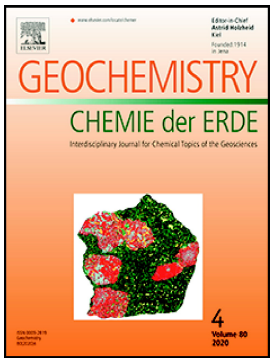
Journal Pre-proof



Machine learning models to predict rare earth elements distribution in Tethyan phosphate ore deposits: Geochemical and depositional environment implications

Nasreddine Tahar-Belkacem, Ouafi Ameur-Zaimeche, Rabah Kechiched, Abdelhamid Ouladmansour, Salim Heddam, David A. Wood, Roberto Buccione, Giovanni Mongelli

PII: S0009-2819(24)00052-7

DOI: https://doi.org/10.1016/j.chemer.2024.126128

Reference: CHEMER 126128

To appear in: *Geochemistry*

Received date: 25 February 2024

Revised date: 19 April 2024

Accepted date: 3 May 2024

Please cite this article as: N. Tahar-Belkacem, O. Ameur-Zaimeche, R. Kechiched, et al., Machine learning models to predict rare earth elements distribution in Tethyan phosphate ore deposits: Geochemical and depositional environment implications, *Geochemistry* (2023), https://doi.org/10.1016/j.chemer.2024.126128

Machine Learning Models to Predict Rare Earth Elements Distribution in Tethyan Phosphate Ore Deposits: Geochemical and Depositional Environment Implications

Nasreddine Tahar-Belkacem[a], Ouafi Ameur-Zaimeche[a]*, Rabah Kechiched[a], Abdelhamid Ouladmansour[b], Salim Heddam[c], David A. Wood[d], Roberto Buccione[e], Giovanni Mongelli[e]

[a] Laboratoire des Réservoirs Souterrains : Pétroliers, Gaziers et Aquifères, Université Kasdi Merbah Ouargla, 30000, Algeria
E-mail: taharbelkacem.nasraddine@univ-ouargla.dz (N. Tahar-Belkacem)
https://orcid.org/0009-0000-4556-5521
E-mail : ouafigeology@gmail.com (O. Ameur-Zaimeche)
https://orcid.org/0000-0003-1886-3034
E-mail: kechiched.rabah@univ-ouargla.dz (R. Kechiched)
https://orcid.org/0000-0002-4729-7126
[b] Department of earth and Geo-environmental sciences, University of Bari, Italy.
E-mail : abdelhamid.ouladmansour@uniba.it (A. Ouladmansour)
https://orcid.org/0000-0002-0522-4316
[c] Faculty of Science, Agronomy Department, University 20 Août 1955 Skikda, Route El Hadaik, BP 26, Skikda, Algeria. E-mail: heddamsalim@yahoo.fr (S. Heddam)
https://orcid.org/0000-0002-8055-8463
[d] DWA Energy Limited, Lincoln, LN5 9JP, United Kingdom
E-mail: dw@dwasolutions.com (D.A. Wood)
orcid.org/0000-0003-3202-4069
[e] Department of Sciences, University of Basilicata, Viale dell'Ateneo Lucano 10, 85100 Potenza, Italy
E-mail : roberto.buccione@unibas.it (R. Buccione)
https://orcid.org/0000-0003-4264-531X
E-mail : giovanni.mongelli@unibas.it (G. Mongelli)
https://orcid.org/0000-0001-8411-2154

*Corresponding author: ouafigeology@gmail.com

Abstract

The global market for rare earth elements (REE) is growing rapidly, driven by rising demand and limited production sources, prompting interest in recovering REE from secondary sources such as phosphate deposits. The Tethyan belt, extending across North Africa and the Middle East contains substantial Upper Cretaceous to Eocene REE-rich phosphorite deposits but with limited geochemical data available. This study provides a novel machine-learning (ML) method to predict REE contents in these deposits and verify a useful geochemical classification based on the concentrations of nine major element oxides. Four ML models are developed to achieve this: eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Regression (SVR), and Decision Tree (DT). The datasets are divided geochemically into oxic and sub-oxic patterns and these are evaluated with the ML models separately and in combination to accurately predict light REE (LREE), heavy REE (HREE), and total REE contents (∑REE). For the oxic pattern dataset, Fe_2O_3 and K_2O exhibit the highest feature importance consistent with a glauconite influence. For the sub-oxic pattern dataset, MnO and SiO_2 exhibit the highest feature importance consistent with high

terrigenous inputs (MnO), and silicification. The ML results support the importance of the local deposition environment in determining REE distributions in these deposits. Paleogeography, ocean-margin tectonics, sea-level oscillations, and marine currents exert influence on the local depositional environments. The eXtreme Gradient Boosting model generates the lowest REE prediction errors for all the datasets evaluated.

Keywords

Tethyan phosphorites; Oxic and sub-oxic REE patterns*; glauconite influence; depositional/ tectonic environments; carbonate-fluorapatite crystallo-chemical control.*

Highlights

- ML prediction using major-element inputs accurately predict REE concentrations
- Fe_2O_3 and K_2O exert highest feature influence in ∑REE for Oxic-pattern deposits
- MnO and SiO_2 exert highest feature influence in ∑REE for sub-Oxic-pattern deposits
- Paleogeography, geodynamics and depositional environment are key influences
- ML based REE predictions offers cost-effective classification of phosphorites

1. Introduction

Critical raw materials are defined by two factors: economic importance and security of supply risks (European Commission, 2020)**.** Rare earth elements (REE) qualify as critical materials because they are essential components in various crucial technologies used extensively in the energy, telecommunications and electronics, defense, aerospace, and transportation sectors (Humphries, 2010; Fortier et al., 2018)**.** REE are comprised of seventeen chemical elements with atomic numbers between 57 (lanthanum; La) to 71 (lutetium; Lu), (Long et al., 2010). Their global market continues to grow rapidly, reaching $7.29 billion with a 10.8% compound annual growth rate (CAGR) in 2023 (Research and Markets, 2023)**.** In addition to their industrial applications, REE have geochemical applications for determining depositional conditions and as dependable proxies for defining redox and diagenetic conditions (Picard et al., 2002; Martin and Scher, 2004; Baioumy et al., 2014; Kechiched et al., 2018).

Ore-grade concentrations of REE are typically associated with bastnäsite, xenotime, monazite, synchysite, eudialyte, loparite, and parisite minerals as well as ion-adsorbing clay minerals (Castor, 2008; Yanfei et al., 2016; Wang et al., 2017; Mongelli et al., 2018; Ahmadinejad and Mongelli, 2022). However, such conventional REE ore reserves are decreasing due to high demand and monopolized production. This has stimulated increased

interest in recovering REE from alternative sources including sedimentary deposits (Condie, 1991; Emsbo et al., 2015; Zhou et al., 2016; Mastalerz et al., 2020). Recently, coal has been proposed as a novel source of REE (Mastalerz et al., 2020; Huang et al., 2021; Ji et al., 2022). Phosphate-rich deposits offer an alternative REE resource because of their REE uptake capability during syn- and post-depositional processes (Emsbo et al., 2015; Kechiched et al., 2016, 2018; Ferhaoui et al., 2022; Linares et al., 2023).

The Tethyan province contains major phosphorite deposits formed episodically throughout the Phanerozoic epoch (Glenn et al., 1994) but particularly during the Upper Cretaceous to Eocene periods. From a geodynamic standpoint, when convergence occurred between the Eurasian and the Afro-Arabian plates, it caused the neo-Tethys Ocean to close (Brew et al., 2001; Sharland et al., 2001; Haq and Al-Qahtani, 2005; Y. Zhang et al., 2021; Ji et al., 2022). This event paved the way for the buildup of huge phosphorite concentrations throughout the Tethys domain (Sheldon, 1988; Abed, 1989; Dill et al., 2008; Dong et al., 2019). This phosphogenic zone contains over 50% of the globe's phosphorite reserves (Notholt et al., 1989; Jasinski, 2011), and stretches from the Middle East through southeast Turkey and across northern Africa (Halalat and Bolourchi, 1994; Baioumy and Farouk, 2022). These regionally distributed phosphorite deposits contain economically viable amounts of REE (Baioumy, 2011; Garnit et al., 2017; Kechiched et al., 2020; Buccione et al., 2021; Ahmed et al., 2022; Ferhaoui et al., 2022).

Machine learning (ML) algorithms have been applied in the field of geosciences since the 1950s (Dramsch, 2020). ML applications have advanced significantly in recent decades, outperforming conventional approaches in terms of accommodating multiple geological variables and approximating complex nonlinear interactions among geological parameters (Cracknell and Reading, 2014; Li et al., 2017; Saporetti et al., 2018; Scott B, 2019; Dumakor-Dupey and Arya, 2021; Zuo et al., 2021). Iron (P. Zhang et al., 2023), copper (Esmaeiloghli et al., 2023), and gold (Silva dos Santos et al., 2022) deposits are the most analyzed ore deposits using ML techniques (Dumakor-Dupey and Arya, 2021). There is a wide range of supervised ML algorithms available but recent published applications suggest that the most accurate and successful algorithms adopted in geochemistry investigation are extreme gradient boosting (XGboost) and random forest (RF) (Parsa, 2021; B. Zhang et al., 2021; Ibrahim et al., 2022, 2023; Chen et al., 2023; Ye et al., 2023; Zhang et al., 2023). These two ML models are applied in the present investigation together with decision tree (DT) and support vector regression (SVR) models.

There is an increased demand for rapid and dependable REE distribution estimates in geological systems. Unfortunately, traditional laboratory methods, such as inductively coupled plasma mass spectrometry (ICP-MS) are time-consuming and expensive to apply. Moreover, ICP-MS typically provides only a few point measurements that may not accurately

represent spatial compositional diversity throughout a deposit. A recent study has demonstrated the value of Al_2O_3, Fe_2O_3, TiO_2, and SiO_2 as input variables in a data-centric ML method to effectively predict the distribution of HREE within karst bauxite deposits in southern Italy (Buccione et al., 2023). A novel connection has also been identified between the TiO_2 content and the provenance and regional tectonic settings of Tethyan bauxite deposits using RF and logistic regression ML models (Zhou et al., 2023). Additionally, nine major oxides were used with ML models to determine the amount of REE within ocean island basalts (Hong et al., 2019). Major elements have also been used to predict LREE patterns in the Choghart iron oxide-apatite deposit (Iran) (Zaremotlagh and Hezarkhani, 2017).

The objectives of this study are to develop accurate, efficient prediction models of REE distributions for existing and new phosphorite exploration programs, utilizing commonly measured and easily accessible geological databases with a particular emphasis on major element distributions. To achieve this a novel ML approach is developed for predicting REE contents at a regional scale in Tethyan phosphorite deposits from North Africa and the Middle East using SiO_2, MnO, Al_2O_3, Fe_2O_3, MgO, CaO, K_2O, TiO_2, and P_2O_5 distributions as inputs.

2. Geological setting

The Tethyan phosphorite belt extends over 7,000 km in length and 800 km in width, oriented east to west (**Figure 1**). The associated geological sequence was deposited during the Late Cretaceous to Eocene interval, spanning from 90 to 45 million years ago. These sediments are widely distributed occurring in parts of the Middle East, the eastern Mediterranean, north Africa, northern South America, and several Caribbean islands (Bentor, 1980; Notholt, 1985; Glenn et al., 1994; Föllmi, 1996; Lucas and Prévôt-Lucas, 1996). They were deposited in paleo-latitudes of 10–20 N. Given their economic importance and scientific relevance, phosphorite ores from North Africa and the Middle East have attracted considerable research attention. This investigation specifically concentrates on phosphorite deposits identified in Morocco, Tunisia, Egypt, Jordan, Saudi Arabia, and Turkey, as illustrated in **Figure 1**, with their details summarized in **Table 1**.

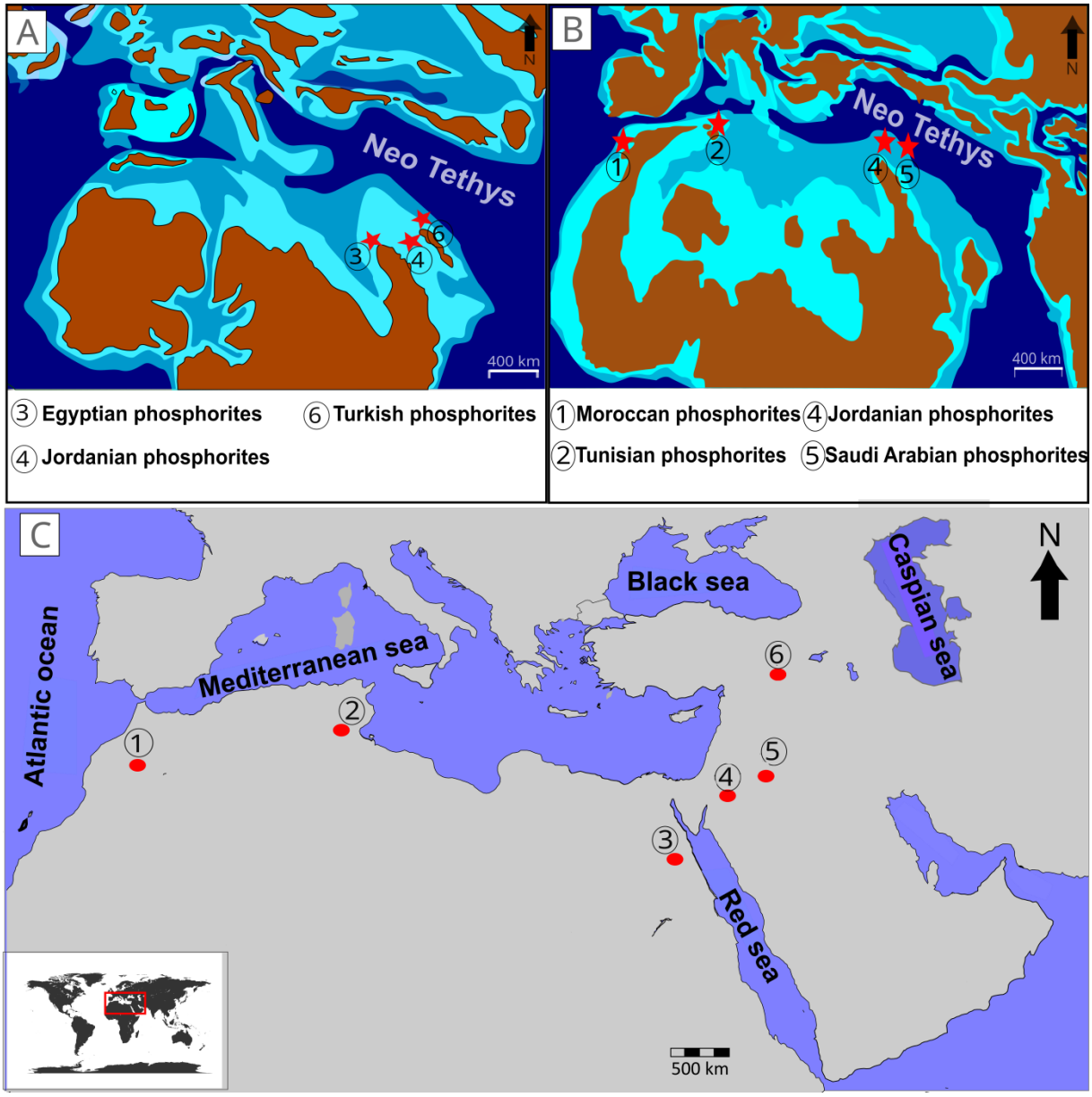


Figure. 1: (A), (B) Paleogeographic map depicting the extent and paleo locations of the Tethyan phosphorite deposits studied; **(A)** during the Upper Cretaceous; **(B)** during the Early Eocene. Maps modified after https://deeptimemaps.com; **(C)** Actual locations map of studied areas. More details about deposits are shown in **Table 2.**

2.1 Moroccan deposits

The Moroccan phosphate deposit was formed during the late Cretaceous-Paleogene epoch (Trappe, 1991; El Chellai et al., 1995; Marzoqi and Pascal, 2000). During this time, the Atlas phosphate basins were situated to the south of the Anti-Atlas paleo high. To the north and east, these basins were bounded by the Variscan basement of the western Meseta (Herbig and Trappe, 1994). The thickening of the geological formations toward the west

(Boujo, 1976) was a result of increased subsidence close to the passive plate-margin edge (El Bamiki et al., 2020).

In a comprehensive sedimentological investigation carried out by (El Bamiki et al., 2020) nine phosphate sections in the High Atlas region of Morocco were thoroughly studied. The current study focuses on five of these areas (Amizmiz, Talentloute, Moulay Brahim, Ait Fars, and Taifest) for which comprehensive data are available.

2.2 Tunisian deposits

The tectonic setting of the late Cretaceous had a significant impact on the development of phosphorites in Tunisia. In the Late Paleocene to Early Eocene, this framework led to a paleogeography that is represented by a shallow marine margin close to the open Tethys Ocean (Garnit et al., 2012). Three prominent basins are situated on the outer borders of Kasserine Island: the northern basin, which has a shallow platform setting; the Gafsa-Metlaoui; and, the eastern basins (Meknassy-Mezzouna basins), which are situated to the south and east of Kasserine Island, respectively. These basins correspond to an environment that was more basinal than coastal (Belayouni et al., 1990; Garnit et al., 2012; Kocsis et al., 2014; Garnit et al., 2017).

2.2.1 Northern Basin

These incorporate the Sra Ouertane and Sekarna deposits in which phosphorites accumulated in the embayment situated between Kasserine Island and the Algerian Promontory over the Late Paleocene and Early Eocene, whereas, to the north and northeast, the successions consists mostly of marls and shales with reduced phosphorite content and an increased presence of glauconite. These changes are indicative of less constrained, deeper water, and better-oxygenated environments (Zaïer et al., 1998).

2.2.2 Eastern and southern basins:

These basins include the Naguess, central Kef Eddour, Table Metlaoui, Mzinda, and Jellabia, Jebel Jebs, Meknassy and Gafsa-Metlaoui sites. These deposits are associated with more basinal conditions characterized by a sub-oxic to reducing depositional environment (Garnit et al., 2017).

2.3 Egypt deposits

The exposed phosphorite layers within the Duwi Formation are conformably overlain by the marine laminated Maastrichtian grey to black shales and marls of the Dakhla Formation. This unit unconformably overlays the Campanian Qusseir shale member marked by an erosional discontinuity contact. The Duwi Formation in the Sibaiya region primarily

consists of gray-colored non-oxidized phosphorite, with other areas also exhibiting non-oxidized phosphorite. Oxidized phosphorite has coarse to fine grain sizes and is relatively hard. In general, thin black shale, dolomitic lenses, and oyster phosphatic limestone strata are mixed with Sibaiya phosphorites (El-Shazly et al., 2022). According to geological estimates, the Red Sea region may contain as much as 50 million tons, the Nile Valley at Sibaiya up to 70 million tons, and the Abu Tartur region potentially holds as much as 7,000 million tons (N.M.A. Report, 2010)

(Banerjee et al., 2020) categorized the phosphorite deposits within the Abu-Tartur basin, specifically the Duwi Formation, into three members: Abu-Tartur phosphorites, Liffiya-Shale, and Maghrabi-Shale. These members collectively exhibit a total thickness of 36 meters. The lowermost member, approximately 11 meters thick, comprises lower phosphorite strata at its base and a dark grey layer characterized by intercalations of shale and siltstone toward the top. This layer also includes economically valuable phosphorite beds at its base, making it the most significant member in the Abu-Tartur area. (Baioumy, 2002; Essa et al., 2016).

2.4 Jordanian deposits

The phosphorite ores in Jordan are a portion of the Al-Hisa phosphorite Formation (AHP) within the Belqa Group. These formations have their origins in the Late Cretaceous-Eocene period (Amireh et al., 2019). The Late Cretaceous-Eocene Belqa group, which disconformably overlies the Ajlun Group, is distinguished by an abundance of layered chert, phosphorite, and organic-rich deposits. The retreat of the Neo-Tethys from the majority of the eastern part of the Mediterranean marks the top of the group and is associated with the end of marine deposition in Jordan during the Late Eocene (Powell, 1989; Amireh et al., 2019).

(Amireh et al., 2019) focused on three areas, which provide the basic data used in this study: the Batn El-Ghoul escarpment on the Ma'an-Mudawwara highway; Nagb Etayyeg, 35 km SE of the Eshidiya mine; and Zgaimat Al-Hasah.

2.5 Saudi Arabian deposits

The Hazm Al-Jalamid area is linked to the northern region's greatest phosphorite deposit. It possesses estimated reserves of 213 million tons with an average P_2O_5 amount of 21% (Collenette and Grainger, 1994). The Umm Wu'al region hosts the second major phosphorite deposit in the north of the country with potential reserves of 537 million tons and 19.35% of P_2O_5 as an average concentration (Ahmed et al., 2022). The Paleocene-aged Jalamid Formation, Lower Eocene Mira Formation, and Upper Eocene Umm Wu'al Formation are the three principal formations that make up the Turayf Group, which hosts nearly all potential phosphorite beds (Riddler et al., 1989), and each formation has a

phosphorite member at its base. The Thaniyat, Ghinnah, and Arqah phosphorite members are each referred to as the formations' basal phosphorite members.

Data from two well-studied phosphorite localities are used in this study: the Hazm Al-Jalamid and Umm Wu'al formations (Ahmed et al., 2022). From a lithostratigraphic standpoint, the "*Phosphorite strata in the two areas are mainly formed from phosphatic materials, which occur mainly as grain-supported phosphatic pellets with apparent internal structure, implying deposition in fluctuating hydrodynamic circumstances*" (Ahmed et al., 2022). The remains of the phosphatic constituents, which are found as entire or fragmented teeth and bone fragments, are formed by bioclasts or skeletal particulates.

