Geochimica et Cosmochimica Acta 364 (2024) 1–9

Contents lists available at ScienceDirect

Geochimica et Cosmochimica Acta

journal homepage: www.elsevier.com/locate/gca

Trace element evidence for diverse origins of superheavy pyrite in Neoproterozoic sedimentary strata

Junyao Kang [a,b], Daniel D. Gregory [c], Benjamin Gill [a,b], Shiqiang Huang [d], Changxin Lai [e], Zhaoshan Chang [d], Huan Cui [f], Ivan Belousov [g], Shuhai Xiao [a,b,*]

[a] Department of Geosciences, Virginia Tech, Blacksburg, VA, USA
[b] Global Change Center, Virginia Tech, Blacksburg, VA, USA
[c] Department of Earth Sciences, University of Toronto, Toronto, Ontario, Canada
[d] Department of Geology and Geological Engineering, Colorado School of Mines, Golden, CO, USA
[e] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
[f] Department of Geosciences, Mississippi State University, Mississippi State, MS, USA
[g] CODES (ARC Centre of Excellence in Ore Deposits), School of Physical Sciences, University of Tasmania, Hobart, Tasmania, Australia

ARTICLE INFO

Associate editor: Eva Stüeken

Keywords:
Pyrite
Trace Element
LA-ICP-MS
Machine Learning

ABSTRACT

Sedimentary pyrite has long been used as an archive of marine environments in Earth history. To capture reliable paleoenvironmental signals, however, we need to first evaluate pyrite in sedimentary strata as it can be altered and masked by later diagenetic and/or hydrothermal processes. Here, we trained two supervised machine learning algorithms on a large LA-ICP-MS pyrite trace element database to distinguish pyrite of different origins. The analysis validates that two models built on the co-behavior of 12 trace elements (Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Au, Tl, and Pb) can be used to accurately predict pyrite origins. Further statistical analysis suggests four trace element clusters behaving differently among sedimentary (syngenetic and early diagenetic), synsedimentary hydrothermal (syngenetic hydrothermal), and post-sedimentary hydrothermal (epigenetic hydrothermal) pyrite, which is probably driven by chemical and physical properties of source fluids, interactions between elements, competition among coprecipitating minerals, and pyrite growth rate. Armed with this initial success and aided by new LA-ICP-MS trace element data from 9 samples, we then demonstrated the efficacy of this approach in identifying the origins of pyrite from two Neoproterozoic sedimentary successions in South China. The first set of samples contain isotopically superheavy pyrite (i.e., whose bulk-sample $\delta^{34}S$ values greater than those of contemporaneous seawater sulfate and whose origins remain controversial) from the Cryogenian Tiesi'ao and Datangpo formations. The second set of samples contain pyritic rims (associated with fossiliferous chert nodules and thought to be critical in exceptional fossil preservation) from the Ediacaran Doushantuo Formation. For the superheavy pyrite, the models consistently show high confidence levels (mostly $>$ 80 % probability) in identifying its genesis type, and three out of four samples were given sedimentary origins. For the pyritic nodule rims, the models suggest that early diagenetic pyrite was subsequently altered by hydrothermal fluids and therefore shows mixed signals. The study highlights the importance of pyrite trace elements in deciphering and distinguishing the origins of pyrite in sedimentary strata.

1. Introduction

Pyrite is a ubiquitous sulfide mineral in organic-rich marine sediments and sedimentary rocks. Its morphology, abundance, trace element content, and sulfur isotope composition have been widely used to reconstruct oceanic redox conditions and the sulfur cycle (Fike et al., 2015; Gregory, 2020; Lyons et al., 2003; Wilkin et al., 1996), which have played important roles in the evolution of life (Anbar and Knoll, 2002). However, the origins of pyrite can be diverse and controversial. Specifically, the presence of hydrothermal pyrite and hydrothermally overprinted sedimentary pyrite can compromise the reliability of the pyrite as a paleoenvironmental proxy, because the chemical and isotopic compositions of such pyrite may reflect in part or in whole the hydrothermal fluids rather than ocean waters.

* Corresponding author.
E-mail address: xiao@vt.edu (S. Xiao).

https://doi.org/10.1016/j.gca.2023.11.005
Received 19 May 2023; Accepted 4 November 2023
Available online 7 November 2023

One prominent case is the origin of isotopically superheavy pyrite — pyrite with bulk-sample $\delta^{34}S$ values greater than those of contemporaneous seawater sulfate — that is widely present in Neoproterozoic and particularly Cryogenian sedimentary rocks (Cui et al., 2018; Fike et al., 2015; Lang et al., 2021; Ries et al., 2009). Sedimentary pyrite formation is driven by microbial sulfate reduction where microorganisms use dissolved sulfate as a terminal electron acceptor to oxidize organic matter, producing sulfide (Berner, 1970). This sulfide then goes on to react with iron to eventually form pyrite. Sedimentary pyrite typically has bulk-sample $\delta^{34}S$ values lower than seawater sulfate because microbial sulfate reduction prefers the lighter sulfur isotope (Habicht and Canfield, 1997). Thus, although rare occurrences of superheavy pyrite have been reported from Phanerozoic rocks and modern environments (Cadeau et al., 2022; Ferrini et al., 2010; Hu et al., 2022b), its common occurrence in the Neoproterozoic has been puzzling and several hypotheses have been proposed to explain its origins (Cai et al., 2022; Liu et al., 2006; Peng et al., 2014; Ries et al., 2009; Shen et al., 2008; Tostevin et al., 2017). For example, superheavy pyrite in the cap dolostone atop late Cryogenian (Marinoan) glacial diamictite was interpreted as a result of oceanic stratification where the bottom-water sulfate inherited high $\delta^{34}S$ values derived from *syn*-glacial isotopic distillation, but the surface-water sulfate had lower $\delta^{34}S$ values derived from post-glacial continental weathering (Shen et al., 2008). In this scenario, carbonate-associated sulfate and pyrite in the cap dolostone were derived from the surface- and bottom-water, respectively, and are thus isotopically decoupled (Shen et al., 2008). Other researchers attributed superheavy pyrite in Ediacaran carbonate rocks to low seawater sulfate concentrations and intense aerobic reoxidation of sulfide (Ries et al., 2009). Still others proposed that superheavy pyrite in the Cryogenian Datangpo Formation overlying Sturtian (early Cryogenian) glacial diamictite was related to the oxidative loss of early ^{34}S-depleted pyrite and the formation of diagenetic pyrite from ^{34}S-enriched sulfate diffusing downward from the deglaciation-facilitated euxinic waters (Cai et al., 2022). On the other hand, the sedimentary origin of Neoproterozoic superheavy pyrite has been questioned by some researchers. New in-situ sulfur isotope measurements suggest that some Cryogenian superheavy pyrite was derived from thermogenic sulfate reduction, hence it is likely late diagenetic or hydrothermal in origin and irrelevant to the oceanic environment (Cui et al., 2018). Therefore, it is important to independently test the origin of Neoproterozoic superheavy pyrite; if it is of hydrothermal origin, then it does not reflect the Neoproterozoic sulfur cycle and marine environments.

Trace element concentrations in pyrite can be used to constrain its origin since trace element enrichment or depletion reflects the formation mechanism and the composition of the source fluid (Gregory, 2020; Gregory et al., 2015; Gregory et al., 2019). Furthermore, these trace elements are relatively stable in unaltered pyrite and can be preserved even in mid-greenschist facies (Large et al., 2009). The challenge, however, is that the behavior of trace elements in pyrite is complex and no simple criteria can be used to distinguish the different origins of pyrite. In other words, different types of pyrite often have overlapping trace element concentrations. For example, gold can have high concentrations in both early diagenetic pyrite and late metamorphic or hydrothermal pyrite (Large et al., 2009). To overcome this challenge, multiple trace elements should be compared simultaneously, which can be implemented using rigorous statistical analysis (Gregory et al., 2019).

