Earth Science Informatics (2026) 19:30
https://doi.org/10.1007/s12145-026-02094-x

RESEARCH

A machine learning framework to lanthanide element distribution and predictability from the northwestern Iranian karst bauxite deposits

Ali Abedini[1] · Maryam Khosravi[2] · Ouafi Ameur-Zaimeche[3] · Rabah Kechiched[3] · Abdelhamid Ouladmansour[4] · Giovanni Mongelli[5]

Received: 31 May 2025 / Accepted: 21 February 2026

Abstract

This study applies Extreme Gradient Boosting (XGBoost) to examine how geological age clustering affects the predictability of light rare earth elements (LREEs: La–Sm) and heavy rare earth elements (HREEs: Eu–Lu) in the northwestern Iranian karst bauxites. Major oxides and weathering indices (CIA and CIW) were used as predictors, revealing contrasting behaviors between the Paleozoic and Mesozoic deposits. LREEs exhibited high accuracy in the Mesozoic deposits, whereas HREEs achieved the best predictability in the Paleozoic samples, reflecting distinct mineralogical controls. Variability in cerium, linked to paleoredox conditions, reduces LREEs model performance in the Paleozoic, while phosphate phases strongly influence HREEs enrichment. Feature importance consistently identifies P_2O_5 as the dominant predictor for both LREEs and HREEs across all deposits, highlighting the key role of phosphate minerals in lanthanide incorporation, whereas major oxides contribute less. CIA and CIW further enhance predictive accuracy, indicating that subtle variations in paleoweathering conditions affect lanthanide distribution and model performance. The LREEs and HREEs prediction models demonstrate promising potential, and further cross-validation across global karst bauxite deposits could improve understanding of the factors controlling REEs distribution, ultimately supporting more efficient and cost-effective exploration strategies for these critical metals.

Keywords Northwestern Iranian karst bauxite deposits · XGBoost · Cerium · P_2O_5 · LREEs and HREEs predictability

Introduction

Rare earth elements (REEs), spanning atomic numbers 57 to 71 from lanthanum to lutetium, constitute a group of metals with remarkably similar chemical and physical properties. REEs are classified as critical raw materials in the latest Study on Critical Raw Materials for the EU (Grohol and Veeh 2023) and as critical minerals in both the 2022 U.S. Geological Survey report and the China's 2023 Critical Minerals List (Li et al. 2023). These elements play an essential role in advanced technologies and/or materials, including metallurgical processes, permanent magnets for electric motors and generators, lighting phosphors, catalysts, batteries, fiber optics, optical amplifiers, lasers, glass, and ceramics. China overwhelmingly dominates global REEs production, supplying 85% of light REEs (LREEs: La–Sm) and 100% of heavy REEs (HREEs: Eu–Lu+Y) (Grohol and Veeh 2023). The global REEs market is experiencing rapid growth, with a market value reaching $7.29 billion in 2023 and a compound annual growth rate of 10.8%. Despite their

Communicated by Hassan Babaie

✉ Abdelhamid Ouladmansour
abdelhamid.ouladmansour@uniba.it

[1] Department of Geology, Faculty of Sciences, Urmia University, 57561-51818 Urmia, Iran

[2] Department of Mining Engineering, Isfahan University of Technology, Isfahan 8415683111, Iran

[3] Laboratoire des Réservoirs Souterrains: Pétroliers, Gaziers et Aquifères, Université Kasdi Merbah Ouargla, Ouargla 30000, Algeria

[4] Dipartimento di Scienze della Terra e Geoambientali, Università degli Studidi Bari Aldo Moro, via Orabona 4, Bari 70125, Italy

[5] Department of Applied and Basic Sciences, University of Basilicata, Viale dell'Ateneo Lucano 10, 85100 Potenza, Italy

Published online: 02 March 2026

relatively high total abundance in the Upper Continental Crust (146.4 ppm, McLennan et al. 2006), REEs primarily occur as substituent in common rock-forming minerals, rather than forming their own distinct mineral phases (Kesler et al. 2015). Economically significant REE-bearing minerals include bastnäsite, xenotime, monazite, synchysite, eudialyte, loparite, and parisite, along with ion-adsorption clay deposits (Castor 2008; Yanfei et al. 2016; Wang et al. 2017; Mongelli et al. 2018; Ahmadnejad and Mongelli 2022).

The growing demand for REEs and the geographically restricted production of these resources have spurred heightened interest in evaluating their distribution and fractionation models from alternative sources, such as sedimentary deposits. Among these, karst bauxites are particularly promising, due to their potential for concentrating metals that are critical to modern technologies (e.g., Liu et al. 2016, 2020; Mongelli et al. 2016, 2017, 2021; Khosravi et al. 2017, 2025; Gamaletsos et al. 2019; Salamab-Ellahi et al. 2019; Abedini et al. 2021, 2024; Ahmadnejad and Mongelli 2022; Sofis et al. 2025).

Over the past decade, extensive research has been conducted on the distribution and geochemical fractionation processes that influence critical metals, particularly REEs, in the Iranian karst bauxites (Khosravi et al. 2017, 2025; Salamab-Ellahi et al. 2019; Abedini et al. 2020a, b, 2021, 2022a, b, 2024; Ahmadnejad and Mongelli 2022). These deposits are part of the Irano–Himalayan karst bauxite belt and exhibit similarities to the Mediterranean-type deposits. The karst bauxite deposits in northwestern Iran primarily formed through long-term weathering of a mafic protolith and are predominantly composed of diaspore and/or boehmite, with no gibbsite occurrence (Khosravi et al. 2017; Abedini and Khosravi 2020; Abedini et al. 2020b, 2021, 2022b, 2024).

Machine learning (ML) algorithms have been applied in geosciences since the 1950s (Dramsch 2020). In recent decades, ML applications have made significant advancements, surpassing traditional methods in their ability to handle multiple geological variables, and thus model complex nonlinear interactions among geological parameters (e.g., Cracknell and Reading 2014; Saporetti et al. 2018; Dumakor-Dupey and Arya 2021; Zuo et al. 2021). Iron (Zhang et al. 2023), copper (Esmaeiloghli et al. 2023), and gold (Silva dos Santos et al. 2022) deposits are among the most frequently analyzed ore deposits using ML techniques. Recently, Buccione et al. (2024) successfully predicted the distribution of HREEs (Eu–Lu) in the Cretaceous Mediterranean-type karst bauxite deposits of southern Italy. They used Al_2O_3, Fe_2O_3, TiO_2, and SiO_2 as input variables with an Extreme Gradient Boosting (XGBoost) model, achieving high accuracy in predicting HREEs ($R^2 \sim 0.830$, RMSE ~ 7.299, MAE ~ 5.091). Additionally, feature importance inputs were interpreted.

Although a broad range of supervised ML algorithms are available, recent studies suggest that the most accurate and successful algorithms used in geochemical investigations are XGBoost and random forest (RF) (Parsa 2021; Zhang et al. 2021, 2023; Ibrahim et al. 2022, 2023; Chen et al. 2023; Ye et al. 2023; Buccione et al. 2024). In this study, XGBoost models were employed to improve the estimation of REEs distribution in the northwestern Iranian karst bauxites. Unlike previous research on lithology (e.g., Buccione et al. 2024), which utilized all major elements to predict only HREEs, the present investigation explores the impact of clustering data based on geological age to examine changes in performance for both LREEs (La–Sm) and HREEs (Eu–Lu) through XGBoost. This study aims to provide insights into the factors influencing REEs distribution. Furthermore, the inclusion of additional parameters, such as the Chemical Index of Alteration (CIA) and the Chemical Index of Weathering (CIW), is assessed.

Geology of the deposits

The Iranian bauxite deposits belong to the Irano–Himalayan karst bauxite belt, exhibiting genetic similarities to the Mediterranean-type bauxites (Khosravi et al. 2021, 2025). These deposits are distributed across four major structural zones: the Northwestern Iran (NWI), the Alborz Mountain Chain (AMC), the Zagros Fold and Thrust Belt (ZFTB), and the Central Iranian Plateau (CIP). Their formation spans a wide temporal range, from the Middle–Upper Permian to the Upper Cretaceous. In the NWI, bauxite deposits occur across multiple stratigraphic intervals, including the Middle–Upper Permian, the Permo–Triassic, the Middle–Upper Triassic, and the Upper Triassic–Lower Jurassic (Abedini et al. 2021, 2022a, b; Khosravi et al. 2021). Notably, the presence of Ce-rich, Ti-rich, and high-field-strength elements (HFSEs) bauxite ores is a distinguishing feature of these deposits (Khosravi et al. 2017; Abedini et al. 2020b, 2024).

The Middle–Upper Permian NWI deposits

The Zagros orogenic belt of Iran, as part of the Alpine–Himalayan orogenic system, comprises three parallel NW-trending structural units: the Urumieh–Dokhtar Magmatic Belt (UDMB), the Sanandaj–Sirjan Zone (SaSZ), and the Zagros Fold and Thrust Belt (ZFTB) (Stöcklin 1968; Khosravi et al. 2019) (Fig. 1). The Middle–Upper Permian bauxite deposits within the SaSZ include the Shahindezh, Kanigorgeh, Darzi-Vali, Sheikh-Marut, and Badamlu deposits. These deposits formed within karstic depressions

