

Review

Not peer-reviewed version

---

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

---

[Faisal Latif](#) , [Muhammad Bilal](#) , Mahnoor Saeed , [Raziya Nadeem](#) \*

Posted Date: 27 February 2026

doi: 10.20944/preprints202602.1893.v1

Keywords: machine learning; data science; heavy metal remediation; environmental remediation; predictive modeling; artificial neural networks; sustainable environmental engineering



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

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

Faisal Latif, Muhammad Bilal, Mahnoor Saeed and Raziya Nadeem \*

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

\* Correspondence: raziyaanalyst@uaf.edu.pk

## Abstract

Heavy metal contamination of soil and water continues to pose persistent environmental and public health challenges, particularly in regions affected by rapid industrialization, mining, and intensive agriculture. Conventional remediation strategies—such as chemical precipitation, adsorption, soil washing, and bioremediation—have contributed significantly to pollution control; however, their implementation often relies on empirical optimization, prolonged experimentation, and site-specific trial-and-error approaches. These limitations restrict scalability, increase operational costs, and slow the development of sustainable remediation solutions. In recent years, the integration of machine learning (ML) and data science has emerged as a transformative direction in environmental engineering, offering predictive, data-driven alternatives to traditional remediation planning. This review critically examines the application of supervised, unsupervised, and deep learning models in metal remediation systems. Emphasis is placed on regression algorithms, artificial neural networks, support vector machines, ensemble techniques, clustering methods, and advanced deep neural architectures that enable prediction of metal removal efficiency, optimization of operational parameters, and modeling of adsorption and kinetic behaviors. The review further explores how data science workflows—including data acquisition, preprocessing, feature engineering, and multi-source data integration—support robust environmental decision-making. Particular attention is given to machine learning applications in bioremediation and phytoremediation, where predictive modeling enhances understanding of microbial performance and plant-metal interactions while reducing experimental time and cost. Despite promising advancements, significant challenges remain, including data scarcity, model interpretability, overfitting risks, lack of standardized environmental datasets, and computational constraints. Addressing these issues will require integration with real-time monitoring systems, Internet of Things (IoT) technologies, explainable artificial intelligence (XAI), and global environmental databases. The review concludes that transitioning from empirical remediation frameworks to predictive and adaptive systems represents a crucial step toward sustainable, scalable, and intelligent heavy metal management strategies. By synthesizing current developments and identifying research gaps, this article provides a comprehensive foundation for future interdisciplinary innovation at the intersection of environmental science, machine learning, and sustainable remediation engineering.

**Keywords:** machine learning; data science; heavy metal remediation; environmental remediation; predictive modeling; artificial neural networks; sustainable environmental engineering

---

## 1. Introduction

Heavy metal contamination has become one of the most pressing environmental concerns due to rapid industrialization, urban expansion, and intensified agricultural practices [1,6]. Toxic metals such as Pb, Cd, Hg, Cr, As, Ni, and Cu are persistent, non-biodegradable, and capable of

accumulating in soil and aquatic systems [6,10]. Unlike organic pollutants, these metals do not degrade into harmless products and can remain in environmental matrices for extended periods [6].

Natural sources of heavy metals include geological weathering, volcanic eruptions, forest fires, and transformation of soil parent materials [20,25]. However, anthropogenic activities such as mining, smelting, fossil fuel combustion, wastewater irrigation, electroplating industries, and excessive fertilizer and pesticide application are the dominant contributors to elevated metal concentrations [1,6,10]. The mobility, chemical speciation, and bioavailability of heavy metals significantly influence their environmental behavior and toxicity [7,11].

Heavy metals pose serious ecological and human health risks due to their toxicity and bioaccumulation potential [6,10]. In soil systems, elevated metal concentrations disrupt microbial diversity, inhibit enzymatic activity, and reduce soil fertility [7,11]. In aquatic ecosystems, metals alter water chemistry and accumulate in organisms, leading to biomagnification across trophic levels [6].

Chronic exposure to toxic metals through contaminated water, food crops, or inhalation can cause severe health disorders. For example, lead affects neurological development, cadmium causes kidney and bone damage, mercury impacts the nervous system, and arsenic is linked to carcinogenic effects [6,10]. Even low-level, long-term exposure can result in cumulative toxicity, making heavy metal pollution a significant global public health issue [6].

Traditional remediation strategies for heavy metal contamination include physical methods (excavation, stabilization/solidification), chemical treatments (precipitation, ion exchange, chemical immobilization), and biological approaches such as bioremediation and phytoremediation [3,4,24]. Although these methods can reduce metal concentrations, they often suffer from limitations, including high operational costs, long treatment durations, and potential secondary pollution [3,24].

Furthermore, remediation efficiency is highly dependent on site-specific parameters such as soil physicochemical properties, pH, temperature, and metal speciation [1,7,11]. Historically, remediation planning has relied on empirical trial-and-error methodologies, which increase uncertainty and experimental cost while limiting scalability and optimization in complex environmental systems [1,10].

The integration of machine learning (ML) and data science has introduced a paradigm shift in environmental metal remediation [1,10,11]. Data-driven frameworks enable the integration of historical experimental datasets and real-time monitoring data to predict remediation performance and support decision-making processes [1,20].

Supervised and unsupervised ML models—including artificial neural networks, support vector machines, random forests, and clustering techniques—have been applied to predict heavy metal removal efficiency, optimize operational parameters, and identify critical controlling factors [1,3,4,23]. These models capture complex nonlinear relationships among environmental variables, reducing experimental time, cost, and uncertainty [1,20,24].

The transition from empirical methodologies to predictive optimization frameworks enhances remediation efficiency, supports sustainable environmental management, and enables scalable smart remediation systems [1,10,11].

This review provides a comprehensive overview of the application of machine learning and data science in heavy metal remediation. It summarizes data-driven workflows, including data acquisition, preprocessing, and integration, and evaluates the role of supervised, unsupervised, and deep learning models in predicting removal efficiency, optimizing operational parameters, and supporting risk assessment. Applications in adsorption processes, bioremediation, and phytoremediation are discussed, along with current challenges such as data limitations and model interpretability. Finally, future perspectives, including IoT integration, explainable artificial intelligence, and the development of standardized environmental databases, are highlighted to guide the advancement of intelligent and sustainable remediation strategies.

## 2. Sources and Characteristics of Heavy Metal Contamination

### 2.1. Natural Sources (Geogenic Processes)

Heavy metals naturally occur in the Earth's crust and are released into the environment through various geogenic processes [7,25]. Weathering of parent rocks, volcanic eruptions, forest fires, and atmospheric deposition contribute to baseline concentrations of metals such as Pb, Cd, Cr, Ni, and As in soils and water bodies [20,25]. The mineralogical composition of bedrock and soil parent materials largely determines the natural background levels of trace metals in a given region [7].

Geochemical processes such as erosion, sediment transport, and groundwater–rock interactions can mobilize these metals into aquatic systems [20]. Although naturally occurring, these geogenic inputs are generally low and become environmentally significant only when amplified by anthropogenic activities [6,25].

### 2.2. Anthropogenic Sources (Industrial, Mining, Agricultural Activities)

Anthropogenic activities are the dominant contributors to elevated heavy metal concentrations in soil and water systems [6,14,24]. Industrial operations such as mining, smelting, electroplating, battery manufacturing, leather tanning, and fossil fuel combustion release substantial quantities of toxic metals into the environment [6,24]. Mining activities, in particular, expose metal-rich ores to oxidation and leaching processes, resulting in acid mine drainage and subsequent contamination of surrounding ecosystems [24].