In this study, we used supervised machine learning, in combination with detailed petrographic analysis, to decipher the complex relationships between trace element patterns and the origin of pyrite. First, as pyrite grains from the same rock sample can have different origins, the study of pyrite origins should focus on individual pyrite grains. Hence, in-situ trace element analysis techniques, such as laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS) and secondary ion mass spectrometry (SIMS), were used for geochemical analysis of individual pyrite grains. Second, we used 12 trace elements (Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Au, Tl, and Pb) as input features to develop an optimal classification model to identify the genesis type of each pyrite. These 12 elements were selected based on data availability and previous machine learning applications (Gregory et al., 2019); incorporating a greater number of elements would reduce the database size and degrade model performance (Gregory et al., 2019). Two widely used supervised classification algorithms, Random Forests (Breiman, 2001) and XGBoost (Chen and Guestrin, 2016), were trained/optimized on a large LA-ICP-MS pyrite trace element database (Gregory et al., 2019). The present study followed the standard machine learning practices, similar to those utilized in other investigations in the field of geochemistry (Chen et al., 2022; Hu et al., 2022a; Li et al., 2023; Mukherjee et al., 2023).

The labeled training dataset for supervised machine learning is adapted from LA-ICP-MS pyrite trace element data compiled by Gregory et al. (2019). This compilation includes sedimentary pyrite through geological history and hydrothermal pyrite from different ore deposits. For the purpose of this study, all data are grouped and labeled into three classes: sedimentary, synsedimentary hydrothermal, and post-sedimentary hydrothermal pyrite. First, sedimentary pyrite in this database consists of syngenetic and early diagenetic pyrite. Samples have been rigorously screened using independent petrographic and geochemical criteria to exclude large inclusions, later hydrothermal overprint, recrystallization, or metamorphism (Gregory et al., 2019). Second, synsedimentary hydrothermal pyrite, also known as syngenetic hydrothermal pyrite, is defined as pyrite precipitating on or near the seafloor but partially influenced by hydrothermal fluids. Here hydrothermal fluids refer to either fluids from volcanic-dominated systems (volcanogenic massive sulfide, VMS) (Hannington, 2014) or sedimentary basinal brines (sedimentary exhalative deposits, SEDEX) (Emsbo et al., 2016). Lastly, post-sedimentary hydrothermal pyrite, also known as epigenetic hydrothermal pyrite, refers to pyrite found in igneous or metamorphosed rocks without contemporaneous seawater inputs. Hydrothermal pyrite in iron oxide copper gold (IOCG), orogenic Au, porphyry Cu, and hydrothermal breccia deposits are included in this category. We note that this is not an exhaustive labeling of pyrite precipitating in different conditions, but it is sufficiently diverse to help us understand the differences among these three classes of pyrite. Meanwhile, both in-sample-locality and out-of-sample-locality test datasets were constructed from the same trace element database. The in-sample-locality test dataset includes data from geologic units/deposits that were part of the training dataset. The out-of-sample-locality test dataset includes pyrite analyses from geologic units/deposits that were not part of the training dataset, which can help us better evaluate the generalizability of the models on unseen instances, including unencountered geologic time intervals and ore deposits.

After training, classification models (Random Forests and XGBoost) were used to determine the origins of two sets of ambiguous Neoproterozoic pyrite samples with existing petrographic and SIMS $\delta^{34}S$ data, and to understand their Neoproterozoic marine environments. The first set came from the Cryogenian Tiesi'ao and Datangpo formations in South China. The SIMS $\delta^{34}S_{pyrite}$ data were previously reported in Cui et al. (2018), who proposed a hydrothermal origin of the superheavy pyrite in these units. The second sample set came from the Ediacaran Doushantuo Formation in South China. Its SIMS $\delta^{34}S_{pyrite}$ data were previously reported in Xiao et al. (2010). These authors argued that rapid microbial sulfate reduction during early diagenesis facilitated the precipitation of pyrite and silica, forming the pyritic rims and fossiliferous chert nodules. The proposed mechanism may improve the understanding of the taphonomic role of silicification in the preservation of Ediacaran fossils. In this study, we used LA-ICP-MS to analyze these two sample sets for trace element concentrations in pyrite, and then used machine learning models to assess the origins of pyrite grains in the two sets of Neoproterozoic samples.

2. Materials and methods

2.1. Ambiguous pyrite from Neoproterozoic strata

A total of nine best-preserved and most-representative samples were analyzed in this study for pyrite trace element concentrations using LA-ICP-MS (see Appendix A. Supplemental Material for geological settings). Four of these samples (Hy1, Hy31, Hy55a, and Hy59) came from a drill core (4.5 cm in diameter) of the Cryogenian Tiesi'ao and Datangpo formations in South China with SIMS $\delta^{34}S_{pyrite}$ data reported in Cui et al. (2018). Hy1 is a shale sample from the Datangpo Formation that overlies the Sturtian-age Tiesi'ao glacial diamictite. Two types of pyrite are present in this sample. The first was named pyrite flowers (Cui et al., 2018), with a framboidal pyrite core surrounded by zoned pyrite overgrowth (Fig. 1 and Fig. S1). The second is characterized by a framboidal pyrite core, a thin pyrite rim, and an iron oxide layer in-between (Fig. 1 and Fig. S1). Pyrite flowers show abnormally high $\delta^{34}S$ values (mean: +61.6 ‰) and were interpreted to be hydrothermal in origin (Cui et al., 2018), whereas pyrite cores within iron oxide show much lower $\delta^{34}S$ values (mean: +22.2 ‰) and were thought to have formed through syngenetic or early diagenetic processes (Cui et al., 2018). Hy31 is also a shale sample from the Datangpo Formation. There is no framboidal pyrite in this sample. Instead, most pyrite grains are subhedral and some grains can reach millimeters in size, with micrometer-sized sub-grains cemented by later-stage pyrite (Fig. 1 and Fig. S2). High $\delta^{34}S$ values of pyrite grains (mean: +66.3 ‰) and cement (mean: +62.0 ‰) were suggested to reflect hydrothermal processes (Cui et al., 2018). Hy55a is a Mn-rich carbonate sample from the basal Datangpo Formation. In this sample, nodule- and sausage-shaped pyrite replaced (and thus post-dates) pre-existing rhodochrosite layers and lath-shaped illite crystals, and inherited the texture of the hosting rhodochrosite (Fig. S3). The pyrite is composed of individual framboids and lacy overgrowth (Fig. 1 and Fig. S3G). On the basis of petrographic observations (e.g., pyrite postdating rhodochrosite and illite) and high $\delta^{34}S$ values of both the framboids (mean: +56.9 ‰) and the overgrowth (mean: +57.8 ‰), Cui et al. (2018) argued for a hydrothermal origin of pyrite in Hy55a. Hy59 came from the uppermost Tiesi'ao Formation, which consists of glacial diamictite deposited during the Sturtian snowball Earth glaciation.