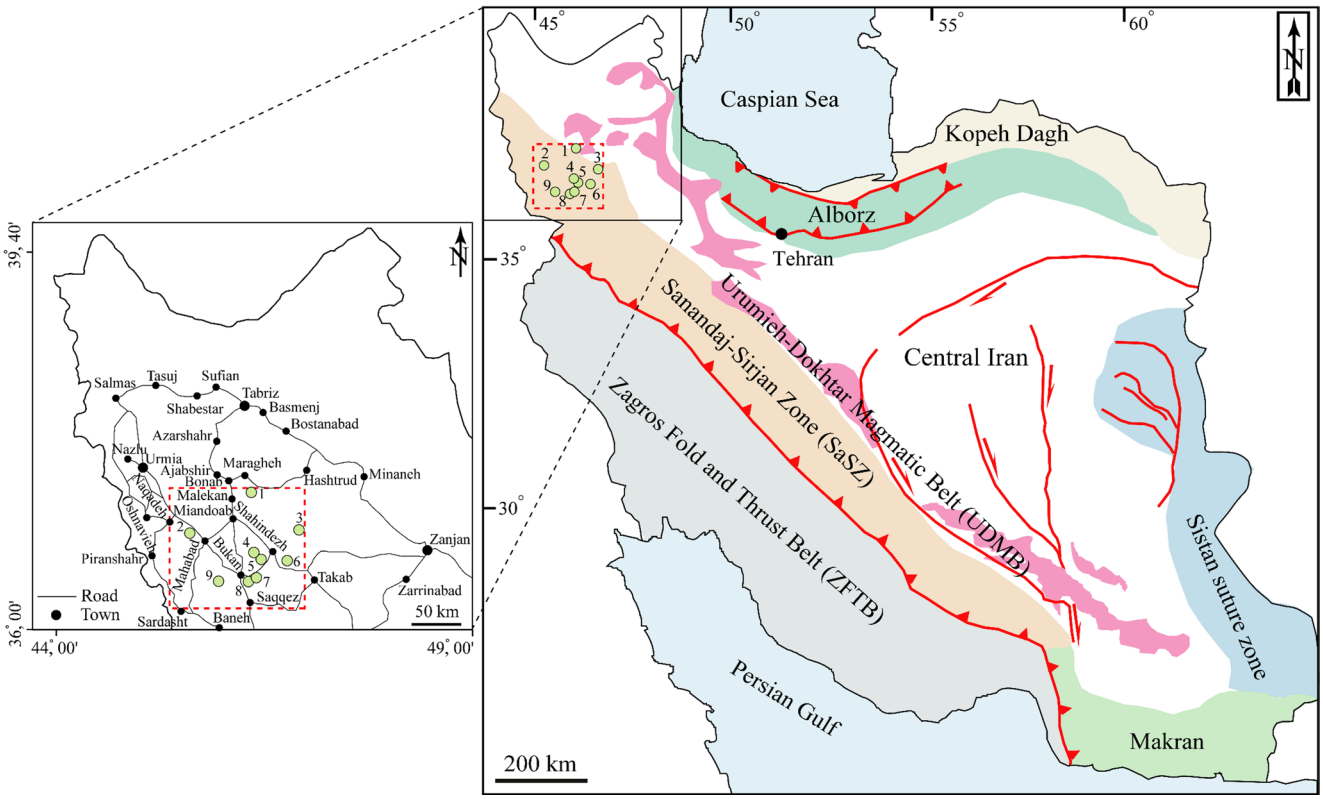


Fig. 1 Simplified tectonic map of Iran (Stöcklin 1968), showing the main tectonic domains and major strike-slip fault systems. The locations of the studied bauxite deposits in northwestern Iran are indicated by filled circles and include: (1) Amir-Abad, (2) Sheikh-Marut, (3) Badamlu, (4) Arbanos, (5) Kanigorgeh, (6) Shahindezh, (7) Darzi-Vali, (8) Soleiman Kandi, and (9) Kani Zarrineh

and sinkholes developed in the carbonate bedrock of the Middle–Upper Permian Ruteh and Nessen Formations (Figs. 2 and 3), occurring as irregular lenses and pockets of variable thickness. Based on carbonate facies stratigraphy and field observations, the relative chronological order of these deposits from the oldest to the youngest is as follows: Shahindezh, Kanigorgeh, Darzi-Vali, Sheikh-Marut, and Badamlu (Abedini et al. 2020b).

The Permo-Triassic Arbanos deposit

The oldest rock units in the Arbanos area belong to the Lalun Formation (Cambrian), composed of shale and sandstone, and the Mila Formation (Cambro-Ordovician), consisting of dolomite and quartzite. These units are unconformably overlain by the shale and sandstone of the Dorud Formation (Lower Permian) and limestone and dolomite with intercalations of shale of the Ruteh Formation (Upper Permian). The bauxite ores in this deposit occur as layers and lens-shaped pockets along the contact between the carbonate rocks intercalated with shales of the Ruteh Formation and the dolomite of the Elika Formation (Fig. 2). They exhibit a range of colors, including brownish-red, red, and greenish-cream. Two primary morphological types of bauxite ores are observed: stratified and massive. The massive ores predominantly developed in close proximity to the carbonate bedrock, whereas the stratified ores are more dispersed within the sequence.

The middle Triassic Kani Zarrineh deposit

The stratigraphic succession in the Kani Zarrineh area consists of, from the oldest to the youngest, the Lower Precambrian rhyolite, gneiss, schist, slate, phyllite, shale, and sandstone, followed by the Lower Precambrian shale and sandstone of the Kahar Formation, and the Upper Precambrian dolomite of the Soltanieh Formation (Fig. 2). The bauxite ores formed within karstic depressions and sinkholes in the Lower Triassic carbonate units of the Elika Formation (Fig. 3) and occur as layer-shaped and lenticular bodies, ranging in thickness from 4 to 22 m and extending approximately 2 km along a dominant N–S trend. The bauxite samples primarily exhibit earthy and massive textures, with rare occurrences of ooidic and pisolitic forms, displaying a brownish-red to red coloration towards the upper part of the succession.

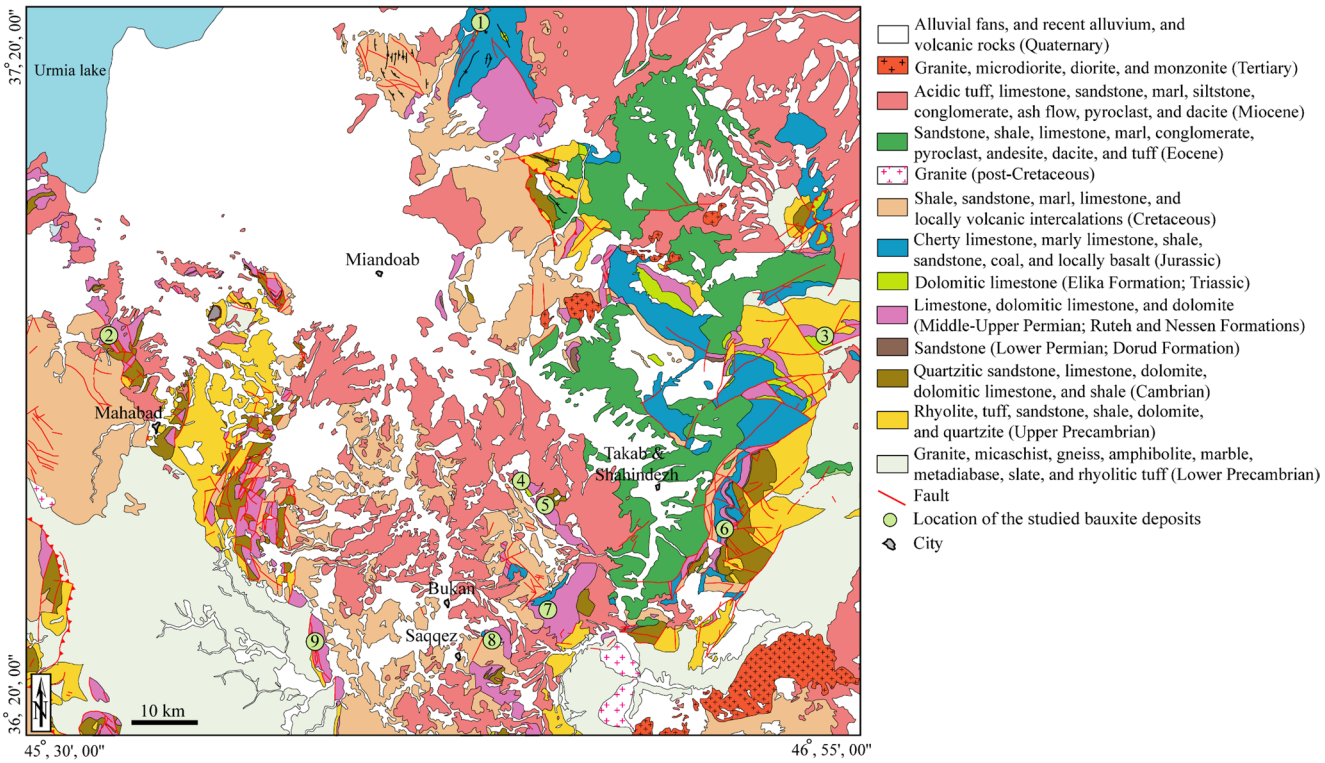


Fig. 2 Simplified geological map of the studied bauxite deposits in northwestern Iran. Locations correspond to those shown in Fig. 1 (modified after Abedini et al. 2022)

The Triassic–Jurassic Amir-Abad deposit

In the Amir-Abad area, the oldest exposed rock units belong to the Triassic dolomitic limestone of the Elika Formation and the Lower Jurassic sandstone, siltstone, shale, and coal of the Shemshak Formation. These units are overlain by the Middle Jurassic marly limestones of the Dalichay Formation, the Upper Jurassic limestones and dolomitic limestones of the Lar Formation, and the Cretaceous limestone of the Tizkuh Formation. The bauxite ores occur as layer-shaped and lenticular pockets at the contact between the dolomitic limestone of the Elika Formation and the sandstone, shale, and coal of the Shemshak Formation (Fig. 3). The deposition of bauxite ores within depressions of the carbonate footwall, the presence of karstified bedrock, and the sharp lithological contact between the bauxite layers and the dolomitic limestone bedrock of the Elika Formation are key geological features of this deposit.

The lower Jurassic Soleiman Kandi deposit

Bauxite ores in the Soleiman Kandi area can be classified into two categories based on their stratigraphic position: bauxites developed on a Permian substrate and those formed on a Lower Jurassic substrate. The bauxites associated with the Permian substrate occur as lenticular pockets, ranging in thickness from 4 to 25 m, within karstic depressions and sinkholes in the carbonate footwall of the Ruteh Formation. These bauxite layers exhibit a NW–SE trend, extend approximately 2.5 km in length, and display a range of colors from reddish-brown to grey, dark green, and cream. The bauxites developed on the Lower Jurassic substrate extend over 4 km with a thickness of 6–25 m. They formed within a synclinal depositional system at the contact between the underlying dolomite of the Elika Formation and the overlying sandstone, siltstone, and shale of the Shemshak Formation (Fig. 3).

Compiled dataset description and analysis

Data source

A comprehensive statistical summary of the whole dataset (i.e., Paleozoic and Mesozoic), which includes 200 samples (Khosravi et al. 2017; Abedini et al. 2019, 2020b, 2021, 2022b; Abedini and Khosravi 2020), is presented in Table 1. The geochemical data reveal significant compositional variability. SiO_2 has a mean concentration of 25.94 wt% with a standard deviation of 11.85 wt%, indicating substantial heterogeneity among the samples. Al_2O_3 and Fe_2O_3 exhibit mean concentrations of 37.92 wt% and 18.87 wt%,