2.6 Turkey deposits

The investigated phosphates are located on the Arabian platform, one of Turkey's major tectonic units. (Şengör and Yilmaz, 1981). From the bottom upward, the Mazda phosphorites appear on the moderately dipping northern edge of the Derik anticline (Berker, 1989; Ghasemian et al., 2022) and include the Tasit phosphate (Turonian), the Kasrik phosphate (East Kasrik), the Şemikan phosphate (West Kasrik), and the Akras phosphate (Maastrichtian). The Şemikan member contains the most phosphorites, including 25%-32% of P_2O_5 contents (Berker, 1989).

The Şemikan phosphate ore, used in this study, is one of the Mazda phosphate formations with the highest economic potential. It is divided into four sections. In these sections, the average phosphate thickness is roughly 0.75 cm. The region's phosphorite level is marked by interlayers that alternate with white-cream-colored chert and carbonate bands. While the oolitic phosphate layer at this level is occasionally 20-30 cm thick. The entire cherty, calcareous, and oolitic phosphate level in some areas reaches up to 4 m thick. The persistence of each layer varies over the entire area, and the overall slope of the platform varies from ~3^o- to ~5^o, with an N-NE dip direction (Gundogar and Sasmaz, 2022).

3. Material and methods

3.1 Data collection

In the database compiled for this study, only those elements for which comprehensive analytical records exist are considered for ML modeling. Dataset preprocessing was performed to remove any samples with less than 18 wt% of P_2O_5 to only retain pure phosphorite *sensu stricto*, samples and avoid the inclusion of less concentrated phosphatic rocks samples (Nathan, 1984; Slansky, 1986; Jarvis et al., 1994). Hence, the compiled

dataset includes samples from six distinct phosphorite deposits belonging to the Tethyan phosphate province **(Figure 1)**.

Using whole rock analysis, 161 samples contain analyzed concentrations of nine major elements: SiO_2, MnO, Al_2O_3, Fe_2O_3, MgO, CaO, K_2O, TiO_2, and P_2O_5. Additionally, each data record includes REE (La to Lu) concentrations that are configured as the dependent variable for ML analysis. Geographically, the phosphorite sample numbers are distributed: Moroccan (15), Tunisian (45), Egyptian (43), Jordanian (5), Saudi Arabian (45), and Turkish (8) (**Table 1)**. Major element concentrations were obtained, according to the data source, mostly through inductively coupled plasma–atomic emission spectrometry (ICP-AES) and X-ray fluorescence (XRF), while inductively coupled plasma mass spectrometry (ICP-MS) was used to analyze REE concentrations.

In terms of analysis accuracy, the precision of major elements and REE across the chosen samples from source data, determined through replicate analyses, indicates that major elements show precision exceeding ±0.5 %, whereas REE exhibit variability spanning from 2 % to 5 %. Detection limits for most major elements stand at 0.01%, except for MnO and TiO_2, which are limited to 0.001%. Additionally, the detection limits for REE are measured at 0.04 ppm. Note that some phosphorite locations, including Algerian, Iraqi, and Syrian are excluded from the dataset compiled due to the limited set of major elements analyzed.

3.2 Dataset characterization combined with geochemical classification

Our primary goal is to provide a better understanding of REE distributions in Tethyan phosphorites on a regional scale. A geochemical-based classification approach is developed for this purpose, which can distinguish three distinct REE patterns: oxic, sub-oxic, and combined classes, which consist of 78, 83, and 161 samples, respectively. The combined REE pattern includes phosphorites displaying both oxic and sub-oxic patterns. The sample REE values are normalized to values of the Post-Archean Australian Shale (PAAS) standard (Taylor and McLennan, 1985; Mclennan, 1989). The normalized geochemical REE signatures exhibit two distinctive patterns reflecting the redox conditions of their depositional environments: a linear pattern and a slightly hat-shaped pattern. **Tables 1** and **2**, describe the linear (oxic seawater) pattern characteristic of Moroccan, northern Tunisian, Jordanian, Saudi Arabian, and Turkish phosphorite deposits, and the slightly hat-shaped (sub-oxic) pattern characteristic of Southern Tunisian, and Egyptian phosphorite deposits (**Figure 2).**

The oxic (***seawater-like) pattern*** displays a relative enrichment in heavy REE (HREE) and a substantial depletion in cerium (Ce) (Bertram et al., 1992; Sholkovitz et al., 1994; Zhang and Nozaki, 1996; Nozaki et al., 1997; Shields and Webb, 2004). This REE pattern is caused by two main processes: (1) increased carbonate complexation of HREE in

saltwater and preferential adsorption of light REE (LREE) onto particles (Ohta and Kawabe, 2000; Tang and Johannesson, 2010); and, (2) a decrease in particle adherence across the REE series, which results in less HREE removal (Sholkovitz et al., 1994). The preferential removal of Ce and LREEs by oxyhydroxide particles therefore is responsible for the characteristic oxic pattern observed (Reynard et al., 1999; Sholkovitz et al., 1994).

The Sub-oxic (hat-shaped) pattern in phosphorites is marked by mid-atomic-weight REE (MREE) enrichment and relative HREE depletion, combined with a slight Ce anomaly (Alibo and Nozaki, 1999; Zhu et al., 2014). Multiple investigations have connected this pattern to diagenetic processes (Reynard et al., 1999; Shields and Stille, 2001; Shields and Webb, 2004; Bright et al., 2009; Kechiched et al., 2018). Given that Ca and MREE share analogous ionic radii, MREE ions preferentially substitute for Ca^{2+} during the diagenetic recrystallization of apatite (Abedini and Calagari, 2017; Y. Zhang et al., 2021). Furthermore, MREE enrichment has also been linked to the selective adsorption of MREE by certain organisms (Wright et al., 1987). The Fe-Mn redox cycle, as well as extensive exposure of porous particles to REE uptake from seawater, may also lead to MREE enrichment (Auer et al., 2017). The latter interpretation allows for these particles to be allochthonous or para-autochthonous (reworked phosphorites) (Auer et al., 2017). Moreover, post-depositional glauconitization of reworked phosphatic particles in semi-restricted, micro-environments in the presence of dissolved Fe^{2+} can lead to REE enrichment, with MREE-enriched patterns (Kechiched et al., 2018).

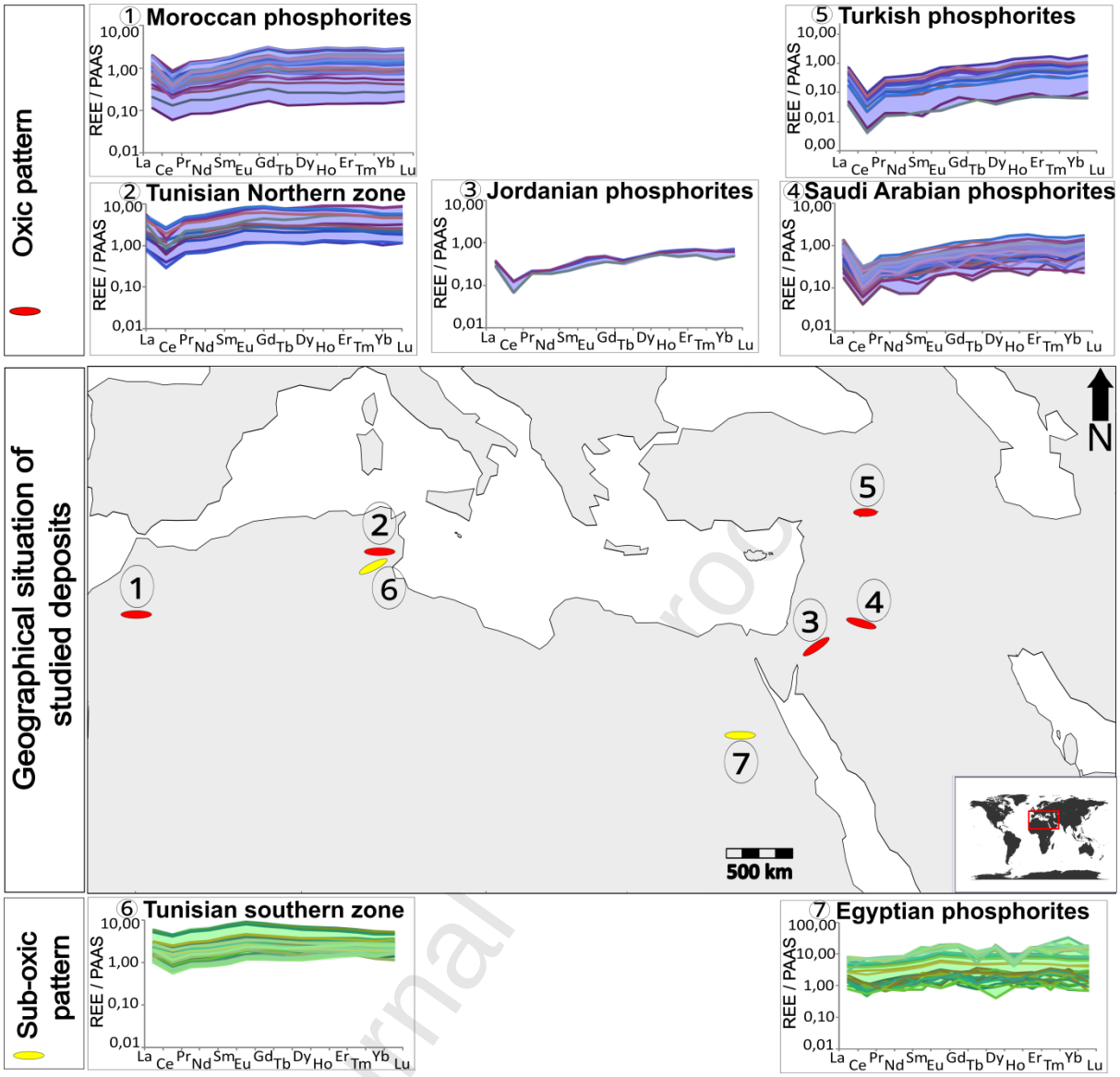


Figure. 2. REE geochemical signatures of Tethyan phosphorites and their geographical distributions.

Table 1. Summary of Tethyan phosphorites studied, with their country, locality, age, geochemical signature, and average content of P_2O_5 (w%), and ∑REE (ppm).

Country	Locality	Age	Geochemical Signature	P_2O_5	∑REE	References
Morocco (15)	Amizmiz	Up. Cretac./ Pal.	Oxic.p	22.12	181.64	(El Bamiki et al., 2023)
	Moulay Brahim	Up. Cretac./ Pal.	Oxic.p	19.84	108.5	
	Talentloute	Up. Cretac./ Pal	Oxic.p	19.06	149.1	

	Taifest	Up. Cretac./ Pal.	Oxic.p	20.38	159.71	
Tunisia (45)	Sra Ouartane	Pal/ Eocene	Oxic.p	24.03	545.65	(Garnit et al., 2017)
	Sekarna	Pal/ Eocene	Oxic.p	20.88	715.9	
	Jbel Jebs	Pal/ Eocene	Sub-oxic. p	28.7	973.83	
	Naguess	Pal/ Eocene	Sub-oxic. p	27.15	345.39	
	Central Kef Eddour	Pal/ Eocene	Sub-oxic. p	26.53	320.41	
	Table Metlaoui	Pal/ Eocene	Sub-oxic. p	26.24	325.49	
	Mzinda	Pal/ Eocene	Sub-oxic. p	25.47	296.24	
	JELLABIA	Pal/ Eocene	Sub-oxic. p	27.91	335.37	
Egypt (43)	Sibaiya Duwi Nile Valley	Up. Cretac.	Sub-oxic. p	25.21	228.82	(El-Shazly et al., 20
	Abu Tartur	Up. Cretac.	Sub-oxic. p	25.87	1266.33	(Awadalla, 2010; Edress et al., 2023)
Jordan (5)	Bat El Ghoul	Up. Cret./ Eoc	Oxic.p	21.83	65.17	(Amireh et al., 2019)
	Zgaimat El Hasah	Up. Cret./ Eoc	Oxic .p	21.8	35.45	
Saudi Arabia (45)	Hazm Al-Jalamis Phosphorites.	Paleocene	Oxic.p	23.4	48.09	(Ahmed et al., 2022)
	Umm Wu'al Phosphorites.	Eocene	Oxic.p	27.13	83.92	
Turkey (8)	Mazıdağı Phosphates,	Up. Cretac.	Oxic.p	28.59	27.58	(Gundogar and Sasmaz, 2022)

3.3 Data processing

The compiled geochemical datasets were formatted as a two-dimensional matrix with each row representing a sample and each column representing a geochemical feature (measured major-element and ∑REE concentrations). The datasets were screened to identify

and remove any data records with missing feature values. The filtered datasets were used for ML modeling and are statistically summarized in **Table 2**.

Table 2. Statistical summary of major-element (wt%) and REE (ppm) concentrations by country.

Location	Patterns	Stat	SiO_2	MnO	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	TiO_2	P_2O_5	∑LREE (La-Eu)	∑HREE (Gd-Lu)	∑REE
MOROCCO	Oxic	**Mean**	**17.72**	**0.001**	**1.19**	**0.5**	**1.94**	**39.94**	**0.27**	**0.07**	**21.12**	**129.06**	**29.46**	**158.53**
		std	4.37	0.004	0.44	0.25	0.97	1.95	0.08	0.02	1.97	43.24	10.5	53.36
		Min	11.24	0	0.66	0.27	0.87	36.38	0.15	0.04	18.46	53.08	11.67	64.75
		Max	27.15	0.010	2.07	0.99	3.96	42.39	0.43	0.11	24.4	218.59	51.05	266.79
		Skewness	0.59	2.40	1.06	0.84	0.69	-0.36	0.74	0.98	0.28	0.76	0.94	0.79
		Kurtosis	0.02	4.35	0,20	-0.65	-0.76	-1.04	-0.07	-0.26	-1.02	0.56	0.71	0.54
TUNISIA	Oxic	**Mean**	**13.53**	**0.015**	**0.89**	**1.35**	**2.38**	**43.24**	**0.33**	**0.04**	**22.14**	**532.22**	**115.58**	**647.8**
		std	6.4	0.010	0.33	0.84	2.18	5.03	0.17	0.01	2.89	137.02	25.14	157.31
		Min	8.23	0.002	0.48	0.42	0.23	38.18	0.11	0.04	19.14	318	82.8	400.8
		Max	22.45	0.025	1.22	2.17	5.12	48.6	0.5	0.05	26.12	669.5	144.9	799.7
		Skewness	0.83	-0.26	-0.18	-0.43	0.17	0.21	-0.53	1.42	0.46	-1.04	-0.33	-1.13
		Kurtosis	-1.90	-1.97	-2.12	-3.09	-2.23	-3.00	2.15	2.07	-1.30	1.03	-1.58	0.93
	Sub-oxic	**Mean**	8.56	0.00	1.11	0.49	0.84	44.90	0.21	0.05	26.78	324.36	49.09	373.44
		std	4.33	0.00	0.38	0.17	0.50	2.35	0.15	0.02	1.85	176.35	22.33	198.36
		Min	2.64	0.00	0.43	0.16	0.39	39.04	0.06	0.02	20.81	128.20	22.60	150.80
		Max	21.96	0.01	2.07	0.91	3.17	49.22	0.69	0.10	30.32	895.20	123.20	1018.40
		Skewness	1.19	0.69	0.59	0.51	3.12	-0.46	2.07	0.99	-0.77	2.08	2.10	2.09
		Kurtosis	1.24	-0.06	0.05	0.30	12.23	0.07	4.32	1.20	1.52	4.65	4.69	4.69
EGYPT	Sub-oxic	**Mean**	**6.33**	**0.117**	**1.39**	**4.18**	**3.3**	**40.15**	**0.18**	**0.04**	**25.54**	**631.2**	**128.44**	**759.64**
		std	3.96	0.102	0.51	1.06	2.75	4.01	0.09	0.02	2.96	477.15	107.75	583.29
		Min	1.03	0.020	0.31	2.4	0.28	30	0.07	0.01	19.2	110.7	16.3	127
		Max	13.5	0.610	2.6	7.69	9.5	46.2	0.41	0.09	30.96	1332	331	1606
		Skewness	0.06	3.09	0.16	0.69	0.63	-0.64	1.27	0.72	-0.26	0.36	0.53	0.38
		Kurtosis	-1.52	14.38	-0.16	1.13	-0.71	-0.19	1.68	-0.35	-0.59	-1.68	-1.43	-1.66
JORDAN	Oxic	**Mean**	**24.1**	**0.010**	**0.17**	**0.18**	**0.26**	**41.34**	**0.04**	**0.02**	**21.82**	**41.65**	**11.63**	**53.28**
		std	11.46	0	0.14	0.09	0.12	7.52	0.04	0.01	1.96	28.9	6.98	35.82
		Min	8.66	0.010	0.05	0.09	0.15	34.1	0.01	0.01	18.8	24.5	6.33	30.83
		Max	34.6	0.010	0.4	0.33	0.44	51.2	0.11	0.04	23.7	92.94	23.8	116.74
		Skewness	-0.71	/	1.62	1.23	0.94	0.60	1.81	0.54	-1.07	2.15	1.95	2.13
		Kurtosis	-2.20	/	2.91	1.18	0.63	-2.33	3.25	1.49	0.38	4.68	4.06	4.63
SAUDI ARABIA	Oxic	**Mean**	**1.84**	**0.020**	**0.25**	**0.16**	**0.9**	**51.75**	**0.04**	**0.03**	**25.22**	**53.31**	**12.3**	**65.6**
		std	1.44	0.01	0.58	0.11	1.25	2.54	0.03	0.01	4.65	19.92	3.91	23.53

				3									
		Min	0.35	0	0.01	0.04	0.14	44.4	0.01	0.01	18.48	21.27	4.73	26
		Max	6.35	0.060	4	0.71	5.12	56.9	0.12	0.07	36.7	102.16	25.37	127.53
		Skewness	1.33	0.93	6.41	2.71	1.80	-0.83	0.92	1.28	0.65	0.49	0.95	0.55
		Kurtosis	1.21	0.50	42.27	12.47	2.40	0.88	0.30	1.12	-0.04	-0.51	1.84	-0.24
		Mean	**5.77**	**0.011**	**0.05**	**0.11**	**0.18**	**53**	**0.01**	**0.01**	**28.59**	**20.57**	**7.02**	**27.58**
		std	11.28	0.004	0.05	0.09	0.03	6.51	0	0	3.26	6.25	2.21	8.43
TURKY	**Oxic**	Min	0.53	0.010	0.01	0.01	0.13	37	0.01	0.01	22.3	11.71	3.87	15.58
		Max	33.5	0.020	0.14	0.25	0.22	56.3	0.02	0.01	31.7	31.7	11.3	43
		Skewness	2.75	2.83	1.17	0.75	-0.70	-2.74	1.44	/	-1.18	0.44	0.78	0.53
		Kurtosis	7.67	8.00	0.20	-0.54	0.52	7.63	0.00	/	0.65	0.34	1.60	0.68

Note: The detection limit value has been used to replace values recorded as “below detection limit” (bdl)

3.4 Variable selection

In order to explore the REE distributions in phosphorite deposits (dependent variables), nine major oxides, key components of phosphatic rocks minerals, were chosen as input features (independent variables): SiO_2, MnO, Al_2O_3, Fe_2O_3, MgO, CaO, K_2O, TiO_2, and P_2O_5 (**Table 3**).

Table 3. Statistical summary of three modeled datasets: major-element oxides (wt%) and REE (ppm).

