Journal of Environmental & Earth Sciences
https://journals.bilpubgroup.com/index.php/jees

REVIEW

Next-Generation Approaches for Predicting Soil Heavy Metal Contamination: Insights from Geospatial and AI-Based Methods

Mingxiong Zhu

Chongqing Jiulongpo District Environmental Monitoring Station, Chongqing 400050, China

ABSTRACT

Contamination of soil with heavy metals poses a serious risk to ecosystems, agriculture, and human health. The spatial and temporal constraints limit the use of traditional techniques of predicting contamination, such as soil sampling and geochemical mapping. New developments in geospatial methods and artificial intelligence (AI) provide promising options for large-scale, real-time monitoring and prediction. Geospatial tools, such as remote sensing and Geographic Information Systems (GIS), can provide important spatial information for mapping contamination patterns. In contrast, AI-based applications, such as machine learning and deep learning, can process multifaceted datasets to forecast contamination trends. A combination of these two methods increases the predictability of measuring soil contamination dynamically and at high resolution. Nevertheless, there are still difficulties with data quality, model interpretability, and computational complexity. Notwithstanding these issues, the interrelation between geospatial and AI approaches represents a major breakthrough in environmental monitoring, offering a more in-depth and efficient tool for regulating heavy-metal pollution in soil. The review highlights the potential of these next-generation approaches and offers insights into their use, limitations, and future directions for enhancing soil health and environmental sustainability.

Keywords: Soil Contamination; Heavy Metals; Geospatial Techniques; Artificial Intelligence; Environmental Monitoring

*CORRESPONDING AUTHOR:
Mingxiong Zhu, Chongqing Jiulongpo District Environmental Monitoring Station, Chongqing 400050, China; Email: 13594528992@163.com

ARTICLE INFO
Received: 22 May 2026 | Revised: 28 June 2026 | Accepted: 4 July 2026 | Published Online: 25 August 2026
DOI: https://doi.org/10.30564/jees.v8i8.13540

CITATION
Zhu, M., 2026. Next-Generation Approaches for Predicting Soil Heavy Metal Contamination: Insights from Geospatial and AI-Based Methods. Journal of Environmental & Earth Sciences. 8(8): 1175–1191. DOI: https://doi.org/10.30564/jees.v8i8.13540

1. Introduction

Heavy metal contamination of soil is a serious worldwide environmental problem that poses significant threats to human health, agriculture, and the environment [1]. Lead (Pb), cadmium (Cd), arsenic (As), and mercury (Hg) are non-biodegradable heavy metals whose accumulation in soil may have long-term adverse effects on plant growth, water quality, and food safety [2]. Industrial activities, agricultural practices, and urbanization often lead to soil contamination, creating a complex, pervasive issue that is difficult to address. This has made the proper prediction and monitoring of soil heavy metal contamination a major agenda for environmental scientists and policymakers worldwide.

The conventional methods for measuring soil contamination have relied on direct soil sampling and laboratory analysis [3]. Although these are effective for identifying contamination at a particular site, they are laborious, time-consuming, and expensive, especially when dealing with extensive geographic areas. In addition, traditional sampling techniques can only provide a snapshot of contamination levels at a given time and are usually insufficient for forecasting the spatial distribution and temporal variation of pollutants. Given the growing urgency to combat soil contamination on a larger scale, there is a need to develop more effective, less expensive, and more accurate methods for forecasting and mapping contamination risks.

The emergence of next-generation technologies, especially geospatial techniques and artificial intelligence, holds promise for addressing the shortcomings of conventional methods. Geospatial technologies, including remote sensing and Geographic Information Systems, have transformed environmental monitoring by providing high-resolution, large-scale spatial information [4]. An example of this is remote sensing, which enables continuous observation of soil conditions through satellite imagery, drones, and other aerial vehicles, and enables the identification of contamination patterns across large territories. GIS, in turn, provides an effective means for spatial data analysis, enabling the mapping of areas where contamination occurs and the integration of various environmental data influencing soil pollution, including land use, climate, and industrial activity [5].

Simultaneously, AI-based techniques, such as machine learning (ML) and deep learning (DL), have become influential predictive tools for soil contamination and complex environmental data [6]. These artificial intelligence methods can discover patterns and correlations in large datasets that would otherwise be undetectable or require a long time to identify by humans. Using AI models to predict the odds of soil heavy metal contamination would be highly accurate, aided by ML algorithms such as decision trees, support vector machines, and neural networks, even with large datasets containing more than a few variables. Moreover, AI-based approaches are dynamic and continually improve as more data emerge, making them highly applicable to dynamic environmental circumstances, where the level of contamination may change over time.

The combination of geospatial techniques and AI-based solutions is especially innovative, as it can be used not only to make spatial predictions about contamination but also to make temporal ones [7]. As an illustration, geospatial data can provide information on pollutant distribution at a given moment, and AI models could predict how contamination will evolve or diffuse over the next several years, accounting for environmental and anthropogenic factors. Such spatial and temporal prediction is a significant step forward compared to classical approaches, which fail to capture the dynamics of soil contamination.

Furthermore, blending geospatial and AI strategies can promise greater predictive accuracy by leveraging numerous data sources [8]. For example, remote sensing data can be combined with soil properties, weather conditions, historical land-use data, and socio-economic variables to produce more detailed and robust predictive models. These models can provide more detailed information about the factors causing contamination and the areas at highest risk of pollution, enabling targeted interventions to prevent further pollution.

Although the application of geospatial and AI-based tools to environmental monitoring is not novel, predicting the occurrence of heavy metal contamination in soil remains a relatively recent area. Current studies have shown the potential of these methods in other fields, including air pollution forecasting, climate research, and biodiversity conservation [9]. Nonetheless, their use in addressing soil pollution, especially heavy metals, has not yet been fully explored, and multiple obstacles must be addressed, including data quality, model generalizability, and integration complexities. Specifically, a critical examination of existing review literature on geospatial-AI integration reveals three persistent deficiencies: (1) the "black box" nature of most AI models severely limits their interpretability for environmental decision-making; (2) existing approaches rarely integrate all relevant exposure pathways (e.g., ingestion, inhalation, and dermal contact) within a unified predictive framework; and (3) the lack of user-friendly, automated tools for cross-scale prediction hinders the translation of research findings into routine environmental management practice. Recent studies have begun to address these gaps, for example, by combining explainable AI (SHAP analysis) with multi-heuristic optimization to clarify the decision logic of contamination risk models [9], and by developing GIS-based automated hazard index tools that integrate multiple exposure pathways in a reproducible, timely manner [10,11]. Nevertheless, the integration of geospatial data, AI-driven prediction, and interpretable risk assessment remains incomplete in the context of soil heavy metal contamination, particularly for cross-scale applications that link local sampling data with regional prediction models.

The goal of this review is to address this gap and discuss current developments in geospatial and AI-based approaches for predicting soil heavy metal contamination. We hope to give us a comprehensive overview of these innovative approaches by synthesizing the existing state of knowledge, and bring up the strengths and limitations of the approaches and suggest ways in which they can be used to enhance soil health monitoring and management. The unusual feature of the review is not only the use of the latest technologies in the specific issue of soil heavy metal contamination, but also the integration of these methods to form more effective, scalable, and accurate predictive systems [12]. It is on this synthesis that we hope to stimulate further research and development in this essential area, ultimately safeguarding soil ecosystems and human health.

