



引文格式: 牛警徽, 吴明刚, 田福泉, 等. 基于机器学习和黄铁矿微量元素识别金矿床类型[J]. 西北地质, 2026, 59(4): 20–29. DOI: 10.12401/j.nwg.2026004

Citation: NIU Jinghui, WU Minggang, TIAN Fuquan, et al. Classification of Gold Deposit Types Based on Machine Learning and Pyrite Trace Element Geochemistry[J]. Northwestern Geology, 2026, 59(4): 20–29. DOI: 10.12401/j.nwg.2026004

## 基于机器学习和黄铁矿微量元素识别金矿床类型

牛警徽<sup>1</sup>, 吴明刚<sup>2</sup>, 田福泉<sup>3</sup>, 邱敦方<sup>3</sup>, 宋子崇<sup>3</sup>, 张永林<sup>4</sup>, 范德江<sup>1</sup>, 钟世华<sup>1,\*</sup>

(1. 中国海洋大学海洋地球科学学院, 山东 青岛 266100; 2. 江西工业贸易职业技术学院, 江西 南昌 330038; 3. 山东黄金矿业(金星)有限公司, 山东 青岛 266748; 4. 山东黄金矿业(莱西)有限公司, 山东 青岛 266715)

**摘要:** 黄铁矿是金矿床中最常见的载金矿物, 其地球化学成分能有效指示成矿信息。笔者系统收集了全球浅成低温热液型、斑岩型和造山型金矿床中黄铁矿的微量元素数据, 选取 9 种常见的地球化学特征, 采用 K 近邻算法、随机森林和支持向量机 3 种机器学习模型进行模型训练, 旨在建立一种基于黄铁矿成分高效、准确识别金矿床类型的新方法。模型分类性能评估结果显示, 训练后的 K 近邻算法、随机森林和支持向量机模型的分​​类准确率分别为 0.89、0.96 和 0.95。此外, 文中采用未参与训练的东昆仑造山型金矿黄铁矿成分对训练模型进行独立验证, 以评估其在实际应用中的适用性与准确性。结果表明, 3 种模型均能正确识别出这些黄铁矿主要来源于造山型金矿, 其中支持向量机模型效果最佳, 准确率达 0.88。为克服机器学习的“黑匣子”问题, 文中还引入机器学习可解释性方法, 以揭示不同黄铁矿微量元素输入特征对分类结果的贡献。SHAP 分析表明, 在识别斑岩型和造山型金矿时, 对模型输出结果影响最大的 3 个黄铁矿微量元素特征依次为 As、Co 和 Cu, 而在识别浅成低温热液型金矿时, 黄铁矿的 Cu、As 和 Ni 含量对模型输出结果影响最大。研究结果表明, 黄铁矿微量元素可作为区分金矿成因类型的有效标志。未来, 结合机器学习方法, 黄铁矿成分有望进一步拓展至其他类型矿床的识别与分类, 为矿床成因研究与找矿预测提供新的技术途径。

**关键词:** K 近邻值算法; 随机森林; 支持向量机; 黄铁矿; 微量元素

中图分类号: P618.51

文献标志码: A

文章编号: 1009-6248(2026)04-0020-10

### Classification of Gold Deposit Types Based on Machine Learning and Pyrite Trace Element Geochemistry

NIU Jinghui<sup>1</sup>, WU Minggang<sup>2</sup>, TIAN Fuquan<sup>3</sup>, QIU Dunfang<sup>3</sup>, SONG Zichong<sup>3</sup>, ZHANG Yonglin<sup>4</sup>,  
FAN Dejiang<sup>1</sup>, ZHONG Shihua<sup>1,\*</sup>

(1. College of Marine Geosciences, Ocean University of China, Qingdao 266100, Shandong, China; 2. Jiangxi Vocational Technical College of Industry and Trade, Nanchang 330038, Jiangxi, China; 3. Shandong Gold Mining Industry (Jinxing) Co., LTD, Qingdao 266748, Shandong,

收稿日期: 2025-12-01; 修回日期: 2026-04-22; 责任编辑: 吕鹏瑞

基金项目: 山东省自然科学基金青年科学基金(B类)项目(ZR2025QB19), 深地国家科技重大专项(2025ZD1009710), 山东省泰山学者青年专家项目(tsqn202507105), 国家自然科学基金项目(42472104)联合资助。

作者简介: 牛警徽(1996–), 男, 博士研究生, 地质学专业。E-mail: niujinghui@163.com。

\* 通讯作者: 钟世华(1989–), 男, 博士, 副教授, 从事矿床学和地质大数据研究。E-mail: zhongshihua@ouc.edu.cn。

© Editorial Office of *Northwestern Geology*. OA under CC BY-NC-ND

China; 4. Shandong Gold Mining Industry (Laixi) Co., LTD, Qingdao 266715, Shandong, China)

**Abstract:** Pyrite is the most common gold bearing mineral in gold deposits, and its geochemical composition can effectively indicate mineralization processes. This study systematically compiles trace element data of pyrite from epithermal, porphyry, and orogenic gold deposits worldwide. Nine commonly used geochemical features are selected, and three machine learning models including the K nearest neighbors algorithm, random forest, and support vector machine are employed for model training, with the aim of establishing an efficient and accurate method for identifying gold deposit types based on pyrite composition. The evaluation results of model classification performance show that the classification accuracies of the trained K nearest neighbors, random forest, and support vector machine models are 0.89, 0.96, and 0.95, respectively. In addition, pyrite compositional data from the East Kunlun orogenic gold deposits that are not involved in model training are used for independent validation to assess the applicability and accuracy of the models in practical scenarios. The results indicate that all three models correctly identify these pyrites as mainly derived from orogenic gold deposits, among which the support vector machine model achieves the best performance with an accuracy of 0.88. To address the black box issue in machine learning, an interpretable machine learning method is further introduced to quantify the contributions of different pyrite trace element features to the classification results. The SHAP analysis reveals that As, Co, and Cu are the three most influential trace element features for distinguishing porphyry and orogenic gold deposits, whereas Cu, As, and Ni exert the greatest influence on the identification of epithermal gold deposits. The results demonstrate that trace elements in pyrite can serve as effective indicators for discriminating genetic types of gold deposits. In the future, combined with machine learning approaches, pyrite composition has the potential to be further extended to the identification and classification of other types of mineral deposits, providing new technical approaches for studies on ore genesis and mineral exploration prediction.