Patterns	**Stat**	**SiO_2**	**MnO**	**Al_2O_3**	**Fe_2O_3**	**MgO**	**CaO**	**K_2O**	**TiO_2**	**P_2O_5**	**∑LREE**	**∑HREE**	**∑REE**
	N	78.00	78.00	78.00	78.00	78.00	78.00	78.00	78.00	78.00	78.00	78.00	78.00
	mean	7.47	0.01	0.45	0.3	1.08	48.39	0.1	0.03	24.36	94.47	21.64	116.11
	std	9.21	0.01	0.63	0.39	1.31	6.34	0.12	0.02	4.44	126.51	27	153.17
	Cv	123.29	100.00	140.00	130.00	121.30	13.10	120.00	66.67	18.23	133.92	124.77	131.93
	min	0.35	0.00	0.01	0.01	0.13	34.10	0.01	0.01	18.46	11.71	3.87	15.58
	25%	1.04	0.01	0.10	0.09	0.18	42.24	0.02	0.02	21.07	34.35	9.39	43.83
	50%	2.74	0.01	0.20	0.20	0.27	51.35	0.04	0.03	23.34	54.46	12.61	67.47
	75%	13.44	0.02	0.61	0.28	1.62	53.27	0.12	0.05	27.59	95.14	23.04	119.74
Oxic pattern													

		34.60					56.90			36.70	669.50	144.90	799.70
	max		0.06	4.00	2.17	5.12		0.50	0.11				
	skew-ness	1.42	1.20	3.02	3.30	1.44	-0.72	1.66	1.32	0.77	3.33	3.26	3.30
	Kurto-sis	1.03	1.42	12.63	12.05	1.14	-0.89	1.84	1.85	0.01	11.26	10.70	10.95
		83.00	83.00	83.00	83.00	83.00	83.00	83.00	83.00	83.00			
	N										83.00	83.00	83.00
							42.44			26.14	483.32	90.2	573.52
	mean	7.86	0.05	1.25	2.4	2.11		0.19	0.04				
											393.95	88.18	480.28
	std	4.17	0.09	0.47	2.01	2.35	4.07	0.12	0.02	2.55			
		53.05	180.00	37.60	83.75	111.37	9.59	63.16	50.00	9.76	81.51	97.76	83.74
	Cv												
							30.00			19.20	110.70	16.30	127.00
	min	1.03	0.00	0.31	0.16	0.28		0.06	0.01				
Suboxic pat-tern							40.00			24.17	201.05	37.00	236.75
	25%	4.37	0.00	0.92	0.52	0.60		0.11	0.03				
							43.15			26.50	299.40	45.40	339.80
	50%	7.26	0.02	1.20	2.81	0.81		0.16	0.04				
		11.00					45.28			27.95	722.75	120.45	852.90
	75%		0.08	1.57	4.21	2.98		0.20	0.06				
		21.96					49.22			30.96	1332.00	331.00	1606.00
	max		0.61	2.60	7.69	9.50		0.69	0.10				
	skew-ness	0.69	3.56	0.48	0.37	1.56	-0.90	2.22	0.60	-0.64	1.15	1.47	1.20
	Kurto-sis	0.43	19.12	-0.03	-1.24	1.48	0.56	6.11	0.01	0.16	-0.21	0.68	-0.10
		161.00	161.00	161.00	161.00	161.00	161.00	161.00	161.00	161.00	161.00	161.00	161.00
	N.												
							45.32			25.28	294.93	56.98	351.92
	mean	7.67	0.03	0.86	1.38	1.61		0.15	0.04				
											353.89	74.28	426.72
	Std.	7.05	0.07	0.69	1.8	1.98	6.06	0.13	0.02	3.69			
		91.92	233.33	80.23	130.43	122.98	13.37	86.67	50.00	14.60	119.99	130.36	121.25
	Cv												
							30.00			18.46	11.71	3.87	15.58
Combined patterns	***Min.***	0.35	0.00	0.01	0.01	0.13		0.01	0.01				
							40.90			22.60	56.17	12.95	69.12
	25%	2.15	0.00	0.20	0.20	0.28		0.04	0.02				
							44.91			25.64	172.60	31.00	196.30
	50%	5.70	0.01	0.87	0.44	0.67		0.12	0.04				
		11.00					51.20			27.86	318.00	51.05	369.20
	75%		0.03	1.29	2.93	2.43		0.20	0.05				
		34.60					56.90			36.70	1332.00	331.00	1606.00
	max.		0.61	4.00	7.69	9.50		0.69	0.11				
	skew-ness	1.51	4.91	0.86	1.31	1.90	-0.04	1.53	0.84	0.18	1.79	2.20	1.85
	Kurto-sis	2.57	35.68	1.49	0.38	3.37	-0.76	3.05	0.59	-0.05	2.19	3.96	2.43

3.5 Major oxides

The P_2O_5 content of the examined phosphorites varies, with the highest average concentration observed in Turkish deposits (28.59 ± 3.26 wt%). Moroccan deposits contain the lowest average of P_2O_5 concentration (21.12 ± 1.97 w%). The most abundant element occurring in phosphorites is CaO. Moroccan and Egyptian phosphorites have the lowest CaO contents (39.94 ± 1.95 wt% and 40.15 ± 4.01 wt%, respectively), whereas Turkish and Saudi Arabian samples have the highest CaO averages (53.00 ± 6.51 wt% and 51.75 ± 2.54 wt%, respectively). Jordanian phosphorites display the highest average SiO_2 content (24.1 ± 11.46 wt%), followed by Moroccan samples (17.72 ± 4.37 wt%). MnO reaches its highest average concentration (0.117 ± 0.102 wt%) in Egyptian samples. MgO content varies greatly, with Turkish deposits having the lowest (0.18 ± 0.03 wt%) and Egyptian samples displaying the highest mean values (3.30 ± 2.75 wt%). Notably, Egyptian phosphorites have highest values of Al_2O_3, Fe_2O_3, and MgO (1.39 ± 0.59 wt%, 4.18 ± 1.06 wt% and 3.30 ± 2.75 wt%, respectively) when compared to other samples. The K_2O content is higher in the central and western Tethyan deposits (Egypt, Northern and Southern Tunisia, and Morocco, with values of 0.18 ± 0.09, 0.33 ± 0.17, 23 ± 0.15, and 0.27 ± 0.08 wt%, respectively). The TiO_2 content tends to decrease from West to East deposits, with values of 0.07 ± 0.02 wt% in Moroccan deposits, 0.04 ± 0.01 to 0.05 ± 0.02 wt% in Tunisian samples, and 0.04 ± 0.02, 0.02 ± 0.01, 0.03 ± 0.01 and 0.01 ± 0.00 wt% in Egyptian, Jordanian, Saudi and Turkish phosphorites, respectively, as indicated in **Table 2**.

Concerning the statistical distribution of the major oxides across all the deposits evaluated skewness and kurtosis values vary from oxide to oxide and region to region. In Moroccan deposits, for instance, SiO_2 demonstrates a skewness of 0.59 and a kurtosis of 0.02, indicating a slightly right-skewed distribution with a near-normal kurtosis. Contrastingly, MnO exhibits a high skewness of 2.40 and a kurtosis of 4.35, suggesting a highly right-skewed distribution with heavy tails. Al_2O_3 displays a skewness of 1.06 and a kurtosis of 0.20, indicating a moderately right-skewed distribution with a slightly heavier tail than normal. Meanwhile, Fe_2O_3 and MgO present skewness values of 0.84 and 0.69, respectively, with corresponding kurtosis values of -0.65 and -0.76, indicating slightly right-skewed distributions with lighter tails than normal. CaO and K_2O exhibit negative skewness values of -0.36 and -0.53, respectively, suggesting moderately left-skewed distributions. TiO_2 and P_2O_5 demonstrate skewness values of 0.98 and 0.28, respectively, with corresponding kurtosis values of -0.26 and -1.02, indicating slightly right-skewed distributions with relatively lighter tails. In Northern Tunisia, significant variations in skewness and kurtosis are observed across major oxides, suggesting diverse distributional characteristics. Similarly, samples from

southern Tunisia, Egypt, Saudi Arabia, and Turkey exhibit notable variations in skewness and kurtosis values, indicating different distribution patterns across these locations..

3.6 Rare earth elements (REEs)

Tethyan phosphorites are known for their substantial REE concentrations. The phosphorites represented in the compiled dataset show an average concentration ranging from a minimum of 27.58 ppm recorded in Turkish phosphorites to maximum average 759.64 ppm in Egypt, with the highest values of ∑REE recorded in the Abu Tartur plateau (western Egypt). The Abu Tartur samples exhibit a maximum content ∑REE of 1266.33 ppm. The Saudi Arabian and Jordanian phosphorites contain low REE concentrations (average = 65.60 ppm and 53.28 ppm, respectively), whereas Moroccan deposits contain moderate REE contents ranging from 64.75 to 266.79 ppm (average = 158.53 ppm). Tunisian phosphorites display an average REE of 647.80 ppm in the northern basins and 403.93 ppm in the southern basins. HREEs (Gd-Lu) achieve the highest average content in Egyptian deposits, with 128.44 ppm (the maximum value = 331.00 ppm recorded in the Abu Tartur plateau), followed by northern Tunisian sites (115.58 ppm of HREEs average content). The LREEs (La-Eu) average contents show the highest concentration in Egyptian ores, reaching 631.20 ppm (with the highest recorded value of all the Tethyan phosphorites sampled reaching 1606 ppm in the Abu Tartur Area). Samples from the Northern Tunisian sites are associated with the next highest LREE contents, with an average content of 532.22 ppm. Saudi Arabian, Jordanian, and Turkish phosphorite samples display average LREE contents of 53.31 ppm, 41.65 ppm, and 20.57 ppm, respectively (**Table 2**).

The statistical examination of skewness and kurtosis for REE also reveals some regional differences. In Morocco, skewness and kurtosis values for LREE, HREE, and REE components remain consistent, ranging from 0.76 to 0.94 and 0.54 to 0.71, respectively. This suggests a relatively symmetric distribution with moderate tail heaviness. In the Northern Tunisian phosphorites, however, the picture changes as skewness varies from -1.04 to -0.33 and kurtosis from 0.93 to 1.03 across REE components, indicating non-normal distributions with varying degrees of asymmetry and tail heaviness. Contrastingly, Southern Tunisia exhibits fairly consistent skewness and kurtosis values across REE components, approximately ranging from 2.08 to 2.10 and 4.65 to 4.69, respectively. These findings suggest highly right-skewed distributions with heavy tails. Conversely, deposits from Egypt, Jordan, Saudi Arabia and Turkey demonstrate relatively consistent skewness and kurtosis values across REE components and locations, with skewness ranging from approximately 0.36 to 2.15 and kurtosis from around -1.68 to 4.68.

3.7 Dataset separation

For application in the ML models, the dataset needs to be split into training and testing subsets. Based on trial-and-error sensitivity analysis, a split of training (80%) and testing (20%), an 80:20 split, was found to deliver the most reliable prediction results.

3.8 Sensitivity analysis and Different combinations of input attributes

Correlation matrices are widely used to characterize feature relationships and interactions (Sadeghi, 2022; Esmaeiloghli, Tabatabaei, & Carranza, 2023; Ribeiro et al., 2023). These are displayed as heat maps (**Figure 3)** with Pearson correlation coefficient value ranges indicated by the size and color of each circle.

In addition to Pearson coefficient correlation, the p-value (probability relating to the null hypothesis) indicates confidence in specific feature correlations (where a p-value of <0.05 is considered significant, and the null hypothesis suggests an inconclusive correlation is rejected). Within the oxic pattern, Fe_2O_3 demonstrates a robust positive correlation with LREE, HREE, and ∑REE (R=0.84, p=0), (R=0.77, p=0), and (R=0.83, p=0) respectively. The second-highest correlating element is K_2O, which exhibits correlations of (R=0.73, p=0), (R=0.69, p=0), and (R=0.72, p=0) with LREE, HREE, and ∑REE, respectively. On the other hand, MnO displays a negative correlation across all three patterns, with less significant R values (R=-0.14, p=0.227), (R=-0.16, p=0.169), and (R=-0.14, p=0.215), respectively. REE correlations with P_2O_5 (R=-0.17, p=0.159), (R=-0.16, p=0.171), and (R=-0.17, p=0.143), respectively, are similar to those with MnO. SiO_2, in contrast, demonstrates consistent positive correlations with all patterns, with values of (R=0.30, p=0.007) for ∑REE and HREE, and (R=0.30, p=0.008) for LREE. Additionally, SiO_2 exhibits a strong inverse correlation with CaO (R=-0.92, p=0) across all patterns.

In the sub-oxic pattern dataset, MnO exhibits a notably strong positive correlation with LREE (R=0.65, p=0), HREE (R=0.68, p=0), and ∑REE (R=0.66, p=0). In contrast, SiO_2 demonstrates significant negative correlations with LREE (R=-0.57, p=0), HREE (R=-0.59, p=0), and ∑REE (R=-0.58, p=0). CaO is positively correlated with P_2O_5 and exhibits a robust negative correlation with MgO (R=-0.82, p=0), Fe_2O_3 (R=-0.69, p=0), and Al_2O_3 (R=-0.71, p=0). Additionally, Fe_2O_3 shows a positive correlation with LREE, HREE, and ∑REE (R=0.36, p=0.001), (R=0.42, p≈0.00), and (R=0.37, p=0.001), respectively. In the sub-oxic pattern dataset, SiO_2 exhibits an inverse relationship, displaying strong negative correlations with LREE, HREE, and ∑REE of (R= -0.57, p=0), (R= -0.59, p=0) and (R= -0.58, p=0), respectively (**Figure 3**).

In the combined pattern dataset, LREE, HREE, and ∑REE exhibit the strongest correlations with MnO, with values of (R=0.65, p=0), (R=0.68, p=0), and (R=0.66, p=0) respectively. Additionally, these REE groupings also show notable correlations with Fe_2O_3,

with values of (R=0.58, p=0), (R=0.59, p=0), and (R=0.58, p=0) respectively. Moreover, SiO_2 exhibits slight negative correlations for LREE, HREE, and ∑REE, with (R= -0.11, p=0.161), (R= -0.13, p=0.102) and (R= -0.11, p=0.148). TiO_2 displays significant positive correlations with K_2O (R=0.51, p=0) and Al_2O_3 (R=0.45, p=0) within the combined pattern dataset (**Figure 3**).

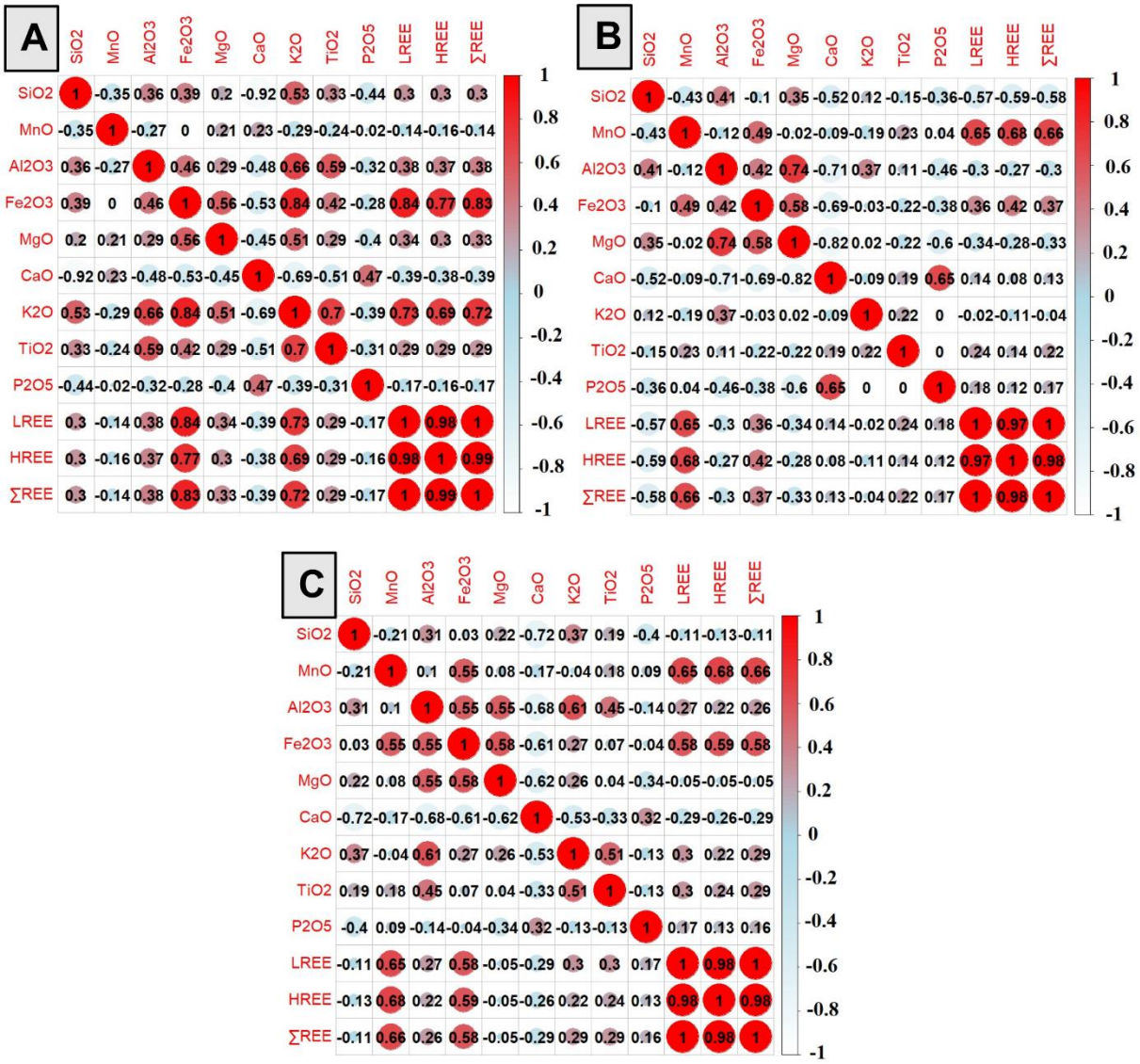


Figure 3. Correlation matrix between input / output variables expressed as a heat map for samples within the: **(A)** Oxic pattern dataset; **(B)** Sub-oxic pattern dataset; and, **(C)** In the dataset combining samples from both patterns.

3.9 Machine learning model configuration

The ML models were configured to assess each of the three pattern datasets using the nine of the major-element concentrations available (SiO_2, MnO, Fe_2O_3, MgO, CaO, Al_2O_3,

K_2O, TiO_2, and P_2O_5) in up to eight different combinations to predict LREE, HREE, ∑REE separately (**Table 4**). This data-centric approach made it possible to systematically develop efficient tools, ensuring high-quality predictions for the oxic, sub-oxic, and combined pattern datasets are generated independently. The four ML models evaluated (DT, RF, SVR, and XGBoost) were each configured to process the three data combinations.

Table 4. Input variable combinations based on feature importance selection.

Pattern	Comb. Ref.	Input Variables (w%)	Output(ppm)
Oxic pattern	1	[' Fe_2O_3', 'K$_2$O','SiO_2','TiO_2']	*LREE/HREE/∑REE*
	2	[' Fe_2O_3',' K_2O ',' SiO_2',' TiO_2','Al_2O_3']	*LREE/HREE/∑REE*
	3	[' Fe_2O_3', K_2O ',' SiO_2',' TiO_2','MnO', ' Al_2O_3']	*LREE/HREE/∑REE*
	4	[' Fe_2O_3',' K_2O ',' SiO_2',' TiO_2', 'P_2O_5', 'MnO', ' Al_2O_3']	*LREE/HREE/∑REE*
	5	[' Fe_2O_3', 'K$_2$O ',' SiO_2',' TiO_2','CaO', 'MgO', 'P_2O_5', 'MnO']	*LREE/HREE/∑REE*
	6	[' Fe_2O_3', 'K$_2$O ',' SiO_2',' TiO_2','CaO', 'MgO','MnO', ' Al_2O_3']	*LREE/HREE/∑REE*
	7	['Fe_2O_3', 'K$_2$O ',' SiO_2',' TiO_2','MgO', 'P_2O_5', 'MnO', ' Al_2O_3']	*LREE/HREE/∑REE*
	8	[' Fe_2O_3', 'K$_2$O ',' SiO_2',' TiO_2','CaO', 'MgO', 'P_2O_5', 'MnO', ' Al_2O_3']	*LREE/HREE/∑REE*
Sub-oxic pattern	1	[' SiO_2','MnO',' Fe_2O_3',' MgO ', 'P_2O_5']	*LREE/HREE/∑REE*
	2	[' SiO_2','MnO',' Fe_2O_3',' MgO ', 'P_2O_5',' K_2O ']	*LREE/HREE/∑REE*
	3	[' SiO_2','MnO',' Fe_2O_3',' MgO ', 'P_2O_5',' TiO_2', ' K_2O ']	*LREE/HREE/∑REE*
	4	[' SiO_2','MnO',' Fe_2O_3',' MgO ', 'P_2O_5','CaO', ' TiO_2', ' K_2O ']	*LREE/HREE/∑REE*
	5	[' SiO_2','MnO',' Fe_2O_3','MgO',' Al_2O_3', 'CaO', ' TiO_2', ' K_2O ']	*LREE/HREE/∑REE*
	6	[' SiO_2','MnO',' Fe_2O_3','MgO',' P_2O_5',' Al_2O_3', 'CaO', ' TiO_2', ' K_2O ']	*LREE/HREE/∑REE*
Combined patterns	1	[' MnO',' SiO_2',' K_2O ',' Fe_2O_3']	*LREE/HREE/∑REE*
	2	[' MnO',' SiO_2',' K_2O ',' Fe_2O_3',' Al_2O_3']	*LREE/HREE/∑REE*
	3	[' MnO',' SiO_2',' K_2O ',' Fe_2O_3','MgO', ' Al_2O_3']	*LREE/HREE/∑REE*
	4	[' MnO',' SiO_2',' K_2O ',' Fe_2O_3',' P_2O_5','MgO', ' Al_2O_3']	*LREE/HREE/∑REE*
	5	[' MnO',' SiO_2',' K_2O ',' Fe_2O_3','CaO',' P_2O_5','MgO', ' Al_2O_3']	*LREE/HREE/∑REE*
	6	[' MnO',' SiO_2',' K_2O ',' Fe_2O_3',' TiO_2','CaO','MgO', ' Al_2O_3']	*LREE/HREE/∑REE*

7	[' MnO',' SiO_2',' K_2O ',' Fe_2O_3',' TiO_2',' P_2O_5','MgO', ' Al_2O_3']	*LREE/HREE/∑ REE*
8	[' MnO',' SiO_2',' K_2O ',' Fe_2O_3',' TiO_2','CaO',' P_2O_5','MgO', ' Al_2O_3']	*LREE/HREE/∑ REE*

3.9.1 Decision Tree (DT)

During training, the DT model seeks to learn patterns in the training subset of data records based on minimizing the entropy between the tree's subgroups (Quinlan, 1993). DTs are a non-parametric, supervised ML model that can be configured for regression analysis (i.e. continuous distributions of all variable values) required for this study, enabling it to divide the dataset into groups of data records at each tree node (Kadavi, 2019). The DT displays its divisions at specific nodes. A DT contains three types of nodes: root nodes, internal nodes, and leaf nodes. The root node is the initial node that is subsequently divided to form multiple internal nodes. The data divisions at the internal nodes are driven by the model's data properties and decision rules. The leaf nodes terminate each tree branch and represent the DT's ultimate data separations. Sklearn's DecisionTreeRegressor function was adapted in Python code to train the DT model. The DT hyperparameters applied were those provided as default options (**Table 5**).