2. Traditional Methods for Predicting Soil Heavy Metal Contamination

The traditional methods for assessing soil heavy metal contamination have been the cornerstone of environmental monitoring for decades [3]. These approaches primarily rely on direct soil sampling and laboratory analysis, which provide valuable insights into the contamination levels of specific sites. While effective in certain contexts, these methods have significant limitations for large-scale monitoring, real-time predictions, and understanding the spatial distribution of contaminants (**Table 1**). In this section, we will discuss traditional techniques for predicting soil contamination, assess their advantages and drawbacks, and explore why newer, more advanced methods are increasingly being considered.

Table 1. Comparison of Traditional and Next-Generation Approaches for Predicting Soil Heavy Metal Contamination.

Approach	Methodology	Advantages	Limitations	Cost	Scalability
Soil Sampling	Collection of soil samples followed by laboratory analysis	High accuracy, direct measurement	Time-consuming, labor-intensive	High	Low
Geochemical Mapping	Mapping of contaminants using soil sample data and spatial analysis	Provides an overview of contamination areas	Limited by sampling density and resolution	Moderate	Moderate
Risk Assessment Models	Quantification of risks based on soil and environmental factors	Predicts potential impacts on ecosystems and human health	Uncertainty in predicting mobility and bioavailability of contaminants	Moderate	Moderate
Geospatial & AI Approaches	Integration of remote sensing, GIS, and machine learning models	High accuracy, scalable, real-time predictions	Data quality issues, computational complexity	Low to High	High

2.1. Soil Sampling and Laboratory Analysis

Laboratory analysis and soil sampling have been regarded as the gold standard for measuring heavy metal contamination of the soil. In this technique, soil samples are collected from different locations within a particular area, and laboratory analysis is conducted to determine levels of heavy metals such as lead, cadmium, and mercury. Sampling is usually performed at varying depths and locations within a contaminated site to obtain a representative sample of contamination levels [13].

The major benefit of the technique is its accuracy. Atomic absorption spectroscopy (AAS) or inductively coupled plasma mass spectrometry (ICP-MS) may be used to measure heavy metal concentrations in soil with high accuracy, even at very low levels [14]. This great accuracy renders soil sampling a useful method of establishing contamination in such areas where there is little existing information or where contamination is suspected but not yet proved.

Nevertheless, reliance on soil sampling has several disadvantages. One, it is a time-consuming and labor-intensive process, particularly when extensive areas must be surveyed [15]. It may take hundreds or even thousands of soil samples to build a complete picture of contamination in a particular area. Such assessments are very costly and time-consuming due to the extensive fieldwork, sample transport, and laboratory analysis involved. Also, the process is mostly manual and relies on human judgment, and thus, it is always possible to have sampling errors and biases [16]. For example, when samples are not taken at representative sites, the findings may be biased and will not accurately reflect the true picture of contamination.

Moreover, the conventional method of soil sampling gives a picture of contamination at a particular time [17]. The levels of heavy metals in the soil may vary over time due to factors such as rainfall, human activities, and land use. Consequently, a single sampling is unlikely to be sufficient to bring the dynamic nature of soil contamination into focus. This time constraint will complicate any attempt to forecast future levels of contamination or to determine how soil pollution trends change over time.

2.2. Geochemical Mapping

Another conventional method which is used in the evaluation of soil contamination is geochemical mapping. This is done by mapping the geographic location of chemical elements, such as heavy metals, on a large scale. The geochemical data is usually obtained by sampling using soil samples at routine intervals and displayed on maps to determine areas that contain high concentrations of heavy metals. These maps may be used to identify contamination hotspots and areas of concern, leading to further investigation or remediation [18].

Geochemical mapping is advantageous in that it helps give a wide view of contamination in large regions, which is more beneficial in tracing trends or patterns of pollution in different landscapes. Geochemical maps can be used to identify priority areas that need remediation or further analysis by creating a spatial image of soil contamination [19].

Similar to soil sampling, however, geochemical mapping has the disadvantages of sampling density and the absence of time [20]. The frequency and distribution of soil samples are extremely important to the effectiveness of geochemical mapping. In low-density sampled areas, the maps cannot provide the exact degree of contamination, and incomplete or erroneous conclusions can be drawn. Also, the approach fails to provide a real-time or dynamic picture of contamination, which is critical for forecasting future contamination. Consequently, geochemical mapping can serve as an initial step in identifying regions that may be contaminated, but it cannot provide a long-term monitoring solution. As shown in **Figure 1**, the usual workflow of traditional methods for predicting soil contamination is summarized as: soil sampling to risk assessment.

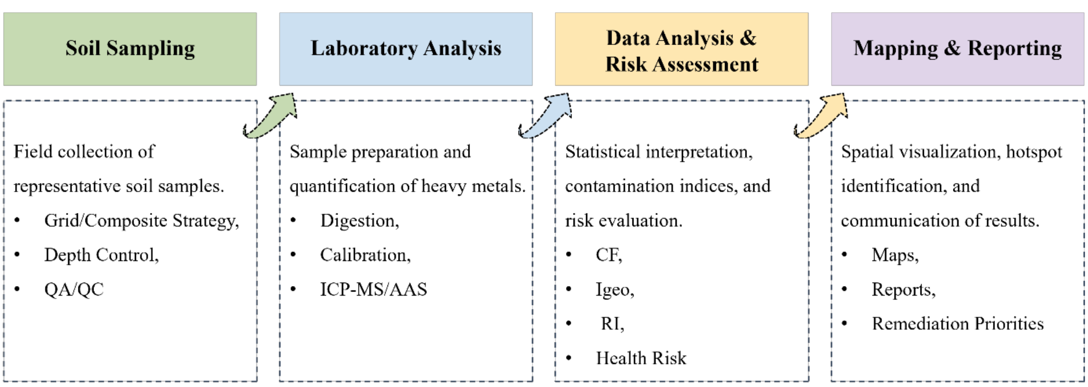

Figure 1. Soil heavy metal contamination prediction workflow.

Note: Traditional linear workflow used for soil heavy metal contamination assessment and prediction.

2.3. Risk Assessment Models

Risk assessment models have been developed to quantify the potential risks posed by heavy metal contamination of soil to human health, agriculture, and ecosystems. These models depend on soil properties, pollutant concentrations, and environmental conditions (including soil pH, organic matter content, and moisture levels) to forecast the potential effects of heavy metals on the environment and human health. Models that quantify the bioavailability and mobility of heavy metals in soil, as well as their potential leaching into groundwater or absorption by plants, are usually used in risk assessment [21].

The beauty of risk assessment models is that they can be used to forecast the possible impacts of contamination under different environmental conditions [22]. Such models may be used to focus remediation efforts on areas based on risk severity and the local population's vulnerability. In other words, highly contaminated areas may be prioritized and addressed immediately around agricultural zones or residential neighborhoods, while lower-risk areas can be tracked.

Nevertheless, risk assessment models face several challenges. Among the major constraints is the uncertainty that is involved in forecasting the movement and bioavailability of heavy metals in soil [23]. The behavior of contaminants in soil is a dynamic phenomenon and can vary under the influence of numerous factors, such as soil type, weather conditions, and anthropogenic influences. Consequently, these models may require extensive calibration and validation, which can be difficult to obtain in the absence of sufficient empirical data. Moreover, the precision of risk assessment models depends on the quality and accessibility of data on soil characteristics and contamination levels, which are not always available, especially in remote or underdeveloped locations.