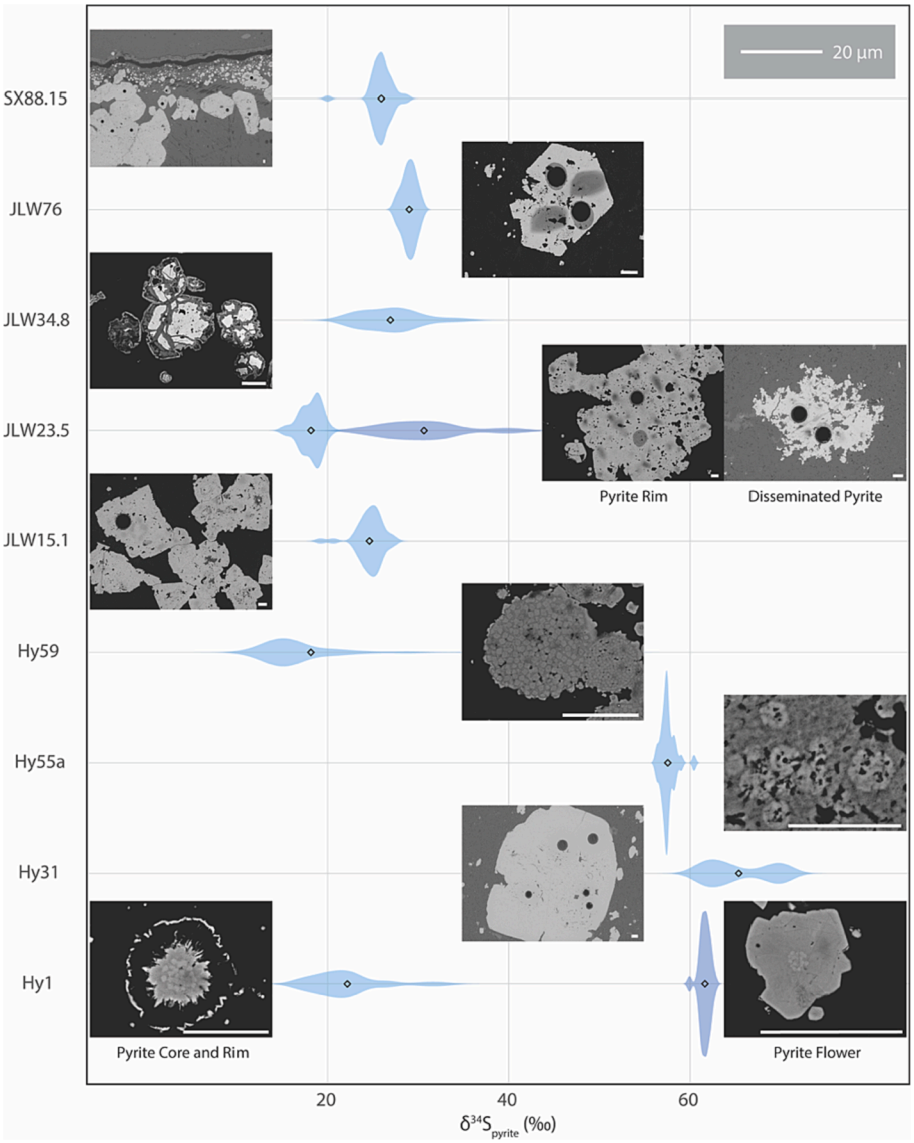

Fig. 1. SIMS $\delta^{34}S_{pyrite}$ data and representative BSE-SEM images of samples analyzed in this study. $\delta^{34}S_{pyrite}$ data were summarized as kernel density plots with mean $\delta^{34}S_{pyrite}$ marked as diamonds. See Section 2.1. for detailed petrographic descriptions. Rim pyrite in JLW15.1, JLW23.5, JLW34.8, and SX88.15 was analyzed. Disseminated subhedral pyrite in JLW23.5 and JLW76 was analyzed. Black dots and dark gray shades in SEM images are, respectively, LA-ICP-MS and SIMS analytical spots. All scale bars = 20 μm. See Appendix A. Supplementary Material for detailed SEM images and elemental maps (Figs. S1–S9).

Pyrite framboids and nodules with low $\delta^{34}S$ values (mean, framboids: +16.4 ‰; nodules: +26.5 ‰) are common in this specimen (Fig. 1 and Fig. S4). Unlike pyrite in Hy31 or Hy55a, these framboids and nodules were thought to have formed through syngenetic and early diagenetic processes (Cui et al., 2018).

The remaining five samples, JLW15.1, JLW23.5, JLW34.8, JLW76, and SX88.15, are outcrop samples of argillaceous dolostones and calcareous mudstones from members II and III of the Ediacaran Doushantuo Formation exposed at the Jiulongwan (JLW) and Sixi (SX) sections in the Yangtze Gorges area of South China. These samples were selected for comparison with the superheavy pyrite described above. They had previously been described in Xiao et al. (2010). Pyrite in these samples occurs as pyritic rims around chert nodules or as disseminated pyrite in the carbonate matrix (Fig. 1 and Fig. S5 – 9). Chert nodules are typically nucleated on a microbial mat fragment, which is surrounded by a silica cortex and then a pyritic rim. Rim pyrite is immersed in a groundmass of silica and occurs as subhedral pyrite grains. A centrifugal decrease in pyrite crystal size is observed in the pyritic nodule rim. The disseminated pyrite is subhedral to anhedral in shape and randomly distributed in the matrix. Compared with the superheavy pyrite in the Cryogenian Tiesi'ao and Datangpo Formations[7], both rim pyrite (mean, JLW15.1: +24.6 ‰; JLW23.5: +18.1 ‰; JLW34.8: +26.6 ‰; SX88.15: +25.9 ‰) and disseminated pyrite (mean, JLW23.5: +30.6 ‰; JLW76: +29.1 ‰) in the Doushantuo samples have much lower $\delta^{34}S$ values. Based on petrographic and isotopic evidence, Xiao et al. (2010) suggested that both the rim pyrite and chert nodules formed during early diagenesis. They argued that bacterial sulfate reduction of an organic-rich nucleus not only provided a H_2S source for pyrite formation but also decreased porewater pH to facilitate silicification (Xiao et al., 2010).

2.2. LA-ICP-MS trace element measurement

LA-ICP-MS pyrite trace element analyses were performed at the LA-ICP-MS Laboratory of the Department of Geology and Geological Engineering, Colorado School of Mines of the USA, using a RESOlution SE 193 nm ArF excimer laser ablation system equipped with an S155 sample chamber, and an Agilent 8900 ICP-MS. Laser spot analysis was carried out with an on-sample fluence of 2.7 J/cm^2, a repetition rate of 5 pulses/sec (5 Hz), and a laser spot size of 14 to 30 μm chosen according to pyrite grain sizes. The ablated material was carried by He gas, then mixed with Ar gas in a funnel sitting right above the sample cell, and subsequently introduced to the ICP-MS. The analyzed isotopes included ^{57}Fe, ^{59}Co, ^{60}Ni, ^{65}Cu, ^{66}Zn, ^{75}As, ^{95}Mo, ^{107}Ag, ^{121}Sb, ^{125}Te, ^{197}Au, ^{205}Tl, and ^{208}Pb. A dwell time of 10 ms was used for all the above isotopes. In addition, ^{29}Si, ^{31}P, ^{43}Ca, ^{49}Ti were also analyzed with a dwell time of 5 ms to monitor the potential intersection of mineral inclusions. Each laser spot was pre-ablated with one laser shot to eliminate possible surface contamination. That was followed by 15 s of washing time for the aerosol to pass through the ICP-MS. A gas blank was collected for 20 s, followed by 30 s of sample signal collection. The primary external standard was STDGL3 glass made at CODES, University of Tasmania (Belousov et al., 2023). The internal standard was Fe assuming a stoichiometric concentration of 46.55 wt%. Data reduction was performed using the software Iolite v. 4.0 (Paton et al., 2011). When selecting an interval on the time-resolved signal (counts per second) plot of each laser spot for data reduction, mineral inclusions were avoided to minimize contaminations. For the same purpose, the data were further screened to ensure that no analyses had higher than 1 % Zn, 2 % As, 1 % Cu, 1 % Ni, and 2 % Co (Gregory et al., 2019). When an element was below the detection limit, half of the detection limit was taken as the concentration. The standard error for each analysis was reported in Supplementary Data S2.

2.3. SEM and EDS observation

SEM and EDS observations were performed at the Virginia Tech Institute of Critical Technology and Applied Science, on a JEOL IT-500HR SEM via a backscattered electron detector at an accelerating voltage of 15 or 20 keV and a working distance of ~ 10.0 mm. Elemental maps were acquired using an AZtecLive Automated Microanalysis System with an UltimMax100 silicon drift detector at an accelerating voltage of 20 keV.

2.4. Random Forests and XGBoost

The LA-ICP-MS pyrite trace element database compiled by Gregory et al. (2019) was used in this study. To train and evaluate models, the labeled trace element dataset was divided into a training dataset, an in-sample-locality test dataset, and an out-of-sample-locality test dataset. A total of 240 pyrite analyses from each of the three labels (sedimentary, synsedimentary hydrothermal, and post-sedimentary hydrothermal) were randomly selected to form a balanced training dataset. For sedimentary pyrite, 16 different geological units each contributed 15 analyses. For synsedimentary hydrothermal pyrite, 120 analyses were from SEDEX deposits and the remaining 120 analyses were from VMS deposits. For post-sedimentary hydrothermal pyrite, hydrothermal breccia, IOCG, orogenic Au, and porphyry Cu deposits each contributed 60 analyses. The rest of analyses within these geologic units/deposits were grouped together to form an in-sample-locality test dataset. Meanwhile, an out-of-sample-locality test dataset was constructed using pyrite analyses from geologic units/deposits that were not part of the training dataset.