3.9.2 Support Vector Regression (SVR)

The SVR algorithm identifies the optimal support vectors in a multi-dimensional hyperspace to effectively segregate data, following the principles of Chervonenkis (Cortes and Vapnik, 1995). SVR can be employed with diverse kernel functions, with linear, polynomial, and radial basis function (RBF) kernels being the most commonly applied. In datasets featuring multiple nonlinear relationships among the variables, SVR models tend to perform well with and minimize overfitting (Smola, A., 1998) but to do so requires careful tuning of their control parameters (**Table 5**).

3.9.3 ***Random Forest (RF)***

RF is an effective tree-ensemble ML method based on multiple decision trees (Breiman, 2001). Each of the trees included in an RF model uses different subsets of data records to fit the relationships of the training dataset, resulting in partial overfitting at the individual tree level. However, the combination of these tree results produces an average conclusion that provides more accurate predictions while effectively minimizing overfitting problems. The "Scikit-learn" Python library applies the CART regression trees process to execute the RF algorithm. This technique builds binary trees with exactly two output branches, picking the best numerical characteristic to partition the data in a specific tree. This

selection procedure aims to reduce the mean squared error (MSE) between the predicted and measured dependent variable (REE) values in the training set, to improve the prediction accuracy of the trained model when applied to the testing subset.

The main hyperparameters regulating the RF algorithm's behavior (**Table 5**) are the number of trees (n_estimators), the maximum depth allocated to individual trees, and the maximum fraction of features available for selection by each tree. A trial-and-error sensitivity analysis of the n_estimator parameter was carried out across a range of 50 to 500, and for max_depth over a range of 3 to 20 to identify the optimum values listed in **Table 5**.

3.9.4 *Extreme Gradient Boosting (XGBoost)*

XGBoost is another ensemble ML algorithm involving multiple decision trees (Chen, 2016). The algorithm strives to improve cost function *m* along with the aid of regularization term α. The tree structure is written as t_p in equation (1), in which the algorithm's goal is represented as *K* evaluated for function space *g*:

$$K^{(g)} = \sum_{i=0}^{n} m(y_i + \hat{y}_i) + \sum_{n=0}^{p} \alpha(t_p) \quad (1)$$

Where y_i represents the measured REE values compiled in the dataset and $\hat{y}_i$ represents the REE values predicted by XGBoost.

The 'XGBoost regressor' module in Python from the 'Scikit-Learn' library was employed to train the XGBoost model with the key hyperparameter values listed in Table 5. The number of trees (n_estimators) was set at 500. The learning rate (eta) was evaluated across the range of 0.01 to 0.2, with eta = 0.09 achieving the best results. To avoid overfitting, the maximum tree depth (max_depth) was set to 5, the subsample limit was set to 0.5, and the maximum fraction of features included in each tree was limited to 0.8. Additionally, an early stopping criterion was applied to halt the algorithm's iterations if no improvements were achieved after 10 repetitions.

3.9.5 Hyperparameters tunning

The performance of machine learning (ML) algorithms is significantly impacted by the choice of hyperparameters (Schratz et al., 2019). Thorough and precise sensitivity analyses were conducted to tune the hyperparameter values of each of the four ML models applied in this study.

Table 5. Tuned hyperparameters values applied to the four ML developed for REE prediction.

Algorithms	Hyperparameter	Optimum value selected
	max_depth	10

RF	max_features	1.0
	n_estimators	130
SVR	kernel	'linear'
XGBoost	n_estimators	500
	learning_rate	0.09
	max_depth	5
	Colsample_bytree	0.8
	Subsample	0.5
	reg_alpha	0.4
	gamma	1
	early_stopping_rounds	10
Decision Tree	splitter	"best"
	max_depth:	"None" (meaning no limit)
	min_samples_split	"2"
	min_samples_leaf	"1"
	max_features	"auto"
	max_leaf_nodes	"None"

3.10 Prediction Performance Evaluation Metrics

The REE prediction performance of each ML model applied to the compiled datasets is assessed applying statistical parameters that quantify prediction errors. The mean absolute error (MAE; equation (2)), root mean square error (RMSE; equation (3)), and coefficient of determination (R^2; equation (4)) are the metrics employed.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|REE_{measured.i} - REE_{predicted.i}\right| (0 \le MAE < +\infty) \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N}\left(REE_{measured.i} - REE_{predicted.i}\right)^2}. (0 \le RMSE < +\infty) \quad (3)$$

$$R^2 = \left[\frac{\frac{1}{N}\sum_{i=1}^{N}(REE)\cdot(\overline{REE_{measured.i}})}{\sqrt{\left(REE_{measured.i}-\overline{REE_{measured.i}}\right)^2}\cdot\sqrt{\left(REE_{predicted.i}-\overline{REE_{predicted.i}}\right)^2}}\right]^2 \quad (4)$$

Where

$(0 < R^2 < +1)$,

N is the number of data records in the subsection being assessed.

$\overline{REE}_{measured}$ and $\overline{REE}_{predicted}$ are the mean real REE and mean predicted REE, respectively.

Figure 4 displays a flowchart illustrating the ML modeling architecture as it is applied to the compiled datasets. The architectural framework of the study is structured as follows: initially, there is a compilation of data gathered from the most exhaustive analyses found in prior studies on Tethyan phosphorites. Subsequently, this collected data undergoes statistical processing with geochemical classification. The subsequent phase involves using this data

as input for selected algorithms, starting with the splitting of the data into training and testing sets. Following this, the model for REE values is constructed. Ultimately, the evaluation of our models is conducted using specific evaluation metrics.

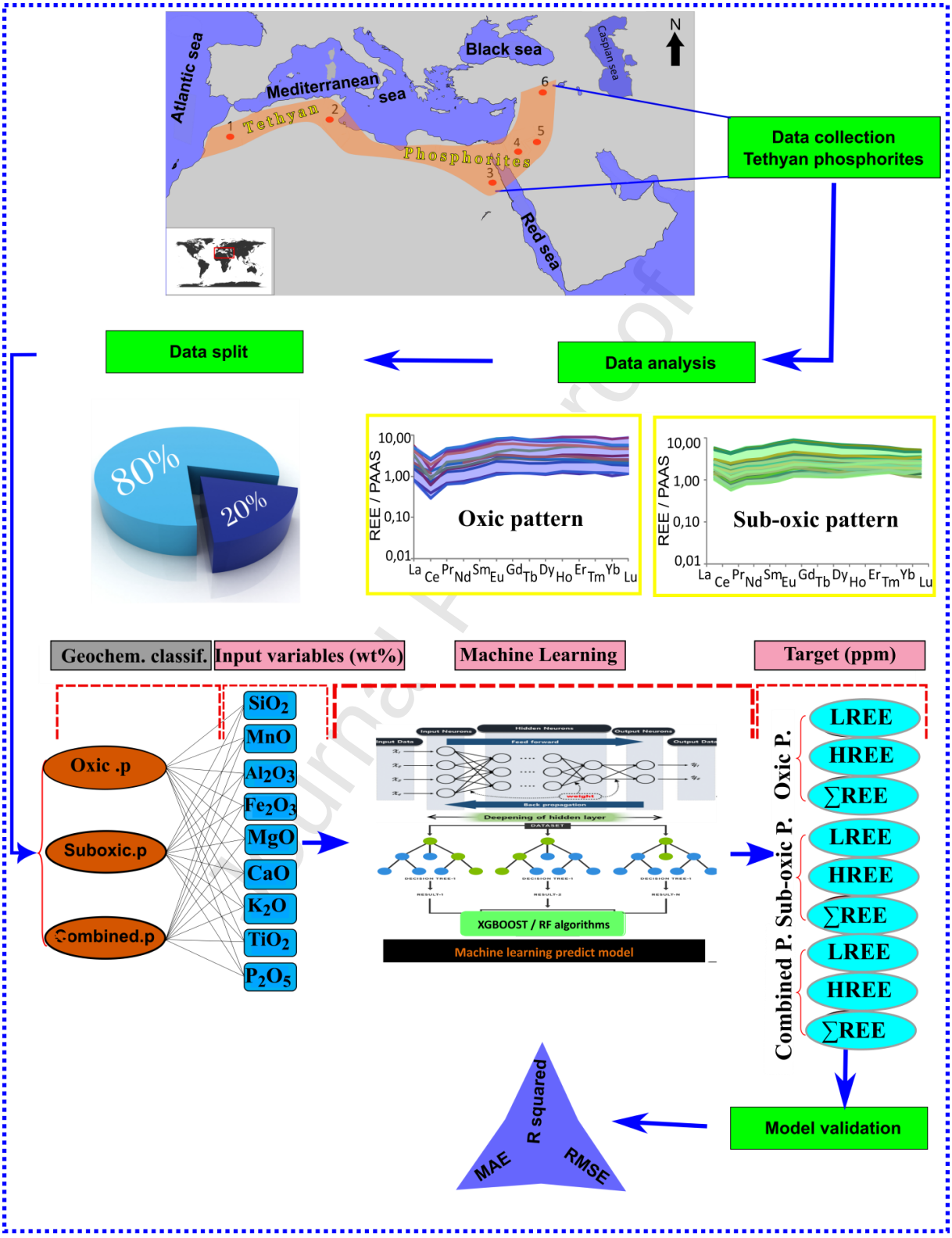


Figure.4. Flowchart illustrating the ML model architecture as applied to compiled Tethyan phosphorite datasets.

4. Results

Multiple distinct feature combinations were assessed with each of the ML models to predict REE concentrations based on the major elements content of the three geochemical patterns distinguished in Tethyan phosphorite. Eight feature combinations were assessed for the oxic and combined patterns, and six feature combinations were assessed for the sub-oxic pattern **(Table 4)**. Across all feature combinations and most geochemical patterns, XGBoost consistently demonstrated the best REE prediction performance (**Tables 6, 7,** and **8**). For LREE, HREE, and ∑REE predictions in the ***oxic*** pattern, models XGBoost8, XGBoost1 and XGBoost3 provided the best prediction performance with high R^2 values and relatively low RMSE: (R^2=0.979, RMSE=20.660 ppm, MAE=15.731 ppm), (R^2=0.961, RMSE=5.435 ppm, MAE=3.604 ppm), and (R^2=0.978, RMSE=25.206 ppm, MAE=17.293 ppm), respectively. For the ***sub-oxic*** pattern, model XGBoost5 provided the best REE prediction performance: (R^2=0.921, RMSE=100.137 ppm, MAE=62.326 ppm) for LREE, (R^2=0.937, RMSE=19.897 ppm, MAE=13.526 ppm) for HREE, and (R^2=0.942, RMSE=105.040 ppm, MAE=72.590 ppm) for ∑REE. Within the combined patterns, model XGBoost3 provided the best prediction performance for both LREE and ∑REE, with R^2 values of 0.925 and 0.928, RMSE values of 108.804 ppm and 128.986 ppm, and MAE values of 59.696 ppm and 78.672 ppm respectively. In the case of HREE prediction, However, for HREE in the combined patterns, model RF1 outperformed other models (R^2=0.930, RMSE=21.998 ppm, MAE=14.280 ppm) (**Table 8**).

The SVR and DT models generated more prediction errors with all geochemical patterns compared to the XGBoost and RF models. The SVR and DT models consistently overfitted each dataset as indicated by the RMSE and MAE values associated with the training subsets being much lower than those of the testing subsets.

Table.6. Prediction evaluation metric values for the best-performing ML models applied to the ***oxic*** pattern, for the feature combinations defined in **Table 4.** The feature combination providing the best performance overall with the testing subset is highlighted in bold type.

Pat-tern	Sub	Comb. Ref.	Best ML Algorithm	Training subset			Test subset		
				MAE	RMSE	R^2	MAE	RMSE	R^2
	LREE	1	XGBoost	7.344	12.06	0.99	18.23	24.13	0.97

					1	0	3	9	1
		2	XGBoost	8.261	14.549	0.986	17.697	23.670	0.972
		3	XGBoost	9.964	19.515	0.974	15.376	21.680	0.977
Oxic pat-tern		4	XGBoost	4.099	7.906	0.996	14.739	21.738	0.977
		5	XGBoost	5.303	10.452	0.993	22.618	36.744	0.933
		6	XGBoost	3.233	6.480	0.997	16.317	21.237	0.978
		7	XGBoost	1.470	2.228	1.000	15.460	21.171	0.978
		8	**XGBoost**	**2.538**	**4.646**	**0.999**	**15.731**	**20.660**	**0.979**
		1	**XGBoost**	**2.225**	**3.781**	**0.980**	**3.604**	**5.435**	**0.961**
		2	XGBoost	2.621	4.696	0.969	4.056	5.710	0.957
		3	XGBoost	2.431	4.411	0.973	3.666	5.570	0.959
	HREE	4	XGBoost	3.237	6.016	0.949	4.254	6.513	0.944
		5	XGBoost	1.490	2.764	0.989	5.021	8.533	0.903
		6	XGBoost	0.982	1.774	0.996	3.617	5.439	0.961
		7	XGBoost	3.237	6.016	0.949	4.254	6.513	0.944
		8	XGBoost	2.493	4.930	0.966	5.133	8.905	0.895
		1	XGBoost	9.153	15.125	0.989	21.644	28.847	0.971
	∑REE	2	XGBoost	7.827	13.979	0.991	21.271	28.238	0.972
		3	**XGBoost**	**9.587**	**18.405**	**0.984**	**17.293**	**25.206**	**0.978**

4	XGBoost	8.703	17.857	0.985	17.544	25.897	0.977
5	XGBoost	6.768	13.051	0.992	28.822	46.384	0.925
6	XGBoost	4.298	8.256	0.997	19.814	26.494	0.976
7	XGBoost	1.690	2.515	1.000	23.013	33.218	0.962
8	XGBoost	2.840	5.197	0.999	18.785	26.067	0.976

Table.7. Prediction evaluation metric values for the best-performing ML models applied to the ***sub-oxic*** pattern, for the feature combinations defined in **Table 4**. The feature combination providing the best performance overall with the testing subset is highlighted in bold type.

Pat-tern	Sub	Comb. Ref.	Best ML Algorithm	Training subset			Test subset		
				MAE	RMSE	R^2	MAE	RMSE	R^2
Sub-oxic pat-tern	LREE	1	XGBoost	34.736	54.489	0.981	100.676	159.901	0.799
		2	XGBoost	52.985	78.362	0.962	98.391	139.330	0.848
		3	XGBoost	49.784	77.458	0.962	94.584	135.536	0.856
		4	XGBoost	33.250	49.194	0.985	79.302	119.733	0.887
		5	**XGBoost**	**24.152**	**36.039**	**0.992**	**62.326**	**100.137**	**0.921**
		6	XGBoost	49.161	75.576	0.964	85.958	128.164	0.871
	HREE	1	XGBoost	6.738	11.712	0.983	17.759	26.325	0.890
		2	XGBoost	9.535	15.332	0.971	17.069	23.997	0.909
		3	XGBoost	7.068	12.316	0.981	16.195	25.699	0.896
		4	XGBoost	7.248	13.595	0.977	17.211	24.709	0.903

		5	**XGBoost**	**7.981**	**13.861**	**0.976**	**13.526**	**19.897**	**0.937**
		6	XGBoost	2.498	4.184	0.998	16.947	24.343	0.906
		1	XGBoost	46.797	73.389	0.977	108.059	169.533	0.848
		2	XGBoost	56.654	83.581	0.971	111.773	150.058	0.881
		3	XGBoost	61.818	94.663	0.962	100.761	143.342	0.891
	∑REE	4	XGBoost	16.297	22.909	0.998	98.269	131.542	0.908
		5	**XGBoost**	**17.825**	**25.265**	**0.997**	**72.590**	**105.040**	**0.942**
		6	XGBoost	60.200	91.152	0.965	91.283	137.568	0.900

Table.8. Prediction evaluation metric values for the best-performing ML models applied to the ***combined*** pattern, for the feature combinations defined in **Table 4.** The feature combination providing the best performance overall with the testing subset is highlighted in bold type.

Pat-tern	Sub	Comb. Ref.	Best ML Algorithm	Training subset			Test subset		
				MAE	RMSE	R^2	MAE	RMSE	R^2
		1	RF	34.336	53.822	0.975	73.854	108.944	0.925
		2	XGBoost	25.608	38.752	0.987	67.136	127.518	0.897
		3	**XGBoost**	**27.538**	**43.062**	**0.984**	**59.696**	**108.804**	**0.925**
		4	XGBoost	23.119	33.893	0.990	65.416	114.301	0.917
	LREE	5	RF	32.181	53.731	0.975	75.039	112.284	0.920
Com-bined patterns		6	XGBoost	6.855	9.267	0.999	71.541	128.370	0.896
		7	RF	30.321	51.881	0.977	74.866	118.286	0.911
		8	RF	31.515	53.103	0.976	75.646	115.739	0.915
		1	**RF**	**7.077**	**11.479**	**0.974**	**14.280**	**21.998**	**0.930**
	HREE	2	RF	6.405	10.751	0.977	15.927	24.937	0.910
		3	RF	6.110	10.44	0.97	14.43	23.088	0.92

				4	9	6		3
	4	RF	6.147	11.210	0.975	14.334	23.189	0.922
	5	RF	6.201	10.870	0.977	14.048	23.064	0.923
	6	RF	6.244	10.644	0.978	14.079	23.198	0.922
	7	RF	6.062	10.751	0.977	14.171	23.312	0.921
	8	RF	6.308	11.075	0.976	14.248	23.831	0.918
	1	RF	40.584	64.724	0.975	87.440	131.114	0.925
	2	XGBoost	28.904	43.843	0.989	79.868	144.327	0.909
	3	**XGBoost**	**27.821**	**42.547**	**0.989**	**78.672**	**128.986**	**0.928**
∑REE	4	XGBoost	25.522	39.191	0.991	79.364	135.777	0.920
	5	RF	38.008	64.873	0.975	91.633	141.738	0.912
	6	RF	37.715	63.138	0.976	93.515	145.896	0.907
	7	RF	35.962	61.564	0.977	89.840	141.810	0.912
	8	RF	38.182	64.162	0.976	90.231	139.649	0.915

Graphical displays of the ML model prediction results achieved with the testing subset are presented in **Figures 5** to **10**. The following scatterplots serve as visual representations illustrating the relationship between the measured and predicted values of REE in (ppm). These plots specifically focus on the oxic, sub-oxic, and combined pattern testing subsets. They are designed to highlight and elucidate the results obtained through the application and evaluation of four distinct algorithms within the scope of this research or study.

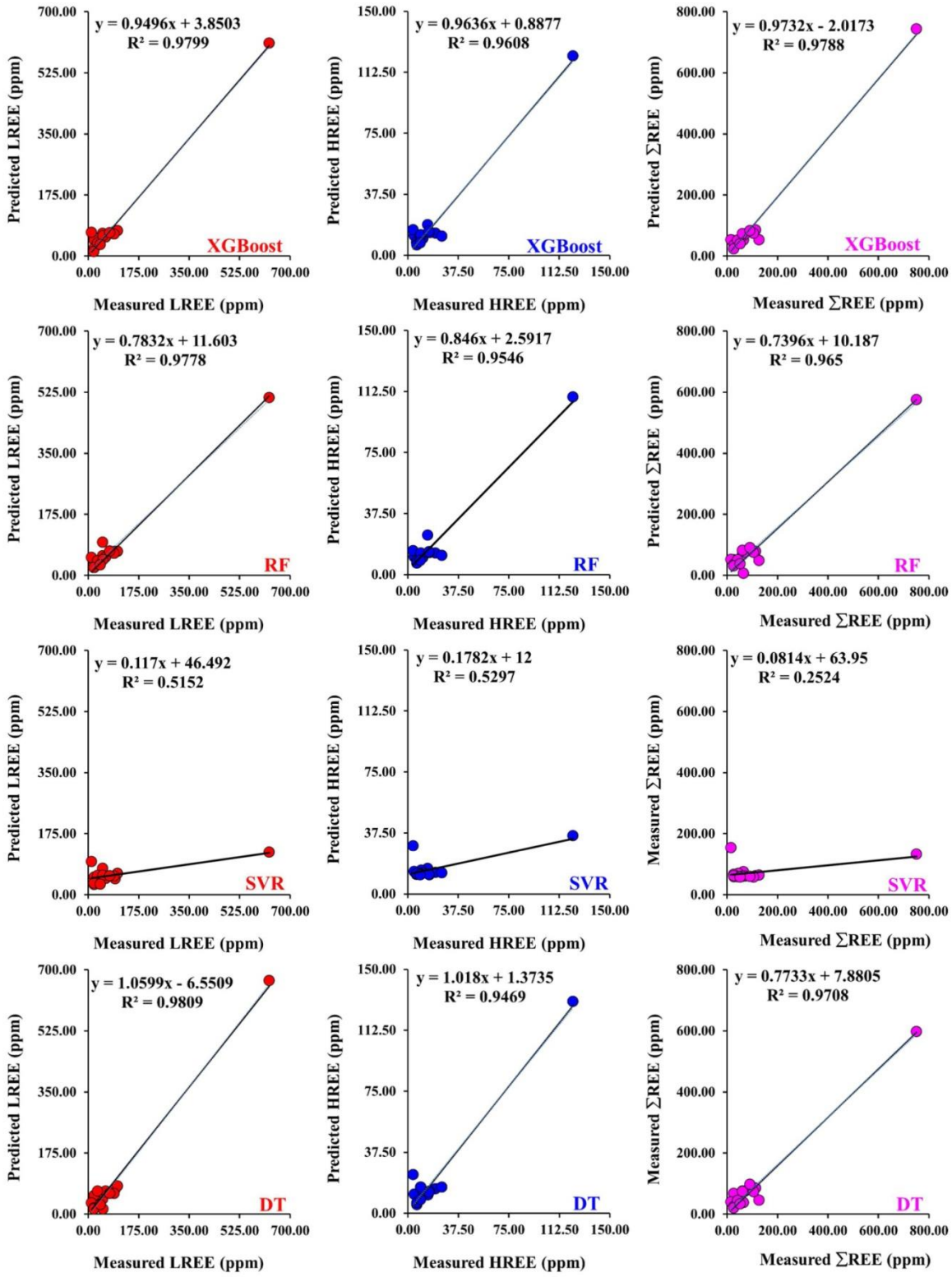


Figure 5. Scatterplots of measured versus predicted values for the ***oxic*** pattern testing subset: (*a*) left panel for the LREE, (*b*) middle panel for the HREE, and (*c*) right panel for the REE.