Urbanization and traffic emissions further contribute to heavy metal accumulation in roadside soils and street dust, particularly Pb, Zn, and Cu [14]. In agricultural systems, the long-term application of phosphate fertilizers, pesticides, sewage sludge, and wastewater irrigation introduces metals such as Cd, As, and Hg into soils [6,25]. These anthropogenic inputs often exceed natural background levels, leading to chronic contamination and ecological imbalance [6,24].

### 2.3. Behavior and Speciation of Heavy Metals in Soil and Water

The environmental behavior and toxicity of heavy metals depend largely on their chemical speciation rather than total concentration [7,11]. In soils, metals may exist in different fractions, including exchangeable, carbonate-bound, Fe–Mn oxide-bound, organic-bound, and residual forms [7]. These fractions influence metal mobility, solubility, and bioavailability.

Factors such as soil pH, redox potential, organic matter content, clay minerals, and cation exchange capacity significantly affect metal speciation and transport [7,11]. In aquatic systems, metals may occur as free ions, complexed species, or adsorbed onto suspended particles [6]. Changes in environmental conditions—such as pH fluctuations or microbial activity—can alter metal speciation and enhance their mobility or toxicity [11].

Understanding speciation is critical for designing effective remediation strategies, as bioavailable fractions pose the greatest ecological and health risks [7].

### 2.4. Persistence and Bioaccumulation

A defining characteristic of heavy metals is their persistence in environmental systems due to their non-biodegradable nature [6]. Unlike organic pollutants, metals cannot be chemically degraded into less toxic forms; instead, they may undergo transformation between oxidation states or binding phases [6,25].

Heavy metals can accumulate in soil over time, especially under continuous anthropogenic input [24]. In aquatic and terrestrial food chains, metals bioaccumulate in organisms and biomagnify at higher trophic levels, increasing exposure risks for wildlife and humans [6]. Chronic exposure, even at low concentrations, may lead to cumulative toxicity due to their long biological half-lives [6,10].

The persistent and bioaccumulative properties of heavy metals underscore the urgent need for efficient and sustainable remediation strategies.

### 3. Conventional Metal Remediation Techniques

#### 3.1. Physical Remediation Methods

Physical remediation techniques are among the earliest approaches used to manage heavy metal contamination. Common methods include soil excavation, landfilling, soil washing, thermal treatment, and solidification/stabilization processes [24]. Excavation involves the removal of contaminated soil followed by off-site disposal or treatment, while soil washing separates metals from soil particles using physical and chemical extractants [24].

Solidification and stabilization techniques aim to immobilize heavy metals within a solid matrix to reduce their mobility and leachability [24]. Although physical methods can provide rapid reduction in contaminant concentration, they are often expensive, energy-intensive, and disruptive to the environment. Additionally, these approaches typically transfer contaminants from one location to another rather than eliminating them completely [24].

#### 3.2. Chemical Treatment Methods

Chemical remediation methods are widely applied for the removal or immobilization of heavy metals from contaminated soil and wastewater. Techniques such as chemical precipitation, coagulation–flocculation, ion exchange, membrane filtration, and adsorption are commonly used in wastewater treatment systems [3,15].

Chemical precipitation converts dissolved metal ions into insoluble forms, which can then be separated from water [3]. Ion exchange resins and membrane technologies allow selective removal of specific metal ions; however, they may require high operational and maintenance costs [3,15]. Adsorption using activated carbon, biochar, or other engineered materials has gained significant attention due to its efficiency and operational simplicity [15].

Despite their effectiveness, chemical treatments may generate secondary sludge, require strict pH control, and involve high reagent consumption, which can increase overall treatment costs and environmental burden [3,15].

#### 3.3. Biological Approaches (Bioremediation and Phytoremediation)

Biological remediation methods utilize microorganisms and plants to remove, stabilize, or transform heavy metals into less toxic forms [4,11].

**Bioremediation** involves the use of bacteria, fungi, or algae capable of bioaccumulation, biosorption, or biotransformation of metals [11]. Microorganisms can alter metal speciation through redox reactions or immobilization processes, thereby reducing toxicity and mobility [11].

**Phytoremediation** employs plants to extract (phytoextraction), stabilize (phytostabilization), or volatilize (phytovolatilization) heavy metals from contaminated soils [4]. Certain hyperaccumulator plant species have demonstrated the ability to accumulate high concentrations of metals in their biomass [4].

Biological approaches are considered environmentally friendly, cost-effective, and suitable for large-scale applications. However, they often require longer remediation periods and are influenced by environmental conditions such as soil properties, climate, and metal bioavailability [4,11].

#### 3.4. Limitations of Traditional Trial-and-Error Approaches

Conventional remediation strategies have historically relied on empirical trial-and-error methodologies to determine optimal treatment conditions [1,3]. This approach requires extensive laboratory experimentation and field validation, which can be time-consuming and costly.

Remediation efficiency varies significantly depending on site-specific parameters such as soil physicochemical properties, metal speciation, pH, temperature, and biological activity [7,11]. The complex interactions among these variables make it difficult to predict outcomes using traditional linear models.

Moreover, conventional methods lack predictive optimization capabilities and real-time adaptive decision-making frameworks [1,10]. As environmental systems are inherently dynamic and multivariate, the absence of data-driven analytical tools limits scalability, cost efficiency, and sustainability. These limitations have paved the way for integrating machine learning and data science approaches into modern remediation strategies [1,20].

## 4. Introduction to Data Science in Environmental Remediation

### 4.1. Concept of Data-Driven Environmental Engineering

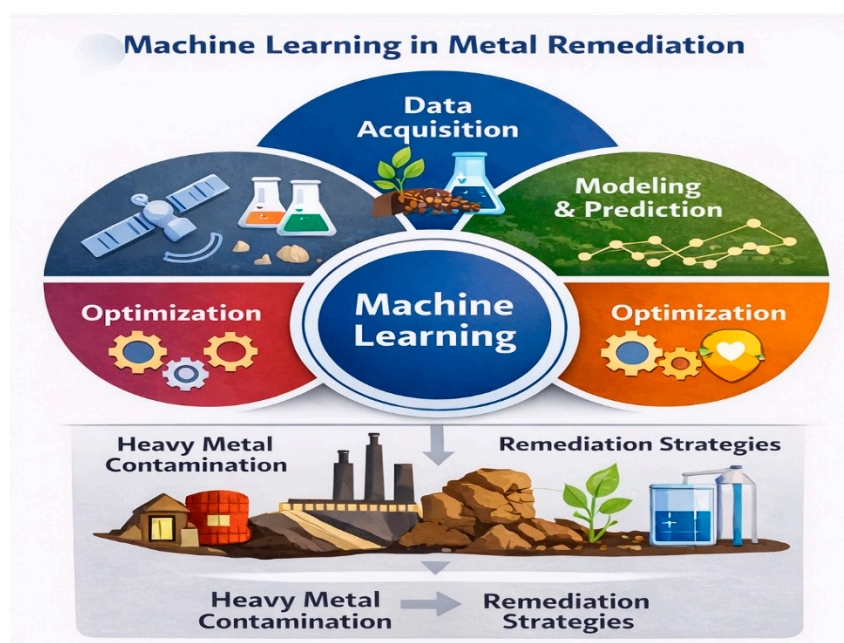
Environmental remediation has traditionally relied on empirical models and laboratory-scale experimentation; however, the growing complexity of contaminated systems has necessitated the adoption of data-driven approaches [1,12]. Data-driven environmental engineering integrates statistical learning, computational modeling, and advanced analytics to interpret complex environmental datasets and support decision-making processes [8,12].

Machine learning (ML) algorithms can capture nonlinear relationships between multiple environmental variables and remediation performance indicators, enabling predictive modeling beyond conventional regression methods [1,17]. Supervised and unsupervised learning techniques allow the identification of critical controlling factors such as pH, temperature, adsorbent dosage, microbial activity, and metal speciation that influence remediation efficiency [1,23].

