practices in diverse spheres. *Grail of Science*, (26), 151–156. <https://doi.org/10.36074/grail-of-science.14.04.2023.027>
10. Obadimu, C. O., Ekwere, I. O., Shaibu, S. E., Essien, U. B., Adelagun, R. O. A., & Adewusi, S. G. (2025). Application of a machine learning framework on heavy metals in the coastal environment. *International Journal of Advanced Research in Chemical Science*, 12(1), 21–29. <https://doi.org/10.20431/2349-0403.1201003>
11. Mustapha, L. S., Obayomi, O. V., & Obayomi, K. S. (2026). A comprehensive review on potential heavy metals in the environment: Persistence, bioaccumulation, ecotoxicology, and agricultural impacts. *Ecological Frontiers*, 46(2), 434–449. <https://doi.org/10.1016/j.ecofro.2025.10.009>
12. Briseño-Bugarín, J., Araujo-Padilla, X., Escot-Espinoza, V. M., Cardoso-Ortiz, J., Flores de la Torre, J. A., & López-Luna, A. (2024). Lead (Pb) pollution in soil: A systematic review and meta-analysis of contamination grade and health risk in Mexico. *Environments*, 11(3), Article 43. <https://doi.org/10.3390/environments11030043>
13. Hou, D., Jia, X., Wang, L., McGrath, S. P., Zhu, Y.-G., Hu, Q., Zhao, F.-J., Bank, M. S., O'Connor, D., & Nriagu, J. (2025). Global soil pollution by toxic metals threatens agriculture and human health. *Science*, 388(6744), 316–321. <https://doi.org/10.1126/science.adr5214>
14. Kumar, A., Kumar, V., Bakshi, P., Parihar, R. D., Radziemska, M., & Kumar, R. (2024). Mercury in the natural environment: Biogeochemical cycles and associated health risks. *Journal of Geochemical Exploration*, 267, Article 107594. <https://doi.org/10.1016/j.gexplo.2024.107594>
15. Gheibi, M., Masoomi, S. R., Uzairu Magala, M., Fathollahi-Fard, A. M., Ghazikhani, A., & Behzadian, K. (2024). The application of artificial intelligence (AI) in adsorption process of heavy metals: A systematic review. *Environmental Industry Letters*, 2(2), 57–78. <https://doi.org/10.15157/eil.2024.2.2.57-78>
16. Punniyakotti Varadharajan, G., Ramar, K., Jayaseelan, A., Packiyadhas, P., Narasiman, N., PanneerSelvam, S., Subbaiyan, N., Kannappan Panchamoorthy, G., Palaniappan, S. K., & Siengchin, S. (2026). AI and ML in mitigating membrane fouling in heavy metal wastewater treatment: A review on recent trends and future industrial outlook. *Archives of Computational Methods in Engineering*, 33(3), 4217–4236. <https://doi.org/10.1007/s11831-025-10434-1>
17. Wuana, R. A., & Okieimen, F. E. (2011). Heavy metals in contaminated soils: A review of sources, chemistry, risks and best available strategies for remediation. *ISRN Ecology*, 2011, Article 402647. <https://doi.org/10.5402/2011/402647>
18. Liu, L., Li, W., Song, W., & Guo, M. (2018). Remediation techniques for heavy metal-contaminated soils: Principles and applicability. *Science of the Total Environment*, 633, 206–219. <https://doi.org/10.1016/j.scitotenv.2018.03.161>
19. Sierra, M. J., Millán, R., López, F. A., Alguacil, F. J., & Cañadas, I. (2016). Sustainable remediation of mercury contaminated soils by thermal desorption. *Environmental Science and Pollution Research*, 23(5), 4898–4907. <https://doi.org/10.1007/s11356-015-5688-8>
20. Castro, K., & Abejón, R. (2024). Removal of heavy metals from wastewaters and other aqueous streams by pressure-driven membrane technologies: An outlook on reverse osmosis, nanofiltration, ultrafiltration and microfiltration potential from a bibliometric analysis. *Membranes*, 14(8), Article 180. <https://doi.org/10.3390/membranes14080180>

21. Dian, X., Hao, J., Zhang, Z., Chen, Z., & Yao, L. (2024). Heavy metal removal performance of capacitive deionization technology studied by machine learning. *Engineering Research Express*, 6(3), 035002. <https://doi.org/10.1088/2631-8695/ad612c>
22. Fu, F., & Wang, Q. (2011). Removal of heavy metal ions from wastewaters: A review. *Journal of Environmental Management*, 92(3), 407–418. <https://doi.org/10.1016/j.jenvman.2010.11.011>
23. Kumpiene, J., Lagerkvist, A., & Maurice, C. (2008). Stabilization of As, Cr, Cu, Pb and Zn in soil using amendments—A review. *Waste Management*, 28(1), 215–225. <https://doi.org/10.1016/j.wasman.2006.12.012>