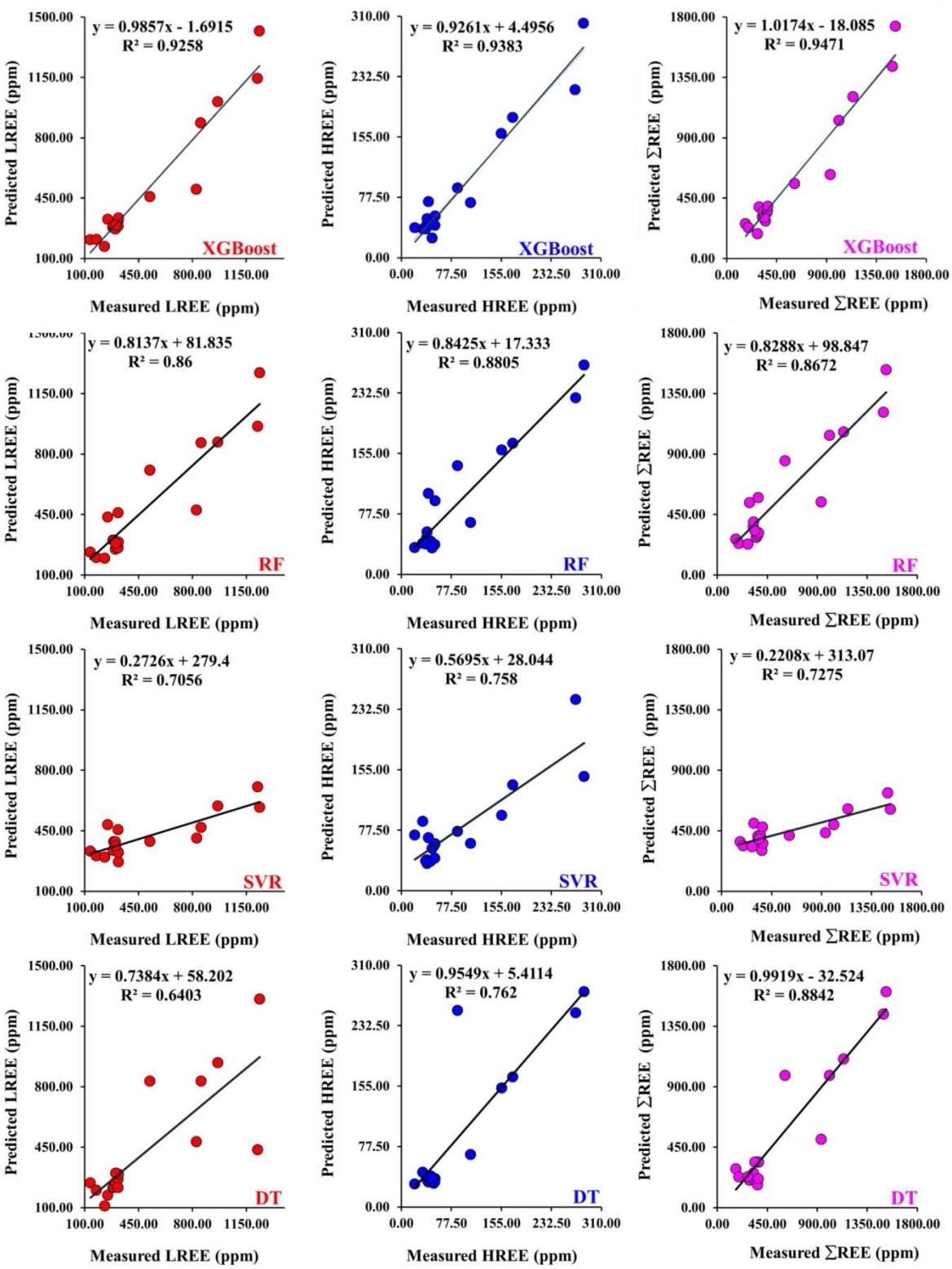


Figure 6. Scatterplots of measured versus predicted values for the ***sub-oxic*** pattern testing subset: (*a*) left panel for the LREE, (*b*) middle panel for the HREE, and (*c*) right panel for the REE.

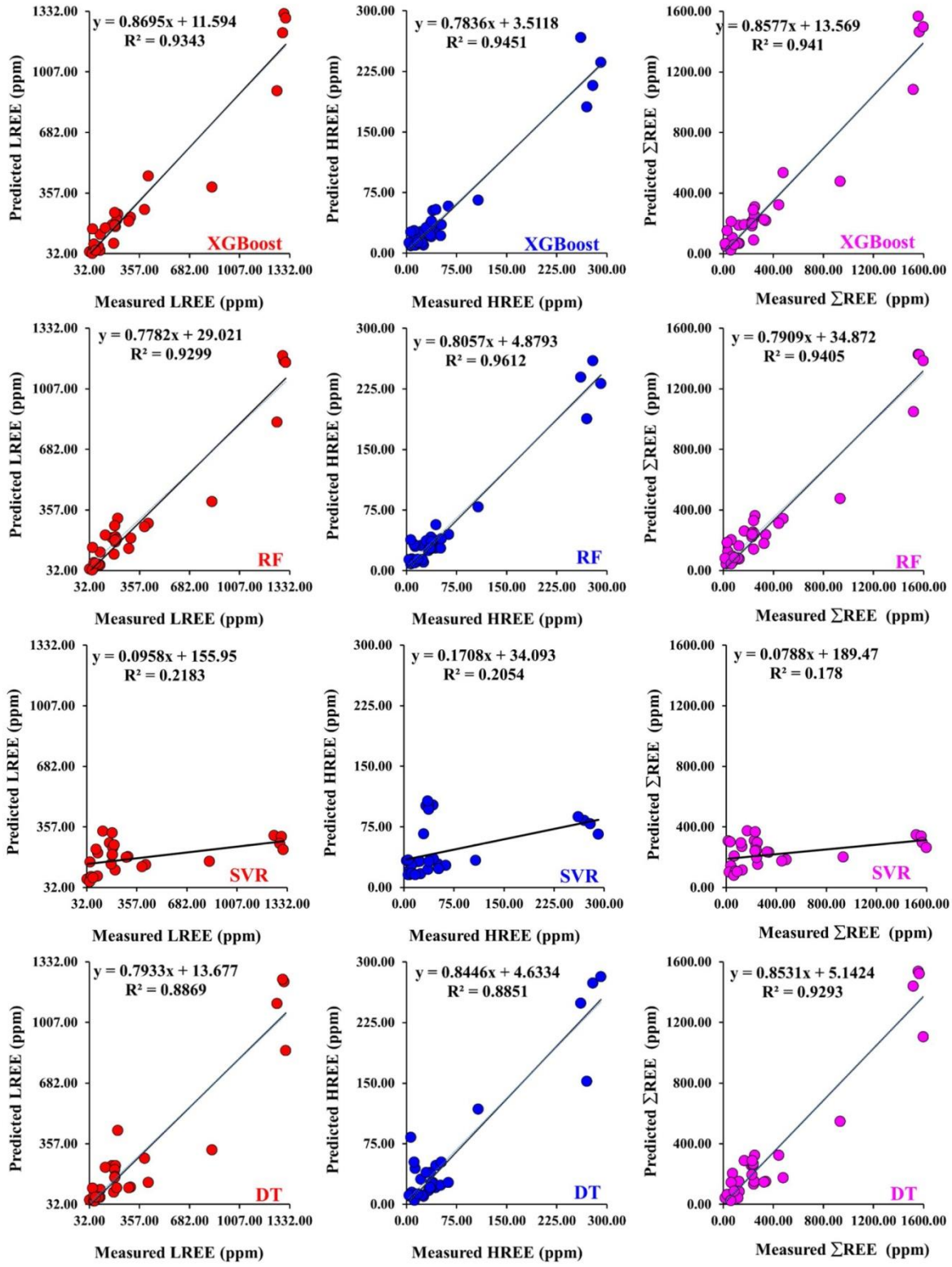


Figure 7. Scatterplots of measured versus predicted values for the ***combined*** pattern testing subset: (*a*) left panel for the LREE, (*b*) middle panel for the HREE, and (*c*) right panel for the REE.

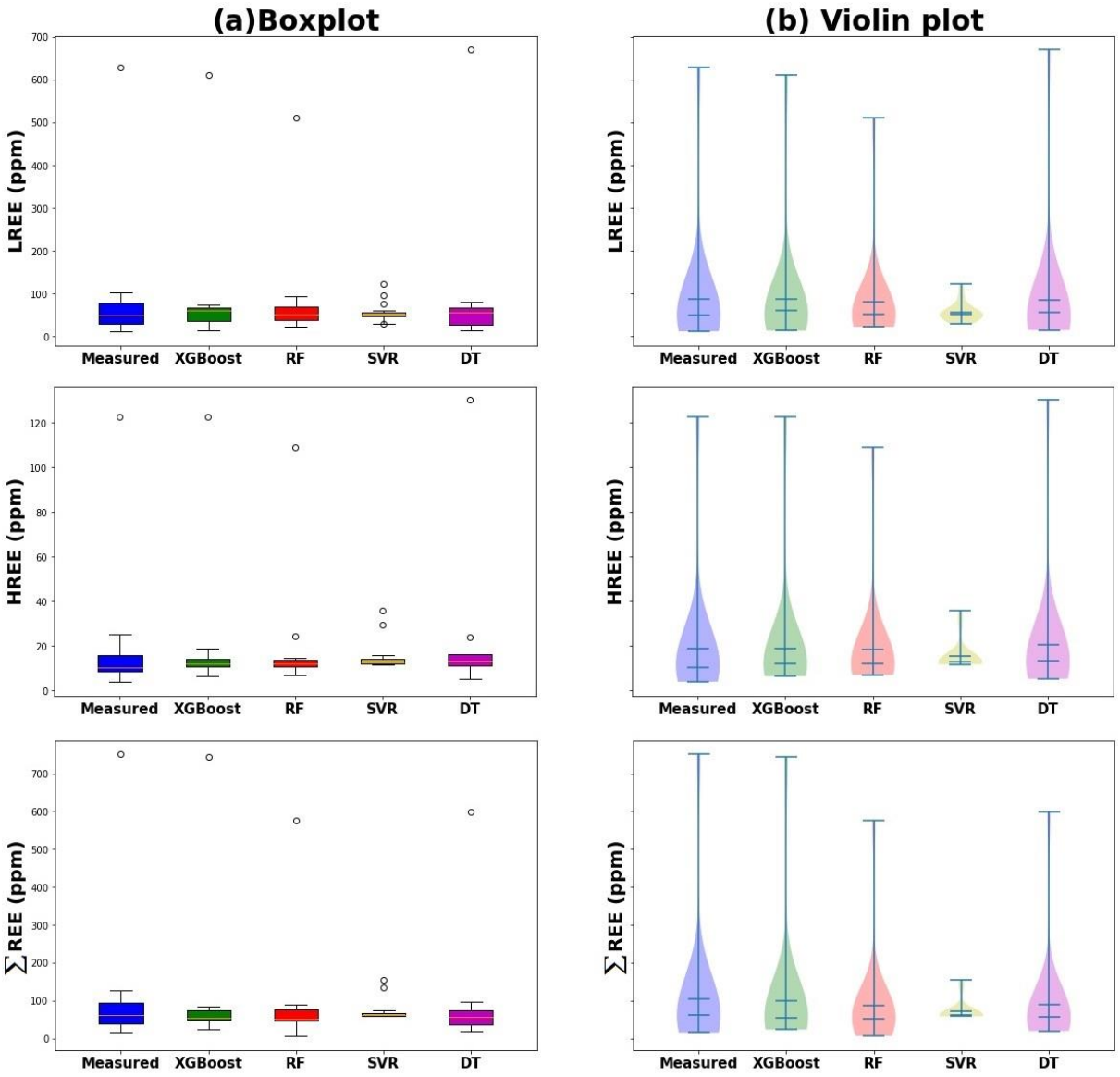


Figure 8. Graphical comparisons between measured and predicted REE (ppm) for the four developed ML models applied to the testing subsets for the ***oxic*** pattern: **(a)** Boxplot, **(b)** Violin plot.

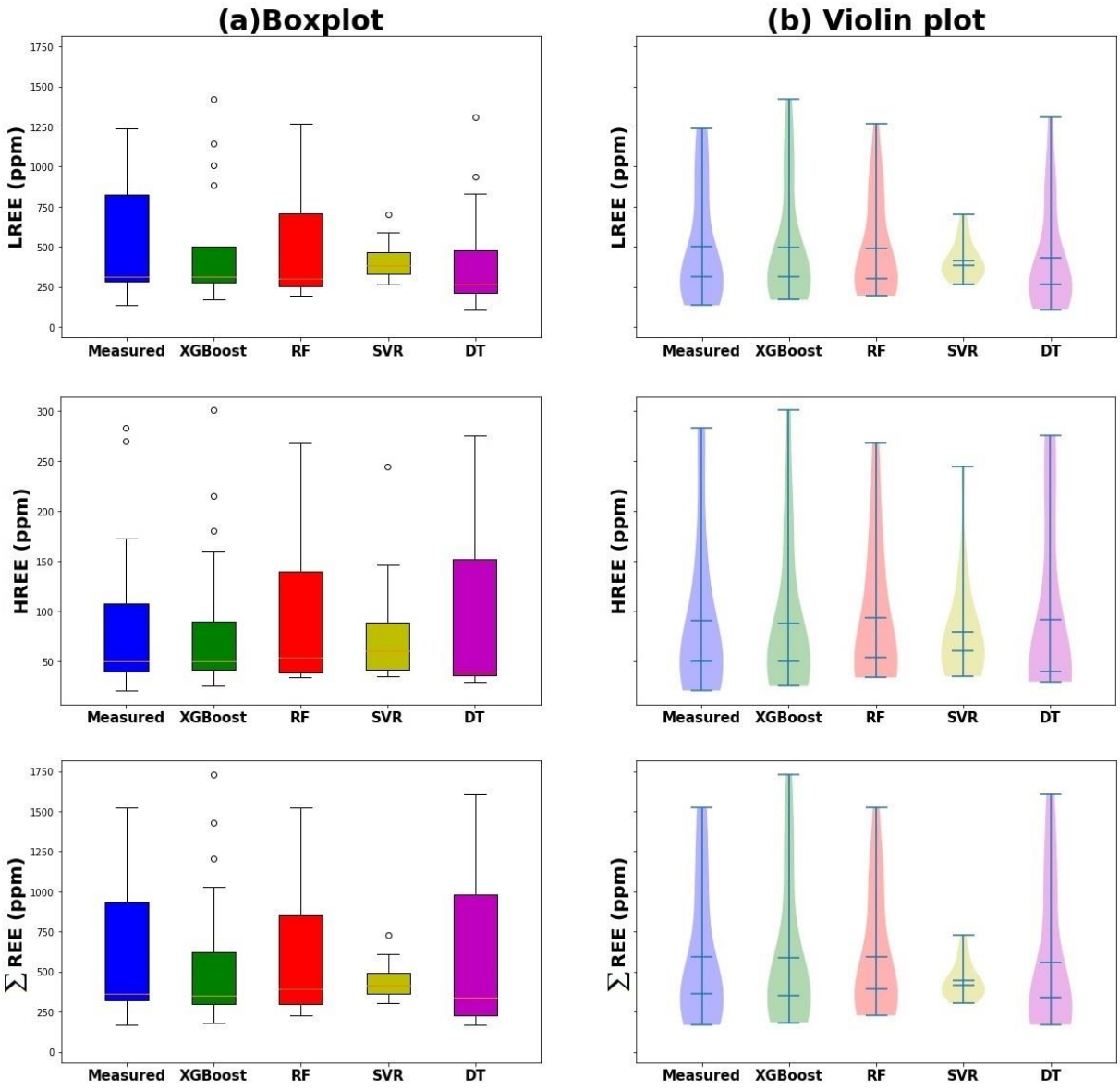


Figure 9. Graphical comparisons between measured and predicted REE (ppm) for the four developed ML models applied to the testing subset for the ***sub-oxic*** pattern: **(a)** Boxplot, **(b)** Violin plot.

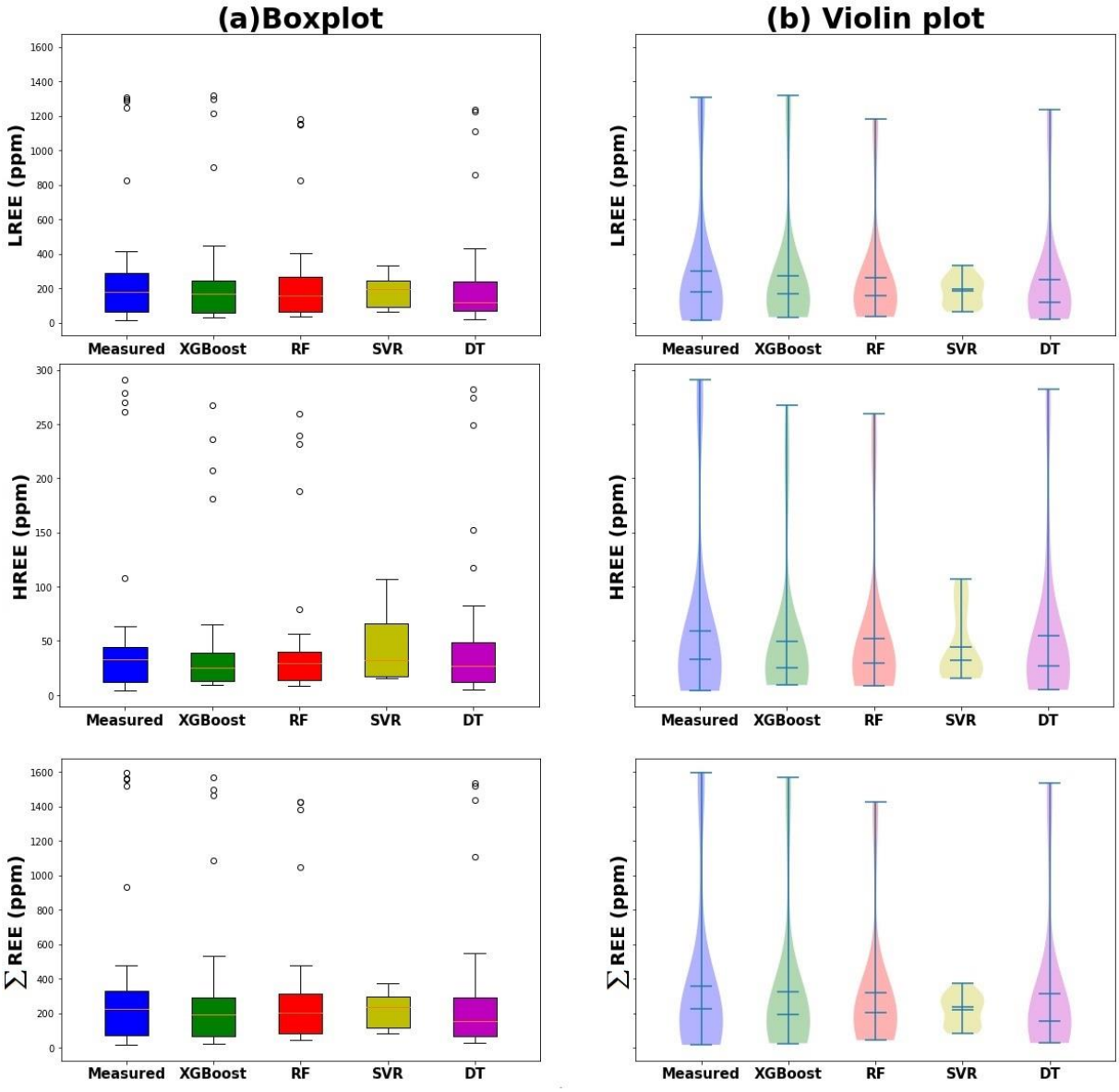


Figure 10. Graphical comparisons between measured and predicted REE (ppm) for the four developed ML models applied to the testing subset for ***combined*** pattern: **(a)** Boxplot, **(b)** Violin plot.

