

Article

Machine Learning-Assisted Biomonitoring of Heavy Metal Accumulation in *Pinus nigra* Needles Across Urban, Industrial, and Pristine Sites in Adiyaman, Türkiye

Turgay Dere ^{1,*}, Sebghatullah Jueyendah ² and Zeynep Yaman ^{2,3} ¹ Department of Environmental Engineering, Adiyaman University, Adiyaman 02040, Turkey² Department of Civil Engineering, Sakarya University, Sakarya 54050, Turkey; sebghatullah.jueyendah@ogr.sakarya.edu.tr or sebghatullahjueyendah@gmail.com (S.J.); zdyaman@sakarya.edu.tr (Z.Y.)³ Faculty of Engineering and Architecture, International Vision University, 1230 Gostivar, North Macedonia

* Correspondence: tdere@adiyaman.edu.tr

Abstract

Heavy metals are persistent environmental contaminants that accumulate in soils and vegetation, posing significant risks to ecological systems and human health. *Pinus nigra* needles are widely recognized as effective biomonitors for reflecting spatial and temporal variations in atmospheric heavy metal deposition. However, the complex, nonlinear interactions among multiple pollutants, environmental factors, and site-specific conditions limit the effectiveness of conventional statistical approaches in accurately modeling and predicting contamination patterns. This study investigated the spatial and seasonal distribution of heavy metals in soils and *Pinus nigra* needles across different environmental settings in Adiyaman, Türkiye, including urban traffic zones, an organized industrial area, a cement factory vicinity, and a clean reference site. Metal concentrations were determined using inductively coupled plasma mass spectrometry (ICP–MS) following standardized acid digestion procedures. To address the limitations of traditional methods and capture complex nonlinear relationships, advanced machine learning (ML) algorithms—multilayer perceptron, Random Forest, XGBoost, LightGBM, CatBoost, and Gradient Boosting—were employed to model elevation based on heavy metal concentrations. The dataset was divided into training (80%) and testing (20%) subsets, and model performance was evaluated using R^2 , RMSE, MAE, MAPE, and EVS. Among the models, XGBoost exhibited superior predictive performance. Excluding Cd, Cr, and Cu, it achieved $R^2 = 0.9996$ (RMSE = 0.068) in training and $R^2 = 0.9526$ (RMSE = 17.77) in testing. Including these metals further improved performance to $R^2 = 0.9999$ (RMSE = 0.054) for training and $R^2 = 0.9890$ (RMSE = 5.55) for testing. The results confirm that *Pinus nigra* needles are reliable bioindicators of heavy metal accumulation. More importantly, the integration of biomonitoring data with ML techniques provides a powerful framework for capturing complex environmental interactions and improving predictive accuracy, thereby supporting more effective environmental monitoring, risk assessment, and sustainable management strategies.



Academic Editor: Guining Lu

Received: 13 March 2026

Revised: 6 April 2026

Accepted: 13 April 2026

Published: 21 July 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 \(CC BY\)](#) license.

Keywords: *Pinus nigra*; black pine; bioindicator species; trace element contamination; inductively coupled plasma mass spectrometry; ensemble learning; spatial distribution; urban–industrial pollution; soil–plant interaction

1. Introduction

Heavy metals are persistent environmental pollutants with well-documented toxicological effects, capable of accumulating in soils, entering the food chain, and posing significant risks to human and ecological health. Among these, lead (Pb) and cadmium (Cd) are of particular concern due to their non-biodegradability and long-term health impacts, including neurotoxicity and disruption of hematopoietic systems [1]. Essential trace elements such as copper, iron, and zinc are vital for biological processes; however, both deficiency and excess can lead to adverse health effects, emphasizing the need for accurate environmental monitoring [2]. Anthropogenic activities, including vehicular emissions, industrial operations, and urban heating, significantly contribute to heavy metal accumulation in soils, vegetation, and the atmosphere. Vegetation-based biomonitoring has emerged as a reliable and cost-effective method for assessing environmental contamination. Tree components such as leaves, bark, and needles have been widely used to reflect spatial and temporal variations in heavy metal deposition [3–27]. Among these, *Pinus nigra* needles are particularly suitable due to their morphological characteristics, longevity, and ability to accumulate airborne pollutants over time. Despite extensive studies on biomonitoring, several important challenges remain. Heavy metal accumulation is governed by complex interactions among multiple environmental variables, including emission sources, meteorological conditions, topography, and physicochemical processes. These interactions are often nonlinear and highly interdependent, making them difficult to accurately capture using conventional statistical methods, which typically rely on linear assumptions and limited variable interactions. As a result, traditional approaches may fail to provide reliable predictive models for environmental contamination and risk assessment. In addition, although some studies have investigated heavy metal pollution in various regions of Turkey, research in Adıyaman remains limited, particularly in terms of integrated analysis combining biomonitoring and advanced predictive modeling. The environmental complexity of this region, influenced by urban traffic, industrial activities, and localized emission sources, further necessitates more robust analytical approaches capable of handling nonlinear and multivariate relationships. In this study, machine learning (ML) techniques offer a significant advantage by enabling the modeling of complex, nonlinear relationships without predefined assumptions [28,29]. ML algorithms efficiently process high-dimensional datasets, identify hidden patterns, and improve predictive accuracy, making them well-suited for environmental systems characterized by high variability and uncertainty [30–34]. Recent advances in ML have enabled the development of predictive models capable of capturing complex nonlinear relationships among pollutants and environmental variables, thereby enhancing the accuracy of contamination assessments and risk evaluations [35–47]. Therefore, this study aims to address the limitations of conventional approaches by integrating *Pinus nigra* biomonitoring with advanced ML algorithms to analyze and predict heavy metal distribution in Adıyaman. Elevation was considered in this study as a surrogate indicator of spatial environmental gradients, as it governs atmospheric transport, pollutant dispersion, and deposition dynamics, leading to systematic variation in pollutant concentrations across topographic gradients [48,49]. Accordingly, modeling elevation based on heavy metal concentrations enables the assessment of whether machine learning algorithms can effectively capture the underlying spatial variability embedded within biomonitoring data. The primary aim of this study is to evaluate heavy metal distribution in Adıyaman using *Pinus nigra* as a bioindicator and to enhance predictive accuracy through ML modeling. The specific objectives are to:

1. Quantify heavy metal concentrations in soils and plant samples across urban, industrial, and reference environments.
2. Assess the effectiveness of *Pinus nigra* needles as bioindicators of spatial and temporal variations in heavy metal accumulation.
3. Apply ML algorithms to capture complex nonlinear relationships among environmental variables and improve predictive modeling for environmental monitoring and risk assessment.

2. Materials and Methods

The study was conducted in Adiyaman, Turkey, at sites representing urban traffic, industrial, and clean-air conditions, including a small industrial zone, university and boulevard vicinities, the organized industrial zone (OIZ), the cement factory area, and the Seyirtepe recreational area (Figure 1). Two environmental matrices were sampled seasonally: *Pinus nigra* needles and co-located surface soils. Samples were dried, digested, and analyzed for arsenic (As), lead (Pb), copper (Cu), chromium (Cr), nickel (Ni), cadmium (Cd), zinc (Zn), iron (Fe), and manganese (Mn) using ICP-MS. Quality control procedures included analytical blanks, sample duplicates, and certified reference materials to ensure measurement accuracy and reproducibility. Heavy metal concentrations served as input variables, while elevation was selected as the target variable because it acts as a proxy for spatial deposition gradients, influencing metal accumulation across diverse environmental conditions. Surface soil samples (0–10 cm depth) were collected at each site in close proximity to the sampled *Pinus nigra* trees to ensure spatial consistency. Sampling locations were selected to represent different environmental conditions while avoiding disturbed areas, and composite samples were prepared from multiple subsamples to ensure representativeness. The dataset comprised 140 observations across four seasons [50]. Data preprocessing included outlier removal via the isolation forest method (5% of observations) and normalization using the standard scaler. Two analytical scenarios were evaluated: one excluding Cd, Cu, and Cr, and another including all metals. The dataset was partitioned into training (80%) and testing (20%) subsets to enhance model robustness and mitigate overfitting. Hyperparameter optimization for multilayer perceptron (MLP), CatBoost, random forest, XGBoost, LightGBM, and gradient boosting models was performed using both grid search and Bayesian optimization techniques [51]. Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and explained variance score (EVS), and predictions were validated against experimental measurements to confirm the models' ability to capture complex, nonlinear relationships among heavy metal concentrations in soils and *Pinus nigra* needles [52]. Table 1 provides a detailed overview of the input and output parameters, including their statistical properties.

Table 1. Descriptive statistics of heavy metal concentrations in *Pinus nigra* needles and elevation.

Abbreviation	Units	Mean	StDev	Minimum	Median	Maximum
Cr	mg/kg	3.914	3.825	0.358	2.322	19.277
Cu	mg/kg	5.771	2.467	2.809	5.159	16.218
Pb	mg/kg	0.539	0.421	0.000	0.494	1.504
Zn	mg/kg	24.073	8.688	9.705	22.422	58.228
As	mg/kg	0.281	0.213	0.091	0.227	1.391
Ni	mg/kg	5.853	3.397	1.860	4.601	16.867
Cd	mg/kg	0.053	0.112	0.000	0.000	0.558
Fe	mg/kg	67.079	29.901	16.981	61.976	169.015
Al	mg/kg	1690.745	393.709	879.859	1714.003	2418.241
Mn	mg/kg	110.164	58.601	31.798	99.261	367.002
Elevation	m	721.898	142.127	505.437	696.637	930.873

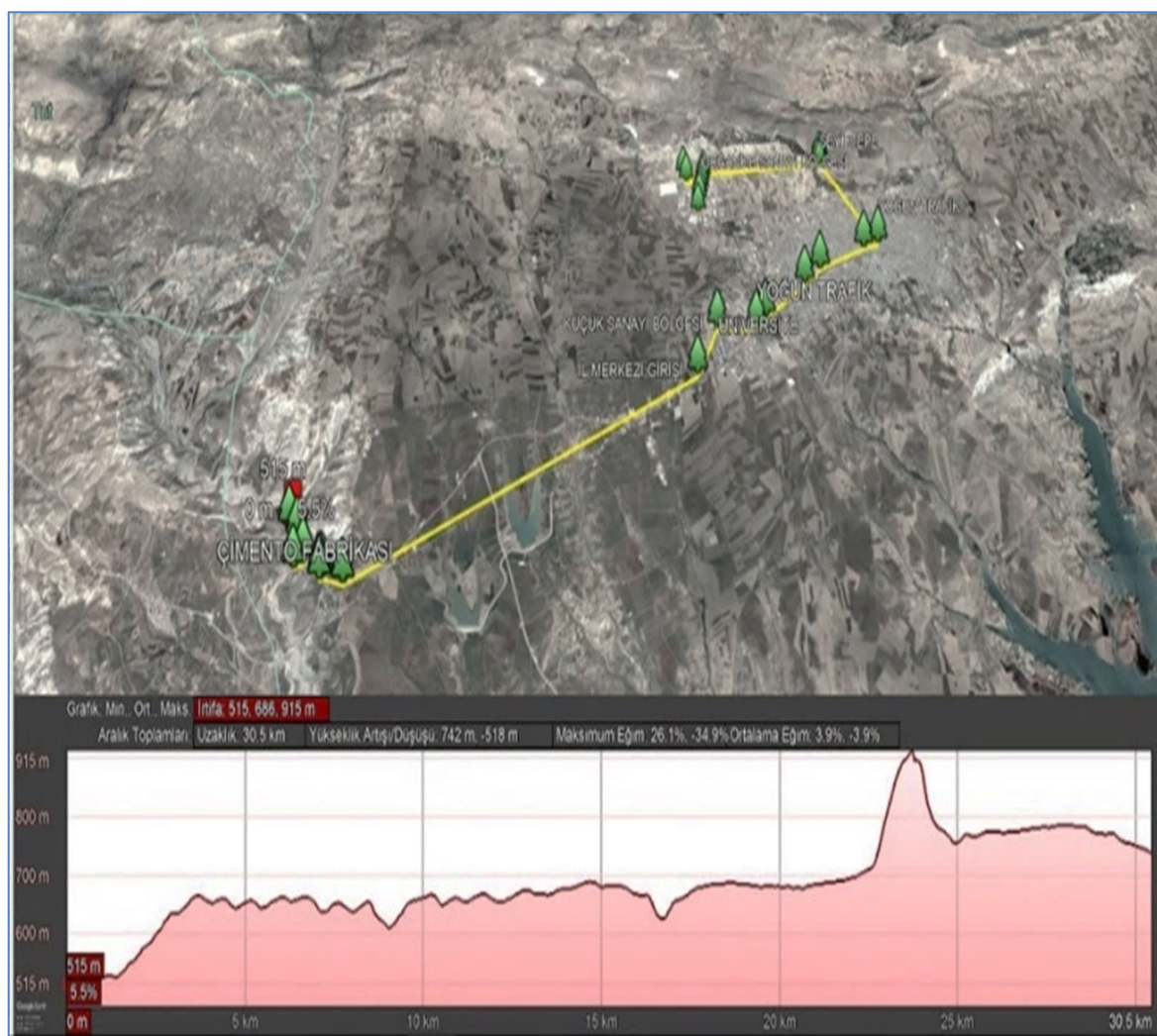


Figure 1. Views of sampling points.

Table 1 summarizes descriptive statistics of heavy metal concentrations and elevation across the study sites. Aluminum exhibited the highest mean (1690.745 mg/kg) and cadmium the lowest (0.053 mg/kg), with Fe, Mn, Cr, and Ni showing notable variability and localized accumulation, while elevation ranged from 505.437 to 930.873 m, providing a basis for data preprocessing, feature evaluation, and modeling of heavy metal distribution. Figure 1 shows the sampling sites across Adıyaman, including the city center (traffic corridors), organized industrial zone, cement factory vicinity, and Seyirtepe recreational area. Sites were selected to represent varying anthropogenic influences, from urban and industrial emissions to relatively clean-air conditions. The figure highlights traffic and industrial sources as key contributors to contamination and reflects how geographic and topographic factors guided the sampling strategy. This spatial framework provides a basis for evaluating and comparing heavy metal accumulation across diverse environmental settings. Figure 2 presents the probability density functions of predictor variables, showing that most exhibit positively skewed distributions with peaks at lower concentrations, indicating predominantly low-to-moderate contamination levels, while some display broader or multimodal patterns, reflecting spatial variability and heterogeneous pollution sources.

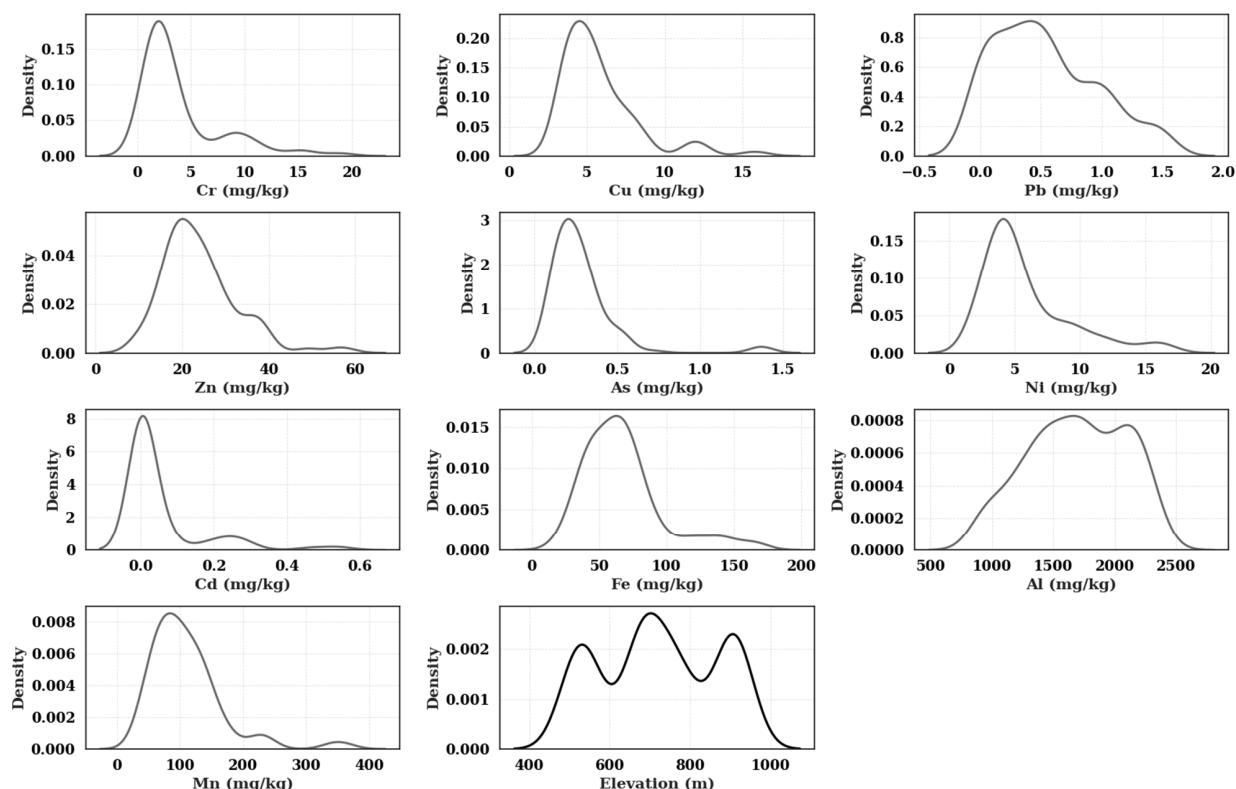


Figure 2. Probability density functions of predictor variables.

Figure 3 presents the correlation heatmap between predictor variables and elevation, illustrating pairwise relationships among heavy metals. Elevation demonstrates weak to moderate associations, with Cu exhibiting the strongest negative correlation (-0.32) and Zn the strongest positive correlation (0.38). Pb, Cd, and Al show modest positive trends, whereas As, Cr, Ni, and Mn display negative trends. These patterns indicate that topography exerts subtle influences on metal distribution, which interact with additional environmental and geochemical factors. Figure 4 presents box-and-whisker plots of the variables, highlighting their central tendency, variability, and outliers. Most heavy metals exhibit right-skewed distributions with notable extreme values, indicating heterogeneous concentrations across sampling locations. Figure 5 illustrates the relationships between heavy metal concentrations and elevation, showing predominantly linear trends, with both positive and negative correlations depending on the metal, reflecting varying environmental influences on their distribution.