2.4. Challenges of Traditional Methods

Although conventional practices such as soil sampling, geochemical mapping, and risk assessment models have been useful for providing information on soil contamination, these techniques are increasingly viewed as insufficient to address the magnitude and complexity of the issue. The space and time coverage of these methods is one of the primary issues. Soil contamination is usually not uniform, and concentrations may exhibit a wide range of variation within a local area [24]. Consequently, conventional approaches based on a small number of samples or fixed maps may be unable to capture the actual contamination across a large region.

In addition, conventional techniques are not particularly effective at forecasting future patterns of contamination or responding to a dynamic environment. Various factors can affect the amount of heavy metals in soil, including land use, industrial activity, and climate [25]. To anticipate how these factors interact and influence future contamination levels, more sophisticated methods are needed that can accommodate complex, multidimensional data and account for changes over time.

The other major disadvantage of the traditional methods is that they are not able to offer real-time monitoring [26]. It may take weeks or even months to collect soil samples, send them to a laboratory, and analyze the results. The level of contamination can vary during this period, and the data collected might not provide a true picture of the environmental situation. It is especially problematic when there is a need to respond to contamination immediately, and the data is not available in real-time.

2.5. The Need for Next-Generation Approaches

Against the backdrop of the limitations of conventional methods, the need for next-generation techniques to anticipate soil heavy metal contamination is increasingly recognized. Emerging technologies such as remote sensing, Geographic Information Systems, and artificial intelligence offer new opportunities to overcome the shortcomings of conventional methods. These sophisticated technologies will be able to offer more detailed, real-time, and predictive information about soil contamination, enabling more efficient decision-making and environmental control [27].

The combination of geospatial and AI-related solutions, e.g., can lead to the precision and scalability of predictions of soil contamination, providing the opportunity to analyze large territories with high spatial and time resolution. Moreover, AI-powered models can be trained on available data to discern patterns and infer future contamination trends, which are dynamic given the nature of soil pollution [7].

Conclusively, although conventional approaches to predicting soil heavy metal contamination have proved invaluable over the years, they are increasingly viewed as inadequate to address the demands of extensive, dynamic contamination. This need for more advanced, efficient, and real-time methods has led to the emergence of next-generation technologies that will revolutionize environmental monitoring [28]. These technologies will enhance our capacity to predict, manage, and remediate soil contamination, ultimately leading to healthier soils and a more sustainable environment.

3. Geospatial Approaches for Heavy Metal Prediction

Geospatial technologies have transformed the way environmental scientists learn about and forecast soil contamination, especially heavy metal contamination [27]. They have made these tools essential for monitoring and managing the environment, as they can map, analyze, and visualize data across large regions. Geospatial techniques are vital for understanding the distribution and concentration of soil contaminants and can enable large-scale assessments that were previously impossible with traditional soil sampling techniques. **Table 2** provides an overview of geospatial methods used to predict soil contamination, along with their strengths and weaknesses. In this section, the authors discuss the use of geospatial technologies, including remote sensing, Geographic Information Systems, and spatial modelling, for predicting soil heavy metal contamination.

Table 2. Summary of common geospatial techniques used in soil contamination prediction.

Technique	Description	Advantages	Applications	Limitations
Remote Sensing	Use of satellite imagery or drones to capture environmental data	Non-invasive, large-scale monitoring, real-time data	Detecting vegetation health, mapping contamination hotspots	Limited by weather conditions and resolution
Geographic Information Systems (GIS)	Software for spatial data analysis and mapping	Data integration and visualization of contamination patterns	Mapping contamination areas, risk assessment	Requires high-quality data for accuracy
Spatial Modeling	Mathematical models predicting the spread of contaminants	Allows for prediction of future contamination, scalable	Predicting contaminant migration over time	Complex to calibrate, requires high computational power

3.1. Remote Sensing Technologies

Remote sensing is a method of obtaining data about the surface of the Earth without directly touching it, that is, in most cases, through satellite imagery, drones, or aerial surveys [4]. Remote sensing could be used in the context of soil contamination to identify contamination patterns over large areas, helping environmental scientists identify the source of pollution, measure trends, and determine spatial changes in contamination. Increased access to high-resolution satellite imagery, along with advances in drone technology, has enhanced the ability to monitor soil contamination at global and local scales.

Among the main benefits of remote sensing is that it can provide real-time, non-invasive information about soil conditions [29]. Satellite imagery can measure several environmental parameters linked to heavy metal contamination, including vegetation health, soil moisture, and temperature. Satellite data, such as the Normalized Difference Vegetation Index (NDVI), could be used to assess the well-being of vegetation that may be affected by heavy metal toxicity. A major decline in vegetation health may indicate contaminated soil. Likewise, the presence of heavy metals can be detected in multispectral and hyperspectral imagery due to changes in vegetation and soil properties.

Although it has potential, remote sensing has certain limitations in terms of direct detection of soil contamination. Most remote sensing systems cannot directly measure heavy metal concentrations in soil but instead use proxies (e.g., vegetation health or surface reflectance) to assess contamination levels. In some cases, this indirect method may cause errors or misinterpretation, especially in regions where land cover is highly complex or where soil properties vary. Also, because the resolution of satellite images is not always fine enough, it may not always be possible to detect contamination in smaller, more localized regions. However, remote sensing is a vital tool for large-scale environmental monitoring and offers invaluable information for predicting potential hotspots of contamination [30].

3.2. Geographic Information Systems (GIS)

Geographic Information Systems have established themselves as indispensable components of environmental management, enabling the storage, analysis, and visualization of geospatial data. With regard to soil heavy metal contamination, GIS can integrate a broad spectrum of data, such as soil properties, land-use patterns, industrial activities, climatic factors, and remote sensing data, to forecast and assess the extent of contamination. GIS can be used to develop extremely detailed maps that indicate areas at risk of contamination, and it is an invaluable resource for both predictive modeling and decision-making [5].

Visualization of complex data relations in space is one of the key benefits of GIS in contamination prediction [31]. As an example, GIS may be applied to superimpose data on heavy metal levels with other environmental conditions like location near industrial areas, farmlands and garbage disposal sites. It is this spatial relationship that allows a researcher to identify potential sources of contamination and determine which areas might be at higher risk.

Furthermore, GIS-based spatial analysis, including buffer analysis, spatial interpolation, and hotspot analysis, can be used to identify areas with high levels of contamination and highlight those requiring remediation or further analysis. As a case in point, predicting heavy metal concentrations at unmeasured locations can be done using spatial interpolation methods such as kriging or inverse distance weighting (IDW) of sample points in the surrounding area. This enables the formation of continuous maps of contamination, even when sample data are sparse or unevenly distributed [31].

GIS has its limitations, although it has some strengths. The availability of high-quality, accurate data is one of the major challenges in using GIS to predict contamination. The success of GIS models largely depends on the quality of the input data, including soil samples, land-use data, and environmental factors [32]. The data used may be inaccurate or too old, potentially leading to misleading results that could undermine the accuracy of the predictions. Moreover, GIS models tend to be computationally intensive, especially when performing spatial analysis on large datasets and intricate operations.

