



引文格式: 樊松浩, 王达, 朱国韬, 等. 机器学习在金矿床成因类型识别中的应用: 基于黄铁矿微量元素地球化学特征的探索研究[J]. 西北地质, 2026, 59(4): 30–47. DOI: 10.12401/j.nwg.2025187

Citation: FAN Songhao, WANG Da, ZHU Guotao, et al. Application of Machine Learning in the Genetic Classification of Gold Deposits: An Exploratory Study Based on Pyrite Trace Element Geochemistry[J]. Northwestern Geology, 2026, 59(4): 30–47. DOI: 10.12401/j.nwg.2025187

机器学习在金矿床成因类型识别中的应用: 基于黄铁矿微量元素地球化学特征的探索研究

樊松浩^{1,2,3}, 王达^{2,*}, 朱国韬², 杨彪^{1,3}, 苏攀云^{1,3}, 侯秀宏^{1,3},
吕汉秦^{1,3}, 韩燕东^{1,3}, 陈磊^{1,3}

(1. 中国地质调查局呼和浩特自然资源综合调查中心, 内蒙古 呼和浩特 010010; 2. 中国地质大学(北京)地球科学与资源学院, 北京 100083; 3. 中国地质学会华北克拉通北缘金矿勘查技术创新基地, 内蒙古 呼和浩特 010010)

摘要: 黄铁矿是金矿床中最常见的金属硫化物, 其微量元素组成能够有效记录成矿流体性质与成矿环境信息, 在金矿床成因类型判别中具有重要指示意义。随着信息技术的迅速发展, 地球科学研究已进入数据密集化阶段, 多源、多尺度地球化学数据的积累为矿床类型的定量识别提供了新的契机。然而, 现有研究仍存在样本规模有限、矿床类型覆盖不足及判别精度不高等问题。针对上述不足, 笔者系统汇集铁氧化物-铜-金 (IOCG) 型、火山成因块状硫化物 (VMS) 型、斑岩型、低硫浅成低温热液型、高硫浅成低温热液型、卡林型及造山型 7 种类型金矿床的黄铁矿 LA-ICP-MS 微量元素数据并构建大数据集。在完成缺失值控制与中心对数比 (CLR) 变换的基础上, 引入支持向量机 (SVM) 与随机森林 (RF) 模型开展矿床类型判别, 并通过交叉验证与多指标评价对模型性能进行系统对比。结果显示, 基于原始 CLR 特征空间的 SVM 模型在分类精度与稳定性方面表现最优; RF 模型在保持较高判别能力的同时, 结合 SHAP 分析揭示了不同矿床类型判别所依赖的关键微量元素组合特征。研究表明, 机器学习方法能够有效提取具有地质意义的成矿指示信息, 将高性能分类模型与可解释性分析相结合, 是开展复杂成矿体系中金矿床类型识别与成因研究的有效技术路径。

关键词: 机器学习; 黄铁矿; 微量元素特征; 金矿成因类型

中图分类号: P618; P628

文献标志码: A

文章编号: 1009-6248(2026)04-0030-18

Application of Machine Learning in the Genetic Classification of Gold Deposits: An Exploratory Study Based on Pyrite Trace Element Geochemistry

FAN Songhao^{1,2,3}, WANG Da^{2,*}, ZHU Guotao², YANG Biao^{1,3}, SU Panyun^{1,3}, HOU Xiuhong^{1,3},

收稿日期: 2026-02-24; 修回日期: 2026-04-07; 责任编辑: 曹佰迪

基金项目: 地球深部探测与矿产资源勘查国家科技重大专项 (2025ZD1006500), 内蒙古自治区自然科学基金 (2024QN04015), 中国地质调查局项目 (DD20230384、DD202602109403、DD20251185) 联合资助。

作者简介: 樊松浩 (1994–), 男, 工程师, 从事基础地质调查研究与矿产资源勘查工作。E-mail: 625040570@qq.com。

* 通讯作者: 王达 (1990–), 男, 副教授, 博士生导师, 研究方向为关键金属成矿研究与勘查评价。E-mail: 13311532737@163.com。

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

LYU Hanqin^{1,3}, HAN Yandong^{1,3}, CHEN Lei^{1,3}

(1. Hohhot Natural Resources Comprehensive Survey Center of China Geological Survey, Hohhot 010010, Inner Mongolia, China; 2. School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China; 3. Innovation Base for Gold Exploration Technology in the Northern Margin of North China Craton, Geological Society of China, Hohhot 010010, Inner Mongolia, China)

Abstract: Pyrite is the most common metallic sulfide in gold deposits, and its trace element composition effectively records the physicochemical properties of ore-forming fluids and mineralization environments, thus serving as an important indicator for discriminating genetic types of gold deposits. With the rapid development of information technology, geoscience research has entered a data-intensive era, and the accumulation of multi-source, multi-scale geochemical data provides new opportunities for quantitative classification of deposit types. However, existing studies are still limited by small sample sizes, insufficient coverage of deposit types, and relatively low classification accuracy. To address these issues, this study systematically compiles LA-ICP-MS trace element data of pyrite from seven types of gold deposits, including iron oxide-copper-gold (IOCG), volcanogenic massive sulfide (VMS), porphyry, low-sulfidation and high-sulfidation epithermal, Carlin-type, and orogenic deposits, and constructs a large geochemical dataset. After missing value control and centered log-ratio (CLR) transformation, Support Vector Machine (SVM) and Random Forest (RF) models are applied to classify deposit types, and their performance is systematically evaluated using cross-validation and multiple metrics. The results show that the SVM model based on the original CLR feature space achieves the best performance in terms of classification accuracy and stability, whereas the RF model, while maintaining strong predictive capability, reveals the key trace element combinations controlling classification through SHAP analysis. Overall, this study demonstrates that machine learning methods can effectively extract geologically meaningful mineralization signatures, and that the integration of high-performance classification models with interpretability analysis provides an effective approach for the identification and genetic study of gold deposit types in complex mineralization systems.

Keywords: machine learning; pyrite; trace element characteristics; genetic types of gold deposits

自21世纪以来,人工智能技术迅速发展,机器学习作为其核心方法之一,已在医疗、金融和交通等多个领域取得显著应用成效(Ngai et al., 2011; Esteva et al., 2019; Haydari et al., 2023)。在地球科学领域,机器学习凭借其在复杂数据处理、模式识别及非线性关系刻画方面的优势,近年来逐渐应用于矿床预测、地质灾害评估等领域(翟明国等, 2018; 郭广慧等, 2023; 刘嘉情等, 2023; 胡祥祥等, 2025; 张茂省等, 2026)。随着激光剥蚀-电感耦合等离子体质谱(LA-ICP-MS)、电子探针显微分析(EMPA)、遥感探测及地球物理成像等高精度技术的不断发展与广泛应用,地球化学、遥感及地质结构等多源数据得以高精度、大规模获取,推动地球科学研究逐步进入“数据密集化”与“智能化”并行发展的新阶段(黄少芳等, 2016; Merghadi et al., 2020)。在此背景下,如何有效整合多源、多尺度地学数据并开展定量分析,已成为当前矿产资源研究的重要发展方向。

矿产资源是支撑国家经济建设和工业化进程的重要物质基础(赵洋等, 2011)。随着浅部矿产资源的逐步消耗,找矿工作正逐渐向深部及隐伏矿体拓展(郑超杰, 2021),勘查难度显著增加,对找矿理论与技术体系提出了更高要求(赵东博, 2021)。传统依赖经验判断及单一指标异常识别的方法,在复杂成矿体系中往往存在信息利用不足、判别精度有限等问题(左仁广等, 2021),难以满足深部找矿与多类型矿床识别的需求。因而,亟需引入能够处理高维、多变量数据的定量分析方法,以提升矿床识别的准确性与科学性。

在金属矿产勘查中,金矿床因其重要的经济价值始终是研究重点。黄铁矿作为金矿床中最为常见的金属硫化物,在多种成因类型矿床中广泛发育(高振敏等, 2000)。大量研究表明,黄铁矿中微量元素组成能够有效记录成矿流体来源、温压条件及成矿过程中的物质迁移与演化信息,是解析成矿过程的重要地球化学载体(徐国凤等, 1979)。随着 LA-ICP-MS 和 EM-

PA 等技术的快速发展,黄铁矿微量元素能够实现高精度原位测定,为矿床成因研究与类型判别提供了可靠的数据支撑(Arehart et al., 1996; Gregory et al., 2019)。不同成因类型金矿床在成矿流体来源、物化条件及演化路径方面存在显著差异,这些差异会直接影响黄铁矿中微量元素的富集与分配特征(Steadman et al., 2021),并在其元素组合中形成具有指示意义的地球化学信息。因此,基于黄铁矿微量元素特征开展矿床成因判别,具有重要的理论意义与应用价值。

然而,现有研究多集中于局部区域或单一矿床类型,样本规模有限,矿床类型覆盖不足,所涉及微量元素种类相对受限,导致已有判别模型在复杂多类型体系中的适用性与泛化能力仍然不足(严育通等, 2012; 李洪梁等, 2019)。在地学大数据背景下,如何系统整合不同成因类型金矿床中黄铁矿的微量元素数据,并结合机器学习方法构建高精度、高稳定性的分类模型,从而实现矿床类型的定量判别,已成为当前亟待解决的关键科学问题。

基于此,笔者以不同成因类型金矿床中黄铁矿为研究对象,系统汇集其微量元素地球化学数据,构建多类型矿床综合数据集,在数据预处理基础上引入支持向量机(SVM)和随机森林(RF)两种典型机器学习模型开展成因类型判别研究,并对模型分类性能及稳定性进行对比分析。研究结果旨在为金矿床成因类型识别提供新的量化方法,同时为复杂成矿体系中机器学习方法的应用提供理论依据与技术参考。

1 数据收集与预处理

1.1 数据收集

笔者针对 LA-ICP-MS 技术测得的不同成因类型金矿床黄铁矿微量元素进行系统收集,进而将整理后的 6 731 条黄铁矿微量元素数据构建成黄铁矿地球化学数据集(图 1)。尽管文中所汇编的数据来源于不同文献,但通过中心对数比(CLR)变换将分析重点从绝对含量转向元素相对比例,可有效削弱不同实验条件带来的尺度差异。多元素组合在不同矿床类型中表现出的稳定结构特征,以及模型在交叉验证中的良好稳定性,表明文中识别的差异主要受地质过程控制,而非分析条件差异所致。由于 SEDEX 型矿床主要富集铅锌银,金品位含量较低,并非典型的金矿(Emsbo et al., 2000; Leach et al., 2005; 朱昊磊等, 2023; 邓依等, 2024),故本研究着重收集数据包含除 SEDEX 型矿床外的铁氧化物铜金(IOCG)型、火山块状硫化物(VMS)型、斑岩型、低硫浅成低温热液型、高硫浅成低温热液型、卡林型、造山型等 7 类金矿床成因类型,重点研究数据齐全的 Co、Ni、Cu、Zn、As、Se、Mo、Ag、Sb、Te、Au、Pb、Bi 等元素(图 1)。

黄铁矿微量元素地球化学大数据集中的各个元素的统计学特征(表 1, 图 2)显示,不同成因类型的金矿床存在微量元素富集特征差异。初步的原始数据特征分析发现: IOCG 型矿床表现出明显的 Te-Co-