Python 3.11 was used for data preprocessing, model development, and performance evaluation, primarily through the Spyder IDE, which facilitated efficient dataset management, model training, and systematic assessment of predictive performance. Python's intuitive syntax, extensive libraries, cross-platform compatibility, and computational efficiency enabled the implementation of complex modeling workflows [53]. Python, with libraries such as NumPy and Pandas and ML frameworks including MLP, CatBoost, Random Forest, XGBoost, LightGBM, and Gradient Boosting, enabled efficient data processing, modeling, and visualization. Its versatility ensures reproducibility and is well-suited for analyzing heavy metal concentrations and environmental processes in civil and environmental engineering [54].

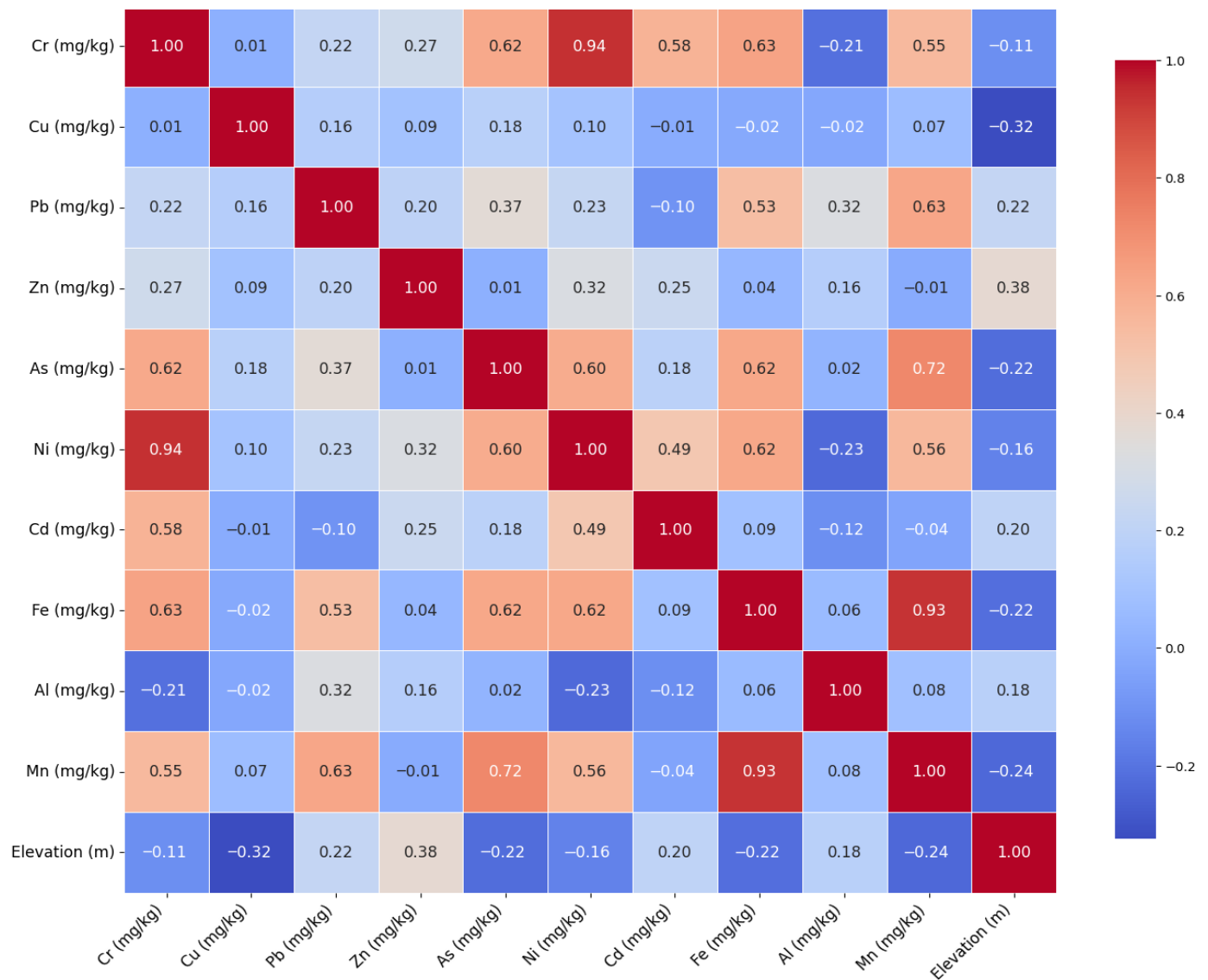


Figure 3. Feature correlation matrix for predictor variables and target.

Plant and soil samples were oven-dried at 105 °C for 24 h. For plant analysis, 0.25 g of each sample was digested with 5 mL of 65% HNO₃ in a Berghof MSW-4 microwave system at 50 bar and 190 °C for 5 min. Aliquots of 0.5 mL (for As, Pb, Cu, Cr, Ni, Cd, Zn) and 0.2 mL (for Fe, Mn, Al) were diluted to 10 mL with ultrapure water and analyzed using a PerkinElmer NexION 350X ICP-MS (PerkinElmer Inc., Waltham, MA, USA). Soil samples (0.2 g) were digested with 6 mL 65% HNO₃ and 2 mL 40% HF at 40 bar and 180 °C for 25 min, followed by the addition of 20 mL saturated boric acid and further digestion at 40 bar and 190 °C for 10 min. After dilution (0.1 mL to 10 mL), elemental concentrations were determined via ICP-MS. All results are expressed in mg/kg.

Outlier detection was conducted using the Isolation Forest algorithm, which identifies anomalous observations based on random partitioning, whereby data points with shorter average path lengths are more likely to be classified as outliers. A contamination rate of 5% was adopted to exclude extreme values while preserving the intrinsic statistical structure of the dataset. This threshold was selected considering the moderate variability typically observed in environmental datasets, ensuring an appropriate balance between noise reduction and information retention. Furthermore, sensitivity analysis confirmed that the exclusion of outliers did not significantly influence model performance or alter the

overall predictive trends, thereby supporting the robustness and stability of the proposed modeling framework.

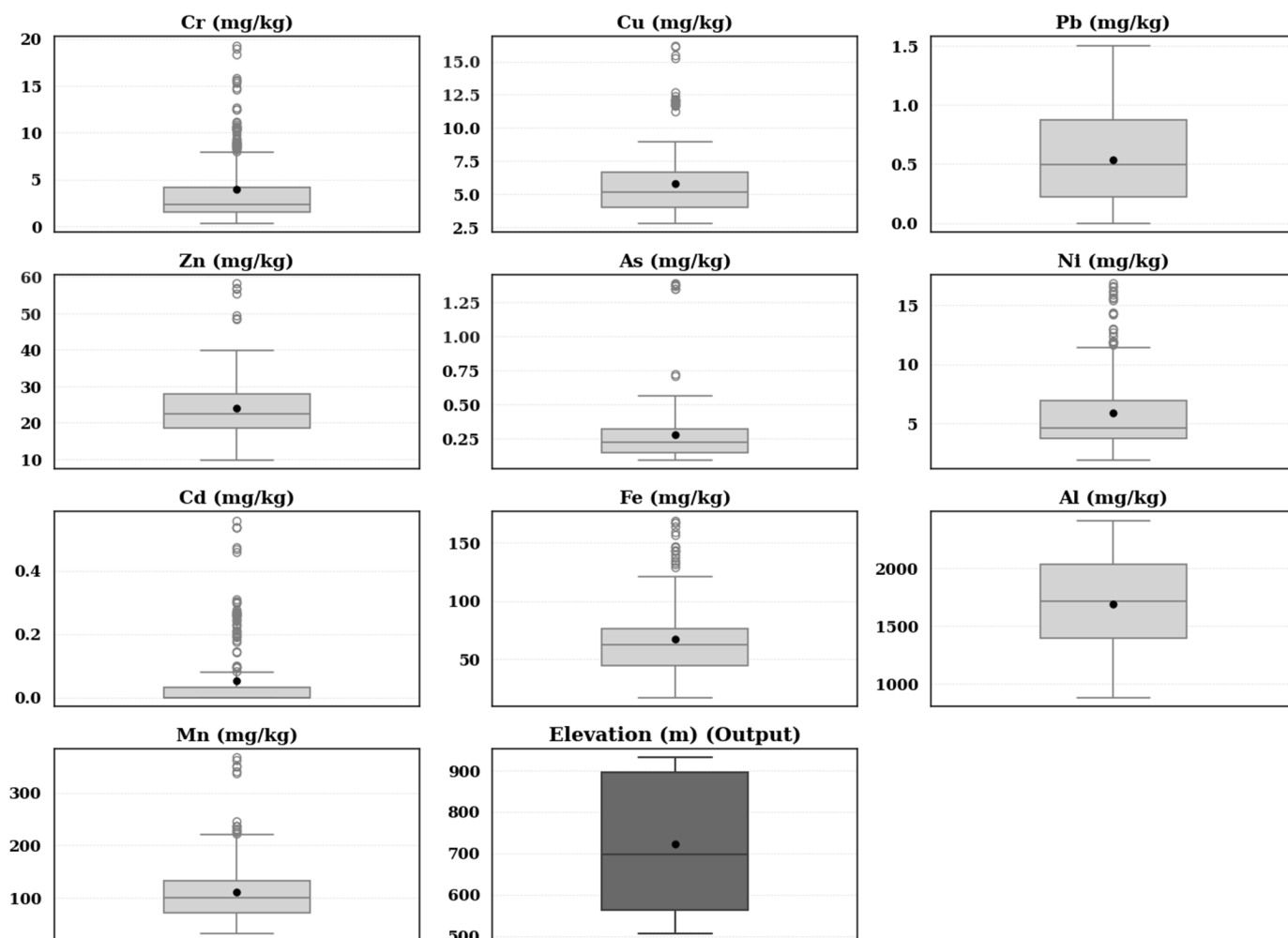


Figure 4. Box-and-whisker plots of variables.

Principal Component Analysis (PCA) is a multivariate statistical technique used to reduce data dimensionality and identify patterns by transforming correlated variables into a set of uncorrelated components. In this study, PCA was applied to the standardized dataset to group variables with similar behavior and identify potential sources of heavy metals, such as industrial emissions, traffic, and geogenic contributions. Table 2 presents the PCA loadings and explained variance of heavy metals. The first four principal components account for 82.72% of the total variance, indicating a strong representation of the dataset. PC1 is dominated by Cr, Ni, Fe, Mn, and As, suggesting anthropogenic sources related to industrial and traffic emissions. PC2 is mainly influenced by Cd and Pb, indicating mixed sources, while PC3 is associated with Zn and Al, reflecting combined industrial and geogenic contributions. PC4 is strongly dominated by Cu, indicating a distinct source. Overall, the results demonstrate that heavy metal distribution is controlled by multiple anthropogenic and natural sources.

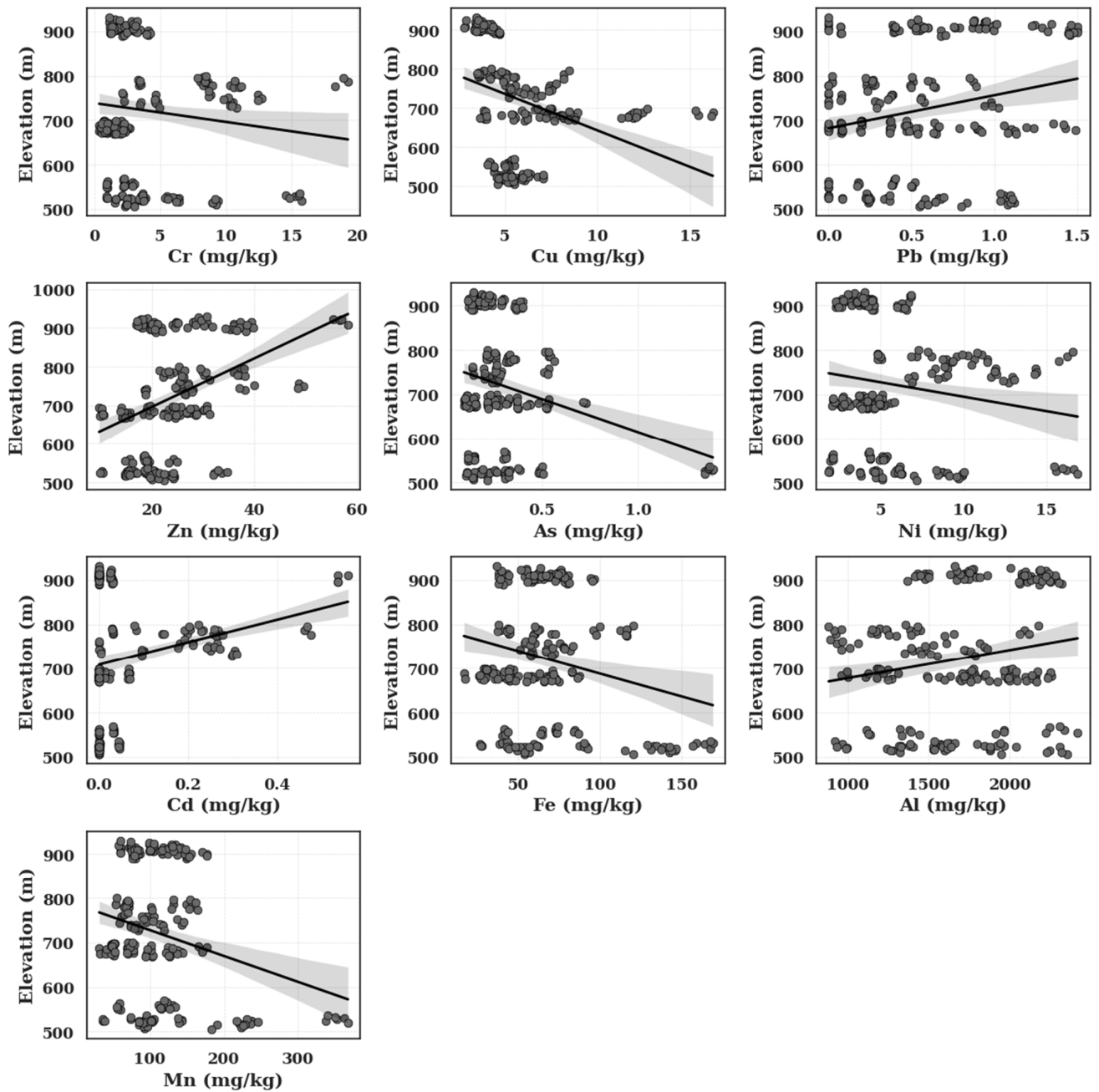


Figure 5. Input–output relationships between heavy metals and elevation.

2.1. Multilayer Perceptron

MLP is a type of feedforward neural network capable of modeling complex nonlinear relationships between inputs and outputs. Its main advantage is flexibility, allowing it to capture subtle patterns in high-dimensional datasets [55]. In civil and environmental engineering, MLP is widely applied for predicting concrete compressive strength, water quality parameters, air pollution levels, and structural behavior under variable loads [56]. It can handle both continuous and categorical data if properly encoded. MLP requires careful hyperparameter tuning to avoid overfitting but provides high predictive accuracy when optimized. Its ability to model nonlinear and interaction effects makes it particularly useful for environmental risk assessment and sustainable infrastructure modeling.

Table 2. PCA loadings and explained variance of heavy metals.

Variable	PC1	PC2	PC3	PC4
Cr (mg/kg)	0.429	0.309	−0.013	−0.071
Cu (mg/kg)	0.060	−0.060	0.214	0.924
Pb (mg/kg)	0.267	−0.442	0.251	0.020
Zn (mg/kg)	0.118	0.168	0.718	−0.059
As (mg/kg)	0.393	−0.086	−0.137	0.138
Ni (mg/kg)	0.427	0.290	0.015	0.027
Cd (mg/kg)	0.177	0.522	0.217	−0.115
Fe (mg/kg)	0.426	−0.207	−0.159	−0.141
Al (mg/kg)	−0.013	−0.420	0.501	−0.290
Mn (mg/kg)	0.422	−0.309	−0.180	−0.029
Variance (%)	41.33	18.31	12.73	10.35
Cumulative (%)	41.33	59.64	72.37	82.72

2.2. CatBoost

CatBoost is a gradient boosting algorithm optimized for categorical data, reducing the need for extensive preprocessing [57]. Its advantages include fast training, high accuracy, and robustness to overfitting, especially with heterogeneous datasets. In civil and environmental engineering, CatBoost is used for predicting pollutant concentrations, assessing soil contamination, modeling groundwater quality, and forecasting urban air pollution [58]. It handles missing values and categorical variables naturally, which is common in environmental datasets. The algorithm's efficiency allows it to process large-scale monitoring data efficiently. Its robustness and interpretability support decision-making in infrastructure planning and environmental risk mitigation.

2.3. Random Forest

Random forest is an ensemble of decision trees that provides stable and reliable predictions by averaging multiple models [59]. Its advantages include robustness to noise, low risk of overfitting, and interpretability through feature importance. Applications in civil and environmental engineering include predicting material properties (e.g., concrete, asphalt), flood risk mapping, soil contamination assessment, and water quality prediction. Random forest can handle large datasets with many features and is effective for both regression and classification tasks. It can also identify key factors influencing structural or environmental performance [60]. Its simplicity and strong predictive performance make it widely used for environmental monitoring and sustainable infrastructure design.

2.4. XGBoost

XGBoost is a gradient boosting framework optimized for speed and performance with built-in regularization [61]. Advantages include high predictive accuracy, handling missing data, and scalability for large datasets. In civil and environmental engineering, XGBoost is applied to predict structural loads, air pollution levels, water quality indices, and the strength of construction materials. It is particularly useful for datasets with nonlinear relationships and interactions among features. The algorithm's efficiency allows processing of high-frequency environmental monitoring data. Its interpretability and reliability support informed decision-making for hazard assessment and infrastructure resilience.