3.3. Spatial Models for Contaminant Distribution

Spatial models are mathematical or statistical models that are used to predict the distribution of contaminants, e.g., heavy metals, in an area [33]. These models use geospatial data and environmental variables to simulate the spread and accumulation of contaminants over time. Spatial models are important for predicting contaminated risk areas and determining factors that lead to heavy metal contamination in soil. Contamination prediction models can be broadly divided into two categories: deterministic and stochastic.

Deterministic models are models whose output is predetermined by a set of fixed rules and assumptions, with the output directly determined by the input parameters. As an illustration, a deterministic spatial model could be used to forecast the propagation of contamination based on predefined environmental conditions such as soil type, rainfall, and land use [34]. Such models are also fairly straightforward and capable of making rapid estimates, although they may fail to capture the full spectrum of uncertainties that can occur within an environmental system.

Stochastic models, on the other hand, incorporate randomness and uncertainty into their forecasts, which is more appropriate for modeling the unpredictability of environmental systems [35]. It is possible to use stochastic models to simulate scenarios based on different inputs and evaluate the range of possible outcomes under varying conditions. The models are mainly used in situations involving complex environmental systems with multiple interacting variables that behave unpredictably.

The Geographically Weighted Regression (GWR) is a spatial model widely applied to contamination prediction and a local regression model that enables the study of spatially varying relationships between variables [36]. GWR has found extensive applications in environmental research for forecasting the distribution of pollutants, including heavy metals, by accounting for global and local spatial patterns. Spatial dependence can also be explained using other spatial models, such as spatial autoregressive (SAR) and spatial lag (SLM), to examine the impact of adjacent areas on contamination levels.

The biggest problem with spatial models is that they are complex and require precise and high-resolution data [37]. Although these models can be very useful for predicting contamination, they are frequently computationally expensive and require substantial expertise to calibrate and validate. Also, assumptions used in certain spatial models are not necessarily true in real-life situations, and therefore, predictions can be erroneous.

3.4. Integration of Geospatial and Environmental Data

The ability to integrate multiple sources of environmental information is one of the most powerful features of geospatial methods for predicting contamination [5]. By integrating geospatial data and other environmental variables, the researchers can develop more precise and detailed models of heavy metal contamination. This integration also enables a subtler interpretation of the factors that lead to contamination and provides a more comprehensive view of the contamination process.

For example, GIS may be used with remote sensing data to monitor variations in vegetation health and associate them with heavy metal pollution [30]. Moreover, environmental factors such as soil pH, the presence of organic matter, and temperature can be incorporated into spatial models to assess their influence on the mobility and bioavailability of heavy metals. Climatic information, such as precipitation and temperature distributions, can also be included to estimate how these factors may affect the propagation of contamination, especially in regions where heavy metals could be washed away by rain or carried by groundwater.

The incorporation of socioeconomic data is also becoming a significant consideration in predicting contamination. With the inclusion of land-use data, industrial activity records, and population density data, geospatial models can be used to determine where human activities contribute to contamination and to forecast the social and economic consequences of heavy metal pollution. Such holistic data integration would help make contamination predictions more accurate and relevant and may inform more effective policy and remediation plans [38].

Although this has potential, there are challenges encountered in data integration in geospatial modeling. Information found in various sources may differ in resolution, accuracy, and format, making it difficult to combine and analyze. Also, the combination of several variables adds further complexity to the modeling process, necessitating more complicated computational methods and a high level of spatial analysis expertise. However, the ability to integrate diverse data sources is one of the main advantages of geospatial modalities, and it is necessary to develop more precise and comprehensive predictions of contamination [8].

3.5. Challenges and Limitations of Geospatial Approaches

Geospatial methods, such as remote sensing, GIS, and spatial modeling, offer significant benefits for predicting soil heavy metal contamination, but they also have their challenges. Dependence on data quality is one of the major limitations. The quality of geospatial models is no better than the data on which they are built, and poor or incomplete data may lead to poor predictions. E.g., remote sensing data can be constrained by cloud cover, atmospheric degradation, or satellite resolution, and it is hard to detect contamination in some regions. Likewise, GIS models require good-quality soil data, which is not always available, particularly in remote areas or under-researched areas [39].

The complexity of the environmental systems is another problem. Myriad factors affect soil contamination, which include soil properties, land use, climate, and human activities [5]. Geospatial models must account for such factors in a way that best captures their interactions in real life. Although one may model several variables using spatial models, their interactions are not always linear and may not be easily modeled.

Lastly, there is the issue of the scalability of geospatial techniques. Remote sensing and GIS are very useful for large-scale assessment, but may not be as useful for fine-scale, local predictions. At small, highly specific sites, ground-truth data may still require traditional soil sampling to obtain accurate results [27].

4. Artificial Intelligence (AI) and Machine Learning in Soil Contamination Prediction

Artificial Intelligence and Machine Learning are revolutionary technologies that have been increasingly used for environmental monitoring and prediction [6]. Such technologies, especially in relation to soil contamination, are a great improvement over traditional methods, as they enable more precise, efficient, and dynamic models that can predict the occurrence and spread of heavy metals in soil. AI and ML can process complex data, identify trends and correlations, and make predictions that are challenging or impossible for conventional statistical algorithms. The sub-topic discusses how AI and ML can be used to predict soil heavy metal contamination, the types of models, the incorporation of big data, successful case studies, and the benefits and limitations of AI in environmental monitoring.

4.1. AI-Based Predictive Models

The last few years have seen remarkable growth in AI-based predictive models because of their ability to handle large datasets and learn complex patterns without explicit programming. Within the context of soil pollution, machine learning, deep learning, and neural networks as AI techniques have demonstrated their effectiveness in predicting contamination levels, identifying factors that affect pollution, and forecasting future trends in contamination [4].

Machine Learning (ML) models, especially supervised learning methods, have been used to predict soil contamination by training them on historical soil data, including heavy metal concentrations and environmental factors such as soil type, land use, temperature, and precipitation [40]. The usual ML models include support vector machines (SVM), decision trees, random forests, and k-nearest neighbors (KNN), which classify areas based on the likelihood of contamination or estimate the amount of heavy metals present in unmeasured locations. Using these techniques, researchers can develop models that generalize from current data and forecast contamination levels in areas where no soil samples have been obtained. **Table 3** outlines the popular AI methods in soil contamination prediction, their application, strengths, and weaknesses.

Table 3. AI and machine learning techniques applied to soil contamination prediction.

AI/ML Technique	Description	Applications	Advantages	Limitations
Support Vector Machines (SVM)	A supervised learning algorithm used for classification and regression tasks	Classifying contaminated vs. non-contaminated areas, predicting contamination levels	Effective in high-dimensional spaces, robust against overfitting	Requires careful parameter tuning, computationally expensive for large datasets
Decision Trees	Tree-like model used for classification and regression tasks	Predicting contamination levels based on environmental variables	Easy to interpret, handles both numerical and categorical data	Prone to overfitting with small datasets
Random Forests	An ensemble method that constructs multiple decision trees and merges their outputs	Predicting soil heavy metal concentrations based on large datasets	Reduces overfitting, handles large data well	Less interpretable than individual decision trees
Neural Networks (Deep Learning)	Multi-layered networks that learn patterns from complex datasets	Analyzing remote sensing data for contamination patterns, soil contamination forecasting	Can handle complex, non-linear relationships	Requires large datasets and high computational power

Multi-layered neural networks, known as deep learning, which is a subdivision of AI, are especially helpful where the correlation between variables and contamination levels is very non-linear [41]. Deep learning algorithms, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are particularly useful in modeling soil contamination on a large scale, capturing intricate interactions, and learning with very large volumes of data. By way of illustration, satellite images or remote sensing data have been analyzed using CNNs to detect heavy metal pollution. In contrast, RNNs can be used to identify temporal variations, such as seasonal changes in contamination levels.