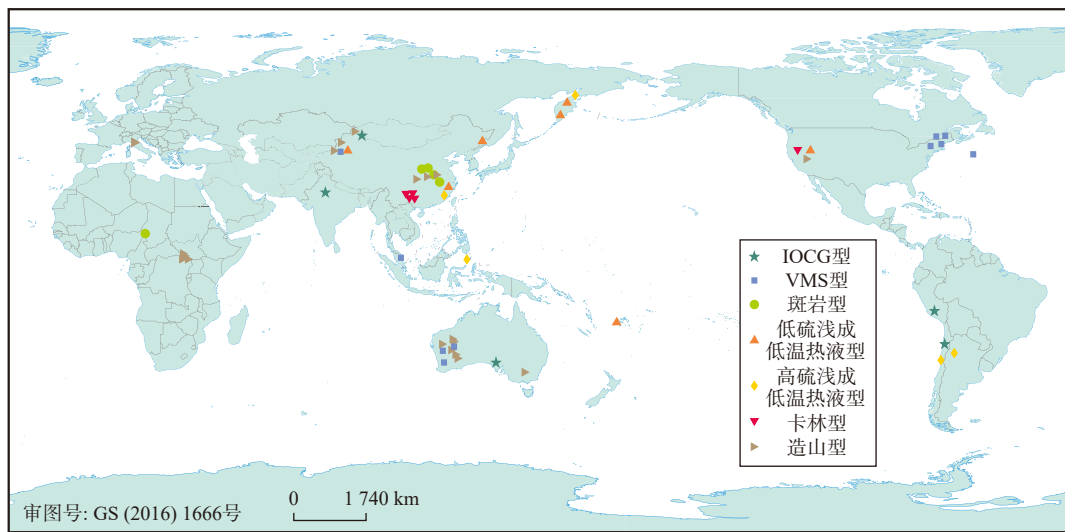


图1 黄铁矿数据采集矿点分布图

Fig. 1 Distribution map of pyrite data sampling points

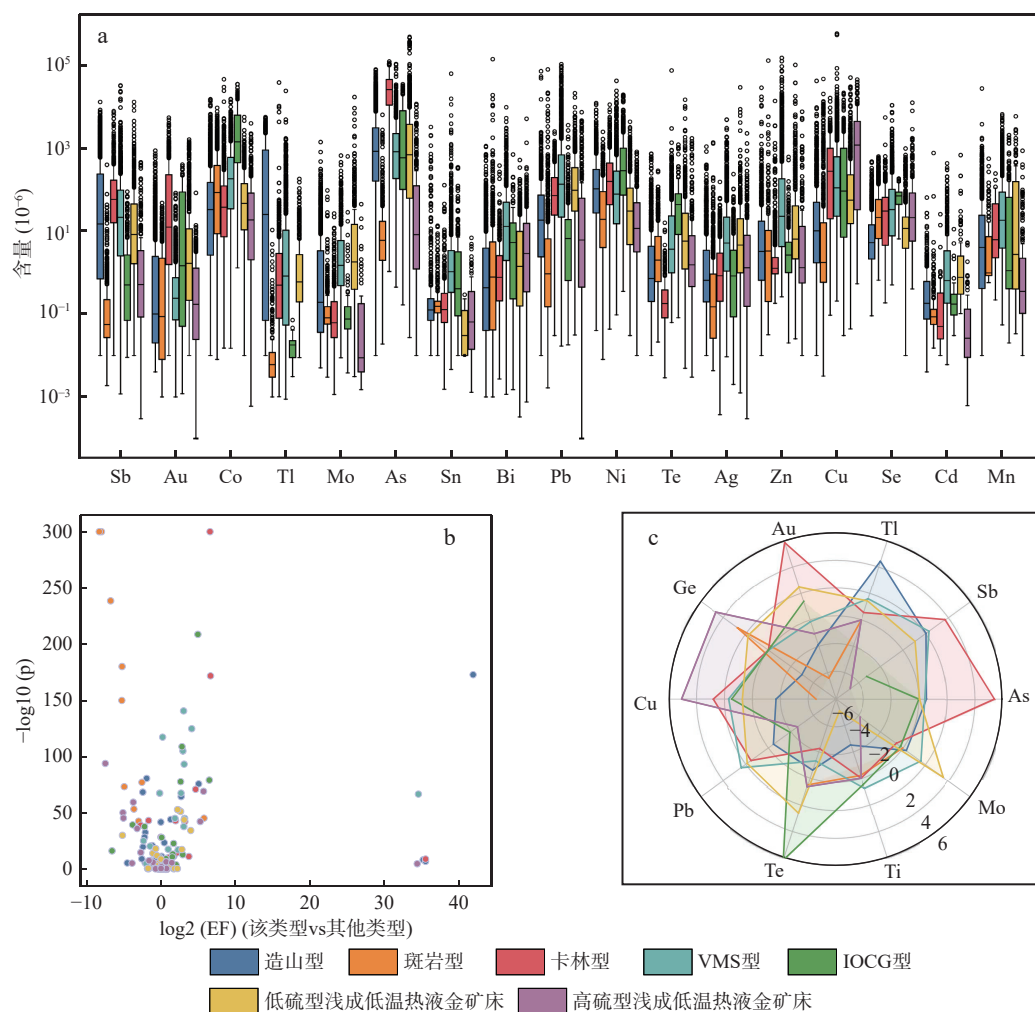
Ni-V-Se-Cu-Au 元素组合富集特征, 其中 Te 的富集程度最为显著($\log_2 EF > 6$; EF 为各矿床类型元素中位数与全样本中位数之比), 同时 Co、Ni 与 V 呈现稳定增强, 反映出深源或幔源流体对成矿体系的显著贡献, Co-Ni-V 的协同富集通常指示铁质-幔源物质输入及深大断裂控制, 而 Te、Bi 与 Au、Cu 的伴生则表明高温、强氧化性流体条件下铁氧化物-硫化物共存的成矿环境, 该元素组合是 IOCG 型矿床的典型地球化学指纹, 在多矿床成因类型对比中具有良好的判别性。

VMS 型矿床显示出高度集中的 Zn-Pb-Cu-Ag-Sb-Bi-In 富集特征, 且多数元素缺失率极低, 说明该类矿床具有稳定而一致的元素组合。Zn、Pb 与 Cu 的显著富集反映了火山喷流环境中硫化物的大规模沉淀过程, 而 Ag、Bi、In 与 Sn 的同步富集则指示高温火

山热液体系对挥发性和亲硫元素的有效富集。Mn 的相对增强进一步表明热液与水混合过程在成矿中的重要作用, 该元素组合构成了典型且易于识别的 VMS 型矿床地球化学特征。

低硫型矿床以 Au-Ag-Te-Mo 的显著富集为主要特征, Pb、Tl 与 Sb 呈现次级增强, 而 As 基本接近背景水平。Au 与 Ag 的强烈富集反映了中低温条件下贵金属的有效沉淀, Te 的同步富集则指示岩浆流体对挥发分元素的贡献。As 缺乏明显富集是该类矿床区别于高硫型浅成矿床和卡林型金矿的重要判据, 整体元素组合表明成矿流体以中性至弱碱性、低硫逸度环境为主。

高硫型矿床以 Cu 的极强富集为最显著特征 ($\log_2 EF > 5$), Ge、Bi、Te、Se 与 Ag 呈现中等至强烈增



a. 17 种元素分布箱线图; b. 综合火山图; c. 元素富集雷达图; 箱线图为对数坐标 (仅保留 > 0 数据); EF 为各矿床类型元素中位数与全样本中位数之比; 火山图 p 值采用曼-惠特尼 (Mann-whitey) U 检验 (双侧); 雷达图纬度为全局区分度较高的前 K 个元素

图2 不同成因类型各元素富集特征对比图

Fig. 2 Comparison of element enrichment characteristics of different genetic types

表 1 不同矿床类型元素富集特征一览表

Tab. 1 Summary of element enrichment characteristics of different deposit types

矿床类型	主要富集元素组合	次级富集元素	低富集/背景元素	关键地球化学指示意义
IOCG型	Te,Co,Ni,V,Se	Cu,Au,Bi,Sn	Fe,S	通常与铁氧化物-硫化物共生体系相关，可能反映深源或幔源流体参与的高温成矿环境(与前人研究结果总体一致)与海底火山喷流沉积过程密切相关，亲硫元素富集特征明显，元素组合通常表现出较好的稳定性 可能对应中性-弱碱性、相对低硫逸度流体环境，Au-Ag-Te组合较为典型，可与高硫型及卡林型进行区分 通常与高硫逸度、酸性岩浆流体体系有关，Cu富集程度较高，是重要判别特征之一与富砷黄铁矿中“不可见金”赋存密切相关，As-Au-Sb组合显著，可能反映低温还原性热液体系 元素富集整体较弱，更多表现为多元素组合特征，通常不以单一元素异常为主要判别依据 可能与变质流体控制的剪切带成矿体系有关，表现为中低温、挥发分参与的地球化学特征
VMS型	Zn,Pb,Cu,Ag,Sb,Bi	In,Sn,Mn,Tl	—	
低硫浅成低温热液型	Au,Ag,Te,Mo	Pb,Tl,Sb	As	
高硫浅成低温热液型	Cu	Ge,Bi,Te,Se,Ag	Fe,S	
卡林型	As,Au,Sb	Cu,W,Se	Zn,Pb	
斑岩型	Ge,Cr,V(弱)	Te,Se	Fe,S,Co	
造山型	Sb,Hg,Tl	Ni,W,Zn,As	Cu,Pb	

强,反映高硫逸度、强酸性岩浆流体环境。Cu 的异常富集是高硫型体系中硫化物(如黄铜矿、斑铜矿)大量沉淀的直接表现,而 Bi-Te-Se 的伴生特征则指示高温、富挥发分的岩浆-热液成矿条件。由于本研究地球化学数据均使用 LA-ICP-MS 以黄铁矿为分析对象获得,因此 S 与 Fe 接近背景水平,不影响该类矿床的判别特征。

斑岩型矿床整体表现为元素富集程度较弱,多数元素 EF 接近或略低于背景值,仅 Ge、Cr、V、Te 与 Se 显示轻微增强。这种特征反映了斑岩系统中金属元素以大规模、弥散式分布为主,而非单元素极端富集。Fe 与 S 接近背景水平,符合斑岩矿床以硫化物含量变化不显著、而依赖空间结构和多元素组合识别的成矿特征。因此,斑岩型矿床更适合通过多变量模式识别而非单一富集指标进行判别。

造山型金矿床以 Sb-Hg-Tl 的显著富集为主要特征,Ni、W、Zn 与 As 呈现轻至中等增强,反映了深部剪切带控制下的变质热液成矿过程。Sb、Hg 与 Tl 的

组合通常指示中低温、富含挥发分的变质流体环境,而 Ni 与 W 的出现则可能与深源物质输入或深部构造通道有关。As 的富集程度相对有限,有助于将其与 As 极为富集的卡林型金矿区分开来。

1.2 数据预处理

1.2.1 数据异常值及缺失值处理

笔者对原始数据中的零值与缺失值进行了机制划分。依据数据分布及多来源文献数据特征,将缺失划分为未测型缺失(NM)、低于检出限(BDL)及真实缺失(NaN)3 类。其中,NM 指某元素在全部样品中均未出现正值,反映其未被测定;BDL 指部分样品为 0 而其他样品为正值,0 值代表低于分析检出限;NaN 则表示数据记录缺失。针对不同缺失机制,笔者采取差异化处理策略:剔除 NM 变量;在剩余元素中基于总体缺失率、分矿床类型最大缺失率及零值比例构建综合缺失风险评分,筛选出 17 个低风险元素(Ni、Co、Sb、As、Pb、Ag、Cu、Zn、Se、Au、Bi、Te、Mo、Mn、Tl、Sn 和 Cd);对 BDL 数据采用最小正值一半进行替代,

