

Article

Machine Learning-Based Prediction of Heavy Metal Contamination and Ecological Risk in Karst Agricultural Soils

Zhe Liu ^{1,2,†}, Juan Wu ^{3,†}, Jie Li ^{1,2,*}, Guodong Zheng ^{1,2}, Jianxun Qin ^{1,2}, Wenbo Gu ^{1,2} and Jiakai Li ⁴

¹ Geological Survey of Guangxi Zhuang Autonomous Region, Nanning 530023, China; liuzhe@gxdzdcy.cn (Z.L.); zhengguodong@gxdzdcy.cn (G.Z.); qinjianxun@gxdzdcy.cn (J.Q.); guwenbo15994461331@163.com (W.G.)

² Medical Geological Engineering Center of Guangxi Zhuang Autonomous Region, Nanning 530023, China

³ Guangxi Water and Power Design Institute Co., Ltd., Nanning 530023, China; wujuan1706@smail.nju.edu.cn

⁴ China Nonferrous Metals Geology and Mining Co., Ltd., Guilin 541004, China; lijikai@cnmc.com.cn

* Correspondence: lijie@gxdzdcy.cn

† These authors contributed equally to this work.

Abstract

Investigating multiple source apportionment methods and quantitatively characterizing heavy metal contamination in soils are of critical importance for effective pollution control and prevention. This study systematically investigates multiple source apportionment methods for soil heavy metals, with quantitative characterization of contamination features crucial for effective pollution control. Taking Jingxi City in Guangxi, China, as a case study, we conducted a comprehensive analysis of 8816 soil samples using multi-source big data integration. By synergistically applying machine learning algorithms, the potential ecological risk index, and bivariate local Moran's index, we achieved dual objectives: quantitative inversion of eight heavy metal concentrations and simultaneous ecological risk assessment with pollution source identification. Through comparative model evaluation, the XGBoost algorithm demonstrated optimal predictive performance. Contribution analyses revealed that soil properties (Fe_2O_3 , Al_2O_3 , and phosphorus content), road distribution, and elevation significantly regulate heavy metal accumulation. Spatial risk mapping identified cadmium, mercury, and arsenic contamination hotspots as critical environmental threat zones. The bivariate local Moran's index model elucidated spatial coupling characteristics between ecological risks and environmental drivers, providing spatially explicit decision-making support for precision environmental management. Our multidimensional analytical framework incorporates spatial visualization of heavy metal distribution, hierarchical ecological risk assessment, and pollution source contribution analysis, ultimately establishing a scientific decision-making system for land safety utilization and pollution risk management. This integrated approach offers methodological references for regional heavy metal pollution control in karst environments.



Academic Editor: Nick B. Comerford

Received: 14 January 2026

Revised: 29 January 2026

Accepted: 7 February 2026

Published: 11 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: karst area; soil heavy metals; machine learning; risk assessment

1. Introduction

Heavy metals are typical cumulative pollutants characterized by significant biotoxicity, non-degradability, and persistence [1]. According to the China Soil Pollution Bulletin, soil heavy metal pollution in China exhibits significant regional disparities, with a relatively extensive range of exceedance observed in the southwestern region [2]. Karst refers to distinctive geomorphological and hydrological landscape phenomena formed by soluble

rocks under hydrothermal conditions in the Earth's crust [3,4]. Karst areas, despite their abundant agricultural, forestry, mineral, water, and tourism resources, exhibit relatively low environmental carrying capacity and are recognized as a fragile ecosystem [5]. China possesses the world's largest karst area in terms of territorial extent. The contiguous exposed karst areas in Yunnan, Guizhou, and Guangxi provinces collectively account for 23.4% of China's total karst coverage and represent 39.7% of the land area within these three provinces [6]. The karst region spanning Yunnan, Guizhou, and Guangxi provinces constitutes a critical headwater area for the Pearl River's upper tributaries and serves as an ecological buffer zone for the Pearl River Delta. However, decades of intensive mining development in this region have led to substantial environmental degradation. The elevated geological background, coupled with anthropogenic activities, undermines the soil environmental carrying capacity, exacerbating regional soil contamination. This dual stress not only compromises the stability of agricultural soil ecosystems but also diminishes crop yield security, quality, and safety, thereby posing significant risks to human health. Consequently, environmental pollution and ecological risks associated with heavy metals in karst areas have garnered increasing attention. Investigating the concentration characteristics of heavy metals in agricultural soils, identifying their sources, and assessing potential ecological risks holds critical importance for local agricultural environmental protection and human health.

Traditional methods for source apportionment of soil heavy metal pollution primarily rely on the integration of chemical analysis, statistical methods, and spatial analysis to establish a fundamental framework for pollution source identification [7–9]. For instance, Wang et al. [10] applied a combined positive matrix factorization (PMF) and Bayesian stable isotope mixing model to trace heavy metal sources in soils of an e-waste processing area in China. Their findings revealed that industrial activities constituted the predominant contamination source, with industrial zones contributing 64.2% compared to 35.6% from agricultural lands. Source apportionment identified specific contributions from e-waste dismantling (25.1%), manufacturing processes (23.7%), leaded/unleaded gasoline emissions (9.1%/20.2%), and natural processes (21.9%). Ha et al. [11] integrated principal component analysis (PCA) and Kriging interpolation to analyze 11 metals in 2046 soil samples from industrial zones. Their results demonstrated significant spatial heterogeneity in heavy metal concentrations across sampling areas except for nickel, with lead (Pb), cadmium (Cd), copper (Cu), and zinc (Zn) primarily associated with anthropogenic activities such as foundry operations and railway transportation. In contrast, cobalt (Co), manganese (Mn), and vanadium (V) were predominantly governed by natural sources, including soil texture, pedogenic processes, and hydrological conditions. Huang et al. [12] developed an enhanced PCA-MLRD model integrating principal component analysis (PCA) and multiple linear regression with distance (MLRD), which was applied to arsenic (As)-, cadmium (Cd)-, mercury (Hg)-, and lead (Pb)-contaminated soils in peri-urban areas of Southeastern China. This model demonstrated high precision in both source identification and quantification, revealing Zn-Pb mining activities as the dominant anthropogenic pollution source. However, these methods exhibit dependencies on prior studies [13] and subjective interpretations [14], which constrain the accuracy of their assessments regarding the impacts of critical environmental factors. Additionally, certain approaches incur high operational costs, while their data analysis frameworks demonstrate limitations in resolving nonlinear relationships and deciphering complex intercorrelations.

In recent years, machine learning models have emerged as essential tools for predicting the spatial distribution of soil heavy metals [15–17]. Models such as Random Forest (RF) [18], Support Vector Machine (SVM) [16], Gradient Boosting Decision Tree (GBDT) [19] and eXtreme Gradient Boosting (XGBoost) [20] have demonstrated superior performance in

identifying influencing factors and predicting concentrations of heavy metals. For instance, Wang et al. [18] employed a Random Forest (RF) model to predict the concentrations and spatial distributions of Pb, Cd, Cr, As, Hg, and Zn in soils. The model demonstrated that anthropogenic activities and transportation constituted the primary sources of Pb, while sewage irrigation was identified as the dominant source of Cd, Cr, and Zn. Atmospheric deposition from coal-fired power plants was recognized as a significant contributor to Hg, and parent materials were determined as the main source of As. Wang et al. [16] utilized visible–near-infrared spectroscopy to predict soil heavy metals, comparing three nonlinear machine learning methods—Cubist regression tree, Gaussian process regression (GPR), and support vector machine (SVM)—with partial least squares regression (PLSR) to identify the optimal prediction model. The study revealed that nonlinear machine learning models significantly outperformed PLSR in most scenarios. Specifically, the SVM model demonstrated superior predictive accuracy and enhanced generalization capability for Zn, Cu, Cr, Pb, the Nemerow integrated pollution index (NIPI), and the potential ecological risk index (RI). However, the accumulation processes of heavy metals in soils are often co-regulated by natural environmental factors [21], socioeconomic drivers [21], and their complex interactions, including hydrothermal cycling [22], atmospheric circulation patterns [23], urbanization processes [24], topographical features [25], and soil properties [26]. This multifaceted regulation necessitates the development of interpretable and multidimensional analytical approaches to more precisely identify the driving mechanisms underlying heavy metal distribution patterns. Such methodological advancements could provide robust scientific foundations for targeted remediation strategies and comprehensive risk assessment of soil heavy metal contamination.