One of the advantages of AI-based models is flexibility. These models can continually learn and improve as new information becomes available, and they are particularly well-suited to dynamic settings where soil conditions and the extent of contamination can vary over time. In contrast to conventional techniques, which rely on predetermined rules or assumptions, AI models are updated and become more precise as they are exposed to more data, leading to more reliable predictions over time [4].

4.2. Data Integration and Big Data Analytics

The success of AI and ML models in predicting soil contamination is highly dependent on the quality and diversity of the datasets. Conventional soil contamination research is usually based on small, localized samples, which limits the extrapolation of findings. Nevertheless, AI models perform best with big data, which enables the incorporation of diverse data sources to make more precise predictions [6].

The term big data analytics refers to the extraction, analysis, and processing of meaningful patterns from large, complex datasets [41]. Big data for soil contamination prediction can originate from various sources, including soil samples, remote sensing images, meteorological data, land-use data, and socioeconomic factors. For example, satellite images can provide more information on the health of vegetation, land cover, and other environmental factors, which can be compared with levels of contamination. Equally, climate information like precipitation, temperature and humidity may affect the mobility and bioavailability of heavy metals and therefore, the factors are paramount in predicting contamination accurately.

This huge volume of data is especially well-suited to AI models. The machine learning algorithms are able to combine heterogeneous data, to factor in the temporal and spatial variations and to discover patterns in complex multi-dimensional data. To illustrate, AI models can be built using data on soil properties, climate conditions, and land use, combined with data on heavy metal concentrations, to create predictive maps indicating the probability of contamination across regions. Moreover, historical data can be used to predict how things will go in the future, enabling proactive management and remediation [40].

Among the main benefits of big data analytics for predicting soil contamination are the ability to discover hidden patterns and relationships that are not easily observed with conventional tools. AI models can reveal latent relationships between numerous variables (e.g., soil structure, industrialization, and proximity to water bodies) and contamination rates, enabling more precise risk analysis and targeted interventions [36]. Moreover, AI models can improve over time and adjust their predictions as new data emerges, continuously refining them.

4.3. Case Studies of AI in Soil Contamination Prediction

The use of AI and machine learning methods in a number of case studies dedicated to the prediction of soil heavy metal contamination have been a success [6]. An example is a study that was carried out in a mining-affected area where the distribution of heavy metals like arsenic, lead and cadmium in the soil was predicted using ML algorithms. The analysis used soil sample data, remote sensing images, and environmental variables (including vegetation cover and land use) to train machine learning models, achieving high accuracy in predicting contamination levels over a large area. These models could identify soils more likely to be contaminated, enabling more focused soil sampling and remediation.

Another interesting use of AI in predicting soil contamination is in agriculture, where AI models have been employed to identify crops capable of absorbing heavy metals [42]. Using information on soil characteristics, crop types, irrigation, and heavy metal concentrations, machine learning algorithms can predict the likelihood of high heavy metal levels in crops, which is vital for food safety. These models are very handy in areas that have high agricultural practices where heavy metal pollution is becoming a major issue as a result of excessive use of pesticides, fertilizers, and industrial waste.

Convolutional neural networks (CNNs) and other deep learning methods have also been applied to satellite and drone data to identify heavy metal contamination in agricultural and urban soils [43]. CNNs can detect vegetation stress and land-cover changes in high-resolution images, which may indicate soil contamination. These models, when used with other data sources, including soil samples and climate data, can provide real-time forecasts of soil quality and contamination levels.

These case studies demonstrate the potential of AI and ML models not only to predict levels of contamination but also to enhance the accuracy, efficiency, and scalability of environmental monitoring [6]. These models can be useful because they provide details on the distribution and dynamics of heavy metal contamination, enabling better decision-making and more effective management by processing large datasets.

4.4. Advantages of AI in Soil Monitoring

The use of AI for predicting soil contamination offers several notable benefits over conventional approaches. The fact that it can process large, intricate datasets is a major advantage [41]. AI models can process large volumes of data in different forms, such as soil samples, remote sensing data, and environmental variables, to provide more detailed and precise predictions. Conventional approaches, in turn, are often constrained by smaller, less heterogeneous data sets, which may lead to less accurate findings.

The second major benefit of AI models is that they enable real-time predictions. Conventional methods of soil sampling can be time-consuming, as laboratory tests can take a long time to detect contamination and therefore slow down decision-making. Conversely, AI models have the potential to continually update with new information, offer real-time evaluations of soil contamination, and respond more promptly to emerging pollution threats. This is especially crucial in places where contamination levels may change rapidly, such as industrial areas or fields [44].

AI models are also highly flexible and can be enhanced over time as more data is provided. AI models do not rely on fixed rules and assumptions, unlike traditional models, which learn from new data and update their predictions. This adaptability enables AI models to be useful in dynamic settings where contamination patterns and environmental conditions evolve [26].

Moreover, resource allocation and decision-making can be optimized using AI models. By accurately forecasting zones at risk of contamination, AI models can be used to implement specific measures, eliminating expensive and time-consuming soil testing across extensive regions. Such a focused strategy can result in more effective remediation activities and the management of the polluted locations [6].

4.5. Challenges of AI in Soil Contamination Prediction

Even though AI and ML offer numerous benefits for predicting soil contamination, several issues must be overcome before these technologies can realize their potential. Data quality is one of the key challenges. The success of AI models depends on high-quality, accurate data, and faulty predictions can result from faulty or missing data. In other places, especially in developing nations or isolated locations, high-quality soil information is not readily available, making it challenging to create precise AI models [6].

The other difficulty is the complexity of environmental systems. The factors that affect soil contamination are numerous, including soil properties, climatic conditions, land-use practices, and human activities [36]. These complicated interactions are not easily modeled, and much remains to be understood regarding the contribution of these factors to contamination. Also, AI models can be data-intensive to operate effectively, and gathering data to train them can be time- and resource-intensive.

Another typical problem of AI models is overfitting. When an AI model is trained with a small dataset, it can overfit to that particular data, resulting in poor generalization to new, unseen data. It is important to ensure that AI models are trained on diverse, representative datasets to avoid overfitting and enable them to make trustworthy predictions across a wide range of settings [45].

Lastly, there is the issue of interpretability. Many AI models, especially deep learning models, are often called black boxes because they make predictions without revealing the underlying logic. Such non-transparency can be problematic in environmental monitoring, where stakeholders may need to know how a model arrived at a specific conclusion. Attempts to enhance the interpretability of AI models, including explainable AI, are underway and could serve to solve this problem [46].

5. Synergy of Geospatial and AI Approaches

Geospatial techniques have a strong synergy with artificial intelligence-based techniques in predicting and controlling soil heavy metal contamination. Although both methods have their advantages, combining them would increase predictive accuracy, scalability, and real-time monitoring. Geospatial tools provide spatial context for contamination, and AI models can absorb complex, multidimensional data to help detect patterns and forecast future trends [8]. This section discusses the complementary nature of these two approaches, the benefits of integrating them, and the challenges that should be overcome to achieve effective and efficient contamination predictions.