对剩余 NaN 缺失采用基于随机森林的迭代插补方法 (MissForest-like) 进行补全。需指出, 该方法仅用于 BDL 处理后的 NaN, 而不适用于 NM 变量, 以避免结构性缺失引入偏差。

结果表明(表 2), MissForest-like 在分类性能与稳定性上优于 KNN (Balanced Accuracy $\approx$ 0.91, F1-macro $\approx$ 0.92), 因此作为主插补方案。缺失率阈值敏感性分析显示, 当阈值由 0.2 提高至 0.5 时模型性能持续提升,

而在 0.5 以上趋于稳定, 最终选取 0.5 作为推荐阈值。对照实验表明, 在不插补及完整变量子集条件下模型仍具有较好分类能力 (F1-macro $\approx$ 0.87), 而引入缺失指示变量虽显著提高精度 (F1-macro $>$ 0.98), 但可能放大结构性缺失信息, 因此仅作为敏感性分析。综上所述, 笔者构建了兼顾地球化学合理性与统计稳健性的缺失值处理流程, 为后续 CLR 变换及机器学习建模提供了可靠数据基础。

表 2 不同缺失值处理方法下随机森林模型分类性能对比

Tab. 2 Comparison of classification performance of random forest model under different missing value imputation methods

缺失值处理方法+分类模型	平衡准确率 (Balanced Accuracy)	平衡准确率标准差	宏平均 F1 值 (F1-macro)	F1 标准差
MissForest-like 插补+随机森林 (RF)	0.909 3	0.013 3	0.915 7	0.011 4
KNN 插补+随机森林 (RF)	0.894 6	0.013 0	0.903 1	0.010 1

1.2.2 数据变换处理

由于数据集数量级差异 (例如, Co 元素含量常达到数百 ppm, 而 Se、Te 等元素则通常仅为个位数 ppm 甚至更低), 因此为了消除数据差异带来的闭合效应, 消除对模型训练的主导效应, 有效的揭示各元素的指示意义 (Zhang et al., 2023)。

CLR 变换: 对每个样本中的元素浓度向量取自然对数; 计算该对数向量的几何均值 $g(\ln(x))$; 执行中心化处理 $CLR(x) = \ln(x) - g(\ln(x))$ 。变换后的每个元素值表示其相对于整个样本中元素的对数比例, 从而实现变量间的归一化和独立性增强。通过 CLR 变换后, 元素间的相关性强度显著下降。CLR 变换有助于将绝对浓度转化为相对组成信息, 减少闭合效应对相关结构和分类模型的干扰, 从而更适合开展多变量统计分析和监督学习建模 (Aitchison et al., 1982)。

1.2.3 相关性分析

基于黄铁矿微量元素地球化学大数据集, 对全体样本及不同矿床成因类型金矿床分别开展了斯皮尔曼 (Spearman) 等级相关分析, 以揭示元素间的单调相关结构及其在不同成矿体系中的差异 (图 3)。

在全体样本尺度上, 共获得 136 组元素对相关关系, 其相关系数绝对值中位数为 0.137, 90 分位值为 0.340, 整体表现为弱-中等相关为主, 表明多数元素之间不存在显著的单调依赖关系, 仅少数元素对呈现明显协同变化。其中, Ni-Co ($\rho=0.633$) 为最强正相关组合, 而 Sb-Se ($\rho=-0.580$) 为最显著负相关, 反映不同元素组合在区域尺度上可能受相似或差异成矿过程控制。

不同成因类型矿床内部的元素相关结构存在明

显差异, 体现出成矿环境与元素迁移机制的多样性。其中, IOCG 型中 Mo-Tl ($\rho=0.800$) 表现出最强正相关; VMS 型矿床以 Sb-Pb ($\rho=0.629$) 为主要相关特征; 低硫型中 Ni-Co ($\rho=0.714$) 相关性最强; 卡林型矿床中 Ni-Co ($\rho=0.864$) 表现出极强正相关, Co-Mo ($\rho=-0.734$) 呈显著负相关; 斑岩型矿床中 Ni-Co ($\rho=0.787$) 相关性突出; 造山型矿床中以 Sb-Tl ($\rho=0.739$) 最为显著; 高硫型中 Mo-Sn ($\rho=0.894$) 表现出最强相关性。上述差异表明, 不同矿床类型具有各自特征性的元素协同变化模式。总体而言, Spearman 相关分析表明, 不同成因类型矿床具有各自特征性的元素协同变化模式, 而全体样本中仅表现出有限的强相关元素对。这一结果说明, 矿床成因类型判别与成矿特征识别不能依赖单一或少数元素间的简单相关关系, 而需结合多元素联合特征开展多变量分析。相关结构分析结果为后续主成分分析、因子分析及监督分类模型提供了重要的统计依据与变量组合线索。

2 机器学习

2.1 支持向量机 (SVM)

支持向量机 (SVM) 通过构建最大间隔最优分离超平面并结合核函数实现非线性映射, 在小样本高维数据中具有良好的泛化能力与稳定性, 但对超参数敏感且在大规模数据下计算成本较高 (Guyon et al., 2003; Schober et al., 2018; Mahesh et al., 2020)。

2.2 随机森林 (RF)

随机森林 (RF) 通过自助采样 (Bagging) 构建多棵

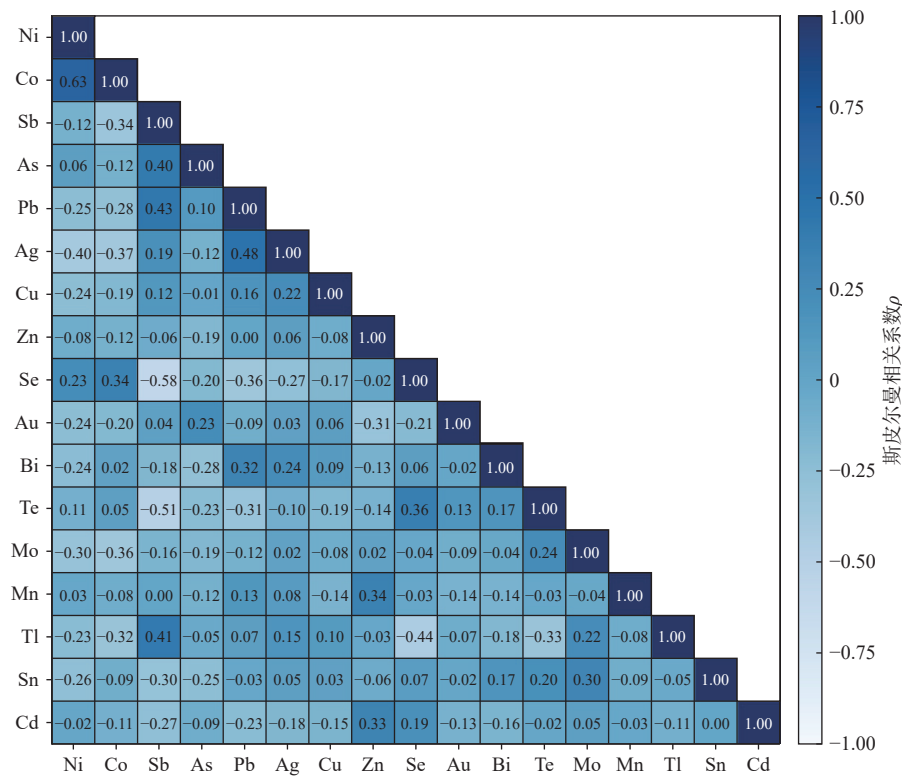


图3 黄铁矿数据集微量元素斯皮尔曼相关性分析图

Fig. 3 Spearman correlation analysis of trace elements in pyrite data

决策树并对其结果进行集成,在降低方差的同时提升模型的稳定性与泛化能力,从而实现高效可靠的分类与回归(Cortes et al., 1995; Cai et al., 2010)。

2.3 模型评估

笔者基于混淆矩阵及其衍生指标(准确率、精确率、召回率和 F1 分数)对模型分类性能进行综合评价,并通过十折交叉验证提高评估结果的稳健性(图 4),从而有效降低数据划分偏差并增强模型泛化能力(Breiman et al., 2001; Breiman et al., 2004; Powers et al., 2020; Nathwani et al., 2022)。

2.4 特征重要性分析

为克服 SVM 与 RF “黑箱”特性带来的可解释性不足,文中引入 SHAP 算法,从全局与局部层面量化各特征对模型预测的贡献,从而揭示模型判别结果背后的地球化学意义(Hawkins et al., 2003; Kotsiantis et al., 2006; Lundberg et al., 2017; Rudin et al., 2019; Pal et al., 2024)。

通过对所有样本 SHAP 值求平均,可获得特征的全局重要性排序;同时,也可针对特定矿床类型分析关键微量元素对预测结果的局部贡献。这一方法不仅提升了模型的可解释性,也为机器学习结果与成矿

地球化学机制之间的联系提供了定量支撑。

3 模型构建与结果

3.1 基于主成分分析(PCA)算法的判别图解

为降低变量维度并提取多元素协同变化结构,在 SVM 与 RF 建模之前,对 CLR 转换后的地球化学变量进行了主成分分析(PCA)。需要说明的是,PCA 所使用的数据与后续机器学习建模保持完全一致,均基于统一预处理流程所得数据,即在 NM、BDL 数据进行最小正值一半替代,并采用 MissForest-like 方法完成剩余 NaN 插补后,再进行 CLR 转换得到的最终数据集。在 PCA 过程中未进行额外插补,仅对变量进行 Z-score

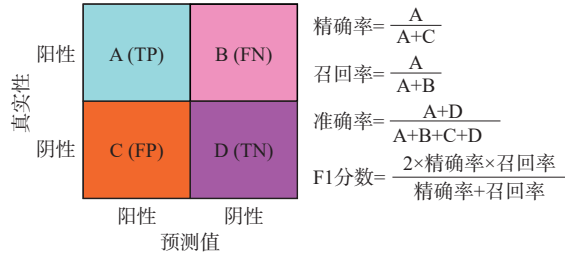


图4 模型评估指标系统(据王智宇等, 2025)

Fig. 4 Evaluation index system for model performance

标准化处理,以消除量纲差异对主成分提取的影响。

PCA 结果表明,前 10 个主成分累计解释了 86.02% 的总体方差。其中,PC1 解释方差为 18.16%,PC2 为 15.69%,前两个主成分累计解释 33.85% 的数据变异性;当前 7 个主成分时,累计解释率已达 73.42%,能够较好表征多元素地球化学结构特征。

从变量载荷贡献角度看,PC1 中 Sb、Se、Te、Tl 和 Pb 的贡献相对较高,同时 Sn、As、Cd、Cu 和 Ag 也表现出一定载荷,表明该主成分主要表征了重金属及硫亲和元素的协同变化特征(图 5a)。PC2 中 Ni 和 Co 的贡献最为突出,Ag、Pb、Mo、Sn 和 Bi 也表现出较高载荷,说明该主成分主要表征了过渡金属与部分亲硫元素之间的联合变化关系,可能反映不同矿床体系中元素赋存状态及流体—矿物相互作用的差异。

在主成分空间中,不同矿床类型样品表现出一定程度的分布差异(图 5b)。部分矿床类型在 PC1 或 PC2 方向上呈现一定分离趋势,但不同类型之间仍存在明显重叠,反映出不同成因类型既具有差异化的地球化学特征,也存在一定连续性与过渡性。这一结果表明,仅依赖低维主成分空间难以实现复杂多类型矿床的完全区分,也进一步说明引入监督学习方法进行判别分析的必要性。