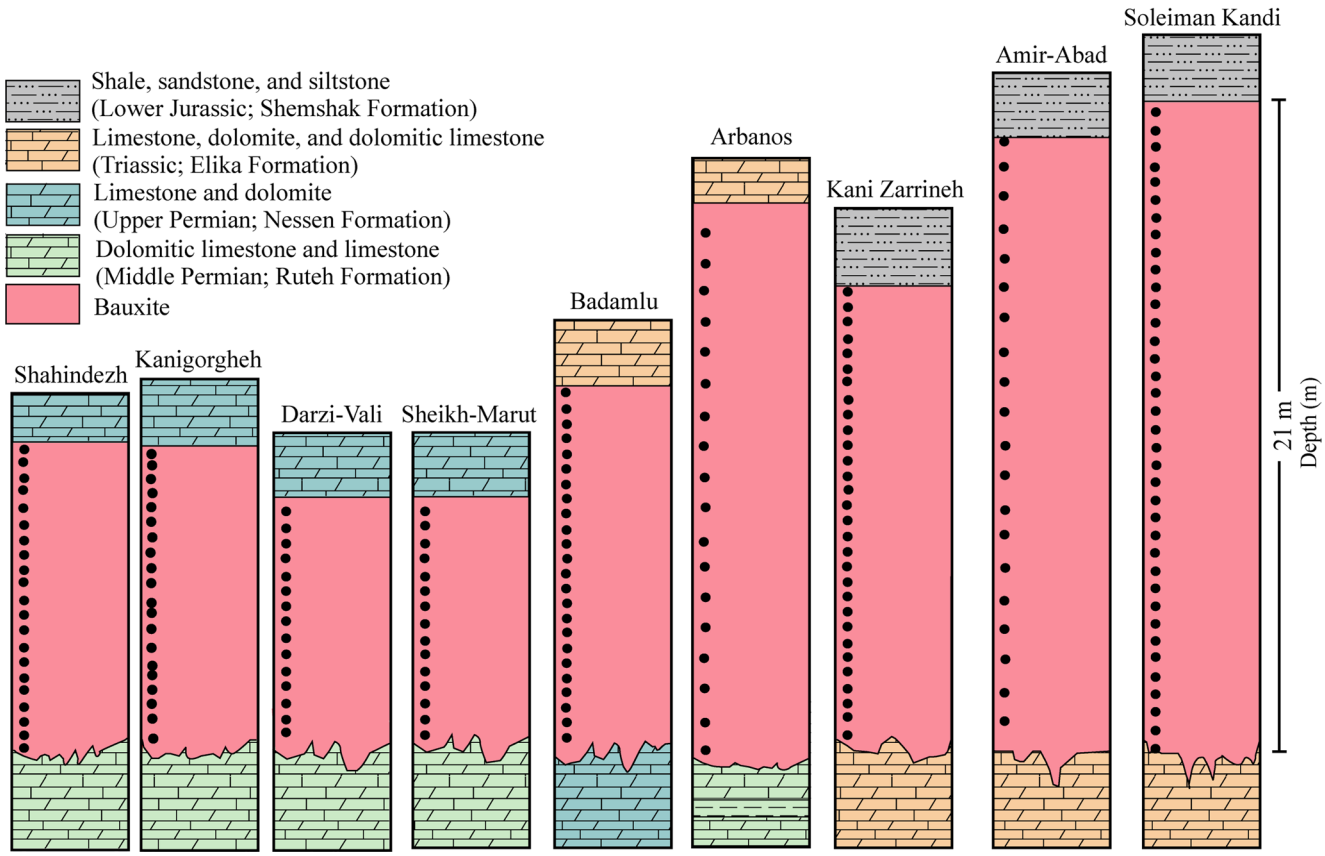


Fig. 3 Stratigraphic columns of the Paleozoic–Mesozoic bauxite deposits in northwestern Iran, with the positions of the analyzed samples indicated by filled circles

Table 1 Statistical summary of the entire dataset. The number of samples is 200. Data source: (Khosravi et al. 2017; Abedini et al. 2019, 2020b, 2021, 2022b; Abedini and Khosravi 2020)

SiO_2 (wt%)	Mean	Std	Min	25%	50%	75%	Max
	25.94	11.85	1.30	15.86	24.20	32.55	59.95
Al_2O_3	37.92	10.69	9.81	30.08	39.12	46.38	56.60
Fe_2O_3	18.87	14.29	0.28	6.10	18.06	25.28	69.40
CaO	0.17	0.21	0.01	0.03	0.13	0.21	1.61
MgO	0.34	0.28	0.02	0.15	0.25	0.49	2.18
Na_2O	0.12	0.14	0.01	0.03	0.07	0.12	0.79
K_2O	0.47	0.46	0.01	0.13	0.27	0.77	2.65
TiO_2	4.75	1.76	1.57	3.54	4.62	5.56	10.32
P_2O_5	0.25	0.20	0.02	0.09	0.23	0.32	0.99
LOI (wt%)	10.60	2.41	6.53	8.90	10.23	11.82	20.22
CIA (%)	98.12	1.43	90.47	97.28	98.27	99.34	99.89
CIW (%)	99.46	0.62	94.35	99.24	99.70	99.88	99.97
LREEs (La–Sm) (ppm)	349.43	403.40	1.16	76.76	214.82	425.80	2304.24
HREEs (Eu–Lu) (ppm)	39.45	31.16	0.23	14.53	32.50	55.33	157.48
(REEs+Y) (ppm)	433.77	451.07	18.71	104.91	316.56	534.21	2389.51

CIA (%) = $100 \times [(Al_2O_3)/(Al_2O_3 + CaO^* + Na_2O + K_2O)]$ (Nesbitt and Young 1982); CIW (%) = $100 \times [(Al_2O_3)/(Al_2O_3 + CaO^* + Na_2O)]$ (Harnois 1988)

respectively, reflecting the Al- and Fe-rich nature of the deposits. The loss on ignition (LOI), representing volatile content, has a mean value of 10.60 wt%. The CIA and CIW both have mean values close to 100%, suggesting intense weathering conditions. The concentrations of LREEs and (REEs+Y) exhibit high variability, with mean values of 349.43 ppm and 433.77 ppm, respectively, and standard deviations of 403.40 ppm and 451.07 ppm. These variations

Table 2 Statistical summary of the Paleozoic dataset. The number of samples is 95. Data source: (Abedini et al. 2020b)

SiO_2 (wt%)	Mean	Std	Min	25%	50%	75%	Max
	32.24	11.16	1.30	23.90	31.50	40.65	59.95
Al_2O_3	35.60	12.86	9.81	22.69	38.00	46.56	56.60
Fe_2O_3	14.59	15.41	0.28	1.79	9.74	22.01	69.40
CaO	0.18	0.20	0.01	0.02	0.13	0.23	0.64
MgO	0.29	0.23	0.02	0.16	0.21	0.37	1.70
Na_2O	0.18	0.17	0.02	0.07	0.11	0.20	0.79
K_2O	0.32	0.29	0.01	0.14	0.26	0.40	1.46
TiO_2	4.82	2.32	1.57	2.91	4.19	6.55	10.32
P_2O_5	0.28	0.17	0.02	0.19	0.26	0.32	0.97
LOI (wt%)	10.95	2.66	6.99	8.87	10.70	11.88	18.59
CIA (%)	98.05	1.45	90.47	97.49	98.24	99.11	99.78
CIW (%)	99.15	0.72	94.35	98.78	99.32	99.67	99.92
LREEs (La–Sm) (ppm)	422.84	518.31	7.78	58.48	204.58	571.06	2304.24
HREEs (Eu–Lu) (ppm)	40.60	35.69	1.13	13.87	25.51	56.07	157.40
(REEs+Y) (ppm)	508.24	579.26	18.71	91.71	249.84	721.18	2389.51

Table 3 Statistical summary of the Mesozoic dataset. The number of samples is 105. Data source: (Abedini et al. 2019, 2021, 2022b; Abedini and Khosravi 2020)

SiO_2 (wt%)	Mean	Std	Min	25%	50%	75%	Max
	20.24	9.31	3.37	14.02	18.01	25.61	45.82
Al_2O_3	40.01	7.67	21.64	34.29	39.75	45.91	52.66
Fe_2O_3	22.74	11.94	0.57	16.32	20.45	28.06	62.65
CaO	0.17	0.22	0.01	0.06	0.12	0.18	1.61
MgO	0.39	0.31	0.02	0.12	0.42	0.53	2.18
Na_2O	0.06	0.07	0.01	0.02	0.04	0.07	0.60
K_2O	0.60	0.54	0.01	0.13	0.42	1.02	2.65
TiO_2	4.68	1.01	2.16	3.87	4.78	5.21	7.05
P_2O_5	0.23	0.22	0.02	0.05	0.15	0.33	0.99
LOI (wt%)	10.29	2.13	6.53	9.12	10.01	11.55	20.22
CIA (%)	98.18	1.41	94.18	97.26	98.46	99.46	99.89
CIW (%)	99.75	0.28	98.04	99.71	99.84	99.92	99.97
LREEs (La–Sm) (ppm)	283.01	241.05	1.16	80.11	237.54	398.71	1298.55
HREEs (Eu–Lu) (ppm)	38.40	26.36	0.23	14.79	37.51	53.65	157.48
(REEs+Y) (ppm)	366.40	278.47	38.89	109.11	336.35	500.67	1486.23

highlight significant differences in REEs concentrations across the dataset.

Table 2 summarizes the geochemical characteristics of only the 95 Paleozoic samples from the total dataset of 200 samples. Compared to the entire dataset, these samples exhibit a higher mean SiO_2 concentration (32.24 wt%), suggesting silica enrichment in older deposits. The mean concentrations of Al_2O_3 (35.60 wt%) and Fe_2O_3 (14.59 wt%) are slightly lower than those observed in the whole dataset, with relatively higher standard deviations, indicating greater compositional variability. The CIA and CIW indices remain high, supporting the interpretation of intense weathering processes. The LREEs and (REEs+Y) concentrations in the Paleozoic samples have higher mean values (422.84 ppm and 508.24 ppm, respectively), with greater variability, which may be attributed to prolonged geochemical evolution over time.

Table 3 presents the statistical summary of only the 105 Mesozoic samples from the total dataset of 200 samples. In contrast to the Paleozoic data, SiO_2 exhibits a lower mean concentration (20.24 wt%), indicating reduced silica content in younger deposits. Conversely, Al_2O_3 and Fe_2O_3 show higher mean concentrations of 40.01 wt% and 22.74 wt%, respectively, suggesting increased Al and Fe enrichment. The LOI remains relatively stable, with a mean value of 10.29 wt%. The CIA and CIW indices continue to display high values, but with slightly lower variability than in the Paleozoic dataset, implying consistent weathering conditions. The LREEs and (REEs+Y) concentrations are lower in the Mesozoic samples, with mean values of 283.01 ppm and 366.40 ppm, respectively, exhibiting reduced variability. These differences between the Paleozoic and Mesozoic bauxites may reflect variations in source material

and/or paleoenvironmental conditions during bauxitization processes.

ML model

Various ML algorithms have been employed for geochemical modelling; however, recent findings suggest that XGBoost and RF are among the most accurate and effective methods in geochemical investigations (e.g., Parsa 2021; Ibrahim et al. 2022, 2023; Chen et al. 2023; Ye et al. 2023; Zhang et al. 2023; Buccione et al. 2024; Tahar-Belkacem et al. 2024). The geochemical data presented in Tables 1, 2 and 3 were utilized to develop the ML model, with a primary focus on predicting LREEs and HREEs under different scenarios. The XGBoost algorithm was selected for this task, due to its demonstrated superiority in regression problems, particularly in subsurface datasets, including both physical (Ouladmansour et al. 2023) and, more recently, geochemical data (Buccione et al. 2024; Tahar-Belkacem et al. 2024).

The performance of ML models, including XGBoost, is highly sensitive to hyperparameter selection (Schratz et al. 2019). In this study, hyperparameter tuning was performed using grid search, a widely employed technique known for its mathematical simplicity and robustness (Bergstra and Bengio 2012). Although computationally intensive, grid search systematically explores various hyperparameter combinations to optimize model performance. The tuning process included adjustments to key parameters, such as a base score of 0.9, a learning rate between 0.001 and 0.2, a number of estimators ranging from 40 to 300, maximum depths between 5 and 30, a minimum split loss of 0, a subsample rate of 0.9, and a column sampling ratio per tree of 0.8 (Table 4). This systematic approach ensured the identification of optimal hyperparameter combinations to enhance predictive accuracy.