5.1. Integration for Enhanced Predictive Accuracy

A combination of geospatial and AI-based approaches provides a more robust model to predict soil heavy metal contamination [9]. **Table 4** provides a comparative overview of geospatial tools (remote sensing and GIS) that can be used with AI models to enhance soil contamination forecasting.

Through geospatial techniques, including remote sensing and Geographic Information Systems, researchers can obtain comprehensive spatial information, enabling them to understand how contaminants are distributed across large areas. These techniques are known to be effective for hotspot detection, contamination pattern mapping, and high-level spatial analysis. Nonetheless, they generally use indirect proxies (e.g., vegetation health or land use) to determine the contamination levels, which may limit their predictive power [30].

Table 4. Integration of geospatial and AI models for soil contamination prediction.

Integration Approach	Geospatial Component	AI Component	Predicted Outcome	Advantages
Remote Sensing & ML	Satellite imagery or drone data	Machine learning algorithms (e.g., SVM, CNN)	Predicts contamination levels from vegetation stress and surface features	High accuracy, scalable, real-time predictions
GIS & Neural Networks	Geographic data layers (soil type, land use)	Deep learning models (e.g., CNN, RNN)	Identifies contamination hotspots and predicts future contamination	Can process large datasets and capture complex relationships
Spatial Models & Random Forests	Spatial data for contamination spread	Random forest algorithms	Models contamination spread over time, and predict risk zones	Robust and interpretable, effective in handling complex datasets
Big Data Integration	Remote sensing, weather, land use, soil properties	Ensemble models, AI-driven data fusion	Real-time, dynamic prediction of contamination across large regions	Dynamic, up-to-date predictions, multi-dimensional data integration

Instead, AIs are efficient at examining extensive, complex data and can discover hidden trends that conventional models can hardly detect. By combining AI with geospatial data, researchers can leverage both the spatial context and the predictive capabilities of machine learning algorithms to produce more accurate predictions. As an illustration, AI models can process geospatial data on soil characteristics, vegetation health, land use, and environmental factors such as temperature and precipitation to forecast heavy metal levels [42]. Such a combination can provide more in-depth insight into the factors affecting contamination, resulting in greater predictive power.

The relationship between AI-based and geospatial methods is especially relevant in predictive mapping. For example, a geospatial model that identifies high-contamination regions can be improved with an AI model trained on historical data to predict future levels of contamination [40]. This makes it possible to create dynamic, real-time maps that not only show existing contamination but also forecast its future development. These types of models can inform policy responses, cleanup actions, and land-use plans, providing a more practical and precise view of the risks of soil contamination. **Figure 2** shows how AI models can be used alongside geospatial tools, including remote sensing and GIS, to improve the prediction of soil contamination.

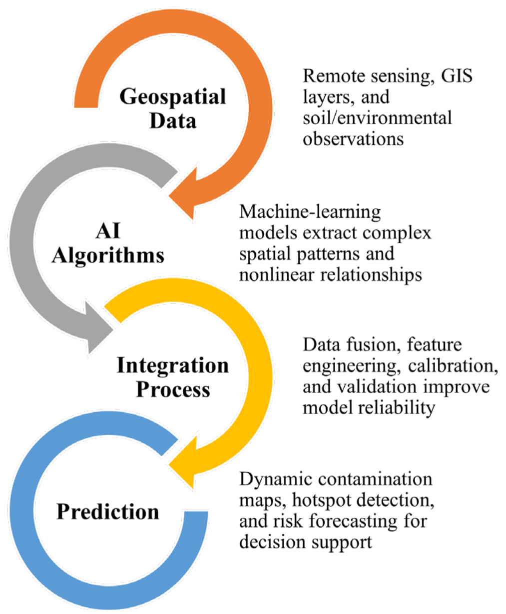

Figure 2. Integration of geospatial and AI models for soil contamination prediction.

5.2. Multidimensional Data Fusion

The ability to integrate multiple data sources to enhance predictive models is one of the most important benefits of combining geospatial and AI approaches [5]. Various factors that affect soil contamination include soil composition, land use, climatic conditions, and industrial activity. Geospatial methods can deliver useful spatial information, whereas AI models can combine this information with other environmental variables to generate multidimensional models that account for the complex interactions among these variables.

An example is using satellite data with soil properties, including organic matter content and pH, to determine the risk of contamination [47]. Some of the environmental variables that can also be factored in the model include rainfall patterns, temperature, and land use history because these factors affect the mobility and bioavailability of the heavy metals in the soil. By fusing these datasets, AI models will be able to forecast contamination hotspots and the most vulnerable areas to heavy metals much more accurately. This data fusion method is not only more accurate in making predictions but also allows researchers to discover factors that could cause soil pollution they had not previously considered.

Furthermore, geospatial and AI data can be combined to create decision-support systems that provide real-time forecasts [9]. With a constant stream of information from sources such as remote sensing platforms, weather stations, and soil sensors, AI models can refresh contamination forecasts in real time and respond more quickly to new contamination hazards. This dynamic methodology offers a significant advantage over conventional techniques, which generally provide a single, static evaluation of contamination rates.

5.3. Case Studies of Geospatial + AI Models

A number of case studies have shown how geospatial and AI methods can be successfully integrated to predict heavy metal contamination in soil. An example of this is the use of machine learning algorithms to forecast the spatial distribution of heavy metals in areas affected by mining. In an experiment in a mining region, the researchers used remote sensing data, including multispectral imagery and soil sample data, to train machine learning models that forecast the presence of arsenic, cadmium, and lead in the soil. The combination of geospatial data and AI made it possible to identify contamination hotspots and create predictive models that forecast future contamination levels based on environmental factors such as precipitation and land use [9].

In a different study, the contamination of agricultural soils by heavy metals was predicted using AI-based models. Researchers combined GIS data on land use and crop types with soil sample data and climate variables to develop a predictive model of contamination risk. This model employed machine learning algorithms to understand the multifaceted relationships among land use, soil properties, and weather conditions, producing risk maps that could be used by farmers and policymakers to reduce contamination and safeguard food safety [43].

These case studies underscore the power of integrating geospatial and AI to enhance the accuracy and scalability of soil contamination prediction. By integrating geospatial methods with AI models, researchers will be able to develop more dynamic, real-time predictions that detect contamination risks and forecast future trends. This combined method provides a more holistic view of the factors that cause soil contamination, with subsequent effective strategies for addressing and remediation [12].

5.4. Challenges in Integration

Even though integrating geospatial and AI solutions has numerous advantages, there are several obstacles to overcome for successful integration. The variability and quality of the data are among the major challenges. Low resolution, cloud cover, and atmospheric interference are common problems in geospatial data, such as satellite imagery and remote sensing data, that can reduce prediction accuracy [39]. Likewise, data on soil samples may be sparse, particularly in remote or under-researched areas, making it challenging to develop credible models.

Combining these various data sources, with their respective advantages and constraints, requires advanced data processing methods. A typical problem is that the spatial resolution of geospatial data (satellite imagery, etc.) and other environmental data (soil samples, etc.) differs. Combining high-resolution geospatial data with lower-resolution datasets may lead to a mismatch in spatial scales, which can impair prediction accuracy. Moreover, the combination of various data types, including satellite, soil, and meteorological data, needs to be carefully harmonized so that the models can efficiently handle and analyze the data.