总体而言,PCA 主要用于揭示黄铁矿微量元素之间的多变量协同变化结构及其潜在地球化学意义,而并非作为提升分类性能的必然降维步骤。其对后续矿床类型判别性能的影响,将在监督分类模型中通过对比分析进行系统评估。基于累计方差解释率与模型复杂度的综合考虑,后续机器学习建模中将采用 80%~95% 累计解释率对应的主成分数量作为 PCA 降维方案,并与使用原始标准化变量的模型结果进行对比,以评估 PCA 对分类性能的影响。

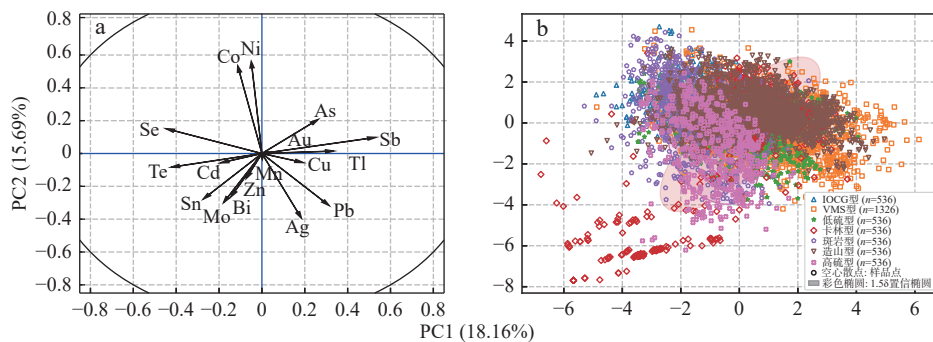
3.2 矿床成因判别模型构建

3.2.1 SVM 模型

基于前述 PCA 分析结果,采用 SVM 对多类型矿床地球化学数据进行了监督分类实验。模型分别基于 CLR 原始标准化特征(RAW, 17 维)和 PCA 分析降维特征(累计解释方差 ≥ 0.85 ,对应 11 个主成分,PCA11) CLR 构建。所有模型在统一的数据划分条件下进行训练与验证,其中训练集样本数为 5 384,测试集样本数为 1 347。考虑到不同矿床类型样本规模存在差异(536~1 541),最大类与最小类样本量比值为 3.61,数据表现为中等程度类别不平衡。因此,设置类别权重平衡策略(class weight='balanced'),并采用平衡准确率(balanced accuracy)与宏平均 F1(macro-F1)作为核心评价指标,以减弱类别分布差异对模型评价的影响。

SVM 采用径向基核函数(RBF),并通过 5 折交叉验证对惩罚参数 C 与核函数参数 γ 进行网格搜索优化(图 6)。交叉验证结果表明,基于 RAW 特征空间的 SVM 模型在整体分类性能与稳定性方面均优于 PCA11 特征方案。结果表明,基于 RAW 特征空间的 SVM 模型在分类性能与稳定性上均优于 PCA 降维方案。训练集 10 折交叉验证结果显示,SVM(RAW)模型的平衡准确率为 0.9212 ± 0.0126 ,准确率为 0.9220 ± 0.0123 ,宏平均 F1 为 0.9219 ± 0.0134 ;而 SVM(PCA11)对应指标分别为 0.8947 ± 0.0133 、 0.8908 ± 0.0122 和 0.8900 ± 0.0123 。在测试集上,SVM(RAW)取得准确率 0.9146、平衡准确率 0.9168、宏平均 F1=0.9143,明显优于 SVM(PCA11)(分别为 0.8656、0.8723、0.8639),说明原始多元素特征能够更充分保留判别信息。

误分类分析进一步揭示了模型判别行为的差异(图 7)。在 SVM(RAW)模型中,误分类主要集中于成矿过程具有过渡性或地球化学特征相近的矿床类型



a. PC1-PC2 得分图; b. PC1-PC2 载荷双标图

图5 不同矿床类型样品地球化学数据主成分分析结果

Fig. 5 Principal component analysis results of geochemical data of samples from different deposit types

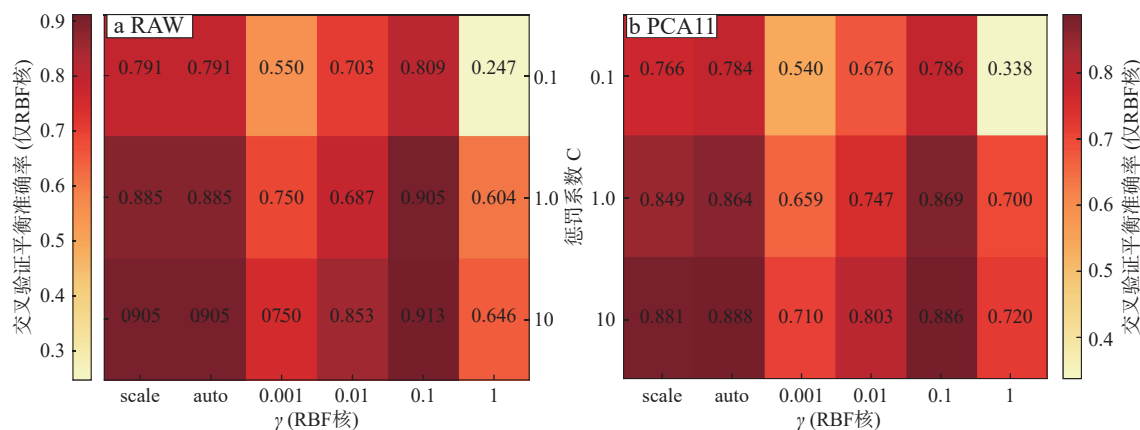


图6 SVM 网格热力搜索图(PC11-RAW)

Fig. 6 Grid heatmap search of SVM (PC11-RAW)

之间,其中低硫浅成低温热液型向造山型的误判最为突出(count=8, proportion=0.051),此外,IOCG 向 VMS、斑岩型及高硫型向斑岩型亦存在一定误判比例。总体来看,误分类主要发生于元素组合相似或流体演化连续的矿床类型之间,表明模型已能够有效刻画主要判别边界。

相比之下,SVM(PCA11)模型的误分类范围明显扩大,不仅低硫型与造山型之间的混淆进一步增强,还出现斑岩型-VMS 型、造山型-卡林型及 VMS 型-造山型等多类型交叉误判,且比例整体升高。这说明 PCA 降维在压缩维度的同时削弱了关键判别信息,使不同矿床类型在低维空间中的可分性下降。

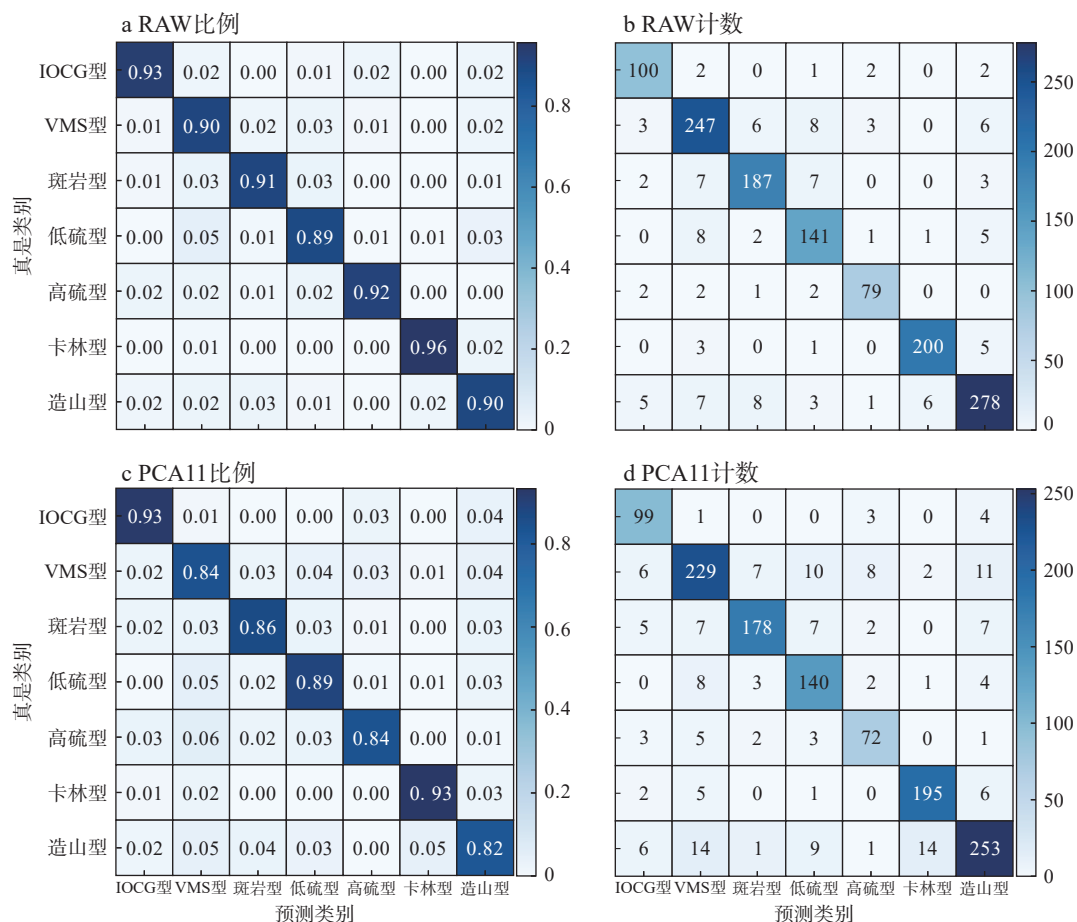


图7 黄铁矿微量元素 SVM 模型混淆矩阵

Fig. 7 Confusion matrix of SVM model for pyrite trace element data

基于30次分层随机重复实验结果, SVM(RAW)的平均平衡准确率与宏平均F1分别为 0.9188 ± 0.0078 和 0.9189 ± 0.0069 , 变异系数均小于1%, 表现出较高稳定性; 而PCA方案对应指标介于 $0.8596 \sim 0.9102$ 。Wilcoxon配对检验表明, RAW模型相对于PCA方案在主要指标上均达到极显著水平($p < 0.001$), 说明性能差异具有统计显著性。因此, SVM(RAW)被选定为后续模型对比分析的基准方案。

3.2.2 RF模型

在相同的数据划分条件下, 采用RF模型对多类型矿床地球化学数据进行监督分类实验。模型使用数据集、样本分布和类别顺序与SVM模型保持一致, 同时设置class_weight='balanced'以确保模型间结果的可比性。

在模型参数设置方面, RF通过10折交叉验证(以准确率为优化指标)对决策树数量进行优化, 最终确定最优树数; 同时固定最大树深为20(图8)。在该参数配置下, RF模型在测试集上取得准确率0.9087、平衡准确率0.8972及宏平均F1为0.9076; 训练集交叉验证结果为 0.9024 ± 0.0194 (balanced accuracy)、 0.9118 ± 0.0125 (accuracy)和 0.9113 ± 0.0166 (macro-F1)。总体来看, RF模型在多数矿床类型上具有较高识别能力, 但在地球化学特征重叠明显的类型之间仍存在一定混淆。

误分类分析表明(图9), RF模型的误判主要集中在低硫型浅成低温热液矿床、VMS型及造山型之间, 其中低硫型向VMS型(count=8, proportion=0.051)及向造山型的误判较为突出, 同时斑岩型与VMS型之