2.5. LightGBM

LightGBM is a fast, efficient gradient boosting method that grows trees leaf-wise, achieving lower loss than traditional approaches [62]. Advantages include computational efficiency, high accuracy, and native handling of categorical features. In civil and environmental engineering, LightGBM is used for flood prediction, traffic-induced air pollution modeling, groundwater contamination forecasting, and concrete durability assessment [63,64]. Its leaf-wise growth enables it to capture complex nonlinear relationships effectively. LightGBM is scalable to very large datasets, making it suitable for real-time environmental monitoring applications. Proper regularization ensures prevention of overfitting while maintaining model interpretability.

2.6. Gradient Boosting

Gradient boosting builds sequential models to minimize residual errors, offering flexibility and high predictive power [65]. Its advantages include adaptability to both regression and classification tasks, the ability to capture nonlinear effects, and interpretability through feature importance. Applications in civil and environmental engineering include predicting pollutant transport, estimating structural reliability, forecasting urban air quality, and modeling soil and water contamination [66]. Gradient boosting handles complex interactions among variables effectively. It requires careful tuning of the learning rate and tree depth to avoid overfitting. Its ability to generate accurate predictions in heterogeneous datasets makes it valuable for environmental decision support and infrastructure management.

2.7. Performance Metrics for Prediction Models

Performance metrics quantitatively assess a model's accuracy, reliability, and ability to generalize the relationship between inputs and outputs [67]. Common metrics such as R^2 , RMSE, MAE, MAPE, and EVS provide complementary insights into model performance. R^2 quantifies the proportion of variance explained by predictors, while RMSE measures the magnitude of squared errors, penalizing large deviations. MAE computes average absolute deviations, offering robust interpretability, and MAPE expresses errors as percentages, enabling scale-independent comparisons. EVS indicates the proportion of observed variance captured, reflecting the model's ability to reproduce data variability. Collectively, these statistical indices provide a rigorous and multidimensional framework for evaluating the accuracy, robustness, and generalizability of predictive models. Model performance was indicated by R^2 , RMSE, MAE, MAPE, and EVS (Equations (1)–(5)).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - p_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i| \quad (3)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - p_i|}{y_i} \quad (4)$$

$$EVS = 1 - \frac{Var(y - p_i)}{Var(y)} \quad (5)$$

where y_i and p_i are the observed and predicted values, $\bar{y}$ is the mean of observed values, and n is the number of observations. Var represents the variance, quantifying the dispersion of data around the mean. R^2 , RMSE, MAE, MAPE, and EVS collectively assess model performance, with R^2 and EVS indicating explained variance and fit, while RMSE, MAE, and MAPE quantify prediction errors and accuracy.

3. Results and Discussion

In needle samples near the cement industry, maximum concentrations occurred in autumn for Cr, Ni, Fe, Mn, and Pb; in spring for Al; in winter for Zn and Cu; and for Cd in autumn and winter, while minimum levels were mostly observed in summer or across all seasons for Cd (Figure 6). Near Seyirtepe, autumn maxima were observed for Cr, Ni, Cu, and Pb; winter for Zn; summer for Al, Fe, Mn, As, and Cd; with minima occurring variably across elements and seasons (Figure 7). Needles near the boulevard showed autumn maxima for Cr, As, Zn, Ni, Cd, and Al, winter maxima for Cu, Pb, Mn, and Fe, and minima mostly in summer or winter, depending on the element (Figure 8). In the organized industrial zone, autumn maxima were observed for Cr, Cu, As, Cd, and Pb; winter for Mn; summer for Fe, Al, and Zn, while minima occurred variably across elements, with no spring maxima and winter minima absent for several elements (Figure 9).



Figure 6. Sampling locations near the cement industry in Adıyaman, Türkiye.

In needle samples collected near the cement industry, maximum concentrations were observed in autumn for Cr, Ni, Fe, Mn, and Pb; in spring for Al; in winter for Zn and Cu; and for Cd in both autumn and winter, while minimum levels generally occurred in summer or remained consistently low for Cd (Figure 6).



Figure 7. Sampling locations in the Seyirtepe area.



Figure 8. Sampling locations in the Boulevard (traffic-influenced area).



Figure 9. Sampling locations in the organized industrial zone.

In the Seyirtepe area, autumn maxima were recorded for Cr, Ni, Cu, and Pb; winter for Zn; and summer for Al, Fe, Mn, As, and Cd, with variable minima across seasons (Figure 7).

Needle samples from the boulevard exhibited autumn maxima for Cr, As, Zn, Ni, Cd, and Al, and winter maxima for Cu, Pb, Mn, and Fe, with minima typically occurring in summer or winter depending on the element (Figure 8). In the organized industrial zone, autumn maxima were observed for Cr, Cu, As, Cd, and Pb; winter for Mn; and summer for Fe, Al, and Zn, while minima varied across elements, with no spring maxima recorded (Figure 9). These seasonal patterns indicate that heavy metal accumulation is strongly influenced by anthropogenic emissions and meteorological conditions, with elevated concentrations during autumn and winter likely associated with increased emissions and reduced atmospheric dispersion.

These patterns indicate industrial activities strongly influence metal deposition in autumn and winter, with Cr exceeding the 0.5 mg/kg WHO/FAO plant limit [68]. Cr in *Pinus nigra* needles exceeded the limit in Seyirtepe, peaking in early–mid autumn, with lower but similar trends along the boulevard; extremely high Cr was observed in the industrial zone and elevated near the cement factory, while Cu remained below the 5 mg/kg WHO/FAO limit [68]. Cu concentrations in Seyirtepe needles remained below the permissible limit with minimal seasonal variation, while boulevard, cement factory, and industrial zone samples peaked in spring, summer, and autumn; Figure 10 shows seasonal Cr and Cu fluctuations across all sites. Table 3 presents the seasonal variation of heavy metal concentrations in soil samples collected from the cement industry area. Overall, Ni and Cr exhibit relatively high concentrations across all seasons, indicating strong anthropogenic influence, while Cd remains consistently low. Seasonal fluctuations are evident, with higher concentrations generally observed in autumn and spring, reflecting

the combined effects of emission intensity and environmental conditions. These patterns highlight the influence of industrial activities on soil contamination.

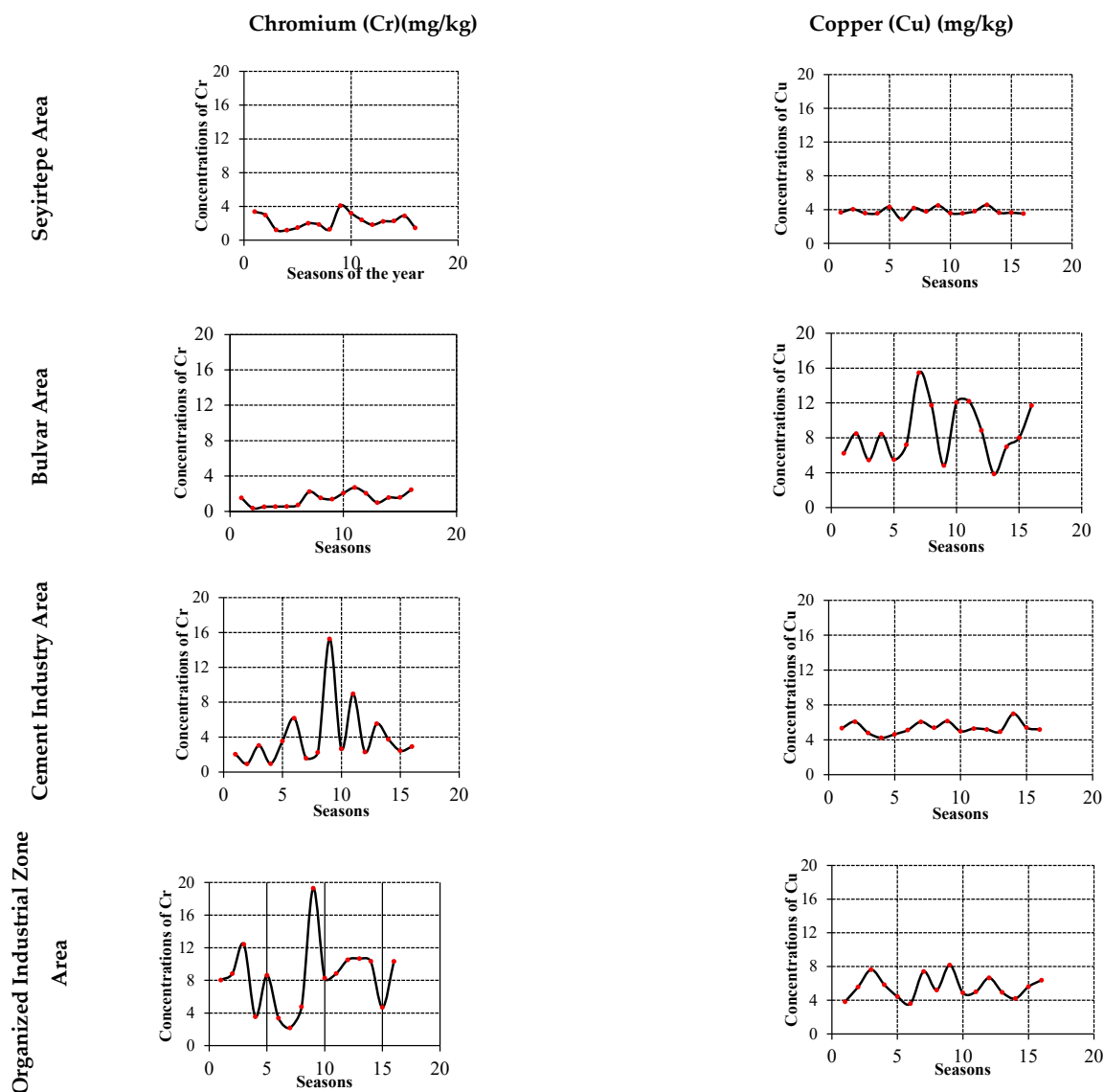


Figure 10. Seasonal variation of Cr and Cu concentrations in black pine needle samples collected from different locations.

Figure 10 shows the seasonal variation of Cr and Cu concentrations in black pine needles collected across all sampling sites. Chromium exhibited pronounced seasonal peaks, especially in industrial and high-traffic areas, reflecting localized emission intensity. Copper concentrations also displayed seasonal variability, frequently reaching maximum levels during colder months, likely driven by increased anthropogenic emissions. In contrast, levels in clean-air sites generally remained below WHO/FAO limits, indicating minimal contamination. The observed seasonal fluctuations underscore the combined influence of local pollution sources, atmospheric conditions, and the physiological activity of the needles. Industrial and traffic-impacted areas consistently showed elevated Cr and Cu concentrations, highlighting the spatial dimension of contamination. Figure 10 demonstrates that both seasonal and spatial factors jointly govern the accumulation patterns of Chromium and Copper in urban and peri-urban vegetation.

Table 3. Seasonal heavy metal concentrations in soil samples collected from the Cement Industry area.

Season	Cu (mg/kg)	Zn (mg/kg)	As (mg/kg)	Ni (mg/kg)	Cd (mg/kg)	Cr (mg/kg)	Pb (mg/kg)	Fe (mg/kg)	Al (mg/kg)	Mn (mg/kg)
Summer	14.83	34.08	4.95	106.44	0.00	69.89	0.33	3.00	13.73	7.49
	14.88	38.29	5.45	108.88	0.00	71.97	0.49	2.91	11.63	7.20
	22.33	47.83	4.78	155.31	0.00	64.07	2.27	4.14	9.13	10.42
	23.78	59.57	4.07	140.43	0.00	76.85	1.83	4.35	14.18	11.03
	24.60	69.46	4.39	158.31	0.00	76.33	2.12	4.02	13.21	10.06
	18.03	44.71	3.86	122.50	0.00	51.83	1.33	3.20	6.53	8.14
	19.97	42.71	5.96	138.37	0.00	68.75	2.18	3.48	9.92	8.87
	17.35	39.89	5.86	127.92	0.00	79.88	1.39	2.55	12.39	6.32
Spring	19.56	74.80	5.70	147.42	0.93	101.79	1.29	2.66	11.25	6.63
	19.65	51.04	7.11	174.75	0.00	83.58	3.43	3.24	6.29	8.42
	26.02	64.43	3.64	162.45	0.00	81.65	2.74	3.66	13.28	9.22
	17.49	34.37	3.70	129.38	0.00	70.32	2.03	2.76	6.23	6.93
	21.50	42.38	4.97	154.63	0.00	79.94	1.97	3.71	12.33	9.43
	19.61	40.69	5.38	139.68	0.00	61.11	3.44	3.12	5.34	7.96
	19.99	56.24	4.89	147.77	0.00	90.50	0.92	3.10	12.95	5.50
	22.67	48.78	6.80	154.13	0.00	83.33	3.12	3.33	6.73	5.93
Fall	25.15	72.57	3.91	187.95	0.00	82.62	2.94	3.14	11.42	5.65
	24.40	105.11	5.49	170.53	1.38	117.11	1.22	2.28	9.87	4.08
	26.34	47.28	6.34	162.54	0.00	72.01	3.50	3.61	5.95	6.78
	22.37	47.91	5.95	156.55	0.00	74.52	3.08	3.48	6.45	6.43
	15.40	41.49	5.19	113.40	0.00	62.61	0.89	2.35	11.72	5.08
	20.49	41.39	3.89	121.95	0.00	57.23	0.86	2.58	12.09	5.31
	28.01	53.48	7.93	129.57	0.00	65.21	8.40	3.17	8.17	6.88
	22.54	61.19	3.63	145.03	0.00	69.46	1.89	3.44	13.82	7.30
Winter	24.54	36.80	3.50	134.74	0.00	87.62	0.27	3.57	18.09	7.82
	19.19	43.37	5.69	140.35	0.00	62.07	3.10	3.10	5.75	6.75
	23.47	45.14	5.33	140.58	0.00	66.21	3.73	3.73	6.24	8.29

According to WHO/FAO, the permissible limit of lead in plants is 2 mg/kg [68].

In the Seyirtepe area, copper concentrations in black pine needles remained below the WHO/FAO permissible limits for plants. Lead concentrations increased during mid-winter and mid-spring, while early and mid-summer and autumn values approached the permissible lead threshold (Figure 11). Along the boulevard, copper levels were also well below the permissible limits, whereas lead concentrations rose during winter, spring, and mid-summer and autumn (Figure 11). In the cement industry and organized industrial zone, lead concentrations in needles were generally below WHO/FAO limits, although elevated values were observed at the beginning and end of autumn (Figure 11). The highest arsenic concentration was recorded at the beginning of autumn in the cement industry area. Figure 11 presents the seasonal variation of Pb and As concentrations in black pine needle samples collected from different sites, illustrating how Pb and As levels fluctuate annually in response to local environmental exposure and pollution sources.

Figure 11 illustrates the seasonal variation of Pb and As concentrations in pine needles across all sampling sites. Pb levels generally peaked during the colder months, reflecting enhanced deposition from heating and traffic-related emissions. Arsenic also exhibited clear seasonal peaks, particularly in the vicinity of the cement factory, indicating the impact of industrial activities. In the Seyirtepe and Boulevard areas, Pb concentrations approached permissible limits during certain seasons, highlighting potential local exposure risks. Although industrial zones generally recorded Pb and As below regulatory thresholds, autumn exhibited noticeable increases for both metals. Arsenic reached its highest concentration at the onset of autumn near the cement plant, emphasizing site-specific accumulation. Figure 11 demonstrates that the seasonal and spatial distribution of Pb and As is strongly influenced by industrial operations, vehicular traffic, and meteorological conditions.

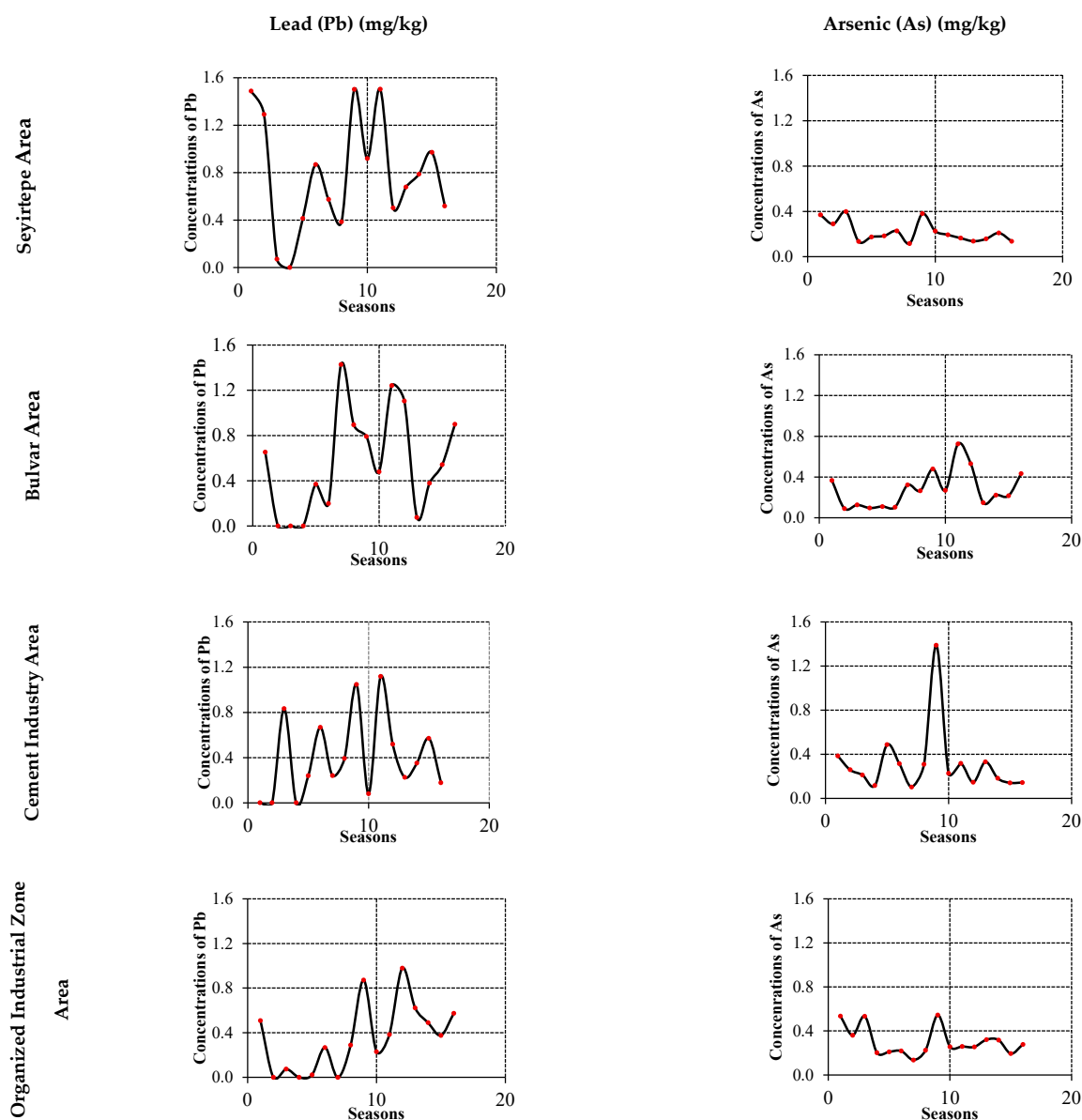


Figure 11. Seasonal variation of Pb and As concentrations in black pine needle samples collected from different locations.

