https://doi.org/10.3799/dqkx.2026.084

基于岩浆锆石微量元素的机器学习模型对斑岩铜矿成矿潜力的评估

吴川川[1,2],孟旭阳[1,2*],刘 品[3*],辛茹月[3],成嘉伟[1,2]

1. 中国地质大学(北京)地球科学与资源学院,自然资源部战略性金属矿产找矿理论与技术重点实验室,北京 100083
2. 中国地质大学(北京)地球科学与资源学院,地质过程与成矿预测全国重点实验室,北京 100083
3. 中国地质大学(北京)人工智能学院,北京 100083

摘 要: 针对传统二元判别图在处理锆石微量元素数据时维度低、难以揭示非线性成矿机制的问题,本研究引入机器学习方法以提升斑岩型铜矿成矿潜力判别的准确性与客观性.基于全球斑岩型矿床锆石微量元素数据,系统对比了XGBoost、随机森林、支持向量机、逻辑回归和神经网络五种模型的分类性能,并结合SHAP可解释性框架解析模型决策机制.结果表明:XGBoost模型分类准确率最高(86.85%),其次为随机森林(83.84%)和逻辑回归(83.84%),支持向量机(78.90%)和多层感知器(69.86%)表现相对较低.SHAP分析显示,Eu_N/Eu_N^*、10 000(Eu_N/Eu_N^*)/Y、Ti、Y、△FMQ和Gd/Dy是判别成矿潜力的关键指标.本研究构建的XGBoost、随机森林等模型能有效识别斑岩型铜矿成矿潜力并筛选关键勘查标识,不仅为智能化矿产勘查提供了新思路,也为实际应用中的模型选择和优化提供了重要参考.

关键词: 斑岩型铜矿;机器学习;锆石微量元素;大数据.

中图分类号: P612 **文章编号:** 1000-2383(2026)09-3530-17 **收稿日期:** 2025-12-15

Application of Machine Learning Models Based on Magmatic Zircon Trace Elements for Discriminating the Metallogenic Potential of Porphyry Copper Deposits

Wu Chuanchuan[1,2], Meng Xuyang[1,2*], Liu Pin[3*], Xin Ruyue[3], Cheng Jiawei[1,2]

1. *Ministry of Natural Resources Key Laboratory for Exploration Theory and Technology of Critical Mineral Resources, School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing* 100083, *China.*
2. *State Key Laboratory of Geological Processes and Mineral Resources, School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing* 100083, *China*
3. *School of Artificial Intelligence, China University of Geosciences (Beijing), Beijing* 100083, *China*

Abstract: To address the limitations of traditional binary discrimination diagrams, which arise from low dimensionality and difficulty in revealing nonlinear metallogenic mechanisms when processing zircon trace element data, this study introduces machine learning methods to enhance the accuracy and objectivity of porphyry copper mineralization potential assessment.

基金项目: 国家深地重大专项(No.2024ZD1001401);国家自然科学基金项目(No.42572091).

作者简介: 吴川川(2001-),男,博士研究生,地质资源与地质工程专业.ORCID: 0009-0006-7692-0665. E-mail: 3001250147@email.cugb.edu.cn

* **通讯作者:** 孟旭阳,E-mail: mengxuyang1991@163.com;刘品,E-mail: liupin@cugb.edu.cn

引用格式: 吴川川,孟旭阳,刘品,辛茹月,成嘉伟,2026.基于岩浆锆石微量元素的机器学习模型对斑岩铜矿成矿潜力的评估.地球科学,51(9):3530-3546.

Citation: Wu Chuanchuan, Meng Xuyang, Liu Pin, Xin Ruyue, Cheng Jiawei, 2026. Application of Machine Learning Models Based on Magmatic Zircon Trace Elements for Discriminating the Metallogenic Potential of Porphyry Copper Deposits. *Earth Science*, 51(9): 3530-3546.

Based on global zircon trace element data from porphyry deposits, we systematically compared the classification performance of five models - XGBoost, Random Forest, Support Vector Machine, Logistic Regression, and Neural Network - and interpreted the model decision-making mechanisms using the SHAP explainable framework. The results show that XGBoost achieves the highest classification accuracy (86.85%), followed by Random Forest (83.84%) and Logistic Regression (83.84%), while Support Vector Machine (78.90%) and Multilayer Perceptron (69.86%) exhibit relatively lower performance. SHAP analysis identifies Eu_N/Eu_N^*, 10 000(Eu_N/Eu_N^*)/Y, Ti, Y, △FMQ, and Gd/Dy as key indicators for discriminating mineralization potential. The models constructed in this study, such as XGBoost and Random Forest, can effectively identify porphyry copper mineralization potential and screen critical exploration indicators, offering not only new insights into intelligent mineral exploration but also important references for model selection and optimization in practical applications.

Key words: porphyry copper deposit; machine learning; zircon trace elements; big data.

0　引言

斑岩型铜矿床是全球铜、钼、金等金属资源的重要来源,提供了全世界约75%的铜、50%的钼和20%的金(Sillitoe, 2010).这类矿床主要形成于与板块俯冲或碰撞造山相关的中酸性弧岩浆岩带(Sillitoe, 1972, 1993; Hedenquist and Richards 1998; Hezarkhani, 2006; Hou *et al.*, 2007, 2011; Meng *et al.*, 2021; Luo *et al.*, 2024;孟旭阳和毛景文, 2025).其成矿岩浆通常具有中等氧化、富水和挥发分的特点(Piquer *et al.*, 2021; Stonadge *et al.*, 2023),岩浆岩全岩地球化学特征通常表现出高Sr/Y和La/Yb比值,并缺乏显著的Eu负异常(Richards and Kerrich, 2007).

锆石因其稳定的晶体结构和较强的抵抗变质、侵蚀和风化能力,是硅酸盐岩浆中常见的副矿物(Lu *et al.*, 2016; Loucks *et al.*, 2020; Meng *et al.*, 2022; Luo *et al.*, 2024),其微量元素(如Ce、Eu、Nb、Ta等)常被用作指示岩浆氧逸度、分异程度及流体活动等与成矿有关的关键信息(Belousova *et al.*, 2002; Houston and Dilles, 2013; Zhong *et al.*, 2023).前人研究已将若干锆石微量元素比值(如$Eu_N/Eu_N^*\geqslant0.4$,$(Ce/Nd)/Y>0.01$,$10\,000\times(Eu_N/Eu_N^*)/Y>1$和$Dy/Yb<0.3$等)作为判断岩浆斑岩型铜矿成矿潜力的有效指标,并成功应用于找矿勘查实践(Lu *et al.*, 2016; Pizarro *et al.*, 2020; Carrasco-Godoy *et al.*, 2024).例如,Lee *et al.*(2021)和Pizarro *et al.*(2024)应用锆石中的地球化学指标(如$Eu_N/Eu_N^*\geqslant0.4$、$10\,000\times(Eu_N/Eu_N^*)/Y>1$、$(Ce/Nd)/Y>0.01$等)分别在不列颠哥伦比亚中南部Highland Valley斑岩区和智利北部Centinela District发现显著的斑岩铜矿成矿潜力.

传统的二维判别图解通常基于固定的元素比值或组合,其分析思路本质上是线性的,难以充分处理高维、非线性的地球化学数据.具体而言,这类方法常面临以下局限:一是投图区域往往存在显著重叠,导致判别边界模糊,难以实现精细分类(Zhou *et al.*, 2022; Zou *et al.*, 2022);二是其判别过程缺乏量化评估标准,无法客观衡量模型的有效性与准确性(Zou *et al.*, 2022;滕文航等,2025);三是传统图解多建立于较小样本基础上,对当前日益积累的海量地球化学数据难以充分挖掘与利用(Petrelli and Perugini, 2016;赵伟光等,2021;王瑀等,2022).相比之下,机器学习作为近年来迅速发展的数据分析方法,其算法能够适应高维、非线性关系,并从复杂的微量元素组合中识别出潜在的深层规律与特征,从而为构建普适性更强、判别精度更高的模型提供了有效途径(Zhou *et al.*, 2022; Zou *et al.*, 2022; Zhong *et al.*, 2023; Wen *et al.*, 2024a; Yuan *et al.*, 2025).因此,引入机器学习方法不仅有助于克服传统图解的上述缺陷,也为处理大规模、高维度地球化学数据提供了重要的技术支撑.

机器学习按照训练数据是否有分类标签,主要分为监督学习和无监督学习(张润和王永滨, 2016;姜娜等, 2019).监督学习基于已分类数据训练模型,实现可解释的分类判别(陈凯和朱钰, 2007).无监督学习则利用无标签数据,挖掘其中隐藏的特征和结构(Sahli, 2020; Luo *et al.*, 2024).目前,Extreme Gradient Boosting(XGBoost)(Chen and Guestrin, 2016)、随机森林(Breiman, 2001)、逻辑回归(Hosmer Jr *et al.*, 2013)、支持向量机(Boser *et al.*, 1992)、神经网络(李苍柏等, 2020; Zhong *et al.*, 2023)等模型是常用的监督学习方法,已被广泛应用在地球科学领域.例如,Zhong *et al.*(2021)利用神经网络(多层感知器)对玄武岩的构造环境进行判别分类,并将该模型应用在熊儿火山岩群.

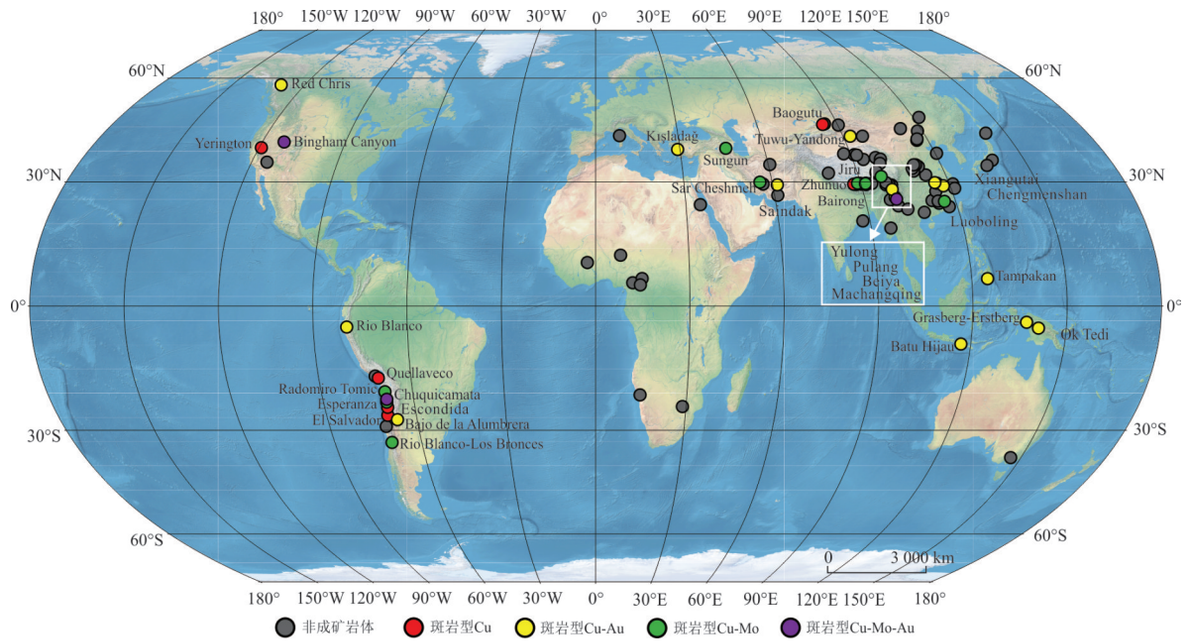

图 1 本文搜集锆石微量数据的斑岩型矿床和不成矿岩体位置

Fig.1 Locations of porphyry deposits and barren intrusions from which zircon trace element data were collected in this study

