DOI:10. 19657/j. geoscience. 1000-8527. 202509300086

基于机器学习的人工智能模型可解释性在矿产预测领域的研究进展及思考

白立科[1,2]，张玉茜[1,2*]，冯启伟[1,3]，刘丰武[1,2]，韩晓彤[1]，李积涛[1]，张文斌[1]，孟 圆[1,2]，徐文鹏[1,2]，曹睿卿[1,2]

（1. 山东省地质矿产勘查开发局第一地质大队（山东省第一地质矿产勘查院），山东 济南 250100；2. 自然资源部地质信息工程技术创新中心山东分中心，山东 济南 250100；3. 富铁矿勘查开发技术山东省工程研究中心，山东 济南 250100）

摘要：近年来，机器学习技术在矿产预测领域的应用显著提升了矿产资源勘探效率，但基于此构建的智能模型可解释性不足已成为制约其规模化应用和落地的核心挑战之一。本研究系统梳理了面向智能成矿预测的模型可解释性研究前沿进展，探讨了可解释性提升在矿产预测领域的研究意义及其在地质科学中的必要性和紧迫性，并基于不同阶段提出不同机器学习算法得出的预测结果可信度提升的方法和建议。现有研究主要围绕三个维度展开：重视地质先验知识的前期嵌入、利用可视化技术揭示全局或增强局部输出与输入值的关系和权重值、研发多种混合解释框架等。但当前高维非线性关系解译困难、机器学习算法普遍架构复杂、多源异构数据融合机理不透明等问题在矿产预测领域仍不容忽视，亟需开发兼顾计算效能与结果逻辑性的新型解释范式，所以本文基于对不同类型机器学习提出应考虑扩展时空框架、重视损失函数和可视化技术、合理选择评价参数等具体建议。未来发展方向应聚焦于构建建模全流程可解释性更高的智能预测系统，以提升基于不同阶段、不同模型所构建预测体系的逻辑置信度，推动机器学习模型与前沿成矿理论的优化创新。

关键词：机器学习；成矿预测；可解释性；人工智能；模型构建

中图分类号：P61 **文献标志码**：A **文章编号**：1000-8527(2026)02-0590-17

Advancements and Considerations on the Interpretability of Machine Learning-Based Artificial Intelligence Models in Mineral Prediction

BAI Like[1,2], ZHANG Yuqian[1,2*], FENG Qiwei[1,3], LIU Fengwu[1,2], HAN Xiaotong[1], LI Jitao[1], ZHANG Wenbin[1], MENG Yuan[1,2], XU Wenpeng[1,2], CAO Ruiqing[1,2]

(1. *No.* 1 *Geological Team of Shandong Provincial Bureau of Geology and Mineral Resources* (*Shandong No.* 1 *Institute of Geology and Mineral Resources Exploration*), *Jinan*, *Shandong* 250100, *China*; 2. *Technology Innovation Center of Geological Information*, *MNR Shandong Subdivision*, *Jinan*, *Shandong* 250100, *China*; 3. *Shandong Engineering Research Center of Rich Iron Ore Exploration and Development Technology*, *Jinan*, *Shandong* 250100, *China*)

Abstract: In recent years, the application of machine learning technology in mineral prospecting has significantly improved the efficiency of mineral resource exploration. However, insufficient model interpretability has

收稿日期：2025-10-12；**修回日期**：2026-01-03；**编辑**：袁茂文。

作者简介：白立科，男，硕士，工程师，1994年出生，主要从事矿产勘查与人工智能矿产预测研究工作。Email：495093084@qq. com。

***通信作者**：张玉茜，女，硕士，正高级工程师，1982年出生，主要从事测绘地理信息、计算机大数据、人工智能地质找矿等方向研究。Email：kittyzyq@163. com。

引用格式：白立科，张玉茜，冯启伟，等. 基于机器学习的人工智能模型可解释性在矿产预测领域的研究进展及思考［J］. 现代地质，2026，40（2）：590-606. DOI：10. 19657/j. geoscience. 1000-8527. 202509300086.

BAI Like, ZHANG Yuqian, FENG Qiwei, et al. Advancements and Considerations on the Interpretability of Machine Learning-Based Artificial Intelligence Models in Mineral Prediction［J］. GEOSCIENCE, 2026, 40（2）: 590-606. DOI: 10. 19657/j. geoscience. 1000-8527. 202509300086.

become one of the core challenges restricting its large-scale application and implementation. This study systematically reviews the cutting-edge progress of model interpretability research, discusses the research significance of improving interpretability in mineral prospecting and its necessity and urgency in geological science, and proposes methods and recommendations for enhancing the credibility of prediction results derived from different machine learning algorithms at various stages. Existing research primarily revolves around three dimensions: focusing on geological prior knowledge; utilizing visualization techniques to reveal relationships and weight values between global/local outputs and input features; and developing diverse hybrid interpretation frameworks. Nevertheless, critical challenges remain in mineral prospecting applications, including difficulties in decoding high-dimensional nonlinear relationships, inherent complexity of machine learning architectures, and opaque mechanisms of multi-source heterogeneous data fusion. There is an urgent need to develop novel explanatory paradigms that balance computational efficiency with logical consistency. Based on those approaches, this paper proposes specific recommendations such as extending the space-time framework, emphasizing loss functions and visualization techniques, and reasonably select evaluation metrics. The future development direction should prioritize building intelligent prediction systems with enhanced end-to-end interpretability throughout the modeling process, to enhance the logical confidence of prediction frameworks built on different environment, thereby facilitating the optimization and innovation of machine learning models in synergy with frontier metallogenic theories.

Key words: machine learning; metallogenic prediction; interpretability; artificial intelligence; model construction

0　引　言

随着人工智能技术的快速发展，机器学习与深度学习算法已逐步成为地学成矿预测领域的重要工具，特别是在目前国内大部分地区各类矿床的工作重点已从浅部、易识别、成矿模式研究成熟地区找矿逐步转向深部、边部隐伏矿探索挖掘的背景下，机器学习因其能够充分捕捉细小、多元、复杂且非线性的成矿特征，而被广泛应用于各种矿产的成矿预测和资源评价中[1-9]。传统基于统计学的成矿规律研究受限于地质系统非线性特征和数据维度约束，而人工智能算法通过挖掘多源异构地质数据中的隐式关联，深度分析非结构化数据，显著提升了矿产资源预测精度[10-15]，为智能找矿预测提供了全新的思路和视角。然而，当前研究更多聚焦模型性能优化，却忽视了对各类模型“黑箱”决策机制的解读——这一矛盾不仅关系各类模型在实际应用过程的推广开发，而且直接影响到成矿理论认知深化与找矿靶区实际验证[16-19]。

模型可解释性作为人工智能伦理的核心议题，已在医疗诊断[20-21]、金融风控[22-23]及自动驾驶[24]等领域形成系统性研究框架。但在地质勘查领域，现有解释性方法仍面临双重挑战：一方面，基于地学思想及成矿理论方面，地质过程多尺度耦合特性要求解释结果必须符合构造演化、元素运移等先验理论，各区域都具有特征成矿模式指导机器学习和模型建立，但在实际数据处理及算法构建过程中深度融合了复杂学习模型和数学理念，难免脱离地学先验思维[25-28]；另一方面，面对地质信息匮乏地区或在研究前期，稀疏采样条件下生成的特征图谱易产生地质学无意义的伪相关，模型欠拟合或过拟合现象经常出现，机器“幻觉”现象难以避免，并且隐式建模中的样本随机插值算法对于模型可解释性评价提出更高的要求[29-32]，过程的过于复杂及不透明性导致推广难度较大，所以构建可信度较高、可解释性强的人工智能模型成为人们关注的重点。但对于如何提高模型可解释性这一问题未进行归纳总结和梳理，且尚未建立适用于二维或三维成矿系统的统一评估标准[33-36]。在此背景下，开展人工智能成矿预测模型的可解释性研究，不仅有助于突破当前“知其然而不知其所以然”的技术瓶颈，更能为人工智能赋能下的成矿模型创新提供知识-数据双驱动的新范式[4-7]。

本文系统梳理了该领域的研究现状，分析现

有解释方法的适用边界，分类总结目前各类人工智能成矿预测模型的可解释性解决方法和存在问题，并根据机器学习的复杂程度，基于人工智能建模不同阶段提出提升模型可解释性的思路、改进建议和对于未来建模优化的思考，讨论面向地质矿产预测领域的可解释性人工智能框架的发展方向。

1 人工智能模型的可解释性研究

在人工智能领域，可解释性与可信赖程度的相关研究仍然处于初级阶段，可解释性定义为让人们能够理解和信任人工智能系统做出决策过程和行为的性能[37]。早期的白盒模式人工智能模型以及半智能化模型具备一定的可解释性，但是近期随着各个领域对于知识数据处理的需求大量增加，其内部运算过程越来越无法平衡透明性和效率[38]。机器学习算法通常具有较低的模型复杂度，在处理简单、线性等中等复杂程度数据时表现较好，但可能无法捕捉高维、非线性关系；而深度学习人工智能模型因其网络由大量非线性复杂网络层组成，能够通过各种内部组合对多源异构数据基于各种标准提取特征向量，因而能够在图像识别、自然语言分析与处理、特征提取与分类等相关领域取得重大进展[39-41]。然而正是因为这种复杂程度异常高、非透明化、大量数学参数计算融合的特点让其变得如“黑盒”一样（图1），通常让人无法理解模型做出预测结果的动机和机理，也就无法进行下一步的验证和推广[42]。