According to WHO/FAO, the permissible limit of zinc in plants is 50 mg/kg [68].

In the Seyirtepe area, black pine needles exhibited both sub-threshold and above-threshold zinc concentrations relative to the WHO/FAO limit. Increases in zinc values were observed at the beginning of summer and spring, as well as at the end of autumn, while concentrations exceeded the permissible limit during mid-autumn (Figure 12). Along the Boulevard area, black pine needles showed zinc concentrations below the WHO/FAO permissible limit. However, slight increases were recorded in mid-spring, mid-summer, and mid-autumn (Figure 12). In the cement industry area, black pine needles displayed zinc concentrations well below the WHO/FAO permissible limit, although an increase was observed during mid-winter (Figure 12). In the organized industrial zone area, black pine needles exhibited zinc concentrations slightly below the WHO/FAO permissible limit. A notable increase was recorded during mid-summer (Figure 8). According to WHO/FAO, the permissible limit of nickel in plants is 5 mg/kg [68].

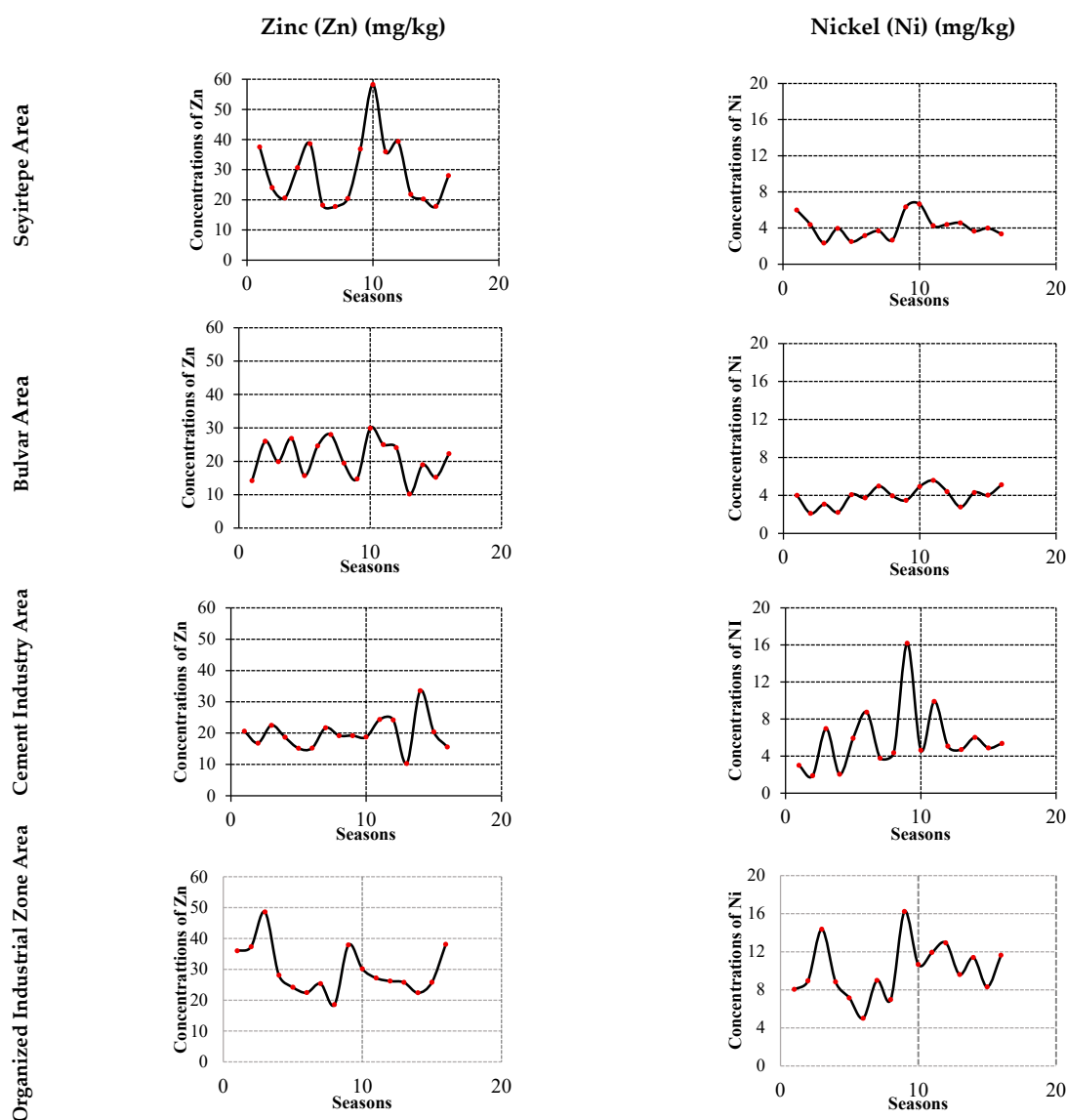


Figure 12. Seasonal variation of Zn and nickel Ni concentrations in black pine needle samples collected from different locations.

In the Seyirtepe and Boulevard areas, nickel concentrations in black pine needles remained close to the WHO/FAO permissible limits, exhibiting relatively consistent seasonal fluctuations (Figure 12). In the cement industry area, nickel levels exceeded the permissible threshold, with particularly pronounced increases at the beginning of autumn, while concentrations during other seasons approached the limit (Figure 12). In the organized industrial zone, nickel concentrations were substantially above the WHO/FAO permissible limit, peaking during mid-summer and early autumn (Figure 12). Figure 12 presents the seasonal variation of Zn and Ni concentrations in black pine needle samples collected from various sites. Zn levels were generally elevated in winter, whereas Ni concentrations peaked in autumn, reflecting the combined influence of local pollution sources and seasonal environmental conditions. Copper (Cu) concentrations are reported in terms of seasonal and spatial trends without reference to permissible limits, as Cu is an essential micronutrient for plants, and its sufficiency rather than toxicity is of primary concern. Similarly, iron (Fe) and manganese (Mn) concentrations remained within physiologically normal ranges for black pine (*Pinus nigra*), ensuring adequate photosynthetic function. These clarifications provide context for interpreting metal accumulation patterns and avoid

conflating essential nutrients with strictly toxic elements. Sampling was performed on healthy *Pinus nigra* trees with similar morphological characteristics, and soil samples were collected from the immediate vicinity of each sampled tree to ensure spatial correspondence. Although detailed meteorological and geological parameters were not directly measured, their effects are inherently reflected in the spatial and seasonal patterns of heavy metal accumulation, which are influenced by emission sources, atmospheric transport, and local geogenic background.

Figure 12 illustrates the seasonal variation of Zn and Ni concentrations in pine needles across all sampling sites. Zn levels were generally higher in winter, particularly at Seyirtepe and along the boulevard, with occasional exceedances of WHO/FAO permissible limits. Ni concentrations peaked in autumn and surpassed threshold values in the cement industry and organized industrial zone, reflecting substantial industrial influence. In Seyirtepe and boulevard areas, Ni levels approached but generally remained near permissible limits, whereas industrial sites consistently exhibited elevated Ni concentrations compared to cleaner areas. Seasonal variations for both metals are closely linked to emission intensity, atmospheric transport, and local meteorological conditions. These patterns demonstrate the combined effects of industrial activities, traffic, and environmental factors on metal accumulation, highlighting Zn and Ni as pollutants of concern and underscoring the importance of continuous monitoring. Cadmium is currently utilized in nickel–cadmium batteries, steel coatings in the shipbuilding industry due to its corrosion resistance under marine conditions, the paint industry, as a PVC stabilizer, in alloys, and in the electronics sector. Additionally, cadmium impurities are present in phosphate fertilizers, detergents, and refined petroleum products, and the widespread use of these materials can contribute to substantial environmental cadmium contamination. According to WHO/FAO, the permissible limit of cadmium in plants is 0.5 mg/kg [68]. In the Seyirtepe area, along the boulevard, and near the cement industry, *Pinus nigra* needles exhibited cadmium concentrations consistently below WHO/FAO limits, with minor seasonal variation, while the organized industrial zone showed a slight increase in early autumn (Figure 13). Iron concentrations exceeded the 30 mg/kg permissible limit across all sites, with relatively consistent seasonal patterns (Figure 14). Seasonal fluctuations of Cd and Al reflect environmental exposure and local pollution sources.

Figure 13 illustrates the seasonal variation of Cd and Al concentrations in pine needles across all sampling sites. Cadmium levels remained consistently below WHO/FAO permissible limits, indicating limited atmospheric deposition, although minor increases were observed in autumn within the industrial zone, reflecting episodic inputs. In contrast, aluminum concentrations were substantially higher, particularly during summer, suggesting enhanced dust resuspension and soil-derived contributions. Seasonal fluctuations of Al likely reflect both environmental exposure and physiological activity of the needles, with spatial analysis indicating elevated levels near industrial and high-traffic areas, highlighting the influence of local emission sources. Cadmium's consistently low concentrations contrast with aluminum's pronounced seasonal and spatial variability, emphasizing the differential behavior of metals and the need to account for both metal-specific properties and environmental factors in assessing atmospheric contamination. Figure 14 presents the seasonal variation of Fe and Mn concentrations in black pine needles across all sampling sites. Both metals exhibited seasonal fluctuations, with higher concentrations in industrial and traffic-affected areas and lower levels in spring, reflecting the combined influence of local emissions, atmospheric transport, and meteorological conditions.

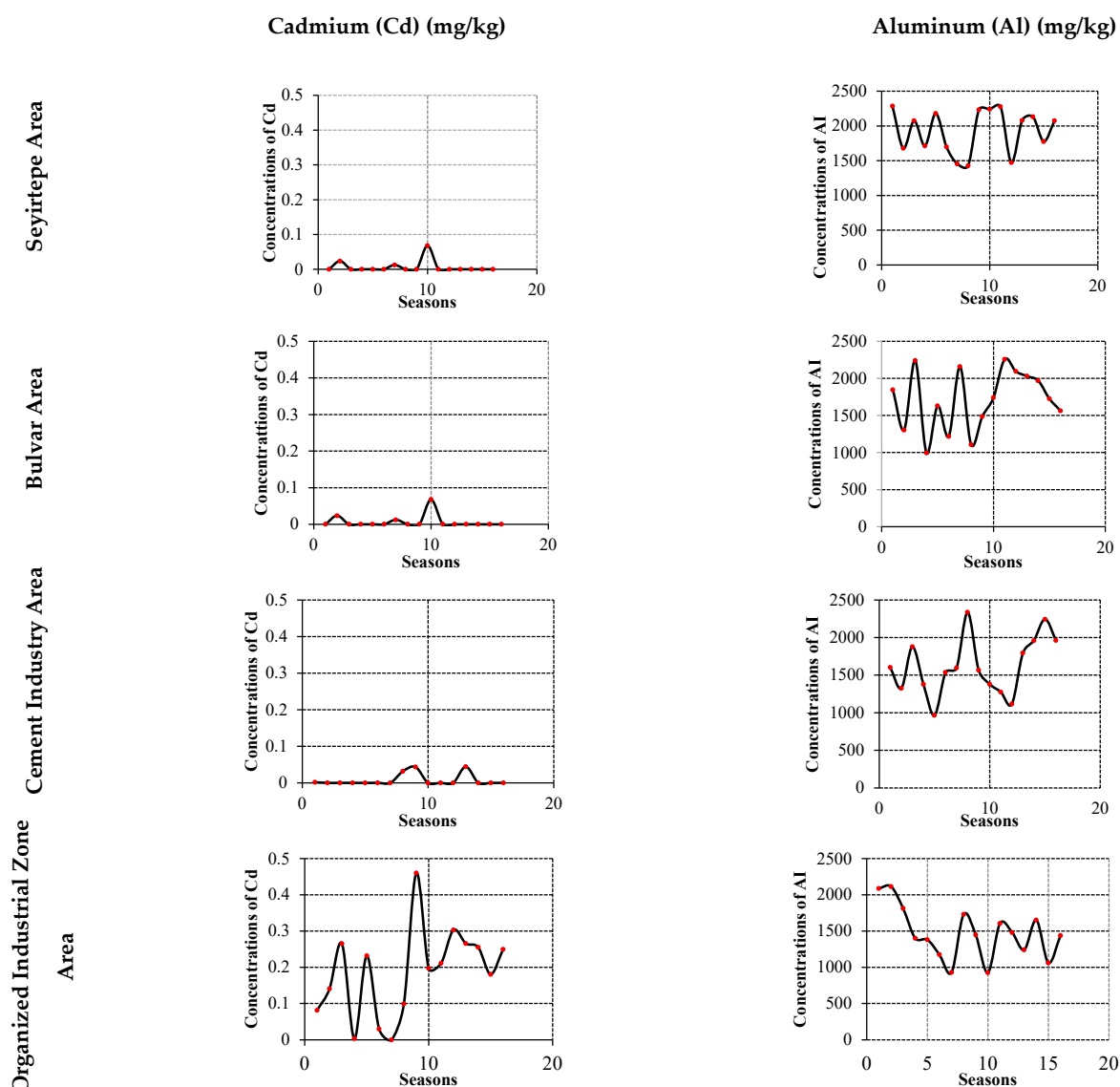


Figure 13. Seasonal variation of Cd and Al concentrations in black pine needle samples collected from different locations.

Figure 14 illustrates seasonal Fe and Mn concentrations in pine needles across all sites, with values exceeding WHO/FAO limits at most locations. Concentrations peaked in colder seasons, particularly winter, due to enhanced deposition from industrial emissions, traffic, heating, and stagnant air, while summer levels were lower, likely influenced by rainfall and needle physiology. Industrial and high-traffic areas, especially the cement factory and organized industrial zone, exhibited the highest levels, highlighting localized emission impacts. Seasonal fluctuations across sites indicate the combined influence of meteorology, natural soil inputs, and anthropogenic sources. Table 3 summarizes seasonal heavy metal concentrations in soils from the cement factory area, with Cu (14.83–28.01 mg/kg), Zn (34.08–105.11 mg/kg), Ni (106.44–187.95 mg/kg), Cr (51.83–117.11 mg/kg), Cd (0–1.38 mg/kg), As (3.50–7.93 mg/kg), Pb (0.27–8.40 mg/kg), Fe (0.89–8.40 g/kg), Al (2.28–18.09 g/kg), and Mn (4.08–11.03 mg/kg). Ni and Cr peaked in spring and autumn, Al and Mn in summer, and Cd remained negligible. Maximum and minimum concentrations varied by season, generally mirroring trends in pine needles. Soil samples were prepared using sequential heating (170–190 °C) and cooling to 50 °C. Table 4 summarizes the microwave digestion parameters for soil samples, comprising

two controlled heating stages with defined temperature, pressure, power, ramp, and hold times to ensure complete digestion. In Step 1, the temperature was increased from 140 to 180 °C at 40 bar and 80% power, followed by cooling to 50 °C, while Step 2 reached 170–190 °C under similar conditions. Soil samples (200–300 mg) were digested with 5 mL of 65% HNO_3 , ensuring efficient and reproducible preparation for accurate ICP-MS analysis. These results highlight the influence of seasonality, site, and emission sources on heavy metal accumulation, supporting targeted planting of black pine to improve air quality in polluted areas.

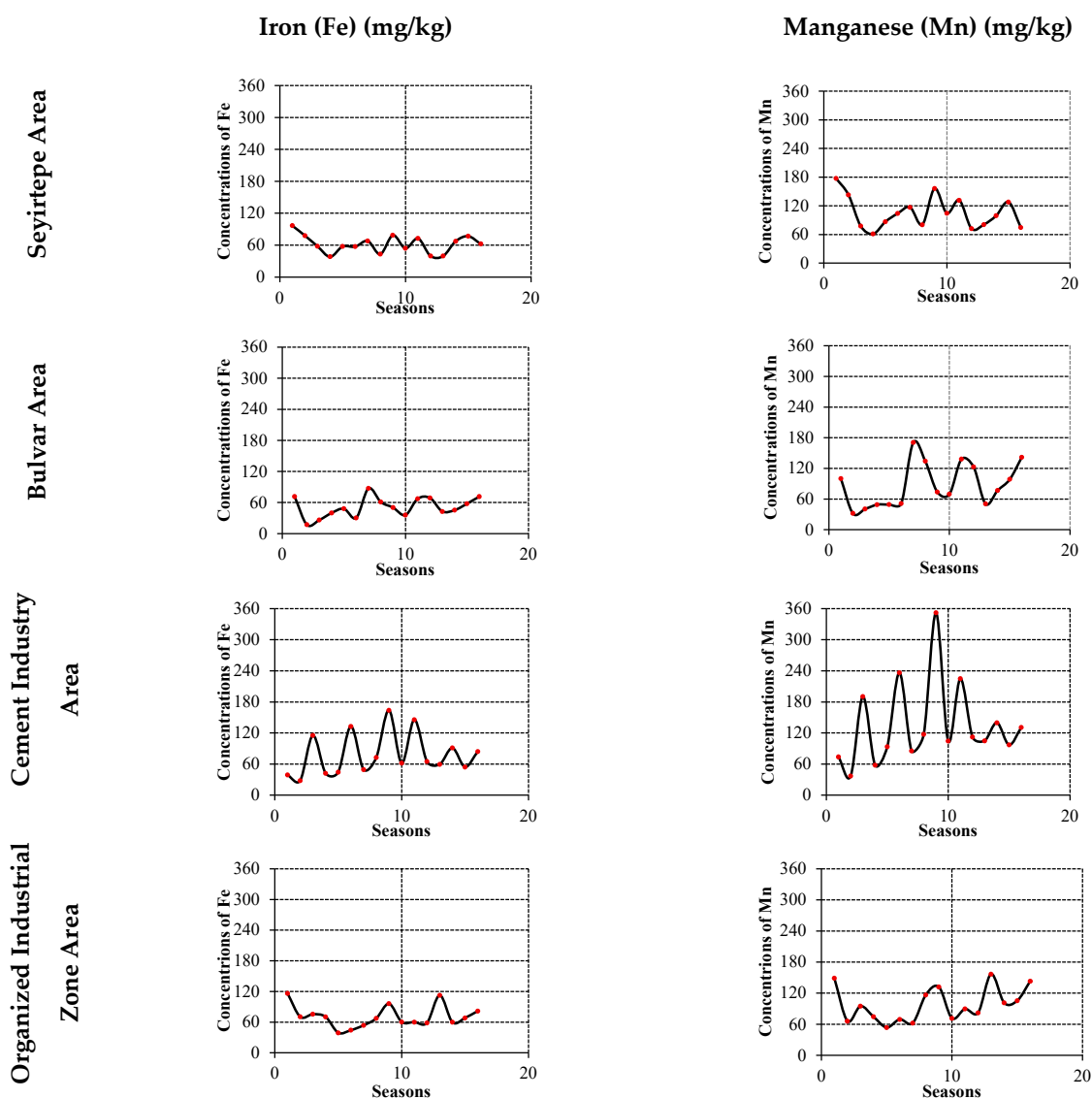


Figure 14. Seasonal variation of Fe and Mn concentrations in black pine needle samples collected from different locations.