Random Forests and XGBoost, two powerful decision-tree ensemble methods (see Appendix A for more details), were performed in Python using the sklearn (Pedregosa et al., 2011) and XGBoost (Chen and Guestrin, 2016) packages, respectively. All data were standardized by subtracting the mean and scaling to unit variance prior to inputting into the models. A five-fold cross-validation method (GridSearchCV in sklearn), which randomly and evenly splits the training dataset into five parts for training and validation, was used to determine optimal hyperparameters. For Random Forests, the feature importance is decided by the mean decrease in impurity when a feature is used to build a split. For XGBoost, the feature importance is decided by the gain, which is the improvement in accuracy brought by a feature. Another widely-used supervised machine learning method, support vector machine (SVM), was also used to analyze the data, although its performance was not as good as Random Forests and XGBoost.

2.5. Hierarchical cluster analysis (HCA) and principal component analysis (PCA)

Two unsupervised machine learning algorithms, HCA and PCA, were used to investigate trace element patterns of sedimentary and hydrothermal pyrite in the database, similar to analyses of Emmings et al. (2022) for origins of sedimentary pyrite through geologic time and Hu et al. (2022a) for origins of high titanium magnetite. HCA is a method of cluster analysis that aims to group similar observations together into several clusters. In order to identify elements sharing similar behaviors, data of each trace element were treated as one observation. To be specific, HCA in this study focused on partitioning 12 observations (elements) into different clusters. PCA is a dimensionality reduction method that aims to increase data interpretability while preserving the maximum amount of information. In PCA, each pyrite analysis was treated as one observation with 12 features (trace elements). The high-dimensional data were projected onto several principal components capturing most of the variance of trace element concentrations and likely recording trace element patterns, which can be used to validate the HCA results. Since LA-ICP-MS trace element data vary over several orders of magnitude and show high skewness, all data were power

transformed using the Yeo-Johnson method (Yeo and Johnson, 2000) before HCA and PCA. Analyses were performed in Python using the sklearn (Pedregosa et al., 2011) package.

3. Results

3.1. New LA-ICP-MS data for Hy, JLW, and SX samples

A total of 177 in-situ pyrite trace element analyses of 9 samples (Hy, JLW, and SX specimens) passed the screening (see Section 2.2.). For each analyzed element, concentrations vary up to 4–5 orders of magnitude. For samples Hy31, JLW15.1, JLW23.5, JLW76, and SX88.15, most analyses of Zn and Te were below detection limits. For JLW34.8, only Te analyses were below detection limits. For Hy31, in addition to Zn and Te, the majority of Mo, Ag, Sb, Au, and Tl analyses were also close to or below detection limits. Analyses below detection limits were replaced by half the detection limit as machine learning inputs (Gregory et al., 2019).

3.2. The LA-ICP-MS database for machine learning analysis

A total of 3632 analyses from 71 different deposits and sedimentary units were used for supervised machine learning model training and testing (Table S1) (Gregory et al., 2019). Of these, 1123 analyses of sedimentary pyrite are from 28 sedimentary units covering the geological interval of 2572–125 Ma and therefore recording the variability of trace element concentrations over geologic time, 1345 analyses of synsedimentary hydrothermal pyrite are from 18 ore deposits, and 1164 analyses of post-sedimentary hydrothermal pyrite are from 25 ore deposits. A total of 720 analyses (240 analyses from each label), from 16 sedimentary units ranging from 2572 to 125 Ma and 33 ore deposits covering 6 different ore types, were randomly selected as the training dataset. The remaining 2487 analyses from these deposits/units were used for initial in-sample-locality testing. Furthermore, 425 analyses from 12 sedimentary units ranging from 2170 to 180 Ma in age and 10 ore deposits covering 3 different ore types, which were not seen in the training and initial testing datasets, were used for out-of-sample-locality testing. By creating such diverse training and testing datasets, we believe that a representative variation in pyrite forming environments has been taken into account. Again, for each analyzed element, concentrations vary up to 5–6 orders of magnitude. When analyses were below detection limits, either half the detection limit or values from the literature were used as an estimation. Considering a significant number of Te and Au analyses are affected by this estimation, we also examined machine learning models on a refined dataset that excluded the Te and Au data.

3.3. Machine learning model performance and predictions on Hy, JLW, and SX samples

Random Forests and XGBoost models both yielded robust performance on in-sample-locality and out-of-sample-locality test datasets (Table 1, Fig. 2A – D), which is significantly better compared with the SVM model (Fig. S10). Average precision, recall, and F1-score are all > 90 %. For sedimentary pyrite, the precision even reaches 96 % (Random Forests) and 98 % (XGBoost) for the out-of-sample-locality test dataset. Mo is the most important feature in both models (Fig. 2E – F). Meanwhile, Ni, Cu, Zn, Ag, Sb, Te, Au, Tl, and Pb all have significant contributions. Additionally, models trained by the dataset without Au and Te, whose concentrations are below the detection limit in many of the samples in the dataset, show lower but still promising precision, recall, and F1-score (Table S2, Fig. S11).

Random Forests and XGboost models, trained with and without Au and Te, show consistent predictions on pyrite analyses of Hy, JLW, and SX samples (Fig. 3 and Fig. S12). The predictions are provided in the ternary plots as the probability of being identified as one of the three types of pyrite (Fig. 3). Pyrite grains in Hy1 (9 grains analyzed), Hy55 (3 nodules analyzed), and Hy59 (2 nodules and 5 aggregates of framboids analyzed) were identified as showing sedimentary signals (mostly > 80 % probability for both models) whereas pyrite grains in Hy31 were classified as post-sedimentary hydrothermal pyrite (mostly > 80 % probability for both models). For JLW and SX samples, the majority of the pyritic rims in JLW15.1, JLW23.5, and SX88.15 and disseminated pyrite in JLW23.5 (4 grains analyzed) and JLW76 (20 grains analyzed) were identified as post-sedimentary hydrothermal pyrite with a few exceptions (pyrites identified as sedimentary or synsedimentary hydrothermal pyrite). A significant portion of these analyses (53 % for the Random Forest model, 13 % for the XGBoost model) falls into the gray triangle in ternary plots because they were given a < 50 % probability of being identified as a leading label. Conversely, all analyses on JLW34.8 show synsedimentary hydrothermal origins (>50 % probability for the Random Forest model, >80 % probability for the XGBoost model).

Table 1
Supervised machine learning model evaluation. Simple arithmetic averages of precision, recall, and F1 score across labels (sedimentary, synsedimentary hydrothermal, post-sedimentary hydrothermal) on in-sample-locality test and out-of-sample-locality test datasets. For each label, the precision is calculated by dividing the true positives by the sum of true positives and false positives, recall by dividing the true positives by the sum of true positives and false negatives, and F1 score is the harmonic mean of precision and recall.

Model	Testing Dataset	Average Precision	Average Recall	Average F1 score
Random Forests	In-sample-locality test	0.91	0.91	0.91
	Out-of-sample-locality test	0.92	0.92	0.92
XGBoost	In-sample-locality test	0.93	0.93	0.93
	Out-of-sample-locality test	0.92	0.93	0.92

4. Discussion

4.1. Trace element patterns of sedimentary and hydrothermal pyrite

Despite the considerable variability in trace element signals within each pyrite type (e.g., variation in trace element compositions in sedimentary pyrite due to changes in ocean chemistry; Emmings et al., 2022), both models exhibited robust predictive capabilities for determining the provenance of pyrite. The consistent precision, recall, and F1-score values obtained from both the in-sample-locality and out-of-sample-locality tests indicate a high level of generalizability in the models. Based on feature importance, ten out of twelve elements show significant contributions to both models, suggesting a complex relationship between trace element concentrations and the origin of pyrite. In other words, there is no single element that can be used to differentiate sedimentary and hydrothermal pyrite, which further supports the necessity of using machine learning modeling of a suite of elements to identify the origins of pyrite.