This paradigm shift from deterministic to predictive frameworks enhances process optimization, reduces experimental uncertainty, and improves overall remediation effectiveness [1,12].

### 4.2. Role of Big Data in Remediation Planning

The expansion of environmental monitoring networks, remote sensing technologies, and laboratory analytical techniques has generated large and heterogeneous datasets commonly referred to as “big data” [20]. These datasets include spatial, temporal, physicochemical, and biological parameters that are essential for understanding contamination patterns and remediation outcomes.



**Figure 1.** Conceptual framework of machine learning integration in heavy metal remediation.

Big data analytics enables the processing and interpretation of high-dimensional datasets to identify contamination hotspots, predict pollutant transport pathways, and evaluate remediation efficiency under varying environmental conditions [20]. Advanced computational techniques such as gradient boosting, random forests, and deep learning algorithms have demonstrated strong performance in modeling complex environmental systems [1,9,22].

The integration of big data analytics into remediation planning supports risk assessment, scenario analysis, and sustainable resource allocation, ultimately improving decision-making accuracy [20].

#### 4.3. Integration of Experimental and Monitoring Data

One of the major strengths of data science in environmental remediation lies in its ability to integrate historical experimental data with real-time monitoring information [1,20]. Laboratory-scale adsorption, bioremediation, and phytoremediation studies generate structured datasets that can be used to train predictive models [3,23].

Simultaneously, field-based monitoring systems provide dynamic information regarding contaminant concentrations, environmental parameters, and treatment performance [20]. By combining these data sources, ML models can enhance predictive reliability and enable adaptive optimization of remediation processes [1,16].

Such integration reduces redundancy in experimentation, shortens project timelines, and facilitates site-specific customization of remediation strategies [1,23].

#### 4.4. Digital Transformation in Environmental Sciences

The adoption of artificial intelligence (AI), machine learning, cloud computing, and open-source analytical tools marks a significant digital transformation in environmental sciences [16,21]. Software platforms and programming libraries such as Scikit-learn enable efficient implementation of predictive models for environmental datasets [21].

Deep learning techniques have further expanded modeling capabilities by analyzing complex spatial and temporal patterns in environmental data [22]. These advancements allow environmental engineers and scientists to transition from static analytical methods toward intelligent, adaptive systems capable of real-time decision support [1,22].

The digital transformation of remediation science promotes transparency, scalability, and sustainability while fostering interdisciplinary collaboration between environmental engineering, data science, and computational modeling [16,20].

## 5. Data Science Workflows in Metal Remediation

The successful application of machine learning in heavy metal remediation depends on a structured and systematic data science workflow. This workflow typically includes data acquisition, feature engineering, model development, validation, and interpretation [12,17]. preprocessing Among these steps, data acquisition and preprocessing are critical, as model performance is highly dependent on data quality and representativeness [1,12].

### 5.1. Data Acquisition and Preprocessing

#### 5.1.1. Sources of Environmental Data

Environmental datasets used in metal remediation studies originate from multiple sources, including laboratory experiments, field surveys, industrial monitoring systems, and public environmental databases [20]. These datasets may contain physicochemical parameters (pH, temperature, conductivity), contaminant concentrations, soil characteristics, hydrological variables, and biological indicators [7,20].

In adsorption and wastewater treatment studies, experimental datasets often include variables such as initial metal concentration, adsorbent dosage, contact time, agitation speed, and removal efficiency [3,23]. The availability of diverse and high-dimensional datasets enables robust machine learning model development but also introduces challenges in data integration and consistency [20].

#### 5.1.2. Laboratory and Field Monitoring Data

Laboratory-generated data provide controlled and reproducible conditions for modeling heavy metal removal processes such as adsorption, ion exchange, and bioremediation [3,23]. These datasets are typically structured and suitable for supervised learning applications, where input variables are mapped to removal efficiency or adsorption capacity.

Field monitoring data, on the other hand, are often dynamic, heterogeneous, and influenced by environmental variability [20]. Real-time monitoring systems may generate temporal datasets that capture fluctuations in contaminant levels and environmental conditions. Integrating laboratory and field data enhances model generalization and predictive accuracy, supporting site-specific remediation planning [1,16].

#### 5.1.3. Data Cleaning and Normalization

Raw environmental data frequently contain inconsistencies, outliers, duplicate records, and measurement errors [12]. Data cleaning involves identifying and correcting these issues to ensure reliability and consistency prior to model development.

Normalization and scaling are essential preprocessing steps, particularly for machine learning algorithms sensitive to variable magnitude differences such as support vector machines and neural networks [2,17]. Techniques such as min-max scaling and standardization help improve convergence rates and model stability [12]. Proper preprocessing ensures that model training is not biased toward variables with larger numerical ranges.

#### 5.1.4. Handling Missing and Noisy Data

Missing and noisy data are common challenges in environmental datasets due to instrument malfunction, sampling limitations, or human error [20]. Incomplete datasets can reduce model accuracy and introduce bias if not properly addressed.

Statistical imputation methods and predictive modeling techniques can be used to estimate missing values, while noise reduction approaches such as smoothing or filtering improve signal clarity [12]. Robust machine learning algorithms, including ensemble methods like random forests, are particularly effective in handling noisy environmental data [1].

Effective treatment of missing and noisy data enhances model robustness and ensures reliable prediction of heavy metal remediation performance.

#### 5.1.5. Feature Engineering and Selection

Feature engineering involves transforming raw environmental variables into meaningful inputs that improve model performance [12]. For example, interaction terms, ratio variables, or derived parameters such as distribution coefficients can enhance predictive capability in adsorption modeling [23].

Feature selection techniques are employed to identify the most influential variables affecting remediation efficiency. Methods such as principal component analysis (PCA), recursive feature elimination, and tree-based importance ranking help reduce dimensionality and prevent overfitting [8,12].

By identifying critical controlling factors—such as pH, metal speciation, microbial activity, and adsorbent properties—feature selection improves interpretability and computational efficiency of machine learning models in metal remediation applications [1,7].

## 5.2. Data Integration and Management

Effective machine learning applications in heavy metal remediation require not only high-quality data but also efficient integration and management of heterogeneous datasets. Environmental remediation projects typically involve multi-source, multi-scale, and multi-format data, necessitating systematic data fusion and database structuring frameworks [20]. Proper data integration enhances model reliability, scalability, and decision-support capabilities [1,16].

### 5.2.1. Multi-source Data Fusion

Heavy metal remediation studies often combine datasets from laboratory experiments, field monitoring campaigns, remote sensing platforms, industrial process controls, and public environmental repositories [20]. These datasets differ in spatial resolution, temporal frequency, measurement techniques, and uncertainty levels.

Multi-source data fusion techniques integrate these heterogeneous datasets into a unified analytical framework, enabling comprehensive system-level understanding [20]. Statistical learning methods and ensemble algorithms can effectively combine diverse input variables to improve predictive performance and reduce model variance [1,9].

Data fusion enhances contamination mapping, risk assessment, and remediation optimization by capturing complex interactions among soil properties, hydrological parameters, contaminant speciation, and treatment variables [7,20]. The integration of structured experimental datasets with unstructured environmental monitoring data strengthens generalization capacity in predictive modeling applications [1,16].

### 5.2.2. Real-Time Monitoring Systems

Advancements in sensor technologies and digital monitoring networks have significantly transformed environmental data acquisition [20]. Real-time monitoring systems generate continuous data streams related to pH, dissolved oxygen, redox potential, metal concentration, temperature, and conductivity.

The integration of real-time monitoring data with machine learning algorithms enables dynamic prediction and adaptive control of remediation systems [1]. For example, supervised learning models such as artificial neural networks and support vector machines can process time-series data to predict removal efficiency under varying operational conditions [2,22].