**Keywords:** K-nearest neighbor; random forest; support vector machine; Pyrite; trace elements

矿物原位微量元素记录了温度、氧逸度、成矿流体等关键信息,是研究矿床成因、揭示矿床类型的重要地球化学标志(Liu et al., 2023; Qiu et al., 2024; 钟世华等, 2025a, 2025b)。近年来,分析仪器的迅猛发展使得利用矿物原位分析技术研究精细成矿过程成为可能(范宏瑞等, 2018)。黄铁矿是各类金矿床中最重要的载金矿物(Li et al., 2021),其原位微量元素特征已被广泛用于揭示成矿过程和判定成因类型(Voute et al., 2019; Cao et al., 2023; Li et al., 2025; Ribeiro et al., 2026)。然而,海量研究数据的迅速积累使得如何高效处理和解读大规模、多维度的分析数据成为当前研究中的新挑战。机器学习(Machine Learning, 简称 ML)作为处理和分析大量多维数据的有效工具,能够提取数据潜在信息并做出预测或决策(He et al., 2022)。目前,机器学习已广泛应用于地球科学多个领域,例如花岗岩类型识别(Zhong et al., 2023)、矿产勘探(Shirmard et al., 2022)、找矿建模(Fu et al., 2021)等复杂问题。已有研究表明,借助机器学习模型,黄铁矿微量元素特征可以有效区分金矿床类型(Li et al., 2024; Wang et al., 2024;

Gul et al., 2025),显示出较大的应用潜力。然而,现有研究仍面临样本来源分布不均、数据预处理方法有限以及模型解释性与地质关联不足等问题。

在此背景下,笔者系统收集了全球范围内浅成低温热液型(125条)、斑岩型(735条)和造山型(909条)矿床中黄铁矿的微量元素数据,并采用K近邻算法(K-Nearest Neighbor, 简称 KNN)、随机森林(Random Forest, 简称 RF)与支持向量机(Support Vector Machine, 简称 SVM)等3种常用机器学习算法进行模型训练。针对传统二维判别图解因受限于元素维度而导致分类精度不足的问题,文中不仅构建了高维分类模型,还引入了机器学习可解释性算法——SHapley Additive exPlanations(简称 SHAP 值),定量评估了黄铁矿微量元素在矿床分类中的重要性排序及其地质指示意义。在此基础上,进一步通过独立验证集对所构建的最优模型进行性能测试。该环节旨在客观评估模型在未参与训练样本上的泛化能力与判别精度,以期为准确识别金矿成因归属并指导找矿勘查提供可靠的量化判别工具。

## 1 数据收集与处理

笔者收集了全球范围内不同成因类型矿床的黄铁矿微量元素数据(表 1)。由于测试工作由不同实验室完成,涉及多种分析方法(如 LA-ICP-MS、EPMA 等)及不同的测试条件,因此不可避免地存在一定的

数据异质性。这种差异可能对部分微量元素的绝对含量产生影响,进而对模型训练过程引入潜在的不确定性。然而,机器学习算法通常更依赖于高维特征之间的相对关系与数据分布模式,而非单一元素的绝对值,这在本质上有助于减弱不同测试条件所带来的偏差。为进一步降低多源数据合并对分类结果的影响,在数据整理过程中对异常值进行了筛选,并优先选取分析精度较高、数据质量可靠的文献数据。经过清洗,最

表 1 文中使用的黄铁矿微量元素数据来源表

Tab. 1 The source of trace element data for pyrite used in the paper

| 地区    | 矿床名称                       | 矿床类型    | 数据量 | 文献来源                    |
|-------|----------------------------|---------|-----|-------------------------|
| 伊朗    | Chah Zard金银矿               | 浅成低温热液型 | 62  | Kouhestani et al., 2017 |
| 印度尼西亚 | Bawone金矿, Binebase金矿       | 浅成低温热液型 | 63  | King et al., 2014       |
| 中国江西  | 德兴铜矿                       | 斑岩型     | 4   | Reich et al., 2013      |
| 澳大利亚  | Chalkidiki斑岩铜矿             | 斑岩型     | 117 | Gregory et al., 2019    |
| 澳大利亚  | Cadia斑岩铜矿                  | 斑岩型     | 614 | Gregory et al., 2019    |
| 意大利   | Apuan Alps矿床               | 造山型     | 77  | George et al., 2018     |
| 中国山东  | 新城金矿, 焦家金矿                 | 造山型     | 25  | Mills et al., 2015      |
| 中国江西  | 大桥金矿                       | 造山型     | 13  | Wu et al., 2018         |
| 印度    | Hutti金矿, Hira-Buddini金矿    | 造山型     | 14  | Hazarika et al., 2018   |
| 澳大利亚  | Bendigo矿床                  | 造山型     | 24  | Thomas et al., 2011     |
| 加拿大   | Spanish Mountain矿床         | 造山型     | 27  | Large et al., 2009      |
| 中国陕西  | 金龙山金矿                      | 造山型     | 26  | Zhang et al., 2014      |
| 澳大利亚  | Darlot矿床                   | 造山型     | 22  | Belousov et al., 2016   |
| 澳大利亚  | East Repulse矿床             | 造山型     | 44  | Belousov et al., 2016   |
| 澳大利亚  | Fortnum矿床                  | 造山型     | 56  | Belousov et al., 2016   |
| 澳大利亚  | Golden Mile矿床              | 造山型     | 86  | Belousov et al., 2016   |
| 澳大利亚  | Granny Smith矿床             | 造山型     | 24  | Belousov et al., 2016   |
| 澳大利亚  | Lancefield矿床               | 造山型     | 18  | Belousov et al., 2016   |
| 澳大利亚  | Mars Pit矿床                 | 造山型     | 34  | Belousov et al., 2016   |
| 澳大利亚  | Meekatharra Prohibition矿床  | 造山型     | 20  | Belousov et al., 2016   |
| 澳大利亚  | Micky Doolan Meekatharra矿床 | 造山型     | 18  | Belousov et al., 2016   |
| 澳大利亚  | Minjar矿床                   | 造山型     | 20  | Belousov et al., 2016   |
| 澳大利亚  | Nathans Labouchere矿床       | 造山型     | 22  | Belousov et al., 2016   |
| 澳大利亚  | Paddington矿床               | 造山型     | 18  | Belousov et al., 2016   |
| 澳大利亚  | Songvang矿床                 | 造山型     | 16  | Belousov et al., 2016   |
| 澳大利亚  | Sunrise Dam矿床              | 造山型     | 24  | Belousov et al., 2016   |
| 澳大利亚  | Victory矿床                  | 造山型     | 50  | Belousov et al., 2016   |
| 澳大利亚  | Hill50矿床                   | 造山型     | 40  | Belousov et al., 2016   |
| 澳大利亚  | Wallaby矿床                  | 造山型     | 40  | Belousov et al., 2016   |
| 澳大利亚  | Wiluna矿床                   | 造山型     | 124 | Belousov et al., 2016   |