Zhou *et al.*(2022)应用支持向量机和随机森林模型,基于全球火成岩锆石微量元素数据有效判别了与成矿和非成矿相关的锆石.该研究进一步通过碎屑锆石分析,揭示冈底斯带东部成矿潜力在 50 Ma后显著升高,并指出高成矿潜力百分比与高奇异性指数的岩浆爆发事件密切相关.Zou *et al.*(2022)则利用随机森林和深度神经网络模型,构建了基于全球锆石微量元素数据的斑岩铜矿成矿潜力判别模型.并在俯冲(加拿大 Highland Valley)和碰撞后(西藏冈底斯带)两种构造背景的矿床中验证了模型不受区域构造背景差异的影响,确认了 Eu_N/Eu_N^*、Ce/Nd 等元素比值以及 Dy、Hf、Ti 含量是关键判别指标.Luo *et al.*(2024)通过锆石微量元素数据来训练支持向量机、随机森林、神经网络等模型,以此来对斑岩型矿床的构造背景(大陆弧、岛弧、非弧等)进行分类.以上成果表明,机器学习在地学领域具有很强的应用前景.

然而,目前研究仍存在一些不足.不同算法在斑岩铜矿预测分类方面缺乏系统性对比;模型训练数据多局限于特定区域,导致泛化能力有限.此外,尽管机器学习模型在锆石数据的分类与判别中表现出色,但由于其“黑箱”特性(即模型决策过程不透明),机器学习模型虽能够有效识别数据中的复杂模式并作出高精度预测,却通常无法阐明微量元素分异与岩浆过程、源区属性等之间的成因联系,因此难以支持深入的机理讨论.

针对上述不足,本研究构建了全球斑岩型铜矿成矿与非成矿岩浆锆石微量元素数据库,系统整合了岩浆氧逸度、含水量及结晶分异等关键控矿要素的多因素及比值指标.在此基础上,对比 XGBoost、随机森林、支持向量机、逻辑回归、多层感知器共 5 类机器学习模型的判别性能,并引入 SHAP 框架对优选模型进行解释,揭示微量元素与成矿潜力间的可解释关联,突破“黑箱”限制.该方法不仅提升了复杂非线性关系下的成矿识别精度,也使决策过程可追溯.最后进一步评估各算法在地球化学判别中的适用性与差异,明确关键判别指标,旨在构建兼具高精度、强泛化能力与可解释性的成矿潜力评价模型,为矿产勘探提供科学工具,推动机器学习在地学领域的可信应用.

1 数据与方法

1.1 数据来源与筛选

在本次研究中,成矿岩浆锆石数据均系统收集自有明确矿化描述的斑岩型铜矿相关文献(详细来源见附表 1).为构建有效的判别对比,本研究同时选取了具有对比意义的非成矿岩浆锆石数据,其选取依据主要包括:(1)空间上位于同一或邻近矿集区;(2)形成时代与成矿岩体相近;(3)与成矿岩体属于同一构造-岩浆旋回或处于相同的大地构造环境,但未发生成矿

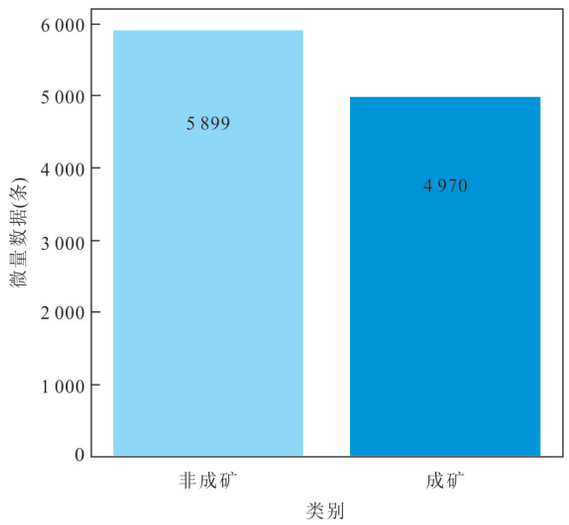

图2　斑岩型矿床成矿岩浆锆石和非成矿岩浆锆石数量统计直方图

Fig.2　Statistical histogram of metallogenic and barren magmatic zircon in porphyry deposits

作用.在此基础上,笔者将所有锆石数据划分为成矿与非成矿两类,以支撑后续的机器学习建模.

为了提高模型的准确性,笔者使用La<1×10^{-6},Ti<50×10^{-6}以及(Dy/Nd+Dy/Sm)>10×10^{-6}来筛选未遭受蚀变和未受矿物/熔融包裹体影响的锆石数据(Bell *et al.*, 2016; Lu *et al.*, 2016; Meng *et al.*, 2024).其次,剔除了缺失率大于20%的特征指标,并对剩余锆石微量元素的缺失值用同类型(成矿岩浆锆石或非成矿岩浆锆石)及同特征指标的均值进行填充.最终获得了成矿岩体锆石的微量数据4 970条,非成矿岩体锆石的微量数据5 899条(如图2).本次研究选取了Hf、Th、U、Y、Ti、Ce、Pr、Nd、Sm、Eu、Gd、Tb、Dy、Ho、Er、Tm、Yb、Lu共18种原始锆石微量元素作为分类指标,并使用Eu_N/Eu_N^*、Ce_N/Ce_N^*(Loader *et al.*, 2017)、10 000(Eu_N/Eu_N^*)/Y、(Th+U)/Y、(Ce/Nd)/Y、Gd/Dy、Ce/Nd、Dy/Yb、Th/U、△FMQ(Loucks *et al.*, 2020)共10项锆石微量元素比值及衍生指标作为本研究的分类特征.这些指标从岩浆氧逸度、含水量以及结晶分异过程共三个关键控制因素出发,有效捕捉与成矿岩浆相关的特征(Ballard *et al.*, 2002; Dilles *et al.*, 2015; Shen *et al.*, 2015; Lu *et al.*, 2016; Pizarro *et al.*, 2020).

氧逸度指标Ce_N/Ce_N^*是基于锆石内变价元素Ce中Ce^{4+}和Ce^{3+}的比值来反映岩浆的相对氧化还原状态(Ballard *et al.*, 2002; Lu *et al.*, 2016).而Eu_N/Eu_N^*、10 000(Eu_N/Eu_N^*)/Y、(Ce/Nd)/Y和Dy/Yb等衍生指标,可识别岩浆演化过程对矿物分离结晶的影响(Lu *et al.*, 2016; Pizarro *et al.*, 2020).

$$Eu_N/Eu_N^* = Eu_N/\left(Sm_N \times Gd_N\right)^{0.5}, \tag{1}$$

$$Ce_N/Ce_N^* = Ce_N/\left(Nd_N^2/Sm_N\right), \tag{2}$$

$$\Delta FMQ = 3.998(\pm 0.124) \times \lg\left[\left(\frac{Ce}{\sqrt{U_i \times Ti}}\right)\right] + 2.284(\pm 0.101). \tag{3}$$

除以上数据集外,本研究还单独将美国的Bingham Canyon斑岩型Cu-Au-Mo矿床、土耳其的Kışladağ斑岩型Cu-Au矿床、伊朗的Sungun斑岩型Cu-Mo矿床、美国的Yerington斑岩型Cu矿这四个矿床成矿岩体的锆石微量数据,以及中国芒康地区的不成矿岩体的锆石微量数据整合为一组新数据集作为未知样本数据,目的是作为一个实例验证本研究训练后的机器学习模型.这些数据与上文提到的全球锆石微量数据采用相同的标准进行筛选后,成矿岩浆锆石微量数据635条、非成矿岩浆锆石微量数据95条.其详细来源见附表1.

1.2　机器学习方法

机器学习主要包括监督学习、无监督学习两种学习方法(张润和王永滨, 2016).在监督学习中,需要使用已有分类标签的数据对模型进行训练(张润和王永滨, 2016; 姜娜等, 2019; 杨剑锋等, 2019).无监督学习则可直接利用未有分类标签的数据对模型进行训练(杨剑锋等, 2019; 孙会玲等, 2026).本研究为了横向对比不同模型的差异性,选取五种典型的监督机器学习算法,利用已有标签的锆石微量数据对其进行训练.

1.2.1　XGBoost模型　该模型是由陈天奇等人于2016年开发的,基于Boosting集成技术以及梯度提升框架而形成的一种算法(Chen and Guestrin, 2016).该模型首先构建一棵树作为初始模型,计算出预测值与真实值之间存在的误差近似值,如公式(4)所示,将其作为新的目标值,并对下一颗树进行训练,通过重复性迭代构建更多的树模型(图3).在迭代的过程中,通过一个损失函数和正则化步骤对目标函数进行优化,以此来减小预测过程中存在的误差以及防止模型出现过拟合化.最后将所有树模型得到的预测结果相加得到XGBoost模型最终预测结果(Chen and Guestrin, 2016; Wen *et al.*, 2024b).误差近似值公式如下:

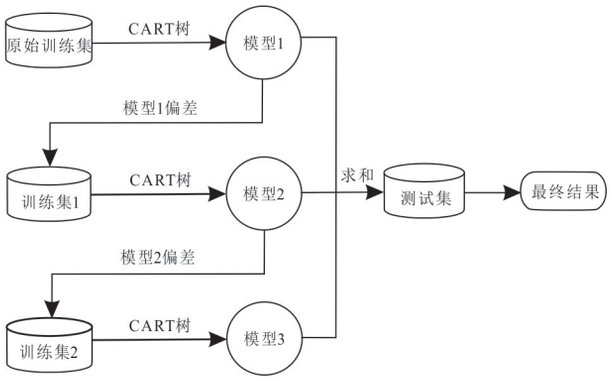

图 3 XGBoost模型原理(据Chen and Guestrin, 2016修改)

Fig.3 XGBoost model principle (modified after Chen and Guestrin, 2016)

$$r_{it}=-\left[\frac{\partial L\left(y_i, F_{t-1}(x_i)\right)}{F_{t-1}(x_i)}\right], \tag{4}$$

其中,$F_{t-1}(x_i)$是前$t-1$轮的集成模型,L为损失函数,r_{it}表示第t轮迭代中第i个样本的误差近似值,该误差近似值将作为下一个模型的拟合目标.