Real-time analytics facilitate early detection of system inefficiencies, rapid identification of contamination spikes, and proactive decision-making in remediation projects [20]. Such adaptive frameworks reduce operational costs and enhance process sustainability.

### 5.2.3. Database Development for Remediation Projects

A well-structured database is fundamental for managing environmental datasets and supporting reproducible research [12]. Remediation databases typically include physicochemical properties, contaminant concentrations, treatment parameters, and performance indicators derived from both laboratory and field studies [3,23].

The use of standardized data formats and metadata documentation ensures consistency, interoperability, and long-term usability of remediation datasets [20]. Open-source analytical platforms and machine learning libraries facilitate data storage, processing, and model deployment within integrated digital systems [21].

Comprehensive remediation databases enable historical trend analysis, benchmarking of treatment technologies, and development of predictive models for site-specific applications [1,16]. By establishing centralized and scalable data management systems, environmental engineers can accelerate the transition toward intelligent, data-driven remediation strategies.

## 6. Machine Learning Models for Metal Remediation

Machine learning (ML) models have emerged as powerful tools for modeling nonlinear, multivariate, and dynamic processes involved in heavy metal remediation [1,12]. These models enable prediction of removal efficiency, adsorption capacity, contaminant transport, and environmental risk with higher accuracy than conventional empirical approaches. ML techniques applied in remediation can be broadly categorized into supervised learning, unsupervised learning, and deep learning approaches [17,22].

### 6.1. Supervised Learning Models

Supervised learning algorithms are trained using labeled datasets, where input variables (e.g., pH, temperature, metal concentration, adsorbent dosage) are mapped to known outputs such as removal efficiency or adsorption capacity [17]. These models are widely used in adsorption modeling, wastewater treatment optimization, and bioremediation performance prediction [23,26].

#### 6.1.1. Linear Regression Models

Linear regression is one of the simplest supervised learning techniques used to model relationships between independent variables and remediation performance indicators [12]. It provides interpretable coefficients that quantify the influence of input parameters on metal removal efficiency.

Although linear regression is computationally efficient, it assumes linear relationships and may fail to capture complex nonlinear interactions common in environmental systems [12]. Therefore, its application is often limited to preliminary modeling or baseline comparisons.

#### 6.1.2. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are among the most widely applied ML models in environmental engineering due to their ability to model nonlinear and multivariate relationships [13,22]. ANNs consist of interconnected layers of neurons that learn patterns from input–output datasets through iterative training.

In heavy metal remediation, ANNs have been successfully used to predict adsorption capacity, removal efficiency, and process optimization in wastewater treatment systems [23,26]. Their flexibility allows modeling of complex interactions between operational parameters and environmental variables. However, ANNs may require large datasets and careful hyperparameter tuning to prevent overfitting [22].

#### 6.1.3. Support Vector Machines (SVM)

Support Vector Machines (SVM) are powerful supervised learning algorithms capable of handling nonlinear regression and classification problems using kernel functions [2,18]. SVM models are particularly effective when dealing with high-dimensional and small-to-medium-sized environmental datasets.

In metal remediation studies, SVM has demonstrated strong predictive performance in modeling adsorption processes and removal efficiencies [26]. Compared to ANN, SVM often provides better generalization with fewer training samples, although kernel selection and parameter tuning remain critical for optimal performance [2,18].

#### 6.1.4. Random Forest (RF)

Random Forest (RF) is an ensemble learning algorithm that constructs multiple decision trees and aggregates their outputs to improve predictive accuracy and robustness [1]. RF is particularly advantageous for environmental datasets due to its ability to handle nonlinear relationships, noisy data, and missing values.

In heavy metal remediation modeling, RF has been applied to predict contamination distribution, removal performance, and risk assessment outcomes [1,16]. Additionally, RF provides feature importance rankings, enabling identification of critical parameters influencing remediation efficiency.

#### 6.1.5. Gradient Boosting Algorithms

Gradient boosting methods iteratively build weak learners (typically decision trees) to minimize prediction error through sequential optimization [9]. Algorithms such as Gradient Boosting Machines (GBM) have shown high accuracy in complex environmental modeling tasks.

In remediation applications, boosting techniques enhance predictive capability for adsorption modeling and wastewater treatment optimization [9,23]. Although highly accurate, gradient boosting models may require careful tuning to avoid overfitting and increased computational cost.

### 6.2. Unsupervised Learning Models

Unsupervised learning techniques analyze unlabeled datasets to identify hidden patterns, groupings, or structural relationships within environmental data [17]. These methods are particularly useful in contamination assessment and exploratory data analysis.

#### 6.2.1. Clustering Techniques (K-Means, Hierarchical Clustering)

Clustering algorithms group contaminated sites or samples based on similarities in physicochemical characteristics and metal concentrations [12].

- **K-means clustering** partitions data into predefined clusters by minimizing within-cluster variance.
- **Hierarchical clustering** builds nested clusters based on similarity measures, enabling visualization of contamination relationships.

These techniques assist in identifying pollution hotspots, categorizing contamination levels, and supporting targeted remediation strategies [20].

#### 6.2.2. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a dimensionality reduction technique used to transform correlated environmental variables into orthogonal principal components [8]. PCA helps identify dominant factors influencing metal contamination and remediation efficiency.

By reducing data dimensionality, PCA enhances computational efficiency and improves visualization of multivariate environmental datasets [8,12]. It is commonly used for feature extraction prior to supervised learning model development.

#### 6.2.3. Pattern Recognition in Contaminated Sites

Pattern recognition techniques enable identification of spatial and temporal contamination trends in soil and water systems [20]. By analyzing correlations among multiple variables, ML-based pattern recognition supports source identification, pollution classification, and ecological risk assessment.

These approaches provide a data-driven basis for selecting appropriate remediation technologies and prioritizing contaminated areas for intervention.

### 6.3. Deep Learning Approaches

Deep learning represents an advanced subset of machine learning that employs multi-layer neural networks capable of learning hierarchical feature representations [22].

### 6.3.1. Deep Neural Networks (DNN)

Deep Neural Networks (DNNs) extend traditional ANNs by incorporating multiple hidden layers to model highly complex nonlinear relationships [22]. DNNs are particularly useful when handling large-scale environmental datasets derived from monitoring networks and sensor arrays.

In metal remediation, DNNs can model adsorption kinetics, contaminant transport processes, and treatment system performance with high predictive accuracy [22].

### 6.3.2. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are primarily designed for spatial data analysis and pattern recognition tasks [22]. In environmental remediation, CNNs can analyze spatial contamination maps, remote sensing imagery, and geospatial datasets to identify pollution distribution patterns.

Their ability to extract spatial features makes CNNs valuable for site characterization and large-scale contamination assessment.

### 6.3.3. Hybrid ML Models

Hybrid models combine multiple machine learning algorithms to enhance predictive performance and robustness [1,26]. Examples include ANN-SVM hybrids, RF-boosting ensembles, and PCA-integrated neural networks.

These hybrid frameworks leverage complementary strengths of individual models, improving accuracy, generalization, and interpretability in remediation modeling [1,26]. Hybrid approaches are increasingly applied in adsorption process optimization and environmental risk prediction.

## 7. Applications of Machine Learning in Metal Remediation

The integration of machine learning (ML) into heavy metal remediation has significantly enhanced predictive accuracy, operational optimization, and environmental risk assessment. ML models enable the analysis of complex, nonlinear relationships among physicochemical parameters, contaminant concentrations, and treatment performance indicators [1,16]. These applications support sustainable remediation planning and reduce reliance on conventional trial-and-error experimentation.

### 7.1. Prediction of Heavy Metal Removal Efficiency

One of the most prominent applications of ML in remediation is the prediction of heavy metal removal efficiency in soil and wastewater treatment systems [23,26]. Supervised learning models such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), and Gradient Boosting have demonstrated strong predictive performance for estimating adsorption capacity and percentage removal under varying operational conditions [1,2,23].