To address these challenges, this study aims to achieve three primary objectives through machine learning models and comprehensive data analytics: (1) develop high-precision predictive models for heavy metal concentrations in arable soils; (2) identify key natural and anthropogenic drivers via feature importance analysis; and (3) generate spatial ecological risk maps to inform zonal remediation strategies. Methodological innovations include integrating multisource heterogeneous data (geological, climatic, and land-use variables) and developing interpretable machine learning frameworks that overcome traditional models' limitations in complex system analysis. We selected Jingxi City in Guangxi Zhuang Autonomous Region, China, as the study area. This region serves as a vital rice-producing base in Guangxi and is nationally recognized as a prominent cultivation center for *Panax notoginseng* (Sanqi). We measured the concentrations of Cr, Pb, Zn, Cu, Ni, Cd, As, and Hg in soils and collected 16 potential drivers through field sampling and geospatial data acquisition. This dataset will lead to the development of a science-driven framework for arable soil management. This study has three primary objectives: to select and develop the optimal machine learning model from multiple candidates for accurately predicting heavy metal concentrations in arable soils; to identify and quantify the primary drivers governing heavy metal accumulation in these soils; and to utilize the developed model for assessing ecological risks associated with elevated heavy metal levels.

2. Methods and Materials

2.1. Study Area

Jingxi City, located in the Guangxi Zhuang Autonomous Region (Figure 1), is situated in the southwestern part of Guangxi and is characterized by typical karst terrain. It spans from 105°56' to 106°48' east longitude and 22°51' to 23°34' north latitude. Jingxi City borders Vietnam, with a boundary line of 152.5 km, and covers a total area of approximately 3327 square kilometers. The city experiences a subtropical monsoon climate, with an average annual temperature of 19.1 °C. It has a total population of 488,200 and a cultivated

land area of 503,100 mu (approximately 33,540 hectares). Jingxi is not only a significant grain-producing area in Guangxi but also renowned as the “hometown of *Panax notoginseng*” (a traditional Chinese medicinal herb). Therefore, this study focuses on investigating the pollution of eight heavy metals in the cultivated land of Jingxi City.

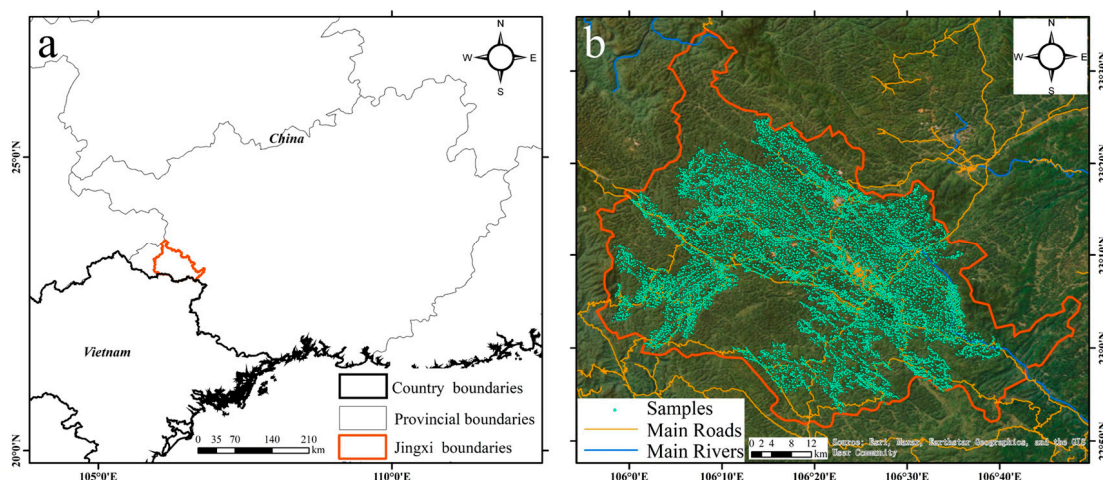


Figure 1. The geographical location of Jingxi City (a) and the distribution map of sampling points (b).

2.2. Sample Collection and Analysis

From December 2024 to June 2025, a total of 8816 topsoil samples (0–20 cm depth) were collected across Jingxi City, covering representative farmland, vegetable plots, forest (orchard) lands, grasslands, and mountainous/hilly areas with thick soil layers. Sampling sites were deliberately positioned away from ditches, forest belts, and field ridges. The geographic coordinates of each sampling point were recorded using a handheld GPS device, which was calibrated prior to fieldwork. At each designated GPS location, 4–6 sub-samples were collected within a 5–10 m radius around the central point and thoroughly mixed to form a composite sample. Post collection, visible impurities (e.g., plant roots, stones, and insects) were manually removed. Approximately 1.0 kg of soil was placed in a clean cotton bag, with each bag further protected by an outer polyethylene plastic sleeve to prevent cross-contamination during transport. Air-dried samples were subsequently passed through a 2 mm sieve for homogenization.

The concentrations of heavy metals in soil were determined following established methods in previous studies [26]. Specifically, Cr, Pb, and Zn were measured using X-ray fluorescence spectrometry (XRF). Cu was analyzed by inductively coupled plasma optical emission spectrometry (ICP-OES). Ni and Cd were quantified via inductively coupled plasma mass spectrometry (ICP-MS). As and Hg were detected using atomic fluorescence spectrometry (AFS). Soil pH was measured using a pH meter (PHS-2C, Shanghai, China). The limits of detection for As, Cd, Cr, Cu, Hg, Ni, Pb, and Zn were 0.5, 0.02, 3, 1, 0.0005, 1.5, 2, and 4 mg/kg, respectively. The recoveries of all heavy metals for the certified reference materials were within the range of 90–110%.

2.3. Data Collection

A total of 22 environmental factors were selected to evaluate the sources and contributions of soil heavy metals, including natural factors (e.g., pH, soil type, normalized difference vegetation index [NDVI]) and anthropogenic factors (e.g., population density, distance to roads, mining activities). Eleven soil chemical properties (e.g., pH, organic matter content, metal oxides) were determined by laboratory analysis, and the results are presented in Supplementary Table S1. Soil type and digital elevation model (DEM) data were recorded in the field during sampling, while the remaining eight environmental

factors (e.g., NDVI, slope, population density) were obtained from databases. The methods for measuring soil chemical properties and the procedures for acquiring environmental data are detailed in Supplementary Text S1 and Table S2.

2.4. Machine Learning Patterns

To predict soil heavy metal concentrations more rapidly and efficiently, we applied and compared the performance of ten machine learning algorithms—namely, Artificial Neural Network (ANN), Decision Tree (DT), Gradient Boosting Decision Tree (GBDT), k-Nearest Neighbors (KNN), Light Gradient Boosting Machine (LightGBM), Neural Network (NN), Random Forest (RF), Support Vector Machine (SVM), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost)—on a dataset comprising 8816 soil samples. Data preprocessing, feature selection, model training, and evaluation were conducted using the Scikit-learn library in Python (<https://scikit-learn.org/stable/index.html>, accessed on 28 July 2025). The hyperparameters for each model were optimized using exhaustive grid search. Source code (XGBoost) is publicly available in the Supplementary Text S2.

Given the large sample size of the dataset, a 60:40 ratio was adopted for the training and prediction sets. The coefficient of determination (R^2) and root mean square error (RMSE) were calculated to evaluate the predictive performance of each model. Our findings indicate that a higher R^2 and a lower RMSE reflect better model fitting. The R^2 and RMSE values were calculated using Python 3.13 code (see Supplementary Text S2).

2.5. Potential Ecological Risk Models

The Potential Ecological Risk Index (PERI) method has been widely employed due to its straightforward calculation and intuitive evaluation results [27]. The index is calculated using the following equations:

$$I_R = \sum_{i=1}^n E_r^i = \sum_{i=1}^n T_r^i \times C_f^i \quad (1)$$

$$C_f^i = \frac{C_s^i}{C_n^i} \quad (2)$$

where I_R is the integrated potential ecological risk index; E_r^i denotes the ecological risk index for heavy metal i ; and T_r^i represents the toxic response coefficient for heavy metal i , reflecting its toxicity level and ecological sensitivity. The T_r^i values for Hg, Cd, Pb, Cr, Cu, Zn, As and Ni are 40, 30, 5, 2, 5, 1, 10, and 5, respectively [28]; C_f^i is the contamination factor of heavy metal i ; C_n^i refers to the reference concentration of heavy metal i (mg/g), typically the soil background values of Guangxi [29,30]. The classification criteria for ecological risk levels are presented in Supplementary Table S3.

2.6. Spatial Bivariate Correlation Analysis

To explore the local spatial relationship between two variables, we applied the Bivariate Local Moran's I statistic, which allows the detection of spatial clusters and outliers involving different attributes across neighboring spatial units. The spatial analysis was conducted using Python 3.11, leveraging the *esda* and *libpysal* modules within the *PySAL* library.

Specifically, variable x (heavy metals) and variable y (e.g., environmental factors) were extracted from the spatial dataset. A spatial weights matrix was constructed based on Queen contiguity and subsequently row-standardized. The Bivariate Local Moran's I statistic was computed using the *Moran_Local_BV* class, which evaluates the local correlation between the value of variable x at a given location and the average of variable y in its neighboring areas.