Decision Tree ensembles (i.e., Random Forest and XGBoost), however, cannot offer further insights into trace element behavior. To investigate any existing trace element patterns of sedimentary and hydrothermal pyrite, we applied hierarchical cluster analysis (HCA) on the power transformed (Yeo and Johnson, 2000) trace element database. HCA reveals four major clusters of elements (Fig. 4A and Fig. S13). Firstly, Te and Au are enriched in post-sedimentary and VMS-type synsedimentary hydrothermal pyrite. The enrichment is possibly related to changes in the chemical and physical properties of fluids, zone refinement (i.e., enrichment through dissolution and reprecipitation), and direct input from magmatic and metamorphic hydrothermal fluids (Fuchs et al., 2019; Large et al., 2009). A coupled Au-Te transfer in fluids

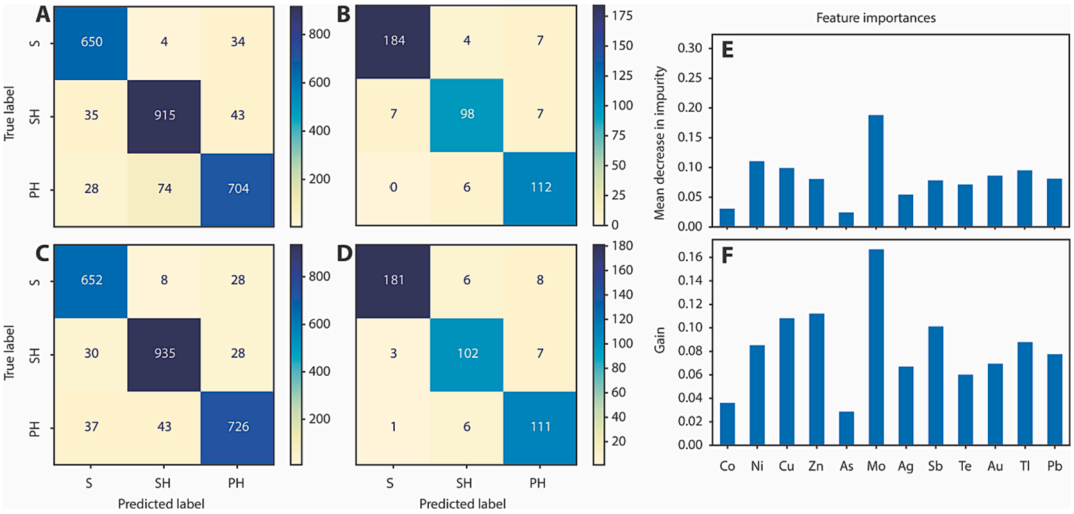

Fig. 2. Supervised machine learning model confusion matrix and feature importance. **A** and **B** are confusion matrixes of the Random Forests model on in-sample-locality test and out-of-sample-locality test datasets, respectively. **C** and **D** are confusion matrixes of the XGBoost model on in-sample-locality test and out-of-sample-locality test datasets, respectively. Confusion matrixes visualize model performance on identifying sedimentary (S), synsedimentary hydrothermal (SH), and post-sedimentary hydrothermal (PH) pyrite. **E** and **F** show the feature importance of the Random Forests and XGBoost models, respectively. For Random Forests, the feature importance is decided by the mean decrease in impurity when a feature is used to build a split. For XGBoost, the feature importance is decided by the gain, which is the improvement in accuracy brought by a feature.

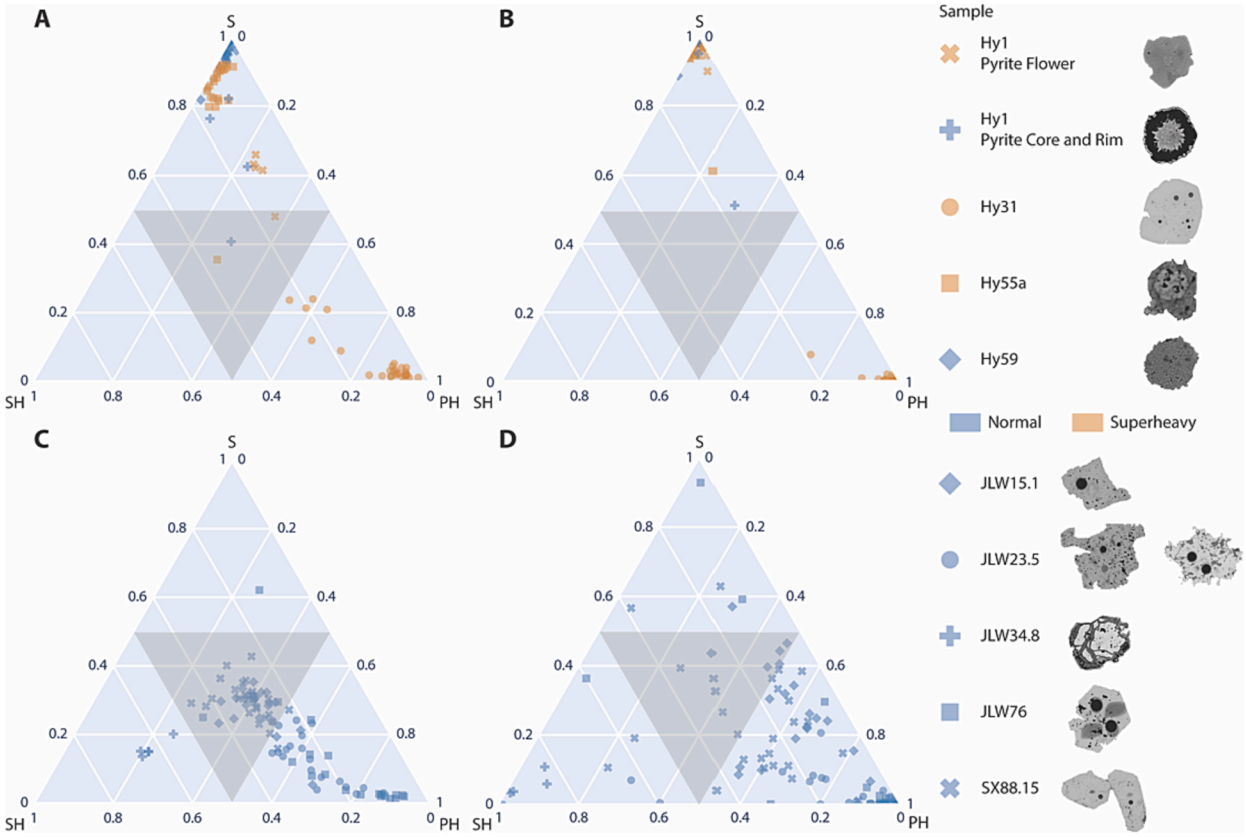

Fig. 3. Ternary plots of model predictions on Hy, JLW, and SX samples. Model predictions on pyrite analyses of Hy, JLW, and SX samples are expressed as the probability of being identified as sedimentary (S), synsedimentary hydrothermal (SH), or post-sedimentary hydrothermal (PH) pyrite. Each data point represents one single LA-ICP-MS analysis. A and B are model outputs of Hy samples based on the Random Forests (A) and XGBoost (B) models. C and D are model outputs of JLW and SX samples based on the Random Forests (C) and XGBoost (D) models. Blue data points represent isotopically normal pyrite and brown data points represent isotopically superheavy pyrite. Superheavy pyrite is defined by $\delta^{34}S > 50$ ‰, which is substantially higher than that of contemporaneous seawater (Fike et al., 2015). Gray triangular shades in ternary plots represent the ambiguous zone. Pyrite analyses in the ambiguous zone have a less than 50 % probability of being identified to any single label.