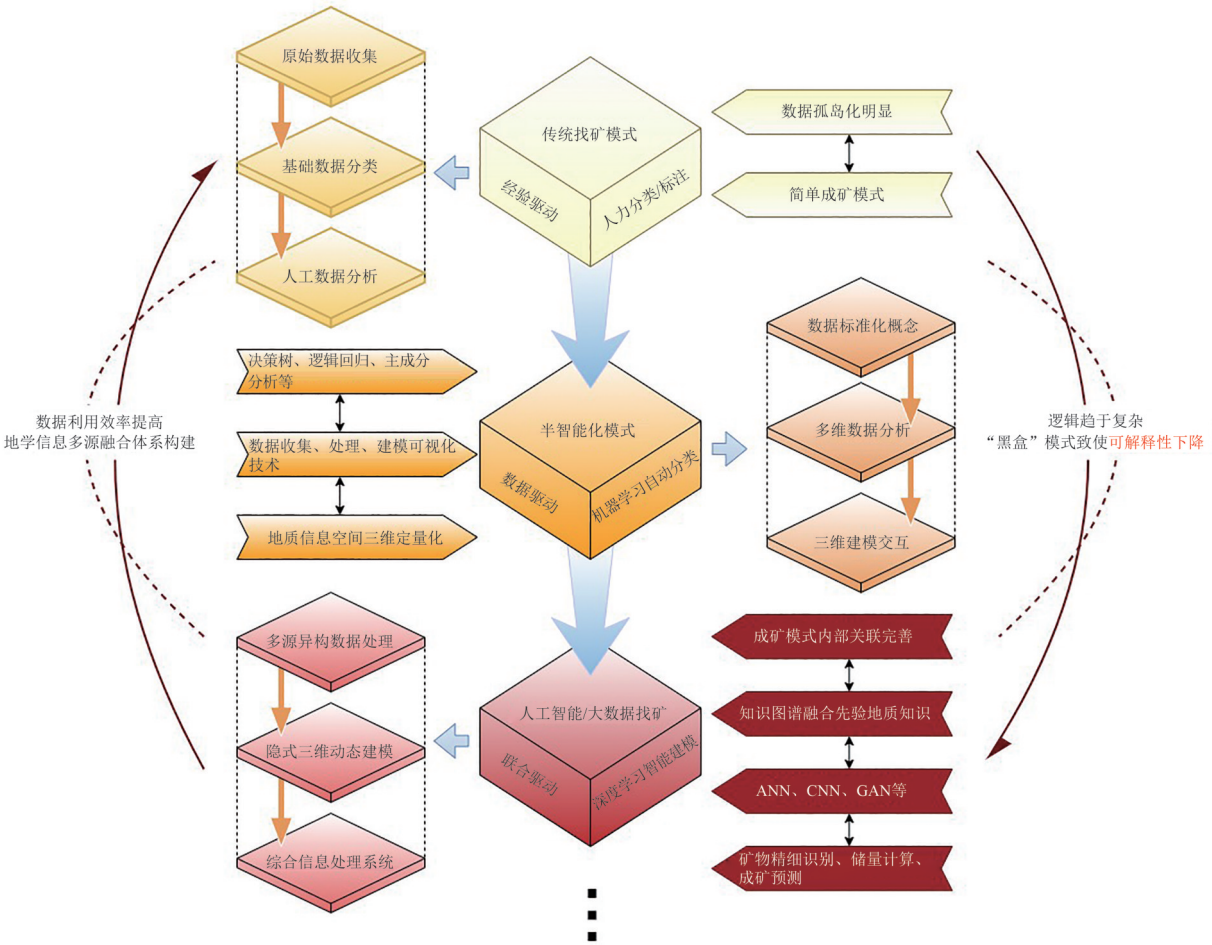


图1 人工智能应用在地学领域的发展

Fig. 1 Developments in artificial intelligence applications for geosciences

王冬丽等[43]将可解释性模型分为机器学习模型可解释性、深度学习模型可解释性及混合模型可解释性，未来发展和改进方向包括双驱动决策模型建立、跨膜太可解释框架及基于微分方程建模等；廖勇等[17]认为，人工智能下的机器学习始终存在“黑箱”效应，并将可解释性人工智能关键技术划分为四大类：模型、解释方法、安全测试以及实验验证。随着深度学习方法的流行，研

究者们开始探索如何增强深度学习模型的可解释性，并逐渐从简单结构机器学习模型发展至混合模型以兼顾高性能和可解释性。闵超[44]等通过对油气领域中的机器学习特征和进展研究，认为机器学习必须保证建模前、中、后期的全过程可解释性；王培晓[45]等深入探讨地理空间智能预测时空依赖机制，引入“地理空间智能预测基础算子”概念，通过不同基础算子分析模型内部结构与功能关系，并指出黑盒特征使得模型内部表征学习过于复杂，缺乏有效神经元隐语义解释，且可解释性评估指标也极度缺乏。

2　矿产预测领域中可解释性模型构建意义

目前数据驱动的人工智能成矿预测模型和机器学习算法在勘探程度较高、数据丰富、成矿规律明确、成矿模式清晰的地区应用较为成熟，以监督学习和半监督学习为主[46-47]。

比如随机森林算法（RF）具有出色的抗噪能力和异常值处理能力，也因其良好的泛化性、稳定性和可解释性，被广泛地应用于矿产定量预测领域[34][39]，简富源等[48]基于多尺度异常解析的控矿要素特征空间，重点针对多源异构地质数据内的隐式矿化关联信息和特征进行算法设计，利用能谱密度-面积多重分形滤波方法有效分离地质背景和致矿异常，构建RF集成学习框架实现矿化信息融合，生成矿产预测效果图。该算法无需对分类对象进行预处理，从而简化了相关数据的预处理流程。在许多方面，RF表现出色，但仍存在一些不足之处。例如，在特征选择过程中，RF会从数据集中随机选取特征，且其参数选择需要手动设置。这些功能虽然不会直接导致分类结果的误差增加，但从实际应用的角度来看，有必要进一步提升算法输出高质量特征的能力，优化参数选择，进一步减少模型的泛化误差，提高分类精度；决策树（DT）是一种基于树结构框架的学习模型，由内部节点和叶节点组成，分别对应于判断条件和决策结果，结构直观，易于理解和实现模型建立，且面对多源异构数据有很好的兼容性，因而具有一定的可解释性[49]，但在面对复杂非线性关系的数据以及数据量较大且特征维度较高时其建模效率有所下降；支持向量机（SVM）则是一种在统计学习理论框架和结构风险最小化原则下提出的机器学习方法[34]，能够通过最大化分类间隔的超平面来提升线性分类器的泛化能力，引入核函数技巧可以解决非线性分类问题，目前更多运用于解决小样本和高维度分类问题[15]。

在面对不同地质背景的成矿研究区，不同算法的适应度以及泛化能力也会有很大差别，这对于研究者来说，同时考虑算法参数的合理性、数据集构建的原则以及模型的过拟合问题，每个环节融合更多的数学算法来提高预测准确性和效率的同时，都对算法可解释性提出了挑战。娄渝明等[50]基于隐式三维建模，通过设计逻辑回归算法（LR）、SVM、人工神经网络算法（ANN）以及RF对从三维矿床块体模型提取的地质成矿信息进行训练和预测，并指出RF训练数来的模型方差小、泛化能力较强，且预测精度更高；Maitre[51]等在矿物颗粒识别过程中对比了RF、DT、K值邻近法等算法分类效率和精度，认为RF与超像素分割法结合具有稳定性、准确率高且可避免欠拟合；Kalashnikov等[52]利用了多元线性回归、ANN、RF及多元自适应回归（MARSpline）以通过全岩地球化学数据预测矿物特性，证实在内部关联性强的岩浆-热液型矿床中MARSpline方法效率最高；Rodriguez等[11]则在西班牙Rodalquilar金矿区矿物潜力预测中对比了ANN、回归树（RT）、RF及SVM的准确性、敏感性、泛化能力及可解释性，认为RF模型稳定性较高、调参敏感性低、预测性能最优。

然而，随着研究的深入和找矿领域需求的不断增加，且矿床形成本身为小概率事件，所以在各类不同区域、不同成矿体系和成矿条件下，基于机器学习的人工智能矿产预测模型面临的问题趋于多样化和多元化[53-54]，机器学习中的深度学习比如ANN、卷积神经网络（CNN）、生成对抗网络（GAN）等在不同阶段被研究者们重视起来并逐渐参与地质建模和矿产预测的各个环节[55]，包括数据整理、数据标准化、数据分类和标注、数据降维、样本补充、算法训练集的有效构建等，以应对各类复杂环境下的模型建立和找矿预测[56]，但同时建立可解释性模型成为不可避免的难题，传统因果关系驱动的地质思维难以理解复杂度过高的数理逻辑，各种预测结果的特征化表达也因数学复杂函数计算而变得难以解释[57]，比如CNN中卷积层的特征筛选与池化层的特征保留（图2）。

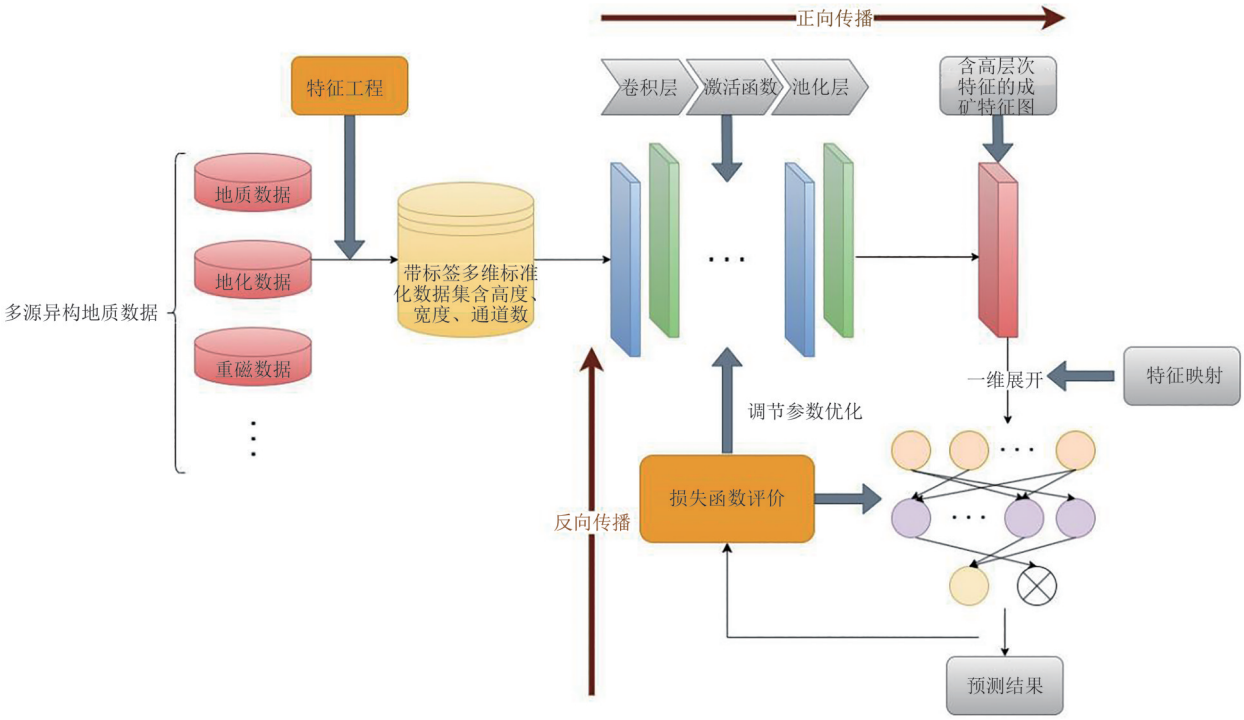


图2 基于卷积神经网络的矿产预测方法示意图（修改自文献［57］）

Fig. 2 Schematic diagram of CNN mineral prediction method （modified from reference［57］）