The computational complexity of combining large, multidimensional data is another obstacle [48]. AI-based and geospatial models can be computationally intensive and require substantial resources to compute and analyze data. This may be especially tricky when large geographic areas are involved or when real-time model updates are required. Researchers and policymakers must strike a balance between the need to achieve high accuracy and the practical constraints of computing power and resources.

And lastly, there is the issue of interpretability [46]. In many cases, AI models, in particular deep learning algorithms, can be black-box, i.e., it can be hard to know how the model came to its conclusions. Such intransparency may be critical when making environmental decisions, as stakeholders have to know what underlies the predictions of a model. Projects to develop explainable AI (XAI) are still in progress and could help resolve this problem by providing insights into how AI models make decisions.

5.5. Future Directions for Geospatial and AI Integration

Nevertheless, the future of combining geospatial and AI methods in predicting soil heavy metal contamination is bright. Technological improvements, especially those related to data gathering and processing, will likely enhance data quality and accessibility, enabling the combination of various datasets. As an illustration, the increasing use of Internet of Things (IoT) sensors for soil monitoring is likely to provide real-time, high-frequency data on soil conditions that can be integrated into AI models to enhance prediction precision [9].

Moreover, the development of faster computational tools and cloud computing will enable the processing and analysis of big data in real time, allowing predictions to be updated more often. It will increase the chances of monitoring soil contamination in real time and responding more promptly to emerging contamination threats. High-performance computing (HPC) and distributed computing systems will also enable the management of large-scale models that combine many data sources and account for complex interactions among variables [49].

One more point where growth can be achieved is the use of citizen science and crowdsourced data [50]. With the increased accessibility of smartphones and remote sensing platforms, individuals and local communities will be able to provide data on soil contamination, which will help fill data gaps, especially in remote or underserved regions. This crowdsourced data can be incorporated into AI models, further enhancing the accuracy and scalability of contamination predictions.

Lastly, explainable AI will enhance the interpretability and transparency of AI models as it continues to develop [46]. AI will assist in developing trust among the stakeholders, including policymakers and environmental managers, and allow them to make more effective decisions by shedding light on how AI models generate predictions. Data-driven, informed decision-making for soil contamination management and remediation will be enabled by integrating AI with geospatial models.

6. Conclusions

Heavy metal contamination of soil is a growing environmental challenge that poses serious threats to the environment, human health, and food safety. Conventional techniques for predicting and monitoring soil contamination, however important, are proving inadequate to serve the needs of large, dynamic environmental systems. With the increased complexity and spread of soil pollution in the future, the next-generation technologies, specifically geospatial tools and AI-driven models, provide a good perspective to fill the gaps in the existing monitoring practices.

The combination of geospatial techniques, including remote sensing, GIS, and spatial modeling, with AI technologies represents a new approach to predicting heavy metal contamination in soil. Geospatial tools can provide spatial context, allowing hotspots of contamination to be identified and the distribution of contamination to be mapped. In contrast, AI models are good at processing complex, multi-dimensional data, identifying patterns, and making predictions. The combination of these technologies provides greater accuracy, improved prognostic prediction, and scalability, making environmental management and intervention measures more effective.

Nevertheless, several challenges remain, including data quality, computational complexity, and the interpretability of AI models. Recent efforts to integrate explainable AI (XAI) into health risk modeling and to develop automated GIS-based hazard assessment tools represent important steps toward overcoming these barriers, yet cross-scale prediction, linking local sampling data with regional-scale models—remains a persistent challenge. It will be important to overcome these barriers to realize the full potential of these next-generation methods. With the future progress of remote sensing, big data analytics, and AI, geospatial and AI synergy will provide more valuable data to inform proactive, effective action to maintain soil health, safeguard human health, and ensure sustainable land-use practices.

In summary, the combination of geospatial and AI technologies represents a revolutionary change in how soil heavy metal contamination is predicted and controlled. Further research and development in this field will certainly help make environmental monitoring more efficient, providing the necessary means to fight soil pollution on an international level.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

All the data is presented in the article.

Conflicts of Interest

The author declares no conflict of interest.

AI Use Statement

The author declares that no artificial intelligence (AI) tools were used in the preparation of this manuscript.

References

[1] Mishra, S., Bharagava, R.N., Yadav, A., et al., 2018. Heavy metal contamination: An alarming threat to environment and human health. In: Sobti, R., Arora, N., Kothari, R. (Eds.). Environmental Biotechnology: For Sustainable Future. Springer: Singapore. pp. 103–125.

[2] Khalef, R.N., Hassan, A.I., Saleh, H.M., 2022. Heavy metal's environmental impact. In: Saleh, M., Hassan, I. (Eds.). Environmental Impact and Remediation of Heavy Metals. IntechOpen: London, UK.

[3] Desaules, A., 2012. Critical evaluation of soil contamination assessment methods for trace metals. Science of the Total Environment. 426, 120–131.

[4] Pranjal, P., Soni, A., 2025. Next-Generation Geographical Information Systems: Integrating Artificial Intelligence, Geographic Information Systems, and Satellite Imagery for Mapping and Surveying. In: Burnett, C.M., Pradhan, B. (Eds.). Global Land-Ocean Geospatial Applications. CRC Press: Boca Raton, FL, USA. pp. 216–233.

[5] Hou, D., O'Connor, D., Nathanail, P., et al., 2017. Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. Environmental Pollution. 231, 1188–1200.

[6] Rosca, C.-M., Stancu, A., 2025. Emerging trends in AI-based soil contamination monitoring and prevention. Agriculture. 15(12), 1280.

[7] Howari, F., 2025. AI–Remote Sensing for Soil Variability Mapping and Precision Agrochemical Management: A Comprehensive Review of Methods, Limitations, and Climate-Smart Applications. Agrochemicals. 5(1), 1.

[8] Dritsas, E., Trigka, M., 2025. Remote sensing and geospatial analysis in the big data era: A survey. Remote Sensing. 17(3), 550.

[9] Wu, C., Lu, S., Tian, J., et al., 2024. Current situation and prospect of geospatial AI in air pollution prediction. Atmosphere. 15(12), 1411.

[10] Adıgüzel, F., Tabar, M.E., Yetis, C., et al., 2026. Health Risk Modeling of Heavy Metal Contamination in Urban Playground Dust in Ankara (Türkiye): Combining Explainable AI with Multi-Heuristic Optimization. Process Safety and Environmental Protection. 213, 108991.

[11] Adıgüzel, F., Karadeniz, E., Karlı, R.G., et al., 2026. An advanced GIS based hazard index tool for an automated health risk assessment framework. Scientific Reports. 16(1), 17249.

[12] Selvam, R., Kalaiyarasi, G., Saritha, B., 2024. Heavy Metal Contamination in Soils: Risks and Remediation. In: Goswami, S.P., Sachan, K., Singh, V., et al. (Eds.). Soil Fertility and Plant Nutrition. Golden Leaf Publisher: Lucknow, India. pp. 141–160.

[13] Zeiner, M., Pirkl, R., Juranović Cindrić, I., 2020. Field-tests versus laboratory methods for determining metal pollutants in soil extracts. Soil and Sediment Contamination: An International Journal. 29(1), 53–68.

[14] Tüzen, M., 2003. Determination of heavy metals in soil, mushroom and plant samples by atomic absorption spectrometry. Microchemical Journal. 74(3), 289–297.