终获得 1 769 条可用于模型训练与测试的样本数据。其中, 包含浅成低温热液型矿床数据 125 条, 斑岩型矿床数据 735 条以及造山型矿床数据 909 条。此外, 尽管研究初期也收集了卡林型矿床的黄铁矿微量元素数据, 但经过数据清洗之后其有效样本数量较少, 为避免模型训练时产生严重的数据不平衡, 故未将其纳入最终数据集。为验证所构建模型的适用性与分类效果, 笔者整理了已公开报道的东昆仑造山带内黄铁矿微量元素数据 133 条, 用作模型的独立验证集。

对于每一条黄铁矿微量元素数据, 由于不同文献报道的元素种类存在差异, 笔者选取了 9 种在所有数据集中均有报道的元素用于机器学习模型训练, 分别为 Au、Co、Ni、Cu、Zn、As、Sb、Te、Pb。汇编了黄铁矿微量元素数据集的统计学特征(表 2, 图 1)。3 种类型矿床的黄铁矿 Au 含量变化范围相对较小, 其中浅成低温热液型矿床为  $0.01 \times 10^{-6} \sim 166.00 \times 10^{-6}$ , 平均值为  $12.35 \times 10^{-6}$ , 斑岩型矿床为  $0.01 \times 10^{-6} \sim 15.53 \times 10^{-6}$ , 平均值为  $0.31 \times 10^{-6}$ , 造山型矿床为  $0.01 \times 10^{-6} \sim 415.00 \times 10^{-6}$ , 平均值为  $10.77 \times 10^{-6}$ 。As 含量变化相对最大, 其中浅成低温热液型矿床为  $0.83 \times 10^{-6} \sim 76\,047.00 \times 10^{-6}$ , 平均值为  $12\,013.62 \times 10^{-6}$ , 斑岩型矿床为  $0.22 \times 10^{-6} \sim 14\,022.31 \times 10^{-6}$ , 平均值为  $699.22 \times 10^{-6}$ , 造山型矿床为  $0.04 \times 10^{-6} \sim 68\,931.00 \times 10^{-6}$ , 平均值为  $3\,700.65 \times 10^{-6}$ 。Co、Cu、Zn 含量变化相对较大, Co 含量在浅成低温热液型矿床和斑岩型矿床要高于在造山型矿床的含量, 依次为  $0.01 \times 10^{-6} \sim 21\,417.67 \times 10^{-6}$  (平均值为  $1\,195.87 \times 10^{-6}$ )、 $0.02 \times 10^{-6} \sim 9\,600.73 \times 10^{-6}$  (平均值为  $1\,097.80 \times 10^{-6}$ )、 $0.02 \times 10^{-6} \sim 8\,082.00 \times 10^{-6}$  (平均值为  $366.98 \times 10^{-6}$ )。Cu 含量

在浅成低温热液型矿床要高于在斑岩型矿床和造山型矿床的含量, 依次为  $3.34 \times 10^{-6} \sim 196\,448.00 \times 10^{-6}$  (平均值为  $9\,844.66 \times 10^{-6}$ )、 $0.01 \times 10^{-6} \sim 8\,294.81 \times 10^{-6}$  (平均值为  $240.67 \times 10^{-6}$ )、 $0.06 \times 10^{-6} \sim 6\,493.16 \times 10^{-6}$  (平均值为  $238.07 \times 10^{-6}$ )。Zn 含量大致相当, 在浅成低温热液型矿床中的含量为  $0.43 \times 10^{-6} \sim 13\,383.00 \times 10^{-6}$ , 平均值为  $277.82 \times 10^{-6}$ , 斑岩型矿床的含量为  $0.01 \times 10^{-6} \sim 6\,218.86 \times 10^{-6}$ , 平均值为  $33.75 \times 10^{-6}$ , 造山型矿床为  $0.02 \times 10^{-6} \sim 13\,609.00 \times 10^{-6}$ , 平均值为  $122.69 \times 10^{-6}$ 。

综上所述, 不同类型矿床黄铁矿微量元素虽然在分布范围上存在一定重合, 但平均含量以及总体的变化趋势在某些元素上存在明显不同, 因此通过这些元素进行矿床类型的区分在理论上可行。

## 2 机器学习模型和评价方法

### 2.1 机器学习模型

本次研究采用 KNN 模型、RF 模型和 SVM 模型 3 种常用的机器学习方法对不同成因类型的金矿床进行分类模型训练。这 3 种模型在处理复杂非线性关系方面表现卓越, 涵盖了从个体判别到集成学习的不同维度。

KNN 模型是一种原理简明、应用广泛的基本分类与回归方法。其核心思想是将待预测样本表示为高维特征空间中的向量, 通过计算其与训练集中样本特征向量之间的距离, 选取距离最小的  $k(k \geq 1)$  个邻近样本, 并根据这些邻居的类别进行分类判断。当进行分类任务时, KNN 模型通过多数投票原则决定待分

表 2 不同矿床类型中黄铁矿微量元素统计表( $10^{-6}$ )

Tab. 2 Statistical table of trace elements in pyrite among three types of ore deposits

| 元素 | 浅成低温热液型         |           | 斑岩型            |          | 造山型            |          |
|----|-----------------|-----------|----------------|----------|----------------|----------|
|    | 含量              | 平均值       | 含量             | 平均值      | 含量             | 平均值      |
| Au | 0.01~166.00     | 12.35     | 0.01~15.53     | 0.31     | 0.01~415.00    | 10.77    |
| Co | 0.01~21 417.67  | 1 195.87  | 0.02~9 600.73  | 1 097.80 | 0.02~8 082.00  | 366.98   |
| Ni | 0.06~3 330.00   | 385.06    | 0.06~9 331.38  | 661.26   | 0.27~8 189.28  | 448.53   |
| Cu | 3.34~196 448.00 | 9 844.66  | 0.01~8 294.81  | 240.67   | 0.06~6 493.16  | 238.07   |
| Zn | 0.43~13 383.00  | 277.82    | 0.01~6 218.86  | 33.75    | 0.02~13 609.00 | 122.69   |
| As | 0.83~76 047.00  | 12 013.62 | 0.22~14 022.31 | 699.22   | 0.04~68 931.00 | 3 700.65 |
| Sb | 0.05~3 459.00   | 249.44    | 0.01~635.65    | 7.01     | 0.01~5 090.00  | 137.09   |
| Te | 0.11~1 430.90   | 88.98     | 0.01~3 706.62  | 13.19    | 0.03~720.70    | 16.92    |
| Pb | 0.05~7 641.00   | 509.07    | 0.01~1 132.00  | 30.86    | 0.01~9 550.00  | 146.37   |