面对成矿预测中的具体问题，多数研究以成矿预测效率评价为主，过程及结果分析中还未重点考虑嵌入可解释性研究。比如针对常见的勘探程度较低的区域，需要处理不平衡样本集，客观系统建立研究区正负样本数据集成为一项重要工作[58]，李斌斌[59]基于多元地球化学综合异常信息，将注意力机制引入到所构建的1D CNN（一维卷积神经网络）模型中，构建基于注意力机制的CNN成矿预测模型；徐凯等[60]提出一种K均值聚类改进的CTGAN的见矿样本（正样本）扩充方法，结合随机欠采样负样本构建平衡数据集，通过Category Boosting（CatBoost）分类器建立成矿预测模型；Kost等[61]提出了WBIC算法，融合了Wald检验公式和贝叶斯变换，提升了传统LR的模型的精度和效率；Tang等[62]利用SMOTE（合成少数类过取样）方法应对样本稀疏问题，结合遥感解译、地球物理及地球化学多源信息，训练了4种机器学习模型进行铜-多金属矿化潜力预测结果对比；桂州等[63]认为不平衡数据分类难点集中于对少数类样本（通常为正样本）的识别问题，比如碎片数据、噪声数据干扰以及归纳偏置不合适，提出利用PCA（主成分分析）-SMOTE融合算法进行插值，通过“先去噪、后插值”的方法解决传统SMOTE插值破坏数据分布形态的技术瓶颈，为高维地质数据分类提供了兼顾降噪和数据平衡的新途径，最后通过分类器得到的分类效果显著提高；参数优化也成为单一模型预测精度和准确度下降常用的改进方法，缺少模型参数优化步骤，仅利用模型内置参数进行成矿预测[64]，往往在应对各种复杂地质条件、高特异性成矿条件，数据资料匮乏情况时还有很大的提升空间。

针对地下结构复杂，多源数据关系异质性高的情况，孔春芳等[65]引入自适应提升模型（Adaptive Boosting）深入分类每一个样本的细微差别，并通过弱学习分类器的分类误差率更新每个样本和弱学习器的权重，增加分类精度，并利用贝叶斯优化算法（BOA）提高模型的准确性、鲁棒性和可靠性，通过对比其他模型的AUC值、准确率、召回率等评价模型预测能力，同时也指出，集成学习的超参数调优被认为是一种黑箱优化，过程不可见使其可解释性有所下降，但BOA通过有效揭示黑盒函数可在一定程度提升可解释性；Jia等[66]将模型分类器分为基分类器和元分类器，利用多个分类器迭代学习分类结果，同时考虑过拟合和泛化问题，通过机器学习的堆叠（Stacking）方法对高维、多源数据进行精准化分类，在分类结果上与参数优化所达到的效果和作用机理基本类似，都是使得数据聚类更准确，并利用

GoCad 软件进行隐式三维建模；刘艳鹏等[9]通过利用CNN分析Pb元素浓度分布特征与矿体地下就位空间的耦合相关性，克服传统地质变量分析的局限性，为未探明区域提供了新找矿靶区指导方向；杨娜[57]提出基于CNN的多尺度融合特征提取方法，搭建智能成矿预测流程框架，对多源控矿因素数据进行多尺度成矿信息特征提取。

3　矿产预测领域中的可解释性研究

针对人工智能模型建模过程中出现的“黑盒”模式，在很多领域都有学者已经开始注意其对于模型开发和结果说明的影响，并尝试从不同维度和不同方向提出对策以提高其结果的可解释性[14][18]。而在地质及成矿预测领域针对模型可解释性研究与分析更多集中于前期地质先验知识的融合、建模过程可视化技术的应用以及混合模型开发等途径[67-70]（表1）。

一是基于先验地质知识嵌入的模型架构设计，通过物理约束条件与地质规律引导模型参数优化，在保证预测精度的同时增强模型可解释性。王成彬等[71]总结知识图谱研究进展并提出全要素知识图谱构建、利用图嵌入与聚类分析及异常信息提取三类融合知识图谱的矿产预测技术方法，认为地质知识增强的深度学习找矿模型和专家知识驱动的地质资源评价是目前主要研究方向；李楠等[72]提出了一种知识-数据联合驱动的可解释矿产资源智能预测方法，建模前期通过嵌入最佳-最差法分析工具，实现先验地质知识嵌入，利用Stacking集成学习算法提升预测准确性，再引入排列重要性（Permutation Importance，PI）、部分依赖图（Partial Dependence Plots，PDP）和与模型无关的局部解释（Local Interpretable Model-agnostic Explanations，LIME）算法生成可解释模型，最后通过准确率、召回率和F1分数评价模型并通过野外验证和样品化验进行评估，在最优化兼顾模型预测准确性和可解释性情况下完成预测研究闭环；叶育鑫等[73]基于综合信息矿产预测理论本体，采用自顶向下的方法构建了斑岩铜矿知识图谱，并通过丰度及可信度评估知识图谱是否存在过多错误信息和噪声干扰等正确性问题；左仁广等[70]提出从三个层面建立矿产预测全流程可解释人工智能模型，包括数据可解释性、模型可解释性以及结果可解释性，并指出地质先验知识的重要性。

表1　矿产预测模型可解释性提升方法

Table 1　Methods for enhancing interpretability of mineral prospectivity prediction models

主要方法	实现路径	可解释性核心
先验地质知识嵌入	知识图谱的应用与数据联合驱动	实现专家知识与数据驱动的结合，达到一定闭环预测框架
特征重要性评估及可视化	特征参数筛选和表征可视化	展示特征提取过程，空间维度与层级交互更透明
混合解释框架与验证	多种机器学习技术融合和改进优化	多重语言优势互补，解决一定过拟合和稀疏参数取样缺点及劣势

二是通过不同阶段特征重要性评估、模型蒸馏与可视化技术揭示全局与局部决策机制，建立矿物分布模式与输入变量间的关联规则。温鸿涛等[74]通过运用图分类技术中的实例分割模型，构建了YOLOv8模型，快速、精准重现锆石-石英假象晶形形态特征，并在特征提取过程中的每个处理阶段均生成特征图张量（Tensor）以更清晰展示特征提取过程，相比类激活图（CAM）通常仅对模型最后一层卷积层进行特征可视化，实例分割模型在可视化和模型可解释方面包含了更多信息，可以更直观揭示模型的识别机制；程艳等[75]为增强卷积神经网络的可解释性，提出LayerRPCAM的类激活映射算法，综合了LayerCAM算法和RelevanceCAM算法的优点，同时考虑空间维度和层级间的相互作用，生成更为清晰的类激活图，突破传统CAM对浅层神经元解释能力弱、背景噪声多以及可视化忠诚度低等问题的限制。

三是开发混合解释框架或者多重解释模型，集成事后解释工具与交互式分析系统，或辅助地质野外实际数据验证模型推理逻辑的合理性。Dong等[76]以DT为基底模型，构建了Deep Forest深度森林模型，通过动态调整层级数和模型复杂性，指出研究区内对于成矿影响最大的特征值，并利用DT本身的可解释性，在一定程度上解决了神经网络的“黑箱”问题；Li等[77]在构建前馈神经网络时嵌入PyBrain代码，并在双隐藏层中使用Sigmoid激活函数以减小目标函数与预测模型之间的均值和方差误差，同时使用PyBrain中内置交叉验证模块对训练集进行交叉验证，定量化表征最小交叉验证误差以提高模型“过拟合”问题和确

定最合适神经元数量；而有些机器学习算法其实考虑了在数据扩维或者插值增补过程中的可解释性，比如GAN包含生成器和判别器两个神经网络[78]，这两个网络在零和博弈中进行对抗训练，特别是利用有条件的生成对抗网络（CGAN）进行数据插值，其运行机制是通过结合地质观测数据生成更符合实际的样本数据以弥补数据集构建过程中正负样本不平衡的问题，在样本生成逻辑方面考虑了地质先验知识和地质逻辑，但其整体建模过程仍然属于黑盒模式，仅凭内在逻辑尝试理解每个特征值或者每个样本训练集的构成对于最终模型生成结果的具体影响是非常困难且耗时的。

4 矿产预测领域中提升模型可解释性方法

在本篇文章中，针对矿产预测领域人工智能模型的可解释性提升重点基于建模不同阶段、成矿预测领域适用性及模型本身透明化等原则，提出对于增强模型可解释性的系统化建议和思路，具体分别从前期知识与数据的处理和分析、模型构建过程中以及建模预测结果三个角度的可解释性进行分析（图3）。

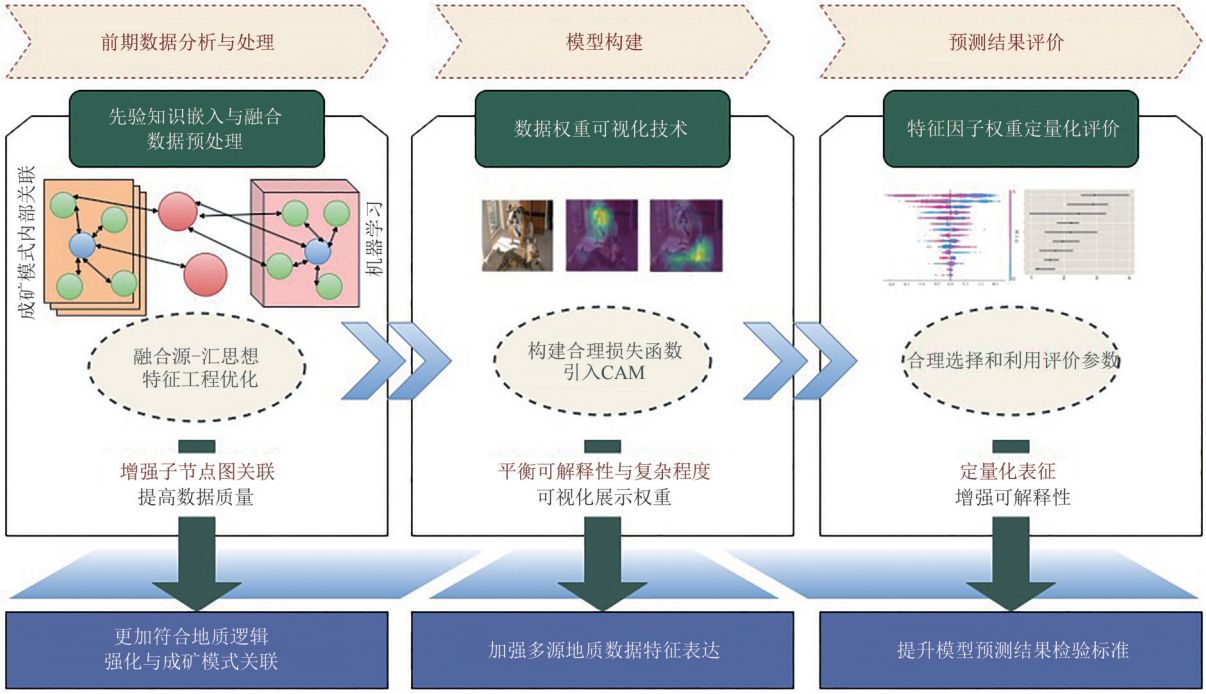


图3 可解释性模型优化示意图（部分图片引自文献［117］）

Fig. 3 Schematic diagram of optimization in interpretable model （figures partially reproduced from reference［117］）