间亦存在明显混淆。这些误分类路径表明, 不同矿床类型在As-Sb-Au-Pb等元素组合上存在连续变化关系, 从而导致判别边界出现重叠。与SVM相比, RF模型的误分类分布更为分散, 说明其在复杂高维连续边界上的刻画能力略逊, 但在整体识别稳定性方面仍具有较好表现。

基于30次重复实验, RF(RAW)的平均平衡准确率为 0.9054 ± 0.0090 , 较SVM(RAW)(0.9188 ± 0.0078)下降约1.3个百分点。Wilcoxon检验结果($p = 1.86 \times 10^{-9}$)表明, 两者差异具有统计显著性, 说明SVM在本研究数据体系下具有更优判别能力, 而RF在模型解释性方面具有优势。

总体而言, RF模型在原始多元素地球化学特征空间中能够稳定识别主要矿床类型, 其分类性能与SVM模型处于同一量级。RF模型在处理复杂非线性关系和高维特征交互方面具有天然优势, 为后续结合

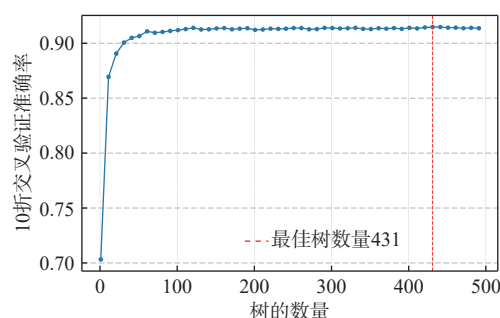


图8 网格搜索超参数 $n_estimators$ 的随机森林(RF)模型的交叉验证精度

Fig. 8 Cross-validation accuracy of random forest model with grid search hyperparameter $n_estimators$

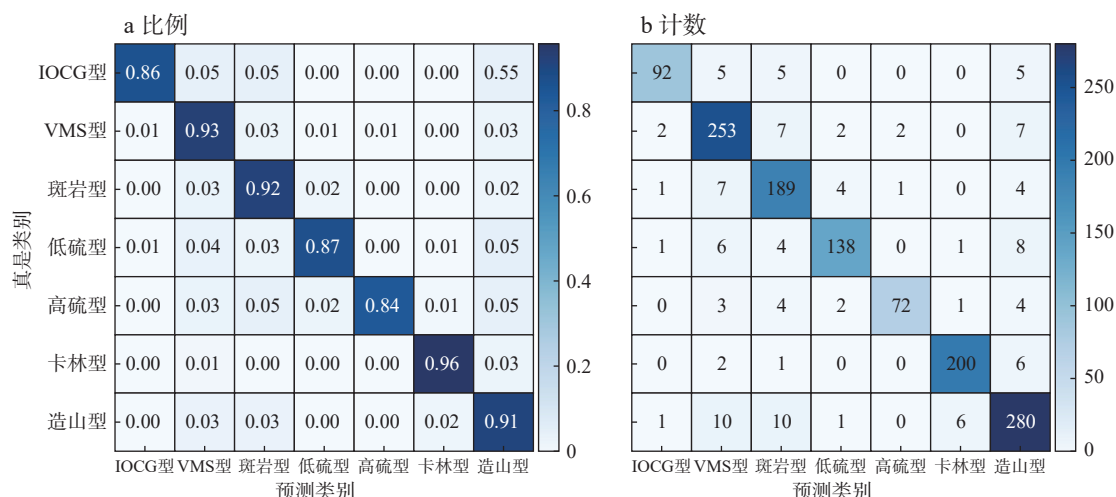


图9 黄铁矿微量元素 RF 模型混淆矩阵

Fig. 9 Confusion matrix of random forest model for pyrite trace element data

SHAP 等解释性分析揭示不同矿床类型的关键控矿元素提供了重要基础。

3.3 模型稳定性分析

结合 One-vs-Rest ROC 曲线(图 10)及 AUC 结果,可对模型稳定性与判别能力作出更加凝练的综合认识。整体而言,RF 模型在不同矿床类型上的表现具有显著分异性,反映出数据内在结构及成矿类型可分性的本质差异。

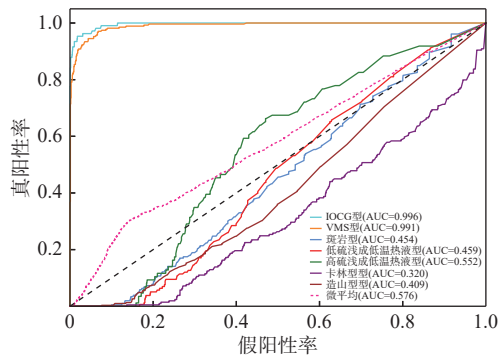


图10 RF模型对不同矿床类型的 ROC 曲线
(One-vs-Rest)

Fig. 10 ROC curves of random forest model for different deposit types (One-vs-Rest)

IOCG 型(AUC=0.996)与 VMS 型(AUC=0.991)具有近乎理想的判别能力,说明其在多元素空间中具有清晰边界;而高硫型(0.552)表现为中等水平,斑岩型(0.454)、低硫型(0.459)、造山型(0.409)及卡林型(0.320)整体较弱,反映出这些类型在地球化学空间中存在显著重叠。

微平均 AUC 为 0.576,表明模型整体性能受少数难分类型制约,呈现“优势类型主导、复杂类型拖累”的特征。这一结果说明,模型性能差异本质上受控于不同成矿体系之间的可分性,而非单一算法差异。进一步分析表明,在 PCA 累计解释率为 0.80~0.95 范围内,各 PCA 方案的分类性能均低于 RAW 模型(0.859 6~0.9102vs0.9188),最大差值约 5.9 个百分点,说明 PCA 对判别信息的削弱具有一致性。

需要指出的是,尽管笔者通过交叉验证、重复随机划分及统计检验增强了模型稳定性论证,但结果仍属于内部验证,尚不能完全替代独立外部数据集的泛化检验。综合分析,RF 模型在地球化学信号强、判别边界清晰的矿床类型中具有极高可靠性,但在多类型、强叠加成矿体系中,其稳定性与泛化能力明显依赖于不同矿床类型之间的本征可分性。这一结果进一步

印证:模型性能差异本质上源于地质过程复杂性,而非单一算法优劣。

3.4 特征重要性分析(SHAP)

3.4.1 全局特征重要性分析

基于 RF 模型的 SHAP 分析结果(图 11、图 12),可以从定量角度揭示不同微量元素在矿床类型判别中的相对贡献,并进一步反映其背后的成矿过程控制机制。从全局 SHAP 重要性来看,各元素对模型的贡献呈现出明显的层级结构。其中,Sb 为最核心的判别因子,其重要性显著高于其他元素,表明挥发性元素在不同成矿体系中的分配行为对矿床类型识别具有关键控制作用。其次为 Ag、As、Cu、Pb 及 Au 等元素,这一组合反映了贵金属-基底金属耦合体系的共同作用,而 Sn、Co、Te 等元素则作为辅助指示因子参与不同成矿过程的调控。该特征说明,模型本质上依赖的是多元素协同变化所构成的高维判别结构,而非单一元素的简单响应。

3.4.2 局部特征重要性分析

从矿床类型的 SHAP 分布特征(图 12)来看,不同成因类型矿床在 SHAP 空间中表现出相对稳定的元素组合特征:IOCG 型矿床,Sb、Pb 及 Ag 表现出较高的重要性,其中较高 Pb 与 Sb 值通常对应正向 SHAP 贡献,这一特征与铁氧化物-铜-金体系中重金属与挥发性元素协同富集的认识相一致。VMS 型矿床中,Ag、Pb 及 Zn 等元素贡献较为突出,高 Ag-Pb 组合对应较强正向 SHAP 值,可能反映喷流沉积环境中多金属同步沉淀的地球化学特征。对于斑岩型矿床,Cu、Mo、Pb 及 Sn 等元素具有较高贡献,其中 Cu 和 Mo 的高值通常

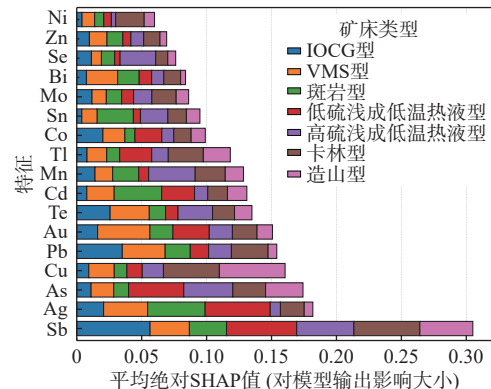


图11 随机森林(RF)模型对不同成因矿床黄铁矿微量元素数据的特征重要性分析

Fig. 11 Feature importance analysis of pyrite trace element data from different genetic types using random forest model

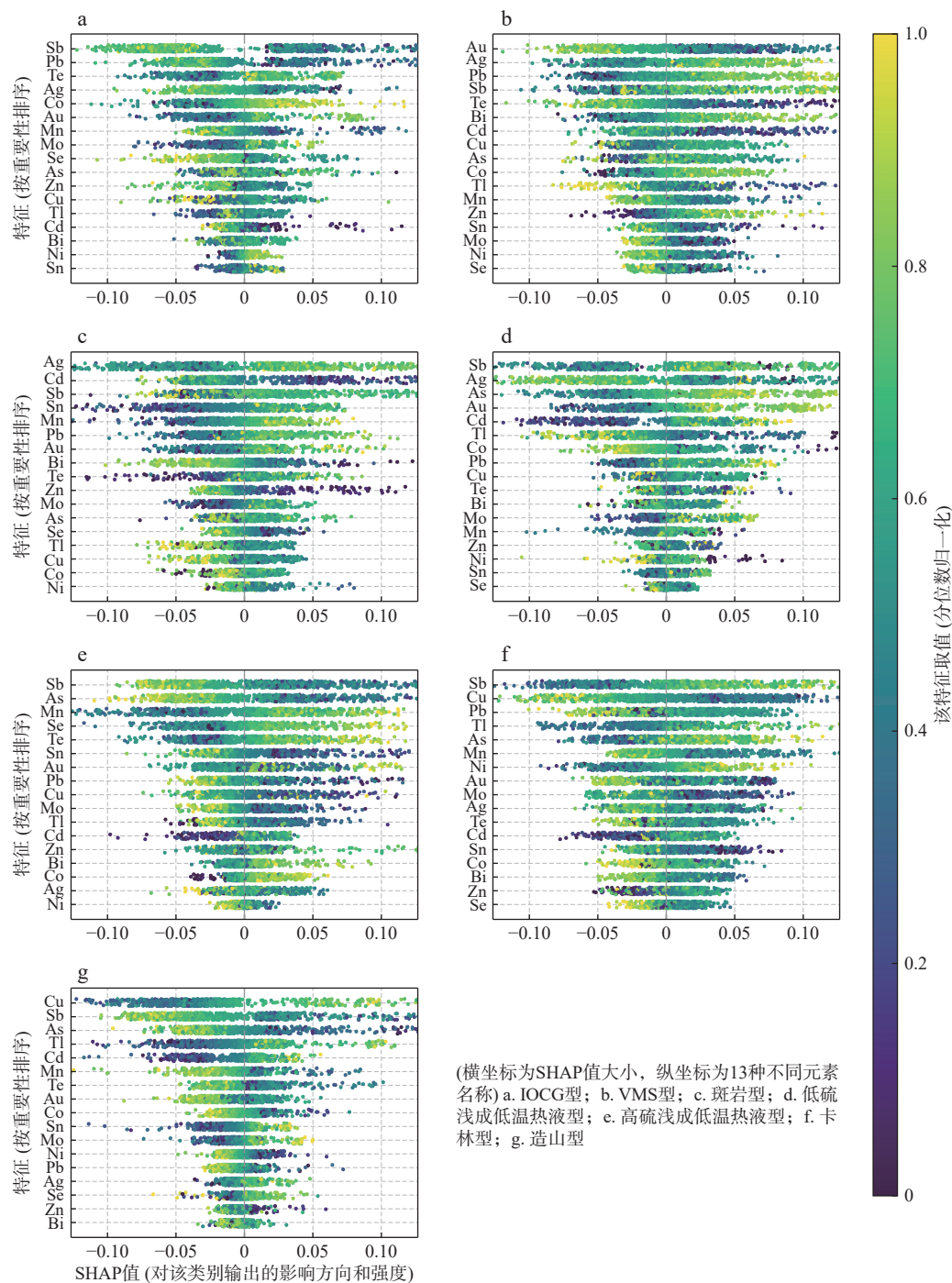


图12 不同成因类型矿床中元素特征重要性排序

Fig. 12 Ranking of element feature importance for different genetic deposit types