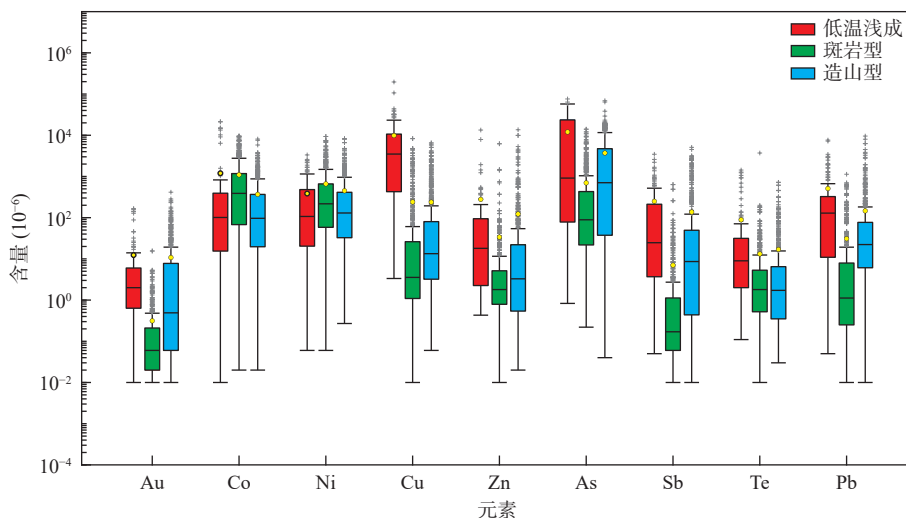


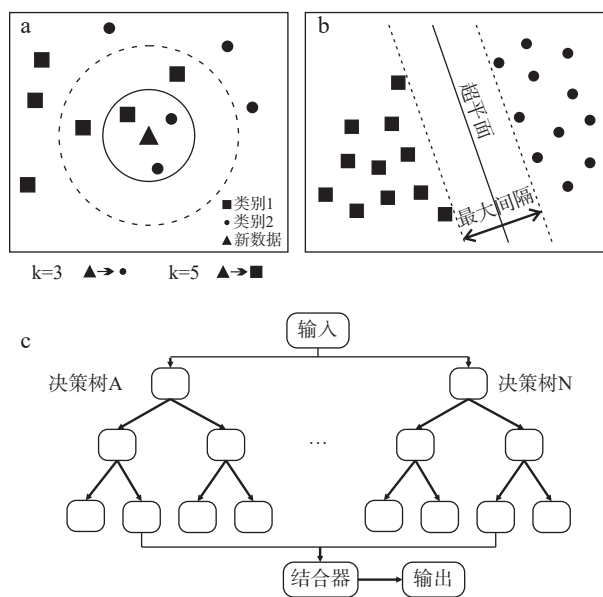
图1 文中用于训练的黄铁矿微量元素箱状图

Fig. 1 Box illustrations of trace elements of pyrite used for training in this paper

类样本的类别。如果  $k=1$ , 则该样本直接被归类为其最近邻样本的类别(图 2a)。较小的  $k$  值易使模型受到噪声数据的干扰, 而较大的  $k$  值则可能导致类别界限模糊, 因此需要通过交叉验证来确定最优  $k$  值。

RF 模型是一种基于集成学习理论的监督式机器学习算法, 其核心机制是通过自助法重采样, 从原始数据集中有放回地随机抽取多个子样本集, 并为每个子样本集构建一棵决策树(胡祥祥等, 2025)。在决策树的生长过程中, RF 模型引入了特征随机选取技术, 即在每个节点分裂时仅从随机抽取的特征子集中选择最优属性, 从而显著降低了各决策树之间的相关性。RF 模型通过对所有决策树的预测结果进行多数投票来确定目标样本的分类归属(图 2b)。相比于单一分类器, 随机森林通过集合多个弱分类器的预测结果, 表现出较强的泛化能力和抗噪声性能, 能有效克服决策树易趋向过拟合的局限。

SVM 模型是一种基于统计学习理论和结构风险最小化原则的监督式算法(Bhavsar et al., 2012; 王本栋等, 2024), 是通过核函数将低维空间的输入向量非线性地映射至高维特征空间, 并在该空间中寻求一个能够将不同类别样本精准分离的最优超平面。该超平面两侧的支持向量与其所对应的平行边界共同确定了分类的决策边界, 当间隔最大化时, 模型获得最优的分类性能(图 2c)。这一机制不仅确保了模型对训练数据的准确分类, 更显著增强了模型的泛化能力。SVM 模型在处理小样本、高维数以及非线性分布的地球化学数据集时表现出极强的鲁棒性。



a. KNN 模型原理; b. RF 模型原理; c. SVM 模型原理

图2 机器学习模型原理简图

Fig. 2 Schematic diagram of KNN, RF, SVM

## 2.2 模型训练与评价

在设计机器学习模型训练过程中, 本文将总数据集( $n=1769$ )按照 8 : 2 的比例随机划分为训练集( $n=1415$ )和测试集( $n=354$ )。由于不同类型矿床中黄铁矿微量元素数据分布不均(浅成低温热液型 : 斑岩型 : 造山型  $\approx 1 : 7 : 9$ ), 为避免数据不平衡对模型训练准确性产生干扰(郭广慧等, 2023; 刘嘉情等, 2023; 2025), 在训练前采用过采样方法对数据集进行平衡处理。此外, 采取五倍交叉验证用以减少单一训练可能产生的误差。在训练过程中同步进行超参数优化, 之后利

用测试集评估模型性能,并输出模型最后的分类评价指标。若机器学习模型在测试集上表现较好,则进一步通过独立验证集进行外部验证,以准确率评估模型在实际应用中的可行性与稳定性。