1.2.2 随机森林模型 它是由Breiman提出的一种基于Bagging技术的集成方法,由多棵决策树共同组成,多使用于分类和回归任务(Breiman, 2001; 姜娜等, 2019; 侯霖莉等, 2024).随机森林在任务中使用bootstrap重抽样方法从数据集中抽取多个样本数据,并随机选择每个bootstrap样本的一些特征建立决策树模型,这些决策树独立对数据进行学习、分类,从而获得判别分类结果(Breiman, 2001; Barry, 2013; Song and Lu, 2015; Lee *et al.*, 2019; 李苍柏等, 2020; Guo and He, 2024).而随机森林模型最终的判别分类结果是由所有决策树集体投票产生(图4; Breiman, 2001).

1.2.3 支持向量机 它是一种常用于分类和回归任务的监督机器学习方法.该方法主要利用核函数将数据映射到高维特征空间中,在这个空间找到一个超平面来分割不同类别的数据(图5),以达到对数据进行分类的目的(Boser *et al.*, 1992; 陈凯和朱钰, 2007; 李苍柏等, 2020).支持向量机生成的超平面是这个空间中最优的平面,它使最近样品点到超平面的间隔最大,而对于一些复杂的非线性分类问题,支持向量机将使用非线性核函数创建分离数据的决策边界(Boser *et al.*, 1992; 刘小生和章治邦, 2019; 李苍柏等, 2020; Yuan *et al.*, 2025).

1.2.4 逻辑回归模型 该模型是一种广义的线性组合,常用于二分类问题,该方法主要是利用Sigmoid函数将线性组合映射到[0,1]区间内,如公式(5)、(6)所示,从而能够直接输出概率值,以达到分类数据的效果(Hosmer Jr *et al.*, 2013; Yuan *et al.*, 2025).

$$Z=\omega_0+\omega_1x_1+\omega_2x_2+\omega_3x_3+\cdots+\omega_nx_n, \tag{5}$$

$$y(Z)=\frac{1}{1+e^{-z}}, \tag{6}$$

其中,ω代表权重,x是输入特征指标(如Ti、Eu、Th/U等).

1.2.5 多层感知器模型 它是一种前馈神经网络模型,有多个层且每层由多个神经元组成,其中包

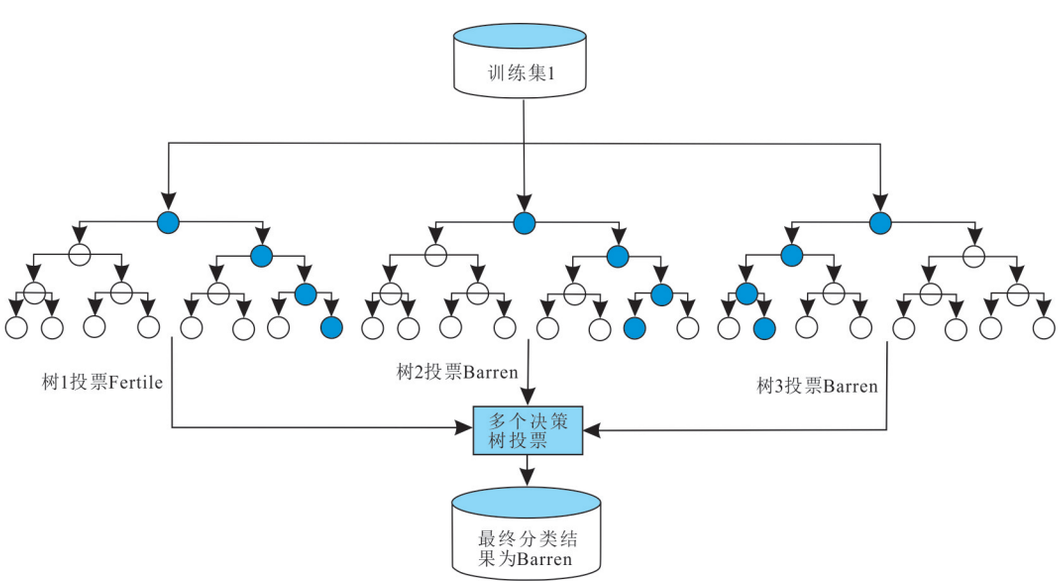

图 4 随机森林模型原理(据Breiman, 2001; Guo and He, 2024修改)

Fig.4 Principle of the random forest model (modified after Breiman, 2001; Guo and He, 2024)

蓝色圆圈代表每棵决策树选择的分类类型结果,最终根据所有决策树投票的结果,获得投票数量最多的类型将作为最终的分类结果

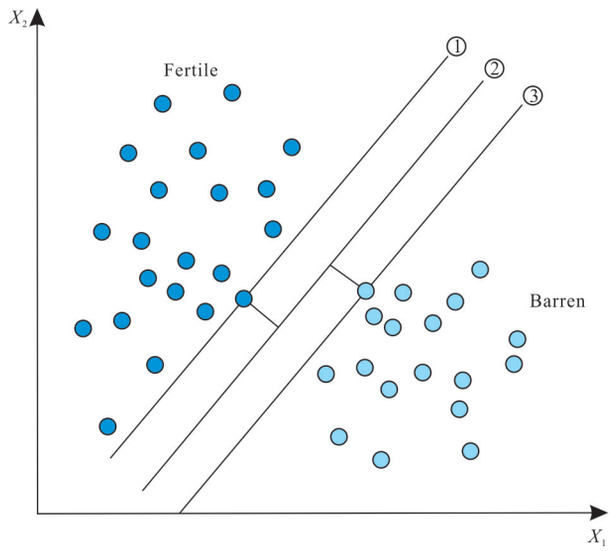

图5　支持向量机原理(据 Boser *et al.*, 1992; Luo *et al.*, 2024 修改)

Fig.5　Principle of support vector machine (modified after Boser *et al.*, 1992; Luo *et al.*, 2024)

深蓝色代表成矿岩体的锆石样本,浅蓝色代表非成矿岩体的锆石样本,X_1、X_2代表着区分两类锆石的关键地质特征(如微量元素 Ti、Y、Th/U 等),直线①②③是支持向量机构建的最优分类超平面(二维下表现为直线),与分类边界最近的样本点是决定分类边界的关键个体,代表两类锆石中最具区分度的典型实例;①$\vec{\omega}\cdot\vec{x}+b=-1$; ②$\vec{\omega}\cdot\vec{x}+b=0$; ③$\vec{\omega}\cdot\vec{x}+b=1$

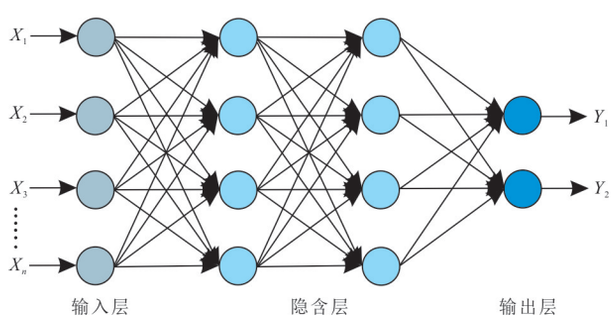

图6　多层感知器原理(据Guo and He, 2024修改)

Fig.6　Principle of multilayer perceptron (modified after Guo and He, 2024)

X_1、X_2、…、X_n代表输入的数据(如微量元素指标 Ti、Y、Th/U 等),输入层、隐含层、输出层代表着神经网络的三层结构,输入层是将数据进行输入,传输到隐含层;隐含层主要通过复杂的权重连接和激活函数(如 Sigmoid)对输入的锆石微量元素指标进行深层次挖掘信息,将得到的结果传输到输出层;对于二分类,输出层将接近 1 表示“成矿岩体锆石”,接近 0 表示“非成矿岩体锆石”进行输出,而Y_1、Y_2则代表输出的结果“成矿岩体锆石”或“非成矿岩体锆石”

括输入层、一个或多个隐藏层以及输出层(图 6).每个神经元通常使用非线性激活函数,因此该模型可以解决更为复杂的非线性问题.多层感知器模型中的上一层的每个神经元都与下一层的神经元相连,进而输入层接收数据传入隐含层处理数据,并在输出层输出最终判别分类结果(He *et al.*, 2015; Zhong *et al.*, 2023).

2　模型训练与评估

为直观对比传统方法与机器学习模型的判别性能,本研究首先采用经典的二元判别图解对数据进行分析.结果表明,成矿与非成矿岩浆锆石数据点在图中均存在显著重叠,难以通过单一或少数线性边界实现有效区分(图 7).这一现象凸显了传统方法在应对高维、非线性地球化学数据时的固有局限.

因此,本研究转向机器学习模型,系统评估更先进的数据驱动方法.为确保模型评估的规范性与结果可比性,将收集的全球锆石微量元素数据集按 7∶3 的比例随机划分为训练集与测试集.其次,使用 XGBoost、随机森林、支持向量机、逻辑回归及多层感知器共五种模型,在相同的训练集上进行学习,并在统一的测试集上评估其泛化性能.旨在客观比较各算法从复杂特征中挖掘判别规律的能力,进而构建泛化性强、精度高的成矿潜力判别模型.

在模型训练阶段,本研究首先在训练集上实施五折交叉验证,具体流程为:将训练集随机划分为五个互不重叠的子集,依次选取其中四个子集作为训练数据用于构建模型,剩余一个子集作为验证集用于评估模型性能;为了减少随机性等影响,该实验过程重复五次(图 8).该方法能够模拟模型在未知数据上的泛化表现,有效降低因单次数据划分的随机性所引起的模型评估偏差,提高模型性能评估的可靠性,同时为参数选择提供更为稳定的依据(Nathwani *et al.*, 2022).随后,为充分挖掘各模型的判别分类潜力,采用随机搜索策略对模型超参数进行优化,确定各模型的最佳参数配置并予以保存,再基于最优参数对模型进行重新训练.最终,将训练完成的模型应用于测试集中的锆石微量元素数据,开展判别分类任务,并通过分析分类结果,系统评估与比较各模型的综合性能.