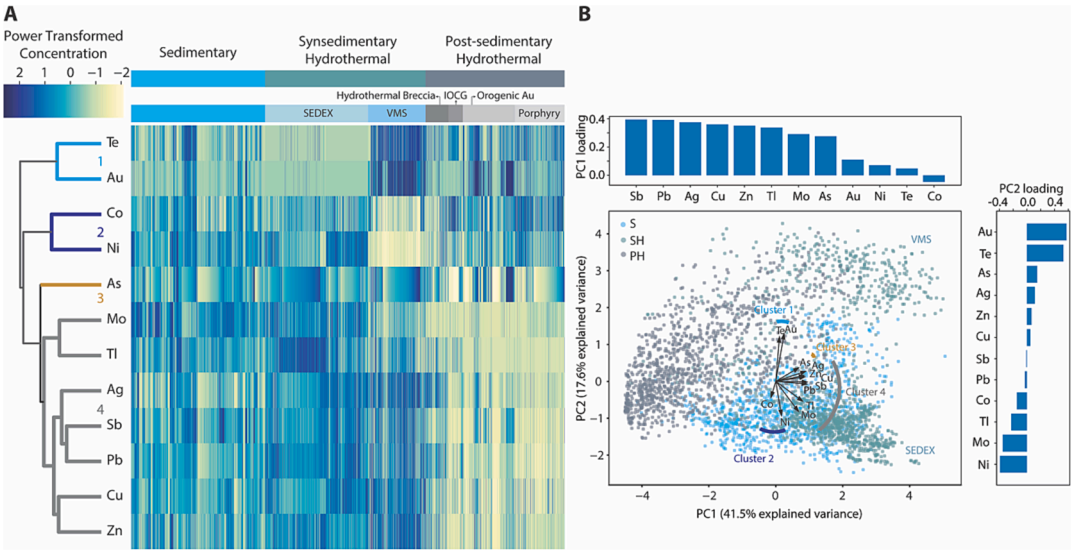

Fig. 4. HCA and PCA for power-transformed pyrite trace element database. A. HCA by elements (n = 12). Four elemental clusters were identified and marked in different colors on the dendrogram. Columns (n = 3632) represent individual pyrite analyses, which were grouped by their origins. Each grid represents the power-transformed concentration of a specific trace element from an individual pyrite analysis. **B.** PCA biplot of the first two principal components (PC1 and PC2). Loadings for each element in PC1 and PC2 are shown as arrows in the biplot and histograms along the × and y axes. The length and direction of the arrows were decided by the loadings in PC1 and PC2. The arrows were also grouped by elemental clusters from HCA. Analyses of synsedimentary hydrothermal pyrite are divided into two clusters in the biplot: the upper cluster is VMS-type pyrite and the lower cluster is SEDEX-type pyrite.

and enhancement of Au partitioning with increasing Tellurium concentrations (Belousov et al., 2016) may also play a role in the enrichment. We note, however, that average Au and Te concentrations from the literature (Gregory et al., 2019) were used for most SEDEX pyrite. These data were assumed to be reasonable, as Au and Te are usually below the detection limit in SEDEX deposits (Gregory et al., 2019). Secondly, Co and Ni are relatively depleted in the majority of VMS-type synsedimentary hydrothermal pyrite. This is probably due to coprecipitating minerals outcompeting pyrite for metal incorporation in their respective structure (Genna and Gaboury, 2015). For example, Co and Ni, with similar ionic radii and charges to Zn, can substitute into sphalerite forming from the hydrothermal fluid (Genna and Gaboury, 2015). Other researchers have also proposed that the depletion in Co and Ni can be attributed to the temperature of pyrite formation. Pyrite formed at high temperatures tends to concentrate Co and Ni. As these local fluids undergo cooling and reach the low-temperature seafloor, both the fluids themselves and the precipitated pyrite are depleted in these elements (Keith et al., 2016). Meanwhile, the solubility of Co and Ni can be affected by the formation of Cl^- complexes in hydrothermal fluids (Keith et al., 2016). Thirdly, As distribution is relatively ubiquitous. No significant difference in As concentration is observed among the three types of pyrite. Lastly, Mo, Tl, Ag, Sb, Pb, Cu, and Zn are depleted in post-sedimentary hydrothermal pyrite. High concentrations of these elements in pyrite forming near the seafloor may be related to the decomposition of particulate and dissolved organic matter and the reduction of Mn-Fe (oxyhydr)oxide (Emmings et al., 2022). On the other hand, for post-sedimentary hydrothermal pyrite, higher temperature and slower pyrite growth rate enable trace elements to be partitioned into other distinct sulfide phases (e.g., sphalerite and galena) rather than existing as solid solutions or micro-inclusions in pyrite (Large et al., 2009). Among the 12 elements investigated, Mo is the only element showing a unidirectional change from sedimentary to post-sedimentary hydrothermal pyrite — Mo concentration decreases from left to right in Fig. 4A. This pattern can be attributed to the higher concentrations of Mo in seawater compared to most hydrothermal fluids (Metz and Trefry, 2000; Trefry et al., 1994). As we transition from sedimentary to post-sedimentary hydrothermal pyrite, the contribution of seawater to pyrite formation progressively diminishes, accompanied with an escalating prominence of hydrothermal fluids in the mineralization process. This unidirectional change likely accounts for the selection of Mo as the most important feature by both machine learning models.

HCA results are further supported by principal component analysis (PCA) (Fig. 4B). The first two principal components (PC1 and PC2) capture 59 % of the total variance (Fig. S14) and accurately record the concentration changes of the four clusters of elements. PC1 records changes in the third and fourth clusters of elements (As, Mo, Tl, Ag, Sb, Pb, Cu, Zn), whereas PC2 records changes in the first and second clusters (Au, Te, Co, Ni) plus Mo and Tl from the fourth cluster. Along the PC1 direction, post-sedimentary hydrothermal pyrite is separated from the sedimentary and synsedimentary hydrothermal pyrite, which is consistent with the insight from HCA that the fourth cluster of elements is enriched in pyrite forming near the seafloor. Additionally, the separation of post-sedimentary hydrothermal and VMS-type synsedimentary hydrothermal pyrite from sedimentary and SEDEX-type synsedimentary hydrothermal pyrite along the PC2 direction confirms the observation that the first cluster of elements is enriched in post-sedimentary hydrothermal and VMS-type synsedimentary hydrothermal pyrite. The differentiation between SEDEX-type pyrite and VMS-type pyrite along the PC2 is further enhanced due to the fact that SEDEX-type pyrite is enriched in Co and Ni while VMS-type pyrite is depleted in these two elements. Meanwhile, due to the observed overlap of pyrite analyses from different origins within the PC1 and PC2 biplot, the viability of employing the first two principal components for discriminating different pyrite is compromised, which further reinforces the complex nature of this target.

4.2. *Origins of Neoproterozoic pyrite*

Our machine-learning models partially confirmed the findings of Cui et al. (2018). Specifically, pyrite grains in Hy31 were identified as post-sedimentary hydrothermal pyrite, and grains in Hy59 were identified as sedimentary pyrite. A post-sedimentary hydrothermal origin of pyrite in Hy31 is consistent with low trace element concentrations of Mo, Ag, Tl, Sb, and Zn. However, both models posit a sedimentary origin for pyrite grains in Hy1 and Hy55a, whereas Cui et al. (2018) argued for a hydrothermal origin. One of their major arguments was based on the

paragenesis of Hy55a, where pyrite replaced and thus postdate rhodochrosite and illite. Key to this argument is the time when rhodochrosite and illite formed. Rhodochrosite can form during early diagenesis as porewaters become anoxic and manganese (oxyhydr)oxides are reduced (Johnson et al., 2016). For example, calcian rhodochrosite has been reported from organic-rich marine muds in the Baltic Sea (Neumann et al., 2002). Recent studies even suggested that manganese-rich carbonate can form in the water column of redox-stratified lakes (Herndon et al., 2018). Furthermore, the preservation of rhodochrosite laminae that warp around pyrite nodules in Hy55a (Fig. S3A – F) suggests a sedimentary or early diagenetic origin. As for the formation of illite, laboratory experiments have demonstrated microbially-induced illitization at room temperature and atmospheric pressure within two weeks (Dong et al., 2022). Field observation also recorded microbially induced smectite-to-illite transformation during diagenesis (Aubineau et al., 2019; Kim et al., 2019). Thus, our model predictions are not necessarily inconsistent with petrographic observations, and the formation of superheavy pyrite could be linked to the contemporaneous oceanic environment. If so, regional and global models of the oceanic sulfur cycle should be considered to seek an explanation for unusual sulfur isotope signals from the Neoproterozoic superheavy pyrite.