5. Discussion

5.1. Factors controlling Tethyan phosphorite REE distributions and their interactions

Phosphorites typically display higher REE contents than other sediments (Emsbo et al., 2015; Kechiched et al., 2016, 2018; Linares et al., 2023). Certain phosphorus-bearing minerals such as francolite ($Ca_5(PO_4)_3F$) can host REE where REE^{3+} tends to replace Ca^{2+} in the crystal lattice (Glenn et al., 1994; Jarvis et al., 1994). This is referred to as the carbonate-

fluorapatite (CFA) crystallo-chemical control of REE. Within the Tethyan phosphorites investigated, P_2O_5 has a high positive correlation with REE in all of the deposits analyzed (Garnit et al., 2017; Kechiched et al., 2020; Ahmed et al., 2022; Edress et al., 2023; El Bamiki et al., 2023), indicating that CFA may represent the dominant process involved in REE enrichment of phosphorites.

For instance, in the Moroccan deposits, REE exhibits a positive correlation with P_2O_5 in all examined phosphatic lithofacies, with R ~ 0.71 (El Bamiki et al., 2023). The Tunisian phosphorite samples display a positive Pearson correlation between REE and P_2O_5 in the Sra Ouertane, Sekarna, and Jebel Jebs deposits with (R = 0.79), (R = 0.59) and (R = 0.97), respectively. These correlations support potential CFA enrichment processes involving francolite in phosphorites within these regions. However, no significant correlation was shown for the Gafsa-Metlaoui (Tunisia) samples (R = 0.05), where the predominant control of REE content is attributed to Fe-Mn oxyhydroxides (Garnit et al., 2017). The Egyptian phosphorites display a strong positive correlation between P_2O_5 and REE (R = 0.82). This supports the shared origin of P_2O_5 and REE in these deposits (Edress et al., 2023). The Jordanian deposits also the Egyptian phosphorites display good positive correlations between P_2O_5 and REE, with R = 0.59 between P_2O_5 and La. This suggests that REE has potentially substituted for Ca within the CFA structure in those deposits (Amireh et al., 2019). The Saudi Arabian phosphorites also display a positive correlation between P_2O_5 and REE (R = 0.62), supporting the influence of CFA on the uptake of REE in those deposits (Ahmed et al., 2022). The same is the case with the Turkish phosphorites with R = 0.92 between P_2O_5 and REE (Gundogar and Sasmaz, 2022).

The content of REE in phosphate-rich rocks tends to be similar within a specific geologic era but varies by up to three orders of magnitude between geologic eras, irrespective of their depositional environments (Emsbo et al., 2015). This implies that secular variations in ocean chemistry may be responsible for such fluctuations (Jarvis et al., 1994; Picard et al., 2002; Lécuyer et al., 2004; Yang et al., 2021). This concept has considerable implications for the distribution of REE-rich sediments. However, not all researchers agree with the overriding control of ocean chemistry in this regard. Some researchers suggest that the primary control of REE loading in phosphorites is associated with the local deposition environment (Valetich et al., 2022). Data from Cambrian phosphorites in Australia's Georgina Basin and Algeria’s Paleocene/Eocene phosphorites support the important role of local depositional conditions.

In North African deposits, numerous studies have confirmed that the primary control of REE enrichment is the local deposition environment (Garnit et al., 2012, 2017; Kechiched

et al., 2018, 2020; Buccione et al., 2021; Ferhaoui et al., 2022). In this context, phosphorites from different basins of the Paleocene-Eocene period exhibit varying REE content and distribution, e.g., those surrounding the paleo-island of Kasserine (Tunisia) (Ferhaoui et al., 2022). The local control is associated with the processes of winnowing and reworking and the extended exposure time that occurs with low sedimentation rates. Both factors have contributed to the REE enrichment of the main phosphorite sub-layer of Kef Essenoun (Algeria) (Kechiched et al., 2020). The samples compiled for this study display a geological age range from the Upper Cretaceous to the Eocene. Comparing the average REE content in Upper Cretaceous phosphorites in the studied dataset, specifically the Abu Tartur deposits in Egypt (1266.33 ppm), the Mazidaq phosphorites in Turkey (27.58 ppm), and the Moroccan deposits (149.1 ppm), reveals a large range of REE concentrations. The large range of average REE content among these Upper Cretaceous deposits is a clear indication that the primary factor influencing REE uptake is the local deposition environment rather than secular variations in ocean water composition.

The dataset compiled for this study specifically focuses on samples with P_2O_5 content exceeding 18 wt% (i.e., restricting the samples to phosphorites sensu stricto) (Nathan, 1984). This deliberate selection has resulted in a relatively narrow range of P_2O_5 values, with an average plus or minus one standard deviation for P_2O_5 of 24.36 ± 4.44 wt% for the oxic pattern dataset and 26.14 ± 2.55 wt% for the sub-oxic pattern dataset (**Table 3**). In comparison, CaO displays an average of 48.39 ± 6.34 wt% for the oxic pattern dataset and 42.44 ± 4.07 wt% for the sub-oxic pattern dataset (**Table 3).** This narrow compositional range has consequently reduced the Pearson correlation coefficient (R) between P_2O_5 and REE in these datasets.

The major elements, for the most part exhibit high coefficients of variation (CV; the ratio of the standard deviation to the mean) for the studied dataset: CV_{SiO2} = 91.92 %, CV_{MnO} = 233.33 %, CV_{Al2O3} = 80.23 %, CV_{Fe2O3} = 130.43 %, CV_{MgO} = 122.98 %, CV_{K2O} = 86.67 %, and CV_{TiO2} = 50.00 %. However, P_2O_5 and CaO display substantially lower coefficients of variation (CV_{P2O5} = 14.60 % and CV_{CaO} = 13.37 %). This observation suggests that although these two elements are important in predicting REE in phosphorites (e.g., the CFA process), their low coefficients of variation may mask the full range of the influences these elements have on REE distribution.

5.2. Feature importance analysis and interpretation

In addition to the REE content control in phosphorites exerted by francolite, the results of this study suggest additional factors that might influence the concentration of REE in Tethyan phosphorites and be exploited for prediction purposes.

5.2.1 Oxic pattern

In the oxic pattern dataset, P_2O_5 exhibits a weak negative correlation with TiO_2 (R = -0.31, p = 0.005) (**Figure 3**). This suggests that upwelling currents potentially play a significant role as the primary source of P_2O_5 within the seawater Tethyan phosphorite deposits. Except for CaO, P_2O_5 exhibits a negative correlation with all the major elements in this dataset, including Al_2O_3 (R = -0.32, p = 0.004), Fe_2O_3 (R = -0.28, p = 0.013), MgO (R = -0.40, p ≈ 0.00), K_2O (R = -0.39, p ≈ 0.00), and TiO_2. However, CaO demonstrates a positive correlation with P_2O_5 (R = 0.41, p ≈ 0.00). Notably, these two elements constitute the primary components of authigenic francolite (Föllmi et al., 1993; Glenn et al., 1994).

P_2O_5 and CaO are not identified as being significant in predicting REE contents in phosphorites in the best-performing ML models based on the XGBoost feature importance score (F-score). This is due to the consistently high contents and relatively uniform levels of P_2O_5 and CaO across all phosphorite samples within this dataset. The concentrations of the major element oxides hosted in silicate minerals are of more use to the ML models, particularly K_2O and Fe_2O_3 **(Figure 11).** These silicate components assume a pivotal role in improving the precision of REE predictions, thereby emphasizing their importance in characterizing the variability among various phosphorite deposits.

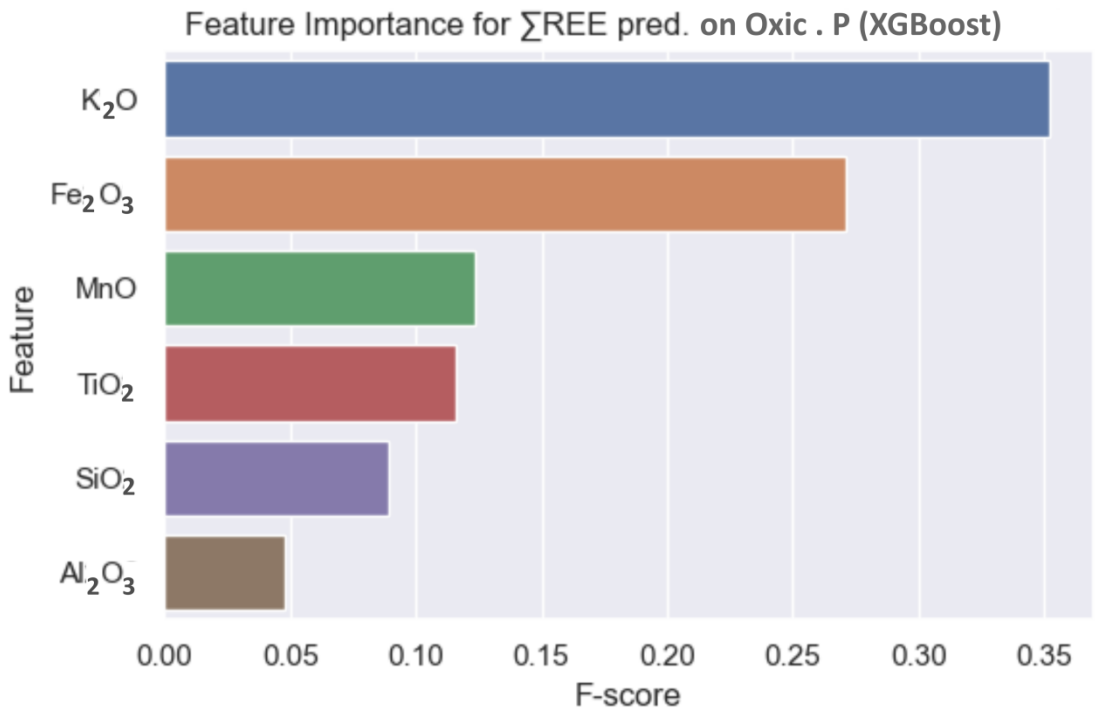


Figure 11. Relative features importance of major element oxides in the prediction of ∑REE in the oxic pattern dataset using the best performing XGBoost3 model.

Iron primarily enters seawater through two mechanisms: as bedload in the form of detrital iron; and, in the form of suspended colloidal iron as Fe-hydroxides gel. These colloidal iron particles are transported by river systems and primarily originate from the extensive chemical weathering of lateritic soils (Adachi et al., 1986; Einsele, 2001).

Al_2O_3, Fe_2O_3, MgO, K_2O, and TiO_2 exhibit positive correlations among themselves. The XGBoost F-score values show that Fe_2O_3 and K_2O are the best predictors of REE concentration (**Figure 11**). The most notable and significant Pearson correlation coefficient is observed between Fe_2O_3 and K_2O, with a value of ($R = 0.84$, $p = 0$). This relationship suggests the existence of a detrital phase or matrix component containing aluminosilicates and clay fractions, such as the glauconite observed in deposits like Sra Ouartane in Tunisia (Garnit et al., 2017). The studied deposits are also characterized by substantial contents of K-feldspar, illite, and smectite minerals (e.g., Garnit et al., 2017).

The ∑REE exhibit a strong and significant correlation with Fe_2O_3 ($R = 0.83$, $p = 0$) and K_2O ($R = 0.72$, $p = 0$). Additionally, a moderate correlation is observed between ∑REE and Al_2O_3 ($R = 0.38$, $p \approx 0.00$), TiO_2 ($R = 0.29$, $p = 0.01$), SiO_2 ($R = 0.30$, $p = 0.007$) and MgO ($R = 0.33$, $p = 0.003$) (**Figure 3**). These associations align with the presence of fine-grained detrital material, often dispersed within the Mg-rich dolomitic matrix. K_2O and Fe_2O_3 are common constituents of glauconite, a mineral known to be a favorable host of substantial REE content (Bansal et al., 2018; Kechiched et al., 2018, 2020). Additionally, iron oxyhydroxides are known to be responsible for HREE and MREE enrichment in certain sedimentary ores such as phosphorite and bauxite (Mongelli et al., 2014; Chen et al., 2015). It seems likely that Fe-oxyhydroxides play a pivotal role in explaining the strong positive correlation between Fe_2O_3 and REE in the samples of this dataset. Indeed, the significant positive correlations of Fe_2O_3 ($R = 0.83$, $p = 0.00$) and K_2O ($R = 0.72$, $p = 0.00$) with ∑REE underline their prominent roles in predicting REE enrichment within the oxic pattern of Tethyan phosphorites.

The low positive correlation coefficient between TiO_2 and ∑REE suggests that the REE content may also be affected by supply from terrestrial sources in at least some of the studied samples. The low negative correlation coefficient between MnO and REEs ($R = -0.14$, $p = 0.215$) suggests that there is no substantial control exerted by MnO on the ∑REE enrichment. However, MnO concentrations are being exploited by the XGBoost3 model when predicting REE concentrations (**Figure 11**).

5.2.2 Sub-oxic pattern

For the sub-oxic pattern dataset, there is a significant positive correlation between P_2O_5 and CaO ($R = 0.65$, $p \approx 0.00$). This alignment is considered to predominantly correspond to the presence of the CFA phase. ∑REE displays a poor positive correlation with

P_2O_5 (R = 0.17, p = 0.125) and also exhibits a low positive correlation with CaO (R = 0.13, p = 0.254). However, these correlations do not accurately represent the actual influence of P_2O_5 and CaO as major controllers of REE content, as previously discussed. This discrepancy can be attributed to the consistently homogeneous content of these elements across all the analyzed deposits. MgO displays a significant negative correlation with P_2O_5 (R = -0.60, p = 0), indicating the absence of substitution between Mg^{2+} and Ca^{2+} within the francolite lattice. A pronounced and significant inverse correlation between MgO and CaO (R = -0.82, p ≈ 0.00) exists within the sub-oxic pattern dataset. This points to a continual replacement of CaO by MgO in the carbonate matrix during diagenesis (dolomitization), with this phenomenon being particularly pronounced in the sub-oxic pattern dataset. In addition to the CFA control, XGBoost F-scores for the sub-oxic pattern dataset, reveal that the secondary drivers of ∑REE enrichment are MnO and SiO_2 concentrations (**Figure 12**).

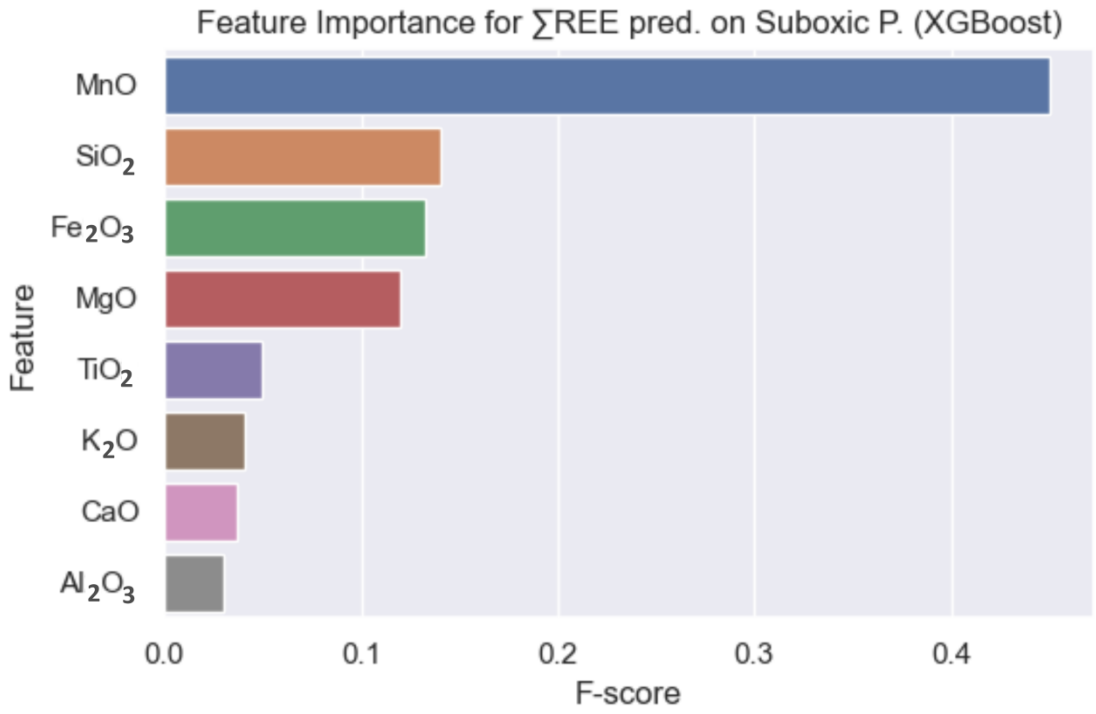


Figure.12. Relative features importance of major element oxides in the prediction of ∑REE in the sub-oxic pattern dataset using the best performing XGBoost5 model.

SiO_2 shows a negative correlation with TiO_2 (R = - 0.15, p = 0.189) Indicating its hydrogenous source (Boström, 1973; Coombs, et al., 1985). This is further supported by the biogenic origin of SiO_2 from siliceous-test plankton microfossils, specifically diatoms and radiolarians (Haddi et al., 2014; Kocsis et al., 2014). A significant negative correlation exists between ∑REE and SiO_2 (R = - 0.58, p = 0) (**Figure 3**). This is interpreted as due to the influence of silicious cementation and a late diagenetic silicification process which act to

reduce the amount of ∑REE-enriched porewater present. The extensive presence of chert-rich layers and nodules supports this interpretation.

∑REE displays a positive correlation with Fe_2O_3 (R = 0.37, p = 0.001). MnO and Fe_2O_3 are positively correlated (R = 0.49, p ≈ 0.00), whereas Fe_2O_3 shows a negative correlation with TiO_2 (R = -0.22, p = 0.045), suggesting ∑REEs uptake by Fe-Mn oxyhydroxides. Fe_2O_3 also constitutes a major component of goethite or hematite mineral phases, both known for their REE-rich characteristics (Mondillo et al., 2019). This aligns with existing models that suggest that Fe-oxyhydroxides play a key role in scavenging and carrying REEs (Jarvis et al., 1994).

The negative relationships of Fe_2O_3 and MgO with TiO_2 (R = -0.22, p = 0.042 and R = -0.22, p = 0.045) (**Figure 3**), suggest a prevailing hydrogenous origin for Fe_2O_3 and MgO. Moreover, Fe_2O_3 and MgO are positively correlated (R = 0.58, p = 0), involving coexistence, possibly within the dolomite fraction or ankerite minerals.

In contrast to the oxic pattern dataset, the sub-oxic pattern dataset displays a negative correlation between MgO and total ∑REE (R = -0.33, p = 0.002). The substantial dolomitization associated with these deposits significantly influences the phosphorite composition. Notably, dolomitic francolite exhibits lower REE concentrations compared to pure francolite. Similar negative correlation trends between ∑REEs and MgO have been observed in REE-enriched phosphorites in Guizhou Province, Southwest China, which exhibit the characteristic sub-oxic hat-shaped REE signature (He et al., 2022).

∑REE displays a significant positive correlation with MnO (R = 0.66, p = 0) whereas no correlation is observed with K_2O (R = -0.04, p=0.72). Glauconite is found in small quantities in the Kef Essenoun deposit (Algeria) (Kechiched et al., 2018, 2020) but contains a relatively minor amount of K_2O (ranging from 1.57 to 3.02 wt%). Nevertheless, despite its low K_2O content, the Kef Essenoun glauconite exhibits a significant amount of ∑REE. This is further supported by the near-zero or negligible negative correlation between K_2O and Fe_2O_3 (R = -0.03, p=0.783). ∑REE display moderate correlations with Al_2O_3 and TiO_2 of (R=-0.30, p=0.006 and R=0.22, p=0.045), respectively. These relationships imply that REE in the sub-oxic pattern dataset are predominantly associated with detrital minerals such as ilmenite, rutile, anatase, and brookite, rather than being primarily contained within aluminosilicates.

5.2.3. Combined pattern

The combined pattern dataset displays a mixture of influences from both the oxic and sub-oxic pattern samples. **Figure 13** shows the XGBoost F-scores for this dataset revealing the dominant influences of MnO and Fe_2O_3 concerning their predictions.

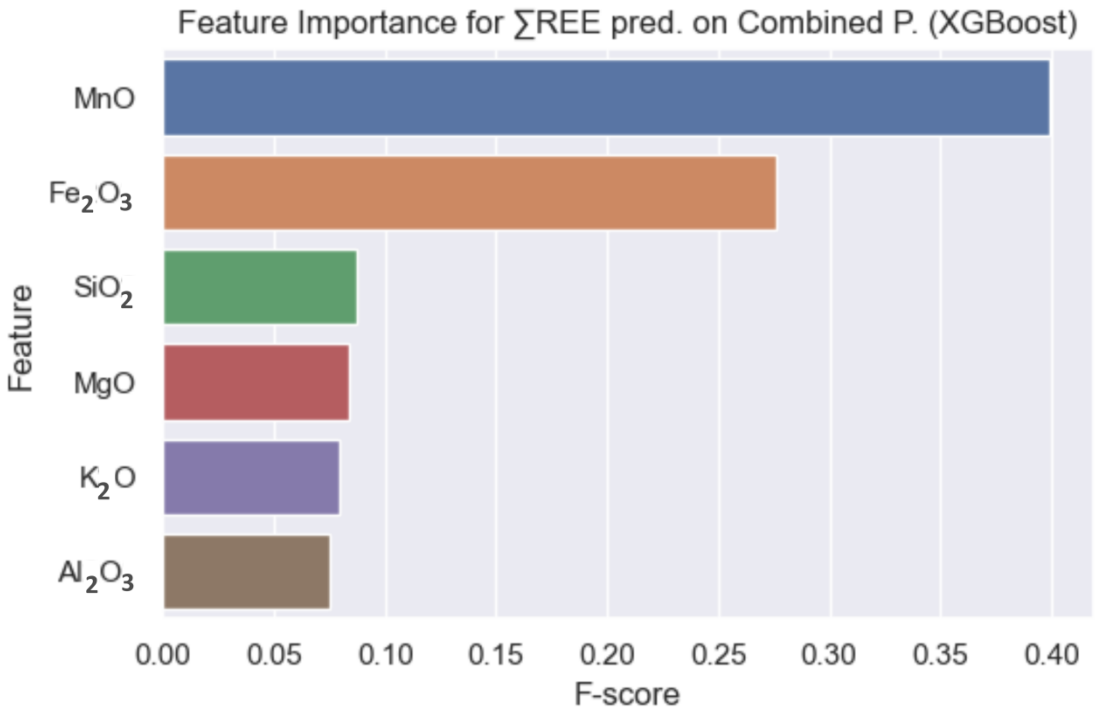


Figure 13. Relative features importance of major element oxides in the prediction of ∑REE in the combined pattern dataset using the best performing XGBoost3 model.