由于机器学习模型可能存在不同程度的欠拟合和过拟合问题,需通过超参数优化以提升模型的泛化能力与预测精度。使用贝叶斯搜索技术对3种机器学习模型的关键超参数进行优化。完成超参数调优后,分别采用优化后的模型对测试集进行预测,并生成相应的混淆矩阵。混淆矩阵用于展示模型预测结果与实际矿床类型之间的对应关系:矩阵的行代表样本的真实类别,每行的总数即为该类别的实际样本数量;列代表模型的预测类别,每列的总数为被预测为该类别的样本数量;而对角线上的元素则代表预测正确的样本数。通过分析混淆矩阵,可以系统评估各模型在不同矿床类型上的分类准确性及泛化能力。在此基础上,进一步利用3种模型的混淆矩阵计算更全面的分类评估指标,包括查准率(Precision)、召回率(Recall)与F1-score。其中,查准率衡量的是被模型预测为正类的样本中实际为正类的比例,召回率衡量的是实际为正类的样本中被模型正确预测的比例,而F1-score则是准确率与召回率的调和平均数,能够综合反映模型在特定类别上的分类性能。

### 3 结果与讨论

#### 3.1 模型训练评价结果

使用测试集对3个已训练模型的性能进行评估,得到相应的分类混淆矩阵(图3)和对应的模型评价指标(表3)。结果表明,3种模型在识别斑岩型和造山型矿床时的性能普遍优于浅成低温热液型矿床。这

一差异可能源于浅成低温热液型矿床的数据量较少,限制了算法对该类型矿床微量元素特征空间的充分提取与学习。

从模型整体性能来看,KNN模型得分相对较低,为0.89;RF模型表现最优,分类准确率达0.96;SVM模型次之,为0.95。上述结果表明,相较于基于距离度量的KNN模型,采用集成学习策略的RF模型和基于高维空间划分的SVM模型更能有效捕捉黄铁矿微量元素之间的复杂非线性耦合关系。

#### 3.2 模型可靠性

测试集各项评估参数表明,3种机器学习模型均表现出优异的分类性能,初步验证了利用黄铁矿微量元素判别金矿床成因类型的可行性。然而,考虑到测试集部分数据可能与训练集样本存在同源或相似的地质背景关联,模型在测试阶段展现出的高精度可能无法完全反映其对完全未知数据的真实预测能力。为客观评价模型的泛化性能与稳健性,本研究进一步引入东昆仑造山带已知成因的黄铁矿样本作为独立验证集。该数据集全程未参与模型的训练与参数调优过程,且其造山型成因归属已获地质证据明确佐证。通过将该独立验证集代入已构建的模型,结合混淆矩阵(图4)分析其外部判别表现。鉴于该验证集样本均属于单一矿床类型,本阶段评估仅以准确率作为核心评价指标(表4)。验证结果显示,KNN模型准确率得分为0.72;尽管RF模型在测试阶段展现出最优的分类精度,但在独立验证集中的准确率仅为0.81。相比之下,SVM模型在验证集表现最为稳健,准确率达0.88,明显优于RF模型和KNN模型。

不同机器学习模型在测试集与独立验证集中表现出性能差异,可能反映了其对数据结构的敏感性的不同。RF模型通过随机采样构建复杂的非线性决策

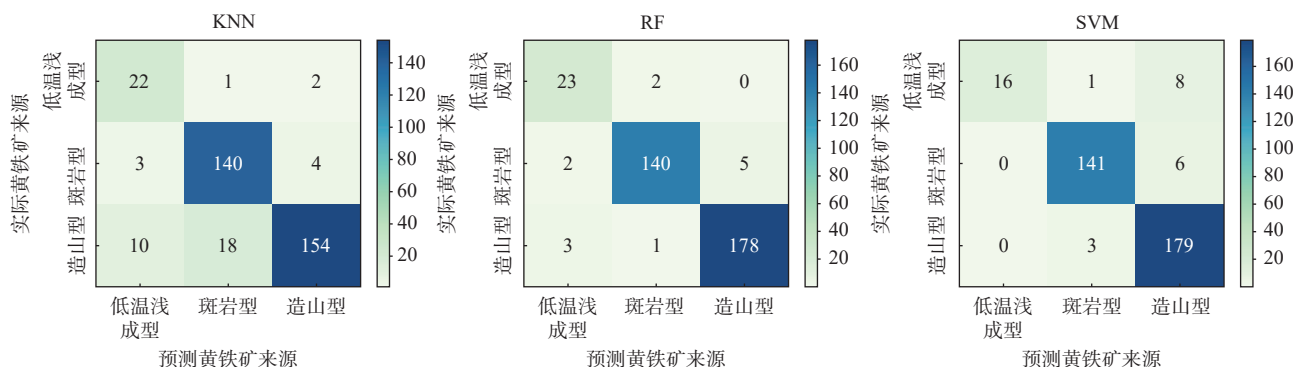


图3 机器学习模型测试集混淆矩阵

Fig. 3 Confusion matrix of machine learning model on test set

表 3 机器学习模型评估指标统计表

Tab. 3 Statistical table of machine learning model evaluation metrics

| 模型    | 数据类型    | 查准率  | 召回率  | F1-score | 准确率  |
|-------|---------|------|------|----------|------|
| KNN模型 | 浅成低温热液型 | 0.63 | 0.88 | 0.73     | 0.89 |
|       | 斑岩型     | 0.88 | 0.95 | 0.92     |      |
|       | 造山型     | 0.96 | 0.85 | 0.90     |      |
| RF模型  | 浅成低温热液型 | 0.82 | 0.92 | 0.87     | 0.96 |
|       | 斑岩型     | 0.98 | 0.95 | 0.97     |      |
|       | 造山型     | 0.97 | 0.98 | 0.98     |      |
| SVM模型 | 浅成低温热液型 | 1.00 | 0.64 | 0.78     | 0.95 |
|       | 斑岩型     | 0.97 | 0.96 | 0.97     |      |
|       | 造山型     | 0.93 | 0.98 | 0.95     |      |