24. Kumar, S., Banerjee, S., Ghosh, A., Pramanik, C., Dey, I., Choudhury, N., Sarkar, S., Das, A., Ghosh, S., Samanta, S., Das, S., Chattopadhyay, N., Bag, M., Talukdar, P., & Das, S. (2026). Biological synergies for heavy metal bioremediation across microbial, algal, and phytoremediation perspectives. *Discover Green Chemistry*, 1, Article 3. <https://doi.org/10.1007/s44509-026-00003-5>
25. Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., Ma, X., Marrone, B. L., Ren, Z. J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B. M., Xiao, X., Yu, X., Zhu, J.-J., & Zhang, H. (2021). Machine learning: New ideas and tools in environmental science and engineering. *Environmental Science & Technology*, 55(19), 12741–12754. <https://doi.org/10.1021/acs.est.1c01339>
26. Liu, X., Lu, D., Zhang, A., Liu, Q., & Jiang, G. (2022). Data-driven machine learning in environmental pollution: Gains and problems. *Environmental Science & Technology*, 56(4), 2124–2133. <https://doi.org/10.1021/acs.est.1c06157>
27. Li, X., Yang, Y., Yang, J., Fan, Y., Qian, X., & Li, H. (2021). Rapid diagnosis of heavy metal pollution in lake sediments based on environmental magnetism and machine learning. *Journal of Hazardous Materials*, 416, 126163. <https://doi.org/10.1016/j.jhazmat.2021.126163>
28. Miller, T., Durlík, I., Kostecka, E., Kozłowska, 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>
29. Wu, J., & Zhao, F. (2023). Machine learning: An effective technical method for future use in assessing the effectiveness of phosphorus-dissolving microbial agromediation. *Frontiers in Bioengineering and Biotechnology*, 11, 1189166. <https://doi.org/10.3389/fbioe.2023.1189166>
30. Obadimu, C. O., Shaibu, S. E., Ekwere, I. O., & Adelagun, R. O. A. (2024). Machine learning-based forecasting of bioaccumulation and histopathological effects in aquatic organisms. *FUDMA Journal of Sciences*, 8(6), 485–496. <https://doi.org/10.33003/fjs-2024-0806-3002>
31. Yan, Z., Li, J., Zhang, W., Wen, H., Yu, H., Yu, M., Liu, Q., & Jiang, G. (2026). A tutorial on best practices and pitfalls in applying machine learning to environmental research. *ACS Environmental Au*, 6(4), 553–565. <https://doi.org/10.1021/acsenvironau.6c00025>
32. Yaseen, Z. M., Doost, Z. H., Khan, R., Abdulraheem, A., Abdulameer, S. F., Falah, M. W., & Farooque, A. A. (2025). Chitosan-based flocculant heavy metal removal prediction using machine learning models. *ACS Omega*, 10(40), 46714–46734. <https://doi.org/10.1021/acsomega.5c04000>
33. Hu, T., Wu, M., Chen, Q., Chai, L., & Qi, C. (2026). Machine learning uncovers dominant fractions of heavy metal(loid)s in global soils. *Communications Earth & Environment*, 7, Article 214. <https://doi.org/10.1038/s43247-026-03221-8>
34. Chen, M.-W., Chang, M.-S., Mao, Y., Hu, S., & Kung, C.-C. (2023). Machine learning in the evaluation and prediction models of biochar application: A review. *Science Progress*, 106(1). <https://doi.org/10.1177/00368504221148842>
35. Wang, C., Zhao, Y., Gao, Y., Chen, H., Li, X., Zhou, B., Fan, D., Fang, Z., & Liu, J. (2024). Interpretable machine learning for predicting heavy metal removal and optimizing biochar characteristics. *Journal of Water Process Engineering*, 68, Article 106484. <https://doi.org/10.1016/j.jwpe.2024.106484>
36. Takarina, N., Matsue, N., Johan, E., Adiwibowo, A., Rahmawati, M., Pramudyawardhani, S., & Wukirsari, T. (2024). Machine learning using random forest to model heavy metals removal efficiency using a zeolite-embedded sheet in water. *Global Journal of Environmental Science and Management*, 10(1), 321–336. <https://doi.org/10.22034/gjesm.2024.01.20>
37. Bi, Z., Sun, J., Xie, Y., Gu, Y., Zhang, H., Zheng, B., Ou, R., Liu, G., Li, L., Peng, X., Gao, X., & Wei, N. (2024). Machine learning-driven source identification and ecological risk prediction of heavy metal pollution in cultivated soils. *Journal of Hazardous Materials*, 476, Article 135109. <https://doi.org/10.1016/j.jhazmat.2024.135109>
38. Zha, Y., & Yang, Y. (2024). Innovative graph neural network approach for predicting soil heavy metal pollution in the Pearl River Basin, China. *Scientific Reports*, 14, Article 16505. <https://doi.org/10.1038/s41598-024-67175-7>

39. Li, H., Zhou, Z., Long, T., Wei, Y., Xu, J., Liu, S., & Wang, X. (2022). Big-data analysis and machine learning based on oil pollution remediation cases from CERCLA database. *Energies*, 15(15), 5698. <https://doi.org/10.3390/en15155698>