5. 3. Evaluation of ML in predicting REEs distribution

Recent review articles (Zuo, 2017; He et al., 2022), strongly support the use of ML as an effective instrument for various geological analyses. However, very few studies have modeled REE contents of sediments using major element oxides as input variables. Examples of such studies are the distribution of LREE patterns in the Chogart deposit in central Iran using DT and ANN models (Zaremotlagh and Hezarkhani, 2017), and the amount of REEs in Ocean Island Basalt (OIB) forecasted by RF and DT models (Hong et al., 2019). Also, Al_2O_3, Fe_2O_3, TiO_2, and SiO_2 were effectively employed as input variables in a data-centric ML approach using XGBoost models to forecast HREE distributions in the karst bauxite deposits of southern Italy (Buccione et al., 2023). **Table 9** compares the results of this study with the three studies mentioned.

Table 9. Comparisons of the REE prediction performance metric values between this work and previously published studies using ML models with major element oxide input variables.

REE-Host rock	Target		Algo-rithms	Metrics evaluation			References
Choghart iron oxide-apatite	LREE			MSE			(Zaremotlagh and Hezarkhani, 2017)
	La		ANN	0.0228			
	Ce		ANN	0.0312			
	Pr		ANN	0.0980			
	Nd		ANN	0.0488			
	Sm		ANN	0.0736			
	Eu		ANN	0.0235			
Ocean Island Basalt (OIB)					RMSE	R^2	(Hong et al., 2019)
	LREE		RF		17,06	0.92	
	HREE		RF		1,98	0.89	
	REE		RF		17,89	0.92	
Karst bauxite deposits in southern Italy				MAE	RMSE	R^2	(Buccione et al., 2023)
	HREE		XGBoost	5,091	7,299	0.83	
Tetyhan phos-phorites				MAE	RMSE	R^2	This study
	Oxic pattern	LREE	XGBoost	15.731	20.660	0.979	
		HREE	XGBoost	3.604	5.435	0.961	
		∑REE	XGBoost	17.293	25.206	0.978	
	Sub-oxic pattern	LREE	XGBoost	62.326	100.137	0.921	
		HREE	XGBoost	13.526	19.897	0.937	
		∑REE	XGBoost	72.590	105.040	0.942	
	Com-bined patterns	LREE	XGBoost	59.696	108.804	0.925	
		HREE	RForest	14.280	21.998	0.930	
		∑REE	XGBoost	78.672	128.986	0.928	

This study conducted on phosphorites, a promising REE-rich lithology (Emsbo et al., 2015; Buccione et al., 2021; Ferhaoui et al., 2022; Valetich et al., 2022), shows high REE prediction performance with ML models using major element oxides as input variables. The findings of this study demonstrate a highly favorable comparison with previously published

results across all geochemical phosphorite patterns (**Table 9)** with more details of the REE prediction results provided in **Tables 6, 7**, and **8**.

The lower error values and higher R^2 values achieved by dividing the dataset into oxic pattern and sub-oxic pattern subsets compared with those relating to the combined dataset (**Table 9**) confirm the value of differentiating the geochemical patterns associated with specific phosphorite deposits in the Tethyan region. The results also strongly support the concept that depositional conditions have exerted the primary control on the REE concentrations in the studied phosphorite deposits.

5.4. REE fractionation in relation paleogeographic distribution

The emergence of two distinct REE patterns in the studied phosphorite samples, the seawater-like pattern and the sub-oxic pattern are attributed to a combination of four factors: paleogeographic conditions, fluctuations in sea levels, geodynamic/tectonic evolution, and the Tethyan Circum-global Current (TCC). The first collision between the Afro-Arabian Plate and the intra-oceanic subduction zone of the Eurasian Plate resulted in a compressional event. That event created intra-plate paleo-highs and land uplift on the northern to northwestern Afro-Arabian platform, such as the Kasserine Island (Svoboda, 1989), the Ha'il, Rutba, and Sirhan highs, the Ga'ara Dome, and the Syrian Arc hinterland (Guiraud and Bosworth, 1997), Obstruction upwelling prevailed along these paleo highs, which are rich in silica and phosphorus. Moreover, the TCC is thought to have moved from east to west over the southern boundary of the Afro-Arabian continental shelf during the Late Cretaceous (Abed, 2013).

These factors suggest that the fractionation of REE across the Tethyan phosphorites was mainly controlled by the paleogeographic setting of the late Cretaceous-Eocene period, the CCT / upwelling currents, and locally by fluctuations in sea levels and paleoclimate. In the studied samples, the sub-oxic pattern was deposited adjacent to a coastline to the south, with high input of terrigenous materials leading to substantial enrichment of the REE content. In contrast, the oxic pattern phosphorite samples are located further to the north of this coastline (**Figures 1** and **2**).

Furthermore, the North African phosphorites deposits coincide with the Paleocene-Eocene Thermal Maximum (PETM), which features a warm and humid climate that increases the availability of dissolved REE ions in seawater via continental weathering (Hessler et al., 2017; Buccione et al., 2021).

In addition to the CFA, The principal controlling factors of REE concentrations predicted by the models are largely related to the roles of paleo-highs caused by tectonic events. Specifically, the influences of terrigenous sources for Fe_2O_3 and K_2O in the oxic pattern dataset, and MnO in the sub-oxic pattern dataset, are clearly distinguished by the XGBoost features importance F- scores related to ∑REE predictions (**Figures 11** and **12**). The CCT/upwelling currents also make a substantial contribution by providing a charge in phosphorus (CFA) and silica, with the latter having a strong negative influence on REE incorporation in the sub-oxic pattern dataset.

The foregoing findings confirm that the geochemical outcomes from this study are interpreted by the mentioned regional paleo-geographical insights involving the Tethyan phosphorites area. The influence of grain size on REE enrichment of phosphorites is less conclusive; although, grain-size influences are reported for Saudi Arabian and Egyptian phosphorite deposits studied (Ahmed et al., 2022; Edress et al., 2023).

In summary, the enrichment of REE in Tethyan phosphorites is significantly influenced by tectonic evolution and weathering processes. The formation of these REE-rich phosphorites is closely linked also to climatic conditions, particularly warm intervals and hyperthermal events (ex, Mazidaq deposits from Turkey), which play a pivotal role in promoting the development of authigenic mineralization (Hessler et al., 2017; Banerjee et al., 2020; Ferhaoui et al., 2022). Throughout the Tethyan province, multiple instances of elevated temperature events have been documented, spanning from the late Cretaceous to the Paleocene. Among these, the Paleocene-Eocene thermal maximum (PETM) period stands out as particularly significant. During this era, warm and wet climatic conditions prevailed. These conditions promoted continental weathering, resulting in an increased supply of dissolved ions, including REE, to the marine environment. The results of the feature importance analysis conducted with the XGBoost models are in agreement with these interpretations. The high feature importance assigned to MnO, SiO_2, and Fe_2O_3 underscores the value of these elements as indicators of tectonic evolution and depositional conditions.

5.5. Economic implications, limitations and future perspectives

The highly promising performance of the REE-prediction ML models developed offers the potential to substantially reduce the cost and time associated with REE resource exploration and exploitation. This is achievable through the rapid and precise mapping of REE contents within existing phosphorite deposits. In light of growing interest in REE-bearing phosphorites, there is a necessity to establish the REE concentration distributions for these resources. However, such information is lacking for several phosphorite deposits worldwide

even though cores exist in some cases. ML techniques offer the potential to successfully solve this issue by delivering accurate REE estimates based on readily available and cheap-to-acquire major element content data (Buccione et al., 2023). This approach trains ML models with data from detailed REE and major-element analysis on an initial, relatively small set of samples, and then applies the trained models to many more samples for which only major-element analysis is required.

By more comprehensively discerning the REE concentrations of phosphorite deposits, it becomes possible to reconstruct paleo-environmental conditions more accurately and reliably. Such reconstructions can be accomplished using major-element oxide measurements as dependable indirect proxies for REE concentrations and paleo-environment interpretations. XGBoost model feature importance scores can be confidently interpreted as a proxy for the depositional conditions that influence REE distributions in phosphorites.

However, the REE prediction models developed and applied to the compiled Tethyan phosphorite dataset, while promising, come with certain limitations. For example, the relatively limited spatial distributions of the available samples and incomplete analysis of input variables associated with certain samples limit its broader application with available data. This is a particular issue for certain known phosphorite deposits located in Algeria, Palestine, Syria, and Iraq. These limitations underscore the need for a more comprehensive regional sampling and geochemical study that integrates major- and trace-element analysis to further enhance the size and geographical coverage of the available datasets to use for ML training. Furthermore, future research should be undertaken to evaluate the Compositional Data (CoDa) approach using machine learning.

By integrating information relating to formation sedimentology, geodynamics, and tectonic evolution at multiple sites, through a multidisciplinary approach, including ML models, it should be possible to improve further the interpretation and prediction of REE distributions of phosphorites across the entire Tethyan region. Such a comprehensive approach planned for future studies should more effectively capture the complex interplays between geological, geochemical, and geodynamic factors, ultimately leading to more accurate and dependable REE mapping and predictions in phosphorites.

6. Conclusion

For the first time this study develops rare earth elements (REEs) concentration prediction models specifically tailored to regional Tethyan phosphorites. It employs a

geochemical classification approach and inverse model based on major element concentrations. It develops and applies machine-learning (ML) techniques to investigate REEs distributions in phosphorites, which has not been previously attempted. Nine major oxides concentrations were utilized as ML input features: SiO_2, MnO, Al_2O_3, Fe_2O_3, MgO, CaO, K_2O, TiO_2, and P_2O_5. These oxides were selected because they are fundamental components of phosphatic rock minerals with their analysis routinely reported. A data-centric approach was adopted to optimize ML model performance, involving the exploration of various combinations of the selected input features. Definitive results were obtained for specific phosphorite patterns. In oxic patterns, the predictions for light REE (LREE), heavy REE (HREE), and total REE (∑REE) demonstrated remarkable accuracy, yielding impressive metrics (expressed in ppm) of (R^2=0.979, RMSE=20.660, MAE=15.731), (R^2=0.961, RMSE=5.435, MAE=3.604 and R^2=0.978, RMSE=25.206, MAE=17.293), respectively.

Similarly, in sub-oxic patterns, the model excelled in predicting LREE, HREE, and ∑REE with relatively low errors (R^2=0.921, RMSE=100.137, MAE=62.326), (R^2=0.937, RMSE=19.897, MAE=13.526 and R^2=0.942, RMSE=105.040, MAE=72.590), respectively. For the combined patterns, the predictions also achieved acceptable accuracy: (R^2=0.925, RMSE=108.804, MAE=59.696), (R^2=0.930, RMSE=21.998, MAE=14.280 and R^2=0.928, RMSE=128.986, MAE=78.672) for LREE, HREE, and ∑REE predictions, respectively. Within the oxic pattern, the best parameters that predict REE distribution are Fe_2O_3 and K_2O. A dominant influence on REEs concentrations in phosphorites stems from the authigenic fraction (glauconite) and the elevated levels of elements linked to terrigenous sources, weathering, erosion, and the composition of soil materials. In sub-oxic conditions, REEs concentration are most positively influenced by MnO, Fe_2O_3 and most negatively influenced to SiO_2. The notable factor controlling ∑REE concentrations within the Tethyan phosphorite deposits is the presence of terrestrial materials with the reduction of Fe-Mn oxyhydroxides. SiO_2 originated primarily from siliceous plankton shells and algae, having a negative impact on REEs enrichment due to the intense late-stage diagenesis associated with silicification in the form of cherty layers and nodules. Additionally, a smaller contribution comes from clays and brine in the pore space.

The results of this study are in agreement with the hypothesis, already postulated in several studies on North African phosphorites, that the primary control of REEs enrichment is linked to the local deposition environments of Tethyan phosphorites (on a regional scale). Additionally, periodic fluctuations in ocean water chemistry may have exerted a marginal or secondary control on their REEs concentrations.

The occurrence of two distinct REEs patterns, the seawater-like pattern and the sub-oxic pattern of Tethyan phosphorites, is the result of a combination of significant triggers.

These factors include paleogeographic settings resulting from the geodynamic history of the Tethyan carbonate shelves, sea-level fluctuations, weathering processes and the effects of the Tethyan Circumglobal Current. The results of features importance established for the XGBoost models and the geochemical approach used are in agreement with the tectonic and regional paleogeographical observations. The importance of MnO, SiO_2, Fe_2O_3 are linked to their sensitivity to paleo-tectonic evolution and depositional conditions.

The REEs prediction models implemented in this paper show promising performance, potentially reducing costs and time for future economic ventures. They accurately map REE content in phosphorite deposits, enabling reconstruction of paleo-environmental conditions using major oxide measurements.

The REEs prediction models created in this study are subject to certain limitations attributed to limited spatial resolution and incomplete input variables within the deposits of certain countries in the Tethyan regions (i.e., Algeria, Palestine, Syria, and Iraq). Future research endeavors and sampling are planned with the intention of filling these data gaps and conducting comprehensive cross-validation studies to enhance the interpretability and reliability of these models. Considering the significance of spatial dependency in geochemical analysis, future research endeavors could focus on integrating spatial data to enhance the robustness of predictive models. Incorporating geographical coordinates of sampling locations (eastings ‘X’, northings ‘Y’, and altitude ‘Z’), into the independent / input features would allow for the exploration of spatial relationships between REE concentrations and geographic features, prove the reliability and improve the accuracy of predictive models. While our current study focuses on REE prediction from major oxides using machine learning, addressing spatial considerations in future research could contribute to a more comprehensive understanding and interpretability of geochemical processes from the REE prediction models.

Acknowledgments

This article is a part of the doctoral thesis research of Nasreddine Tahar-Belkacem. O.A. and R.K. are thankful to the University of Kasdi Merbah Ouargla for providing the financial support of short scientific trips to Italy.

Data availability: the data presented in this study are available on request from the corresponding author.

References

Abed, A.M., 2013. The eastern Mediterranean phosphorite giants: An interplay between tectonics and upwelling. GeoArabia (2013) 18 (2): 67–94.

Abed, A.M., 1989. On the genesis of the phosphorite-chert association of the Amman Formation in the Tel es Sur area, Ruseifa, Jordan. Sur la genèse de l'association phosphorite-chert de la Formation de Amman dans la région de Tel es Sur, Ruseifa, Jordanie. Sciences Géologiques, bulletins et mémoires 42, 141–153.

Abedini, A., Calagari, A.A., 2017. REEs geochemical characteristics of lower Cambrian phosphatic rocks in the Gorgan-Rasht Zone, northern Iran: Implications for diagenetic effects and depositional conditions. J. African Earth Sci 135, 115–124.

Adachi, M., Yamamoto, K., Sugisaki, R., 1986. Hydrothermal chert and associated siliceous rocks from the Northern Pacific: Their geological significance as indication of ocean ridge activity. Sedim. Geol. 47, 125–148.

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.

Ahmed, A.H., Aseri, A.A., Ali, K.A., 2022. Geological and geochemical evaluation of phosphorite deposits in northwestern Saudi Arabia as a possible source of trace and rare-earth elements. Ore Geol. Rev. 144, 104854.

Alibo, D.S., Nozaki, Y., 1999. Rare earth elements in seawater: particle association, shale-normalization, and Ce oxidation. Geochim. et Cosmochim. Acta 63, 363–372.

Amireh, B.S., Amaireh, M.N., Taha, S.A., Abed, A.M., 2019. Petrogenesis, provenance, and rare earth element geochemistry, southeast desert phosphorite, Jordan. J. African Earth Sci.150, 701–721.

Auer, G., Reuter, M., Hauzenberger, C.A., Piller, W.E., 2017. The impact of transport processes on rare earth element patterns in marine authigenic and biogenic phosphates. Geochim. et Cosmochim. Acta 203, 140–156.

Awadalla, G.S., 2010. Geochemistry and microprobe investigations of Abu Tartur REE-bearing phosphorite, Western Desert, Egypt. J. African Earth Sci.57, 431–443.

Baioumy, H., 2011. Rare earth elements and sulfur and strontium isotopes of upper Cretaceous phosphorites in Egypt. Cretac. Res. 32, 368–377.

Baioumy, H., 2002. Forms of iron in the phosphorites of Abu-Tartur Area, Egypt. Chinese Geochem. J. 21, 215–226.

Baioumy, H., Farouk, S., 2022. The geochemical and economic significance of REE in the Upper Cretaceous-Eocene Tethyan phosphorites. J. African Earth Sci.194, 104635.

Baioumy, H.M., Ahmed, A.H., Khedr, M.Z., 2014. A mixed hydrogenous and hydrothermal origin of the Bahariya iron ores, Egypt: Evidences from the trace and rare earth element geochemistry. J. Geochem. Explor. 146, 149–162.

Banerjee, S., Choudhury, T.R., Saraswati, P.K., Khanolkar, S., 2020. The formation of authigenic deposits during Paleogene warm climatic intervals: a review. J. Palaeogeogr. 9, 27.

Bansal, U., Banerjee, S., Ruidas, D.K., Pande, K., 2018. Origin and geochemical characterization of the glauconites in the Upper Cretaceous Lameta Formation, Narmada Basin, central India. J. Palaeogeogr. 7, 99–116.

Belayouni, H., Slansky, M., Trichet, J., 1990. A study of the organic matter in Tunisian phosphates series: Relevance to phosphorite genesis in the Gafsa Basin (Tunisia). Org. geochem. 15, 47–72.

Berker, E., 1989. The Mardin-Mazidagi-Derik phosphate deposits, south-eastern Turkey. Presented at the Phosphate deposits of the world. Phosphate rock resources, pp. 380–386.

Bertram, C., Elderfield, H., Aldridge, R., Morris, S.C., 1992. 87Sr/86Sr, 143Nd/144Nd and REEs in Silurian phosphatic fossils. EPSL 113, 239–249.

Boström, K., 1973. The origin and fate of ferromanganoan active ridge sediments., in: Stockholm contributions in geology. Acta Universitatis Stockholmiensis 27: 149-243.

Boujo, A., 1976. Contribution à l'étude géologique du gisement de phosphate crétacé-éocène des Ganntour (Maroc occidental). Persée-Portail des revues scientifiques en SHS.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Brew, G., Barazangi, M., Al-Maleh, A.K., Sawaf, T., 2001. Tectonic and geologic evolution of Syria. GeoArabia 6, 573–616.

Bright, C.A., Cruse, A.M., Lyons, T.W., MacLeod, K.G., Glascock, M.D., Ethington, R.L., 2009. Seawater rare-earth element patterns preserved in apatite of Pennsylvanian conodonts? Geochim. et Cosmochim. Acta 73, 1609–1624.

Buccione, R., Ameur-Zaimeche, O., Ouladmansour, A., Kechiched, R., Mongelli, G., 2023. Data-centric approach for predicting critical metals distribution: Heavy rare earth elements in cretaceous Mediterranean-type karst bauxite deposits, southern Italy. Geochem J 126026.

Buccione, R., Kechiched, R., Mongelli, G., Sinisi, R., 2021. REEs in the north africa P-bearing deposits, paleoenvironments, and economic perspectives: A review. Minerals 11.

Castor, S., 2008. The Mountain Pass rare-earth carbonatite and associated ultrapotassic rocks, California. Canadian Mineralogist – Can. Mineralog. 46, 779–806.

Chen, J., Algeo, T.J., Zhao, L., Chen, Z.-Q., Cao, L., Zhang, L., Li, Y., 2015. Diagenetic uptake of rare earth elements by bioapatite, with an example from Lower Triassic conodonts of South China. Earth-Sci. Rev. 149, 181–202.

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.

Collenette, P., Grainger, D.J., 1994. Mineral resources of Saudi Arabia: not including oil, natural gas, and sulfur. Ministry of Petroleum and Mineral Resources, Directorate General of Mineral ….

Condie, K.C., 1991. Another look at rare earth elements in shales. Geochim. et Cosmochim. Acta 55, 2527–2531.

Coombs, D. S, D., M., Grapes, R, K., Y., Roser, B. P, 1985. Geochemistry and origin of piedmontite-bearing and associated manganiferous schists from Arrow Junction, western Otago, in: New Zealand. Chem. Geol. 48: 57-78.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297.

Cracknell, M.J., Reading, A.M., 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.

Dill, H.G., Sachsenhofer, R.F., Grecula, P., Sasvári, T., Palinkaš, L.A., Borojević-Šoštarić, S., Strmić-Palinkaš, S., Prochaska, W., Garuti, G., Zaccarini, F., 2008. Fossil fuels, ore and industrial minerals.

Dong, J., Zou, H., Liu, H., Li, Y., Jiang, X., 2019. The Mesozoic Closure of the Paleo-Tethys Ocean: Constraints from Zircon U–Pb Ages, Hf and Nd Isotopes of The Granitic Plutons in the Hoh-Xil-Songpan-Ganzi. J. Coast. Res. 94, 101–106.