对应正向 SHAP 分布, 提示该类矿床中岩浆-热液分异过程对金属元素富集具有重要控制作用。在低硫型矿床中, As、Sb、Ag 及 Au 表现出较强贡献, 高 As-Sb-Ag 组合对应稳定的正向 SHAP 值, 与低温热液体系中挥发性元素与贵金属协同富集的典型特征相一致。相比之下, 高硫型矿床表现出更强的 As、Sb 及 Sn 控制, 高 As 与 Sb 值对模型输出具有持续正向影响, 部分基

性金属元素呈现双向分布, 可能反映流体-岩石反应及硫化作用强度差异对元素分配的影响。

在卡林型矿床中, As、Sb、Tl 及 Au 表现出较高重要性, 高 As-Sb-Tl 组合对模型输出具有显著贡献, 提示该类矿床可能受沉积建造与隐伏热液作用共同控制。造山型金矿床, As、Sb、Cu 及 Au 为主要判别因子, 其 SHAP 分布具有一定离散性, 表明构造控矿背景下多

阶段流体活动可能导致元素响应的复杂性与多样性。

总体而言, SHAP 结果表明 RF 模型在矿床类型判别过程中并非依赖单一元素, 而是通过多元素协同变化模式进行决策。不同矿床类型在 SHAP 空间中呈现出具有物质基础和成矿指向意义的差异化响应特征, 这不仅解释了 RF 模型较高的分类性能, 也为后续从地球化学角度理解不同矿床类型的判别机理提供了直观依据。

4 讨论

4.1 Sb 元素的地球化学特征

从模型解释结果来看, Sb 元素在本研究中表现出最突出的判别贡献, 其平均绝对 SHAP 值显著高于其他元素, 明显超过第二重要元素 Ag, 表明 Sb 元素是控制不同金矿成因类型判别的核心变量之一。该结果不仅体现了数据驱动模型的统计优势, 也具有明确的地球化学意义。已有研究表明, Sb 元素在不同矿床类型中具有显著分异特征, 尤其在卡林型及低温热液体系中富集, 而在斑岩型及岩浆主导体系中相对较低, 同时在多变量分析中表现出较高载荷, 显示出良好的稳定性与敏感性(Large et al., 2007; 邓依, 2024)。结合本研究 SHAP 分布特征, Sb 元素不仅具有最高的重要性, 还在不同矿床类型中呈现方向性差异, 表明其既控制分类边界, 也参与定义矿床的地球化学结构。

从成矿机制来看, Sb 元素属于典型低温亲硫元素, 其迁移主要以硫络合物形式进行, 并与 As、Au、Ag 等元素形成稳定共迁移体系, 构成典型 Au-As-Sb-Ag 组合(Reich et al., 2005; Martin et al., 2019), 广泛发育于卡林型及浅成低温热液矿床中, 指示还原性强、硫活度较高的成矿环境。相关研究表明, Sb 元素的沉淀通常发生于约 150~250 °C 的中低温条件, 并受流体混合、沸腾及 CO₂ 脱气等过程控制, 从而触发与 Au 等元素的协同沉淀(万卫, 2026)。在矿物尺度上, Sb 元素既可进入黄铁矿晶格, 也常沿微裂隙或边部富集, 并与 Au 元素形成复合矿物相(彭渤, 2007; Wang et al., 2024), 记录多阶段热液演化信息。

从矿床类型对比来看, Sb 元素对低温成矿体系具有显著指示意义: 卡林型及浅成低温热液型矿床表现为 Sb-As-Au 协同富集, 而斑岩型及高温岩浆体系中 Sb 含量较低, 反映温度控制下的系统性分异; 在 VMS 及 SEDEX 等沉积喷流体系中, Sb 元素常与 Zn-Pb-Cu

等元素共同富集, 指示海底热液多金属共沉淀过程。总体而言, Sb 元素在“成矿温度-流体性质-矿物赋存”多维耦合中的综合响应, 使其成为连接不同成矿过程的重要指示元素, 其在模型中的主导作用反映了低温热液体系在金矿类型判别中的关键控制地位。

4.2 元素组合对矿床类型的指示作用

综合 PCA、SVM 与 RF 模型结果表明, 不同矿床类型在黄铁矿微量元素空间中的可分性差异是控制模型性能与稳定性的根本因素。SVM(RAW)在原始 CLR 特征空间中表现出最高的分类精度与最小的交叉验证波动, 表明多元素之间存在显著的非线性组合关系; 而 PCA 降维后性能下降, 则说明弱异常信息及高阶非线性交互对矿床类型判别具有关键意义。这与前人认识一致, 即黄铁矿微量元素不仅受单一地球化学过程控制, 而是综合记录成矿流体来源、温度、氧逸度、硫逸度及流体混合、相分离等多因素耦合过程(Mukherjee et al., 2017; Li et al., 2018; Del Real et al., 2020; Liang et al., 2021; Baidya et al., 2021; Ehrig et al., 2023)。RF 的 SHAP 结果进一步表明, 模型判别依赖于 As、Sb、Cu、Sn、Pb、Co、Au 和 Te 等多元素协同变化模式, 说明“特征重要性”本质上对应不同成矿体系的地球化学指纹, 而非单一变量的统计贡献。

从不同矿床类型来看, IOCG 与 VMS 型矿床表现出最高的可分性及 ROC-AUC 值。IOCG 型矿床以 Co-As-Te 组合为核心, 反映高温、相对氧化的岩浆-铁氧化物热液体系, 其元素分布受流体来源转换与氧化还原条件控制(Li et al., 2018; Del Real et al., 2020; Liang et al., 2021; Ehrig et al., 2023); VMS 型矿床则以 Zn-Pb-Cu-Ag-Sb 组合为特征, 指示喷流沉积环境中多金属共沉淀及热液演化过程(Genna et al., 2015; Belousov et al., 2016; Basori et al., 2018; Dehnavi et al., 2018; Sun et al., 2023)。相比之下, 斑岩型与浅成低温热液体系呈现明显连续谱特征, 其中斑岩型矿床主要受 Cu-Mo-Sn 控制, 反映岩浆热液分异、相分离及流体混合作用(Tang et al., 2021; Wu et al., 2021; Keith et al., 2022; Sheng et al., 2022; Chen et al., 2024); 低硫型矿床以 As-Sb-Ag-Au 组合为主, 指示低温条件下贵金属沉淀与挥发性元素协同富集(Keith et al., 2020; Xu et al., 2020; Börner et al., 2021; Cheng et al., 2024; Tolstykh et al., 2025), 而高硫型矿床则表现为 As-Sb-Sn-Cu 组合, 反映岩浆挥发分主导的高硫化环境及斑岩型-浅成低温热液型过渡过程(King et al., 2014; Franchini et al.,

2015; Tanner et al., 2016; Zhao et al., 2020)。

卡林型与造山型矿床则表现出较低的 AUC 值与更强的特征重叠,揭示其成矿过程的复杂性与多阶段演化特征。卡林型矿床以 As-Sb-Tl-Au 组合为典型特征,反映低温隐伏体系中“不可见金”在富砷黄铁矿中的赋存机制及沉积地层-热液流体耦合过程(Liang et al., 2021; He et al., 2021; Li et al., 2021; Wei et al., 2022; Lin et al., 2023; Chen et al., 2025; Xiao et al., 2025);造山型矿床则表现为 As-Cu-Au 及部分 Co 的协同作用,指示变质流体主导、多期流体叠加及流体-岩石反应过程(George et al., 2018; Wu et al., 2019; Tarasova et al., 2020; Zheng et al., 2021; Li et al., 2022; Zhou et al., 2023; Mpaka et al., 2024; Wang et al., 2024)。因此,RF 在这些类型中的较低判别能力并非模型缺陷,而是反映了斑岩-浅成低温热液体系及卡林型-造山型之间客观存在的成矿连续性与叠加改造特征。

研究表明,基于黄铁矿微量元素的机器学习方法可构建“判别-解释-预测”一体化找矿框架。不同矿床类型对应的关键元素组合(如 IOCG 的 Co-As-Te、VMS 的 Sb-Pb-Zn-Cu、斑岩型的 Cu-Mo-Sn 及卡林型的 As-Sb-Tl-Au)可作为区域地球化学筛选与矿床类型快速识别的重要指标;SHAP 所揭示的元素贡献方向为构建向量找矿模型提供了定量依据。例如,利用多元素梯度约束喷流中心或岩浆热液中心位置(Genna et al., 2015; Dehnavi et al., 2018)。更为重要的是,模型识别出的“低可分性区域”往往对应成矿过渡带或多期叠加区,这为深部隐伏矿体预测提供了新的判据。

尽管本研究取得了较好的分类效果与地质解释一致性,但仍存在一定局限性。SHAP 结果反映的是统计贡献而非严格因果关系,其地质解释需结合矿物学与成矿理论综合分析。不同矿床类型样本数量及来源差异可能引入数据偏倚,从而影响模型对弱类的识别能力。此外,黄铁矿微量元素信号可能受到后期改造(如变质重结晶或流体再活化)的影响,从而增加类内离散性(George et al., 2018; Wu et al., 2019)。未来研究可通过引入多矿物组合(如黄铁矿+黄铜矿+闪锌矿)、同位素信息及空间结构数据,构建多模态融合模型,以进一步提高矿床类型判别精度与成矿过程解析能力。结合三维地质建模与机器学习方法,有望实现从“元素判别”向“成矿系统定量重建”的跨越。

5 结论

(1)构建的黄铁矿微量元素大数据集在样本规模、元素维度及矿床类型覆盖方面具有良好代表性。统计结果表明,Se 元素在不同成因类型金矿床中普遍富集且整体高于 Te 元素,显示出显著的成因普适性,该特征从大样本统计角度支持了 Se 元素可能以类质同象方式进入黄铁矿晶格的认识,并指示其在成矿流体中的稳定配位行为。

(2)通过系统的数据预处理(缺失值控制与 CLR 变换)和多变量分析表明,黄铁矿微量元素具有清晰的多元素协同变化结构,但 PCA 降维在一定程度上削弱了对矿床类型判别具有关键意义的弱异常与非线性组合信息。

(3)在多类型金矿床智能判别中,基于原始 CLR 特征空间的支持向量机在分类精度与交叉验证稳定性方面整体表现最优,而随机森林模型在保持较高判别能力的同时,通过 SHAP 分析提供了清晰的模型可解释性。

(4)SHAP 结果揭示了不同矿床类型所对应的关键微量元素组合特征(如 IOCG 的 Co-As-Te、斑岩型的 Cu-Mo-Sn、卡林型的 As-Sb-Tl-Au 等),这些判别特征与经典成矿模型高度一致,表明将高性能机器学习模型与解释性分析相结合,是复杂成矿体系中金矿床类型识别与成因理解的有效技术路径。

致谢:本研究得到中国地质大学(北京)王智宇硕士、马虎超博士、王贵龙博士等人的帮助,得到中国地质调查局呼和浩特自然资源综合调查中心矿产资源勘查室李成元高级工程师的大力支持,谨此致谢。感谢审稿专家提出的建设性修改意见!