The resulting local Moran's I_i values (I_i), and their pseudo-significance levels (p -values) were used to classify spatial association patterns into four categories: High–High (HH), Low–Low (LL), High–Low (HL), and Low–High (LH). These spatial clusters and outliers were then visualized through a choropleth map using the matplotlib and geopandas libraries, facilitating an intuitive understanding of the local bivariate spatial structure.

2.7. Network Analysis

The correlation network among eight heavy metals and twenty-two environmental variables was constructed using Python 3.13. Pairwise Spearman correlation coefficients were computed across all variables, and only those with an absolute correlation coefficient ($|R|$) exceeding 0.5 and a significance level of $p < 0.05$ were retained to generate the network structure.

2.8. Data Processing

Data preprocessing was conducted in ArcGIS 10.7, where categorical variables were transformed into dummy variables and all distance metrics were standardized to Euclidean form. The spatial distributions of the 20 processed environmental factors were individually visualized as thematic maps (Supplementary Figure S1). Additionally, inverse distance weighting (IDW) interpolation was employed to generate a spatial distribution map of the ecological risk posed by soil heavy metals. A suite of machine learning models was developed in the Python 3.11 environment to predict and analyze heavy metal concentrations. Statistical analyses were performed using R software (version 4.4.2). All remaining data processing and figure generation were carried out with Origin 2019 and Microsoft Excel 2019, respectively.

3. Results and Discussion

3.1. Statistical Evaluation of Heavy Metal Concentrations and Chemical Properties

The concentration ranges of eight heavy metals in the soil samples were as follows (Table 1): cadmium (Cd), 26–45,407 $\mu\text{g kg}^{-1}$; mercury (Hg), 46–73,997 $\mu\text{g kg}^{-1}$; arsenic (As), 1.1–367 mg kg^{-1} ; chromium (Cr), 20–1247 mg kg^{-1} ; copper (Cu), 4.3–191 mg kg^{-1} ; nickel (Ni), 6–667 mg kg^{-1} ; lead (Pb), 5.2–216 mg kg^{-1} ; and zinc (Zn), 23.9–1323 mg kg^{-1} . The corresponding mean concentrations (mean $\pm$ standard deviation) were $2491 \pm 2860 \mu\text{g kg}^{-1}$ for Cd, $556 \pm 1016 \mu\text{g kg}^{-1}$ for Hg, $38.6 \pm 27.02 \text{mg kg}^{-1}$ for As, $254 \pm 149 \text{mg kg}^{-1}$ for Cr, $48.25 \pm 19.69 \text{mg kg}^{-1}$ for Cu, $69.8 \pm 39.3 \text{mg kg}^{-1}$ for Ni, $51.4 \pm 23.36 \text{mg kg}^{-1}$ for Pb, and $222 \pm 124 \text{mg kg}^{-1}$ for Zn.

Table 1. Descriptive statistics of heavy metal concentrations.

	Min	Mean	Median	Max	SD	CV	Skewness	Kurtosis
Cd ($\mu\text{g kg}^{-1}$)	26	2491	1603	45,407	2860	1.15	3.39	20.77
Hg ($\mu\text{g kg}^{-1}$)	46	556	479	73,997	1016	1.83	51.28	3374
As (mg kg^{-1})	1.1	38.6	36.1	367	27.02	0.7	2.26	12.66
Cr (mg kg^{-1})	20	254	220	1247	149	0.59	0.86	0.8
Cu (mg kg^{-1})	4.3	48.25	45.6	191	19.69	0.41	1	2.24
Ni (mg kg^{-1})	6	69.8	62.8	667	39.3	0.56	2.59	19.51
Pb (mg kg^{-1})	5.2	51.4	48.5	216	23.36	0.45	0.42	−0.43
Zn (mg kg^{-1})	23.9	222	195	1323	124	0.56	1.24	2.39

Compared with the regional background values for Guangxi Province (Supplementary Table S4), all eight heavy metals exhibited elevated concentrations to varying extents. Specifically, the average levels of Cr, Pb, Cd, As, Hg, Ni, Cu, and Zn were 5.10, 2.30, 17.30, 4.85, 6.71,

4.70, 2.70, and 5.12 times higher than their respective baseline values. These results suggest widespread accumulation of heavy metals in the study area's soils, with pronounced enrichment observed for Cd, Hg, and Cr. Notably, Cd showed the highest enrichment factor (17.30), likely attributable to long-term anthropogenic inputs such as phosphate fertilizers, mining, and soil acidification. Although the enrichment levels of Ni, Cu, and Zn were comparatively moderate, they still exceeded background thresholds, warranting further attention.

The coefficient of variation (CV) serves as a measure of relative dispersion in data distribution. A higher CV indicates greater variability around the mean, whereas a lower CV reflects more uniform concentrations [31]. Among the eight elements studied, the CVs were ranked as follows: Hg (183%) > Cd (115%) > As (70%) > Cr (59%) > Zn (56%) = Ni (56%) > Pb (45%) > Cu (41%).

According to the classification criteria proposed by [32] (Supplementary Table S3), $CV < 10\%$ denotes low variability, $10\% < CV < 90\%$ represents moderate variability, and $CV > 90\%$ indicates high variability. Based on these thresholds, Hg and Cd demonstrated high spatial heterogeneity, suggesting strong influences from point sources or environmental fluctuation. In contrast, Cr, Cu, Zn, and other elements fell within the moderate variability range, implying relatively stable spatial distributions.

3.2. Performance Comparison of Predictive Models

Multiple machine learning models were assessed to determine the optimal approach for heavy metal concentration prediction, with 22 environmental parameters serving as input variables. Model accuracy was evaluated using cross-validation metrics, including R^2 and root mean square error (RMSE) values. Comprehensive model performance metrics are presented in Figure 2.

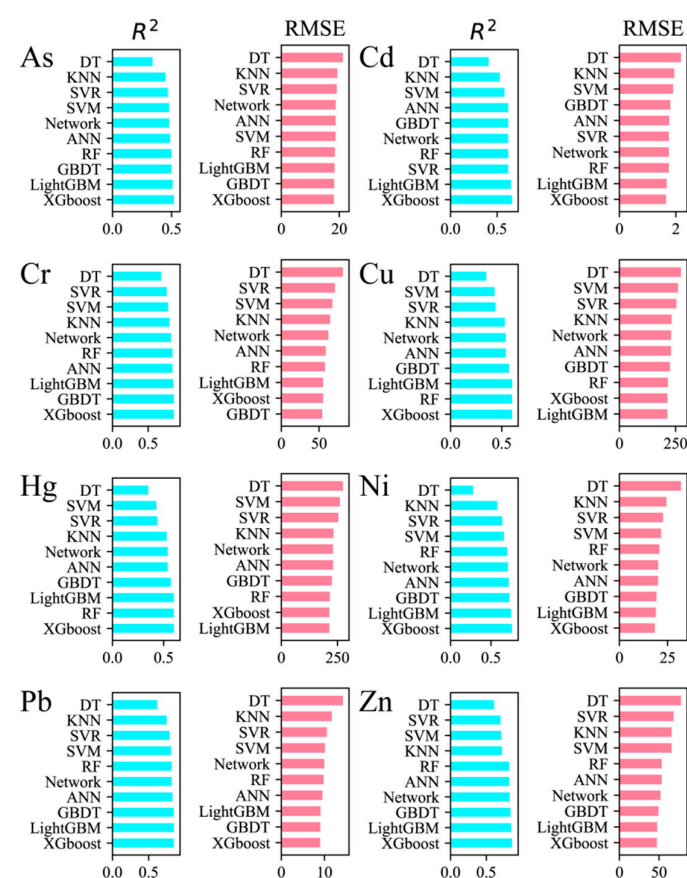


Figure 2. The prediction performance (R^2 and RMSE) of ten machine learning models for eight heavy metals.

The predictive performance of decision tree (DT)-based models for heavy metal concentrations was suboptimal. Gradient boosting ensemble models (e.g., XGBoost and LightGBM) outperformed other algorithms owing to their distinctive optimization architecture and computational design, demonstrating particular advantages in modeling complex nonlinear relationships [33]. XGBoost employs an iterative weak learner training process to progressively reduce prediction residuals. During each iteration, the algorithm specifically targets mispredicted samples (negative gradient direction) from preceding models to systematically minimize prediction bias. This is in contrast to random forest (RF) models, which utilize parallel bagging of independent trees with averaging (primarily for variance reduction and overfitting prevention). Through sequential boosting iterations, XGBoost concurrently minimizes bias and variance while achieving superior complex pattern recognition. Unlike GBDT, which is susceptible to overfitting due to absence of regularization, XGBoost incorporates regularization to substantially enhance model generalizability. Moreover, XGBoost effectively identifies the complex nonlinear interactions among various determinants (e.g., pH, organic content, and anthropogenic influences) governing soil heavy metal concentrations [34].