XGBoost is a gradient-boosting algorithm specifically designed for decision-tree ensembles, initially introduced by Chen and Guestrin (2016). It enhances predictive performance by iteratively refining decision trees to minimize errors from previous iterations. The primary goal of XGBoost is to optimize its objective function, which balances the model's predictive performance and complexity through regularization. The core objective function in XGBoost in the g-th round is defined by Eq. (1):

$$L^{(g)} = \sum_{i=1}^{n} l\left(y_i, y_i^{\prime(g-1)} + F_g(x_i)\right) + \sum_{k=1}^{T} \Omega\ (F_k) \tag{1}$$

where $l\left(y_i, y_i^{\prime}\right)$ represents the loss function, quantifying the difference between the predicted ($y_i^{\prime}$) and actual (y_i) values. $F_g(x_i)$ denotes the prediction generated by the new tree added in the g-th iteration, $\Omega\ (F_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2$ is the regularization term, which penalizes model complexity by controlling the number of leaves (T) and the magnitude of leaf weights (ω_j), and γ and λ are regularization parameters that mitigate overfitting.

Model performance was evaluated using standard regression metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). These metrics, defined in Eqs. (2)–(4), quantify the predictive accuracy of the model by comparing predicted and observed values:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}(XI - YI)^2}, (0 \le \mathrm{RMSE} < +\infty) \tag{2}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |YI - XI|, (0 \le \mathrm{MAE} < +\infty) \tag{3}$$

$$R^2 = \left[\frac{\frac{1}{N}\sum_{i=1}^{N}(X1)\left(\overline{X1}\right)}{\sqrt{(X1-\overline{X1})^2}\sqrt{(Y1-\overline{Y1})^2}}\right]^2, (0 < R^2 < +1) \tag{4}$$

where N is the number of observations, XI represents the actual values, YI represents the predicted values, and $\overline{X1}$ and $\overline{Y1}$ are the mean of the observed X and Y.

Figure 4 shows a flowchart illustrating the ML methodology prediction using XGBoost in the studied bauxite deposits.

Table 4 Hyperparameter settings applied to the configured ML model

Algorithm	Hyperparameter	Optimum value selected
XGB	Base_score	0.9
	Colsample_bytree	0.8
	n_estimators	300
	Max_depth	30
	Learning rate	0.2
	Min_split_loss (γ)	0
	Reg_lambda (λ)	1
	Subsample	0.9

Results and discussion

Predictive performances for LREEs and HREEs

The predictive performance of the model for LREEs was assessed across different geological periods, including the Paleozoic and Mesozoic eras, with the incorporation of geochemical indices, such as CIW and CIA (Tables 5 and 6; Figs. 5, 6, 7 and 8). The results, presented in Table 5; Fig. 5, indicate strong predictive capability in Mesozoic, particularly for the Permo-Triassic, middle Triassic,

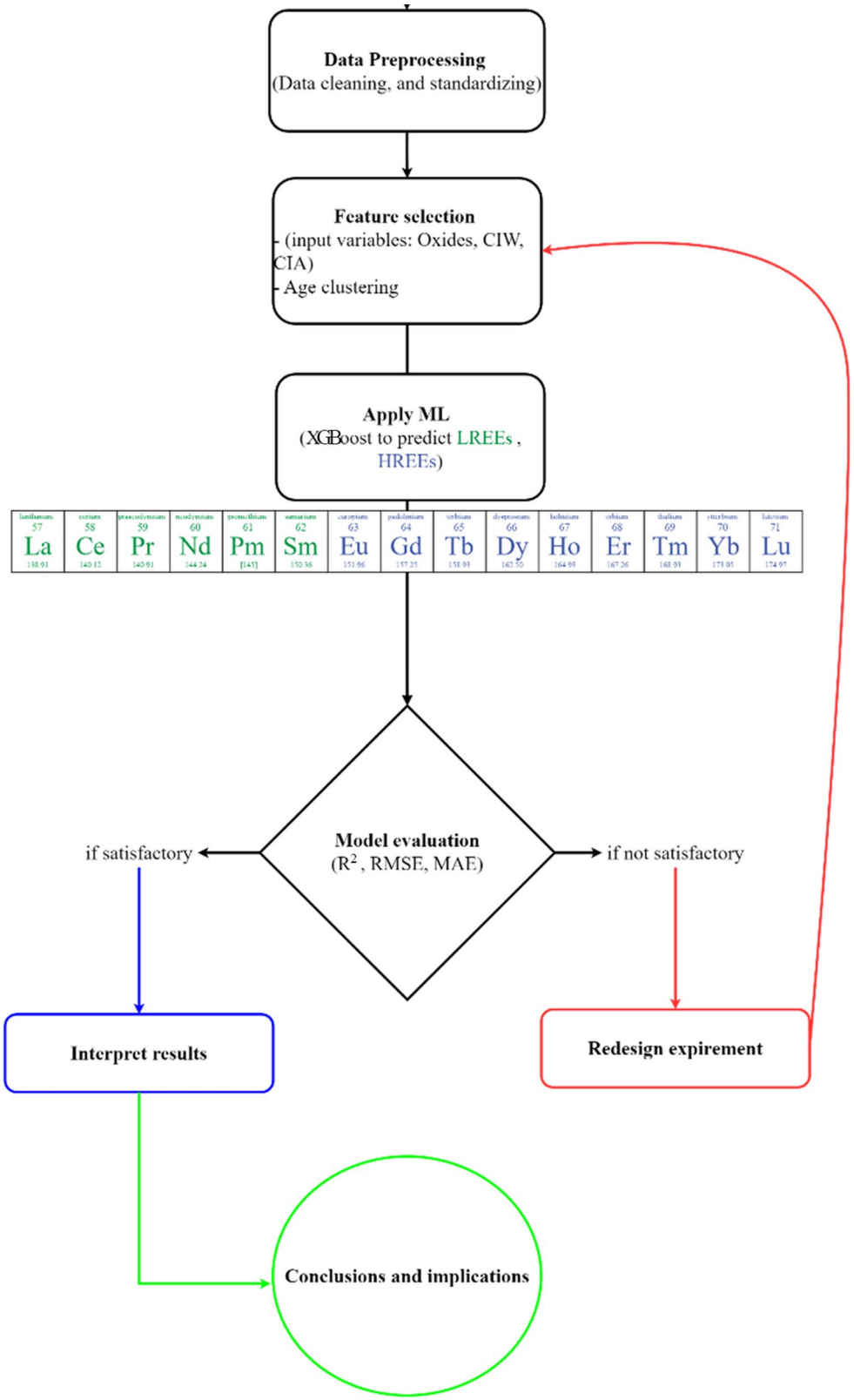


Fig. 4 Workflow for ML prediction using XGBoost in the northwestern Iranian karst bauxite deposits

Table 5 LREEs prediction error for the XGB algorithm (bold values indicate the highest-performing combinations)

Era	Combinations	R^2	RMSE	MAE
Paleozoic	Oxides	**0.853**	**173.040**	**118.294**
	Oxides + CIW	0.783	210.56069	138.148
	Oxides + CIA	0.8238	188.481	118.588
	Oxides + CIW + CIA	0.8297	186.338	121.621
Mesozoic	Oxides	0.934	**51.415**	**40.059**
	Oxides + CIW	**0.951**	51.876	40.888
	Oxides + CIA	0.9096	69.832	53.568
	Oxides + CIW + CIA	0.9099	69.708	53.454

Table 6 HREEs prediction error for the XGB algorithm applied to the entire dataset (bold values indicate the highest-performing combinations)

Era	Phase*	R^2	RMSE	MAE
Paleozoic	Oxides	0.922	8.771	6.928
	Oxides + CIW	0.897	9.835	7.367
	Oxides + CIA	0.9367	7.892	6.2425
	Oxides + CIW + CIA	**0.955**	**6.642**	**5.5065**
Mesozoic	Oxides	**0.778**	**12.034**	**8.908**
	Oxides + CIW	0.620	13.595	9.582
	Oxides + CIA	0.600	13.926	9.841
	Oxides + CIW + CIA	0.666	12.631	9.249

Triassic–Jurassic, and early Jurassic intervals. Notably, the inclusion of the CIW index marginally improved model performance, with values reaching up to 0.951, meaning that the model explains 95.1% of the variance in LREEs predictions. Additionally, the model exhibited low RMSE (51.876) and MAE (40.888) values, further supporting its predictive reliability. Conversely, the model demonstrated relatively weaker performance for the Paleozoic era, particularly the Middle–Upper Permian, where values ranged from 0.783 to 0.853. Furthermore, the inclusion of geochemical indices did not enhance predictive accuracy in these cases; for instance, incorporating both CIA and CIW indices led to a slight decrease in the value from 0.853 (using only major elements) to 0.829.

Similarly, the model's predictive performance for HREEs was evaluated across different geological periods and with the inclusion of CIA and CIW indices. The results, summarized in Table 6; Figs. 7 and 8, indicate a lower correlation compared to LREEs, yet lower error values were observed in the Paleozoic samples, particularly when the CIA and CIW index were included. The values for HREEs predictions

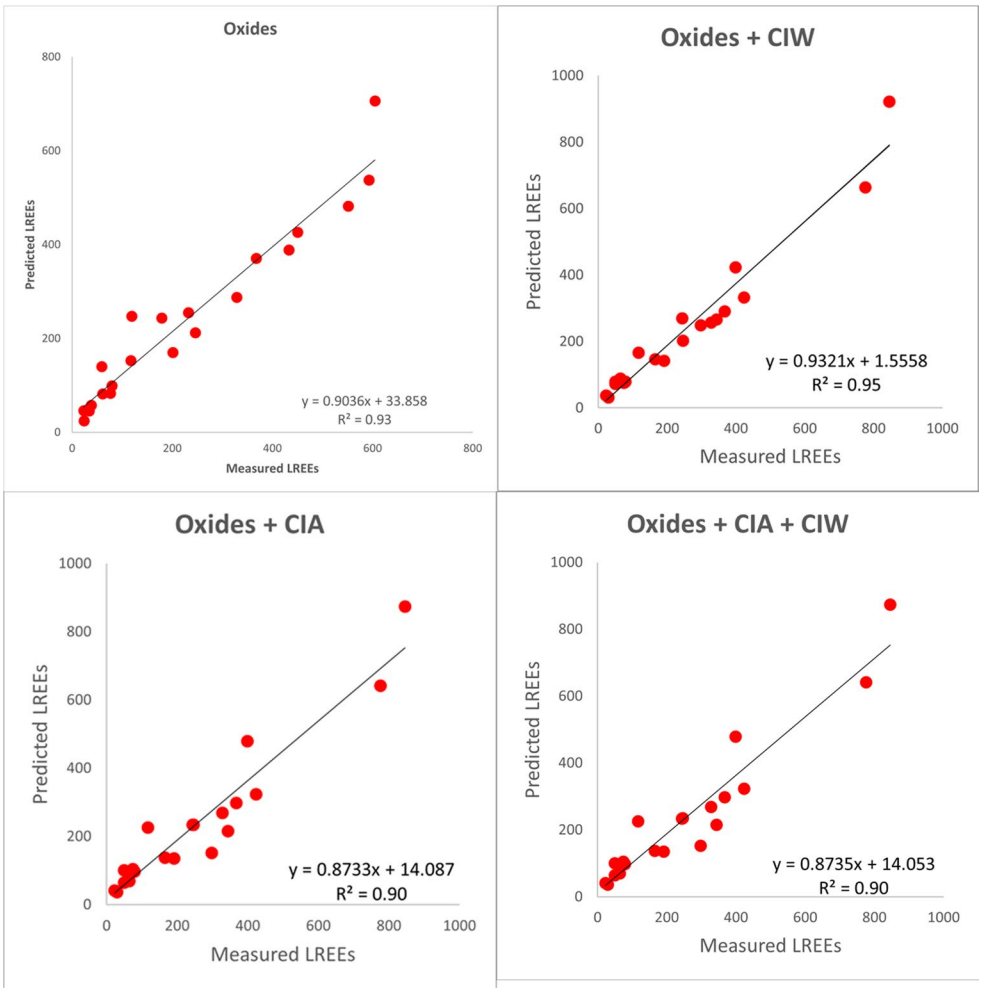


Fig. 5 Measured LREEs versus predicted LREEs for the XGB algorithm in the Mesozoic datasets