4.1 基于前期思维强化的可解释性

在当前地质成矿规律研究与成矿区域预测中，基于先验知识的地学分析旨在总结相关成矿规律，因为缺乏与地学数据的结合而无法独立进行预测结果的可视化表达[38]；而单独的数据驱动矿产建模预测因未整合地学信息导致结果可信度下降[18][79]；同时深度学习过程中虽充分利用现有数据和海量多源数据，但由于缺乏可视化环节和算法解释，使得它们最终在预测结果可解释性上表现一般。

知识图谱将传统地质图中的点、线及多边形区域元素转化为文本属性，实现区域地质地层岩性、地质构造、火山活动、矿化现象等空间关联性与矿产分布特征和规律的有效可视化表征，不仅突破传统几何形状表征，还赋予地质知识的空间属性[73][80-81]。在现阶段矿产勘查与预测研究的过程中知识图谱的有效建立一方面对于地学先验知识的融合和基于不同建模数据要求的训练数据集向量化对齐起到推进作用，另一方面更是对于人工智能模型预测结果的指导检验，一定程度上提升了各种混合模型算法结果的可解释性[82-83]。所以越来越多的学者将目光集中在知识图谱的提升、拓展和可视化研究等相关技术上，以最大程度完善建模前期的系统思维框架和在可信度方面的矿产人工智能预测整体架构。

Qiu等[84]指出目前地质信息具有高度复杂性和多样性，对于知识图谱的构建，传统的人工特

征元素选取和信息挖掘标注准确率和效率都已不适应当前的研究需要，并且针对历史地质数据大量都存在于地质报告中，以非结构化地质文本和地质图件（彩色栅格图像等）形式留存，通过图像矢量化框架和自然语言处理技术（NLP）在地质报告和地质剖面中提取语义信息、语义属性和语义关联关系并建立知识图谱。Yan等[85]提出在矿产预测过程中无论是通过先验地质知识的建模指导还是通过地质数据异常进行的模型预测，都需要与地质知识进行深度融合，并提出一种集成地质知识的矿产预测远景制图方法（MPM），使用node2vec算法将知识图谱节点转换为低维向量，并通过空间网格化与地球化学异常数据进行对齐融合，实现地质知识与地球化学的数据前向融合，最后利用DenseNet模型进行训练和金矿预测，在陕西饶峰研究区测试中表现出优异的效果，而通过嵌入地质知识融合步骤且适当的可视化技术对进行复杂人工智能算法预测前的数据进行预处理可以对提高预测结果的可解释性有一定帮助。

建立基于因果解释框架的知识图谱可以更好地提升节点之间联系的因果关系和关联强弱，同时融合扩展基于地质成因的时空关系，提升整体的可解释性。比如类似多构建于沉积地质学中的源–汇系统[86]，将多个时空体系和成因体系建立联系，从整体历史演化角度论述和展现沉积物从源到汇的具体过程，揭示沉积物经历物源区母岩风化、剥蚀、搬运、沉积、成岩过程中各个体系的耦合关系[87-88]，涵盖了各种地质资料的研究方法。而在知识图谱建立过程中，融合源–汇系统思想，对应增加数据获取范围，不仅仅包括研究区成矿条件及成矿模式，还应包含区域地质构造及地层年代信息等文献，具体可以通过利用GraphRAG框架融合大语言模型LLM筛选和识别更多相关的地质要素，在知识图谱聚类形成语义一致的主题子图时，添加以往各研究区基于机器学习及深度学习构建模型的预测结果，比如提炼“基于机器学习成矿预测”子结构的地质、地球物理、地球化学等信息的利用率、模型准确率及评估效果等，聚焦成矿预测领域，构建更具时效性、完整性、逻辑性和可解释性的知识图谱体系，并且具备地质信息及经验更新的时效性，配置增量构建机制，实时加强整体可信度[89]。如此不仅可以将地质先验知识与地质关联信息更加合理地融合，而且可以更加直观展现构建成矿模式过程中各部分的影响因素，实时补充最新的控矿要素研究成果，构建地质思维更为结构化的智能问答平台和系统，同时也为后续基于机器学习的空间智能模型建立和算法选择提供可解释性更高的建模基础。

另一方面，定量化表达在知识图谱构建过程中有很大的应用空间，定量化参数对比模型构建前的知识图谱更具有可解释性和可信度。周贞云等[90]认为图数据库以节点和边表示数据关系，在基于关系构建知识图谱过程中有一定优势，相关专利数据量较为庞大，针对利用海量专利信息提出了一种基于LLM的专利知识图谱（PKG）构建方法，利用TextRank算法提取领域关键词并通过结合PageRank算法对应用领域的影响力进行定量化打分评级，解决传统PKG构建效率低且依赖人工标注的问题，在知识图谱建立过程中一定程度上做到了海量数据的有效扩充并且实现了关键步骤的知识抽取、知识融合及知识加工的可视化，增强最终模型的可解释性。

现如今隐式三维建模技术已在油气资源开发利用、矿产普查和成矿预测及地下空间研究等领域得到广泛利用[91-93]，特别是面对多源异构数据，人工智能学习模式针对海量数据的学习、分类、管理及标准化的处理相比单纯依靠人工完成效率提升巨大，而隐式三维建模软件和平台的应用推广对于重现重要地质体在空间中的构造特征、物质组成、成岩演化及矿物组分分布起到重要作用，同时随着数学地质及空间分析方法的发展，也推动了三维预测方法不断丰富。但随着建模软件和算法的成熟，矿产预测趋于高效率、低透明化的黑盒模式，模型预测精度和成矿预测算法复杂性的矛盾一直存在。而无论是机器学习、深度学习建模预测还是隐式三维建模成矿预测，特征信息提取和预处理一直是预测的关键因素，其对应的特征工程研究成为数据预处理中不容忽视的部分。

特征工程是指在进行成矿预测机器学习算法构建之前，通过数学算法、统计方法或先验知识对非结构化数据进行清洗、降噪、筛选、优化，针对传统机器学习算法，是减少模型训练时间、降低过拟合风险、提高机器学习模型性能和泛化能力的关键步骤[77]。比如可构建特征向量以应对数据集贫乏和正样本较少的问题，同时可以更好地说明不同地区矿产预测过程中不同影响因素的选择依据和样本构建依据，使得模型更具可解释性。并且虽然深度学习具有自动学习特征表征的优势，但是如果通过构建更加合理的训练集以表

征特征来进行训练和模拟，也可提高深度学习模型的预测能力[94]。

很多特征工程的构建对于深度学习模型来说相对简单，而且其可以用在建模和数据分析的各种阶段，对于提高模型整体的可解释性和透明度具有重要意义。Zhu等[95]认为，如果遗漏了重要特征数据，无论选择何种分类器，最终的分类结果也都不会太精细，所以他们构建了条件特征工程（Conditional Feature）框架，在构建机器学习分类算法训练数据集之前对多组特征数据进行提前分类、构建特征向量，并采用Relief算法和L1正则化逻辑回归进行特征选择，完成特征数据的筛选，提高模型预测的效率和精度；Dai等[96]利用皮尔逊相关系数（Pearson correlation coefficient）先进行数据相关性分析，通过利用简单函数（$|x|^{1/2}$，x^2，x^3，$\log(1+|x|)$）生成新的非线性数据集，并利用递归特征消除法（RFE）进行数据降维和数据集选取，使用交叉验证方法避免机器学习中的过拟合，使得最终LR对比SVM、自适应提升分类、DT以及RF分类的结果准确率提高11%，同时指出选择LR算法也是为了避免模型复杂化。可以看出在特定样本关系不明确且数量较少时，特征工程的利用在线性算法中可以成为一种未来选择，可以在提升预测模型准确率的同时，对扩展的数据集进行精准分类，一定程度上保持原有算法的表征和可解释性。

4.2 基于建模过程的可解释性

随着机器学习算法和人工智能模型的广泛运用，在不同领域各自都衍生出多种多样的机器学习类别，主要包括快速挖掘信息、数据分类及回归分析、数据降维等，但通常来说机器学习以结构风险最小化为优化目标[97]，具体形式为：

$$L = \min_{f} \frac{1}{n} \sum_{i=1}^{n} L\theta(f(xi)) + \lambda R(f) \qquad (1)$$

其中，n是训练集样本的数量，L为损失函数参数，θ是参数向量，第二项$R(f)$则是正则化项，表示模型复杂程度，$\lambda \geqslant 0$是平衡经验风险和模型复杂程度的特征参数。所以从中可以看出损失函数的构建对于算法有效性有很大的影响（图4），同时