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):

邓依,张静,钟日晨,等.基于机器学习的主成分分析方法在金矿类型判别中的应用:以黄铁矿元素地球化学特征为例[J].岩石学报,2024,40(6):1801-1816.

- DENG Yi, ZHANG Jing, ZHONG Richen, et al. Application of principal component analysis method based on machine learning to gold ddepoist type discrimination: A case study of the geochemical characteristics of pyrite[J]. *Acta Petrologica Sinica*, 2024, 40(6): 1801–1816.
- 高振敏, 杨竹森, 李红阳, 等. 黄铁矿载金的原因和特征[J]. *高校地质学报*, 2000, 6(2): 156–162.
- GAO Zhenmin, YANG Zhusen, LI Hongyang, et al. Genesis and characteristics of gold hosted by pyrite[J]. *Geological Journal of China Universities*, 2000, 6(2): 156–162.
- 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 MachineLearning 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.
- 黄少芳, 刘晓鸿. 地质大数据应用与地质信息化发展的思考[J]. *中国矿业*, 2016, 25(8): 166–170.
- HUANG Shaofang, LIU Xiaohong. Thinking about the application of geological big data and geological information development[J]. *China Mining Magazine*, 2016, 25(8): 166–170.
- 李洪梁, 李光明. 不同类型热液金矿床主成矿期黄铁矿成分标型特征[J]. *地学前缘*, 2019, 26(6): 202–210.
- LI Hongliang, LI Guangming. Compositional characteristics of pyrite ore formed in the main metallogenic period of various types of hydrothermal gold deposits[J]. *Earth Science Frontiers*, 2019, 26(6): 202–210.
- 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 pophryry-skarn deposits from the Qimantagh, East Kunlun: based on machine learning and whole-rock compositions[J]. *Northwestern Geology*, 2023, 56(6): 41–56.
- 彭渤, Adam Piestrzynski, 刘升友, 等. 湘西沃溪钨锡金矿床黄铁矿中发现 Au-Sb 矿物相[J]. *地质通报*, 2007, 26(5): 553–559.
- PENG Bo, PIESTRZYNSKI Adam, LIU Shengyou, et al. Discovery of Au-Sb mineral phase in pyrite from the Woxi W-Sb-Au deposit, western Hunan, China[J]. *Geological Bulletin of China*, 2007, 26(5): 553–559.
- 王智宇, 王达, 邱昆峰, 等. 机器学习在矿物岩石地球化学大数据挖掘中的应用与展望[J]. *成都理工大学学报(自然科学版)*, 2025, 52(5): 844–858.
- WANG Zhiyu, WANG Da, QIU Kunfeng, et al. Application and prospect of machine learning in geochemical big data mining of minerals and rocks[J]. *Journal of Chengdu University of Technology (Science & Technology Edition)*, 2025, 52(5): 844–858.
- 万卫, 范会虎, 徐广明, 等. 原生晕及深穿透地气测量对金矿勘查的指示——以荒漠草原覆盖区内蒙古毕力赫金矿为例[J/OL]. *地球学报*, 2026: 1–13. <https://link.cnki.net/urlid/11.3474.P.20260330.1025.002>.
- WAN Wei, FAN Huihu, XU Guangming, et al. Indications of primary halos and deep-penetrating geogas measurement for gold exploration: A case study of the Bilihe gold deposit in the desert steppe-covered area, Inner Mongolia[J/OL]. *Acta Geoscientia Sinica*, 2036: 1–13. <https://link.cnki.net/urlid/11.3474.P.20260330.1025.002>.
- 徐国风, 邵洁涟. 黄铁矿的标型特征及其实际意义[J]. *地质与勘探*, 1979: 541–546.
- XU Guofeng, SHAO Jielian. Typomorphic characteristics of pyrite and their practical significance[J]. *Geology and Exploration*, 1979: 541–546.
- 严育通, 李胜荣, 贾宝剑, 等. 中国不同成因类型金矿床的黄铁矿成分标型特征[J]. *地学前缘*, 2012, 19(4): 214–226.
- YAN Yutong, LI Shengrong, JIA Baojian, et al. Composition typomorphic characteristics and statistics analysis of metallogenic pyrite in gold deposits of different genetic types[J]. *Earth Science Frontiers*, 2012, 19(4): 214–226.
- 翟明国, 杨树锋, 陈宁华, 等. 大数据时代: 地质学的挑战与机遇[J]. *中国科学院院刊*, 2018, 33(8): 825–831.
- ZHAI Mingguo, YANG Shufeng, CHEN Ninghua, et al. Big data epoch: challenges and opportunities for geology[J]. *Bulletin of Chinese Academy of Sciences*, 2018, 33(8): 825–831.
- 张茂省, 郭柱国, 闫雨藤, 等. 基于大语言模型(LLMs)的地质灾害风险防控体系建设[J]. *中国地质*, 2026, 53(1): 136–157.
- ZHANG Maosheng, GUO Zhuguo, YAN Yuteng, et al. Construction of the geological hazard risk mitigation and management system powered by Large Language Models (LLMs)[J]. *Geology in China*, 2026, 53(1): 136–157.
- 赵东博. 矿产地质勘查理论及技术方法的研究[J]. *世界有色金属*, 2021(22): 110–111.
- ZHAO Dongbo. Study on the theory and technical methods of mineral geological exploration[J]. *World Nonferrous Metals*, 2021(22): 110–111.
- 赵洋, 鞠美庭, 沈镭. 我国矿产资源安全现状及对策[J]. *资源与*

- 产业, 2011, 13(6): 79–83.
- ZHAO Yang, JU Meiting, SHEN Lei. Situation and approaches to China's ore resources security[J]. Resources Industries, 2011, 13(6): 79–83.
- 郑超杰. 基于成分数据及机器学习在阿舍勒地区的综合找矿研究[D]. 昆明: 桂林理工大学, 2021.
- ZHENG Chaojie. Research of comprehensive prospecting prediction based on compositional data analysis and machine learning model in Ashele copper-zinc deposit, Xinjiang, NW China[D]. Kunming: Guilin University of Technology, 2021.
- 左仁广, 王健, 熊义辉, 等. 2011–2020 年勘查地球化学数据处理研究进展[J]. 矿物岩石地球化学通报, 2021, 40(1): 81–93.
- ZUO Renguang, WANG Jian, XIONG Yihui, et al. Progresses of researches on geochemical exploration data processing during 2011–2020[J]. Bulletin of Mineralogy, Petrology and Geochemistry, 2021, 40(1): 81–93.
- 朱昊磊, 杨兴科, 何虎军, 等. 基于卷积神经网络和黄铁矿大数据判别金矿类型[J]. 地质学报, 2023, 97(10): 3396–3409.
- ZHU Haolei, YANG Xingke, HE Hujun, et al. Discrimination of gold deposit types based on convolutional neural network and pyrite big data[J]. Acta Geologica Sinica, 2023, 97(10): 3396–3409.
- Aitchison J. The statistical analysis of compositional data[J]. Journal of the Royal Statistical Society: Series B, 1982, 44(2): 139–160.
- Arehart G B. Characteristics and origin of sediment-hosted disseminated gold deposits: A review[J]. Ore Geology Reviews, 1996, 11(6): 383–403.
- Breiman L. Random forests[J]. Machine Learning, 2001, 45: 5–32.
- Breiman L, Cutler A. RFtools—For predicting and understanding data[A]. Interface Workshop[C]. 2004, 1–62.
- Baidya A S, Sen A, Pal, et al. Ore-forming processes in the Khetri Copper Belt, western India: Constraints from trace element chemistry of pyrite and C-O isotope composition of carbonates[J]. Mineralium Deposita, 2021, 56(5): 957–974.
- Belousov I, Large R R, et al. Pyrite compositions from VHMS and orogenic Au deposits in the Yilgarn Craton, Western Australia: Implications for gold and copper exploration (Review)[J]. Ore Geology Reviews, 2016, 79: 474–499.
- Basori M B I, Gilbert, et al. Textures and trace element composition of pyrite from the Bukit Botol volcanic-hosted massive sulphide deposit, Peninsular Malaysia[J]. Journal of Asian Earth Sciences, 2018, 158: 173–185.
- Börner F, et al. Fingerprinting fluid evolution by trace elements in epithermal pyrite, Vatukoula Au-Te deposit, Fiji[J]. Ore Geology Reviews, 2021, 137: 104314.
- Cai S M, Zhang, et al. A method of salt-affected soil information extraction based on a support vector machine with texture features[J]. Mathematical and Computer Modelling, 2010, 51(11): 1319–1325.
- Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20: 273–297.
- Cheng X, Ling M X, Liu P H, et al. Multistage sulfide and hydrothermal quartz as a record of precision ore-forming process in the epithermal Cu–Zn–Au system in the Kalatag arc, East Tianshan, NW China[J]. Ore Geology Reviews, 2024, 166: 105917.
- Chen K, Shao Y J, Liu Z F, et al. Re–Os isotopes and geochemistry of pyrite from the Dongguashan Cu(-Au) deposit, Eastern China: Implications for porphyry-skarn-stratabound mineralization systems[J]. Ore Geology Reviews, 2024, 164: 105827.
- Chen B, Zuo Y J, Zheng L L, et al. Relationship between silicification and gold mineralization and evolution of hydrothermal systems in Nibao Carlin-type gold deposits, SW China: Insights from rock geochemistry and trace elements of quartz and pyrite[J]. Ore Geology Reviews, 2025, 176: 106394.
- Del real I, Thompson J F H, Simon A C, et al. Geochemical and isotopic signature of pyrite as a proxy for fluid source and evolution in the Candelaria-Punta del Cobre iron oxide copper-gold district, Chile[J]. Economic Geology, 2020, 115(7): 1493–1518.
- Dehnavi A S, Mcfarlane C R M, Lentz D R, et al. Assessment of pyrite composition by LA-ICP-MS techniques from massive sulfide deposits of the Bathurst Mining Camp, Canada: From textural and chemical evolution to its application as a vectoring tool for the exploration of VMS deposits[J]. Ore Geology Reviews, 2018, 92: 656–671.
- Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare[J]. Nature Medicine, 2019, 25: 24–29.
- Emsbo P. Gold in SEDEX deposits[A]. Simmons S F, Graham I J, et al. Reviews in Economic Geology (Vol. 13)[C]. Society of Economic Geologists, 2000: 427–437.
- Ehrig K, Ciobanu C L, Verdugo-ihl M R, et al. Lifting the cloak of invisibility: Gold in pyrite from the Olympic Dam Cu-U-Au-Ag deposit, South Australia[J]. American Mineralogist, 2023, 108(2): 259–276.
- Franchini M, Mcfarlane, et al. Trace metals in pyrite and marcasite from the Agua Rica porphyry-high sulfidation epithermal deposit, Catamarca, Argentina: Textural features and metal zoning at the porphyry to epithermal transition[J]. Ore Geology Reviews, 2015, 66: 366–387.
- Gregory D D, Cracknell M J, Large R R, et al. Distinguishing ore deposit type and barren sedimentary pyrite using LA-ICP-MS trace element data and statistical analysis[J]. Economic Geology, 2019, 114: 771–786.
- Guyon I, Elisseeff A. An introduction to variable and feature selection[J]. Journal of Machine Learning Research, 2003, 3: 1157–1182.