Table 5 summarizes the microwave digestion parameters for plant samples in Step 2, consisting of two heating stages at 150 °C and 190 °C (50 bar, 70–90% power), followed by cooling to 50 °C with no power or pressure. Ramp and hold times ensured controlled heating and complete digestion, enabling efficient and reproducible preparation for ICP-MS analysis. These physiological effects are associated with the structural and functional characteristics of coniferous needles. The waxy cuticle enhances the adsorption of airborne particulate matter, while stomatal pathways allow the uptake of certain trace elements into internal tissues. In addition, seasonal physiological processes influence metal concentra-

tions: active growth during warmer periods may dilute metal content, whereas reduced metabolic activity in colder seasons promotes the accumulation of deposited metals. These mechanisms contribute to the observed temporal variations and support the role of *Pinus nigra* needles as effective bioindicators of atmospheric heavy metal deposition.

Table 4. Microwave digestion program parameters applied to soil samples prior to ICP-MS analysis [69].

Step 1	1	2	3	4
T (°C)—Temperature	140 °C	160 °C	180 °C	50 °C
P (bar)—Pressure	40	40	40	0
Power (%)—Power	80	80	80	0
Ta (min)—Ramp time	10	5	5	0
Time (min)—Hold time	2	5	25	0
Step 2	1	2	3	
T (°C)—Temperature	170 °C	190 °C	50 °C	
P (bar)—Pressure	40	40	0	
Power (%)—Power	80	80	0	
Ta (min)—Ramp time	5	10	0	
Time (min)—Hold time	5	10	0	

Table 5. Microwave digestion parameters for plant samples (step 2).

Step 2	1	2	3
T (°C)—Temperature	150 °C	190 °C	50 °C
P (bar)—Pressure	50	50	0
Power (%)—Power	70	90	0
Ta (min)—Ramp time	5	1	0
Time (min)—Hold time	5	5	0

Table 6 summarizes ICP-MS operating conditions for soil, sediment, and plant analyses. Samples were analyzed using a Mainhard concentric nebulizer with a cyclonic spray chamber and nickel triple cone interface, under plasma gas (18.0 L/min), auxiliary gas (1.2 L/min), and nebulizer gas (0.92 L/min) at 1500 W RF power and 1 mL/min uptake. Triplicate measurements were conducted in STD/KED mode with helium collision gas to minimize interferences. Digested volumes were adjusted and diluted appropriately to measure Cu, Pb, Zn, As, Ni, Cd, and Cr at ppb levels, and Fe, Al, and Mn at ppm levels, ensuring accurate, sensitive, and reproducible quantification across all sample types.

Table 6. Instrumental operating conditions for ICP-MS analysis.

Component/Parameter	Type/Value/Mode
Nebulizer	Mainhard (concentric)
Spray Chamber	Glass Cyclonic
Triple Cone Interface Material	Nickel
Plasma Gas Flow	18.0 L/min
Auxiliary Gas Flow	1.2 L/min
Nebulizer Gas Flow	0.92 L/min
Sample Uptake Rate	1 mL/min
RF Power	1500 W
Replicates per Sample	3
Mode of Operation	STD/KED Mode
Nebulizer	Collision (using He gas)

A more detailed evaluation of the results reveals distinct spatial and seasonal patterns in heavy metal accumulation. Industrial and high-traffic areas, particularly the cement factory and organized industrial zone, consistently exhibited higher concentrations of Cr, Ni, Fe, and Pb compared to the Seyirtepe reference site, indicating the dominant influence of anthropogenic emission sources such as industrial processes, vehicular emissions, and road dust resuspension. Spatial differences further demonstrate that proximity to emission sources plays a critical role in determining accumulation levels in *Pinus nigra* needles.

Seasonally, elevated concentrations during autumn and winter can be attributed to increased emissions from residential heating and industrial activities, combined with reduced atmospheric dispersion and thermal inversion conditions, which enhance pollutant accumulation. In contrast, lower concentrations observed during summer are likely associated with increased atmospheric mixing, precipitation-induced wash-off, and dilution effects due to active plant growth. These findings highlight that both spatial proximity to pollution sources and seasonal meteorological conditions jointly control heavy metal deposition patterns, reinforcing the effectiveness of *Pinus nigra* needles as bioindicators of environmental pollution. Although certain metals (e.g., Fe, Al, and Mn) may partly originate from geogenic sources related to local geology, the observed spatial and seasonal patterns indicate that anthropogenic emissions are the dominant contributors to heavy metal accumulation. Only mature *Pinus nigra* trees were sampled, as young plants were excluded to avoid variability associated with growth-related differences in metal accumulation. The study area is predominantly characterized by sedimentary geological formations (e.g., limestone, marl, and clay), which may contribute to the natural background levels of certain elements (e.g., Fe, Al, and Mn) in soils.

This study evaluates the predictive performance of MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting for estimating heavy metal concentrations in black pine needles and predicting elevation from environmental features. Hyperparameters were optimized, and model performance was assessed using R^2 , RMSE, MAE, MAPE, and EVS to ensure accuracy, generalization, and sensitivity to complex patterns. Table 7 indicates the optimized hyperparameters for all models, which govern learning and directly influence predictive accuracy, convergence, and generalization. Key parameters include hidden layer sizes, iterations, learning rate, max_depth, n_estimators, batch size, and optimizer selection, tuned via systematic grid search. MLP used two hidden layers (100 neurons each) and 3000 iterations; CatBoost, 500 iterations and depth 6; random forest, 300 estimators and depth 15; XGBoost and LightGBM, 500 estimators with depths 6 and 15, respectively; and gradient boosting, 500 estimators and depth 5, all with random state 42. Figure 15 shows residual distributions for MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting, indicating prediction errors in black pine needle heavy metal concentrations. Residuals near zero reflect high accuracy, with CatBoost and XGBoost exhibiting the tightest distributions, MLP the widest, and random forest, gradient boosting, and LightGBM exhibiting moderate spreads.

Table 7. Hyperparameters of the ML models.

Model	Key Hyperparameters
MLP	hidden_layer_sizes = (100, 100), max_iter = 3000, random state = 42
CatBoost	iterations = 500, depth = 6, learning_rate = 0.05, verbose = 0, random state = 42
Random forest	n_estimators = 300, max_depth = 15, random state = 42
XGBoost	n_estimators = 500, max_depth = 6, learning_rate = 0.05, tree method = 'hist', random state = 42
LightGBM	n_estimators = 500, max_depth = 15, learning_rate = 0.05, random state = 42
Gradient Boosting	n_estimators = 500, max_depth = 5, learning_rate = 0.05, random state = 42

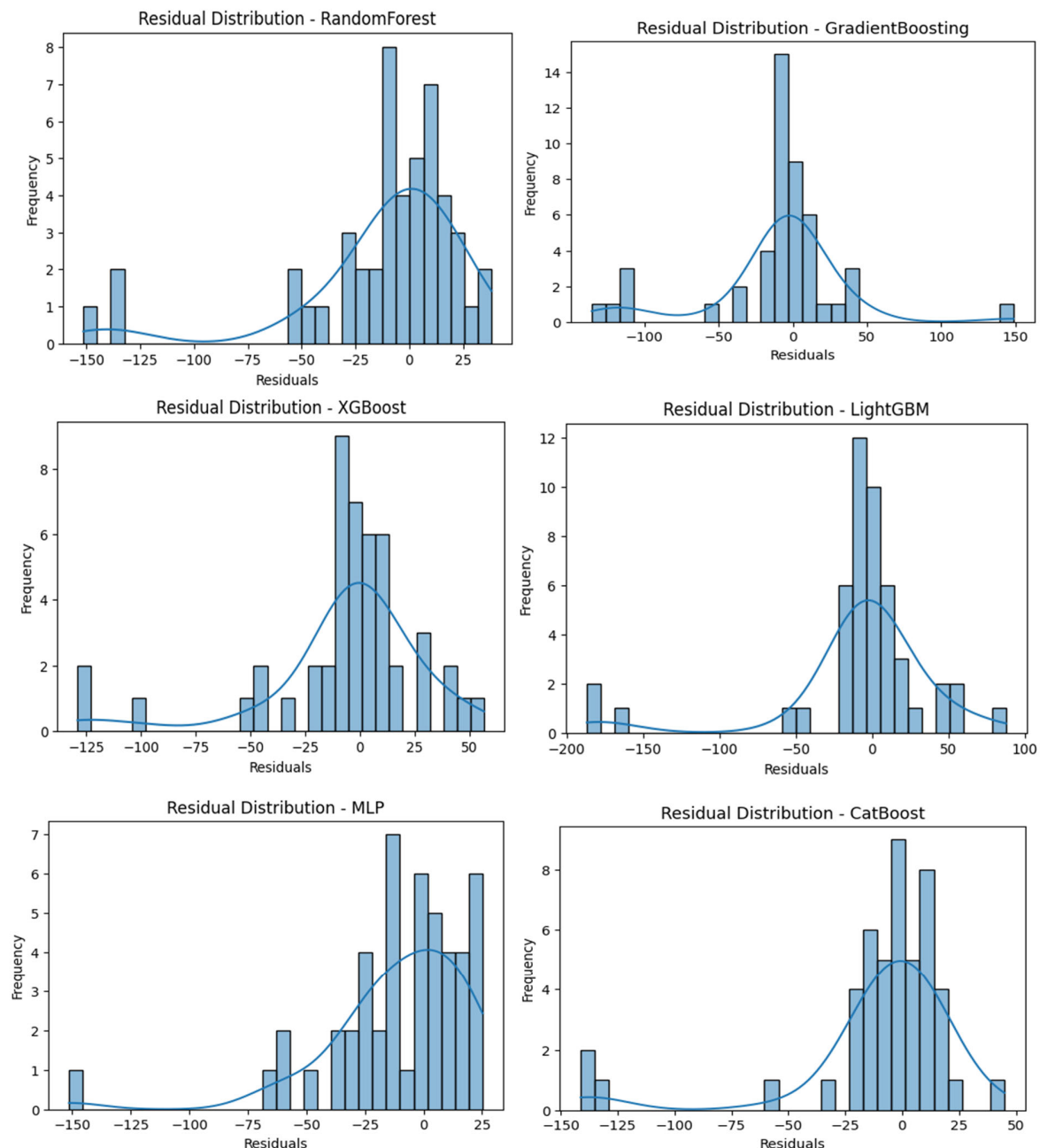


Figure 15. Residual distribution of prediction errors for MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting models.

Figure 16 presents a SHAP (SHapley Additive exPlanations) summary plot illustrating the contribution of each heavy metal variable to the XGBoost model output. Each point represents an observation, where the horizontal position indicates the SHAP value, reflecting both the magnitude and direction of feature influence. Positive values increase the prediction, while negative values decrease it, and the color gradient denotes feature magnitude. Cd, Cu, and Cr exhibit the largest SHAP value dispersion, indicating their dominant and consistent influence, whereas Mn and As show comparatively minor effects. Based on their environmental relevance and statistical behavior, two modeling scenarios were considered: excluding these metals to assess baseline performance and including all variables to evaluate full predictive capacity. The results confirm that feature selection was guided by both domain knowledge and SHAP-based interpretability, ensuring a robust and scientifically justified model. Table 8 summarizes model performance excluding Cd,

Cr, and Cu. Predictive accuracy decreased for all models, with XGBoost achieving the highest test $R^2 = 0.953$, followed by CatBoost ($R^2 = 0.934$), while MLP, LightGBM, random forest, and gradient boosting showed moderate performance. XGBoost and CatBoost thus demonstrated superior accuracy and robustness despite the reduced input set. Figure 17 illustrates the comparative performance of MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting models during the training and testing phases, evaluated using R^2 , RMSE, MAE, MAPE, and EVS, with Cd, Cr, and Cu excluded as input features.

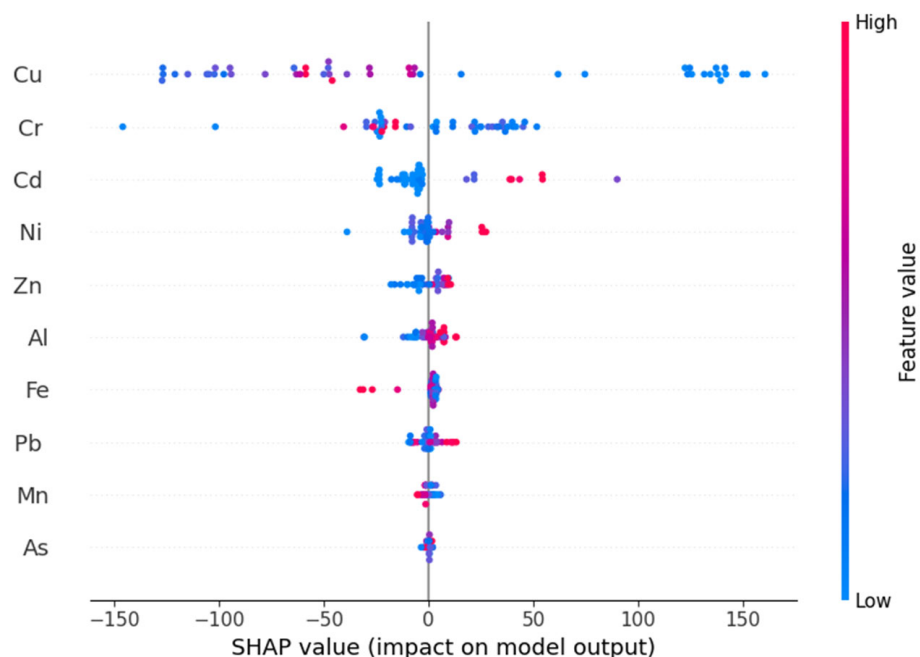


Figure 16. SHAP summary plot illustrating feature importance in the XGBoost model.

Table 8. Model performance without Cd, Cr, and Cu as input parameters.

Algorithms	Training					Testing				
	R^2	RMSE	MAE	MAPE	EVS	R^2	RMSE	MAE	MAPE	EVS
MLP	0.9670	25.9474	17.1155	0.0237	0.9670	0.8725	46.6996	25.9510	0.0362	0.8824
CatBoost	0.9991	4.3769	3.6586	0.0052	0.9991	0.9340	33.6094	22.9954	0.0332	0.9354
Random forest	0.9789	20.7749	14.1088	0.0208	0.9789	0.9167	37.7536	30.2212	0.0440	0.9178
XGBoost	0.9996	0.0681	0.0525	0.0002	0.9996	0.9526	17.7655	19.5467	0.0255	0.9526
LightGBM	0.9926	12.3048	8.9786	0.0126	0.9926	0.8808	45.1428	34.1884	0.0484	0.8829
Gradient Boosting	0.9991	0.1231	0.0840	0.0003	0.9991	0.9093	21.4564	21.8988	0.0299	0.9093

Figure 17 compares model performance with Cd, Cr, and Cu excluded. Boosting models, particularly XGBoost and CatBoost, consistently outperform MLP and random forest across R^2 , RMSE, MAE, MAPE, and EVS, with XGBoost achieving the most balanced and robust results. MLP shows lower accuracy, reflecting sensitivity to limited or skewed features, while LightGBM and gradient boosting perform moderately. Overall, feature exclusion reduces predictive capacity, emphasizing the advantage of advanced boosting algorithms. Figure 18 compares experimental and predicted elevation values for all models under the same feature-constrained scenario.