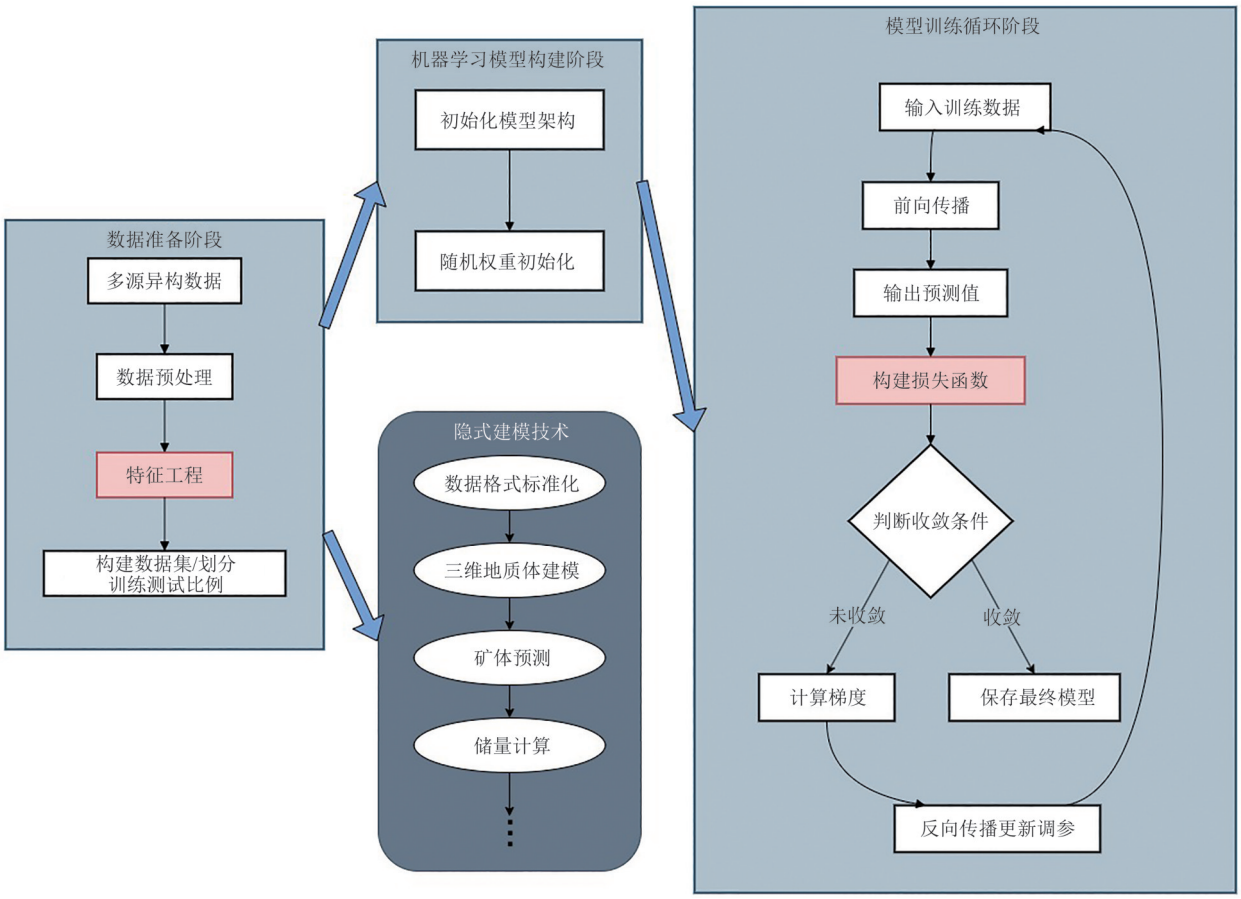


图4 矿产预测领域中的人工智能和机器学习模式

Fig. 4 Artificial intelligence and machine learning paradigms in mineral prediction

对人工智能模型可解释基础产生影响，因为不同机器学习算法不能简单随机选择任一的损失函数，而在当前的研究当中还未经重视。例如在传统机器学习算法中，分类问题是常见的成矿预测结果所面临的问题，0 ~ 1 损失函数的构建意味着预测值是否与标签特征一致，相同则损失值为 0，否则则为 1，未考虑预测值与真实标签表征的数值之间的差距或者是偏离程度，即当样本被误分类到另一边时，无论误差值多少，对应的损失值都相同，从而在一定程度上影响了模型的准确性和可解释性[98]；再比如在深度机器学习算法中，面对结果二分问题，在样本数量较少，正负样本数（表现在成矿预测中为矿点）严重失衡的情况下，大量负样本被分类而导致机器学习方向改变，尽管有很多方法对不平衡样本进行插值模拟或过采样矫正，但其过程的出现仍然在一定程度上降低了模型可解释性。

因此针对深度学习，基于避免公式过于复杂化的原则以及兼顾模型可解释性，同时考虑到我们在矿产预测研究利用深度学习建模过程中，更多利用到神经网络方法通过构建多层非线性计算单元，以地球物理、遥感图像等识别特征值为主进行“端到端”反向推演学习和预测，并且突破了传统克里金插值法等变异函数的依赖。而在其逐步拟合数据的过程中不可避免会面临输入图像会有很多种候选框，对其进行分类时只有少数包含真实个体，正样本远少于大部分容易分类的负样本，导致模型学习方向发生改变，背景影响程度过高。所以我们可以借鉴 Wang 等[97] 对于损失函数的研究，基于传统机器学习损失函数的 Sigmoid 交叉熵损失函数，在其架构中引入注意力机制[99-100]，其核心思想是通过一定途径获取每张特征图的权重差异，并将神经网络计算重点集中于场景预测中的关键部分[101]，即在建模过程中构建焦点损失函数，也可称为注意力函数，来解决因为背景类别不平衡导致的难易程度偏差。具体公式为：

$$L = -\bar{\tau}\,(1-\bar{p})^{\gamma}\log(\bar{p})$$

其中 γ 为用于调整损失减少的速度参数，当其为 0 是正好是 Sigmoid 交叉熵损失函数的表达公式。其中 τ 均表示一种条件概率，p 则表示被分配为正样本的概率，取值均为 0 至 1，从公式可以看出，容易分类的负样本具有较大的 $\bar{p}$ 和较小的损失值，需要注意的正样本则刚好相反。所以其可以使得负样本的权重得以降低，让模型更加关注稀少的正样本，适当降低其忠诚度，从而完成对于深度机器学习中损失函数的构建，适度的透明化以增强模型准确性和可解释性的平衡。

建模可视化技术提升则是另一种可解释性模型构建的改进途径。这里提到的可视化技术是对于人工智能模型可解释性研究较为直观表现方法的综合表达，包含多种不同类型的可视化。程渐[102] 融合文本抓取和 Selenium 网页自动化与改进的隐狄利克雷分布模型（LDA）对文献数据进行筛选、聚类和主题演化分析，绘制可视化海洋知识图谱；贺海洋[103] 等利用 CiteSpace 内置的算法模块，包括共现网络分析、突现性分析、中介中心性分析和聚类分析方法等，进行可视化图谱解读，探究地球化学与人工智能机器学习之间的融合发展现状和趋势。基于可解释性的建模过程，本文将可视化类别主要分为决策树可视化、聚类结果可视化、特征重要性可视化、张量图可视化以及数据增强可视化等，分类方式更加注重过程中对于数据处理和分析的增补解释，因为综合考虑简单机器学习和深度机器学习，所以没有对是否进行数据预处理或者标签标注分类进行划分，此种分类更加突出具体过程的可视化重点。其中决策树和聚类可视化以数据分布整体的可视化为主，更偏于线性分类；张量图可视化以处理高维数值张量的空间关系为主，更多应用于医学或者力学过程中的材料形变分析[104-105]；数据增强可视化更加注重结果的对比分析，与模型可解释性相关性较低[51]。

针对特征重要性可视化，随着 CNN 预测模型在不同领域中的广泛应用，国内外学者基于实现特征重要性可视化[106]，越来越重视不同卷积层数对于数据学习和分类的权重影响程度，因为在黑盒模式中，随着卷积层数的不断加深，特征图像的提取和分析是否准确一直缺少可视化表征，也就是说虽然在机器学习模型构建之前已经进行特征工程对标签进行标注、数据集优化及调整输入图像分辨率等，但如果存在多种特征图或者背景噪声过大，则很难判断特征提取的效率[107-108]，比如地质遥感信息中利用光谱数据提取岩性特征和蚀变特征，存在岩矿与植被混合光谱而被误提取导致矿产勘查预测区域远离成矿带问题[19]，并且通过多层卷积无法得知特征信息是如何被错误

提取的。所以基于矿产预测领域本文认为可以参考积分梯度法（Integrated Gradients）和CAM[109]，来判断模型做出分类和预测的依据是什么，增强模型的可解释性和稳定性（图5）。

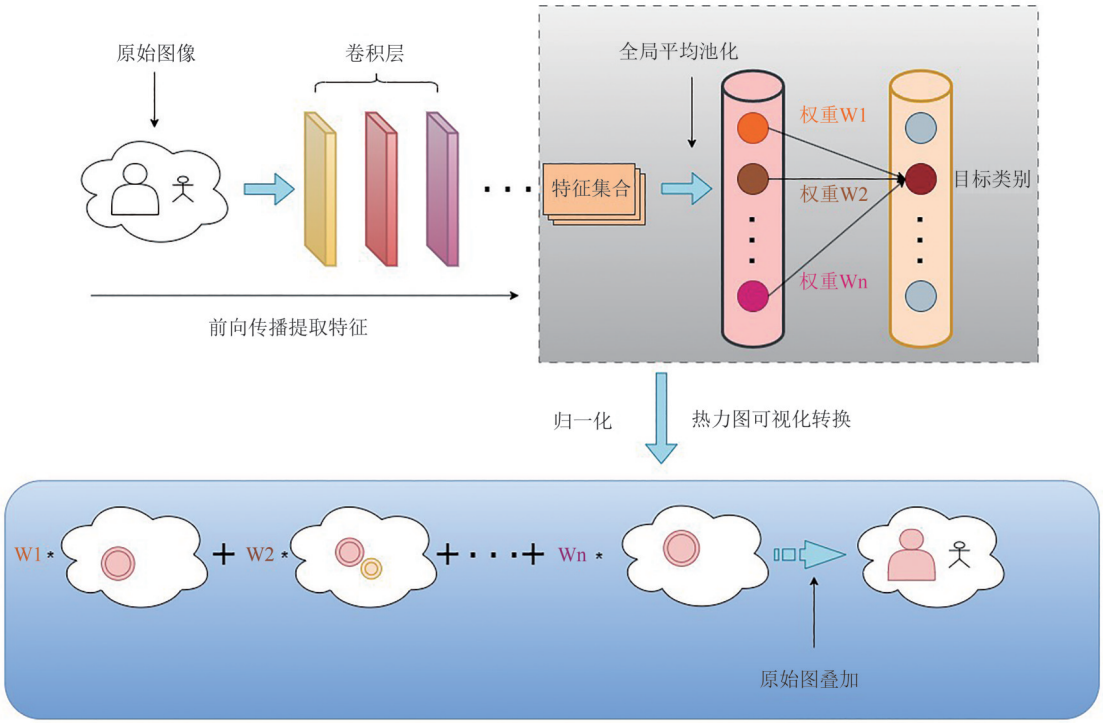


图5 类激活映射技术（CAM）流程示意图

Fig. 5 Class activation mapping （CAM） workflow schematic

IG利用神经网络反向传播过程的梯度信息，量化输入特征对于预测结果的边际贡献，更适合高精度领域[110]；而CAM由Zhou等[111]提出，通过全局平均池化层取代传统CNN的全连接层，与最后一层卷积特征图加权生成分类依据，可以生成热力图突出显示对分类决策和结果产生显著影响的特征区域，实现空间可视化定位，提高了模型可解释性。Fu等[112]通过详细的数理模型提升了梯度加权类激活映射法（Grad-CAM）的敏感性和真实性，构建了XGrad-CAM模型；Wang等[113]构建了Score-CAM新型卷积神经网络架构，旨在解决CAM存在的噪声和普遍存在的虚假置信度问题，引入通道置信度概念增强模型实际的相关性参数表达，抗干扰能力更强，稳定性提升；张国栋等[114]指出在模型可解释性方面，IG的固定反向传播路径可能会在过程中引入与目标无关的梯度，导致更多的无关背景噪声影响结果的准确性和可信度，所以提出一种基于导向积分梯度与多尺度融合的类激活映射方法GIG-CAM，以动态路径替代传统IG中的直线路径，降低噪声干扰，并且利用小尺度显著图递归法优化大尺度显著性图，提升分辨率的同时也保留原始图像尺度的定位能力。