Dramsch JS, 2020. 70 years of machine learning in geoscience in review. In: Moseley B, Krischer L (eds) Machine Learning in Geosciences, ser. Adv Geophys 61:1–55. Elsevier., in: Chapter One.

Dumakor-Dupey, N.K., Arya, S., 2021. Machine Learning—A review of applications in mineral resource estimation. Energies 14.

Edress, N.A.A., Abdel-Rahman, E.A., Abdel-Wahab, M.G.F., 2023. Geochemical significance for the composition and depositional environments of the Campanian carbonate-rich phosphorite, Abu-Tartur plateau, Western Desert, Egypt. J. African Earth Sci. 202, 104938.

Einsele, G., 2001. Sedimentary basins: Evolution, facies, and sediment budget. Serbiula (sistema Librum 2.0) 143.

El Bamiki, R., Séranne, M., Chellaï, E.H., Merzeraud, G., Marzoqi, M., Melinte-Dobrinescu, M.C., 2020. The Moroccan High Atlas phosphate-rich sediments: Unraveling the accumulation and differentiation processes. Sediment. Geol. 403, 105655.

El Bamiki, R., Séranne, M., Parat, F., Aubineau, J., Chellaï, E.H., Marzoqi, M., Bodinier, J.-L., 2023. Post-phosphogenesis processes and the natural beneficiation of phosphates: Geochemical evidence from the Moroccan High Atlas phosphate-rich sediments. Chem. Geol. 631, 121523.

Elchellai, H., Marzoqi, M., Pascal, A., Mouflih, M., 1995. Stratigraphy and evolution of upper Cretaceous-Paleogene sedimentary systems in the Marrakesh high-atlas (Morocco). comptes rendus de l academie des sciences serie ii fascicule a-sciences de la terre et des planetes 321, 745–752.

El-Shazly, M., Mansour, A., Osman, M., Khaleal, F., 2022. Geology and geochemistry of the phosphorite deposits,Sibaiya,area,Nile,valley,Egypt. Arab J. Nucl. Sci.

Emsbo, P., McLaughlin, P.I., Breit, G.N., du Bray, E.A., Koenig, A.E., 2015. Rare earth elements in sedimentary phosphate deposits: Solution to the global REE crisis? Gondwana Res. 27, 776–785.

Esmaeiloghli, S., Tabatabaei, S. H., Hosseini, S., Deville, Y., & Carranza, E. J. M. (2023). Blind Source Separation of Spectrally Filtered Geochemical Signals to Recognize Multi-depth Ore-Related Enrichment Patterns. *Mathematical Geosciences*. https://doi.org/10.1007/s11004-023-10101-w

Esmaeiloghli, S., Tabatabaei, S. H., & Carranza, E. J. M. (2023). Infomax-based deep auto-encoder network for recognition of multi-element geochemical anomalies linked to mineralization. *Computers & Geosciences*, *175*, 105341. https://doi.org/10.1016/j.cageo.2023.105341

Essa, M.A., Ahmed, E.A., Kurzweil, H., 2016. Genesis, maturity and weathering of some Upper Cretaceous Egyptian glauconites: Mineralogical and geochemical implications. J. African Earth Sci. 124, 427–446.

European Commission, 2020. Critical Raw Materials Resilience: Charting a Path towards greater Security and Sustainability.

Ferhaoui, S., Kechiched, R., Bruguier, O., Sinisi, R., Kocsis, L., Mongelli, G., Bosch, D., Ameur-Zaimeche, O., Laouar, R., 2022. Rare earth elements plus yttrium (REY) in phosphorites from the Tébessa region (Eastern Algeria): Abundance, geochemical distribution through grain size fractions, and economic significance. J. Geochem. Explor. 241, 107058.

Föllmi, K., 1996. The phosphorus cycle, phosphogenesis and marine phosphate-rich deposits. Earth-Sci. Rev. 40, 55–124.

Föllmi, K.B., Weissert, H., Lini, A., 1993. Nonlinearities in phosphogenesis and phosphorus-carbon coupling and their implications for global change. Presented at the Interactions of C, N, P and S Biogeochemical Cycles and Global Change, Springer, pp. 447–474.

Fortier, S.M., Nassar, N.T., Lederer, G.W., Brainard, J., Gambogi, J., McCullough, E.A., 2018. Draft critical mineral list—Summary of methodology and background information—U.S. Geological Survey technical input document in response to Secretarial Order No. 3359 (Report No. 2018–1021), Open-File Report. Reston, VA.

Garnit, H., Bouhlel, S., Barca, D., Chtara, C., 2012. Application of LA-ICP-MS to sedimentary phosphatic particles from Tunisian phosphorite deposits: Insights from trace elements and REE into paleo-depositional environments. Geochem J 72, 127–139.

Garnit, H., Bouhlel, S., Jarvis, I., 2017. Geochemistry and depositional environments of Paleocene–Eocene phosphorites: Metlaoui Group, Tunisia. J. African Earth Sci.134, 704–736.

Ghasemian, S., Öztürk, H., Cansu, Z., 2022. Geochemistry of red and cream phosphorites from the Şemikan phosphorite deposit, SE Turkey: Implication for phosphorite deposition conditions in the Upper Cretaceous. J. African Earth Sci. 185, 104398.

Glenn, C., Föllmi, K., Riggs, S., Baturin, G., Grimm, K., Trappe, J., Abed, A., Galli-Olliver, C., Garrison, R., Ilyin, A., 1994. Phosphorus and phosphorites: Sedimentology and Environments of Formation: Eclogae Geol. Helv. v. 87.

Guiraud, R., Bosworth, W., 1997. Senonian basin inversion and rejuvenation of rifting in Africa and Arabia: synthesis and implications to plate-scale tectonics. Tectonophysics 282, 39–82.

Gundogar, D.Y., Sasmaz, A., 2022. Geochemical Approach to Determine the Possible Precipitation Parameters of the Coniacian–Santonian Mazıdağı Phosphates, Mardin, Turkey. Minerals 12.

Haddi, H.E., Benbouziane, A., Mouflih, M., Jourani, E., Amaghzaz, M., 2014. Siliceous Forms of Phosphate Deposits of Cretaceous Age in Oulad Abdoun Basin (Central Morocco). Mineralogy, geochemistry and diagenetic phenomena. Procedia Eng. 83, 60–69.

Halalat, H., Bolourchi, M., 1994. Geology of Iran: phosphate. Geological Survey of Iran, Tehran.

Haq, B., Al-Qahtani, A., 2005. Phanerozoic cycles of sea-level change on the Arabian Platform. GeoArabia 10.

He, S., Xia, Y., Xiao, J., Gregory, D., Xie, Z., Tan, Q., Yang, H., Guo, H., Wu, S., Gong, X., 2022. Geochemistry of REY-Enriched Phosphorites in Zhijin Region, Guizhou Province, SW China: Insight into the Origin of REY. Minerals 12.

Herbig, H.-G., Trappe, J., 1994. Stratigraphy of the Subatlas Group (Maastrichtian-Middle Eocene, Morocco). Newsl. Stratigr. 125–165.

Hessler, A.M., Zhang, J., Covault, J., Ambrose, W., 2017. Continental weathering coupled to Paleogene climate changes in North America. Geology 45, 911–914.

Hong Jin, GAN Chengshi, LIU Jie, 2019. Prediction of REEs in OIB by major elements based on machine learning. Earth Sci. Front.. 2019(04) Page:45-54.

Huang, S., Ning, S., Zhang, J., Zhang, L., Liu, K., 2021. REE characteristics of the coal in the Erlian Basin, Inner Mongolia, China, and its economic value. China Geol. 4, 256–265.

Humphries, M., 2010. Rare earth elements: the global supply chain. Diane Publishing.

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.

Jarvis, I., Burnett, W., Nathan, Y., Almbaydin, F., Attia, A., Castro, L., Flicoteaux, R., Hilmy, M.E., Husain, V., Qutawnah, A., 1994. Phosphorite geochemistry: state-of-the-art and environmental concerns. Eclogae Geol. Helv. 87, 643–700.

Ji, B., Li, Q., Tang, H., Zhang, W., 2022. Rare earth elements (REEs) recovery from coal waste of the Western Kentucky No. 13 and Fire Clay seams. Part II: Re-investigation on the effect of calcination. Fuel 315, 123145.

J.R. Quinlan, 1993. programs for machine learning The Morgan Kaufmann Series in Machine Learning, Morgan Kaufmann, San Mateo, CA (1993) (No. C 4.5:).

Kadavi, S.L.R., 2019. Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models Prima Riza Kadavi, Chang-Wook Lee,.

Kechiched, R., Laouar, R., Bruguier, O., Kocsis, L., Salmi-Laouar, S., Bosch, D., Ameur-Zaimeche, O., Foufou, A., Larit, H., 2020. Comprehensive REE + Y and sensitive redox trace elements of Algerian phosphorites (Tébessa, eastern Algeria): A geochemical study and depositional environments tracking. J. Geochem. Explor. 208, 106396.

Kechiched, R., Laouar, R., Bruguier, O., Laouar-Salmi, S., Ameur-Zaimeche, O., Foufou, A., 2016. Preliminary Data of REE in Algerian Phosphorites: A Comparative Study and Paleo-redox Insights. Procedia Eng. 138, 19–29.

Kechiched, R., Laouar, R., Bruguier, O., Salmi-Laouar, S., Kocsis, L., Bosch, D., Foufou, A., Ameur-Zaimeche, O., Larit, H., 2018. Glauconite-bearing sedimentary phosphorites from the Tébessa region (eastern Algeria): Evidence of REE enrichment and geochemical constraints on their origin. J. African Earth Sci. 145, 190–200.

Kocsis, L., Ounis, A., Baumgartner, C., Pirkenseer, C., Harding, I.C., Adatte, T., Chaabani, F., Neili, S.M., 2014. Paleocene–Eocene palaeoenvironmental conditions of the main phosphorite deposits (Chouabine Formation) in the Gafsa Basin, Tunisia. J. African Earth Sci. 100, 586–597.

Lécuyer, C., Reynard, B., Grandjean, P., 2004. Rare earth element evolution of Phanerozoic seawater recorded in biogenic apatites. Chem. Geol. 204, 63–102.

Li, L., Solana, C., Canters, F., Kervyn, M., 2017. Testing random forest classification for identifying lava flows and mapping age groups on a single Landsat 8 image. J.Volcanol.Geotherm. 345, 109–124.

Linares, E., Velasquez, G., Manrique, J., Monsalve, J., Lo Mónaco, S., Shumlyanskyy, L., 2023. REE + Y signatures of the Navay phosphate deposit, SW Venezuela: Seawater paleoredox conditions and diagenetic implications. J.South Am. Earth Sci. 129, 104532.

Long, K., Gosen, B., Foley, N., Cordier, D., 2010. The Principal rare earth elements deposits of the united states: A summary of domestic deposits and a global perspective, cultural analysis: an interdisciplinary forum on folklore and popular culture.

Lucas, J., Prévôt-Lucas, L., 1996. Tethyan Phosphates and Bioproductites, in: Nairn, A.E.M., Ricou, L.-E., Vrielynck, B., Dercourt, J. (Eds.), The Tethys Ocean. Springer US, Boston, MA, pp. 367–391.

Martin, E., Scher, H., 2004. Preservation of seawater Sr and Nd isotopes in fossil fish teeth: bad news and good news. EPLS 220, 25–39.

Marzoqi, M., Pascal, A., 2000. Séquences de dépôts et tectono-eustatisme à la limite Crétacé/Tertiaire sur la marge sud-téthysienne (Atlas de Marrakech et bassin de Ouarzazate, Maroc). Newsl. Stratigr. 57–80.

Mastalerz, M., Drobniak, A., Eble, C., Ames, P., McLaughlin, P., 2020. Rare earth elements and yttrium in Pennsylvanian coals and shales in the eastern part of the Illinois Basin. International Journal of Coal Geology 231, 103620. Mclennan, S.M., 1989. Rare earth elements in sedimentary rocks; influence of provenance and sedimentary processes. Rev. Mineral. Geochem.21, 169–200.

Mondillo, N., Balassone, G., Boni, M., Chelle-Michou, C., Cretella, S., Mormone, A., Putzolu, F., Santoro, L., Scognamiglio, G., Tarallo, M., 2019. Rare earth elements (REE) in Al- and Fe-(Oxy)-Hydroxides in bauxites of provence and languedoc (Southern France): implications for the potential recovery of REEs as by-products of bauxite mining. Minerals 9.

Mongelli, G., Boni, M., Buccione, R., Sinisi, R., 2014. Geochemistry of the Apulian karst bauxites (southern Italy): Chemical fractionation and parental affinities. Ore Geol. Rev. 63, 9–21.

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.

Nathan, Y., 1984. The Mineralogy and Geochemistry of Phosphorites, in: Nriagu, J.O., Moore, P.B. (Eds.), Phosphate Minerals. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 275–291.

N.M.A. Report, 2010. Economic evaluation of U and REEs in Egyptian phosphorite ores.Nuclear Materials Authority Report, Cairo, Egypt. 72 p.

Notholt, A., 1985. Phosphorite resources in the Mediterranean (Tethyan) phosphogenic province: a progress report. Sciences Géologiques, bulletins et mémoires 77, 9–17.

Notholt, A.J.G., Sheldon, R.P., Davidson, D.F., 1989. Phosphate Deposits of the World: Phosphate Rock Resources. Cambridge University Press Cambridge.

Nozaki, Y., Zhang, J., Amakawa, H., 1997. The fractionation between Y and Ho in the marine environment. EPLS 148, 329–340.

Ohta, A., Kawabe, I., 2000. Theoretical study of tetrad effects observed in REE distribution coefficients between marine Fe-Mn deposit and deep seawater, and in REE(III)-carbonate complexation constants. Geochem. J. 34, 455–473.

Parsa, M., 2021. A data augmentation approach to XGboost-based mineral potential mapping: An example of carbonate-hosted ZnPb mineral systems of Western Iran. J. Geochem. Explor. 228, 106811.

Picard, S., Lécuyer, C., Barrat, J.-A., Garcia, J.-P., Dromart, G., Sheppard, S.M., 2002. Rare earth element contents of Jurassic fish and reptile teeth and their potential relation to seawater composition (Anglo-Paris Basin, France and England). Chem. Geol. 186, 1–16.

Powell, J., 1989. Stratigraphy and sedimentation of the Phanerozoic rocks in central and south Jordan. Division, Directorate, Authority, Ministry.

Research and Markets, 2023. Rare Earth Metals Global Market Report 2023.

Reynard, B., Lécuyer, C., Grandjean, P., 1999. Crystal-chemical controls on rare-earth element concentrations in fossil biogenic apatites and implications for paleoenvironmental reconstructions. Chem. Geol. 155, 233–241.

Ribeiro, B. O. L., Barbuena, D., & de Melo, G. H. C. (2023). Geochemical multifractal modeling of soil and stream sediment data applied to gold prospectivity mapping of the Pi-

tangui Greenstone Belt, northwest of Quadrilátero Ferrífero, Brazil. *Geochemistry*, *83*(2), 125951. https://doi.org/10.1016/j.chemer.2023.125951

Riddler, G., Van Eck, M., Farasani, A., 1989. The phosphorite deposits of the Sirhan-Turafy region, northern Saudi Arabia. Presented at the Phosphate deposits of the world. Phosphate rock resources, pp. 332–337.

Sadeghi, B. (2022). Chatterjee Correlation Coefficient : A robust alternative for classic correlation methods in geochemical studies- (including "TripleCpy" Python package). *Ore Geology Reviews*, *146*, 104954. https://doi.org/10.1016/j.oregeorev.2022.104954

Saporetti, C.M., da Fonseca, L.G., Pereira, E., de Oliveira, L.C., 2018. Machine learning approaches for petrographic classification of carbonate-siliciclastic rocks using well logs and textural information. Appl. Geophys. J. 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.

Scott B, S.N., 2019. Machine learning in geology.

Şengör, A.C., Yilmaz, Y., 1981. Tethyan evolution of Turkey: a plate tectonic approach. Tectonophysics 75, 181–241.

Sharland, P., Archer, R., Casey, D., Davies, R., Hall, S.H., Heward, A., Horbury, A., Simmons, M., 2001. Arabian plate sequence stratigraphy. GeoArabia, GeoArabia Special Publication.

Sheldon, L.E., 1988. Evaluating ELT textbooks and materials. ELT journal 42, 237–246.

Shields, G., Stille, P., 2001. Diagenetic constraints on the use of cerium anomalies as paleoseawater redox proxies: an isotopic and REE study of Cambrian phosphorites. Chem. Geol. 175, 29–48.

Shields, G.A., Webb, G.E., 2004. Has the REE composition of seawater changed over geological time?

Sholkovitz, E.R., Landing, W.M., Lewis, B.L., 1994. Ocean particle chemistry: the fractionation of rare earth elements between suspended particles and seawater. Geochimi. et Cosmochim. Acta 58, 1567–1579.

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). https://doi.org/10.3390/min12080941

Slansky, M., 1986. Geology of sedimentary phosphates.

S.M. Jasinski, 2011. Phosphate rock mineral commodity summaries 2011, United States Geological Survey, United States Government Printing Office, Washington, D.C.

Smola, A., 1998. Learning with kernels.

Svoboda, K, 1989. The lower Tertiary phosphate deposits of Tunisia. In A.J.G. Notholt, R.P. Sheldon and D.F. Davidson (Eds.), Phosphate Deposits of the World,.

Tang, J., Johannesson, K.H., 2010. Ligand extraction of rare earth elements from aquifer sediments: Implications for rare earth element complexation with organic matter in natural waters. Geochim. et Cosmochim. Acta 74, 6690–6705.

Taylor, S.R., McLennan, S.M., 1985. The continental crust: its composition and evolution.

Tianqi Chen, C.G., 2016. XGBoost: A Scalable Tree Boosting System.

Trappe, J., 1991. Stratigraphy, facies distribution and paleogeography of the marine Paleogene from the Western High Atlas, Morocco. Neues Jahrb Geol Palaontol Abh 279–321.

Valetich, M., Zivak, D., Spandler, C., Degeling, H., Grigorescu, M., 2022. REE enrichment of phosphorites: An example of the Cambrian Georgina Basin of Australia. Chem. Geol. 588, 120654.

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.

Wright, J., Schrader, H., Holser, W.T., 1987. Paleoredox variations in ancient oceans recorded by rare earth elements in fossil apatite. Geochim. et Cosmochim. Acta 51, 631–644.

Y. K. Bentor, 1980. Phosphorites—the unsolved problems, in: Marine phosphorites.

Yanfei, X., Zongyu, F., Xiaowei, H., Li, H., Yingying, C., Xiangsheng, L., Liangshi, W., Zhiqi, L., 2016. Recovery of rare earth from the ion-adsorption type rare earths ore: II. Compound leaching. Hydrometallurgy 163, 83–90.

Yang, H., Zhao, Z., Xia, Y., Xiao, J., 2021. REY enrichment mechanisms in the early Cambrian phosphorite from South China. Sedim. Geol. 426, 106041.

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.

Zaïer, A., Beji-Sassi, A., Sassi, S., Moody, R., 1998. Basin evolution and deposition during the Early Paleogene in Tunisia. J. Geol. Soc., London, Special Publications 132, 375–393.

Zaremotlagh, S., Hezarkhani, A., 2017. The use of decision tree induction and artificial neural networks for recognizing the geochemical distribution patterns of LREE in the Choghart deposit, Central Iran. J. African Earth Sci. 128, 37–46.

Zhang, B., Li, M., Li, W., Jiang, Z., Khan, U., Wang, L., Wang, F., 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, 1422–1447.

Zhang, J., Nozaki, Y., 1996. Rare earth elements and yttrium in seawater: ICP-MS determinations in the East Caroline, Coral Sea, and South Fiji basins of the western South Pacific Ocean. Geochim. et Cosmochim. Acta 60, 4631–4644.

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.

Zhang, Y., Li, Ziying, Dini, S.M., Qin, M., Banakhar, A.S., Li, Zhixing, Yi, L., Memesh, A.M., Shammari, A.M., Li, G., 2021. Origin and evolution of the late Cretaceous reworked phosphorite in the Sirhan-Turayf basin, Northern Saudi Arabia. Minerals 11, 350.

Zhang, P., Zhang, Z., Yang, J., & Cheng, Q. (2023). Machine Learning Prediction of Ore Deposit Genetic Type Using Magnetite Geochemistry. *Natural Resources Research*, *32*(1), 99-116. https://doi.org/10.1007/s11053-022-10146-4

Zhou, B., Li, Z., Zhao, Y., Zhang, C., Wei, Y., 2016. Rare earth elements supply vs. clean energy technologies: new problems to be solve. Gospod. 32, 29–44.

Zhou, J., Yu, W., Wei, W., Yang, M., Du, Y., 2023. Provenance and tectonic evolution of Bauxite deposits in the Tethys: Perspective from Random Forest and Logistic Regression analyses. Geochemistry Geophys. Geosystems 24, e2022GC010745.

Zhu, B., Jiang, S.-Y., Yang, J.-H., Pi, D., Ling, H.-F., Chen, Y.-Q., 2014. Rare earth element and SrNd isotope geochemistry of phosphate nodules from the lower Cambrian Niuti-

tang formation, NW Hunan province, south China. Palaeogeogr. Palaeoclimatol. Palaeoecol. 398, 132–143.

Zuo, R., 2017. Machine Learning of mineralization-related geochemical anomalies: A review of potential methods. Nat. Resour. Res. 26, 457–464.

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

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Graphical abstract :