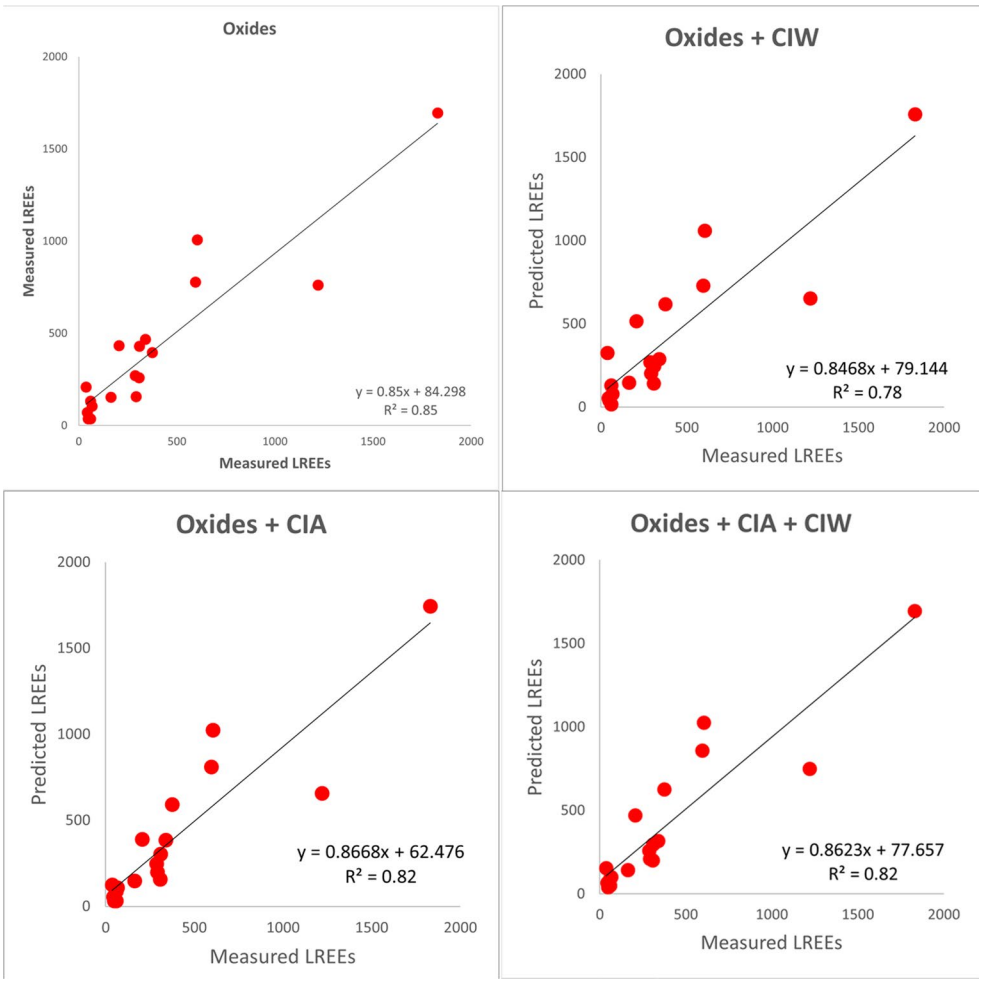


Fig. 6 Measured LREEs versus predicted LREEs for the XGB algorithm in the Paleozoic datasets

reached a maximum of 0.955, suggesting that the model explains 95.5% of the variance in HREEs concentrations.

Evaluation and interpretation of XGBoost predictions

The application of ML in geochemical predictions has been widely explored in recent studies (e.g., Zuo et al. 2019, 2021; Nassif et al. 2021; Huang et al. 2022; Xu and Zuo 2024). For instance, Tahar-Belkacem et al. (2024) demonstrated that XGBoost consistently outperforms RF in predicting REEs concentrations. Data clustering based on geochemical fingerprints further enhances prediction accuracy. Additionally, studies on bauxites (Buccione et al. 2024) showed that XGBoost effectively models the distribution of HREEs in the Cretaceous Mediterranean-type karst bauxite deposits of southern Italy, yielding high predictive performance. In the present study, our developed models extend these capabilities to predict LREEs as well. A comparison of predicted versus actual LREEs and HREEs values (Fig. 9) illustrates the strong predictive power of XGBoost when using major oxides in combination with the CIA and CIW. Furthermore, the feature importance analysis provides insights into the most influential variables in the predictions. Previous geochemical studies have successfully employed feature importance scores to interpret variable significance in ML models (e.g., Sun et al. 2023; Buccione et al. 2024; Zhang et al. 2024). Interestingly, in our case, feature importance scores reveal that P_2O_5 is consistently the most influential predictor across all age-constrained datasets (Fig. 10). This finding underscores the critical role of P_2O_5 in REEs distribution modelling and highlights the strength of XGBoost in geochemical analysis.

Constraints in LREEs prediction

The results of LREEs prediction reveal significant differences in model performance between the Paleozoic and Mesozoic clusters. XGBoost captures approximately 95% of the total variance in LREEs within the Mesozoic deposits when using oxides as predictor inputs ($R^2 = 0.951$). However, its performance decreases in the Paleozoic deposits ($R^2 = 0.853$), which can be attributed to differences in the underlying data structure. Notably, the standard deviation

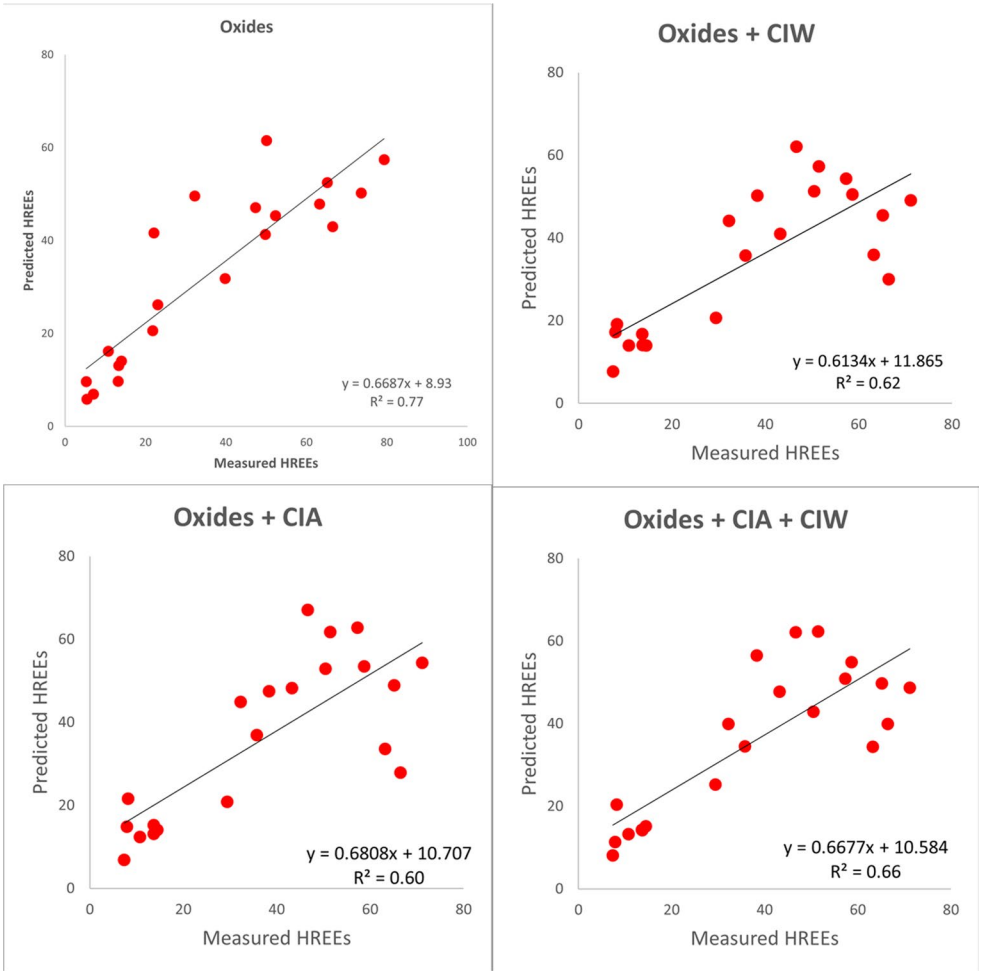


Fig. 7 Measured HREEs versus predicted HREEs for the XGB algorithm in the Mesozoic datasets

of LREEs is higher in the Paleozoic deposits compared to the Mesozoic ones, with mean values of 422.84 ± 518.31 ppm and 283.01 ± 240.05 ppm, respectively. In particular, the high variability of Ce concentrations in the Paleozoic samples (Fig. 11) appears to have a significant impact on LREEs content, especially in the Upper Permian samples, contributing to the model's lower predictive accuracy. Due to its unique redox chemistry, Ce behaves differently from other LREEs. In addition to Ce^{3+}, it can also exist as Ce^{4+}, which may precipitate as cerianite (CeO_2) under slightly acidic conditions (0.38–0.61 V redox potential) during bauxitization and intense weathering (Ahmadnejad and Mongelli 2022). Cerianite, along with Ce-rich phosphates, such as florencite and monazite, has been observed in the northwestern Iranian karst bauxites (Abedini et al. 2020a, 2021, 2022a, b, 2024). Changes in paleoredox conditions can either promote the formation of cerianite or favor Ce-rich phosphate minerals.