XGBoost exhibited superior predictive performance for heavy metal concentrations compared to all other models assessed. For the training dataset, XGBoost yielded coefficient of determination (R^2) values of 0.97 for chromium (Cr), 0.98 for lead (Pb), 0.94 for cadmium (Cd), 0.94 for arsenic (As), 0.95 for copper (Cu), 0.98 for nickel (Ni), 0.98 for zinc (Zn), and 0.94 for mercury (Hg), accompanied by root mean square error (RMSE) values spanning 0.58 to 70.97. In the validation phase, the model maintained robust performance with R^2 values of 0.72 (arsenic), 0.69 (cadmium), 0.89 (chromium), 0.79 (copper), 0.78 (mercury), 0.78 (nickel), 0.86 (lead), and 0.88 (zinc), while RMSE values varied between 1.29 and 129.17. These results confirm the model's consistent predictive accuracy for all target heavy metals, reflecting exceptional training data fitting performance coupled with remarkable generalization capacity on independent validation data (Figure 3).

Comparative analysis between measured and predicted heavy metal concentrations revealed systematic underestimation for cadmium (Cd), arsenic (As), and mercury (Hg) at elevated levels, as evidenced by regression lines persistently deviating below the 1:1 reference line, a finding aligned with previous reports and studies. This underestimation is likely related to the highly skewed distribution of these elements and the relatively limited number of extreme-value samples in the training dataset. As a result, the model may have been biased toward moderate concentration ranges, leading to reduced sensitivity in high-risk zones.

In addition, the absence of data transformation may have further constrained model performance at high concentration levels. Log-transformation of heavy metal concentrations has been reported to effectively reduce distributional skewness and enhance predictive accuracy for extreme values [35]. Furthermore, stratified modeling approaches that separately model high- and low-concentration zones may improve prediction reliability in contaminated areas. These strategies warrant further investigation in future studies.

While the model demonstrated satisfactory predictive accuracy for chromium (Cr) based on RMSE metrics, its performance was notably poorer for cadmium (Cd) estimations. Contrary to the established positive correlation between predictive accuracy and both concentration range width/standard deviation (SD) reported by Jia et al. [36], cadmium, despite exhibiting the broadest concentration span and highest SD in our dataset, paradoxically yielded the least accurate predictions among all analyzed elements. Despite comprehensive inclusion of known environmental determinants of soil pollution, potential unaccounted contamination sources may persist within the model framework. To elucidate residual uncertainties, we implemented a multifaceted approach incorporating

complementary data acquisition strategies. Importantly, the model demonstrated robust overall performance metrics, successfully fulfilling all predefined research objectives with statistically validated accuracy and reliability.

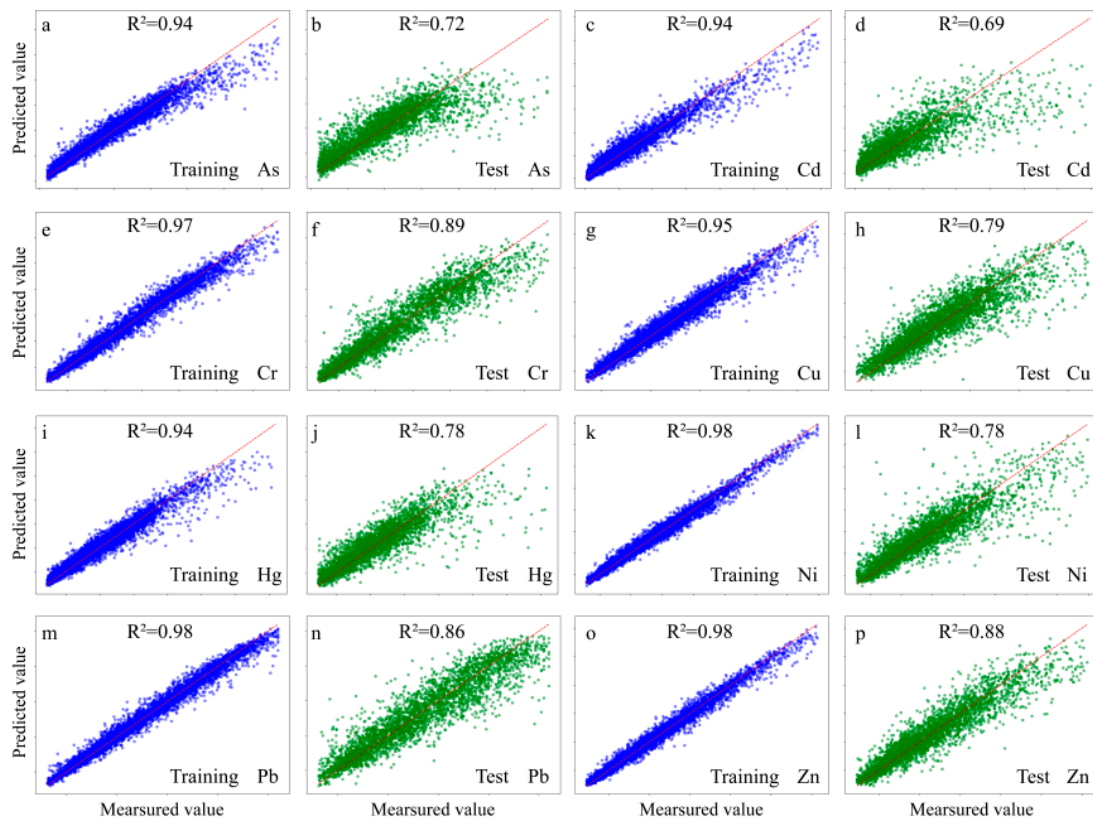


Figure 3. Regression results comparison for training and test datasets across different heavy metals. (a,c,e,g,i,k,m,o) represent the training sets of As, Cd, Cr, Cu, Hg, Ni, Pb, and Zn, respectively. (b,d,f,h,j,l,n,p) denote the prediction sets of As, Cd, Cr, Cu, Hg, Ni, Pb, and Zn, respectively.

3.3. Feature Importance

Figure 4 displays the primary environmental factors influencing soil concentrations of arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), mercury (Hg), nickel (Ni), lead (Pb), and zinc (Zn), along with their relative contributions. The results indicate that iron oxides (Fe_2O_3) exhibit the highest explanatory power for Cr (44.8%) and Ni (26.8%) variations, whereas aluminum oxides (Al_2O_3) are the dominant factors affecting Hg (30.2%) and Pb (36.4%). Phosphorus content (P) significantly influences the distribution of Cd (21.9%), Cu (18.2%), and Zn (28.4%). The digital elevation model (DEM) demonstrates a consistent explanatory capacity (8.6%) for Cr spatial heterogeneity. Additionally, magnesium oxide (MgO, 13.9% for Ni), silicon oxide (SiO_2 , 9.5% for Zn), and calcium oxide (CaO, 6.4% for Cu) also play notable roles. These findings suggest that soil mineral composition (e.g., Fe_2O_3 , Al_2O_3), nutrient elements (e.g., P), and topographic factors (DEM) collectively govern the spatial distribution of heavy metals in the study area.

This study reveals the synergistic control of multiple geochemical factors on heavy metal distribution in soils. The strong association between Fe_2O_3 and Cr/Ni (44.8% and 26.8%, respectively) likely stems from the high surface area and adsorption capacity of iron-bearing minerals (e.g., hematite), consistent with known metal immobilization mechanisms [37]. The pronounced influence of Al_2O_3 on Hg (30.2%) and Pb (36.4%) may arise from specific adsorption by aluminum octahedral structures in clay minerals [38]. The correlation between P and Cd/Cu/Zn (21.9–28.4%) implies regulatory effects from phosphate precipitation and organic phosphorus mineralization on metal bioavailability [39]. Notably,

DEM's explanatory power for Cr spatial variation (8.6%) highlights topography-driven erosion–deposition processes in redistributing heavy metals, particularly in transitional slope zones [40]. Secondary contributions from MgO (13.9% for Ni), SiO₂ (9.5% for Zn), and CaO (6.4% for Cu) further underscore the role of mineral weathering and pedogenesis in controlling metal geochemical behavior. These findings emphasize the necessity of integrating mineral composition (Fe₂O₃/Al₂O₃), nutrient cycling (P), and topographic features (DEM) in a three-dimensional framework for comprehensive environmental risk assessment of heavy metals.