本文将采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1-score、受试者特征曲线(ROC-AUC)、精确率-召回率曲线(PR-AUC)等指标来评价各模型在区分成矿岩浆锆石和非成矿岩浆锆石任务中的性能.其中准确率表示模型对成矿岩体锆石和非成矿岩体锆石做出正确分类的比例,精确率是指所有被分类为成矿岩体锆石中实际为

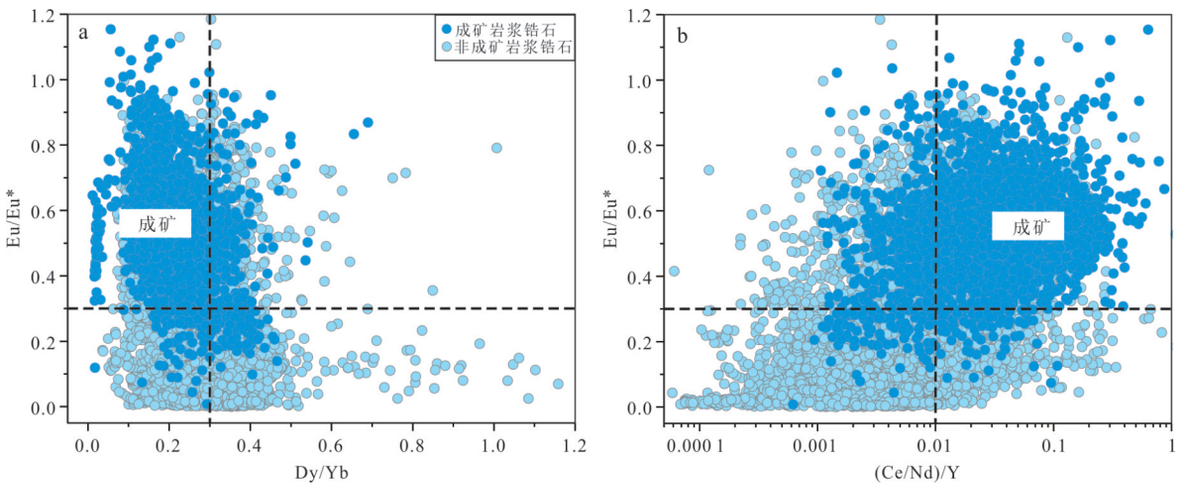

图 7 成矿岩浆锆石和非成矿岩浆锆石散点图(据 Lu *et al.*, 2016 修改)

Fig.7 Scatter plots of mineralized versus non-mineralized magmatic zircons (modified after Lu *et al.*, 2016)

a 图中 Eu/Eu*(>0.3)与 Dy/Yb(<0.3)为成矿岩浆锆石区域;b 图中 Eu/Eu*(>0.3)与 Ce/Nd/Y(>0.01)为成矿岩浆锆石区域

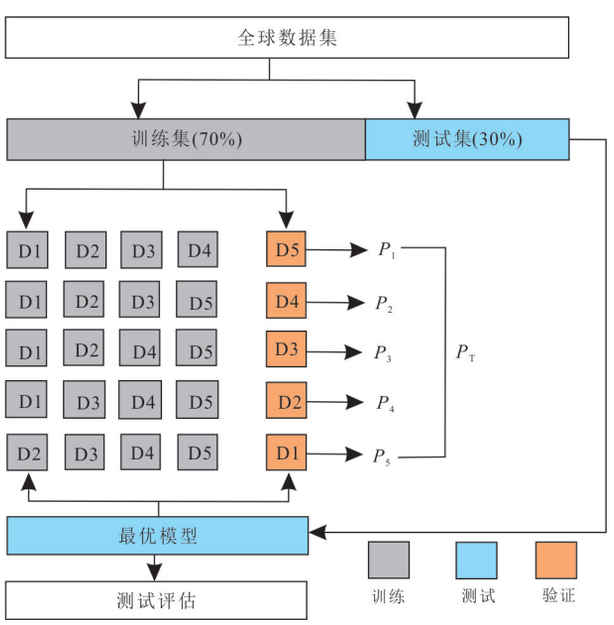

图 8 五折交叉验证原理(Pedregosa *et al.*, 2011)

Fig.8 Principle of five-fold cross-validation (Pedregosa *et al.*, 2011)

D1、D2、D3、D4、D5 代表将训练集 1 分成的 5 个子集;P_1、P_2、P_3、P_4、P_5 代表每次交叉验证之后的性能指标;P_T 代表五次交叉验证之后的平均性能指标

成矿岩体锆石的数量的比例,召回率表示真实成矿岩体锆石样本数量中预测为成矿的数量所占的比例,F1-score 是精确率和召回率的调和平均值,各指标表示见公式(7)~(10).

ROC 曲线是以假阳性率(FPR)为横坐标,真阳性率(TPR)为纵坐标,公式(11)~(12).PR 曲线则是以召回率为横坐标,精确率为纵坐标,主要用来解释召回率和精确率之间的关系,当两个类别的样本量不平衡时,PR 曲线更能反映模型的性能.虽然 ROC 曲线和 PR 曲线都很直观,但无法定量地描述模型的性能,因此可以使用曲线下的面积来评估模型的性能(Yang *et al.*, 2024).而 AUC 值则是表示曲线下的面积大小,能够量化反映模型的性能,并且其值越接近于 1,说明该模型性能越好(潘北斗等, 2022;刘嘉情等, 2023; Carrasco-Godoy *et al.*, 2024).

$$准确率=\frac{TP+TN}{TP+FP+TN+FN}, \tag{7}$$

$$精确率=\frac{TP}{TP+FP}, \tag{8}$$

$$召回率=\frac{TP}{TP+FN}, \tag{9}$$

$$F1\text{-}score=\frac{2\times Precision\times Recall}{Precision+Recall}, \tag{10}$$

$$真阳性率(TPR)=\frac{TP}{TP+FN}, \tag{11}$$

$$假阳性率(FPR)=\frac{FP}{FP+TN}, \tag{12}$$

其中,TP 代表实际为成矿岩体锆石预测结果也为成矿岩体锆石的样本数量,FP 则表示实际为非成矿岩体锆石而预测为成矿岩体锆石的样本数量,TN 则代表实际为非成矿岩体锆石预测结果也为非成矿岩体锆石的样本数量,FN 则表示实际为成矿岩体锆石而预测为非成矿岩体锆石的样本数量(图 9).

本研究使用的编程语言为 Python,主要的库是 Scikit-learn、Pandas Numpy 等,全程主要在 Visual Studio Code 软件运行相关代码.

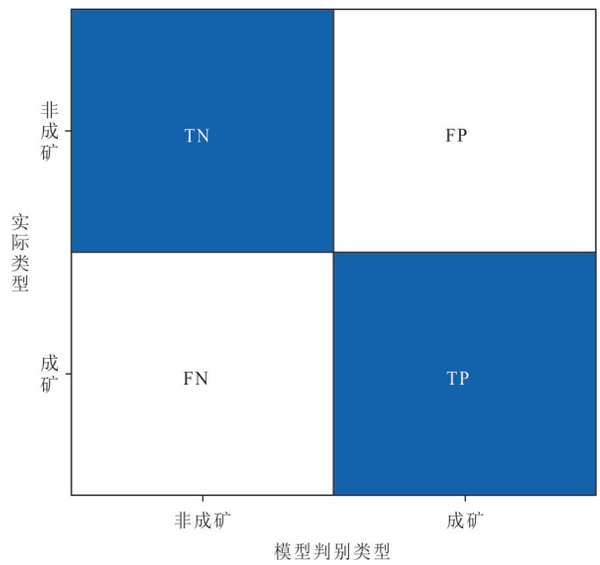

图 9　混淆矩阵图

Fig.9　Confusion matrix

3　结果分析

3.1　机器学习模型性能

在模型训练阶段,本研究首先基于训练集采用五折交叉验证方法进行超参数优化与初步性能评估,结果如表 1 所示.性能对比结果表明,XGBoost 模型训练表现最优,其五折交叉验证准确率为 0.944 4±0.003 7,精确率为 0.943 8±0.003 7.神经网络(多层感知器)次之,准确率与精确率分别为 0.943 7±0.005 7 和 0.943 2±0.005 6.而支持向量机和随机森林模型的准确率分别为 0.931 4±0.008 2 和 0.919 3±0.006 9,训练表现中等.逻辑回归模型训练表现相对较差,准确率与精确率分别为 0.878 9±0.010 2 和 0.878 2±0.009 9,均显著低于其他模型.

在模型通过随机搜索确定最优超参数后,采用该参数组合在整个训练集上重新训练,最终模型在训练集上的学习效果见附图 1、附表 2.混淆矩阵结果表明,XGBoost 模型在训练集上取得了最优的学习性能,其分类准确率与精确率均达到 100%(见附图 1).在测试集上,XGBoost 同样表现出最佳的泛化能力,分类准确率与精确率均为 95.28%(表 2),高于其他模型.其余模型在测试集上的表现从高到低依次为:神经网络(多层感知器)准确率与精确率均为 95.00%;支持向量机均为 94.14%;随机森林分别为 92.79% 与 92.80%;逻辑回归模型其准确率为 89.05%,精确率为 89.33%.

根据各模型在测试集的表现,为了评估各个模型的泛化能力,绘制的模型 ROC 曲线和 PR 曲线结果如图 10、表 2 所示.结果表明,表现最好的 XGBoost 模型其 AUC 和 AP 值分别为 0.990 1 和 0.988 0.神经网络(多层感知器)模型的 AUC 值和 AP 值分别为 0.986 5 和 0.982 1.表现中等的随机森林模型和支持向量机模型其 AUC 和 AP 值分别为 0.982、0.978 和 0.981、0.980.逻辑回归模型的 AUC 值和 AP 值分别是 0.944 8、0.905 5,相对较差.

以上结果表明,各模型在测试集上的性能依次递减:XGBoost 模型性能表现相对最优,多层感知器模型为次优方法,支持向量机与随机森林模型表现中等,而逻辑回归模型性能相对欠佳.

3.2　模型性能独立测试

在此次研究中,XGBoost 模型(准确率 95.28%)、多层感知器(准确率 95%)、支持向量机(准确率 94.14%)、随机森林(准确率 92.79%)等模型在测试集中都表现出很高的分类准确率(表 2).但是,已有研究表明,由于使用的机器学习方法(多层感知器、随机森林、XGBoost 等模型)会出现过拟合或欠拟合以及“黑盒效应”等原因,可能无法准确反映训练后的模型性能.例如,在参数巨量的神经网络中,这种不可解释性直接影响了模型结果的准确性与可靠性(Reunanen, 2003; Rudin, 2019; Zhong *et al.*, 2023).本研究的模型性能虽基于独立的测试集进行评估,但该评估方式可能存在潜在的数据泄漏风险.由于训练集与测试集源自同一总体样本,且按固定比例随机划分,二者可能在岩石样本或地

表 1　模型五折交叉验证平均性能指标

Table 1　Average performance metrics of five-fold cross-validation for the model

模型	准确率	精确率	召回率	F1-score	AUC
XGBoost	0.944 4±0.003 7	0.943 8±0.003 7	0.944 3±0.003 7	0.944 0±0.003 7	0.986 5±0.001 9
多层感知器	0.943 7±0.005 7	0.943 2±0.005 6	0.943 6±0.005 9	0.943 4±0.005 7	0.983 3±0.002 1
支持向量机	0.931 4±0.008 2	0.930 9±0.008 5	0.931 0±0.008 0	0.930 9±0.008 2	0.978 0±0.003 6
随机森林	0.919 3±0.006 9	0.918 5±0.007 1	0.919 2±0.006 8	0.918 8±0.006 9	0.975 1±0.003 4
逻辑回归	0.878 9±0.010 2	0.878 2±0.009 9	0.880 6±0.009 8	0.878 6±0.010 2	0.939 3±0.008 2

表 2 模型在测试集上的性能指标值

Table 2 Model performance metrics on the test set

模型	准确率	精确率	召回率	F1-score	AUC	AP
XGBoost	0.952 8	0.952 8	0.952 8	0.952 8	0.990 1	0.988 0
多层感知器	0.950 0	0.950 0	0.950 0	0.950 0	0.986 5	0.982 1
支持向量机	0.941 4	0.941 4	0.941 4	0.941 4	0.980 9	0.979 8
随机森林	0.927 9	0.928 0	0.927 9	0.928 0	0.981 9	0.977 8
逻辑回归	0.890 5	0.893 3	0.890 5	0.890 7	0.944 8	0.905 5

表 3 模型在未知数据集上的性能指标值

Table 3 Performance metrics of the model on the unseen dataset

模型	准确率	精确率	召回率	F1-score
XGBoost	0.868 5	0.923 6	0.868 5	0.883 8
随机森林	0.838 4	0.914 3	0.838 4	0.859 5
逻辑回归	0.838 4	0.909 3	0.838 4	0.858 9
支持向量机	0.789 0	0.915 7	0.789 0	0.821 4
多层感知器	0.698 6	0.900 2	0.698 6	0.747 9

质背景上高度同源，导致模型在测试集上的性能被高估，无法真实反映其泛化能力.