GIG-CAM已经可以做到有效展示其模型决策结果的依据，但其模型分类错误而导致的预测结果偏差在实际应用中对于整体效果的影响其实难以估量，所以需要构建更精准的训练集。为避免模型复杂化以及突出数据集中各部分特征值权重占比对模型造成的影响，可利用双路径动态融合方法，评估各个分类特征在标签归属过程中的权重问题，在基础CAM架构中并行嵌入通道注意力分支和空间注意力分支，通过可变换参数α/β控制两者的贡献度：

$$F_{att}=\alpha\times F_c+\beta\times F_s \tag{2}$$

其中：F_c为通道增强特征，F_s为空间增强特征；同时考虑了利用互补特性提升模型的鲁棒性和实时性，增强CAM预测精度和分类效率的同时，避免了过于复杂的数理模式带来的更多可解释性问题。但同时应注意，此种嵌入方式也存在原模型架构基础上进行修改以及效率下降的风险。

4.3 基于预测结果验证的可解释性

建模预测结果后验证多源数据的计算权重在

模型可解释方面成为一种适用性很高的途径，一般通过轻微扰动、数值边际效应或数值替换判别和定量化表征[115-118]，主要优势在于可以进行数据定量化表达，直观解释模型做出决策的原因和逻辑。比如，Mou等[119]在矿产预测建模中选择了RF和SVM算法对研究区进行成矿预测，通过迭代自助法量化预测结果的不确定性，并采用模型无关方法加强机器学习模型结果的可解释性，具体包括排列特征重要性（PFI）、PDP值、个体条件期望图（ICE）及SHAP值。

PFI（Permutation Feature Importance）是一种用于评估特征对模型预测结果贡献的关键技术和方法，通过将目标特征的值随机打乱，重新计算模型在扰动后的性能变化幅度，性能下降越剧烈则表明该特征越重要[120]，以此量化单个特征值扰动对模型性能的影响，揭示该特征值的重要性及权重，具有高效性和鲁棒性，尤其在解释高复杂性“黑箱”模型中较为实用。

PDP（Partial Dependence Plot）是一种用于解释机器学习的可视化工具算法[121]，目的是为了展示预测模型中某一特定特征对预测结果的边际影响，一般用于解释RF、ANN等算法中特征与预测结果的关联，同时揭示目标变量与预测结果之间存在的线性、单一或更复杂的关系。

ICE（Individual Conditional Expectation Plot）也是一种用于可视化单一特征对模型预测结果影响的工具[122]，通过对目标特征进行系统性的数值替换，同时保持其他特征不变，重新计算模型对该样本的预测结果，并绘制每个样本在某一特征值变化时的预测结果变化趋势曲线图，揭示该特征对不同个体预测值的具体影响模式。

SHAP（Shapley Additive Explanations）是一种解释黑箱模型的框架，计算每个样本中每个特征的SHAP值来解释模型的预测结果[123]。同时也是一种基于博弈论的机器学习模型解释方法，用于量化每个特征因子对模型单个预测结果的贡献度。其核心思想是将模型的预测结果视为一种“合作博弈”的收益，并通过SHAP值公平分配各特征对预测结果的贡献，可以直观展示单一因子扰动对于模型输出的预测值的干扰程度[124]，在矿产建模和成矿预测中常被用来表征模型和机器学习的可解释性。

对于表征模型可解释性有多种定量化参数，但对于不同机器学习和深度学习模型，需要选择不同的合适参数进行可解释性的评价，比如也可以引入注意力机制，模仿人类大脑选择性专注于有用的信息，以注意力权重热图的形式体现模型在决策过程中赋予最高权重值的区域[43]；陈忠元[125]以揭示微量元素与不同矿床类型之间潜在的非线性、多维关联机制为研究目标，基于TPE-SVM法建立了闪锌矿微量元素分类模型，并利用SHAP方法进行特征重要性分析指出其中Mn、Ge、Co为矿床成因类型判别的三大关键元素。所以基于SHAP值的特征重要性量化[126-127]和注意力机制可视化[128]揭示控矿要素的空间响应关系，同时引入各种可解释模型和定量化指标，可能是对模型进行预测后解释和检验以最大程度平衡模型可信度和准确性的良性选择。

5　讨　论

国家自然科学基金委员会在2024年3月发布的《可解释、可通用的下一代人工智能方法重大研究计划2024年度项目指南》中指出，重大研究计划的核心科学问题包括通过规则与学习结合的方式，建立高精度、可解释、可通用且不依赖大量标注数据的人工智能新方法。开发下一代人工智能方法需要的数据库和模型训练平台，完善下一代人工智能方法驱动的基础设施。充分指出了未来我们不仅仅在地学领域要增强机器学习的可解释性，更要在各个科学研究过程中重视各类模型的可信度和泛化能力。

矿产预测是地学领域的重要分支，是基于多源数据及经验知识的复杂决策过程，在面对分析数据量快速增长、数据格式多元化、数据特征趋于模糊的诸多地学数据难题，人工智能算法在该领域的各个阶段逐渐普及推广，从一开始的证据权重法、模糊逻辑法到简单线性机器学习再到融合大数据、云计算、深度学习等，以加快数据智能化处理、增强数据特征提取、提升模型泛化能力和预测效率为主的矿产预测融合人工智能的新范式逐渐形成。而模型复杂程度和预测精度，与其可解释性和透明化地质逻辑思维一直存在着难以平衡的矛盾，同时也制约着知识-数据双驱动下的人工智能找矿模型探索。

同时，在面对不同区域、不同的地质条件和不同成矿体系时，从数据处理到人工智能算法搭

建以及预测模型建立已可以通过集成式平台或智能化系统完成，包括进行高效、快速前期数据集构建、科学筛选控矿因素、融合多种机器学习算法、智能插值地质建模以及成矿模式优化完善，但目前仍主要通过局部嵌入可解释性定量化计算以表达模型准确性和可行性，以预测值相关参数评价等为主，在不同地区的可解释性评价取得一定效果。相信将来随着人工智能技术发展及机器学习算法优化，集成式矿产预测研究平台在隐式建模的同时，综合考虑预测模型的前期知识图谱构建的成因思路、中期建模过程损失函数的构建和特征参数可视化表征以及后期预测结果的验证等，以期实现全过程可解释性优化提升，结合更多应用实例和前沿技术协助建立完善的评价标准，从而推动具有高效性、泛化强的矿产预测研究体系落地和推广。

无论在哪个研究领域，人工智能模型的可解释性研究始终需要数学、物理学、统计学、计算机等多个科学领域、多方面进行互相辅助验证，因为在构建模型的过程中，基于机器学习人机交互的背景下，不同领域的先验知识和经验对于模型可解释性的最终表征都具有重要意义，最终目的都是帮助研究人员和使用者构建足以使人信赖且预测逻辑合乎领域思维的学习算法。而如何判断模型可解释性的程度高低以及是否需要在各个领域总结关于可解释性相关规则对于不同算法系统的研究和推广至关重要，也是未来如何在以人的逻辑思维下正确推动人工智能发展所要解决的关键问题。

参考文献：

[1] 周永章，张良均，张奥多，等. 地球科学大数据挖掘与机器学习[M]. 广州：中山大学出版社，2018：1－269.

[2] 肖克炎，李程，唐瑞，等. 大数据智能预测评价[J]. 地学前缘，2025，32(4)：20－37.

[3] 周永章，肖凡. 管窥人工智能与大数据地球科学研究新进展[J]. 地学前缘，2024，31(4)：1－6.

[4] 叶舒婉，侯卫生，杨玠，等. 三维地质智能建模研究进展[J]. 地学前缘，2025，32(4)：182－198.

[5] 周永章，黎培兴，王树功，等. 矿床大数据及智能矿床模型研究背景与进展[J]. 矿物岩石地球化学通报，2017，36(2)：327－331，344.

[6] 张起，高静. 基于特征定量化预测的矿产资源勘查技术研究[J]. 中国非金属矿工业导刊，2024(1)：84－87.

[7] KONG Y，CHEN G，LIU B，et al. 3D mineral prospectivity mapping of Zaozigou gold deposit，west Qinling，China：Machine learning-based mineral prediction [J]. Minerals，2022，12(11)：1361.

[8] 肖克炎，孙莉，李楠，等. 大数据思维下的矿产资源评价[J]. 地质通报，2015，34(7)：1266－1272.

[9] 刘艳鹏，朱立新，周永章. 大数据挖掘与智能预测找矿靶区实验研究——卷积神经网络模型的应用[J]. 大地构造与成矿学，2020，44(2)：192－202.

[10] ZUO R，XIONG Y，WANG Z，et al. A new generation of artificial intelligence algorithms for mineral prospectivity mapping [J]. Natural Resources Research，2023，32(5)：1859－1869.

[11] RODRIGUEZ-GALIANO V，SANCHEZ-CASTILLO M，CHICA-OLMO M，et al. Machine learning predictive models for mineral prospectivity：An evaluation of neural networks，random forest，regression trees and support vector machines [J]. Ore Geology Reviews，2015，71：804－818.

[12] 赵忠海，崔晓梦，孙景贵，等. 基于三维地质-地球物理建模的深部成矿预测——以黑河地区永新金矿床为例[J]. 吉林大学学报(地球科学版)，2024，54(2)：498－515.

[13] 赵义来，刘亮明，胡荣国. 基于地质体产状特征及建模数据特点的三维地质建模方法探讨——以安徽月山矿田为例[J]. 桂林理工大学学报，2018，38(4)：752－760.

[14] 李章林，吴冲龙，张夏林，等. 地质科学大数据背景下的矿体动态建模方法探讨[J]. 地质科技通报，2020，39(4)：59－68.

[15] ALGAIARI M，HOSSAIN M，PETROVSKI A，et al. Applications of artificial intelligence in geothermal resource exploration：A review [J]. Deep Underground Science and Engineering，2024，3(3)：269－285.

[16] SHARMA A，SINGH P K，CHANDRA R. SMOTified-GAN for class imbalanced pattern classification problems [J]. IEEE Access，2022，10：30655－30665.

[17] 廖勇，韩小金，刘金林，等. 可解释性人工智能研究进展[J/OL]. 计算机工程，2025：1－28[2025-07-31]. https://doi.org/10.19678/j.issn.1000-3428.0069925.

[18] 周圣荃，李以科，王永志，等. 生成式AI技术在地学研究中的应用现状及发展趋势[J]. 地学前缘，2025，32(4)：303－316.

[19] 李志忠，陈圣波，刘德长，等. 近年国内外遥感地质找矿主要进展[J]. 西北地质，2025，58(3)：183－195.

[20] GU D，LI Y，JIANG F，et al. VINet：A visually interpretable image diagnosis network[J]. IEEE Transactions on Multimedia，2020，22(7)：1720－1729.

[21] YANG C，RANGARAJAN A，RANKA S. Global model interpretation via recursive partitioning[C]//2018 IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE，2018：1563－1570.

[22] 杨涛，黄定江，刘俊，等. "AI浪潮下金融机构的机遇与挑战"主题座谈会专家观点荟萃[J]. 新金融，2025(7)：39－46.