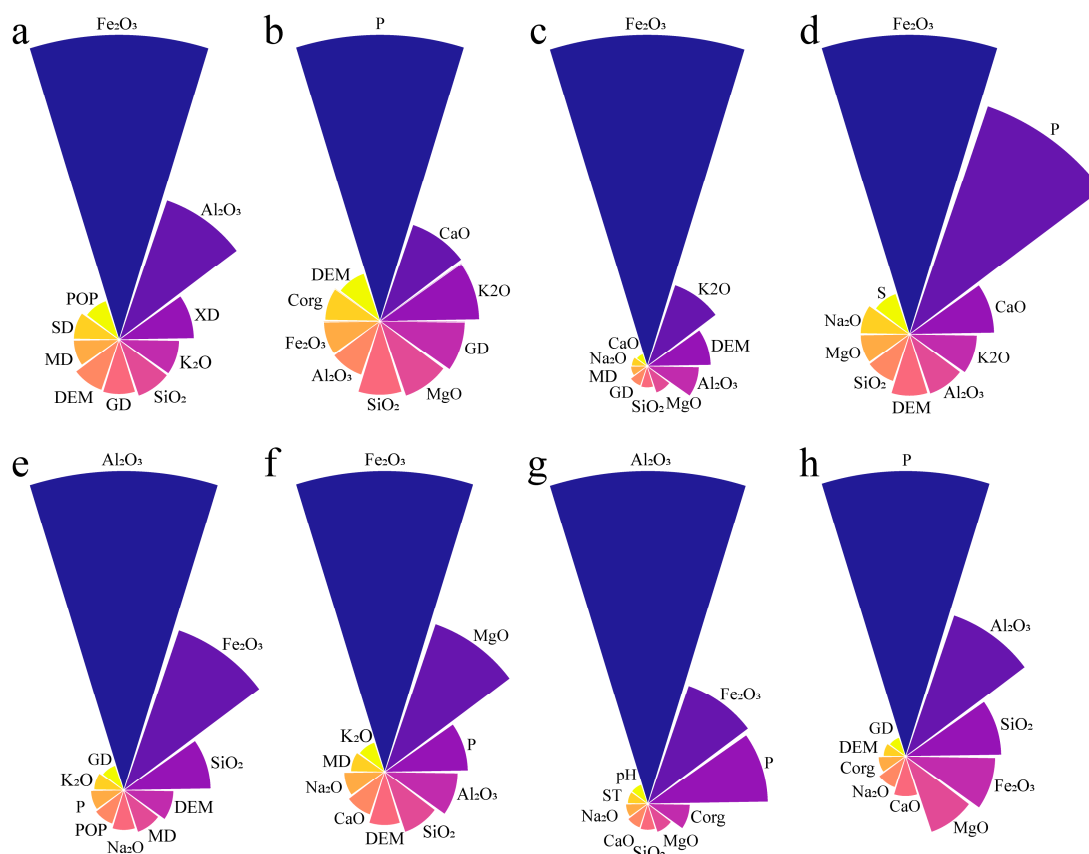


Figure 4. The rose diagram displays contributions of the top 10 factors to As (a), Cd (b), Cr (c), Cu (d), Hg (e), Ni (f), Pb (g), and Zn (h) concentrations in soil by XGBoost.

3.4. Ecological Risk Assessment

Figure 5 presents the calculated single-factor potential ecological risk indices (Er) and comprehensive potential ecological risk index (RI) for the eight heavy metals in agricultural soils. The Er values revealed the following risk hierarchy: Cd > Hg > Pb > As > Ni > Cu > Cr (note: removed duplicate Pb for logical sequence). Of particular concern, cadmium (Cd) and mercury (Hg) demonstrated exceptionally high ecological risks. Spatial analysis identified significant heterogeneity in Cd and Hg distributions: For Cd, most areas exhibited Er values < 40 (indicating low risk), while extreme-risk zones (Er > 320) formed distinct belts in northern and southeastern regions. Hg's extreme-risk concentrations (Er > 160) were predominantly clustered in western sectors. Lead (Pb) showed generally low-risk levels (Er < 40) across most cultivated soils, with elevated values localized in western and southeastern zones. Arsenic (As) maintained a low-risk status (Er < 40) at most sampling sites, though notable moderate-to-high risk pockets emerged in northwestern and central areas. Nickel (Ni) displayed the lowest overall risk, with sporadic low-risk spots (Er 20–40) distributed in southwestern and southern locales. Both Pb and copper (Cu) maintained uniformly low-risk patterns (Er < 20) without spatial clustering. Chromium

(Cr) registered minimal risk indices, with slightly elevated readings concentrated in central and northern sectors.

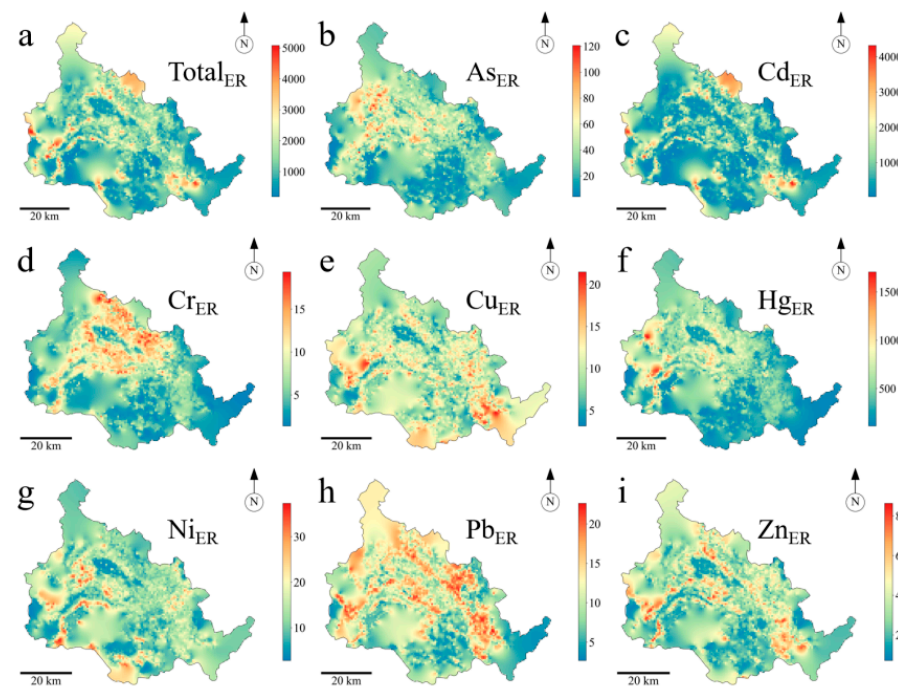


Figure 5. Comprehensive ecological risk (ER) assessment (a) and individual metal-specific potential ER indices (b–i) for eight heavy metals detected in the soils of the investigated region.

The comprehensive potential ecological risk index (RI) across the study area exhibited significantly elevated levels, with maximum values exceeding 5000. Spatial analysis revealed three critical high-risk zones requiring immediate attention: (1) northern regions, (2) southeastern sectors, and (3) southwestern areas, all demonstrating extreme ecological risk levels ($RI > 600$).

Agricultural soils faced predominant ecological threats from cadmium (Cd) and mercury (Hg) contamination, with these two elements collectively contributing $> 75\%$ of the total ecological risk. This alarming situation demands prompt implementation of coordinated mitigation strategies by agricultural and environmental protection authorities, focusing particularly on: Source control of Cd/Hg emissions, soil remediation in identified hotspots, and long-term monitoring programs.

Figure S2 shows a schematic diagram of the overall soil ecological risk overlaying cropland. It can be observed that nearly all high-risk ecological zones overlap with agricultural land, particularly in the northwest, western, and southeastern regions, where the overlap with cropland is more pronounced. Additionally, there is agricultural land surrounding the large high-risk area in the central northern region. However, the distribution of cropland in the Jingxi region is relatively scattered, which may be unfavorable for managing heavy metal pollution.

3.5. Spatial Associations Between Ecological Risk Levels and Environmental Drivers

Figure 6 clearly demonstrates the spatial correlations between the integrated ecological risk index (RI) and natural environmental factors. High risk–high factor (H–H) clustering zones (i.e., areas with simultaneously elevated ecological risk and environmental factor values) exhibit pronounced spatial aggregation characteristics [30]. In detail, the spatial patterns of C_{org} , N, and S show similarities with ecological risk (predominantly L–H clustering zones), where many regions exhibit low intrinsic heavy metal pollution risk (Er)

but are adjacent to areas with elevated C_{org} or N levels, indicating localized spatial negative correlations. This phenomenon may arise because high C_{org} /N/S zones could reduce heavy metal bioavailability through adsorption, complexation, or microbial immobilization, thereby lowering pollution risks in neighboring L-H regions [41]. This suggests that organic matter- or nitrogen-rich environments may act as barriers to heavy metal diffusion. For example, in the concept proposed by Kwiatkowska-Malina [41], organic matter in soil can act as a barrier to reduce the uptake of heavy metals by plants or their migration into groundwater, which is consistent with our research findings. That is, organic matter in soil can reduce the bioavailability of heavy metals by binding (complexing) them.