These models incorporate input variables including pH, temperature, initial metal concentration, contact time, adsorbent dosage, and surface characteristics to generate accurate output predictions [23]. Compared to conventional regression models, ML approaches effectively capture nonlinear interactions among variables, thereby improving prediction reliability and reducing experimental workload [12,26].

### 7.2. Optimization of Operational Parameters

Machine learning techniques facilitate optimization of operational parameters in remediation systems by identifying optimal combinations of variables that maximize metal removal efficiency [1,23]. Ensemble algorithms such as Random Forest and Gradient Boosting can rank feature importance, enabling identification of critical controlling factors [1,9].

ANN and SVM models are commonly used for process optimization in adsorption and wastewater treatment systems, reducing chemical consumption, energy requirements, and operational costs [23,26]. By simulating different operational scenarios, ML-based optimization frameworks minimize time-consuming laboratory experimentation and support real-time adaptive control strategies [1,16].

### 7.3. Selection of Suitable Remediation Agents

Selecting appropriate remediation agents—such as adsorbents, biochar, nanoparticles, microorganisms, or hyperaccumulator plants—is crucial for effective metal removal [4,15]. Machine learning models can analyze historical experimental datasets to predict the performance of different remediation materials under specific environmental conditions [23,26].

Feature selection techniques help identify key properties influencing treatment efficiency, such as surface area, porosity, functional groups, and microbial activity [7,23]. Data-driven material screening accelerates the discovery and optimization of novel remediation agents, reducing research and development time [1,15].

### 7.4. Modeling Adsorption and Kinetic Behavior

Adsorption is one of the most widely used methods for heavy metal removal from aqueous systems [15]. ML models have been extensively applied to predict adsorption isotherms, kinetic parameters, and equilibrium behavior without relying solely on classical mechanistic models [23,26].

ANN and SVM models can accurately simulate nonlinear adsorption kinetics by incorporating multiple operational variables simultaneously [2,23]. Ensemble learning techniques further enhance prediction accuracy for complex adsorption systems involving multiple metals and competing ions [1].

These data-driven models complement traditional isotherm models (e.g., Langmuir and Freundlich) by providing flexible, high-accuracy predictive frameworks for process design and scale-up [15,23].

### 7.5. Risk Assessment and Site Characterization

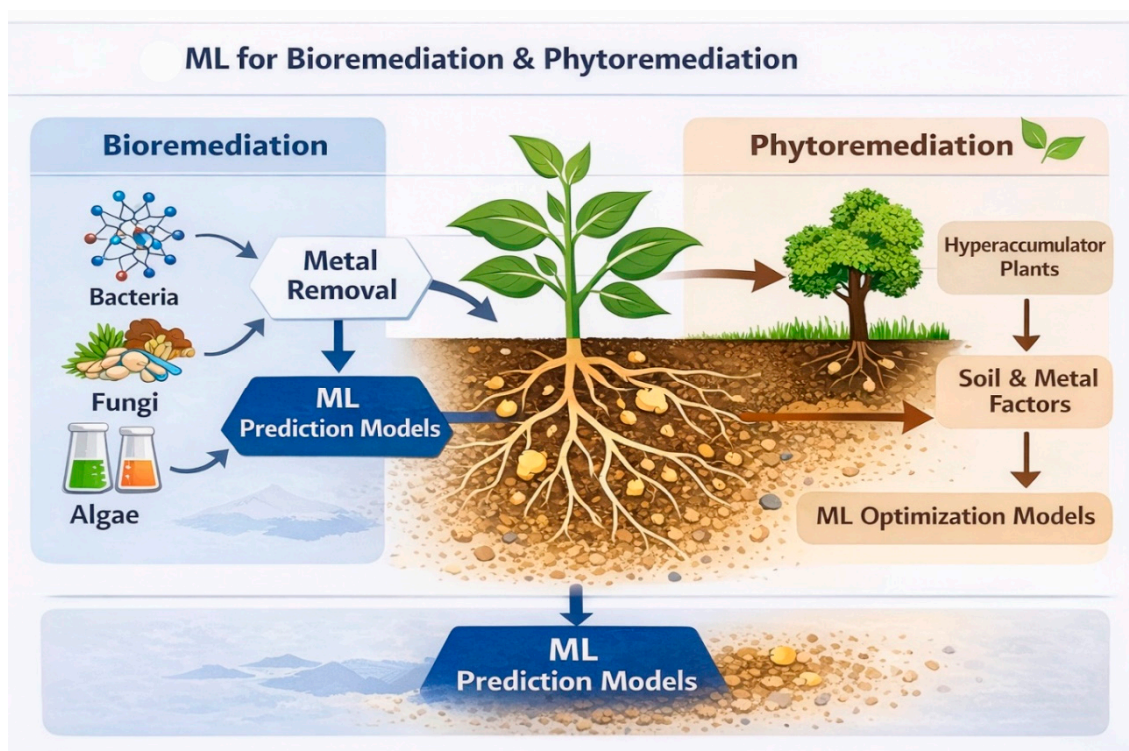
Machine learning plays a critical role in environmental risk assessment and contaminated site characterization [16,20]. Unsupervised learning techniques such as clustering and Principal Component Analysis (PCA) help classify contamination levels and identify pollution sources based on multivariate environmental data [8,20].

Random Forest and other ensemble methods are widely used for spatial contamination mapping, hotspot detection, and prediction of ecological and human health risks [1,16]. Deep learning approaches, including Convolutional Neural Networks (CNN), can analyze spatial datasets and remote sensing imagery to identify contamination patterns across large geographic regions [22].

ML-driven risk assessment frameworks improve decision-making by providing quantitative predictions of contamination spread, exposure risk, and remediation priority ranking [16,20].

## 8. Machine Learning in Bioremediation and Phytoremediation

Bioremediation and phytoremediation are environmentally sustainable approaches for heavy metal removal; however, their performance is strongly influenced by complex biological, physicochemical, and environmental interactions [4,11]. Machine learning (ML) provides advanced analytical tools to model these nonlinear relationships and enhance prediction, optimization, and decision-making processes in biological remediation systems [1,16].



**Figure 2.** Application of machine learning in bioremediation and phytoremediation.

### 8.1. Prediction of Microbial Performance

Microorganisms such as bacteria, fungi, and algae play a critical role in metal transformation, biosorption, bioaccumulation, and redox-mediated detoxification processes [11]. However, microbial remediation efficiency depends on multiple interacting factors, including pH, temperature, nutrient availability, metal concentration, and microbial strain characteristics.

Supervised ML models such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forest (RF) have been applied to predict microbial metal removal efficiency under varying environmental conditions [1,26]. These models can identify key operational parameters influencing microbial growth and metabolic activity, thereby improving process control and scalability.

Ensemble learning techniques are particularly effective in handling noisy and heterogeneous biological datasets, enhancing prediction robustness in field-scale applications [1].

### 8.2. Plant–Metal Interaction Modeling

Phytoremediation relies on plant–metal interactions that govern metal uptake, translocation, accumulation, and stabilization within plant tissues [4]. These processes are influenced by soil properties, metal speciation, root physiology, and environmental conditions [7].

Machine learning techniques help model these complex interactions by integrating soil physicochemical parameters, plant growth indicators, and metal concentration data [1,23]. Principal Component Analysis (PCA) and clustering methods assist in identifying dominant factors affecting plant-based remediation performance [8].

Predictive ML models enable screening of hyperaccumulator plant species and estimation of bioaccumulation factors under diverse site conditions, facilitating informed species selection for contaminated sites [4,16].