[23] WANG D，QUEK C，NG G S. Bank failure prediction using an accurate and interpretable neural fuzzy inference system[J]. AI Communications，2016，29(4)：477－495.

[24] SHAO H，WANG L，CHEN R，et al. Safety-enhanced autono-

mous driving using interpretable sensor fusion transformer[C]// Conference on Robot Learning. PMLR, 2023: 726 - 737.

[25] 张彤, 刘仁宇, 王培晓, 等. 感知物理先验的机器学习及其在地理空间智能中的研究前景[J]. 地球信息科学学报, 2023, 25(7): 1297 - 1311.

[26] 周邓, 曾广亮, 王洪荣, 等. 江西桃山罗布里南部地区三维地质建模与成矿预测[J]. 地质通报, 2022, 41(12): 2256 - 2264.

[27] 王帅. 基于Adaboost算法的成矿预测模型研究[D]. 北京: 中国地质大学(北京), 2016.

[28] 袁峰, 李晓晖, 田卫东, 等. 三维成矿预测关键问题[J]. 地学前缘, 2024, 31(4): 119 - 128.

[29] 王斌, 高永宝, 任涛, 等. 龙门山构造带辛家咀金矿三维原生晕特征及深部找矿预测[J/OL]. 地学前缘, 2025: 1 - 22[2025-07-31]. https://doi. org/10. 13745/j. esf. sf. 2025. 1. 17.

[30] WANG Y, COINT N, MANSUR E T, et al. Leveraging domain expertise in machine learning for critical metal prospecting in the Oslo Rift: a case study for Fe-Ti-P-rare earth element mineralization[J]. Minerals, 2024, 14(4): 377.

[31] ZHANG Z, WANG G, CARRANZA E J M, et al. An uncertainty-quantification machine learning framework for data-driven three-dimensional mineral prospectivity mapping [J]. Natural Resources Research, 2024, 33(4): 1393 - 1411.

[32] 高猛, 王功文, 唐冬梅. 新疆喀拉通克铜镍矿集区三维成矿预测与不确定性评价[J/OL]. 地学前缘, 2025: 1 - 22[2025-07-31]. https://doi. org/10. 13745/j. esf. sf. 2025. 4. 2.

[33] 董旭朝, 孙莉, 肖克炎, 等. 基于证据权和随机森林的山东微山稀土矿三维成矿预测[J/OL]. 地质通报, 2025: 1 - 24[2025-07-31]. https://link. cnki. net/urlid/11. 4648. P. 20250320. 1352. 012.

[34] GRANEK J. Application of machine learning algorithms to mineral prospectivity mapping[D]. Vancouver: University of British Columbia, 2016.

[35] 郭彬,郑颖超. 浅谈三维地质模型在露天矿山的应用[J]. 中国非金属矿工业导刊,2025(5):82 - 84,89.

[36] SHIRAZI A, HEZARKHANI A, SHIRAZY A, et al. Geochemical modeling of copper mineralization using geostatistical and machine learning algorithms in the Sahlabad area, Iran[J]. Minerals, 2023, 13(9): 1133.

[37] LIPTON Z C. The mythos of model interpretability[J]. ACM Queue, 2018, 16(3): 31 - 57.

[38] XIA Y, FENG Y. Prospects of Artificial Intelligence Application in the High-Quality Development of the Geological Exploration Industry[C]//2023 International Conference on Management Innovation and Economy Development (MIED 2023). Atlantis Press, 2023: 94 - 101.

[39] ZHANG B, XU Z, WEI X, et al. Deep subsurface pseudo-lithostratigraphic modeling based on three-dimensional convolutional neural network (3D CNN) using inversed geophysical properties and shallow subsurface geological model[J]. Lithosphere, 2024(1): lithosphere_2023_273.

[40] WASHIZAKI H, YOSHIOKA N. AI Security Continuum: Concept and Challenges[C]//2024 IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI (CAIN). Lisbon, Portugal: IEEE, 2024.

[41] LUO Z, FARAHBAKHSH E, HORE S, et al. An explainable semi-supervised deep learning framework for mineral prospectivity mapping: DEEP-SEAM v1. 0[J]. EGUsphere, 2025, 2025: 1 - 43.

[42] SHAW K O, GOÏTA K, GERMAIN M. Prospectivity mapping of heavy mineral ore deposits based upon machine-learning algorithms: Columbite-tantalite deposits in West-Central Côte d'Ivoire[J]. Minerals, 2022, 12(11): 1453.

[43] 王冬丽, 杨珊, 欧阳万里, 等. 人工智能可解释性: 发展与应用[J]. 计算机科学, 2023, 50(S1): 19 - 25.

[44] 闵超, 文国权, 李小刚, 等. 可解释机器学习在油气领域人工智能中的研究进展与应用展望[J]. 天然气工业, 2024, 44(9): 114 - 126.

[45] 王培晓, 张恒才, 张岩, 等. 地理空间智能预测研究进展与发展趋势[J]. 地球信息科学学报, 2025, 27(1): 60 - 82.

[46] JIANG S, ZHENG Y, SOLOMATINE D. Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning [J]. Geophysical Research Letters, 2020, 47(13): e2020GL088229.

[47] MUMUNI A, MUMUNI F. Automated data processing and feature engineering for deep learning and big data applications: a survey [J]. Journal of Information and Intelligence, 2025, 3(2): 113 - 153.

[48] 简富源, 张子鸣, 董岳霖, 等. 基于多重分形与随机森林的新疆哈巴河金矿成矿预测[J]. 地学前缘, 2025, 32(4): 78 - 94.

[49] BRESLOW L A, AHA D W. Simplifying decision trees: a survey[J]. The Knowledge Engineering Review, 1997, 12(1): 1 - 40.

[50] 娄渝明, 康旭, 赖渊平, 等. 隐式建模和机器学习算法在西藏巨龙斑岩型铜钼矿床三维成矿预测中的应用研究[J]. 地学前缘, 2025, 32(5): 440 - 455.

[51] MAITRE J, BOUCHARD K, BÉDARD L P. Mineral grains recognition using computer vision and machine learning[J]. Computers & Geosciences, 2019, 130: 84 - 93.

[52] KALASHNIKOV A O, PAKHOMOVSKY Y A, BAZAI A V, et al. Rock-chemistry-to-mineral-properties conversion: Machine learning approach [J]. Ore Geology Reviews, 2021, 136: 104292.

[53] MA C, KALE A S, ZHANG J, et al. A knowledge graph and service for regional geologic time standards [J]. Geoscience Frontiers, 2023, 14(5): 101453.

[54] BAI T, TAHMASEBI P. Hybrid geological modeling: Combining machine learning and multiple-point statistics[J]. Computers & Geosciences, 2020, 142: 104519.

[55] SUN K, CHEN Y, GENG G, et al. A review of mineral prospectivity mapping using deep learning [J]. Minerals, 2024, 14(10): 1021.

[56] ZHANG C, ZUO R. Recognition of multivariate geochemical

anomalies associated with mineralization using an improved generative adversarial network [J]. Ore Geology Reviews, 2021, 136: 104264.

[57] 杨娜. 多尺度特征融合算法及其在智能成矿预测中的应用[J]. 矿物岩石地球化学通报, 2025, 44(3): 478 - 491, 452.

[58] HARRIS J R, GRUNSKY E, BEHNIA P, et al. Data-and knowledge-driven mineral prospectivity maps for Canada's North [J]. Ore Geology Reviews, 2015, 71: 788 - 803.

[59] 李斌斌. 区域地球化学元素空间分布模式分析及其定量成矿远景预测[D]. 成都: 成都理工大学, 2021.

[60] 徐凯, 徐城阳, 吴冲龙, 等. 基于样本扩充的黔西北垭都-蝉硐矿区铅锌矿成矿预测研究[J]. 地学前缘, 2025, 32(4): 95-107.

[61] KOST S, RHEINBACH O, SCHAEBEN H. Using logistic regression model selection towards interpretable machine learning in mineral prospectivity modeling[J]. Geochemistry, 2021, 81(4): 125826.

[62] TANG J, ZHANG H, BAI R, et al. Mineral prospectivity mapping for exploration targeting of porphyry Cu-polymetallic deposits based on machine learning algorithms, remote sensing and multi-source geo-information [J]. Minerals, 2025, 15 (10): 1050.

[63] 桂州, 陈建国, 王成彬. 基于PCA-SMOTE-随机森林的地质不平衡数据分类方法——以东天山地球化学数据为例[J]. 桂林理工大学学报, 2017, 37(4): 587 - 593.

[64] 肖立志. 机器学习数据驱动与机理模型融合及可解释性问题[J]. 石油物探, 2022, 61(2): 205 - 212.

[65] 孔春芳, 田倩, 刘健, 等. 基于集成学习模型与贝叶斯优化算法的成矿预测[J]. 地学前缘, 2025, 32(4): 122 - 139.

[66] JIA R, LV Y, WANG G, et al. A stacking methodology of machine learning for 3D geological modeling with geological-geophysical datasets, Laochang Sn camp, Gejiu (China)[J]. Computers & Geosciences, 2021, 151: 104754.

[67] ZHANG Q, ZHOU Y, HE J, et al. A review on global cooperation network in the interdisciplinary research of geochemistry combined with artificial intelligence [J]. Minerals, 2023, 13 (10): 1332.

[68] YANG F, ZUO R, XIONG Y, et al. Dual-branch convolutional neural network and its post hoc interpretability for mapping mineral prospectivity [J]. Mathematical Geosciences, 2024, 56 (7): 1487 - 1515.

[69] YIN S, LI N, XIAO K, et al. A dual-channel iterative method integrating semi-supervised self-training and interpretable deep learning models for mineral prospectivity prediction[J]. Natural Resources Research, 2025, 34(6): 2953 - 2971.

[70] 左仁广, 成秋明, 许莹, 等. 可解释性矿产预测人工智能模型[J]. 中国科学: 地球科学, 2024, 54(9): 2917 - 2928.

[71] 王成彬, 王明果, 王博, 等. 融合知识图谱的矿产资源定量预测[J]. 地学前缘, 2024, 31(4): 26 - 36.

[72] 李楠, 尹世滔, 柳炳利, 等. 知识-数据联合驱动的可解释智能矿产预测研究:以四川可尔因矿集区为例[J]. 地学前缘, 2025, 32(4): 60 - 77.

[73] 叶育鑫, 刘家文, 曾婉馨, 等. 基于本体指导的矿产预测知识图谱构建研究[J]. 地学前缘, 2024, 31(4): 16 - 25.

[74] 温鸿涛, 邱昆峰, 才艺伟, 等. 碱性花岗岩中锆石-石英假象智能识别:基于深度学习的新视角[J/OL]. 地学前缘, 2025: 1 - 18[2025-07-31]. https://doi. org/10. 13745/j. esf. sf. 2025. 1. 29.