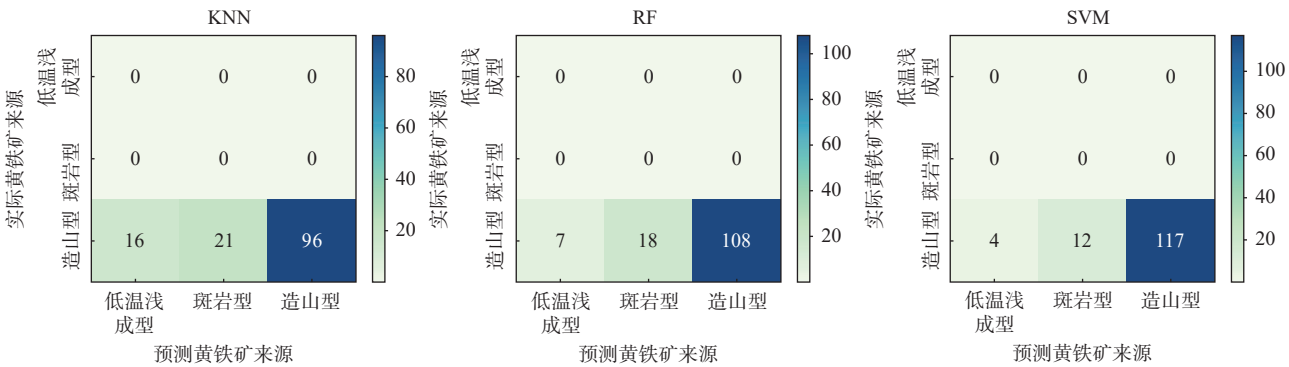


图4 机器学习模型验证集混淆矩阵

Fig. 4 Confusion matrix of machine learning model on validation set

表 4 机器学习模型验证集准确率得分统计表

Tab. 4 Statistical table of accuracy scores for machine learning models on validation set

| 模型 | KNN模型 | RF模型 | SVM模型 |
|----|-------|------|-------|
| 得分 | 0.72  | 0.81 | 0.88  |

边界,虽能捕捉微细特征,但在面对具有区域特异性的新样本时,易受训练集局部特征的干扰而产生过拟合现象,这一局限在先前的研究中亦有报道(Barreñada et al., 2024)。相比之下,SVM 模型通过构建最大间隔分类边界,并引入正则化机制控制模型复杂度,其决策函数相对平滑,对局部异常值或噪声不敏感,因此在跨区域数据应用中通常表现出更稳定的泛化能力。

综上所述,SVM 模型在 3 种模型中整体性能最优,具备较强的稳定性与预测能力,可作为基于黄铁矿微量元素区分浅成低温热液型、斑岩型与造山型矿床的有效工具,具有一定的实际应用价值。

3.3 特征值重要性分析

尽管机器学习已成为地球科学研究的重要工具,

但其复杂的“黑匣子”特性往往限制了研究人员对分类机制的深入理解。为提升决策者对模型的信任度并从中获取有效的科学认知,模型可解释性的重要性日益凸显。因此,笔者引入机器学习可解释方法 SHAP 值对表现最优的 SVM 模型的内部运作过程进行了解释,用以定量评估 Au、Co、Ni、Cu、Zn、As、Sb、Te 和 Pb 等 9 种黄铁矿微量元素对分类结果的影响程度。SHAP 特征贡献图能够直观展示不同特征对模型输出的相对重要性:图中纵坐标按特征重要性由高到低自上而下排列,越靠近上方的特征表明其对模型整体输出的贡献越大。每个数据点代表一个样本,点的颜色反映该样本对应特征的取值高低(蓝色至红色渐变表示特征值由低到高),而点在水平方向上的位置则体现了该特征对模型输出的正向或负向影响强度。

在运用黄铁矿微量元素成分识别斑岩型和造山型金矿时,对判别结果影响最大的 3 个元素依次为 As、Co 和 Cu(图 5)。尽管两种矿床的黄铁矿均以低 Cu

为特征,来自斑岩型和造山型金矿的黄铁矿表现出不同的 As、Co 含量趋势:来自造山型金矿的黄铁矿具有高 As、低 Co 的特征,而来自斑岩型金矿的黄铁矿具有低 As、高 Co 的特征。此外,两种矿床的其他微量元素也表现出截然不同的特征,如来自斑岩型矿床的黄铁矿具有高 Ni、低 Sb、低 Pb 和低 Zn 的趋势,而来自

自造山型金矿的黄铁矿则倾向于低 Ni、高 Sb、高 Pb 和高 Zn。与斑岩型金矿和造山型金矿不同,在运用黄铁矿成分识别浅成低温热液型金矿时最重要的特征依次为 Cu、As 和 Ni,并且 Cu 含量越高,越可能为浅成低温热液矿床。以上这些特征与 3 种矿床的黄铁矿微量元素统计结果一致(图 1)。

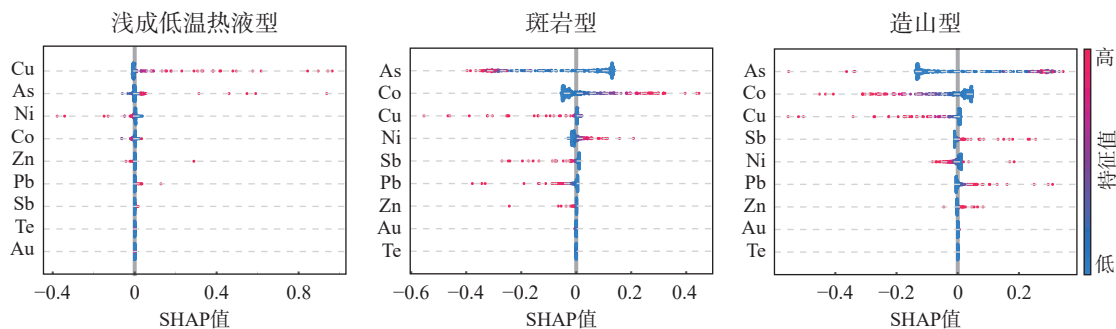


图5 不同类型矿床的 SHAP 特征贡献图

Fig. 5 SHAP feature contribution plot for three types of ore deposits

综上所述,SHAP 特征贡献图系统揭示了黄铁矿微量元素在不同矿床类型判别中的差异化影响模式。上述差异不仅从机器学习可解释性角度量化了各微量元素在不同成矿系统中的判别权重,更与各类型矿床黄铁矿的地球化学特征高度吻合,从而在模型预测能力与地质认知之间建立了可信的桥梁。不过,造成不同矿床之间黄铁矿地球化学差异的地质过程尚未完全查明,需要未来进一步研究。

### 3.4 存在问题和展望

文中基于黄铁矿原位微量元素大数据构建的 SVM 模型,在区分浅成低温热液型、斑岩型和造山型金矿床方面展现出较高的准确率与泛化能力。与传统二维判别图解相比,该高维机器学习模型能够深度挖掘多元素之间的非线性关系与综合特征,有效优化了矿床成因的判别流程,为找矿预测与靶区评价提供了更为客观的定量工具。