### 8.3. Optimization of Phytoremediation Conditions

The efficiency of phytoremediation depends on optimizing variables such as soil amendments, irrigation practices, nutrient supplementation, planting density, and harvesting time [4]. Traditional optimization approaches require long-term field experiments and repeated trials.

Machine learning-based optimization frameworks analyze historical datasets to determine optimal combinations of environmental and agronomic factors that maximize metal uptake and biomass production [1,23]. Gradient boosting and Random Forest models can rank the relative importance of soil and plant parameters, enabling targeted intervention strategies [1,9].

Such predictive optimization reduces uncertainty and enhances the feasibility of phytoremediation as a large-scale remediation technology.

### 8.4. Reduction of Experimental Cost and Time

Bioremediation and phytoremediation studies often require extended monitoring periods due to biological growth cycles and environmental variability [4,11]. ML models significantly reduce experimental workload by predicting remediation performance without exhaustive trial-and-error experimentation [1,12].

By simulating multiple treatment scenarios, ML enables rapid screening of microbial strains, plant species, and operational parameters before field implementation [16,26]. This data-driven approach accelerates decision-making, reduces research costs, and supports sustainable remediation planning.

The integration of ML into biological remediation frameworks ultimately enhances efficiency, scalability, and environmental sustainability while minimizing time-intensive experimental procedures [1,16].

## 9. Implications for Remediation Strategies

The integration of machine learning (ML) and data science into heavy metal remediation has profound implications for remediation planning, system design, and environmental decision-making. By transforming remediation processes from empirical, experience-based methods to predictive and optimization-driven frameworks, ML enhances efficiency, reliability, and sustainability [1,12].

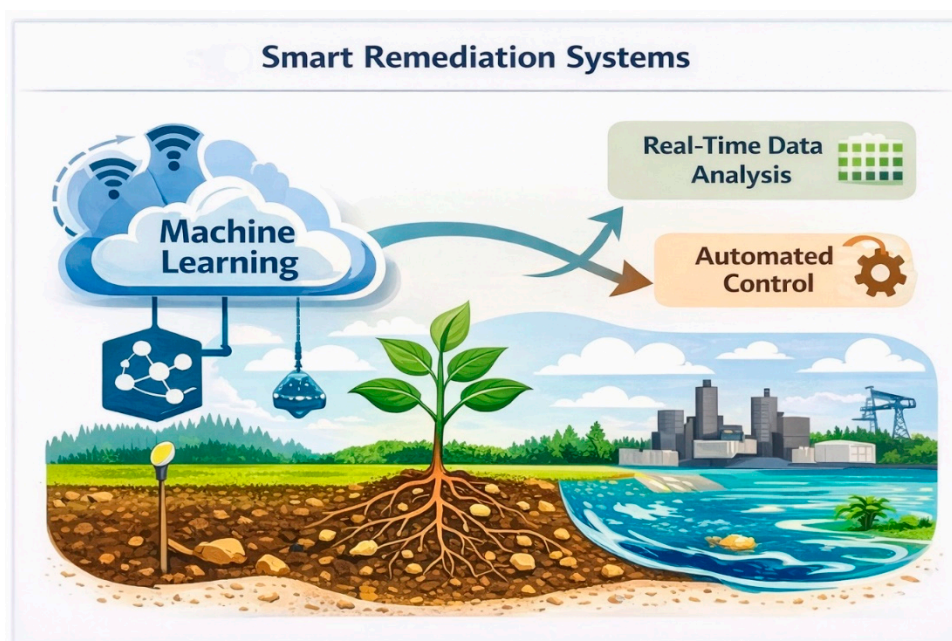


Figure 3. Architecture of smart and adaptive remediation systems.

### 9.1. Shift from Empirical to Predictive Frameworks

Traditional remediation approaches largely relied on empirical experimentation and mechanistic modeling, often requiring extensive laboratory trials and site-specific adjustments [3,15]. While these methods provided valuable insights, they were limited in handling nonlinear and multivariate environmental systems.

Machine learning enables a paradigm shift toward predictive modeling by learning complex relationships directly from historical and real-time datasets [1,17]. Algorithms such as Random Forest, Support Vector Machines, and Artificial Neural Networks can simulate remediation outcomes under diverse operational conditions, reducing uncertainty and experimental dependency [1,2,26].

This transition from deterministic trial-and-error methods to predictive analytics improves planning accuracy and accelerates remediation project implementation [1,12].

### 9.2. Identification of Critical Controlling Factors

Environmental remediation performance is influenced by numerous interacting variables, including soil pH, redox potential, organic matter content, metal speciation, temperature, and biological activity [7,11]. Identifying the most influential parameters is essential for optimizing treatment efficiency.

Machine learning models, particularly ensemble algorithms like Random Forest and gradient boosting, provide feature importance analysis that ranks the relative contribution of input variables [1,9]. Dimensionality reduction techniques such as Principal Component Analysis (PCA) further help identify dominant contamination drivers and remediation determinants [8].

By systematically identifying critical controlling factors, ML supports targeted interventions, efficient resource allocation, and improved treatment outcomes [1,16].

### 9.3. Sustainable Remediation Planning

Sustainability is a central objective in modern environmental engineering. Conventional remediation techniques may involve high energy consumption, chemical usage, and secondary waste generation [3,24].

Data-driven optimization frameworks enable environmentally sustainable planning by minimizing reagent use, reducing operational costs, and enhancing process efficiency [1,23]. ML-assisted selection of appropriate remediation agents—such as bio-based adsorbents or hyperaccumulator plants—supports eco-friendly treatment alternatives [4,15].

Furthermore, predictive risk assessment models improve long-term environmental management by evaluating potential ecological and human health impacts before large-scale implementation [16,20]. This proactive planning approach aligns remediation strategies with sustainability principles.

### 9.4. Scalable and Smart Remediation Systems

The integration of real-time monitoring systems with ML algorithms facilitates the development of smart and adaptive remediation systems [20]. These systems continuously analyze sensor data to adjust operational parameters dynamically, ensuring consistent performance under changing environmental conditions [1].

Deep learning and advanced analytics enable large-scale contamination mapping and predictive modeling across extensive geographic areas [22]. Such scalability is essential for managing industrial zones, mining regions, and urban environments affected by heavy metal pollution.

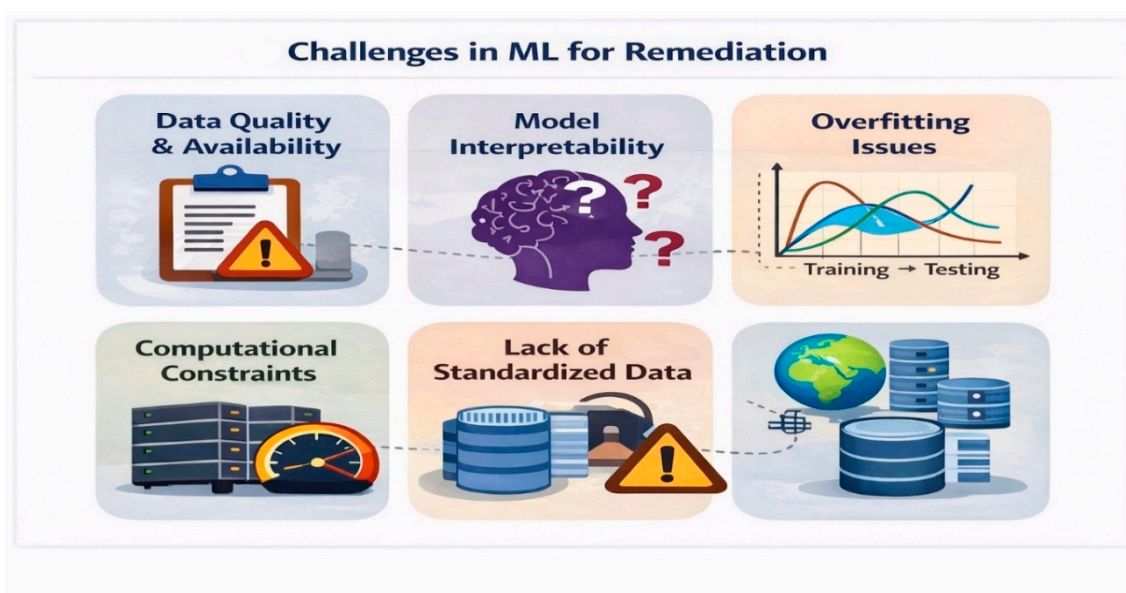
Smart remediation systems combine data acquisition, predictive modeling, and automated control mechanisms to deliver efficient, cost-effective, and scalable environmental solutions [1,16]. This digital transformation marks a significant advancement toward intelligent environmental management frameworks.