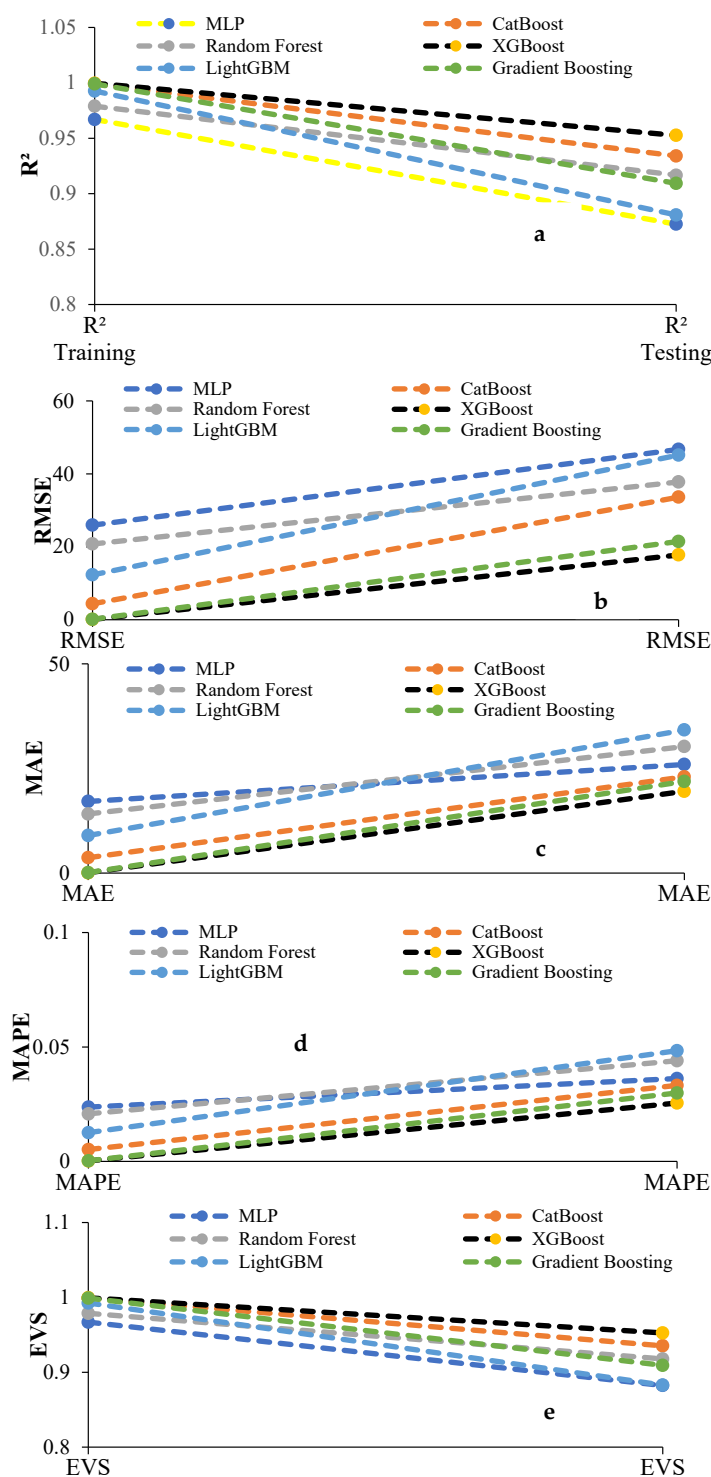


Figure 17. Comparative evaluation of MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting models during training and testing phases based on R^2 , RMSE, MAE, MAPE, and EVS, excluding Cd, Cr, and Cu as input features.

Figure 18 compares actual and predicted elevation values without Cd, Cr, and Cu as inputs. In Figure 18A, predicted values are plotted against actual values, while Figure 18B shows predicted versus actual elevations across training and testing datasets. MLP exhibits high scatter in testing, indicating lower reliability, whereas boosting models (XGBoost, CatBoost, gradient boosting) closely align with observed values. Random forest shows moderate accuracy with occasional overestimation, and LightGBM displays wider deviations. Overall, ensemble methods outperform MLP, demonstrating superior generalization

under feature-constrained conditions. Table 9 summarizes the performance of MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting with Cd, Cr, and Cu included as input features. Inclusion of these metals substantially improved predictive accuracy across all models. XGBoost achieved the highest performance (training $R^2 = 0.9999$, testing $R^2 = 0.9890$), followed by CatBoost and gradient boosting, while MLP remained the least accurate. Boosting models consistently outperformed random forest and LightGBM, highlighting their superior ability to capture complex relationships when critical heavy metals are included.

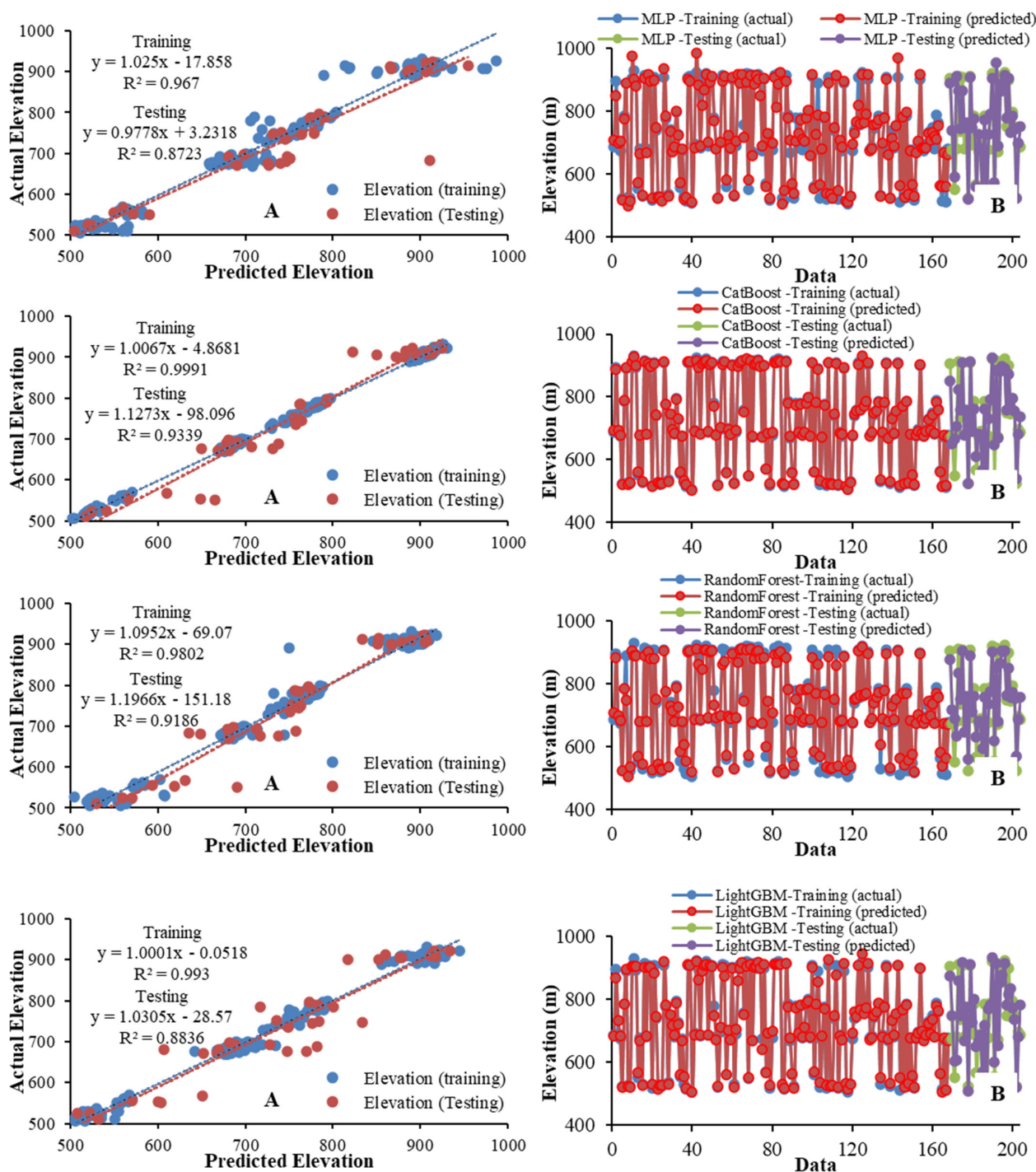


Figure 18. Cont.

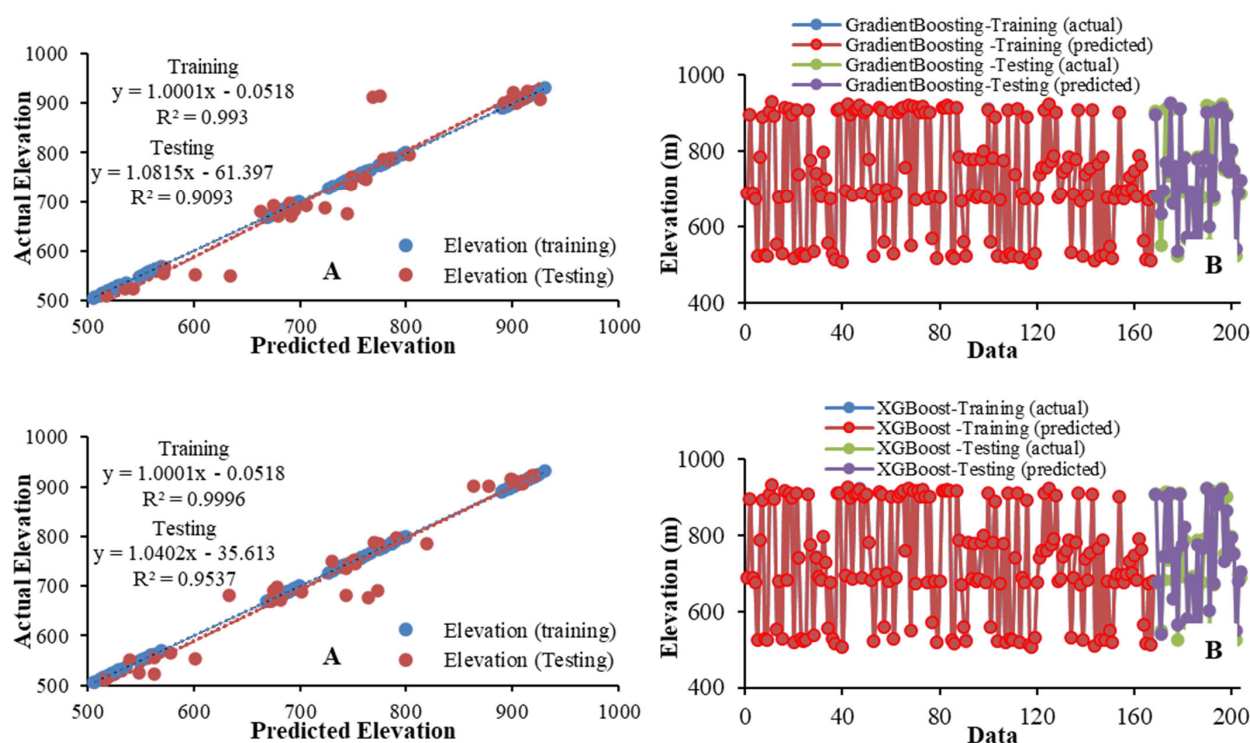


Figure 18. Comparison of experimental and predicted elevation values using MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting models, excluding Cd, Cr, and Cu as input features.

Table 9. Model performance with Cd, Cr, and Cu as input parameters.

Algorithms	Training					Testing				
	R ²	RMSE	MAE	MAPE	EVS	R ²	RMSE	MAE	MAPE	EVS
MLP	0.9816	19.3931	11.3351	0.0154	0.9816	0.885	44.453	24.6043	0.0339	0.885
CatBoost	0.9995	3.2179	2.6214	0.0038	0.9995	0.9770	19.8292	13.9817	0.0208	0.9782
Random forest	0.9893	14.7883	9.1740	0.0137	0.9893	0.9762	20.1879	16.2195	0.0244	0.9766
XGBoost	0.9999	0.0535	0.0512	0.0001	0.9999	0.9890	5.5462	7.5644	0.0114	0.9891
LightGBM	0.9966	8.3047	5.8816	0.0084	0.9966	0.9601	26.1127	17.7540	0.0268	0.9631
Gradient Boosting	0.9999	0.1163	0.0730	0.0001	0.9999	0.9765	7.5654	9.8981	0.0175	0.9765

Figure 19 illustrates the comparative performance of MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting models when Cd, Cr, and Cu are included as input features. Figure 19a illustrates that boosting models achieve R² values approaching or exceeding 0.95, whereas Figure 19b demonstrates notable RMSE reductions, particularly for XGBoost and gradient boosting. Figure 19c shows that XGBoost and Gradient boosting achieve the lowest MAE, while MLP records the highest, and Figure 19d indicates that XGBoost attains the lowest MAPE, whereas MLP exhibits the highest. Figure 19e highlights consistently high EVS scores across boosting methods, with XGBoost demonstrating superior performance in both training and testing. Compared to Figure 17, the inclusion of Cd, Cr, and Cu significantly enhances predictive accuracy, with MLP remaining weakest yet improved, CatBoost showing stable generalization, random forest reducing MAE, gradient boosting excelling in EVS and RMSE, and XGBoost achieving near-perfect performance, confirming the pivotal role of these metals in strengthening predictions.

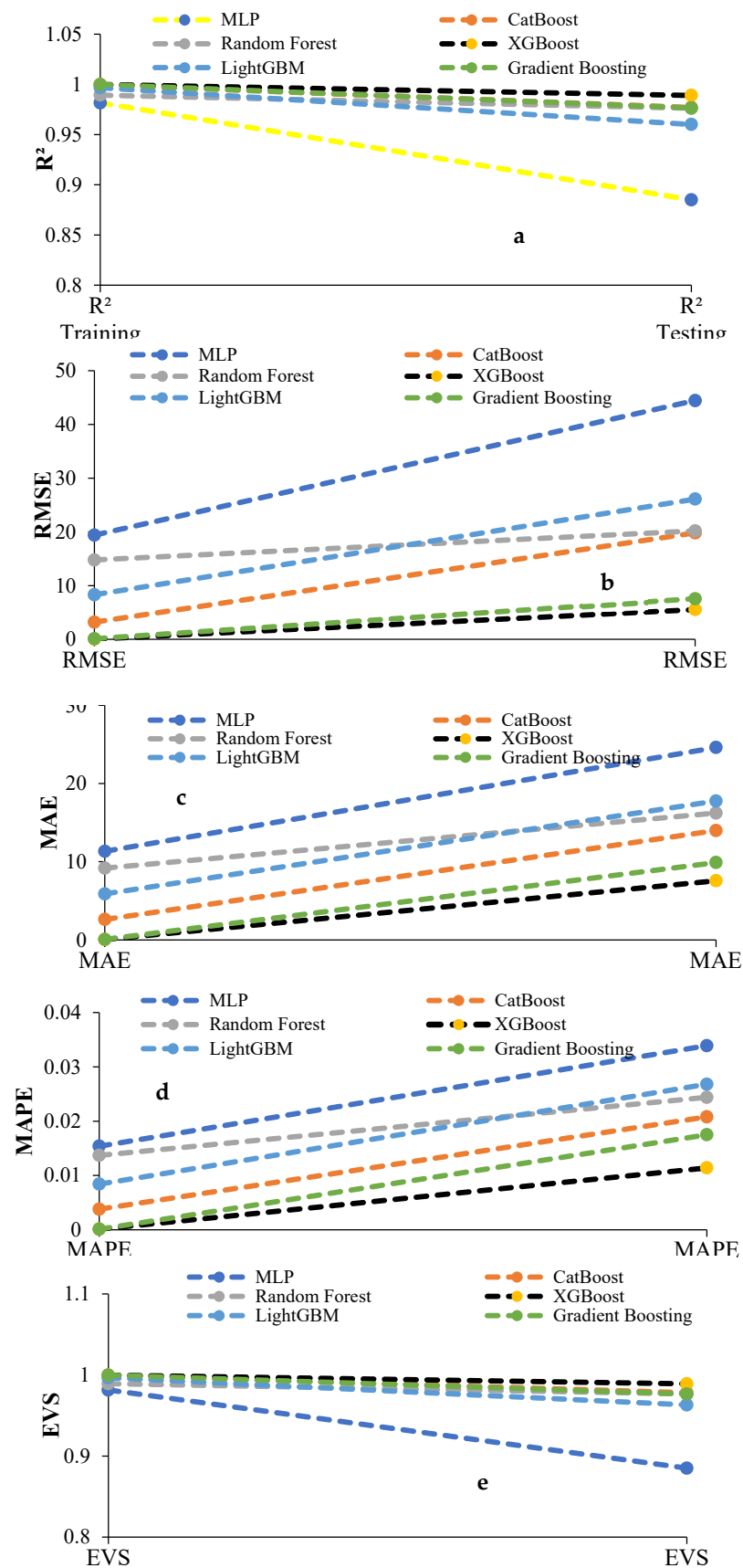


Figure 19. Comparative evaluation of model performance during training and testing phases using R^2 (a), RMSE (b), MAE (c), MAPE (d), and EVS (e).

It should be noted that Table 7 presents the optimized hyperparameters for all models, whereas Tables 8 and 9 provide a comparative summary of model performance without and with Cd, Cr, and Cu as input parameters, respectively.

Figure 20 illustrates the relationship between actual and predicted elevation values, highlighting model performance under different input feature scenarios. In the absence of Cd, Cr, and Cu (Table 8), the MLP model exhibited moderate dispersion, with a testing performance of $R^2 = 0.8725$, whereas CatBoost and XGBoost demonstrated relatively strong predictive capability ($R^2 = 0.9340$ and 0.9526 , respectively). The random forest model achieved $R^2 = 0.9167$, while LightGBM and gradient boosting showed comparatively moderate performance, with R^2 values of 0.8808 and 0.9093 , respectively. With the inclusion of Cd, Cr, and Cu (Table 9), a substantial improvement in predictive accuracy was observed across all models. The MLP model showed a slight increase in performance ($R^2 = 0.8850$), although it remained the least accurate. CatBoost exhibited a marked enhancement, achieving $R^2 = 0.9770$ with reduced RMSE and MAE, while random forest improved to $R^2 = 0.9762$. XGBoost demonstrated the highest predictive performance, achieving near-perfect agreement between predicted and observed values ($R^2 = 0.9890$, RMSE = 5.5462). LightGBM also showed considerable improvement ($R^2 = 0.9601$), and gradient boosting reached $R^2 = 0.9765$. Overall, although all models exhibited strong training performance, the inclusion of Cd, Cr, and Cu significantly reduced prediction errors and enhanced model generalization. The regression fits approached a unity slope with minimal intercept deviation, with CatBoost and XGBoost exhibiting the lowest bias. These findings underscore the critical role of key heavy metal variables in improving predictive reliability, with XGBoost demonstrating the highest accuracy and consistency in capturing the relationship between predicted and actual elevation values.