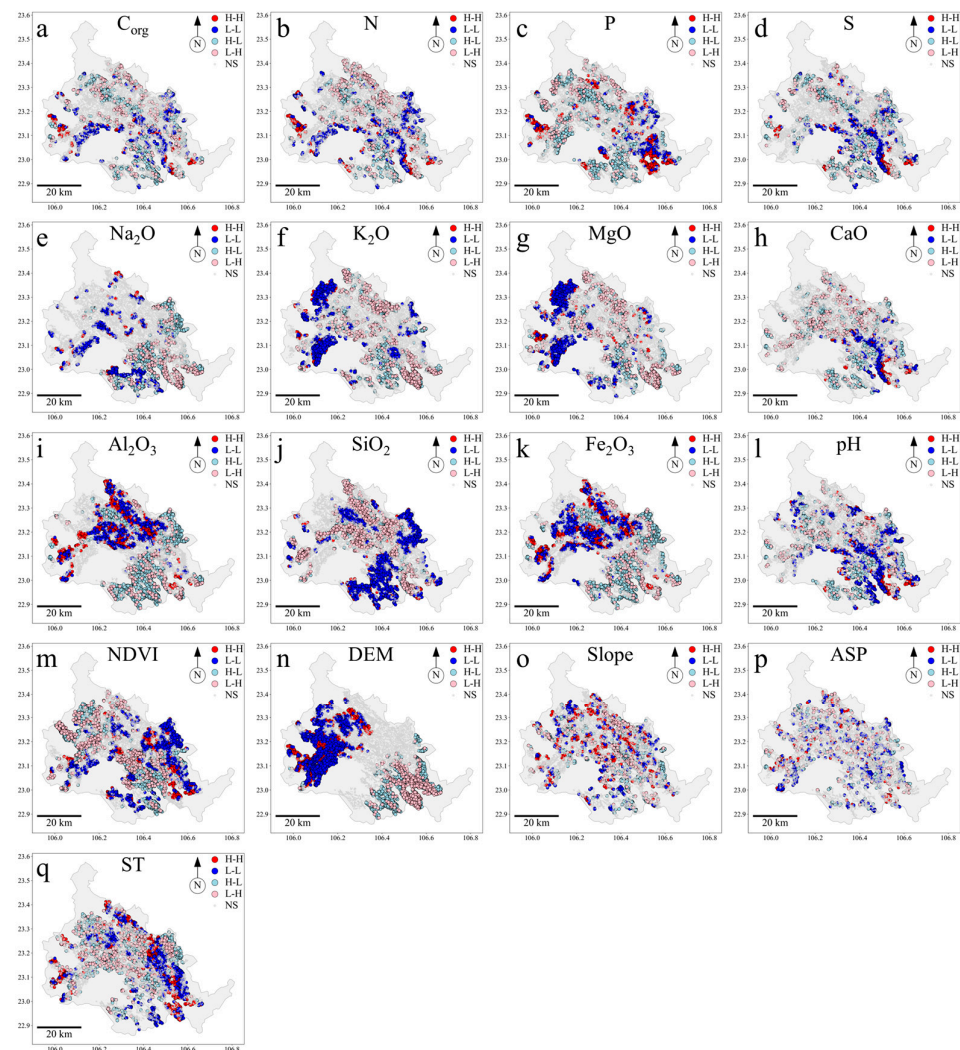


Figure 6. Spatial clustering map based on the bivariate local Moran's index between the comprehensive ecological risk index and natural factors. The four quadrants represent different types of local spatial autocorrelation: High–High (H–H) indicates spatial units with high values surrounded by neighbors with high values, suggesting a positive spatial clustering of high values; Low–Low (L–L) represents spatial units with low values surrounded by neighbors with low values, indicating a positive clustering of low values. In contrast, High–Low (H–L) denotes spatial units with high values surrounded by neighbors with low values, while Low–High (L–H) represents spatial units with low values surrounded by neighbors with high values; both H–L and L–H indicate negative spatial autocorrelation and spatial outliers.

Furthermore, the western study area exhibits low K_2O / MgO content alongside low heavy metal risk, likely due to soils derived from K/Mg-poor parent materials (e.g., sandstone, granite), which inherently possess low heavy metal background levels. Notably,

SiO₂ displays significant L-H clustering, potentially attributed to industrial activities (e.g., smelters) in Jingxi, where acidic deposition (e.g., SO₂) may dissolve and leach SiO₂ while introducing anthropogenic heavy metals, forming a “low Si–high pollution” combination. Additionally, the co-occurrence of low pH (acidity) and low heavy metal ecological risk in this study may reflect soils derived from acidic parent materials (e.g., granite, sandstone), which naturally exhibit low pH but also inherently low heavy metal background concentrations [42].

Figure 7 demonstrates that anthropogenic factors (road proximity and mining distance) predominantly exhibit low–high (L-H) clustering patterns with the ecological risk index. This spatial association arises because road drainage systems serve as rapid conduits for heavy metal transport, and ore spillage during transportation leads to localized heavy metal enrichment in roadside soils. Similar mechanisms apply to mining-adjacent areas, where operational activities and atmospheric deposition contribute to heavy metal accumulation in surrounding soils.

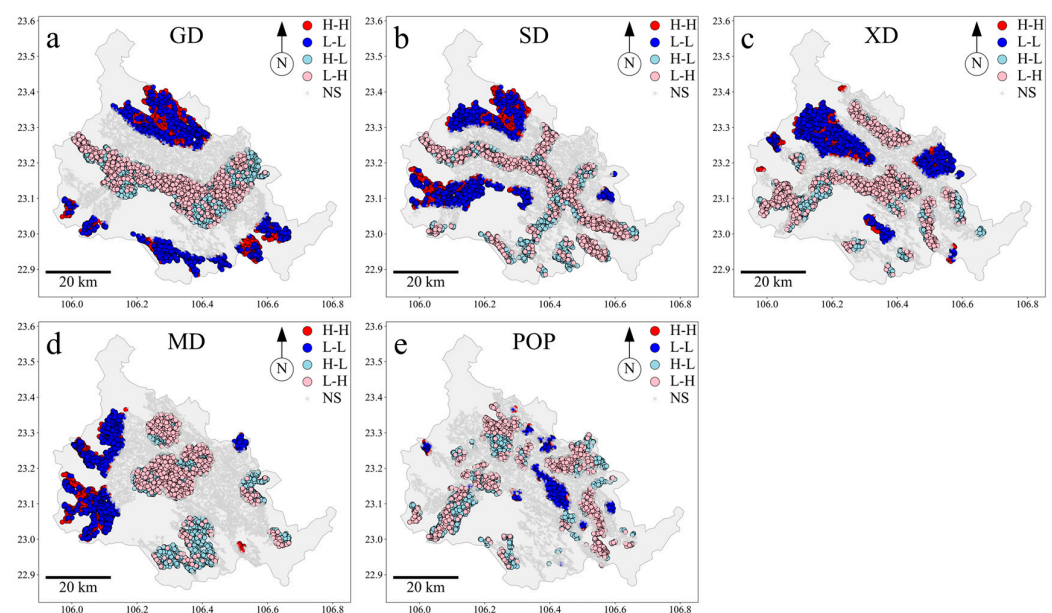


Figure 7. Spatial clustering map based on the bivariate local Moran’s index between the comprehensive ecological risk index and anthropogenic factors. The four quadrants represent different types of local spatial autocorrelation: High–High (H–H) indicates spatial units with high values surrounded by neighbors with high values, suggesting a positive spatial clustering of high values; Low–Low (L–L) represents spatial units with low values surrounded by neighbors with low values, indicating a positive clustering of low values. In contrast, High–Low (H–L) denotes spatial units with high values surrounded by neighbors with low values, while Low–High (L–H) represents spatial units with low values surrounded by neighbors with high values; both H–L and L–H indicate negative spatial autocorrelation and spatial outliers.

3.6. Network Pattern

The network analysis (Figure 8) revealed significant positive correlations among most soil parameters and heavy metals, consistent with known co-accumulation patterns in contaminated systems [43]. Notably, SiO₂ exhibited consistent negative correlations with Fe/Al oxides, sulfur, and multiple heavy metals (Hg, Pb, Zn, Cr, As), likely reflecting geochemical antagonism between silicate minerals and metal(loid)s, where quartz-rich parent materials typically show lower native metal contents [44]. This inverse relationship may also involve competitive sorption effects, as Fe/Al oxides (positively correlated with metals) become less abundant in SiO₂-dominated systems, along with potential pH-mediated effects since SiO₂-rich soils often exhibit higher pH values that reduce metal solubility.

Interestingly, correlations were exclusively observed among quantitative soil elemental indicators, with no detectable relationships for non-quantitative parameters like DEM or soil type, suggesting heavy metal distribution in our study area is primarily governed by geochemical processes rather than topographic or categorical soil characteristics. This implies that SiO_2 content may serve as a useful indicator for lower metal availability areas, and that management efforts should prioritize elemental composition over general soil classifications when assessing metal risks, while recognizing the limited predictive power of terrain-based pollution models in this system. The absence of DEM correlations may reflect either insufficient spatial resolution in our elevation data or the predominance of geochemical over physical transport processes in controlling metal distribution patterns. These findings collectively highlight the importance of mineralogical composition in understanding heavy metal behavior in soils.