For the JLW and SX samples from the Ediacaran Doushantuo Formation, our models reveal a more complex diagenetic history than originally thought (Xiao et al., 2010). Instead of a sedimentary origin, rim pyrite and disseminated pyrite in JLW15.1, JLW23.5, JLW76, and SX88.15 were classified as post-sedimentary hydrothermal pyrite. We note, however, that a large portion of analyses shows ambiguous prediction results (i.e., they are in the gray triangles in ternary diagrams in Fig. 3C–D). They were given a < 50 % probability of being identified to the leading label (Fig. 3 and Fig. S12), which in this case is the post-sedimentary hydrothermal pyrite. We propose two explanations for the low probability. First, since our hydrothermal training data come from a limited number of ore deposits, it is impossible for us to cover every pyrite-forming hydrothermal environment. Accordingly, these subhedral pyrite grains may have formed in a hydrothermal environment not represented in the training dataset. Second, as some Doushantuo pyrite analyses have been identified to sedimentary or synsedimentary hydrothermal origins with some confidence, it is possible that some of these ambiguous pyrite grains were originally sedimentary pyrite but later altered by hydrothermal fluids and therefore show mixed signals. In this scenario, JLW34.8 may have been altered by synsedimentary hydrothermal fluids and the four other Doushantuo samples may have been further altered by post-sedimentary processes. The possibility of hydrothermal alteration is also consistent with petrographic observations: no pyrite framboids have been observed in any of the five Doushantuo samples analyzed in this study. Considering that not all analyses identified as post-sedimentary hydrothermal pyrite (i.e., the label with the highest possibility predicted by the model) fall into the ambiguous zone (probability < 50 %) and more than half (89 % for XGBoost, 53 % for Random Forests) of them were given a high probability (>50 %), we argue the hydrothermal alteration hypothesis may be more plausible. Given the petrographic evidence for an early diagenetic origin of the pyritic rim around Doushantuo chert nodules, the silicification model proposed by Xiao et al. (2010) may still be valid, although the trace element data do show that the pyrite may have been hydrothermally altered at a later time.

5. Conclusions

To sum up, this is the first study applying supervised machine learning to understand the origins of Neoproterozoic pyrite. The two models tested here, Random Forests and XGBoost, show strong performance in distinguishing sedimentary and hydrothermal pyrite using in-situ trace element data. Our analysis shows that superheavy pyrite in the Cryogenian Datangpo Formation can be sedimentary or hydrothermal in origin. Future studies are needed to investigate what the ^{34}S-enriched pyrite can tell us about the sulfur biogeochemical cycle. Our analysis of pyritic rims surrounding fossiliferous chert nodules from the Ediacaran Doushantuo Formation revealed a more complex history than previously thought. Early diagenetic pyrite precipitated around fossiliferous chert nodules and may have played a key role in chert nodule formation and fossil preservation (Xiao et al., 2010), but was subsequently altered and overprinted by hydrothermal fluids. This study highlights machine learning models as a powerful tool to study the origins of pyrite and to effectively screen pyrite samples for paleoenvironmental research.

6. Data availability

Research data has been supplied via a repository: https://doi.org/10.17632/w36v9gx9d6.1. All machine learning codes used in this paper, including the optimised hyperparameters, are avilable on the GitHub repository: https://github.com/junyaok/ML_Pyrite.

Declaration of competing interest

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.

Acknowledgments

This work was supported by the National Science Foundation (EAR-2021207 to S.X.) and a Natural Sciences and Engineering Research Council of Canada Discovery grant to D.G.

Appendix A. Supplementary material

The Supplementary Material contains: a brief introduction to decision tree-based supervised machine learning; geological settings of Hy, JLW, SX samples; Figs. S1–S9, SEM images and elemental maps of Hy, JLW, SX samples; Fig. S10, SVM model performance with Au and Te included in the analysis; Figs. S11–S12, model evaluation and model predictions with Au and Te excluded from the analysis; Fig. S13, correlation matrix of 12 trace elements from the LA-ICP-MS database; Fig. S14, PCA scree plot; Table S1, summary of LA-ICP-MS pyrite trace element database; and Table S2, model evaluation with Au and Te excluded from the analysis. Supplementary material to this article can be found online at https://doi.org/10.1016/j.gca.2023.11.005.

References

Anbar, A.D., Knoll, A.H., 2002. Proterozoic ocean chemistry and evolution: A bioinorganic bridge? Science 297, 1137–1142.

Aubineau, J., El Albani, A., Bekker, A., Somogyi, A., Bankole, O.M., Macchiarelli, R., Meunier, A., Riboulleau, A., Reynaud, J.-Y., Konhauser, K.O., 2019. Microbially induced potassium enrichment in Paleoproterozoic shales and implications for reverse weathering on early Earth. Nat. Commun. 10, 2670.

Belousov, I., Large, R., Meffre, S., Danyushevsky, L., Steadman, J., Beardsmore, T., 2016. Pyrite compositions from VHMS and orogenic Au deposits in the Yilgarn Craton, Western Australia: Implications for gold and copper exploration. Ore Geol. Rev. 79, 474–499.

Belousov, I., Danyushevsky, L., Goemann, K., Gilbert, S., Olin, P., Thompson, J., Lounejeva, E., Garbe-Schönberg, D., 2023. STDGL3, a reference material for analysis of sulfide minerals by laser ablation ICP-MS: An assessment of matrix effects and the impact of laser wavelengths and pulse widths. Geostand. Geoanal. Res. 47, 493–508.

Berner, R.A., 1970. Sedimentary pyrite formation. Am. J. Sci. 268, 1–23.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Cadeau, P., Cartigny, P., Thomazo, C., Jézéquel, D., Leboulanger, C., Sarazin, G., Ader, M., 2022. The Dziani Dzaha Lake: A long-awaited modern analogue for superheavy pyrites. Geobiology 20, 444–461.

Cai, C., Lyons, T.W., Sun, P., Liu, D., Wang, D., Tino, C.J., Luo, G., Peng, Y., Jiang, L., 2022. Enigmatic super-heavy pyrite formation: Novel mechanistic insights from the aftermath of the Sturtian Snowball Earth. Geochim. Cosmochim. Acta 334, 65–82.

Chen T. and Guestrin C. (2016) XGBoost: A scalable tree boosting system, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794.

Chen, G., Cheng, Q., Lyons, T.W., Shen, J., Agterberg, F., Huang, N., Zhao, M., 2022. Reconstructing Earth's atmospheric oxygenation history using machine learning. Nat. Commun. 13, 5862.

Cui, H., Kitajima, K., Spicuzza, M.J., Fournelle, J.H., Denny, A., Ishida, A., Zhang, F., Valley, J.W., 2018. Questioning the biogenicity of Neoproterozoic superheavy pyrite by SIMS. Am. Mineral.: J. Earth Planet. Mater. 103, 1362–1400.

Dong, H., Huang, L., Zhao, L., Zeng, Q., Liu, X., Sheng, Y., Shi, L., Wu, G., Jiang, H., Li, F., 2022. A critical review of mineral–microbe interaction and co-evolution: mechanisms and applications. Natl. Sci. Rev. 9, nwac128.

Emmings, J.F., Poulton, S.W., Walsh, J., Leeming, K.A., Ross, I., Peters, S.E., 2022. Pyrite mega-analysis reveals modes of anoxia through geological time. Sci. Adv. 8, eabj5687.

Emsbo P., Seal R. R., Breit G. N., Diehl S. F. and Shah A. K. (2016) Sedimentary exhalative (sedex) zinc-lead-silver deposit model. US Geological Survey.

Ferrini, V., Fayek, M., Vito, C.D., Mignardi, S., Pignatti, J., 2010. Extreme sulphur isotope fractionation in the deep Cretaceous biosphere. J. Geol. Soc. London 167, 1009–1018.

Fike, D.A., Bradley, A.S., Rose, C.V., 2015. Rethinking the ancient sulfur cycle. Annu. Rev. Earth Planet. Sci. 43, 593–622.

Fuchs, S., Hannington, M.D., Petersen, S., 2019. Divining gold in seafloor polymetallic massive sulfide systems. Miner. Deposita 54, 789–820.

Genna, D., Gaboury, D., 2015. Deciphering the hydrothermal evolution of a VMS system by LA-ICP-MS using trace elements in pyrite: an example from the Bracemac-McLeod deposits, Abitibi, Canada, and implications for exploration. Econ. Geol. 110, 2087–2108.