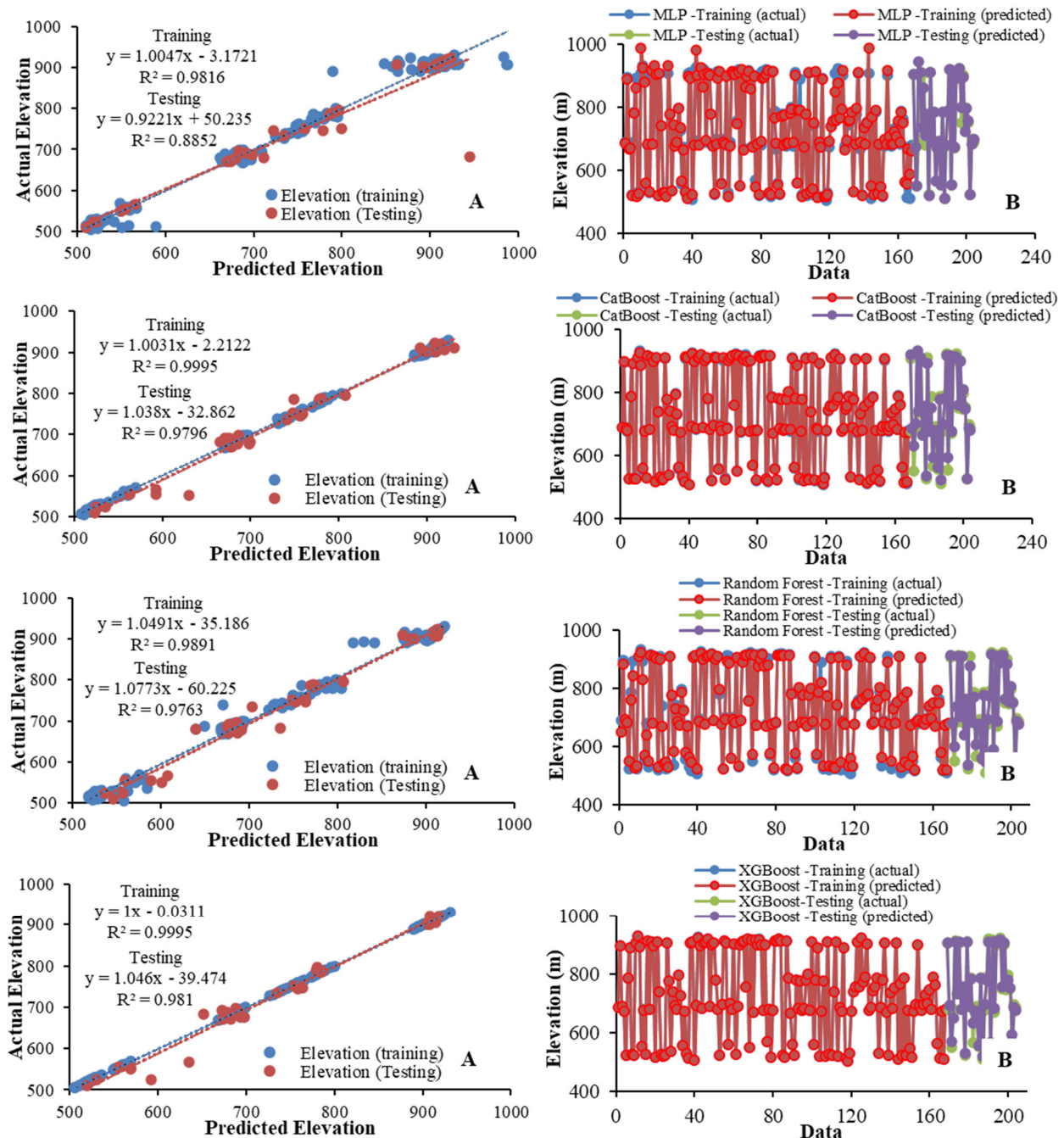
Figure 21 presents the residual analysis of predicted versus experimental elevation values obtained using MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting models.

Figure 21 presents residual plots of predicted versus experimental elevation values for all models. MLP residuals exhibit wide scatter, indicating lower predictive accuracy. CatBoost residuals are compact and symmetrically distributed around zero, reflecting strong generalization, whereas random forest shows moderate dispersion with a slight tendency to overestimate. XGBoost achieves the most balanced residual distribution, demonstrating minimal bias and variance, while LightGBM shows improved residuals compared to MLP but remains more dispersed than CatBoost. Gradient boosting residuals are narrow, confirming robust predictive performance. Across all models, training residuals are consistently smaller than testing residuals, highlighting differences in model generalization. The residual distributions underscore the bias–variance tradeoff, with boosting algorithms clearly outperforming MLP and confirming the superior reliability of ensemble methods for elevation prediction.

Table 10 presents the comparative performance of traditional models and XGBoost in predicting elevation. The results indicate that traditional linear models exhibit relatively low predictive accuracy, reflecting their limited ability to capture nonlinear relationships. In contrast, XGBoost achieved substantially superior performance, with an $R^2 = 0.9890$ (RMSE = 5.5462, MAE = 7.5644, and MAPE = 0.0114) in testing, demonstrating its effectiveness in modeling complex environmental patterns.

Table 10. Performance comparison of traditional models and XGBoost for predicting elevation.

Model	Training				Testing			
	R ²	RMSE	MAE	MAPE	R ²	RMSE	MAE	MAPE
Linear Regression	0.5806	92.182	74.257	11.081	0.5691	91.688	73.643	11.009
Ridge	0.5684	93.510	74.915	11.243	0.5589	92.774	73.218	11.080
Lasso	0.5806	92.182	74.232	11.079	0.5691	91.694	73.646	11.011
ElasticNet	0.5689	93.454	74.871	11.235	0.5592	92.742	73.233	11.079
Decision Tree	0.9604	28.330	14.368	2.240	0.8643	51.447	29.965	4.743
XGBoost	0.9999	0.0535	0.0512	0.0001	0.9890	5.5462	7.5644	0.0114

**Figure 20.** Cont.

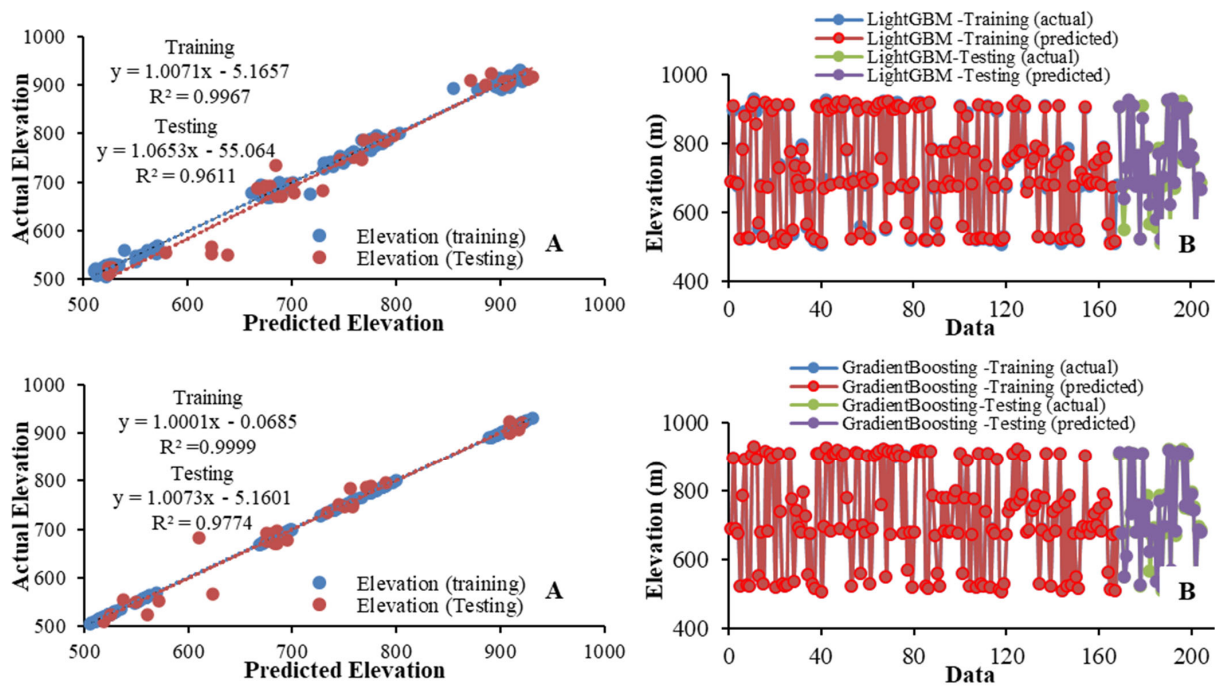


Figure 20. Comparison of experimental and predicted elevation values using MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting.

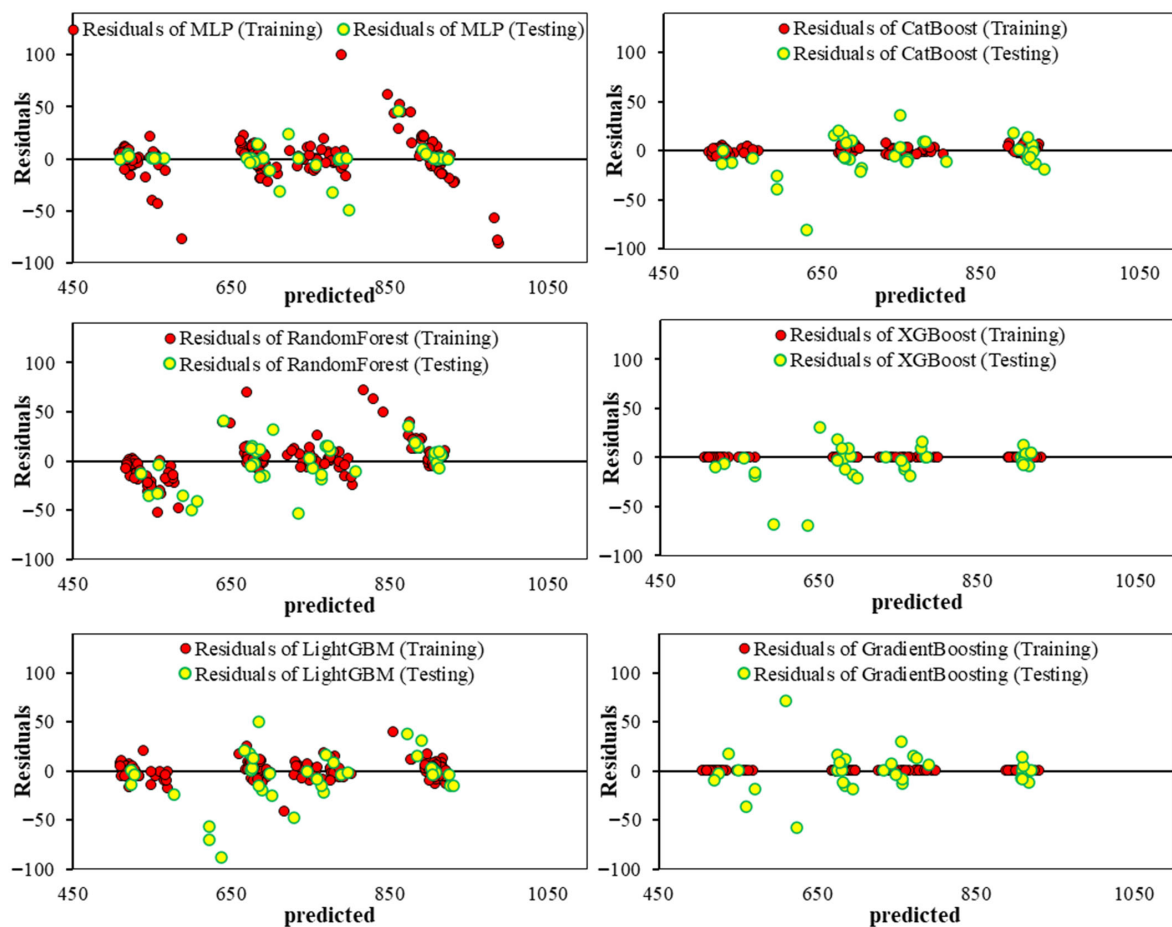


Figure 21. Residual analysis comparing predicted and experimental elevation values for MLP, CatBoost, random forest, XGBoost, LightGBM, and gradient boosting models.

Discussion and Comparison

The findings of this study revealed that heavy metal accumulation in *Pinus nigra* needles varied significantly across seasons and locations in Adıyaman. Industrial zones and traffic corridors exhibited the highest concentrations of Cr, Ni, Fe, and Pb, whereas the recreational Seyirtepe site showed comparatively lower levels. These spatial patterns strongly suggest that vehicular emissions, fossil fuel combustion, industrial processing activities, and road dust resuspension constitute the primary sources of metal deposition in the study area. In particular, elevated Cr and Ni concentrations may be associated with fuel combustion and metallurgical activities, while Pb enrichment reflects continued traffic-related inputs despite regulatory controls. Seasonal variability further indicates that atmospheric transport and deposition processes play a critical role in metal accumulation. Higher autumn–winter concentrations result from increased emissions and reduced dispersion, while lower summer levels reflect enhanced mixing and deposition removal, confirming pine needles as reliable indicators of temporal pollution dynamics. Beyond descriptive statistics, the results reveal clear spatial and seasonal patterns in heavy metal accumulation. Industrial and high-traffic areas exhibited consistently higher metal concentrations compared to Seyirtepe, indicating the dominant influence of anthropogenic emission sources. Seasonally, concentrations increase in autumn–winter due to higher emissions and limited dispersion, and decrease in summer due to wash-off and vegetative dilution.

From an ecological perspective, the exceedance of WHO/FAO permissible limits for Cr, Ni, Fe, and Pb at several sites raises concerns regarding potential phytotoxic effects, soil contamination feedback, and trophic transfer within local ecosystems. Persistent accumulation of these metals may impair photosynthetic efficiency, alter nutrient cycling, and pose indirect risks to herbivorous organisms and, ultimately, human populations through bioaccumulation pathways. Therefore, the observed contamination patterns are not only indicators of air quality degradation but also signals of broader ecosystem vulnerability. Compared with previous national and international studies, the spatial and seasonal distribution patterns observed in Adıyaman align with global trends where traffic density, industrial emissions, and meteorological conditions govern heavy metal deposition. However, this study extends prior work by quantitatively integrating biomonitoring data with advanced machine learning (ML) modeling to enhance predictive capability. While studies such as those by Soylak and Türkoğlu [22], Arslan, and Yılmaz and Zengin [70] primarily focused on concentration assessment, the present research advances the field by linking environmental variables with predictive analytics to support proactive monitoring strategies. The integration of ML, particularly the XGBoost algorithm, significantly enhanced the predictive capability of this study by capturing complex nonlinear interactions among environmental variables that cannot be adequately represented by traditional statistical methods. As demonstrated in Table 9, ML models clearly outperformed conventional approaches, confirming their superior ability to model heterogeneous environmental systems. The high R^2 value (0.9890) obtained when all metals were included highlights the robustness and reliability of the proposed modeling framework. Beyond accuracy improvement, the significance of ML lies in its ability to provide predictive insights rather than solely descriptive analysis. By integrating spatial, seasonal, and environmental variables, the model enables a more comprehensive understanding of pollution dynamics and supports proactive environmental management. These predictive capabilities can assist in early warning systems, optimization of monitoring strategies, and development of targeted pollution mitigation policies in rapidly urbanizing regions.

4. Conclusions

Pinus nigra needles effectively indicate atmospheric heavy metal deposition, with elevated Cr, Ni, Fe, and Pb near industrial and traffic-influenced areas and peak accumulation in autumn–winter due to increased emissions and limited dispersion. Exceedances of permissible limits highlight potential ecological risks and the need for continued monitoring and emission control. ML enhanced the analysis by capturing complex environmental interactions; however, the primary contribution lies in integrating predictive modeling with ecological assessment. Despite strong predictive performance, the relatively limited dataset size ($n = 140$) may affect generalization. To mitigate overfitting, data were split into training and testing subsets (80/20), with preprocessing, hyperparameter optimization, and multi-metric evaluation applied, supported by residual and validation analyses. Nevertheless, limitations include restricted geographic coverage, absence of source apportionment, and unaccounted environmental variables. Future studies should incorporate larger datasets and advanced source-tracing approaches to improve robustness and applicability.

The following conclusions are substantiated by the results of the study:

1. Cadmium concentrations remained consistently low across all sites, suggesting limited atmospheric deposition, whereas Al, Fe, and Mn reflected mixed geogenic and anthropogenic influences, and elevated metal levels were primarily associated with industrial and traffic emissions modulated by seasonal meteorological conditions.
2. *Pinus nigra* needles proved to be practical and sensitive bioindicators for assessing trace metal accumulation in mixed urban–industrial environments, with maximum accumulation occurring in autumn and winter, likely associated with reduced atmospheric dispersion and increased heating-related emissions.
3. The findings highlight the urgent need for stricter regulation of industrial and traffic-related emissions to reduce environmental and public health risks, and emphasize the importance of establishing long-term monitoring programs to assess temporal trends and the effectiveness of pollution control strategies.
4. Future research should integrate isotopic source apportionment and comparative biomonitoring with other plant species to better identify pollution sources while assessing human exposure pathways and applying health risk models to strengthen policy relevance.
5. Excluding Cd, Cr, and Cu reduced model performance, with XGBoost achieving the highest test $R^2 = 0.953$ and MLP the lowest ($R^2 = 0.873$), whereas their inclusion significantly improved predictive accuracy, with XGBoost reaching $R^2 = 0.989$ and strong performance also observed for CatBoost and gradient boosting.
6. Ensemble models outperformed MLP with lower errors and better generalization, and the inclusion of Cd, Cr, and Cu further improved predictive accuracy, confirming the effectiveness of ML for environmental modeling and biomonitoring.
7. The study's limited geographic scope and absence of isotopic analysis may constrain generalizability and precise source identification, highlighting the need for larger datasets and additional variables in future research to enhance model accuracy.
8. Future studies should expand data diversity, integrate additional variables and isotopic analysis, and develop real-time systems while addressing ML limitations such as data dependency, generalization, and interpretability through improved data handling and model updating.