- George L L, Biagioni, 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.
- Genna D, Gaboury D, et al. 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[J]. *Economic Geology*, 2015, 110(8): 2087–2108.
- Hawkins D M, Basak S C, Mills D. Assessing model fit by cross-validation[J]. *Journal of Chemical Information and Computer Sciences*, 2003, 43(2): 579–586.
- Haydari A, Yılmaz Y. Deep reinforcement learning for intelligent transportation systems: A survey[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 23(1): 11–32.
- He X H, Su, et al. In situ multiple sulfur isotopes and chemistry of pyrite support a sedimentary source-rock model for the Linwang Carlin-type gold deposit in the Youjiang basin, southwest China[J]. *Ore Geology Reviews*, 2021, 139: 104533.
- Keith M, Mcfarlane, Haase K M, et al. Phase separation and fluid mixing revealed by trace element signatures in pyrite from porphyry systems[J]. *Geochimica et Cosmochimica Acta*, 2022, 329: 185–205.
- Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets: A review[J]. *GESTS International Transactions on Computer Science and Engineering*, 2006, 30(1): 25–36.
- Keith M, Smith, et al. Pyrite chemistry: A new window into Au-Te ore-forming processes in alkaline epithermal districts, Cripple Creek, Colorado[J]. *Geochimica et Cosmochimica Acta*, 2020, 274: 172–191.
- King J, Williams-jones, et al. High-sulfidation epithermal pyrite-hosted Au (Ag-Cu) ore formation by condensed magmatic vapors on Sangihe Island, Indonesia[J]. *Economic Geology*, 2014, 109(6): 1705–1733.
- Leach D L, Sangster D F, et al. Sediment-hosted lead-zinc deposits: A global perspective[J]. *Economic Geology*, 2005, 100: 561–607.
- Lundberg S M, Lee S I. A unified approach to interpreting model predictions[J/OL]. *arXiv preprint*, 2017: arXiv: 1705.07874.
- Liang P, Chen, et al. In-situ element geochemical and sulfur isotope signature of pyrite and chalcopyrite: Constraints on ore-forming processes of the Laoshankou iron oxide-copper (-gold) deposit, northern East Junggar[J]. *Ore Geology Reviews*, 2021, 139: 104510.
- Li R, Chen H, Xia X, et al. Using integrated in-situ sulfide trace element geochemistry and sulfur isotopes to trace ore-forming fluids: Example from the Mina Justa IOCG deposit (southern Peru)[J]. *Ore Geology Reviews*, 2018, 101: 165–179.
- Lin S R, Hu K, Cao J, et al. Coupling and decoupling of Au and As in pyrite from Carlin-type Au deposits, southwest China[J]. *Journal of Asian Earth Sciences*, 2023, 246: 105582.
- Liang Q L, Xie Z J, Song X Y, et al. Evolution of invisible Au in arsenian pyrite in Carlin-type Au deposits[J]. *Economic Geology*, 2021, 116(2): 515–526.
- Li X H, Bai, et al. Mineralization processes involved in the formation of the Jinya Carlin-type Au deposit, northwestern Guangxi, China: Evidence from in situ trace element and S isotope geochemistry of Au-bearing zoned pyrite[J]. *Ore Geology Reviews*, 2021, 138: 104376.
- Li B, Zhu L M, Xiong X, et al. Genesis of the Longtougou gold deposit, South Qinling Orogen, China: Constraints from ore geology, trace element, and S–Pb–H–O isotopes[J]. *Ore Geology Reviews*, 2022, 149: 105104.
- Large R, Maslennikov V, Robert F, et al. Multistage sedimentary and metamorphic origin of pyrite and gold in the giant Sukhoi Log deposit, Lena gold province, Russia[J]. *Economic Geology*, 2007, 102(7): 1233–1267.
- Merghadi A, Yunus A P, Dou J, et al. Machine learning methods for landslide susceptibility studies[J]. *Earth-Science Reviews*, 2020, 207: 103225.
- Mahesh B. Machine learning algorithms-a review[J/OL]. *International Journal of Science and Research (IJSR)*, 2020, 9(1): 381–386.
- Mukherjee I, Large R R, et al. Application of pyrite trace element chemistry to exploration for SEDEX style Zn-Pb deposits: McArthur Basin, Northern Territory, Australia[J]. *Ore Geology Reviews*, 2017, 81(4): 1249–1270.
- Mpaka Y W, Von der heyden B P, Glynn S, et al. Multistage evolution of gold mineralization in the Kibali gold district: Insights from pyrite analyses[J]. *Journal of African Earth Sciences*, 2024, 214: 105244.
- Ngai E W T, Hu Y, Wong Y H, et al. The application of data mining techniques in financial fraud detection[J]. *Decision Support Systems*, 2011, 50(3): 559–569.
- Martin Andrew J. Trace element systematics and ore-forming processes in mafic VMS deposits: Evidence from the Troodos ophiolite, Cyprus[J]. *Ore Geology Reviews*, 2019, 106: 205–225.
- Nathwani C L, Wilkinson J J, Fry G, et al. Machine learning for geochemical exploration[J]. *Mineralium Deposita*, 2022, 57: 1143–1166.
- Powers D M W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation[J/OL]. *arXiv preprint*, 2020: arXiv: 2010.16061.
- Pal M. Fuzzy matrices[A]. In: *Recent Developments of Fuzzy Matrix Theory and Applications*[C]. Cham: Springer Nature Switzerland, 2024: 1–63.

- Rudin C. Stop explaining black box models[J]. *Nature Machine Intelligence*, 2019, 1: 206–215.
- Reich Martin. Solubility of gold in arsenian pyrite[J]. *Geochimica et Cosmochimica Acta*, 2005, 69(11): 2781–2796.
- Steadman J A, Large R R, et al. Pyrite trace element behavior in magmatic-hydrothermal environments: An LA-ICPMS imaging study[J]. *Ore Geology Reviews*, 2021, 128: 103878.
- Schober P, Boer C, Schwarte L A. Correlation coefficients: Appropriate use and interpretation[J]. *Anesthesia & Analgesia*, 2018, 126(5): 1763–1768.
- Sun B, Lv X, Wang S, et al. Evolution of an ancient VMS ore-forming system recorded by pyrite and sphalerite mineral texture, trace elements, and sulfur isotope: A case study from the Huangtupo Cu-Zn (-Au) deposit, Eastern Tianshan, NW China[J]. *Ore Geology Reviews*, 2023, 158: 105475.
- Sheng Y M, Tang, et al. Distal gold mineralization associated with porphyry system: The case of Hongzhuang and Yuanling deposits, East Qinling, China[J]. *Ore Geology Reviews*, 2022, 142: 104701.
- Tang L, Zhao Y, Zhang S T, et al. Origin and evolution of a porphyry-breccia system: Evidence from zircon U-Pb, molybdenite Re-Os geochronology, in situ sulfur isotope and trace elements of the Qiyugou deposit, China[J]. *Gondwana Research*, 2021, 89: 88–106.
- Tolstykh N, Bortnikov N, Zhukova I, et al. Trace elements in pyrite from AuAg epithermal deposits of Kamchatka, Russia: Comparison with geochemical features of mineral systems[J]. *Journal of Geochemical Exploration*, 2025, 275: 107774.
- Tanner D, Henley R W, Mavrogenes J A, et al. Sulfur isotope and trace element systematics of zoned pyrite crystals from the El Indio Au-Cu-Ag deposit, Chile[J]. *Contributions to Mineralogy and Petrology*, 2016, 171(4): 1–17.
- Tarasova Y I, Budyak, et al. Mineralogical and isotope-geochemical ($\delta^{13}\text{C}$, $\delta^{34}\text{S}$ and Pb-Pb) characteristics of the Krasniy gold mine (Baikal-Patom Highlands): Constraining ore-forming mechanisms and the model for Sukhoi Log-type deposits[J]. *Ore Geology Reviews*, 2020, 119: 103365.
- Wang Xinfu. Characteristics of antimony mineralization in the Yangla polymetallic deposit, northwestern Yunnan, SW China: Insights from calcite Sm-Nd dating and C-O-Sr isotopes[J]. *Ore Geology Reviews*, 2024, 173: 106266.
- Wu Q, Zhao, et al. Origin of the Qiyugou gold deposit in the southern margin of the North China Craton: Insights from trace elements of pyrite and mineralogy of Bi-minerals[J]. *Ore Geology Reviews*, 2021, 133: 104085.
- Wei D T, Zhou T F, Xia Y, et al. Pyrite textural, trace elemental and sulfur isotope signatures of the Badu gold deposit, Youjiang basin (SW China): Implications for ore-fluid source and Au precipitation mechanism[J]. *Ore Geology Reviews*, 2022, 149: 105083.
- Wang Z L, Zhang X B, Cui Z L, et al. The Carboniferous Wuzunbulake orogenic gold deposit in South Tianshan Orogen (NW China): Sericite Rb-Sr geochronology, pyrite geochemistry, and metallogeny[J]. *Journal of Geochemical Exploration*, 2024, 267: 107576.
- Wu Y F, et al. Metal remobilization and ore-fluid perturbation during episodic replacement of auriferous pyrite from an epizonal orogenic gold deposit[J]. *Geochimica et Cosmochimica Acta*, 2019, 245: 98–117.
- Xu N, Wu C L, Li S R, et al. LA-ICP-MS in situ analyses of the pyrites in Dongyang gold deposit, Southeast China: Implications to the gold mineralization[J]. *China Geology*, 2020, 3(2): 230–246.
- Xiao Jingdan, Xie Zhuojun, Xia Yong, et al. Consistent crystal orientation of core and rim pyrites indicates an epitaxial growth of rim in Carlin-type gold deposits[J]. *Earth Science Frontiers (English Edition)*, 2025, 16(1): 223–234.
- Zhang Z C, Long L L, Huang B, et al. The ore-forming process of the Qingshan gold deposit, Qaidam Basin: Evidence from LA-ICP-MS trace element analysis and S-Pb isotopic compositions[J]. *Ore Geology Reviews*, 2023, 162: 105668.
- Zhao X Y, Zhong, et al. Molybdenite Re-Os dating and LA-ICP-MS trace element study of sulfide minerals from the Zijinshan high-sulfidation epithermal Cu-Au deposit, Fujian Province, China[J]. *Ore Geology Reviews*, 2020, 118: 103363.
- Zheng Y, Wu, et al. Gold accumulation in the metavolcanic-hosted orogenic gold deposit constrained by pyrite paragenesis coupled with in-situ trace elements and sulfur isotope: The Sarekuobu example in the Chinese Altay Orogen[J]. *Ore Geology Reviews*, 2021, 138: 104387.
- Zhou Z J, Chen Z L, Weyer S, et al. Metal source and ore precipitation mechanism of the Ashawayi orogenic gold deposit, southwestern Tianshan Orogen, western China: Constraints from textures and trace elements in pyrite[J]. *Ore Geology Reviews*, 2023, 157: 105452.