[15] Ionescu, A., Cristea, M., Matache, M., et al., 2024. Maximising precision and accuracy in soil sampling using automated technological processes. Annals of the Faculty of Engineering Hunedoara-International Journal of Engineering. 22(3), 35–40.

[16] Pennock, D., Yates, T., Braidek, J., 2007. Soil sampling designs. In: Carter M.R., Gregorich, E.G. (Eds.). Soil Sampling and Methods of Analysis, 2nd ed. CRC Press: Boca Raton, FL, USA. pp. 1–14.

[17] Boulding, J.R., 2017. Description and Sampling of Contaminated Soils: A Field Guide. CRC Press: Boca Raton, FL, USA. .

[18] Ripin, S.N.M., Hasan, S., Kamal, M., 2014. Environmental geochemical mapping on distribution of metal contamination in topsoils Perlis, Malaysia. Journal of Medical and Bioengineering. 3(4), 277–281.

[19] Johnson, C.C., Ander, E.L., 2008. Urban geochemical mapping studies: How and why we do them. Environmental Geochemistry and Health. 30(6), 511–530.

[20] Hosseini-Dinani, H., Mokhtari, A.R., Shahrestani, S., et al., 2019. Sampling density in regional exploration and environmental geochemical studies: A review. Natural Resources Research. 28(3), 967–994.

[21] Jiang, H.-H., Cai, L.-M., Wen, H.-H., et al., 2020. An integrated approach to quantifying ecological and human health risks from different sources of soil heavy metals. Science of the Total Environment. 701, 134466.

[22] Ashraf, M.A., Maah, M.J., Yusoff, I., 2014. Soil contamination, risk assessment and remediation. In: Hernandez-Soriano, M.C. (Ed.). Environmental Risk Assessment of Soil Contamination. IntechOpen: London, UK. pp. 3–56.

[23] Peijnenburg, W.J.M., 2025. Bioavailability of heavy metals in soil: A review of tools, models, and regulatory applications. Environmental and Biogeochemical Processes. 1(1), e011.

[24] Markus, J., McBratney, A.B., 2001. A review of the contamination of soil with lead II. Spatial distribution and risk assessment of soil lead. Environment International. 27(5), 399–411.

[25] Huang, S.-W., Jin, J.-Y., 2008. Status of heavy metals in agricultural soils as affected by different patterns of land use. Environmental Monitoring and Assessment. 139(1–3), 317–327.

[26] Fan, Y., Wang, X., Funk, T., et al., 2022. A critical review for real-time continuous soil monitoring: Advantages, challenges, and perspectives. Environmental Science & Technology. 56(19), 13546–13564.

[27] Hussain, A., Bangroo, S.A., Muslim, M., 2023. Assessment of soil contamination using remote sensing and spatial techniques. In: Mushtaq, F., Farooq, M., Mukherjee, A.B., et al. (Eds.). Geospatial Analytics for Environmental Pollution Modeling: Analysis, Control and Management. Springer: Cham, Switzerland. pp. 249–266.

[28] Yaseen, Z.M., 2022. The next generation of soil and water bodies heavy metals prediction and detection: New expert system based Edge Cloud Server and Federated Learning technology. Environmental Pollution. 313, 120081.

[29] Abdulraheem, M.I., Zhang, W., Li, S., et al., 2023. Advancement of remote sensing for soil measurements and applications: A comprehensive review. Sustainability. 15(21), 15444.

[30] Lovynska, V., Bayat, B., Bol, R., et al., 2024. Monitoring heavy metals and metalloids in soils and vegetation by remote sensing: A review. Remote Sensing. 16(17), 3221.

[31] Henshaw, S.L., Curriero, F.C., Shields, T.M., et al., 2004. Geostatistics and GIS: Tools for characterizing environmental contamination. Journal of Medical Systems. 28(4), 335–348.

[32] Jalhoum, M.E., Abdellatif, M.A., Mohamed, E.S., et al., 2024. Multivariate analysis and GIS approaches for modeling and mapping soil quality and land suitability in arid zones. Heliyon. 10(5), e27577.

[33] Shen, F., Xu, C., Wang, J., et al., 2024. A new method for spatial three-dimensional prediction of soil heavy metals contamination. CATENA. 235, 107658.

[34] Liu, G., Zhou, X., Li, Q., et al., 2020. Spatial distribution prediction of soil As in a large-scale arsenic slag contaminated site based on an integrated model and multi-source environmental data. Environmental Pollution. 267, 115631.

[35] Romanowicz, R., Macdonald, R., 2005. Modelling

uncertainty and variability in environmental systems. Acta Geophysica Polonica. 53(4), 401–417.
[36] Li, C., Li, F., Wu, Z., et al., 2017. Exploring spatially varying and scale-dependent relationships between soil contamination and landscape patterns using geographically weighted regression. Applied Geography. 82, 101–114.
[37] Li, J., 2019. A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R. Applied Sciences. 9(10), 2048.
[38] Cui, H.-L., Gao, S.H., Wang, H.C., et al., 2026. Big data integration for environmental risk assessment of emerging contaminants. Nature Sustainability. 9(2), 196–206.
[39] Frew, R.J., 2016. GIS-Based Accessibility Modelling as a Means of Evaluating Geospatial Data Usability [PhD Thesis]. University of South Wales: Cardiff, UK.
[40] Zhao, W., Ma, J., Liu, Q., et al., 2023. Accurate prediction of soil heavy metal pollution using an improved machine learning method: A case study in the Pearl River Delta, China. Environmental Science & Technology. 57(46), 17751–17761.
[41] Szramowiat-Sala, K., 2023. Artificial intelligence in environmental monitoring: Application of artificial neural networks and machine learning for pollution prevention and toxicity measurements. Preprints. 2023071298. DOI: https://doi.org/10.20944/preprints202307.1298.v1
[42] Hu, B., Xue, J., Zhou, Y., et al., 2020. Modelling bioaccumulation of heavy metals in soil-crop ecosystems and identifying its controlling factors using machine learning. Environmental Pollution. 262, 114308.
[43] Liu, Q., Du, B., He, L., et al., 2023. Digital soil mapping of heavy metals using multiple geospatial data: Feature identification and deep neural network. Ecological Indicators. 154, 110863.
[44] Karri, N., 2024. Real-Time Performance Monitoring with AI. International Journal of Emerging Trends in Computer Science and Information Technology. 5(1), 102–111.
[45] Aliferis, C., Simon, G. (Eds.), 2024. Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. In Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls. Springer: Cham, Switzerland. pp. 477–524.
[46] Wang, Q., Li, C., Hao, D., et al., 2023. A novel four-dimensional prediction model of soil heavy metal pollution: Geographical explanations beyond artificial intelligence "black box". Journal of Hazardous Materials. 458, 131900.
[47] Ong, C.C.H., Cudahy, T.J., 2014. Mapping contaminated soils: Using remotely-sensed hyperspectral data to predict pH. European Journal of Soil Science. 65(6), 897–906.
[48] Liakos, L., Panagos, P., 2022. Challenges in the geo-processing of big soil spatial data. Land. 11(12), 2287.
[49] Gathu, S., 2024. High-performance computing and big data: Emerging trends in advanced computing systems for data-intensive applications. Journal of Advanced Computing Systems. 4(8), 22–35.
[50] Welvaert, M., Caley, P., 2016. Citizen surveillance for environmental monitoring: Combining the efforts of citizen science and crowdsourcing in a quantitative data framework. SpringerPlus. 5(1), 1890.