需要指出的是,文中所采用的数据集来源于全球不同区域的浅成低温热液型、斑岩型及造山型金矿床,不同区域在构造背景、岩浆演化及成矿流体性质等方面存在一定差异,这可能在一定程度上引入区域偏倚,从而对模型的泛化能力产生影响。尽管如此,黄铁矿微量元素组成在不同类型矿床中通常表现出相对稳定的判别特征,使得跨区域数据整合具备一定的可行性。以东昆仑地区金矿床作为独立验证数据集,其判别结果与已有地质认识相一致,表明该模型在一定程

度上具备跨区域应用能力。未来研究中,可进一步通过构建区域子数据集、开展分区建模或引入更多具有代表性的样本数据,对模型进行针对性优化,以提升其泛化能力与适用范围。

此外,多源数据融合不可避免地受制于不同实验室分析方法及测试条件的差异。未来需建立统一的数据筛选标准、引入标准参考物质进行归一化处理,或构建同一分析平台下的数据子集模型,以进一步提高数据的一致性和模型的可靠性。用于训练与验证的特征选择主要基于数据可获得性,可能未包含部分对矿床成因判别具有重要意义但报道较少的微量元素,且不同元素之间可能存在一定相关性或信息冗余,可能对模型判别能力产生影响。虽然文中已引入 SHAP 算法对主控元素进行了提取,但在未来研究中,仍需结合更全面的标准化测试数据与先进的降维算法,持续优化关键判别特征库,从而不断提升模型的预测精度与地球化学可解释性。

## 4 结论

(1)汇编了来自全球浅成低温热液型、斑岩型、造山型矿床中的黄铁矿微量元素数据,采用 KNN 模型、RF 模型与 SVM 模型 3 种机器学习方法,构建了基于黄铁矿微量元素判别上述 3 类矿床的机器学习分类模型。

(2)模型分类性能评估结果表明,训练所得 KNN 模型、RF 模型与 SVM 模型的分类准确率分别为 0.89、0.96 和 0.95。将上述模型应用于东昆仑造山带金矿识别,结果显示各模型均具备可靠的识别能力,其中 SVM 模型表现最佳,准确率达 0.88。

(3)SHAP 分析结果显示,在斑岩型和造山型金矿的识别过程中,对模型输出贡献最大的黄铁矿微量元素特征依次为 As、Co 和 Cu,而在识别浅成低温热液型金矿时,Cu、As 和 Ni 对模型输出的影响最为显著。未来工作中,需要进一步深化对控制上述元素地球化学行为的地质过程解释。

## 利益冲突声明(Conflict of Interests):

所有作者声明不存在利益冲突。

All authors disclose no relevant conflict of interests.

## 参考文献(References):

- 范宏瑞,李兴辉,左亚彬,等. LA-(MC)-ICP MS 和 (Nano) SIMS 硫化物微量元素和硫同位素原位分析与矿床形成的精细过程[J]. 岩石学报, 2018, 34(12): 3479-3496.
- FAN Hongrui, LI Xinghui, ZUO Yabin, et al. In-situ LA-(MC)-ICP MS and (Nano) SIMS trace elements and sulfur isotope analyses on sulfides and application to confine metallogenic process of ore deposit[J]. Acta Petrologica Sinica, 2018, 34(12): 3479-3496.
- 郭广慧,钟世华,李三忠,等. 运用机器学习和锆石微量元素构建花岗岩成矿潜力判别图解:以东昆仑祁漫塔格为例[J]. 西北地质, 2023, 56(6): 57-70.
- GUO Guanghui, ZHONG Shihua, LI Sanzhong, et al. Constructing Discrimination Diagrams for Granite Mineralization Potential by Using Machine Learning and Zircon Trace Elements: Example from the Qimantagh, East Kunlun[J]. Northwestern Geology, 2023, 56(6): 57-70.
- 胡祥祥,石亚亚,胡良柏,等. 融合 InSAR 与信息量-机器学习耦合模型的黄土滑坡易发性评价[J]. 西北地质, 2025, 58(2): 159-171.
- HU Xiangxiang, SHI Yaya, HU Liangbai, et al. Evaluation of Loess Landslide Susceptibility by Combining InSAR and Information-Machine Learning Coupling Model[J]. Northwestern Geology, 2025, 58(2): 159-171.
- 刘嘉情,钟世华,李三忠,等. 基于机器学习和全岩成分识别东昆仑祁漫塔格斑岩-矽卡岩矿床成矿岩体和贫矿岩体[J]. 西北地质, 2023, 56(6): 41-56.
- LIU Jiaqing, ZHONG Shihua, LI Sanzhong, et al. Identification of mineralized and barren magmatic rocks for the porphyry-skarn deposits from the Qimantagh, East Kunlun: based on machine learning and whole-rock compositions[J]. Northwestern Geology, 2023, 56(6): 41-56.
- 刘嘉情,钟世华,李三忠,等. 基于机器学习的磷灰石源岩识别方法[J]. 大地构造与成矿学, 2025, 49(2): 467-480.
- LIU Jiaqing, ZHONG Shihua, LI Sanzhong, et al. A Machine Learning-based Method for Identifying Apatite Source Rocks[J]. Geotectonica et Metallogenia, 2025, 49(2): 467-480.
- 钟世华,李三忠,丰成友,等. 原特提斯和古特提斯碰撞岩浆与斑岩成矿:以东昆仑祁漫塔格地区为例[J]. 地质通报, 2025a, 44(4): 511-533.
- ZHONG Shihua, LI Sanzhong, FENG Chengyou, et al. Proto-Tethys and Paleo-Tethys collisional magmas and porphyry mineralization: A case study of the Qimantagh area, East Kunlun Mountains[J]. Geological Bulletin of China, 2025a, 44(4): 511-533.
- 王本栋,李四全,许万忠,等. 基于 3 种不同机器学习算法的滑坡易发性评价对比研究[J]. 西北地质, 2024, 57(1): 34-43.
- WANG Bendong, LI Siquan, XU Wanzhong, et al. A Comparative Study of Landslide Susceptibility Evaluation Based on Three Different Machine Learning Algorithms[J]. Northwestern Geology, 2024, 57(1): 34-43.
- 钟世华,黄宇,刘永乐,等. 东昆仑志留纪——泥盆纪关键金属成矿大爆发[J]. 地质通报, 2025b, 44(4): 574-586.
- ZHONG Shihua, HUANG Yu, LIU Yongle, et al. Silurian-Devonian critical metal mineralization boom of the East Kunlun Orogenic Belt[J]. Geological Bulletin of China, 2025b, 44(4): 574-586.
- Barreñada L, Dhiman P, Timmerman D, et al. Understanding overfitting in random forest for probability estimation: a visualization and simulation study[J]. Diagnostic and Prognostic Research, 2024, 8(1): 14.
- Belousov I, Large R R, Meffre S, et al. Pyrite compositions from VHMS and orogenic Au deposits in the Yilgarn Craton, Western Australia: Implications for gold and copper exploration[J]. Ore Geology Reviews, 2016, 79: 474-499.
- Bhavsar H and Panchal M H. A review on support vector machine for data classification[J]. International Journal of Advanced Research in Computer Engineering Technology, 2012, 1: 185-189.
- Cao G, Zhang Y, Zhao H, et al. Trace element variations of pyrite in orogenic gold deposits: Constraints from big data and machine learning[J]. Ore Geology Reviews, 2023, 157: 105447.
- Fu G, Lü Q, Yan J, et al. 3D mineral prospectivity modeling based on machine learning: A case study of the Zhuxi tungsten deposit in northeastern Jiangxi Province, South China[J]. Ore Geology Reviews, 2021, 131: 104010.
- George L L, Biagioni C, D'Orazio M, et al. Textural and trace element evolution of pyrite during greenschist facies metamorphic