## 10. Challenges and Limitations

Despite the significant advantages of machine learning (ML) in heavy metal remediation, several technical, methodological, and practical challenges limit its widespread implementation. These challenges relate to data quality, model reliability, interpretability, computational requirements, and the absence of standardized environmental datasets [12,16].

### 10.1. Data Availability and Quality Issues

Machine learning models are highly dependent on the availability of large, reliable, and representative datasets [12]. However, environmental datasets are often limited in size, geographically fragmented, or inconsistently recorded. Field monitoring data may contain missing values, measurement errors, and inconsistencies due to instrument malfunction or sampling variability [20].



**Figure 4.** Major challenges in applying machine learning to metal remediation.

In remediation studies, laboratory-scale datasets may not fully represent field-scale complexity, limiting model transferability [3,23]. Additionally, heterogeneous data formats and lack of standardized metadata reduce interoperability across studies and regions [20]. Poor data quality can significantly compromise model accuracy and predictive reliability.

### 10.2. Model Interpretability

Many advanced ML models, particularly Artificial Neural Networks (ANN) and deep learning architectures, are often described as “black-box” models due to limited transparency in their internal decision-making processes [22]. While these models provide high predictive accuracy, their lack of interpretability may hinder regulatory acceptance and stakeholder confidence in remediation planning.

In contrast, simpler models such as linear regression or decision-tree-based approaches offer better interpretability but may sacrifice predictive performance in highly nonlinear systems [12]. Although feature importance ranking in ensemble methods like Random Forest improves transparency [1], further development of explainable artificial intelligence (XAI) frameworks is necessary for environmental applications [16].

### 10.3. Overfitting and Generalization Problems

Overfitting occurs when a model learns noise and specific patterns from training data rather than capturing generalizable relationships [12]. This issue is particularly common in environmental studies where datasets are small or highly variable.

Complex models such as deep neural networks may achieve high training accuracy but perform poorly when applied to new or unseen contaminated sites [22]. Proper cross-validation techniques, regularization methods, and feature selection strategies are essential to ensure model robustness and generalization [12,17].

Ensuring transferability of ML models across different geographic regions and environmental conditions remains a significant challenge in remediation research [16].

### 10.4. Computational Constraints

Advanced ML and deep learning models require substantial computational resources, especially when handling large-scale environmental datasets or real-time monitoring streams [22]. High computational demands may limit practical implementation in regions with limited technical infrastructure.

Training complex ensemble and deep learning models can also be time-consuming and energy-intensive, potentially contradicting sustainability goals [22]. Although open-source tools and optimized algorithms have reduced some computational barriers [21], efficient model design and resource management remain critical considerations.

### 10.5. Lack of Standardized Environmental Datasets

The absence of standardized, open-access environmental remediation datasets poses a major barrier to reproducibility and model benchmarking [20]. Differences in sampling methods, analytical techniques, and reporting standards make cross-study comparisons challenging.

Without standardized datasets, it becomes difficult to evaluate model performance objectively or establish universal predictive frameworks for heavy metal remediation [16]. Developing centralized, harmonized databases with consistent metadata documentation is essential to advance data-driven environmental engineering [20].

Addressing these limitations through improved data governance, model transparency, validation protocols, and collaborative data-sharing initiatives will be crucial for the long-term success of ML applications in remediation science.

## 11. Future Perspectives

The future of machine learning (ML) and data science in heavy metal remediation lies in deeper digital integration, improved transparency, and global collaboration. Emerging technologies such as Internet of Things (IoT)-enabled sensors, explainable artificial intelligence (XAI), and standardized environmental databases are expected to transform remediation from reactive intervention to proactive and intelligent environmental management [16,20,22].

### 11.1. Integration with IoT and Smart Sensors

The proliferation of IoT-based environmental sensors enables continuous monitoring of parameters such as pH, redox potential, dissolved oxygen, temperature, and metal concentration in soil and water systems [20]. These smart sensors generate high-frequency, real-time datasets that can be directly integrated with ML algorithms for predictive analytics and automated control.

The combination of IoT networks and ML facilitates early detection of contamination spikes, dynamic process optimization, and rapid response to environmental changes [1,20]. As sensor technologies become more affordable and accurate, their integration into remediation infrastructures will enhance data availability and improve predictive model performance.



**Figure 5.** Emerging trends and future directions in ML-driven metal remediation.

#### 11.2. Real-Time Adaptive Remediation Systems

Future remediation systems are expected to operate as adaptive, closed-loop frameworks where real-time data streams continuously inform ML models to optimize operational parameters [1]. Such systems can automatically adjust treatment variables—such as adsorbent dosage, aeration rate, nutrient supply, or hydraulic retention time—based on predictive outputs.

Deep learning and ensemble models are particularly suitable for processing large-scale time-series environmental data [22]. The implementation of adaptive remediation systems reduces operational inefficiencies, enhances energy optimization, and improves long-term sustainability [16].

This transition from static treatment designs to dynamic, self-optimizing systems represents a major advancement in intelligent environmental engineering.

#### 11.3. Explainable Artificial Intelligence (XAI)

While advanced ML models offer high predictive accuracy, their limited interpretability remains a concern for environmental decision-making [22]. The development of Explainable Artificial Intelligence (XAI) frameworks aims to enhance transparency by clarifying how input variables influence predictive outcomes.

Feature importance analysis in Random Forest and boosting algorithms provides partial interpretability [1,9], but future research must focus on more comprehensive explanation techniques tailored to environmental systems [16].

Improved interpretability will strengthen regulatory acceptance, stakeholder trust, and scientific credibility of ML-assisted remediation strategies.

#### 11.4. Policy and Regulatory Integration

For ML-driven remediation approaches to achieve widespread adoption, integration with environmental policy and regulatory frameworks is essential [16]. Regulatory bodies require validated, transparent, and reproducible methodologies before approving remediation technologies.

Standardized validation protocols, benchmarking datasets, and clear reporting guidelines will facilitate regulatory compliance and encourage institutional acceptance [20]. Data-driven risk assessment models can further support evidence-based policymaking by quantifying environmental and health impacts prior to remediation implementation [16].

Collaboration between environmental scientists, data engineers, and policymakers will be critical for translating ML innovations into practical regulatory frameworks.

#### 11.5. *Development of Global Environmental Databases*

The establishment of centralized, standardized global environmental databases represents a crucial step toward advancing ML applications in remediation [20]. Harmonized datasets containing physicochemical parameters, contamination levels, treatment outcomes, and geospatial information would enable robust cross-regional model development and benchmarking.

Open-access databases would improve reproducibility, enhance collaboration, and accelerate innovation in remediation technologies [16,20]. Integration of such databases with cloud computing and open-source ML platforms would create scalable, globally accessible decision-support systems.

Developing comprehensive global environmental data infrastructures will significantly enhance the reliability, transferability, and scalability of machine learning models in heavy metal remediation.