为系统评估五种机器学习方法的泛化能力，本研究采用一个独立于训练与测试过程的未知数据集对模型进行验证.基于已训练完成的五个模型对未知数据集进行判别分类，所得结果及性能指标见图 12 和表 3. XGBoost 模型在此独立验证集上仍保持最优性能，准确率为 86.85%，精确率为 92.36%，体现出良好的泛化能力与分类稳定性.随机森林与逻辑回归模型的准确率均为 83.84%，但从精确率来看，随机森林(91.43%)略优于逻辑回归(90.43%).支持向量机与神经网络(多层感知器)的分类性能下降幅度较大，准确率分别为 78.90% 与 69.86%.以上结果表明，XGBoost、随机森林等模型在面对未知数据时仍能保持较高的判别能力，具备较好的稳定性与实际应用潜力，而支持向量机和神经网络(多层感知器)下降幅度较大，泛化能力表现较弱，其模型性能不稳定.

因此，整体来说，XGBoost 模型的性能表现最好且稳定，其次就是随机森林模型，接着是逻辑回归、支持向量机、多层感知器等模型.

3.3 基于 SHAP 的模型可解释性分析与关键特征识别

为揭示机器学习模型的决策机制并识别控制成矿潜力的关键地球化学指标，本研究采用 SHAP(SHapley Additive exPlanations)框架对表现最佳的模型进行了可解释性分析.SHAP 值能够定量评估每个特征对于单个样本预测结果的贡献度与方向(正负影响)，其全局重要性(取绝对值后的均值)则可用来衡量特征的整体影响力.

在所有对比模型中，XGBoost 整体表现出最优的综合性能(表 2).因此，笔者以其为例进行深入的特征解读.SHAP 分析表明(图 11a)，Eu_N/Eu_N^*是判别成矿潜力的首要控制因子.紧随其后的关键指标包括 Lu、Gd/Dy、10 000×(Eu_N/Eu_N^*)/Y、Yb、△FMQ、Y、Pr、Hf、Ti、U 等.这些发现与本文构建特征体系时所依据的氧逸度、含水量与结晶分异三大控制维度完全吻合，从数据驱动角度验证了其理论基础.

为检验上述关键特征是否在不同算法中具有稳健性，笔者选取了性能仅次于 XGBoost 模型且在未知数据集上表现优异的随机森林模型进行对比分析，而多层感知器在未知数据集上泛化能力表现

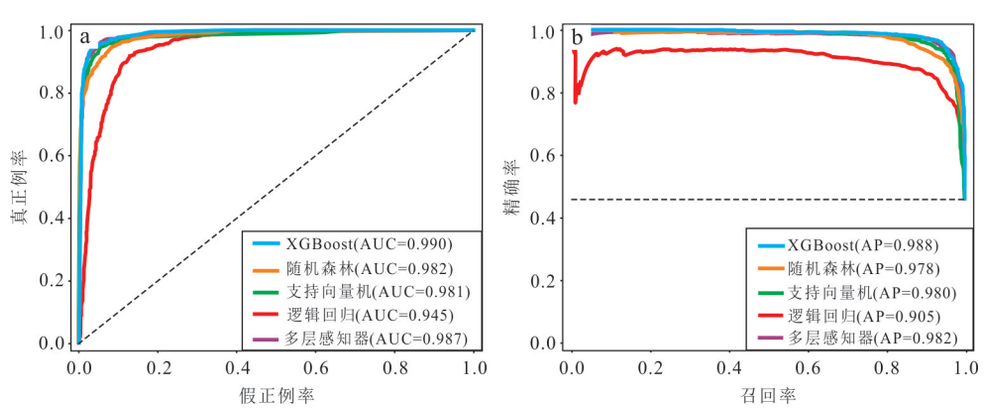

图 10 模型评价曲线

Fig.10 Model evaluation curves

a. ROC 曲线；b. PR 曲线；AUC. 曲线下的面积；AP. 平均精确度

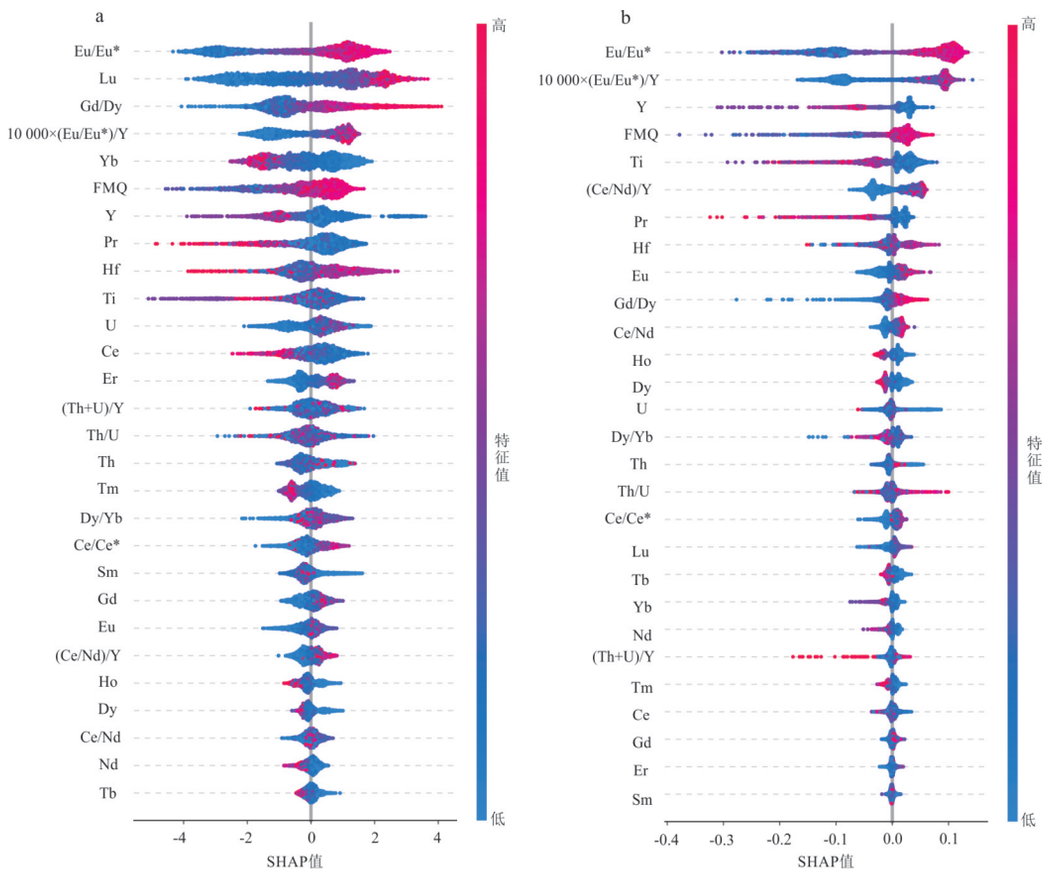

图 11　基于SHAP值的特征重要性及影响方向分析蜂窝图

Fig.11　Honeycomb plot of feature importance and influence direction analysis based on SHAP values

a 为 XGBoost 模型；b 为随机森林模型 .a 和 b 纵轴特征指标按其重要性从大到小排序 . 颜色条深浅的变化代表特征值的高低（红色代表高值，蓝色代表低值）. 图中的每个点代表一个样本，其水平位置表示该样本中特征值对应的 SHAP 值（正值推动预测趋向“成矿”，负值推动预测趋向“非成矿”）

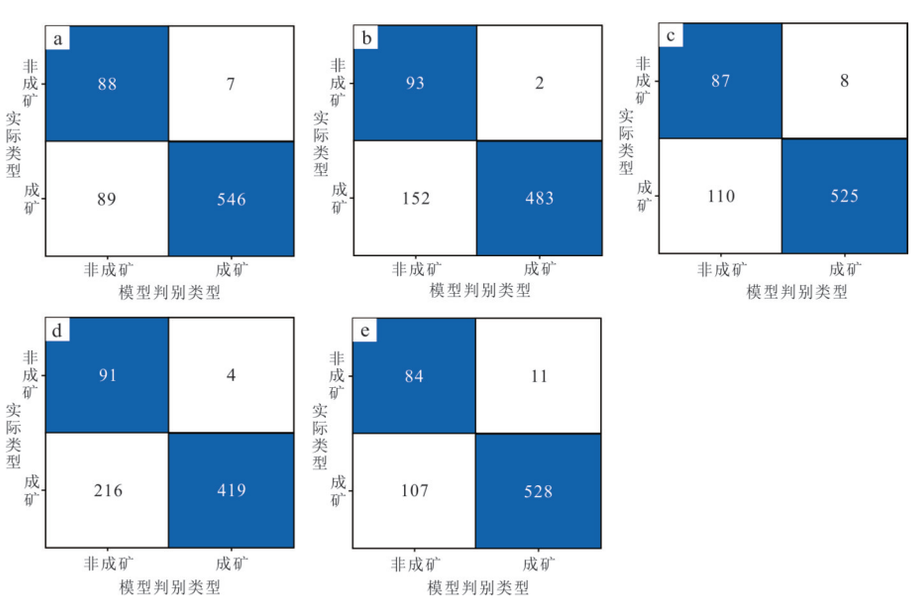

图 12　模型对新数据集的分类结果混淆矩阵

Fig.12　Confusion matrix of the model’s classification results on the new dataset

a. XGBoost模型分类结果；b. 支持向量机模型分类结果；c. 随机森林对新数据集分类结果；d. 神经网络（多层感知器）模型对新数据集分类结果；e. 逻辑回归模型对新数据集分类结果

弱,其性能大幅下降(表3),所以其识别的特征重要性不作为本次研究的参考.SHAP分析结果显示(图11b),尽管两个模型在特征重要性的排序上存在细微差异,但识别出了一组高度一致的核心特征组合,包括:Eu_N/Eu_N^*、10 000(Eu_N/Eu_N^*)/Y、Ti、Y、ΔFMQ与Gd/Dy.其中特征指标Eu_N/Eu_N^*、10 000(Eu_N/Eu_N^*)/Y及ΔFMQ是与含水量和氧逸度密切相关的指标,其SHAP值表现出与特征值明确的正相关关系(即图中点群颜色从蓝到红,其水平位置整体向右移动).这表明,更高的计算水含量及更高的氧逸度,对模型将其判断为成矿锆石产生强烈的正向贡献,这与传统成矿理论认识完全一致.这组特征在多模型中反复出现,其指示意义不受单一算法偏好影响,也证明了它们在判别成矿与非成矿锆石中的普适性与地球化学重要性.