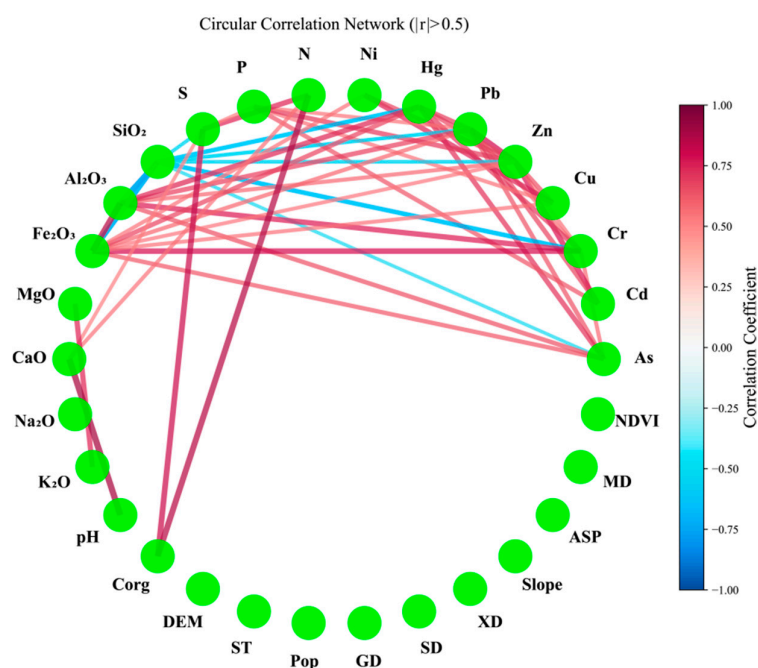


Figure 8. The correlation network diagram between 22 environmental factors and 8 heavy metals, where red lines represent positive correlations and blue lines represent negative correlations.

4. Conclusions

This study systematically evaluated the contamination characteristics and ecological risks of heavy metals in soils from a typical karst region of Guangxi, elucidating multi-scale environmental driving mechanisms. The results demonstrated widespread enrichment of Cd, Hg, and Cr (up to 17.3-fold background levels), with Cd and Hg constituting primary ecological risk sources (combined contribution > 75%). High-risk zones were predominantly clustered in northern, southeastern, and southwestern sectors. The XGBoost model achieved optimal performance in spatial prediction (validation $R^2 = 0.69\text{--}0.89$), with feature importance analysis identifying iron/aluminum oxides (Fe_2O_3 44.8%; Al_2O_3 36.4%) and phosphorus content (P 21.9–28.4%) as key determinants of heavy metal distribution. Spatial association analysis indicated that organic-rich zones (elevated Corg/N/S) formed heavy metal diffusion barriers through adsorption–complexation mechanisms, while typical “low–high” pollution clustering patterns around roads/mining areas reflected anthropogenic inputs. Network analysis revealed significant negative correlations between SiO_2 and heavy metals, confirming the inhibitory effect of silicate minerals on metal bioavailability. We propose a tiered management strategy: implementing source control and soil remediation

in high-risk zones, enhancing organic matter regulation in moderate–low risk areas, and establishing an early warning system based on mineralogical composition ($\text{Fe}_2\text{O}_3/\text{SiO}_2$ ratio). This work provides a theoretical framework integrating natural–anthropogenic dual drivers for heavy metal pollution control in karst regions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/land15020304/s1>, Figure S1: Spatial distribution map of 20 environmental factors: (a) Organic carbon; (b) Nitrogen; (c) Phosphorus; (d) Sulfur; (e) pH; (f) K_2O ; (g) CaO ; (h) Na_2O ; (i) MgO ; (j) Al_2O_3 ; (k) SiO_2 ; (l) Fe_3O_4 ; (m) NDVI; (n) DEM; (o) Slope; (p) Aspect; (q) Mine; (r) Population; (s) Main roads, and (t) Soil type. Figure S2: Soil potential risk index and cropland overlay analysis map; Table S1: Descriptive statistics of soil nutrients and metal oxides (Unit: mg/kg). Table S2: 8 environmental factors data sources. Table S3: Classification of ecological risk levels associated with multi-heavy metal soil contamination. Table S4: Risk-screening value of soil pollutant factors for agricultural land and Guangxi autonomous region's soil background values (Unit: mg/kg); Text S1: Methods for determining soil chemical indicators. Text S2: Machine learning codes used in Python 3.13.

Author Contributions: Conceptualization, Z.L. and J.L. (Jie Li); methodology, J.W.; software, J.W.; validation, G.Z., J.Q. and W.G.; formal analysis, J.L. (Jiacai Li); investigation, W.G.; resources, J.L. (Jie Li); data curation, Z.L.; writing—original draft preparation, Z.L.; writing—review and editing, J.W.; visualization, J.W.; supervision, Z.L.; project administration, J.L. (Jie Li); funding acquisition, J.L. (Jie Li). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Guangxi Key Technologies R&D Program (AB25069152, AB24010136); Preliminary Project of the Guangxi Bureau of Geology & Mineral Prospecting & Exploitation (GXDK202643); Young Elite Scientists Sponsorship Program by GXAST (GXYESS2025085); and Guangxi Geochemistry and Environmental Restoration, Management Research Talent Highland (Guangxi Mining Office (2023) No. 55).

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy restrictions.

Acknowledgments: We gratefully acknowledge the significant contributions of Guohui Yue (Geological Survey of Guangxi) and Bingke Lu (Geological Survey of Guangxi) to the sampling activities.

Conflicts of Interest: Juan Wu is employed by Guangxi Water and Power Design Institute Co., Ltd. Jiakai Li is employed by China Nonferrous Metals Geology and Mining Co., Ltd. The authors declare no conflicts of interest.

References

1. Timothy, N.A.; Williams, E.T. Environmental pollution by heavy metal: An overview. *Int. J. Environ. Chem.* **2019**, *3*, 72–82. [CrossRef]
2. Wan, X.; Yang, J.; Song, W. Pollution status of agricultural land in China: Impact of land use and geographical position. *Soil Water Res.* **2018**, *13*, 234–242. [CrossRef]
3. Zerga, B. Karst topography: Formation, processes, characteristics, landforms, degradation and restoration: A systematic review. *Watershed Ecol. Environ.* **2024**, *6*, 252–269. [CrossRef]
4. De Waele, J.; Gutiérrez, F.; Parise, M.; Plan, L. Geomorphology and natural hazards in karst areas: A review. *Geomorphology* **2011**, *134*, 1–8. [CrossRef]
5. Chen, J.; Yu, J.; Bai, X.; Zeng, Y.; Wang, J. Fragility of karst ecosystem and environment: Long-term evidence from lake sediments. *Agric. Ecosyst. Environ.* **2020**, *294*, 106862. [CrossRef]
6. Wang, Y.F.; Liang, L.S.; Jing, J.L.; Luo, F.L.; Wang, A.N. Characteristic of drought and flood in the Dian-Qian-Gui karst areas based on TRMM-Z Index. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *42*, 917–924. [CrossRef]
7. Micó, C.; Recatalá, L.; Peris, M.; Sánchez, J. Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis. *Chemosphere* **2006**, *65*, 863–872. [CrossRef] [PubMed]
8. Li, J.; He, M.; Han, W.; Gu, Y. Analysis and assessment on heavy metal sources in the coastal soils developed from alluvial deposits using multivariate statistical methods. *J. Hazard. Mater.* **2009**, *164*, 976–981. [CrossRef] [PubMed]