## 12. Conclusions

### 12.1. *Summary of ML Contributions to Metal Remediation*

The integration of machine learning (ML) and data science into heavy metal remediation has fundamentally transformed environmental engineering from empirical, trial-and-error approaches to predictive, optimization-driven frameworks [1,12]. ML models—including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), gradient boosting algorithms, and deep learning architectures—have demonstrated strong capability in predicting heavy metal removal efficiency, modeling adsorption behavior, optimizing operational parameters, and supporting risk assessment [1,2,23,26].

Unlike conventional statistical methods, ML techniques effectively capture nonlinear and multivariate relationships among physicochemical, biological, and operational variables [12,22]. Feature importance analysis and dimensionality reduction techniques such as Principal Component Analysis (PCA) further enhance understanding of critical controlling factors in remediation systems [8,9].

Overall, ML contributes to improved prediction accuracy, reduced experimental workload, enhanced decision-making, and accelerated implementation of remediation technologies across laboratory and field scales [1,16].

### 12.2. *Environmental Sustainability Outlook*

Sustainable remediation requires minimizing environmental footprint while maximizing contaminant removal efficiency. Data-driven optimization frameworks enable efficient resource utilization by reducing chemical consumption, energy demand, and secondary waste generation [3,23].

Machine learning also supports eco-friendly technologies such as bioremediation and phytoremediation by modeling microbial performance and plant–metal interactions under diverse environmental conditions [4,11]. The integration of real-time monitoring systems, IoT-enabled sensors, and adaptive ML models further promotes smart and energy-efficient remediation strategies [20,22].

By enabling predictive risk assessment and proactive environmental management, ML aligns remediation practices with global sustainability goals and long-term ecological protection objectives [16,20].

### 12.3. Recommendations for Future Research

Despite notable advancements, several research gaps remain. Future studies should focus on:

- Development of standardized, high-quality, and open-access environmental datasets to enhance model reproducibility and benchmarking [20].
- Integration of explainable artificial intelligence (XAI) frameworks to improve model transparency and regulatory acceptance [16,22].
- Expansion of hybrid and deep learning architectures capable of handling large-scale, multi-source environmental datasets [1,22].
- Strengthening interdisciplinary collaboration between environmental scientists, data scientists, and policymakers to facilitate practical implementation [16].
- Validation of ML models across diverse geographic regions and environmental conditions to improve generalization and scalability [12,16].

In conclusion, machine learning and data science represent transformative tools in heavy metal remediation. By combining predictive analytics, intelligent optimization, and sustainable engineering principles, ML-driven frameworks offer a robust pathway toward smart, scalable, and environmentally responsible remediation systems for the future [1,16].

## References

1. Hao, L., Zhang, D., Zhou, H., Wang, Z., Zhang, J., Zhao, Z., Li, M. Machine learning applications in prediction and optimization of heavy metal bioremediation: A review.
2. Ju, C., Xu, X., Wang, Q., Park, J., Meng, L., Ruan, Z., Zhou, L. Application of machine learning methods to predict the immobilization rate of heavy-metal-contaminated soils by alkaline solid waste.
3. Barkhordari, M. S., Zhou, N., Li, K., Qi, C. Interpretable machine learning for predicting heavy metal removal efficiency in electrokinetic soil remediation.
4. Barkhordari, M. S., Qi, C. Integrating machine learning and reliability analysis: A novel approach to predicting heavy metal removal efficiency using biochar.
5. El Shafie, M., Mubarak, M. F., Nasr, M., Shaltout, A., El Shahawy, A. Application of machine learning in predicting heavy metal uptake by activated carbon adsorbents.
6. Tymoteusz Miller, Danuta Cembrowska-Lech, Anna Kisiel, Adrianna Krzemińska, Polina Kozlovska, M Jawor .HARNESSING AI FOR ENVIRONMENTAL RESILIENCE: Mitigating heavy metal pollution and advancing sustainable practices in diverse spheres.
7. Obadimu, C. O., Ekwere, I. O., Shaibu, S. E., Essien, U. B., Adelagun, R. O. A., Adewusi, S. G. Application of machine learning framework on heavy metals fate in the coastal environment.
8. Gheibi, M., Masoomi, S. R., Magala, M. U., Fathollahi-Fard, A. M., Ghazikhani, A., Behzadian, K. The application of artificial intelligence in adsorption process of heavy metals: A systematic review.
9. Dian, X., Hao, J., Zhang, Z., Chen, Z., Yao, L. Heavy metal removal performance of capacitive deionization technology studied by machine learning.
10. 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., Zhu, J., Zhang, H. Machine learning, New ideas and tools in environmental science and engineering.
11. Liu, X., Lu, D., Zhang, A., Liu, Q., Jiang, G. Data-driven machine learning in environmental pollution, Gains and problems.
12. Li, X., Yang, Y., Yang, J., Fan, Y., Qian, X., Li, H. Rapid diagnosis of heavy metal pollution in lake sediments based on environmental magnetism and machine learning.
13. Wu, J., Zhao, F. Machine learning as an effective technical method for assessing phosphorus-dissolving microbial agromediation.
14. Obadimu, C. O., Shaibu, S. E., Ekwere, I. O., Adelagun, R. O. A. Machine learning-based forecasting of bioaccumulation and histopathological effects in aquatic organisms.
15. Yaseen, Z. M., Doost, Z. H., Khan, R., Abdulraheem, A., Abdulameer, S. F., Falah, M. W., Farooque, A. A. Chitosan-based flocculant heavy metal removal prediction using machine learning models.

16. Chen, M. W., Chang, M. S., Mao, Y., Hu, S., Kung, C. C. Machine learning in the evaluation and prediction models of biochar application, A review.
17. Takarina, N. D., Matsue, N., Johan, E., Adiwibowo, A., Rahmawati, M. F., Pramudyawardhani, S. A., Wukirsari, T. Machine learning using random forest to model heavy metals removal efficiency using a zeolite-embedded sheet in water.
18. Li, H., Zhou, Z., Long, T., Wei, Y., Xu, J., Liu, S., Wang, X. Big-data analysis and machine learning based on oil pollution remediation cases from CERCLA database.
19. Yaqub, M., Eren, B., Eyupoglu, V. Prediction of heavy metals removal by polymer inclusion membranes using machine learning techniques.
20. Yaseen, Z. M. An insight into machine learning models era in simulating soil, water bodies and adsorption of heavy metals, Review, challenges and solutions.
21. Wei, X., Liu, Y., Shen, L., Lu, Z., Ai, Y., Wang, X. Machine learning insights in predicting heavy metals interaction with biochar.
22. Yaseen, Z. M., Alhalimi, F. L. Heavy metal adsorption efficiency prediction using biochar properties, A comparative analysis for ensemble machine learning models.
23. Yuan, X., Li, J., Lim, J. Y., Zolfaghari, A., Alessi, D. S., Wang, Y., Wang, X., Ok, Y. S. Machine learning for heavy metal removal from water, Recent advances and challenges.
24. Blessing, A. A., Olateru, K. AI-driven optimization of bioremediation strategies for river pollution, A comprehensive review and future directions. *Frontiers in Microbiology*. 2025.
25. Yang, H., Huang, K., Zhang, K., Weng, Q., Zhang, H., Wang, F. Predicting heavy metal adsorption on soil with machine learning and mapping global distribution of soil adsorption capacities.
26. Bhagat, S. K., Tung, T. M., Yaseen, Z. M. Development of artificial intelligence for modeling wastewater heavy metal removal, State of the art, application assessment and possible future research.
27. Malviya, A., Jaspal, D. Artificial intelligence as an upcoming technology in wastewater treatment, A comprehensive review.
28. Alam, G., Ihsanullah, I., Naushad, M., Sillanpää, M. Applications of artificial intelligence in water treatment for optimization and automation of adsorption processes, Recent advances and prospects.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.