4 讨论

4.1 模型性能差异性分析

XGBoost和随机森林两个模型在测试集和未知数据集上都表现出较高的准确率,并展现出卓越且稳定的分类性能与泛化能力.这得益于它们都是集成学习机制(Chen and Guestrin, 2016; 姜娜等, 2019),其中XGBoost模型通过梯度提升策略依次构建多棵决策树,每一棵树都致力于修正前一棵树的残差(公式(4)),这种机制能更高效捕捉锆石微量元素的高维复杂特征,精准挖掘斑岩铜矿成矿数据规律中的非线性关联(姜娜等, 2019).而对于锆石微量元素来说,其成矿信息不仅包含锆石单一元素的含量,更依赖于其元素比值(如10 000(Eu_N/Eu_N^*)/Y、(Ce/Nd)/Y、Ce/Ce*)与岩浆过程(水含量、氧逸度)的非线性关联(Chen and Guestrin, 2016).因此对于XGBoost模型来说,其能从中稳健地识别出对斑岩铜矿成矿潜力最具指示意义的关键特征.同时,XGBoost内置的正则化项有效控制了模型复杂度,避免了过拟合,从而在面对来自不同构造背景、未见过的矿床数据时,仍能保持较高的分类精度,最终在斑岩铜矿成矿潜力分析中展现出比单一模型更稳定的泛化能力与决策可靠性(Chen and Guestrin, 2016; Wang *et al.*, 2023a, 2023b).而且这与模型在测试集和新数据集分类准确率(表2,表3)结果上的表现一致.

随机森林通过Bagging重抽样构建多棵决策树,并以“投票”的方式输出结果(图4).尽管并行决策机制在处理复杂数据且具有弱关联特征时略逊于XGBoost,但依然表现出稳健的判别分类能力(表2,表3).其与XGBoost共同证明了集成学习模型在处理高维、非线性地球化学数据方面的强大优势.

支持向量机模型在测试集上表现较好(94.14%),而在新数据集上准确率(78.9%)却较低.其原因是该模型通过核函数将数据映射至高维空间以构建分类超平面(图5),所以其性能主要依赖核函数选择以及参数调优.并且该模型对核函数以及参数敏感,容易受到未知锆石微量元素数据(如美国Bingham Canyon矿床、土耳其Kışladağ矿床)分布及与训练集存在差异的影响,从而使得模型分类边界偏移,导致其准确率降低.

而逻辑回归模型在对新数据集进行分类时,其分类准确率(83.84%)超过了支持向量机模型(78.9%)和神经网络(多层感知器)模型的准确率(69.86%).其原因可能是逻辑回归作为线性模型,主要通过Sigmoid函数将线性组合映射到区间[0, 1](公式(5)~(6)),其测试集准确率最低(89.05%),这能够反映出锆石微量元素与成矿信息的非线性本质,并非单一元素的简单线性叠加,而是氧逸度、含水量、分异程度的耦合作用,这与传统的二维图解法的局限一致(Zou *et al.*, 2022; Zhong *et al.*, 2023; Wen *et al.*, 2024b).而其模型在新数据集上准确率提升(83.84%),并且与随机森林模型相同,其中的原因是模型结构简单,仅依赖关键特征的线性关系(如Eu_N/Eu_N^*、10 000(Eu_N/Eu_N^*)/Y与成矿呈正相关),而在新数据集中此类线性关联并未发生根本变化,并且逻辑回归对特征线性关系具有良好的稳定性,因此避免了复杂模型对噪声特征的过度拟合,且从训练集学到的线性特征关系刚好能够应用在新数据集上,从而提高了其模型在新数据集上分类的准确率(刘晓丽等, 2022; Yang *et al.*, 2024).以上结果表明,在数据量有限以及线性特征关系明显的场景中,简单的线性模型(如逻辑回归)仍具有一定的应用价值,但是无法替代非线性模型对特征之间关系的深度解析.

神经网络(多层感知器)虽在测试集中表现较好(准确率95%),仅次于XGBoost模型,但在独立新数据集上准确率骤降至69.86%,其性能表现是五个模型中最差的.其结果可能是由于神经网络(多层感知器)出现过拟合所导致,而神经网络具有黑箱效应,其通过多层神经元的权重连接挖掘特征信

息(图 6),但锆石微量数据中部分区域上特有的而并非成矿通用的元素信息被过度学习;从而导致模型在独立的新数据集上泛化能力不足.神经网络模型适用于具有大量样本的判别分类任务,本次研究样本量较小,也会使其分类的准确度降低.虽然训练集覆盖多种构造背景下的锆石微量数据,但新数据集中的锆石微量数据可能具有独有的特征,而神经网络并未学到这种特征,导致该模型误将成矿锆石的部分特征与非成矿锆石的独有特征混淆,从而将大部分成矿锆石判别分类为非成矿锆石,导致在新数据集的准确率下降.

4.2　关键判别指标及其地质意义

基于XGBoost与随机森林模型的SHAP特征重要性分析,本研究识别出一组高度一致的判别成矿与非成矿锆石的核心指标,包括:Eu_N/Eu_N^*、10 000×(Eu_N/Eu_N^*)/Y、Ti、Y、ΔFMQ及Gd/Dy(图 11).这一跨模型共识显著增强了这些指标的地质可靠性.它们并非孤立存在,而是分别从岩浆氧逸度、水含量、结晶温度与分异过程等关键控制维度,共同刻画了有利于斑岩铜矿形成的岩浆系统特征.

Eu_N/Eu_N^*与10 000×(Eu_N/Eu_N^*)/Y是揭示岩浆水含量的关键指标.锆石中Eu的负异常主要受控于斜长石的结晶分离,而富水条件下角闪石的结晶会抑制斜长石分离,从而保留熔体中的Eu异常.因此,显著的负Eu异常(低Eu_N/Eu_N^*值)通常与富水岩浆环境相关(Ballard *et al.*, 2002; Dilles *et al.*, 2015; Lu *et al.*, 2016; Li *et al.*, 2025).当显著的负Eu异常与较低的Y含量相结合时,则指示岩浆经历了以角闪石为主的大量分离结晶,而角闪石的稳定存在要求岩浆具有较高的初始水含量(Loader *et al.*, 2017).该指标的重要性直接印证了“富水岩浆”是形成大规模成矿流体物质基础的核心观点(Lu *et al.*, 2016).

Ti作为锆石结晶温度的可靠地球化学温度计(Watson and Harrison, 2005).斑岩成矿系统的成矿岩浆存在一个明确的有利温度窗口(通常为650~850 ℃),在岩浆固相线温度之上的较窄范围.中酸性岩浆中锆饱和结晶温度随着含水量增加而降低,因此由Ti含量记录的锆石结晶温度也越低,可能反映了岩浆中水含量较高(Richards, 2003; Lu *et al.*, 2016; Loucks *et al.*, 2020; Nathwani *et al.*, 2024).此外,Y作为重稀土元素(HREE)的代表,其含量主要受控于副矿物(如角闪石、磷灰石)的分离结晶.较低的Y含量常指示经历了显著的角闪石分异,这也与富水岩浆条件或者较深的岩浆源区特征一致(Ballard *et al.*, 2002; Loader *et al.*, 2017).

ΔFMQ作为直接的氧逸度量化指标,其作为重要性特征印证了高氧逸度是成矿的先决条件.高氧逸度能确保硫主要以硫酸盐形式而非硫化物形式存在于岩浆中,从而防止成矿金属过早析出,使其得以保留至流体出溶阶段(Mungall, 2002; Sun *et al.*, 2015).在本研究中,ΔFMQ的重要性排序相对靠后,可能源于以下两点原因:第一,Eu_N/Eu_N^*作为一个复合指标,已同时承载了氧逸度、水含量及斜长石分异的多重信号.在高氧逸度条件下,Eu以Eu^{3+}形式存在,易于进入斜长石晶格,从而增强锆石的Eu负异常(Ballard *et al.*, 2002; Trail *et al.*, 2012).因此,ΔFMQ的独立信息可能已被Eu_N/Eu_N^*部分涵盖.第二,氧逸度是成矿的必要非充分条件,需与适宜的水含量、温度等因素协同作用.Eu_N/Eu_N^*与Ti等指标更能直接表征流体出溶与结晶温度等直接控制因素(Shen *et al.*, 2015; Loucks and Fiorentini, 2023).此外,锆石ΔFMQ主要记录岩浆中晚期的氧逸度,而Eu异常可能反映了氧化和含水量特征的综合指标(Loucks *et al.*, 2020).

Gd/Dy等比值指标反映了“以角闪石主导”的结晶分异路径.在俯冲背景的岩浆系统中,角闪石(而非石榴石)作为主要的残留相或分离结晶相,使得中-重稀土元素(如Dy、Yb)以相近的比例进入熔体,此分异路径有利于铜等成矿金属的预富集(Chiaradia, 2014).

值得注意的是,支持向量机、逻辑回归及多层感知器模型所识别的重要特征集合与上述树模型结果存在差异(附图 2).这种差异可能源于算法原理:树模型(如XGBoost、随机森林)通过递归分割数据,擅长捕捉特征间的非线性交互作用,其共识特征可能更直接地对应关键地球化学变量(Chen and Guestrin, 2016);逻辑回归侧重线性关联(Hosmer Jr *et al.*, 2013);支持向量机依赖于核空间的最优边界(Boser *et al.*, 1992);多层感知器则可能构建复杂的隐式特征表示.因此,在树模型间达成高度共识的特征组合,因其在两种高性能且可解释性强的模型中得到交叉验证,很可能代表了控制成矿潜力最稳健、最核心的地球化学“指纹”.在地球化学数据挖掘中,应通过对比不同原理模型的输出,重点关注在多模型中共同凸显的特征,以增强地质结论的稳健性.

综上所述,机器学习模型所揭示的关键指标组合,共同指向一个具备高氧逸度、富水、经历角闪石域分异的弧岩浆系统,该结论与前人基于传统地球化学研究获得的认识高度相符(Lu *et al.*, 2016; Meng *et al.*, 2021; Meng *et al.*, 2022; Loader *et al.*, 2022; Loucks and Fiorentini, 2023).

5 结论

本研究通过构建包含全球斑岩型铜矿成矿岩浆与不成矿岩浆的锆石微量元素数据库,系统训练并评估了XGBoost、随机森林模型、支持向量机模型、逻辑回归模型、神经网络(多层感知器)五种监督学习模型,并展开性能评估.结论如下:在本次斑岩型铜矿成矿岩浆锆石和非成矿岩浆锆石分类任务中,模型性能从高到低依次为XGBoost、随机森林、逻辑回归、支持向量机与多层感知器.这一明确的性能排序,为在实际工作中选择适用于斑岩型铜矿成矿潜力评估的模型提供了关键参考.通过SHAP可解释性分析发现,Eu_N/Eu_N^*、10 000(Eu_N/Eu_N^*)/Y、Ti、Y、ΔFMQ及Gd/Dy等指标对区分成矿与非成矿锆石具有稳定且显著的贡献.从数据驱动角度揭示了斑岩成矿系统的地球化学标志,为基于锆石的成矿潜力识别提供了可靠的指标组合.