Gregory, D.D., Large, R.R., Halpin, J.A., Baturina, E.L., Lyons, T.W., Wu, S., Danyushevsky, L., Sack, P.J., Chappaz, A., Maslennikov, V.V., 2015. Trace element content of sedimentary pyrite in black shales. Econ. Geol. 110, 1389–1410.

Gregory, D.D., Cracknell, M.J., Large, R.R., McGoldrick, P., Kuhn, S., Maslennikov, V.V., Baker, M.J., Fox, N., Belousov, I., Figueroa, M.C., 2019. Distinguishing ore deposit type and barren sedimentary pyrite using laser ablation-inductively coupled plasma-mass spectrometry trace element data and statistical analysis of large data sets. Econ. Geol. 114, 771–786.

Gregory D. D. (2020) The Pyrite Trace Element Paleo-Ocean Chemistry Proxy. *Elements in Geochemical Tracers in Earth System Science*.

Habicht, K.S., Canfield, D.E., 1997. Sulfur isotope fractionation during bacterial sulfate reduction in organic-rich sediments. Geochim. Cosmochim. Acta 61, 5351–5361.

Hannington M. D. (2014) 13.18 - Volcanogenic Massive Sulfide Deposits. In *Treatise on Geochemistry (Second Edition)*, 2nd edn (eds. H. D. Holland and K. K. Turekian). Elsevier. pp. 463-488.

Herndon, E.M., Havig, J.R., Singer, D.M., McCormick, M.L., Kump, L.R., 2018. Manganese and iron geochemistry in sediments underlying the redox-stratified Fayetteville Green Lake. Geochim. Cosmochim. Acta 231, 50–63.

Hu, S.-Y., Wang, X.-C., Tian, L., Martin, L., Schoneveld, L., Barnes, S.J., Guagliardo, P., Ding, W., Rickard, W.D., 2022b. Variability of sulfur isotopes and trace metals in pyrites from the upper oceanic crust of the South China Sea basin, implications for sulfur and trace metal cycling in subsurface. Chem. Geol. 606, 120982.

Hu, B., Zeng, L.-P., Liao, W., Wen, G., Hu, H., Li, M.Y.H., Zhao, X.-F., 2022a. The origin and discrimination of High-Ti magnetite in magmatic-hydrothermal systems: Insight from machine learning analysis. Econ. Geol. 117, 1613–1627.

Johnson, J.E., Webb, S.M., Ma, C., Fischer, W.W., 2016. Manganese mineralogy and diagenesis in the sedimentary rock record. Geochim. Cosmochim. Acta 173, 210–231.

Keith, M., Haase, K.M., Klemd, R., Krumm, S., Strauss, H., 2016. Systematic variations of trace element and sulfur isotope compositions in pyrite with stratigraphic depth in the Skouriotissa volcanic-hosted massive sulfide deposit, Troodos ophiolite, Cyprus. Chem. Geol. 423, 7–18.

Kim, J., Dong, H., Yang, K., Park, H., Elliott, W.C., Spivack, A., Koo, T.-H., Kim, G., Morono, Y., Henkel, S., 2019. Naturally occurring, microbially induced smectite-to-illite reaction. Geology 47, 535–539.

Lang, X., Zhao, Z., Ma, H., Huang, K., Li, S., Zhou, C., Xiao, S., Peng, Y., Liu, Y., Tang, W., Shen, B., 2021. Cracking the superheavy pyrite enigma: possible roles of volatile organosulfur compound emission. Natl. Sci. Rev. 8.

Large, R.R., Danyushevsky, L., Hollit, C., Maslennikov, V., Meffre, S., Gilbert, S., Bull, S., Scott, R., Emsbo, P., Thomas, H., 2009. Gold and trace element zonation in pyrite using a laser imaging technique: Implications for the timing of gold in orogenic and Carlin-style sediment-hosted deposits. Econ. Geol. 104, 635–668.

Li, X.-M., Zhang, Y.-X., Li, Z.-K., Zhao, X.-F., Zuo, R.-G., Xiao, F., Zheng, Y., 2023. Discrimination of Pb-Zn deposit types using sphalerite geochemistry: New insights from machine learning algorithm. Geosci. Front. 14, 101580.

Liu T.-B., Maynard J. B. and Alten J. (2006) Superheavy S isotopes from glacier-associated sediments of the Neoproterozoic of south China: Oceanic anoxia or sulfate limitation? In *Evolution of Early Earth's Atmosphere, Hydrosphere, and Biosphere—Constraints from Ore Deposits: Geological Society of America Memoir 198* (eds. S. E. Kesler and H. Ohmoto). Geological Society of America, Boulder, Colorado. pp. 205-222.

Lyons, T.W., Werne, J.P., Hollander, D.J., Murray, R., 2003. Contrasting sulfur geochemistry and Fe/Al and Mo/Al ratios across the last oxic-to-anoxic transition in the Cariaco Basin, Venezuela. Chem. Geol. 195, 131–157.

Metz, S., Trefry, J.H., 2000. Chemical and mineralogical influences on concentrations of trace metals in hydrothermal fluids. Geochim. Cosmochim. Acta 64, 2267–2279.

Mukherjee, I., Corkrey, R., Large, R., Danyushevsky, L., 2023. Abiotic and biotic constraints on Earth's ancient colonisers in the Proterozoic. Precambr. Res. 393, 107089.

Neumann, T., Heiser, U., Leosson, M.A., Kersten, M., 2002. Early diagenetic processes during Mn-carbonate formation: evidence from the isotopic composition of authigenic Ca-rhodochrosites of the Baltic Sea. Geochim. Cosmochim. Acta 66, 867–879.

Paton, C., Hellstrom, J., Paul, B., Woodhead, J., Hergt, J., 2011. Iolite: Freeware for the visualisation and processing of mass spectrometric data. J. Anal. At. Spectrom 26, 2508–2518.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peng, Y., Bao, H., Pratt, L.M., Kaufman, A.J., Jiang, G., Boyd, D., Wang, Q., Zhou, C., Yuan, X., Xiao, S., Loyd, S., 2014. Widespread contamination of carbonate-associated sulfate by present-day secondary atmospheric sulfate: Evidence from triple oxygen isotopes. Geology 42, 815–818.

Ries, J.B., Fike, D.A., Pratt, L.M., Lyons, T.W., Grotzinger, J.P., 2009. Superheavy pyrite ($\delta^{34}S_{pyr} > \delta^{34}S_{CAS}$) in the terminal Proterozoic Nama Group, southern Namibia: A consequence of low seawater sulfate at the dawn of animal life. Geology 37, 743–746.

Shen, B., Xiao, S., Kaufman, A.J., Bao, H., Zhou, C., Wang, H., 2008. Stratification and mixing of a post-glacial Neoproterozoic ocean: Evidence from carbon and sulfur isotopes in a cap dolostone from northwest China. Earth Planet. Sci. Lett. 265, 209–228.

Tostevin, R., He, T.C., Turchyn, A.V., Wood, R.A., Penny, A.M., Bowyer, F., Antler, G., Shields, G.A., 2017. Constraints on the late Ediacaran sulfur cycle from carbonate associated sulfate. Precambr. Res. 290, 113–125.

Trefry, J.H., Butterfield, D.B., Metz, S., Massoth, G.J., Trocine, R.P., Feely, R.A., 1994. Trace metals in hydrothermal solutions from Cleft segment on the southern Juan de Fuca Ridge. J. Geophys. Res. Solid Earth 99, 4925–4935.

Wilkin, R.T., Barnes, H.L., Brantley, S.L., 1996. The size distribution of framboidal pyrite in modern sediments: An indicator of redox conditions. Geochim. Cosmochim. Acta 60, 3897–3912.

Xiao, S., Schiffbauer, J., McFadden, K., Hunter, J., 2010. Petrographic and SIMS pyrite sulfur isotope analyses of Ediacaran chert nodules: Implications for microbial processes in pyrite rim formation, silicification, and exceptional fossil preservation. Earth Planet. Sci. Lett. 297, 481–495.

Yeo, I.K., Johnson, R.A., 2000. A new family of power transformations to improve normality or symmetry. Biometrika 87, 954–959.