The Ce/Ce* ratio, exceeding 20 in few Upper Permian samples, indicates significant paleoredox fluctuations over geological time (Abedini et al. 2020a). Furthermore, since the northwestern Iranian karst bauxite deposits studied here—including both the Paleozoic and Mesozoic deposits, with the exception of the Arbanos deposit— primarily originate from a mafic protolith (Khosravi et al. 2017; Abedini and Khosravi 2020; Abedini et al. 2020b, 2021, 2022b), variations in paleoredox conditions have likely influenced Ce behavior within and across these deposits. These geochemical differences contribute to the lower predictive accuracy for LREEs in the Paleozoic samples. Conversely, the superior model performance in the Mesozoic deposits likely reflects their more homogeneous geochemical characteristics, including minor fluctuations in the Ce/Ce* ratio, which remains close to unity. This stability may have favored the formation of LREE-rich phosphate minerals, leading to improved predictability. Additionally, the stronger CIA feature importance score for the Mesozoic samples suggests that paleoweathering also plays a role in LREEs distribution and prediction. More intense weathering conditions likely enhanced the breakdown of LREE-bearing detrital minerals, subsequently promoting the formation of authigenic LREE-rich phosphates. These findings highlight the necessity of considering both geological age and key geochemical factors—such as Ce variability, redox conditions, and

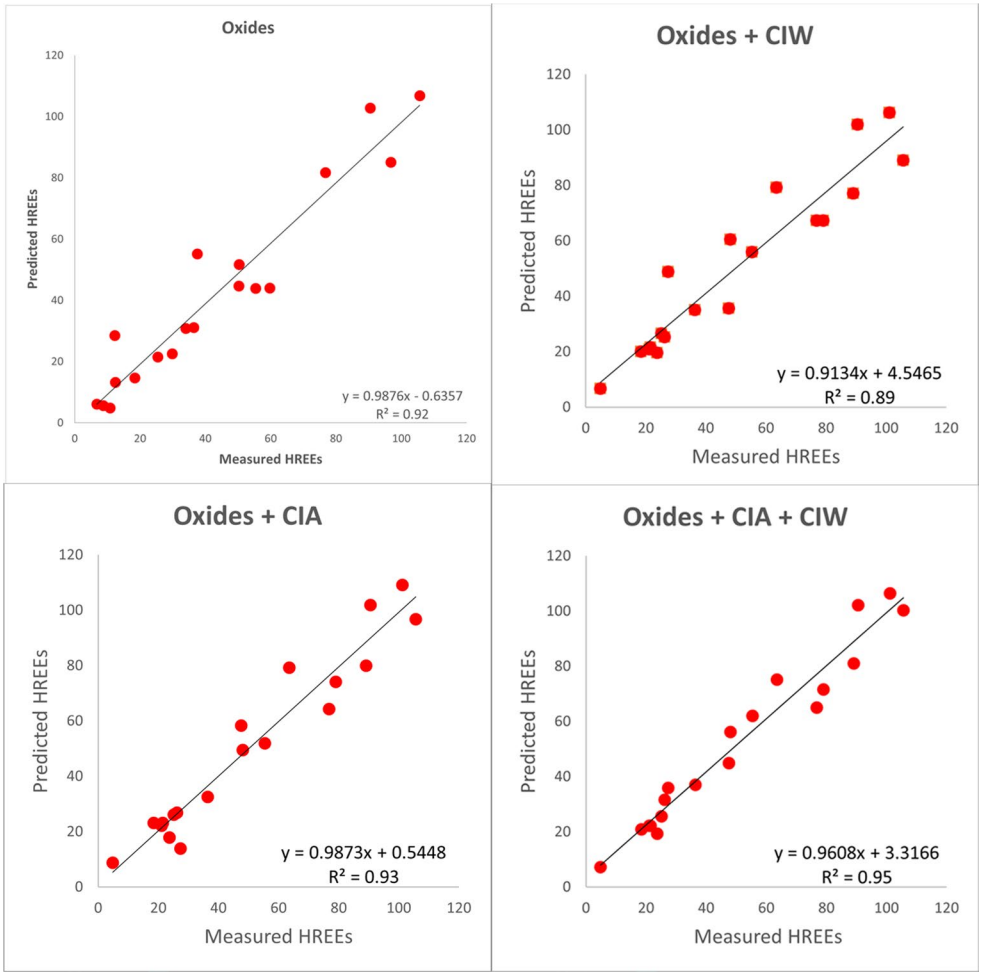


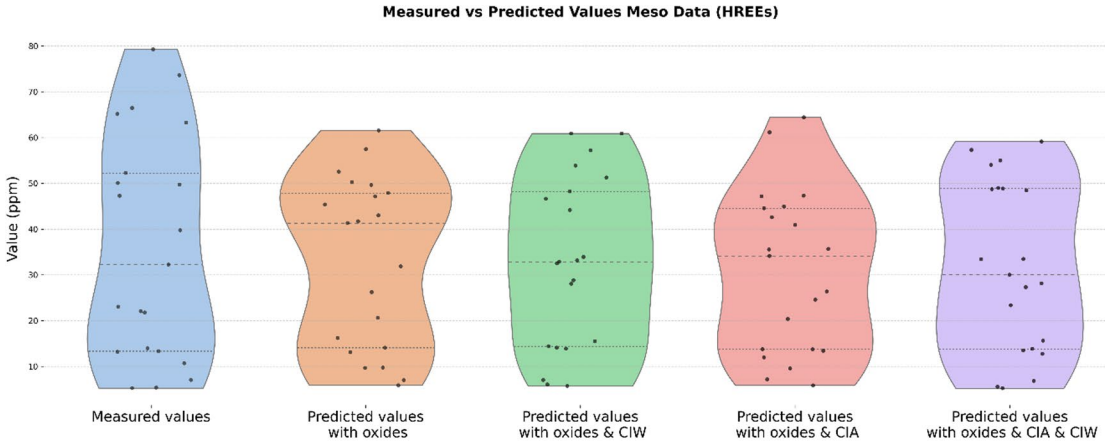
Fig. 8 Measured HREEs versus predicted HREEs for the XGB algorithm in the Paleozoic datasets

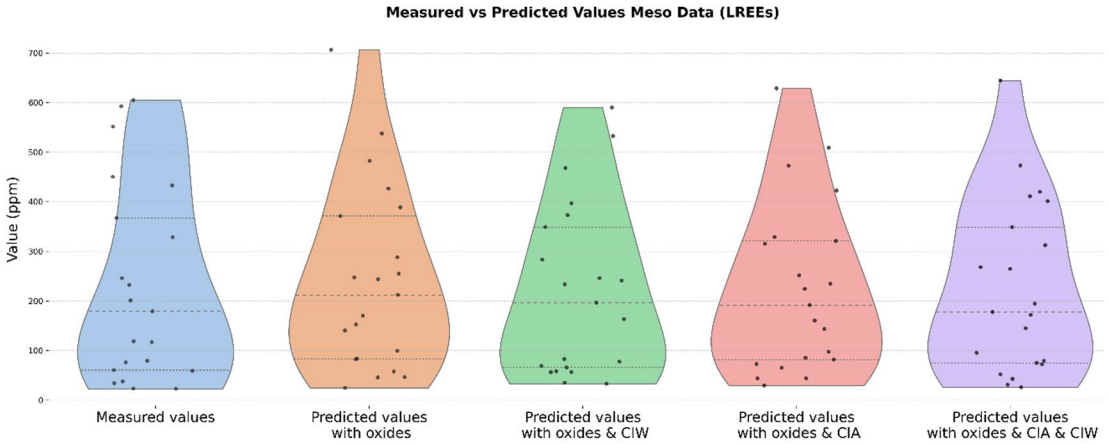
paleoweathering—when applying ML techniques to predict LREEs distribution in karst bauxites.

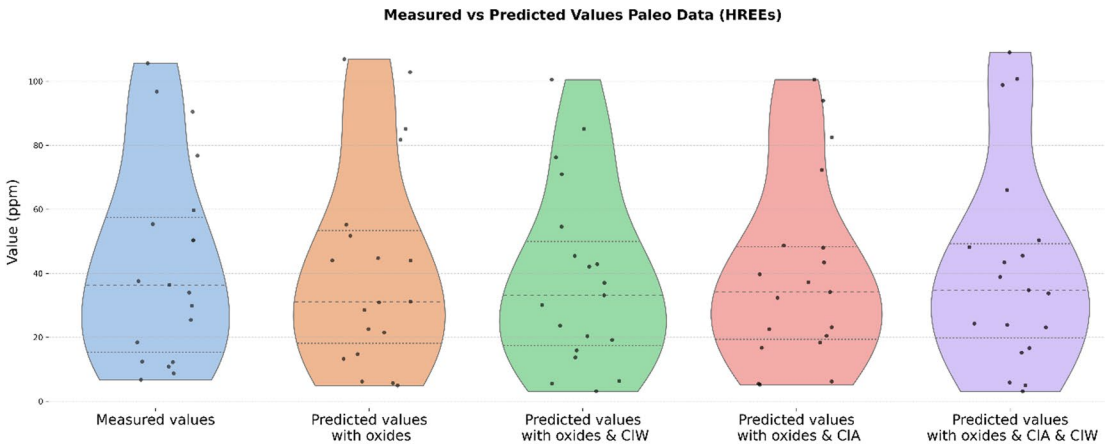
Constraints in HREEs prediction

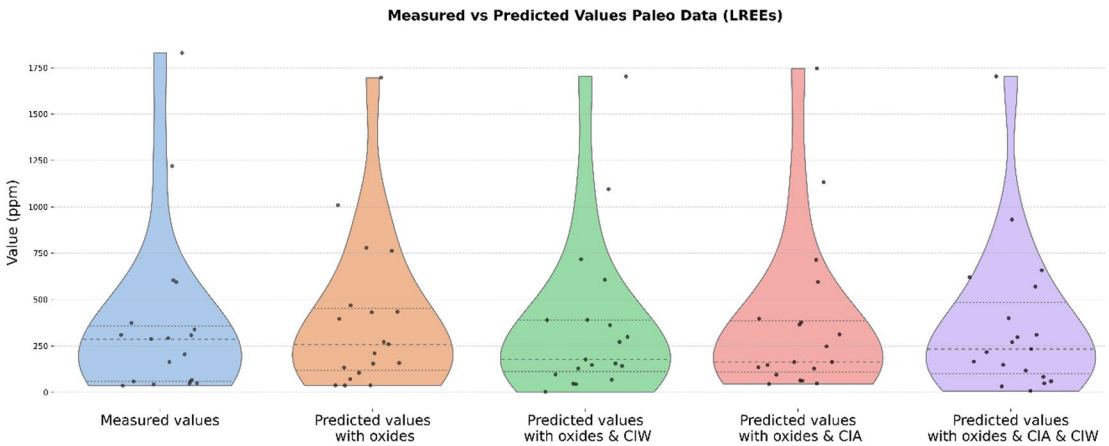
The results of HREEs prediction reveal a significant difference between the Paleozoic and Mesozoic samples. Using oxides as predictor variables, the XGBoost model demonstrates lower performance in the Mesozoic dataset ($R^2 = 0.778$) compared to the Paleozoic dataset ($R^2 = 0.955$). This discrepancy suggests that major oxide-hosted minerals, particularly phosphates, play a more effective role in accommodating HREEs within the Paleozoic samples. Notably, the feature importance score of P_2O_5 is substantially higher for the Paleozoic samples than for the Mesozoic ones, reinforcing the idea that phosphate minerals are the primary hosts of HREEs in these deposits. In the northwestern Iranian karst bauxites, the only HREE-enriched minerals are phosphates, such as churchite and xenotime (Abedini et al. 2020a, 2021). The formation of authigenic HREE-rich phosphates typically occurs under acidic conditions (Li et al. 2013) and is likely associated with HREE-enriched percolating solutions that develop after the precipitation of LREE-rich fluorocarbonates. In contrast, detrital phosphates tend to persist as resistant minerals over geological timescales. However, recent studies suggest that bio-weathering mechanisms may significantly enhance REEs solubilization from minerals, such as xenotime (He et al. 2024). For the Mesozoic dataset, the CIW—a modification of the CIA index that excludes K_2O to eliminate diagenetic or aeolian influences (Martínez et al. 2007)—exhibits the highest feature importance score after P_2O_5. This finding suggests that variations in weathering intensity, as captured by the CIW index, influence the occurrence and distribution of HREE-rich phosphate minerals in the Mesozoic samples.