附表和附图见https://doi.org/10.3799/dqkx.2026.084.

Reference

Ballard, J. R., Palin, J. M., Campbell, I. H., 2002. Relative Oxidation States of Magmas Inferred from Ce(IV)/Ce(III) in Zircon: Application to Porphyry Copper Deposits of Northern Chile. *Contributions to Mineralogy and Petrology*, 144(3): 347−364. https://doi.org/10.1007/s00410-002-0402-5

Barry, D. V., 2013. Decision Trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6): 448−455. https://doi.org/10.1002/wics.1278

Bell, E. A., Boehnke, P., Harrison, T. M., 2016. Recovering the Primary Geochemistry of Jack Hills Zircons through Quantitative Estimates of Chemical Alteration. *Geochimica et Cosmochimica Acta*, 191: 187−202. https://doi.org/10.1016/j.gca.2016.07.016

Belousova, E., Griffin, W., O'Reilly, S. Y., et al., 2002. Igneous Zircon: Trace Element Composition as an Indicator of Source Rock Type. *Contributions to Mineralogy and Petrology*, 143(5): 602−622. https://doi.org/10.1007/s00410-002-0364-7

Boser, B. E., Guyon, I. M., Vapnik, V. N., 1992. A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. Pittsburgh Pennsylvania USA. ACM, 144−152. https://doi.org/10.1145/130385.130401

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1): 5−32. https://doi.org/10.1023/A:1010933404324

Carrasco-Godoy, C., Campbell, I. H., Cajal, Y., 2024. Quantifying the Criteria Used to Identify Zircons from Ore-Bearing and Barren Systems in Porphyry Copper Exploration. *Economic Geology*, 119(5): 1035−1058. https://doi.org/10.5382/econgeo.5086

Chen, K., Zhu, Y., 2007. A Summary of Machine Learning and Related Algorithms. *Journal of Statistics and Information*, 22(5): 105−112 (in Chinese with English abstract).

Chen, T. Q., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco. https://doi. org/10.1145/2939672.2939785

Chiaradia, M., 2014. Copper Enrichment in Arc Magmas Controlled by Overriding Plate Thickness. *Nature Geoscience*, 7(1): 43−46. https://doi. org/10.1038/ngeo2028

Dilles, J. H., Kent, A. J. R., Wooden, J. L., et al., 2015. Zircon Compositional Evidence for sulfur-Degassing from ore-Forming Arc Magmas. *Economic Geology*, 110(1): 241−251. https://doi.org/10.2113/econgeo.110.1.241

Guo, J., He, W. Y., 2024. Application of Machine Learning to Characterize Metallogenic Potential Based on Trace Elements of Zircon: A Case Study of the Tethyan Domain. *Minerals*, 14(9): 945. https://doi. org/10.3390/min14090945

He, K. M., Zhang, X. Y., Ren, S. Q., et al., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV). Santiago. https://doi.org/10.1109/iccv.2015.123

Hedenquist, J. W., Richards, J. P. 1998. The Influence of Geochemical Techniques on the Development of Genetic Models for Porphyry Copper Deposits. *Reviews in Economic Geology*, 235−256. https://doi. org/10.5382/Rev.10.10

Hezarkhani, A., 2006. Hydrothermal Evolution of the Sar-Cheshmeh Porphyry Cu-Mo Deposit, Iran: Evidence

from Fluid Inclusions. *Journal of Asian Earth Sciences*, 28(4－6): 409－422. https://doi. org/10.1016/j. jseaes.2005.11.003

Hosmer, D. W. Jr, Lemeshow, S., Sturdivant, R. X., 2013. Applied Logistic Regression. Wiley, Hoboken. https://doi.org/10.1002/9781118548387

Hou, L. L., Wu, S., Yi, J. Z., et al., 2024. Discriminating Deposit Types Using Chlorite Trace Elements Based on Machine Learning. *Earth Science*, 49(12): 4303－4317 (in Chinese with English abstract).

Hou, Z. Q., Pan, X. F., Yang, Z. M., et al., 2007. Porphyry Cu-(Mo-Au) Deposits No Related to Oceanic-Slab Subduction: Examples from Chinese Porphyry Deposits in Continental Settings. *Geoscience*, 21:332－351.

Hou, Z. Q., Zhang, H. R., Pan, X. F., et al., 2011. Porphyry Cu (-Mo-Au) Deposits Related to Melting of Thickened Mafic Lower Crust: Examples from the Eastern Tethyan Metallogenic Domain. *Ore Geology Reviews*, 39(1－2): 21－45. https://doi. org/10.1016/j. oregeorev.2010.09.002

Houston, R. A., Dilles, J. H., 2013. Structural Geologic Evolution of the Butte District, Montana. *Economic Geology*, 108(6): 1397－1424. https://doi. org/10.2113/econgeo.108.6.1397

Jiang, N., Yang, H. Y., Gu, Q. C., et al., 2019. Machine Learning and Its Algorithm and Development Analysis. *China Computer & Communication*, 31(1): 83－84, 87 (in Chinese with English abstract).

Lee, R. G., Plouffe, A., Ferbey, T., et al., 2021. Recognizing Porphyry Copper Potential from till Zircon Composition: A Case Study from the Highland Valley Porphyry District, south-Central British Columbia. *Economic Geology*, 116(4): 1035－1045. https://doi. org/10.5382/econgeo.4808

Lee, T. H., Ullah, A., Wang, R., 2019. Bootstrap Aggregating and Random Forest. *Macroeconomic Forecasting in the Era of Big Data*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-31150-6_13

Li, C. B., Xiao, K. Y., Li, N., et al., 2020. A Comparative Study of Support Vector Machine, Random Forest and Artificial Neural Network Machine Learning Algorithms inGeochemical Anomaly Information Extraction. *Acta Geoscientia Sinica*, 41(2): 309－319 (in Chinese with English abstract).

Li, H. X., Gao, M. Q., Ji, X. H., et al., 2025. Machine Learning-Based Tectonic Discrimination Using Basalt Element Geochemical Data: Insights into the Carboniferous-Permian Tectonic Regime of Western Tianshan Orogen. *Minerals*, 15(2): 122. https://doi. org/10.3390/min15020122

Liu, J. Q., Zhong, S. H., Li, S. Z., et al., 2023. 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. *Northwestern Geology*, 56(6): 41－56 (in Chinese with English abstract).

Liu, X. L., Li, C, Wang, T. S., 2022. An Improved Support Vector Machine Model for Personal Credit Assessment. *Computer Science and Technology*, 1(1): 27－34 (in Chinese with English abstract).

Liu, X. S., Zhang, Z. B., 2019. Parameter Optimization of Support Vector Machine Based on Improved Grid Search Method. *Journal of Jiangxi University of Science and Technology*, 40(1): 5－9 (in Chinese with English abstract).

Loader, M. A., Nathwani, C. L., Wilkinson, J. J., et al., 2022. Controls on the Magnitude of Ce Anomalies in Zircon. *Geochimica et Cosmochimica Acta*, 328: 242－257. https://doi.org/10.1016/j.gca.2022.03.024

Loader, M. A., Wilkinson, J. J., Armstrong, R. N., 2017. The Effect of Titanite Crystallisation on Eu and Ce Anomalies in Zircon and Its Implications for the Assessment of Porphyry Cu Deposit Fertility. *Earth and Planetary Science Letters*, 472: 107－119. https://doi. org/10.1016/j.epsl.2017.05.010

Loucks, R. R., Fiorentini, M. L., 2023. Oxidation of Magmas during Gain and Loss of H_2O Recorded by Trace Elements in Zircon. *Earth and Planetary Science Letters*, 622: 118377. https://doi.org/10.1016/j.epsl.2023.118377

Loucks, R. R., Fiorentini, M. L., Henríquez, G. J., 2020. New Magmatic Oxybarometer Using Trace Elements in Zircon. *Journal of Petrology*, 61(3): egaa034. https://doi.org/10.1093/petrology/egaa034

Lu, Y. J., Loucks, R. R., Fiorentini, M., et al., 2016. Zircon Compositions as a Pathfinder for Porphyry Cu±Mo±Au Deposits. *Society of Economic Geologists*. 329－347. https://doi.org/10.5382/sp.19.13

Luo, L., Chen, G. X., Li, Z. H., 2024. Identifying Tectonic Settings of Porphyry Copper Deposits Using Zircon Trace Elements-a Semi-Supervised Machine Learning Method. *Ore Geology Reviews*, 171: 106170. https://doi.org/10.1016/j.oregeorev.2024.106170

Meng, X. Y., Kleinsasser, J. M., Richards, J. P., et al., 2021. Oxidized Sulfur-Rich Arc Magmas

Formed Porphyry Cu Deposits by 1.88 Ga. *Nature Communications*, 12: 2189. https://doi.org/10.1038/s41467-021-22349-z

Meng, X. Y., Mao, J. W., 2025. Temporal Distribution of Porphyry Copper Deposits and the Influencing Factors. *Acta Petrologica Sinica*, 41(12): 4164－4177 (in Chinese with English abstract).

Meng, X. Y., Mao, J. W., Simon, A., et al., 2024. Contrasting Tectonomagmatic Conditions for Coexisting Iron Oxide-Apatite Deposits and Porphyry and Skarn Cu±Au Deposits in the Middle-Lower Yangtze River Metallogenic Belt, China. *Economic Geology*, 119(5): 1059－1087. https://doi.org/10.5382/econgeo.5084

Meng, X. Y., Simon, A. C., Kleinsasser, J. M., et al., 2022. Formation of Oxidized Sulfur-Rich Magmas in Neoarchaean Subduction Zones. *Nature Geoscience*, 15(12): 1064－1070. https://doi.org/10.1038/s41561-022-01071-5

Mungall, J. E., 2002. Roasting the Mantle: Slab Melting and the Genesis of Major Au and Au-Rich Cu Deposits. *Geology*, 30(10): 915. https://doi.org/10.1130/0091-7613(2002)0300915:rtmsma＞2.0.co;2

Nathwani, C. L., Wilkinson, J. J., Fry, G., et al., 2022. Machine Learning for Geochemical Exploration: Classifying Metallogenic Fertility in Arc Magmas and Insights into Porphyry Copper Deposit Formation. *Mineralium Deposita*, 57(7): 1143－1166. https://doi.org/10.1007/s00126-021-01086-9

Nathwani, C., Blundy, J., Large, S. J. E., et al., 2024. A Zircon Case for Super-Wet Arc Magmas. *Nature Communications*, 15: 8982. https://doi.org/10.1038/s41467-024-52786-5

Pan, B. D., Li, B., Wang, F. C., et al., 2022. Quantitative Study on Spatial Correlation of Different Types of Deposits (Occurrences): A Case Study of Indochina Peninsula and Adjacent Areas. *Geology and Exploration*, 58(3): 532－544 (in Chinese with English abstract).

Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825－2830. https://doi.org/10.48550/arXiv.1201.0490

Petrelli, M., Perugini, D., 2016. Solving Petrological Problems through Machine Learning: The Study Case of Tectonic Discrimination Using Geochemical and Isotopic Data. *Contributions to Mineralogy and Petrology*, 171(10): 81. https://doi.org/10.1007/s00410-016-1292-2

Piquer, J., Sanchez-Alfaro, P., Pérez-Flores, P., 2021. A New Model for the Optimal Structural Context for Giant Porphyry Copper Deposit Formation. *Geology*, 49(5): 597－601. https://doi.org/10.1130/g48287.1

Pizarro, H., Campos, E., Bouzari, F., et al., 2020. Porphyry Indicator Zircons (PIZs): Application to Exploration of Porphyry Copper Deposits. *Ore Geology Reviews*, 126: 103771. https://doi.org/10.1016/j.oregeorev.2020.103771

Pizarro, H., Rousse, S., Bouzari, F., et al., 2024. Use of Porphyry Indicator Zircons (PIZs) in the Sedimentary Record as an Exploration Tool for Covered Porphyry Copper Deposits in the Atacama Desert, Chile. *Journal of Geochemical Exploration*, 256: 107351. https://doi.org/10.1016/j.gexplo.2023.107351

Reunanen, J. 2003. Overfitting in Making Comparisons Between Variable Selection Methods. *Journal of Machine Learning Research*, 3(3): 1371－1382.

Richards, J. P., 2003. Tectono-Magmatic Precursors for Porphyry Cu-(Mo-Au) Deposit Formation. *Economic Geology*, 98(8): 1515－1533. https://doi.org/10.2113/gsecongeo.98.8.1515

Richards, J. P., Kerrich, R., 2007. Special Paper: Adakite-Like Rocks: Their Diverse Origins and Questionable Role in Metallogenesis. *Economic Geology*, 102(4): 537－576. https://doi.org/10.2113/gsecongeo.102.4.537

Rudin, C., 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, 1(5): 206－215. https://doi.org/10.1038/s42256-019-0048-x

Sahli, H. 2020. An Introduction to Machine Learning. In: Laffly, D., ed., TORUS 1-Toward an Open Resource Using Services : Cloud Computing for Environmental Data. Wiley, Hoboken. https://doi.org/10.1002/9781119720492.ch7

Shen, P., Hattori, K., Pan, H. D., et al., 2015. Oxidation Condition and Metal Fertility of Granitic Magmas: Zircon Trace-Element Data from Porphyry Cu Deposits in the Central Asian Orogenic Belt. *Economic Geology*, 110(7): 1861－1878. https://doi.org/10.2113/econgeo.110.7.1861

Sillitoe, R. H., 1972. A Plate Tectonic Model for the Origin of Porphyry Copper Deposits. *Economic Geology*, 67(2): 184－197. https://doi.org/10.2113/gsecongeo.67.2.184

Sillitoe, R. H., 1993. Gold-Rich Porphyry Copper Deposits: Geological Model and Exploration Implication. Geological *Association of Canada-Special Paper*, 40:465－478.

Sillitoe, R. H., 2010. Porphyry Copper Systems. *Economic*

Geology, 105(1): 3－41. https://doi.org/10.2113/gsecongeo.105.1.3

Song, Y. Y., Lu, Y., 2015. Decision Tree Methods: Applications for Classification and Prediction. *Shanghai Archives of Psychiatry*, 27: 130－135. https://doi.org/10.11919/j.issn.1002-0829.215044

Stonadge, G., Miles, A., Smith, D., et al., 2023. The Volatile Record of Volcanic Apatite and Its Implications for the Formation of Porphyry Copper Deposits. *Geology*, 51(12): 1158－1162. https://doi.org/10.1130/g51461.1

Sun, H. L., Tang, R., Li, Y., et al., 2026. Three-Dimensional Mineral Prospectivity Prediction Based on Mineralization Process Numerical Simulation and Machine Learning: A Case Study of the Maodeng Deposit in Inner Mongolia. Earth Science51(3): 921－939 (in Chinese with English abstract).

Sun, W. D., Huang, R. F., Li, H., et al., 2015. Porphyry Deposits and Oxidized Magmas. *Ore Geology Reviews*, 65: 97－131. https://doi.org/10.1016/j.oregeorev.2014.09.004

Teng, W. H., Pan, Z. F., Wang, R., 2025. Machine-Learning-Based Classification of Igneous Rock Type Using Titanite Trace Element Data. *Journal of Chengdu University of Technology* (*Science and Technology Edition*), 52(5): 873－887 (in Chinese with English abstract).

Trail, D., Watson, E. B., Tailby, N. D., 2012. Ce and Eu Anomalies in Zircon as Proxies for the Oxidation State of Magmas. *Geochimica et Cosmochimica Acta*, 97: 70－87. https://doi.org/10.1016/j.gca.2012.08.032

Wang, L. Y., Zhang, C., Geng, R., et al., 2023a. The Discrimination of Tectonic Settings Using Trace Elements in Magmatic Zircons: A Machine Learning Approach. *Earth Science Informatics*, 16(4): 4097－4112. https://doi.org/10.1007/s12145-023-01142-0

Wang, Y. Z., Hasanipanah, M., Rashid, A. S. A., et al., 2023b. Advanced Tree-Based Techniques for Predicting Unconfined Compressive Strength of Rock Material Employing Non‐Destructive and Petrographic Tests. *Materials*, 16(10): 3731. https://doi.org/10.3390/ma16103731

Wang, Y., Qiu, K. F., Hou, Z. L., et al., 2022. Quartz Ti/Ge-P Discrimination Diagram: A Machine Learning Based Approach for Deposit Classification. *Acta Petrologica Sinica*, 38(1): 281－290 (in Chinese with English abstract).

Watson, E. B., Harrison, T. M., 2005. Zircon Thermometer Reveals Minimum Melting Conditions on Earliest Earth. *Science*, 308(5723): 841－844. https://doi.org/10.1126/science.1110873

Wen, Z. H., Li, L., Kirkland, C. L., et al., 2024a. A Machine Learning Approach to Discrimination of Igneous Rocks and Ore Deposits by Zircon Trace Elements. *American Mineralogist*, 109(6): 1129－1142. https://doi.org/10.2138/am-2022-8899

Wen, Z. H., Xu, B., Kirkland, C. L., et al., 2024b. A Machine Learning Zircon Trace Element Tool to Predict Porphyry Deposit Type and Resource Size. *Journal of Geophysical Research: Machine Learning and Computation*, 1(4): e2024JH000180. https://doi.org/10.1029/2024JH000180

Yang, J. F., Qiao, P. R., Li, Y. M., et al., 2019. A Review of Machine‐Learning Classification and Algorithms. *Statistics and Decision*, 35(6): 36－40 (in Chinese with English abstract).

Yang, S. Q., Yang, W., Cui, T., et al., 2024. Prediction and Practical Application of Bauxite Mineralization in Wuzhengdao Area, Guizhou, China. *PLoS One*, 19(7): e0305917. https://doi.org/10.1371/journal.pone.0305917

Yuan, L. L., Chai, P., Hou, Z. Q., et al., 2025. Machine Learning for Characterizing Magma Fertility in Porphyry Copper Deposits: A Case Study of Southeastern Tibet. *Acta Geologica Sinica (English Edition)*, 99(2): 611－624. https://doi.org/10.1111/1755-6724.15277

Zhang, R., Wang, Y. B., 2016. Research on Machine Learning with Algorithm and Development. *Journal of Communication University of China Science and Technology*, 23(2): 10－18, 24 (in Chinese with English abstract).

Zhao, W. G., Chen, L., Chen, J. L., 2021. Application of Machine Learning Methods in Mineral Exploration. *World Nonferrous Metals*, (1): 1－5 (in Chinese with English abstract).

Zhong, R. C., Deng, Y., Yu, C., 2021. Multi-Layer Perceptron‐Based Tectonic Discrimination of Basaltic Rocks and an Application on the Paleoproterozoic Xiong'er Volcanic Province in the North China Craton. *Computers & Geosciences*, 149: 104717. https://doi.org/10.1016/j.cageo.2021.104717

Zhong, S. H., Liu, Y., Li, S. Z., et al., 2023. A Machine Learning Method for Distinguishing Detrital Zircon Provenance. *Contributions to Mineralogy and Petrology*, 178(6): 35. https://doi.org/10.1007/s00410-023-02017-9

Zhou, Y. Z., Zhang, Z. J., Yang, J., et al., 2022. Machine Learning and Singularity Analysis Reveal Zircon Fertility and Magmatic Intensity: Implications for Porphyry Copper Potential. *Natural Resources Re-*

search, 31(6): 3061－3078. https://doi.org/10.1007/s11053-022-10122-y

Zou, S. H., Chen, X. L., Brzozowski, M. J., et al., 2022. Application of Machine Learning to Characterizing Magma Fertility in Porphyry Cu Deposits. *Journal of Geophysical Research*: *Solid Earth*, 127(8): e2022JB024584. https://doi.org/10.1029/2022JB024584

中文参考文献

陈凯, 朱钰, 2007. 机器学习及其相关算法综述. 统计与信息论坛, 22(5): 105－112.

侯霖莉, 吴松, 易建洲, 等, 2024. 基于机器学习的绿泥石微量元素判别矿床类型. 地球科学, 49(12): 4303－4317.

姜娜, 杨海燕, 顾庆传, 等, 2019. 机器学习及其算法和发展分析. 信息与电脑, 31(1): 83－84, 87.

李苍柏, 肖克炎, 李楠, 等, 2020. 支持向量机、随机森林和人工神经网络机器学习算法在地球化学异常信息提取中的对比研究. 地球学报, 41(2): 309－319.

刘嘉情, 钟世华, 李三忠, 等, 2023. 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. 西北地质, 56(6): 41－56.

刘晓丽, 李程, 王天珊, 2022. 改进的支持向量机的个人信用评估模型. 计算机科学与技术, 1(1): 27－34.

刘小生, 章治邦, 2019. 基于改进网格搜索法的SVM参数优化. 江西理工大学学报, 40(1): 5－9.

孟旭阳, 毛景文, 2025. 斑岩铜矿时间分布规律与影响因素. 岩石学报, 41(12): 4164－4177.

潘北斗, 李宝, 王福臣, 等, 2022. 不同类型矿床(点)空间关联关系定量研究-以中南半岛及临近地区为例. 地质与勘探, 58(3): 532－544

孙会玲, 唐瑞, 李杨, 等, 2026. 基于成矿过程数值模拟与机器学习的三维找矿预测: 以内蒙古毛登矿床为例. 地球科学, 51(3): 921－939.

滕文航, 潘增锋, 王瑞, 2025. 基于机器学习的榍石微量元素特征判别岩浆岩类型. 成都理工大学学报(自然科学版), 52(5): 873－887.

王瑀, 邱昆峰, 侯照亮, 等, 2022. 石英Ti/Ge-P: 基于机器学习的矿床类型判别新图解. 岩石学报, 38(1): 281－290.

杨剑锋, 乔佩蕊, 李永梅, 等, 2019. 机器学习分类问题及算法研究综述. 统计与决策, 35(6): 36－40.

张润, 王永滨, 2016. 机器学习及其算法和发展研究. 中国传媒大学学报(自然科学版), 23(2): 10－18, 24.

赵伟光, 陈磊, 陈军林, 2021. 机器学习方法在矿产勘查中的应用. 世界有色金属, (1): 1－5.