[75] 程艳, 何慧娟, 陈彦滢, 等. 基于空间权重和层间相关性的可解释浅层类激活映射算法研究[J]. 计算机科学, 2025, 52 (S1): 510 - 516.

[76] DONG Y L, ZHANG Z J. Deep forest modeling: an interpretable deep learning method for mineral prospectivity mapping [J]. Journal of Geophysical Research: Machine Learning and Computation, 2024, 1(4): e2024JH000311.

[77] LI Z, MA X, XIN H. Feature engineering of machine-learning chemisorption models for catalyst design [J]. Catalysis Today, 2017, 280: 232 - 238.

[78] FARAHBAKHSH E, MAUGHAN J, MÜLLER R D. Prospectivity modelling of critical mineral deposits using a generative adversarial network with oversampling and positive-unlabelled bagging[J]. Ore Geology Reviews, 2023, 162: 105665.

[79] ZUO R. Geodata science-based mineral prospectivity mapping: A review [J]. Natural Resources Research, 2020, 29 (6): 3415 - 3424.

[80] 周永章, 张前龙, 黄永健, 等. 钦杭成矿带斑岩铜矿知识图谱构建及应用展望[J]. 地学前缘, 2021, 28(3): 67 - 75.

[81] 张前龙, 周永章, 郭兰萱, 等. 找矿知识图谱的智能化应用:以钦杭成矿带斑岩铜矿为例[J]. 地学前缘, 2024, 31(4): 7 - 15.

[82] ZUO R, KREUZER O P, WANG J, et al. Uncertainties in GIS-based mineral prospectivity mapping: Key types, potential impacts and possible solutions [J]. Natural Resources Research, 2021, 30(5): 3059 - 3079.

[83] YU X, YU P, WANG K, et al. Data-driven mineral prospectivity mapping based on known deposits using association rules [J]. Natural Resources Research, 2024, 33(3): 1025 - 1048.

[84] QIU Q, DUAN Y, MA K, et al. Information extraction and knowledge linkage of geological profiles and related contextual texts from mineral exploration reports for geological knowledge graphs construction [J]. Ore Geology Reviews, 2023, 163: 105739.

[85] YAN Q, XUE L, LI Y, et al. Mineral prospectivity mapping integrated with geological map knowledge graph and geochemical data: A case study of gold deposits at Raofeng area, Shaanxi Province[J]. Ore Geology Reviews, 2023, 161: 105651.

[86] 陆威延, 朱红涛, 徐长贵, 等. 源-汇系统级次划分方法及应用[J]. 地球科学, 2020, 45(4): 1327 - 1336.

[87] 刘强虎, 李志垚, 陈贺贺, 等. 陆相裂陷盆地深时源-汇系统关键地质问题及革新方向[J]. 地球科学, 2023, 48(12): 4586 - 4612.

[88] 刘炳强, 王伟超, 张文龙, 等. 陆相盆地河—湖沉积源—汇系统收支分析——以柴北缘中侏罗统石门沟组为例[J]. 沉积学报, 2022, 40(6): 1494 - 1512.

[89] 李博文，王永志，丁正江，等. 基于大模型与GraphRAG的胶东金矿智能搜索技术[J]. 地学前缘，2025，32(4)：155－164.

[90] 周贞云，黄昭昭，邱均平. 基于LLM与可视化的图数据库专利知识图谱构建及分析[J]. 情报杂志，2025，44(10)：155－162，154.

[91] 盛世强，戴黎明，李法坤，等. 基于GOCAD软件的渤中凹陷428构造带三维地质建模[J]. 海洋地质前沿，2023，39(3)：91－100.

[92] 任凤玉，刘欢，何荣兴，等. 基于SURPAC软件构建罗卜岭铜钼矿岩体基本质量的空间分布[J]. 中国矿业，2019，28(9)：80－84.

[93] 王君照，李胜涛，岳冬冬，等. 基于GIS与GOCAD的天津双窑凸起构造区热储三维地质建模[J]. 科学技术与工程，2023，23(14)：5887－5902.

[94] CHAKRABORTY S，DATTA GUPTA S，DEVI V，et al. Using rock physics analysis driven feature engineering in ML-based shear slowness prediction using logs of wells from different geological setup[J]. Acta Geophysica，2024，72(5)：3237－3254.

[95] ZHU Y，ZHONG E，LU Z，et al. Feature engineering for semantic place prediction[J]. Pervasive and Mobile Computing，2013，9(6)：772－783.

[96] DAI D，XU T，WEI X，et al. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys[J]. Computational Materials Science，2020，175：109618.

[97] WANG Q，MA Y，ZHAO K，et al. A comprehensive survey of loss functions in machine learning[J]. Annals of Data Science，2022，9(2)：187－212.

[98] KARBALAEIRAMEZANALI A，YOUSEFI F，LENTZ D R，et al. Machine learning classification of fertile and barren adakites for refining mineral prospectivity mapping：geochemical insights from the Northern Appalachians，New Brunswick，Canada[J]. Minerals，2025，15(4)：372.

[99] HU J，SHEN L，SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018：7132－7141.

[100] SHENG S，WANG Y，TIAN J，et al. Knowledge-data collaboration-driven mineral prospectivity prediction with graph attention networks[J]. Minerals，2025，15(11)：1164.

[101] 张宸嘉，朱磊，俞璐. 卷积神经网络中的注意力机制综述[J]. 计算机工程与应用，2021，57(20)：64－72.

[102] 程渐，周为峰. 基于优化机器学习模型的海洋主题文本聚类研究与可视化分析[J/OL]. 海洋湖沼通报(中英文)，2025：1－9[2025-07-31]. https://link.cnki.net/urlid/37.1141.P.20250716.1642.004.

[103] 贺海洋，李朋刚，李亚林，等. 基于CiteSpace的机器学习在地球化学应用领域的可视化分析[J]. 地球化学，2025，54(6)：1058－1072.

[104] 陈穗芬，郝立巍，江贵平. 磁共振扩散张量成像的可视化方法研究进展[J]. 生物医学工程学杂志，2008，25(3)：724－728.

[105] 张敏娜，刘玉树，林国璋. 地应力三维张量场的可视化技术研究[J]. 北京理工大学学报，2000，20(4)：456－460.

[106] 曾春艳，严康，王志锋，等. 深度学习模型可解释性研究综述[J]. 计算机工程与应用，2021，57(8)：1－9.

[107] SMILKOV D，THORAT N，KIM B，et al. Smoothgrad：removing noise by adding noise[J]. arXiv preprint arXiv：1706.03825，2017.

[108] AGRAWAL N，GOVIL H. A deep residual convolutional neural network for mineral classification[J]. Advances in Space Research，2023，71(8)：3186－3202.

[109] 李妙歌，陈渤，王东升，等. 面向SAR图像目标分类的CNN模型可视化方法[J]. 雷达学报，2024，13(2)：359－373.

[110] 邵国栋，李峥辉，郭松杰，等. 基于梯度下降法直接拟合光谱吸收信号的气体浓度测量方法研究[J]. 光谱学与光谱分析，2021，41(10)：3256－3261.

[111] ZHOU B L，KHOSLA A，LAPEDRIZA A，et al. Learning deep features for discriminative localization[C]//Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2016：2921－2929.

[112] FU R，HU Q，DONG X，et al. Axiom-based grad-cam：Towards accurate visualization and explanation of cnns[J]. arXiv preprint arXiv：2008.02312，2020.

[113] WANG H，WANG Z，DU M，et al. Score-CAM：Score-weighted visual explanations for convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020：24－25.

[114] 张国栋，韩晓红，许衍胜. 基于导向积分梯度与多尺度融合的类激活映射方法[J/OL]. 太原理工大学学报，2025：1－8[2025-07-31]. https://link.cnki.net/urlid/14.1220.N.20250606.1503.004.

[115] GHOLAMI H，MOHAMMADIFAR A，GOLZARI S，et al. Interpretability of simple RNN and GRU deep learning models used to map land susceptibility to gully erosion[J]. Science of the Total Environment，2023，904：166960.

[116] 杜树坤，张晶，韩志军，等. 基于随机森林模型的“网格-月”尺度武装冲突风险预测及影响因素分析——以中南半岛为例[J]. 地球信息科学学报，2023，25(10)：2026－2038.

[117] MCCOY J T，AURET L. Machine learning applications in minerals processing：A review[J]. Minerals Engineering，2019，132：95－109.

[118] ZHANG S，CARRANZA E J M，FU C，et al. Interpretable machine learning for geochemical anomaly delineation in the Yuanbo Nang district，Gansu Province，China[J]. Minerals，2024，14(5)：500.

[119] MOU N，CARRANZA E J M，WANG G，et al. A framework for data-driven mineral prospectivity mapping with interpretable machine learning and modulated predictive modeling[J]. Natural Resources Research，2023，32(6)：2439－2462.

[120] 苏恺明，徐耀辉，徐旺林，等. 鄂尔多斯盆地延长组多油源贡献比例与分布规律：基于机器学习与可解释性研究[J]. 地学前缘，2024，31(3)：530－540.

[121] ROY T，DAS P，JAGIRDAR R，et al. Prediction of mechanical

properties of eco-friendly concrete using machine learning algorithms and partial dependence plot analysis[J]. Smart Construction and Sustainable Cities, 2025, 3(1): 2.

[122] NOHARA Y, WAKATA Y, NAKASHIMA N. Interpreting Medical Information Using Machine Learning and Individual Conditional Expectation[J]. MedInfo, 2015, 1073.

[123] 孙德亮, 陈丹璐, 密长林, 等. 基于随机森林-特征递归消除模型的可解释性缓丘岭谷地貌滑坡易发性评价[J]. 地质力学学报, 2023, 29(2): 202 - 219.

[124] 陈丹璐, 安雪莲, 邵怀勇, 等. 黄土塬地貌区滑坡灾害易发性量化评估及关键驱动因子[J]. 北京师范大学学报(自然科学版), 2025, 61(2): 255 - 267.

[125] 陈忠元, 任涛, 赵冻. 基于TPE-SVM模型和SHAP解释的闪锌矿微量元素特征识别铅锌矿床类型[J]. 地球科学, 2025, 50(11): 4355 - 4369.

[126] KONG B, CHEN Z. Data-driven EUR modeling and optimization in the liquid-rich Duvernay Formation, western Canada sedimentary basin, Canada[J]. Journal of Petroleum Science and Engineering, 2022, 213: 110352.

[127] LUNDBERG S M, ERION G, CHEN H, et al. From local explanations to global understanding with explainable AI for trees[J]. Nature Machine Intelligence, 2020, 2(1): 56 - 67.

[128] LIN Z, FENG M, SANTOS C N D, et al. A structured self-attentive sentence embedding[C]//Proceedings of the 5th International Conference on Learning Representations. Toulon, France, 2017: 1 - 15.