Author Contributions: Conceptualization, S.J.; Methodology, S.J.; Formal analysis, S.J.; Investigation, S.J.; Writing—original draft, S.J.; Writing—review & editing, T.D. and Z.Y.; Supervision, T.D. and Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors acknowledge that no additional administrative, technical, or material support was received for this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

Al	Aluminum
As	Arsenic
Cd	Cadmium
Cr	Chromium
Cu	Copper
Fe	Iron
HF	Hydrofluoric Acid
HNO ₃	Nitric Acid
Mn	Manganese
Ni	Nickel
Pb	Lead
Zn	Zinc
PVC	Polyvinyl Chloride
WHO	World Health Organization
FAO	Food and Agriculture Organization
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
RF	Radio Frequency
DAP	Diammonium Phosphate
ppm	Parts Per Million

References

1. Collin, M.S.; Venkatraman, S.K.; Vijayakumar, N.; Kanimozhi, V.; Arbaaz, S.M.; Stacey, R.S.; Anusha, J.; Choudhary, R.; Lvov, V.; Tovar, G.I. Bioaccumulation of lead (Pb) and its effects on human: A review. *J. Hazard. Mater. Adv.* **2022**, *7*, 100094. [[CrossRef](#)]
2. Agarwal, A.; Khanna, P.; Baidya, D.K.; Arora, M.K. Trace elements in critical illness. *J. Endocrinol. Metab.* **2011**, *1*, 57–63. [[CrossRef](#)]
3. Loganathan, B.G.; Kumar, K.S.; Seaford, K.D.; Sajwan, K.S.; Hanari, N.; Yamashita, N. Distribution of persistent organohalogen compounds in pine needles from selected locations in Kentucky and Georgia, USA. *Arch. Environ. Contam. Toxicol.* **2008**, *54*, 422–439. [[CrossRef](#)]
4. Wen, S.; Yang, F.; Li, J.; Gong, Y.; Zhang, X.; Hui, Y.; Wu, Y.; Zhao, Y.; Xu, Y. Polychlorinated dibenzo-p-dioxin and dibenzofurans (PCDD/Fs), polybrominated diphenyl ethers (PBDEs), and polychlorinated biphenyls (PCBs) monitored by tree bark in an E-waste recycling area. *Chemosphere* **2009**, *74*, 981–987. [[CrossRef](#)]
5. Augusto, S.; Máguas, C.; Matos, J.; Pereira, M.J.; Branquinho, C. Lichens as an integrating tool for monitoring PAH atmospheric deposition: A comparison with soil, air and pine needles. *Environ. Pollut.* **2010**, *158*, 483–489. [[CrossRef](#)]
6. Ratola, N.; Alves, A.; Santos, L.; Lacorte, S. Pine needles as passive bio-samplers to determine polybrominated diphenyl ethers. *Chemosphere* **2011**, *85*, 247–252. [[CrossRef](#)] [[PubMed](#)]
7. Guéguen, F.; Stille, P.; Millet, M. Air quality assessment by tree bark biomonitoring in urban, industrial and rural environments of the Rhine Valley: PCDD/Fs, PCBs and trace metal evidence. *Chemosphere* **2011**, *85*, 195–202. [[CrossRef](#)]
8. Choi SungDeuk, C.S. Time trends in the levels and patterns of polycyclic aromatic hydrocarbons (PAHs) in pine bark, litter, and soil after a forest fire. *Sci. Total Environ.* **2014**, *470–471*, 1441–1449.
9. Bellis, D.J.; Satake, K.; Kagawa, A. Chronological trends in trace metals recorded by a tree bark pocket in Yakushima Island, Japan. *J. Environ. Monit.* **2005**, *7*, 384–387. [[CrossRef](#)]
10. Mandiwana, K.L.; Resane, T.; Panichev, N.; Ngobeni, P. The application of tree bark as bio-indicator for the assessment of Cr (VI) in air pollution. *J. Hazard. Mater.* **2006**, *137*, 1241–1245. [[CrossRef](#)] [[PubMed](#)]
11. Al-Alawi, M.t.M.; Batarseh, M.I.; Carreras, H.; Alawi, M.; Jiries, A.; Charlesworth, S.M. Aleppo pine bark as a biomonitor of atmospheric pollution in the arid environment of Jordan. *CLEAN—Soil Air Water* **2007**, *35*, 438–443. [[CrossRef](#)]
12. Berlizov, A.; Blum, O.; Filby, R.; Malyuk, I.; Tryshyn, V. Testing applicability of black poplar (*Populus nigra* L.) bark to heavy metal air pollution monitoring in urban and industrial regions. *Sci. Total Environ.* **2007**, *372*, 693–706. [[CrossRef](#)] [[PubMed](#)]

13. Dogan, Y.; Durkan, N.; Baslar, S. Trace element pollution biomonitoring using the bark of *Pinus brutia* (Turkish red pine) in the Western Anatolian part of Turkey. *Trace Elem. Electrolytes* **2007**, *24*, 146. [\[CrossRef\]](#)
14. Conkova, M.; Kubiznakova, J. Lead isotope ratios in tree bark pockets: An indicator of past air pollution in the Czech Republic. *Sci. Total Environ.* **2008**, *404*, 440–445. [\[CrossRef\]](#)
15. Schelle, E.; Rawlins, B.; Lark, R.; Webster, R.; Staton, I.; McLeod, C. Mapping aerial metal deposition in metropolitan areas from tree bark: A case study in Sheffield, England. *Environ. Pollut.* **2008**, *155*, 164–173. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Catinon, M.; Ayrault, S.; Clocchiatti, R.; Boudouma, O.; Asta, J.; Tissut, M.; Ravel, P. The anthropogenic atmospheric elements fraction: A new interpretation of elemental deposits on tree barks. *Atmos. Environ.* **2009**, *43*, 1124–1130. [\[CrossRef\]](#)
17. Aksoy, A.; Hale, W.H.; Dixon, J.M. *Capsella bursa-pastoris* (L.) Medic. as a biomonitor of heavy metals. *Sci. Total Environ.* **1999**, *226*, 177–186. [\[CrossRef\]](#)
18. Yilmaz, S.; Zengin, M. Monitoring environmental pollution in Erzurum by chemical analysis of Scots pine (*Pinus sylvestris* L.) needles. *Environ. Int.* **2004**, *29*, 1041–1047. [\[CrossRef\]](#)
19. Celik, A.; Kartal, A.A.; Akdoğan, A.; Kaska, Y. Determining the heavy metal pollution in Denizli (Turkey) by using *Robinia pseudo-acacia* L. *Environ. Int.* **2005**, *31*, 105–112. [\[CrossRef\]](#)
20. Aksoy, A.; Şahin, U.; Duman, F. *Robinia pseudo-acacia* L. as a possible biomonitor of heavy metal pollution in Kayseri. *Turk. J. Bot.* **2000**, *24*, 279–284.
21. Djingova, R.; Kuleff, I. Monitoring of heavy metal pollution by *Taraxacum officinale*. In *Plants as Biomonitor: Indicators for Heavy Metals in the Terrestrial Environment*; Markert, B., Ed.; VCH: Weinheim, Germany, 1993; pp. 435–460.
22. Soylak, M.; Türkoğlu, O. Trace metal accumulation caused by traffic in an agricultural soil near a motorway in Kayseri, Turkey. *J. Trace Microprobe Tech.* **1999**, *17*, 209–217.
23. Arslan, H. Heavy metals in street dust in Bursa, Turkey. *J. Trace Microprobe Tech.* **2001**, *19*, 439–445. [\[CrossRef\]](#)
24. Köleli, N. Determination of heavy metal concentrations in open and protected cropping systems. *Fresenius Environ. Bull.* **2004**, *13*, 1521–1524.
25. Kiliel, F. Heavy metal contamination in soils of garbage area in Van, Turkey. *Asian J. Chem.* **2006**, *18*, 461–468.
26. Badamasi, H. Urban roadside trees as eco-sustainable filters of atmospheric pollution: A review of recent evidence from atmospheric trace elements deposition. In *New Paradigms in Environmental Biomonitoring Using Plants*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 73–94.
27. Yang, D.; Ye, C.; Xu, J. Land-use change and health risks in the process of urbanization: A spatiotemporal interpretation of a typical case in Changzhou, China. *Land* **2021**, *10*, 820. [\[CrossRef\]](#)
28. Jueyendah, S.; Ağcakoca, E. An Interpretable Ensemble Machine Learning Framework for Predicting the Ultimate Flexural Capacity of BFRP-Reinforced Concrete Beams. *Polymers* **2026**, *18*, 601. [\[CrossRef\]](#)
29. Jueyendah, S.; Holly, I. Integrating Explainable Hybrid Machine Learning and Evolutionary Optimization for High-Fidelity Prediction of Ultra-High-Performance Concrete Compressive Strength. *Array* **2026**, *30*, 100774. [\[CrossRef\]](#)
30. Xu, Y.; Liu, X.; Cao, X.; Huang, C.; Liu, E.; Qian, S.; Liu, X.; Wu, Y.; Dong, F.; Qiu, C.-W.J.T.I. Artificial intelligence: A powerful paradigm for scientific research. *Innovation* **2021**, *2*, 100179.
31. Jueyendah, S.; Ağcakoca, E.; Yaman, Z. An interpretable hybrid machine learning and deep learning framework for predicting compressive strength of recycled powder mortar. *Civ. Eng. Beyond Limits* **2026**, *7*, 11118. [\[CrossRef\]](#)
32. Ghanbari, S.; Shahmansouri, A.A.; Akbarzadeh Bengar, H.; Jafari, A. Compressive strength prediction of high-strength oil palm shell lightweight aggregate concrete using machine learning methods. *Environ. Sci. Pollut. Res.* **2023**, *30*, 1096–1115. [\[CrossRef\]](#)
33. Jueyendah, S.; Lezgy-Nazargah, M.; Eskandari-Naddaf, H.; Emamian, S.J.C.; Materials, B. Predicting the mechanical properties of cement mortar using the support vector machine approach. *Constr. Build. Mater.* **2021**, *291*, 123396. [\[CrossRef\]](#)
34. Jueyendah, S.; Martins, C.H.J.C.E.; Modeling, P. Optimal Design of Welded Structure Using SVM. *Comput. Eng. Phys. Model.* **2024**, *7*, 84–107. [\[CrossRef\]](#)
35. Zhu, J.-J.; Yang, M.; Ren, Z.J. Machine learning in environmental research: Common pitfalls and best practices. *Environ. Sci. Technol.* **2023**, *57*, 17671–17689. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Singh, M.; Singh, M.; Singh, J.; Singh, H.; Singh, M. Application of Machine Learning Techniques for Environmental Monitoring and Conservation: A Review. In *Digitization and Manufacturing Performance: An Environmental Perspective*; Wiley-Scrivener Publishing LLC: Salem, MA, USA, 2026; pp. 97–127.
37. Cojbasic, S.; Dmitrasinovic, S.; Kostic, M.; Turk Sekulic, M.; Radonic, J.; Dodig, A.; Stojkovic, M. Application of machine learning in river water quality management: A review. *Water Sci. Technol.* **2023**, *88*, 2297–2308. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Cui, S.; Gao, Y.; Huang, Y.; Shen, L.; Zhao, Q.; Pan, Y.; Zhuang, S. Advances and applications of machine learning and deep learning in environmental ecology and health. *Environ. Pollut.* **2023**, *335*, 122358. [\[CrossRef\]](#)
39. Konya, A.; Nematzadeh, P. Recent applications of AI to environmental disciplines: A review. *Sci. Total Environ.* **2024**, *906*, 167705. [\[CrossRef\]](#)

40. Sirimewan, D.; Bazli, M.; Raman, S.; Mohandes, S.R.; Kineber, A.F.; Arashpour, M. Deep learning-based models for environmental management: Recognizing construction, renovation, and demolition waste in-the-wild. *J. Environ. Manag.* **2024**, *351*, 119908. [\[CrossRef\]](#)
41. Fleming, S.W.; Watson, J.R.; Ellenson, A.; Cannon, A.J.; Vesselinov, V.C. Machine learning in Earth and environmental science requires education and research policy reforms. *Nat. Geosci.* **2021**, *14*, 878–880. [\[CrossRef\]](#)
42. Marabelli, M.; Davison, R.M. Artificial intelligence and the environment: Ethical challenges and strategic opportunities for organizations. *J. Strateg. Inf. Syst.* **2025**, *34*, 101921. [\[CrossRef\]](#)
43. Ajayi, B.O.; Bamişaye, M.E.; Macedo, H.M.M.; Sereewatthanawut, I. Integrating system dynamics and machine learning for environmental impact analysis of building materials in the demolition process. *Sci. Rep.* **2025**, *15*, 27114. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Zhou, S.; Mou, L.; Hua, Y.; Zhang, L.; Kaufmann, H.; Zhu, X.X. Can we use deep learning models to identify the functionality of plastics from space? *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *123*, 103491. [\[CrossRef\]](#)
45. Wang, T.; Zhang, M.; Li, Z. Explainable machine learning links erosion damage to environmental factors on Gansu rammed earth Great Wall. *npj Herit. Sci.* **2025**, *13*, 366. [\[CrossRef\]](#)
46. Zhao, W.; Chen, H.; Yazan, D.M.; Taghavifar, H.; Lyu, Y.; Bulis, A. Few-shot learning and deep predictive models for cost optimization and carbon emission reduction in energy-water management. *J. Environ. Manag.* **2025**, *389*, 126077. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Singh, G.; Moncrieff, G.; Venter, Z.; Cawse-Nicholson, K.; Slingsby, J.; Robinson, T.B. Uncertainty quantification for probabilistic machine learning in earth observation using conformal prediction. *Sci. Rep.* **2024**, *14*, 16166. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Wang, Y.; Okochi, H.; Igawa, M. Long term trends of dry and wet deposition of air pollutants at declining forest site of Mt. Oyama in Japan during 1994–2019. *Water Air Soil Pollut.* **2023**, *234*, 257. [\[CrossRef\]](#)
49. Igawa, M.; Kojima, K.; Yoshimoto, O.; Nanzai, B. Air pollutant deposition at declining forest sites of the Tanzawa Mountains, Japan. *Atmos. Res.* **2015**, *151*, 93–100. [\[CrossRef\]](#)
50. Dere, T.; Çekim, M. *Ağır Metal—Karaçam—Temiz Hava*; Güven Plus Grup A.Ş.: Basaksehir, Turkey, 2022.
51. Jueyendah, S.; Yaman, Z.; Dere, T.; Çavuş, T.F. Comparative Study of Linear and Non-Linear ML Algorithms for Cement Mortar Strength Estimation. *Buildings* **2025**, *15*, 2932. [\[CrossRef\]](#)
52. Ağcakoca, E.; Jueyendah, S.; Yaman, Z.; Sümer, Y.; Maali, M. Advanced Hybrid Modeling of Cementitious Composites Using Machine Learning and Finite Element Analysis Based on the CDP Model. *Buildings* **2025**, *15*, 3026. [\[CrossRef\]](#)
53. Grigorov, D. Harnessing Python 3.11 and Python Libraries for LLM Development. In *Introduction to Python and Large Language Models: A Guide to Language Models*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 303–368.
54. Menke, W. *Environmental Data Analysis with MatLab or Python: Principles, Applications, and Prospects*; Academic Press: Cambridge, MA, USA, 2022.
55. Popescu, M.-C.; Balas, V.E.; Perescu-Popescu, L.; Mastorakis, N. Multilayer perceptron and neural networks. *WSEAS Trans. Circuits Syst.* **2009**, *8*, 579–588.
56. El Badaoui, H.; Abdallaoui, A.; Manssouri, I.; Lancelot, L. Application of the artificial neural networks of MLP type for the prediction of the levels of heavy metals in Moroccan aquatic sediments. *Int. J. Comput. Eng. Res.* **2013**, *3*, 75–81.
57. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for big data: An interdisciplinary review. *J. Big Data* **2020**, *7*, 94. [\[CrossRef\]](#)
58. Xing, F.; Li, H.; Li, T. Deformation Modeling and Prediction of Concrete Dam Using Observed Air Temperature and Enhanced CatBoost Algorithm. *Water* **2024**, *16*, 3341. [\[CrossRef\]](#)
59. Rigatti, S.J. Random forest. *J. Insur. Med.* **2017**, *47*, 31–39. [\[CrossRef\]](#)
60. Liu, X.; Song, Y.; Yi, W.; Wang, X.; Zhu, J. Comparing the random forest with the generalized additive model to evaluate the impacts of outdoor ambient environmental factors on scaffolding construction productivity. *J. Constr. Eng. Manag.* **2018**, *144*, 04018037. [\[CrossRef\]](#)
61. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM Digital Library: New York, NY, USA, 2016; pp. 785–794.
62. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2017**, *30*, 3146–3154.
63. Daniel, C. A robust LightGBM model for concrete tensile strength forecast to aid in resilience-based structure strategies. *Heliyon* **2024**, *10*, e39679. [\[CrossRef\]](#)
64. Wang, P.; Zhou, X.; Zeng, Z. LightGBM-BES-BiLSTM carbon price prediction based on environmental impact factors. *Comput. Econ.* **2025**, *65*, 2891–2917. [\[CrossRef\]](#)
65. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [\[CrossRef\]](#)
66. Lee, S.; Park, J.; Kim, N.; Lee, T.; Quagliato, L. Extreme gradient boosting-inspired process optimization algorithm for manufacturing engineering applications. *Mater. Des.* **2023**, *226*, 111625. [\[CrossRef\]](#)

67. Plevris, V.; Solorzano, G.; Bakas, N.P.; Ben Seghier, M.E.A. Investigation of performance metrics in regression analysis and machine learning-based prediction models. In Proceedings of the 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2022), Oslo, Norway, 5–9 June 2022.
68. Özkan, A. Antakya-Cilvegözü karayolu etrafındaki tarım arazilerinde ve bitkilerdeki ağır metal kirliliği. *Çukurova Üniversitesi Mühendislik-Mimar. Fakültesi Derg.* **2017**, *32*, 9–18. [[CrossRef](#)]
69. Falciani, R.; Novaro, E.; Marchesini, M.; Gucciardi, M. Multi-element analysis of soil and sediment by ICP-MS after a microwave assisted digestion method. *J. Anal. At. Spectrom.* **2000**, *15*, 561–565. [[CrossRef](#)]
70. Arslan, H. Investigation of heavy metal pollution of traffic in Kemalpasa-Turkey. *Fresenius Environ. Bull.* **2001**, *10*, 405–408.

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.