Fig. 9 Violin plot comparisons between measured and predicted values (in ppm) for the XGBoost models


Measured vs Predicted Values Meso Data (HREEs)
Value (ppm)
Measured values
Predicted values with oxides
Predicted values with oxides & CIW
Predicted values with oxides & CIA
Predicted values with oxides & CIA & CIW


Measured vs Predicted Values Meso Data (LREEs)
Value (ppm)
Measured values
Predicted values with oxides
Predicted values with oxides & CIW
Predicted values with oxides & CIA
Predicted values with oxides & CIA & CIW


Measured vs Predicted Values Paleo Data (HREEs)
Value (ppm)
Measured values
Predicted values with oxides
Predicted values with oxides & CIW
Predicted values with oxides & CIA
Predicted values with oxides & CIA & CIW


Measured vs Predicted Values Paleo Data (LREEs)
Value (ppm)
Measured values
Predicted values with oxides
Predicted values with oxides & CIW
Predicted values with oxides & CIA
Predicted values with oxides & CIA & CIW

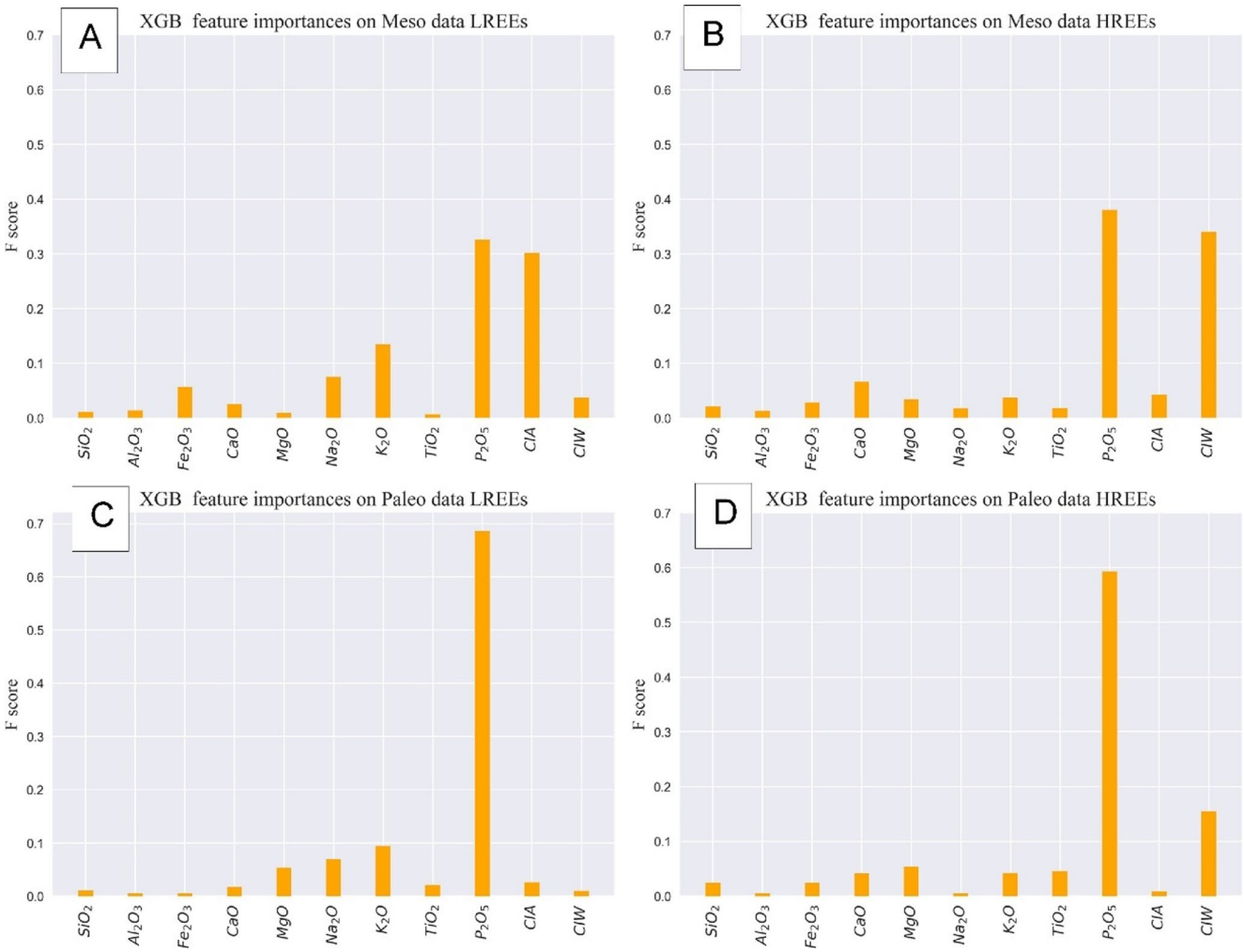


Fig. 10 Feature importance in the prediction of (**A**) LREEs (Mesozoic), (**B**) HREEs (Mesozoic), (**C**) LREEs (Paleozoic), and (**D**) HREEs (Paleozoic)

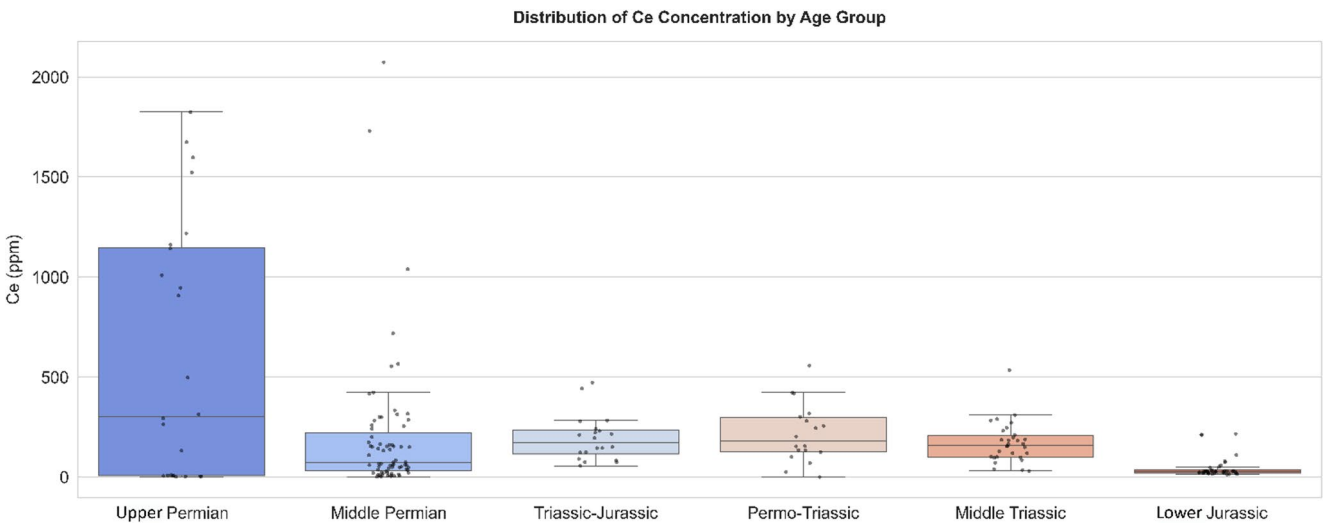


Fig. 11 Boxplot of Ce concentration based on the age of the deposits

Conclusions

This study on the age-constrained northwestern Iranian karst bauxite deposits, employing the XGBoost machine learning model, provides several key insights into the predictive performance of REEs distribution models:

1. Influence of paleoenvironmental factors

The distribution of LREEs and HREEs is strongly controlled by paleoenvironmental conditions, resulting in contrasting predictive outcomes across different geological age clusters.

2. Predictive performance of LREEs vs. HREEs

For LREEs, the model shows high accuracy during the Mesozoic era ($R^2 = 0.951$, RMSE = 51.876, MAE = 40.888).

Predictive accuracy decreases in the Paleozoic samples, due to higher geochemical variability, particularly the fluctuating redox state of Ce and the formation of cerianite and Ce-rich phosphate phases, as indicated by the Ce/Ce* ratio.

In contrast, HREEs achieve the highest predictive accuracy in the Paleozoic dataset ($R^2 = 0.955$), reflecting the strong role of phosphate minerals in hosting these elements.

3. Feature importance and mineralogical controls

The P_2O_5 feature emerges as the most significant predictor for both LREEs and HREEs, emphasizing the role of phosphate minerals in lanthanide hosting during bauxitization.

Other oxides (Al_2O_3, Fe_2O_3, and TiO_2) are abundant but relatively uniform, contributing less to predictive accuracy.

Weathering indices (CIA and CIW) also show substantial importance, indicating that subtle variations in paleoweathering conditions can significantly influence lanthanide distribution.

4. Implications for predictive modelling and resource assessment

The dependency of REEs distribution on geological age and paleoenvironmental factors highlights the need for tailored predictive models.

The LREEs and HREEs prediction models developed here demonstrate promising performance and can inform future resource assessment strategies.

Further cross-validation in other karst bauxite provinces is recommended to refine predictive accuracy and reduce the cost and time of future REEs exploration.

Acknowledgements Special thanks are extended to the Associate Editor for their valuable suggestions and editorial assistance, as well as to two anonymous reviewers for their critical insights and constructive feedback on this manuscript.

Author contributions Ali Abedini: Data Curation, Interpretation of the Results, Writing Review and Editing. Maryam Khosravi: Interpretation of the Results, Writing-original draft preparation, Writing Review and Editing. Ouafi Ameur-Zaimeche: Methodology, Software, Investigation, Visualization, Writing Review and Editing. Rabah Kechiched: Interpretation of the Results, Conceptualization, Writing-original draft preparation, Writing Review and Editing. Abdelhamid Ouladmansour: Methodology, Software, Investigation, Visualization, Writing Review and Editing. Giovanni Mongelli: Interpretation of the Results, Conceptualization, Investigation, Validation, Writing Review and Editing.

Funding Open access funding provided by Università degli Studi di Bari Aldo Moro within the CRUI-CARE Agreement.

Data availability The datasets analysed during the current study are available from the corresponding author upon reasonable request.

Declarations

Financial interests The authors declare that they have no conflicts of interest regarding the information presented in this study. They also confirm that no financial support has been received for this work, and it has not influenced its content in any manner.

Competing interests The authors declare no competing interests.