9. Fu, S.; Wei, C.Y. Multivariate and spatial analysis of heavy metal sources and variations in a large old antimony mine, China. *J. Soils Sediments* **2013**, *13*, 106–116. [\[CrossRef\]](#)
10. Wang, Y.; Li, Y.; Yang, S.; Liu, J.; Zheng, W.; Xu, J.; Liu, X. Source apportionment of soil heavy metals: A new quantitative framework coupling receptor model and stable isotopic ratios. *Environ. Pollut.* **2022**, *314*, 120291. [\[CrossRef\]](#)
11. Ha, H.; Olson, J.R.; Bian, L.; Rogerson, P.A. Analysis of heavy metal sources in soil using kriging interpolation on principal components. *Environ. Sci. Technol.* **2014**, *48*, 4999–5007. [\[CrossRef\]](#)
12. Huang, Y.; Deng, M.; Wu, S.; Japenga, J.; Li, T.; Yang, X.; He, Z. A modified receptor model for source apportionment of heavy metal pollution in soil. *J. Hazard. Mater.* **2018**, *354*, 161–169. [\[CrossRef\]](#)
13. Bonten, L.T.; Groenenberg, J.E.; Weng, L.; van Riemsdijk, W.H. Use of speciation and complexation models to estimate heavy metal sorption in soils. *Geoderma* **2008**, *146*, 303–310. [\[CrossRef\]](#)
14. Badenko, V.; Kurtener, D.; Krueger, E. Utilization of fuzzy set theory for interpretation of data of investigations of soil contamination by heavy metals. *Eur. Agrophysical J.* **2014**, *1*, 25–41. [\[CrossRef\]](#)
15. Yang, S.; Taylor, D.; Yang, D.; He, M.; Liu, X.; Xu, J. A synthesis framework using machine learning and spatial bivariate analysis to identify drivers and hotspots of heavy metal pollution of agricultural soils. *Environ. Pollut.* **2021**, *287*, 117611. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Wang, Y.; Zhao, Y.; Xu, S. Application of VNIR and machine learning technologies to predict heavy metals in soil and pollution indices in mining areas. *J. Soils Sediments* **2022**, *22*, 2777–2791. [\[CrossRef\]](#)
17. Zhang, H.; Yin, A.; Yang, X.; Fan, M.; Shao, S.; Wu, J.; Gao, C. Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils. *Ecol. Indic.* **2021**, *122*, 107233. [\[CrossRef\]](#)
18. Wang, H.; Yilihamu, Q.; Yuan, M.; Bai, H.; Xu, H.; Wu, J. Prediction models of soil heavy metal(loid)s concentration for agricultural land in Dongli: A comparison of regression and random forest. *Ecol. Indic.* **2020**, *119*, 106801. [\[CrossRef\]](#)
19. Zheng, J.; Wang, P.; Shi, H.; Zhuang, C.; Deng, Y.; Yang, X.; Xiao, R. Quantitative source apportionment and driver identification of soil heavy metals using advanced machine learning techniques. *Sci. Total Environ.* **2023**, *873*, 162371. [\[CrossRef\]](#)
20. Li, X.; Gu, H.; Tang, R.; Zou, B.; Liu, X.; Ou, H.; Wen, B. A Fusion XGBoost Approach for Large-Scale Monitoring of Soil Heavy Metal in Farmland Using Hyperspectral Imagery. *Agronomy* **2025**, *15*, 676. [\[CrossRef\]](#)
21. Zhang, Q.; Wang, C. Natural and human factors affect the distribution of soil heavy metal pollution: A review. *Water Air Soil Pollut.* **2020**, *231*, 350. [\[CrossRef\]](#)
22. Cui, X.; Zhang, J.; Wang, X.; Pan, M.; Lin, Q.; Khan, K.Y.; Chen, G. A review on the thermal treatment of heavy metal hyperaccumulator: Fates of heavy metals and generation of products. *J. Hazard. Mater.* **2021**, *405*, 123832. [\[CrossRef\]](#)
23. Liu, P.; Wu, Q.; Hu, W.; Tian, K.; Huang, B.; Zhao, Y. Effects of atmospheric deposition on heavy metals accumulation in agricultural soils: Evidence from field monitoring and Pb isotope analysis. *Environ. Pollut.* **2023**, *330*, 121740. [\[CrossRef\]](#)
24. Zheng, F.; Guo, X.; Tang, M.; Zhu, D.; Wang, H.; Yang, X.; Chen, B. Variation in pollution status, sources, and risks of soil heavy metals in regions with different levels of urbanization. *Sci. Total Environ.* **2023**, *866*, 161355. [\[CrossRef\]](#)
25. Dragović, S.; Mihailović, N.; Gajić, B. Heavy metals in soils: Distribution, relationship with soil characteristics and radionuclides and multivariate assessment of contamination sources. *Chemosphere* **2008**, *72*, 491–495. [\[CrossRef\]](#)
26. Lu, A.X.; Wang, J.H.; Pan, L.G.; Han, P.; Han, Y. Determination of Cr, Cu, Zn, Pb and As in soil by field portable X-ray fluorescence spectrometry. *Spectrosc. Spectr. Anal.* **2010**, *30*, 2848–2852.
27. Hakanson, L. An ecological risk index for aquatic pollution control. a sedimentological approach. *Water Res.* **1980**, *14*, 975–1001. [\[CrossRef\]](#)
28. Zhang, H.; Cai, A.; Wang, X.; Wang, L.; Wang, Q.; Wu, X.; Ma, Y. Risk assessment and source apportionment of heavy metals in soils from Handan City. *Appl. Sci.* **2021**, *11*, 9615. [\[CrossRef\]](#)
29. China Environmental Monitoring Centre. *Values of Soil Elements in China*; China Environmental Science Press: Beijing, China, 1990.
30. Huang, G.; Wang, X.; Chen, D.; Wang, Y.; Zhu, S.; Zhang, T.; Liao, L.; Tian, Z.; Wei, N. A hybrid data-driven framework for diagnosing contributing factors for soil heavy metal contaminations using machine learning and spatial clustering analysis. *J. Hazard. Mater.* **2022**, *437*, 129324. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Hu, B.; Shao, S.; Fu, Z.; Li, Y.; Ni, H.; Chen, S.; Zhou, Y.; Jin, B.; Shi, Z. Identifying heavy metal pollution hot spots in soil-rice systems: A case study in South of Yangtze River Delta, China. *Sci. Total Environ.* **2019**, *658*, 614–625. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Zhang, X.Y.; Sui, Y.Y.; Zhang, X.D.; Meng, K.; Herbert, S.J. Spatial variability of nutrient properties in black soil of Northeast China. *Pedosphere* **2007**, *17*, 19–29. [\[CrossRef\]](#)
33. Niazkar, M.; Menapace, A.; Brentan, B.; Piraei, R.; Jimenez, D.; Dhawan, P.; Righetti, M. Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023). *Environ. Model. Softw.* **2024**, *174*, 105971. [\[CrossRef\]](#)
34. Ye, M.; Zhu, L.; Li, X.; Ke, Y.; Huang, Y.; Chen, B.; Feng, H. Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data. *Sci. Total Environ.* **2023**, *858*, 159798. [\[CrossRef\]](#)

35. Li, S.; Zhang, Y. An integrated model of log-normal ordinary Kriging interpolation-based source-specific human health risk assessment (LSR) for soil heavy metal pollution: Insights from an abandoned industrial area in China. *Environ. Monit. Assess.* **2025**, *197*, 1204. [[CrossRef](#)]
36. Jia, X.; Fu, T.; Hu, B.; Shi, Z.; Zhou, L.; Zhu, Y. Identification of the potential risk areas for soil heavy metal pollution based on the source-sink theory. *J. Hazard. Mater.* **2020**, *393*, 122424. [[CrossRef](#)]
37. Shi, T.; Zhang, J.; Shen, W.; Wang, J.; Li, X. Machine learning can identify the sources of heavy metals in agricultural soil: A case study in northern Guangdong Province, China. *Ecotoxicol. Environ. Saf.* **2022**, *245*, 114107. [[CrossRef](#)] [[PubMed](#)]
38. Fest, E.P.M.J.; Temminghoff, E.J.M.; Comans, R.N.J.; van Riemsdijk, W.H. Partitioning of organic matter and heavy metals in a sandy soil: Effects of extracting solution, solid to liquid ratio and pH. *Geoderma* **2008**, *146*, 66–74. [[CrossRef](#)]
39. Ding, Q.; Cheng, G.; Wang, Y.; Zhuang, D. Effects of natural factors on the spatial distribution of heavy metals in soils surrounding mining regions. *Sci. Total Environ.* **2017**, *578*, 577–585. [[CrossRef](#)] [[PubMed](#)]
40. He, Z.L.; Yang, X.E.; Stoffella, P.J. Trace elements in agroecosystems and impacts on the environment. *J. Trace Elem. Med. Biol.* **2005**, *19*, 125–140. [[CrossRef](#)]
41. Kwiatkowska-Malina, J. Functions of organic matter in polluted soils: The effect of organic amendments on phytoavailability of heavy metals. *Appl. Soil Ecol.* **2018**, *123*, 542–545. [[CrossRef](#)]
42. Zhai, M.; Kampunzu, H.A.B.; Modisi, M.P.; Totolo, O. Distribution of heavy metals in Gaborone urban soils (Botswana) and its relationship to soil pollution and bedrock composition. *Environ. Geol.* **2003**, *45*, 171–180. [[CrossRef](#)]
43. Andra, S.S.; Makris, K.C.; Charisiadis, P.; Costa, C.N. Co-occurrence profiles of trace elements in potable water systems: A case study. *Environ. Monit. Assess.* **2014**, *186*, 7307–7320. [[CrossRef](#)] [[PubMed](#)]
44. Shah, S.A.; Shao, Y.; Zhang, Y.; Zhao, H.; Zhao, L. Texture and trace element geochemistry of quartz: A review. *Minerals* **2022**, *12*, 1042. [[CrossRef](#)]

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.