40. Yaqub, M., Eren, B., & Eyupoglu, V. (2021). Prediction of heavy metals removal by polymer inclusion membranes using machine learning techniques. *Water and Environment Journal*, 35, 1073–1084. <https://doi.org/10.1111/wej.12699>
41. Yaseen, Z. M., & Alhalimi, F. L. (2025). Heavy metal adsorption efficiency prediction using biochar properties: A comparative analysis for ensemble machine learning models. *Scientific Reports*, 15, 13434. <https://doi.org/10.1038/s41598-025-96271-5>
42. Yaseen, Z. M. (2021). An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. *Chemosphere*, 277, 130126. <https://doi.org/10.1016/j.chemosphere.2021.130126>
43. Yang, H., Huang, K., Zhang, K., Weng, Q., Zhang, H., & Wang, F. (2021). Predicting heavy metal adsorption on soil with machine learning and mapping global distribution of soil adsorption capacities. *Environmental Science & Technology*, 55(20), 14316–14328. <https://doi.org/10.1021/acs.est.1c02479>
44. Wei, X., Liu, Y., Shen, L., et al. (2024). Machine learning insights in predicting heavy metals interaction with biochar. *Biochar*, 6, 10. <https://doi.org/10.1007/s42773-024-00304-7>
45. Shi, L., Li, J., Palansooriya, K. N., Chen, Y., Hou, D., Meers, E., Tsang, D. C. W., Wang, X., & Ok, Y. S. (2023). Modelling phytoremediation of heavy metal contaminated soils through machine learning. *Journal of Hazardous Materials*, 441, Article 129904. <https://doi.org/10.1016/j.jhazmat.2022.129904>
46. Yuan, X., Li, J., Lim, J. Y., Zolfaghari, A., Alessi, D. S., Wang, Y., Wang, X., & Ok, Y. S. (2024). Machine learning for heavy metal removal from water: Recent advances and challenges. *ACS ES&T Water*, 4(3), 820–836. <https://doi.org/10.1021/acsestwater.3c00215>
47. Blessing, A.-A., & Olateru, K. (2025). AI-driven optimization of bioremediation strategies for river pollution: A comprehensive review and future directions. *Frontiers in Microbiology*, 16, 1504254. <https://doi.org/10.3389/fmicb.2025.1504254>
48. Bhagat, S. K., Tung, T. M., & Yaseen, Z. M. (2020). Development of artificial intelligence for Modelling wastewater heavy metal removal: State of the art, application assessment and possible future research. *Journal of Cleaner Production*, 250, 119473. <https://doi.org/10.1016/j.jclepro.2019.119473>
49. Khajavian, M., Jang, J.-H., Kwon, J.-Y., Lee, J.-M., Lee, S., Hwang, M.-H., Yang, E., Jang, J. K., & Chae, K.-J. (2026). Harnessing interpretable machine learning: SHapley additive exPlanations (SHAP)-driven insights, transformative impact, and controversies in adsorption-based environmental remediation. *Inorganic Chemistry Communications*, 186, Article 116269. <https://doi.org/10.1016/j.inoche.2026.116269>
50. Xie, Y., & Tummala, S. P. (2025). Machine learning for sensor analytics: A comprehensive review and benchmark of boosting algorithms in healthcare, environmental, and energy applications. *Sensors*, 25(23), Article 7294. <https://doi.org/10.3390/s25237294>
51. Olawade, D. B., Wada, O. Z., Ige, A. O., Egbewole, B. I., Olojo, A., & Oladapo, B. I. (2024). Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions. *Hygiene and Environmental Health Advances*, 12, Article 100114. <https://doi.org/10.1016/j.heha.2024.100114>
52. Malviya, A., & Jaspal, D. (2021). Artificial intelligence as an upcoming technology in wastewater treatment: A comprehensive review. *Environmental Technology Reviews*, 10(1), 177–187. <https://doi.org/10.1080/21622515.2021.1913242>
53. Alam, G., Ihsanullah, I., Naushad, M., & Sillanpää, M. (2022). Applications of artificial intelligence in water treatment for optimization and automation of adsorption processes: Recent advances and prospects. *Chemical Engineering Journal*, 427, 130011. <https://doi.org/10.1016/j.cej.2021.130011>
54. Zhang, et al. (2025). Predictive Modelling of heavy metal lead and cadmium adsorption on biochar based on machine learning. *International Journal of Phytoremediation*, 28(4)748–755. <https://doi.org/10.1080/15226514.2025.2581819>
55. Sahu, S., Mallick, N., & Hossain, I. (2025). Regulating the future: Policy frameworks for AI in environmental monitoring and governance. In M. Azrour, I. Hossain, & A. Haque (Eds.), *Addressing Environmental Challenges With AI, Robotics, and Augmented Reality* (pp. 83–112). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3373-1892-9.ch004>

Received: August 12th 2026Accepted: September 19th 2026