References

Abedini A, Khosravi M (2020) Geochemical constraints on the Triassic–Jurassic Amir-Abad karst type bauxite deposit, NW Iran. J Geochem Explor 211:106489

Abedini A, Khosravi M, Calagari AA (2019) Geochemical characteristics of the Arbanos karst-type bauxite deposit, NW Iran: implications for parental affinity and factors controlling the distribution of elements. J Geochem Explor 200:249–265

Abedini A, Khosravi M, Dill HG (2020a) Rare earth element geochemical characteristics of the late Permian Badamlu karst bauxite deposit, NW Iran. J Afr Earth Sci 172:103974

Abedini A, Khosravi M, Mongelli G (2022a) The middle Permian pyrophyllite-rich ferruginous bauxite, northwestern Iran, Irano–Himalayan karst belt: constraints on elemental fractionation and provenance. J Geochem Explor 233:106905

Abedini A, Khosravi M, Mongelli G (2024) Critical metals distribution in the late Triassic–early Jurassic Nasr-Abad bauxite deposit,

Irano–Himalayan karst bauxite belt, NW Iran. Geochemistry 84(2):126039
Abedini A, Mongelli G, Khosravi M (2021) Geochemical constraints on the middle Triassic Kani Zarrineh karst bauxite deposit, Irano–Himalayan belt, NW Iran: implications for elemental fractionation and parental affinity. Ore Geol Rev 133:104099
Abedini A, Mongelli G, Khosravi M (2022b) Geochemistry of the early Jurassic Soleiman Kandi karst bauxite deposit, Irano–Himalayan belt, NW Iran: constraints on bauxite genesis and the distribution of critical raw materials. J Geochem Explor 241:107056
Abedini A, Mongelli G, Khosravi M, Sinisi R (2020b) Geochemistry and secular trends in the middle–late Permian karst bauxite deposits, northwestern Iran. Ore Geol Rev 124:103660
Ahmadnejad F, Mongelli G (2022) Geology, geochemistry, and genesis of REY minerals of the late Cretaceous karst bauxite deposits, Zagros Simply Folded Belt, SW Iran: constraints on the ore-forming process. J Geochem Explor 240:107030
Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. J Mach Learn Res 13:2
Buccione R, Ameur-Zaimeche O, Ouladmansour A, Kechiched R, Mongelli G (2024) Data-centric approach for predicting critical metals distribution: heavy rare earth elements in Cretaceous Mediterranean-type karst bauxite deposits, southern Italy. Geochemistry 84(2):126026
Castor SB (2008) Rare earth deposits of North America. Resour Geol 58(4):337–347
Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785–794
Chen Y, Li F, Zhou S, Zhang X, Zhang S, Zhang Q, Su Y (2023) Bayesian optimization based random forest and extreme gradient boosting for the pavement density prediction in GPR detection. Constr Build Mater 387:131564
Cracknell MJ, Reading AM (2014) Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Comput Geosci 63:22–33
Dramsch JS (2020) 70 years of machine learning in geoscience in review. Adv Geophys 61:1–55
Dumakor-Dupey NK, Arya S (2021) Machine learning—a review of applications in mineral resource estimation. Energies 14(14):4079
Esmaeiloghli S, Tabatabaei SH, Carranza EJM (2023) Infomax-based deep autoencoder network for recognition of multi-element geochemical anomalies linked to mineralization. Comput Geosci 175:105341
Gamaletsos PN, Godelitsas A, Filippidis A, Pontikes Y (2019) The rare earth elements potential of Greek bauxite active mines in the light of a sustainable REE demand. J Sustain Metall 5:20–47
Grohol M, Veeh C (2023) Study on the Critical Raw Materials for the EU 2023. Publications Office of the European Union, p 160
Harnois L (1988) The CIW index: a new chemical index of weathering. Sediment Geol 55(3):319–322
He Y, Ma L, Liang X, Li X, Zhu J, He H (2024) Resistant rare earth phosphates as possible sources of environmental dissolved rare earth elements: insights from experimental bio-weathering of xenotime and monazite. Chem Geol 661:122186
Huang D, Zuo R, Wang J (2022) Geochemical anomaly identification and uncertainty quantification using a Bayesian convolutional neural network model. Appl Geochem 146:105450
Ibrahim B, Ahenkorah I, Ewusi A, Majeed F (2023) A novel XRF-based lithological classification in the Tarkwaian paleo placer formation using SMOTE-XGBoost. J Geochem Explor 245:107147
Ibrahim B, Majeed F, Ewusi A, Ahenkorah I (2022) Residual geochemical gold grade prediction using extreme gradient boosting. Environ Chall 6:100421
Kesler SE, Simon AC, Simon AF (2015) Mineral resources, economics and the environment. Cambridge University Press
Khosravi M, Abedini A, Alipour S, Mongelli G (2017) The Darzi-Vali bauxite deposit, West Azarbaidjan Province, Iran: critical metals distribution and parental affinities. J Afr Earth Sci 129:960–972
Khosravi M, Rajabzadeh MA, Qin K, Asadi Harooni H (2019) Tectonic setting and mineralization potential of the Zefreh porphyry Cu–Mo deposit, central Iran: constraints from petrographic and geochemical data. J Geochem Explor 199:1–15
Khosravi M, Vérard C, Abedini A (2021) Palaeogeographic and geodynamic control on the Iranian karst-type bauxite deposits. Ore Geol Rev 139:104589
Khosravi M, Yu W, Abedini A, Zhou J (2025) Geochemical constraints on critical metals in the Jajarm bauxite deposit in northeastern Iran. Sci Rep 15:25256
Li JW, Li TJ, Jia HX, Wang AJ (2023) Determination of China's strategic and critical minerals list. Acta Geosci Sin 44(2):261–270
Liu X, Wang Q, Zhang Q, Zhang Y, Li Y (2016) Genesis of REE minerals in the karstic bauxite in western Guangxi, China, and its constraints on the deposit formation conditions. Ore Geol Rev 75:100–115
Liu X, Wang Q, Zhao L, Peng Y, Ma Y, Zhou Z (2020) Metallogeny of the large-scale Carboniferous karstic bauxite in the Sanmenxia area, southern part of the North China Craton, China. Chem Geol 556:119851
Li ZH, Din J, Xu JS, Liao CG, Yin FG, Lu T, Cheng L, Li JM (2013) Discovery of the REE minerals in the Wulong-Nanchuan bauxite deposits, Chongqing, China: insights on conditions of formation and processes. J Geochem Explor 133:88–102
Martínez J, Llamas J, De Miguel E, Rey J, Hidalgo MC (2007) Determination of the geochemical background in a metal mining site: example of the mining district of Linares (South Spain). J Geochem Explor 94:19–29
McLennan SM, Taylor SR, Hemming SR (2006) Composition, differentiation, and evolution of continental crust: constrains from sedimentary rocks and heat flow. In: Brown M, Rushmer T (eds) Evolution and differentiation of the continental crust. Cambridge University Press, Cambridge, pp 92–134
Mongelli G, Boni M, Oggiano G, Mameli P, Sinisi R, Buccione R, Mondillo N (2017) Critical metals distribution in Tethyan karst bauxite: the Cretaceous Italian ores. Ore Geol Rev 86:526–536
Mongelli G, Buccione R, Gueguen E, Langone A, Sinisi R (2016) Geochemistry of the Apulian allochthonous karst bauxite, Southern Italy: distribution of critical elements and constraints on Late Cretaceous Peri-Tethyan palaeogeography. Ore Geol Rev 77:246–259
Mongelli G, Mameli P, Sinisi R, Buccione R, Oggiano G (2021) REEs and other critical raw materials in Cretaceous Mediterranean-type bauxite: the case of the Sardinian ore (Italy). Ore Geol Rev 139:104559
Mongelli G, Sinisi R, Paternoster M, Perri F (2018) REEs and U distribution in P-rich nodules from Gelasian Apulian Tethyan carbonate: a genetic record. J Geochem Explor 194:19–28
Nassif AB, Elnagar A, Shahin I, Henno S (2021) Deep learning for Arabic subjective sentiment analysis: challenges and research opportunities. Appl Soft Comput 98:106836
Nesbitt HW, Young GM (1982) Early Proterozoic climate and plate motion inferred from major element chemistry of lutites. Nature 299:715–717
Ouladmansour A, Ameur-Zaimeche O, Kechiched R, Heddam S, Wood DA (2023) Integrating drilling parameters and machine learning tools to improve real-time porosity prediction of multi-zone reservoirs. Case study: Rhourd Chegga oilfield, Algeria. Geoenergy Sci Eng 223:211511

Parsa M (2021) A data augmentation approach to XGBoost-based mineral potential mapping: an example of carbonate-hosted Zn–Pb mineral systems of Western Iran. J Geochem Explor 228:106811

Salamab-Ellahi S, Taghipour B, Mongelli G (2019) Clayey bauxite from the Irano–Himalayan belt: critical metals, provenance and palaeoclimate in the Upper Cretaceous Semirom ore deposit, Zagros Mountain, Iran. J Asian Earth Sci 172:126–142

Saporetti CM, da Fonseca LG, Pereira E, de Oliveira LC (2018) Machine learning approaches for petrographic classification of carbonate-siliciclastic rocks using well logs and textural information. J Appl Geophys 155:217–225

Schratz P, Muenchow J, Iturritxa E, Richter J, Brenning A (2019) Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol Modell 406:109–120

Silva dos Santos V, Gloaguen E, Hector Abud Louro V, Blouin M (2022) Machine learning methods for quantifying uncertainty in prospectivity mapping of magmatic-hydrothermal gold deposits: a case study from Juruena mineral province, northern Mato Grosso, Brazil. Minerals 12(8):941

Sofis N, Ntouros EP, Kalaitzidis S (2025) REE Mineralogy and Geochemistry of the Lower Karstic Bauxite Strata (b1), in the Parnassos-Ghiona Unit. Greece Minerals 15(8):804

Stöcklin J (1968) Structural history and tectonics of Iran: a review. Am Assoc Petr Geol B 52:1229–1258

Sun J, Dang W, Wang F, Nie H, Wei X, Li P, Zhang S, Feng Y, Li F (2023) Prediction of TOC content in organic-rich shale using machine learning algorithms: comparative study of random forest, support vector machine, and XGBoost. Energies 16(10):4159

Tahar-Belkacem N, Ameur-Zaimeche O, Kechiched R, Ouladmansour A, Heddam S, Wood DA, Buccione R, Mongelli G (2024) Machine learning models to predict rare earth elements distribution in Tethyan phosphate ore deposits: geochemical and depositional environment implications. Geochemistry 126128

Wang L, Huang X, Yu Y, Zhao L, Wang C, Feng Z, Cui D, Long Z (2017) Towards cleaner production of rare earth elements from bastnaesite in China. J Clean Prod 165:231–242

Xu Y, Zuo R (2024) Geochemical survey data cube: a useful tool for lithological classification and geochemical anomaly identification. Geochemistry 84(2):125959

Yanfei XIAO, Huang L, Zhiqi LONG, Zongyu FENG, Liangshi WANG (2016) Adsorption ability of rare earth elements on clay minerals and its practical performance. J Rare Earths 34(5):543–548

Ye M, Zhu L, Li X, Ke Y, Huang Y, Chen B, Yu H, Li H, Feng H (2023) Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data. Sci Total Environ 858:159798

Zhang BY, Li MY, Li WX, Jiang ZW, Khan U, Wang LF, Wang FY (2021) Machine learning strategies for lithostratigraphic classification based on geochemical sampling data: a case study in area of Chahanwusu River, Qinghai Province, China. J Cent South Univ 28(5):1422–1447

Zhang H, Xie M, Dan S, Li M, Li Y, Yang D, Wang Y (2024) Optimization of feature selection in mineral prospectivity using ensemble learning. Minerals 14(10):970

Zhang W, He Y, Wang L, Liu S, Meng X (2023) Landslide Susceptibility mapping using random forest and extreme gradient boosting: a case study of Fengjie, Chongqing. Geol J 58(6):2372–2387

Zuo R, Wang J, Xiong Y, Wang Z (2021) The processing methods of geochemical exploration data: past, present, and future. Appl Geochem 132:105072

Zuo R, Xiong Y, Wang J, Carranza EJM (2019) Deep learning and its application in geochemical mapping. Earth-Sci Rev 192:1–14