- recrystallization in the southern Apuan Alps (Tuscany, Italy): Influence on the formation of Ti-rich sulfosalt melt[J]. *Ore Geology Reviews*, 2018, 102: 59–105.
- Gregory D D, Cracknell M J, Large R R, et al. 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[J]. *Economic Geology*, 2019, 114: 771–786.
- Gul M A, Zhang H, Li Y, et al. Machine learning-driven classification of PbZn ore deposits using pyrite trace elements and isotopic signatures: A case study of the Gunga deposit[J]. *Journal of Geochemical Exploration*, 2025, 272: 107693.
- Hazarika P, Mishra B and Lochan Pruseth K. Trace-element geochemistry of pyrite and arsenopyrite: ore genetic implications for late Archean orogenic gold deposits in southern India[J]. *Mineralogical Magazine*, 2018, 81: 661–678.
- He Y, Zhou Y, Wen T, et al. A review of machine learning in geochemistry and cosmochemistry: Method improvements and applications[J]. *Applied Geochemistry*, 2022, 140: 105273.
- King J, WilliamsJones A, Van Hinsberg V, et al. High-sulfidation epithermal pyrite-hosted Au (Ag-Cu) ore formation by condensed magmatic vapors on Sangehe Island, Indonesia[J]. *Economic Geology*, 2014, 109: 1705–1733.
- Kouhestani H, Ghaderi M, Large R R, et al. Texture and chemistry of pyrite at Chah Zard epithermal gold-silver deposit, Iran[J]. *Ore Geology Reviews*, 2017, 84: 80–101.
- Large R R, Danyushevsky L, Hollit C, et al. 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[J]. *Economic Geology*, 2009, 104: 635–668.
- Li Q, Song H, Chi G, et al. Genesis of visible gold in pyrite in the Zhaoxian gold deposit, Jiaodong gold province, China: Constraints from EBSD micro-structural and LA-ICP-MS elemental analyses[J]. *Ore Geology Reviews*, 2021, 139: 104591.
- Li H, Zhang Z, Zhang R, et al. Geochemical discrimination of pyrite in diverse ore deposit types through statistical analysis and machine learning techniques[J]. *American Mineralogist*, 2024, 109(5): 846–57.
- Li S S, Zang C J, Li L, et al. Geochemistry, isotopes, and morphology of coarse-grained pyrite from the Jinqingding gold deposit, Jiaodong Peninsula: Implications for episodic ore-forming fluid evolution[J]. *Ore Geology Reviews*, 2025: 106742.
- Liu H, Harris J, Sherlock R, et al. Mineral prospectivity mapping using machine learning techniques for gold exploration in the Larder Lake area, Ontario, Canada[J]. *Journal of Geochemical Exploration*, 2023, 253: 107279.
- Mills S E, Tomkins A G, Weinberg R F, et al. Implications of pyrite geochemistry for gold mineralisation and remobilisation in the Jiaodong gold district, northeast China[J]. *Ore Geology Reviews*, 2015, 71: 150–168.
- Qiu K, Zhou T, Chew D, et al. Apatite trace element composition as an indicator of ore deposit types: A machine learning approach[J]. *American Mineralogist*, 2024, 109: 303–314.
- Reich M, Deditius A, Chrysosoulis S, et al. Pyrite as a record of hydrothermal fluid evolution in a porphyry copper system: A SIMS/EMPA trace element study[J]. *Geochimica et Cosmochimica Acta*, 2013, 104: 42–62.
- Ribeiro D, Rottier B, Godet A, et al. Gold mobility in Archean metasedimentary belts: Implications for orogenic gold deposits[J]. *Geology*, 2026, 54(2): 99–104.
- Shirmard H, Farahbakhsh E, Müller R D, et al. A review of machine learning in processing remote sensing data for mineral exploration[J]. *Remote Sensing of Environment*, 2022, 268: 112750.
- Thomas H V, Large R R, Bull S W, et al. Pyrite and pyrrhotite textures and composition in sediments, laminated quartz veins, and reefs at Bendigo gold mine, Australia: Insights for ore genesis[J]. *Economic Geology*, 2011, 106: 1–31.
- Voute F, Hagemann S, Evans N, et al. Sulfur isotopes, trace element, and textural analyses of pyrite, arsenopyrite and base metal sulfides associated with gold mineralization in the Pataz-Parcoy district, Peru: implication for paragenesis, fluid source, and gold deposition mechanisms[J]. *Mineralium Deposita*, 2019, 54: 1077–1100.
- Wang Y, Qiu K F, Telea A C, et al. Interpreting mineral deposit genesis classification with decision maps: A case study using pyrite trace elements[J]. *American Mineralogist*, 2024, 109(12): 2116–26.
- Wu Y, Li J, Evans K, et al. Ore-Forming Processes of the Daqiao Epizonal Orogenic Gold Deposit, West Qinling Orogen, China: Constraints from Textures, Trace Elements, and Sulfur Isotopes of Pyrite and Marcasite, and Raman Spectroscopy of Carbonaceous Material[J]. *Economic Geology*, 2018, 113: 1093–1132.
- Zhang J, Li L, Gilbert S, et al. LA-ICP-MS and EPMA studies on the Fe-S-As minerals from the Jinlongshan gold deposit, Qinling Orogen, China: implications for ore-forming processes[J]. *Geological Journal*, 2014, 49: 482–500.
- Zhong S, Liu Y, Li S, et al. A machine learning method for distinguishing detrital zircon provenance[J]. *Contributions to Mineralogy and Petrology*, 2023, 178: 35.