

基于 SMOTified-BRF 模型的地球化学金矿化异常预测研究

——以陕西汉滨—旬阳地区为例

徐正林¹, 王 晰¹, 薛林福^{1*}, 冉祥金¹, 燕 群¹, 李永胜^{2,3}, 于晓飞^{2,3}

(1. 吉林大学地球科学学院; 2. 中国地质调查局发展研究中心; 3. 自然资源部矿产勘查技术指导中心)

摘要:矿产勘查既是新质生产力的重要组成,又是推动新质生产力发展的重要力量。为了改善矿产勘查中金矿化预测面临的由于已知矿化样本数量少导致的样本类不平衡问题及矿化信息稀缺问题,提出 SMOTified-BRF 模型,该模型使用 SMOTE 方法对极少数已知矿化样本进行数量增强并使用平衡随机森林方法进行预测。以汉滨—旬阳地区为研究区,对水系沉积物地球化学数据分别使用 SMOTified-BRF 模型和 BRF 模型进行金矿化预测效果和模型性能对比研究。研究结果表明: SMOTified-BRF 模型的 AUC 值(0.987 5)高于 BRF 模型的 AUC 值(0.972 6),且在约登指数指示的最优阈值下 SMOTified-BRF 模型预测的矿化面积占比(1.95 %)相较于 BRF 模型(12.23 %)更小,说明 SMOTified-BRF 模型相比于 BRF 模型具有更准确高效的金矿化异常预测性能。

关键词:金矿化;异常;预测;平衡随机森林法;SMOTE;地球化学;汉滨—旬阳地区

中图分类号:TD11 P618.51

文章编号:1001-1277(2025)01-0011-09

文献标志码:A

doi:10.11792/hj20250103

引 言

2023 年 9 月,习近平总书记在主持召开新时代推动东北全面振兴座谈会时提出,积极培育战略性新兴产业和未来产业,加快形成新质生产力,增强发展新动能^[1]。矿产勘查处于矿产资源产业链供应链的最前端,关系到现代化产业体系发展的能量与物质基础,在发展新质生产力中具有重要的基础地位和支撑作用。矿产勘查通过发现和查明矿产资源,既为现代产业体系构建提供物质基础,又是现代产业体系的一部分,自身生产力的提升也是发展新质生产力的重要内容,对于保障资源安全,增强发展新动能具有重要意义^[2-3]。作为矿业行业新质生产力的重要组成,人工智能找矿方兴未艾^[4]。

地球化学数据是一类重要的找矿信息,在矿产勘查与评价中具有非常关键的作用^[5-7]。基于地球化学数据处理的人工智能方法研究一直是地球化学研究领域的热门方向之一。使用先进的方法从大量地球化学数据中挖掘矿致信息并进行矿化预测,有助于推动矿产勘查发展^[8]。

成矿系统的复杂性及成矿作用的多期多阶段特征导致地球化学数据往往具有复杂、未知的空间与频率分布特征。近年来,由于机器学习、深度学习方

在处理具有复杂分布特性非线性数据的分类与回归预测问题中具有良好表现^[9],越来越多的学者将先进的机器学习、深度学习方法引入地球化学矿化预测领域,如支持向量机^[10-11]、随机森林^[12-13]、受限玻尔兹曼机^[14]等。然而,地球化学数据中的已知矿化样本(阳性样本)数量稀缺,背景样本(阴性样本)数量占比大,使用一般的机器学习和深度学习方法对类不平衡的地球化学数据进行矿化预测面临着挑战。已有学者将关注点放在解决地球化学数据类别、数量不平衡问题的方法研究上,此类研究包括从数据处理角度引进各类采样方法平衡数据类别比例和从算法角度引进类不平衡学习集成算法进行数据的分类任务处理。例如,SHAYILAN 等^[15]利用合成少数类过采样技术进行过采样以平衡数据类别比例,CHEN 等^[16-17]引入自定步集成框架来处理类不平衡问题。

考虑到已有模型大多单独关注从数据或者算法角度解决样本类不平衡的问题,因此本研究同时从数据处理角度和算法角度出发,结合合成少数类过采样技术(Synthetic Minority Oversampling Technique, SMOTE)^[18]和平衡随机森林算法(Balanced Random Forest, BRF)^[19],提出采用 SMOTified-BRF 模型进行类不平衡地球化学数据的矿化预测研究。平衡随机森林算法是一种类不平衡学习集成算法,它在随机森

收稿日期:2024-07-02; 修回日期:2024-09-15

基金项目:中国地质调查局矿调项目(DD20190159)

作者简介:徐正林(2000—),男,硕士研究生,研究方向为人工智能找矿预测、人工智能油气应用等;E-mail:2765018131@qq.com

* 通信作者:薛林福(1962—),男,教授,博士,从事地学数据分析、三维地质建模与地质过程模拟等方面的研究工作;E-mail:xuelin@jlu.edu.cn

林算法(Random Forest, RF)的基础上引进了随机欠采样策略^[20]。平衡随机森林算法在训练每棵决策树(Decision Tree, DT)之前通过对数量占多数的类别数据进行随机欠采样,以此来平衡用于训练每棵决策树的训练子集类别比例。平衡随机森林算法建立在传统随机森林算法的基础上,使用多棵决策树的集成来提高预测的准确性和稳定性。平衡随机森林算法不仅具有随机森林算法的准确率高、善于处理离散数据等优点^[21],同时更善于处理类不平衡数据。平衡随机森林算法适合用于类不平衡地球化学数据的异常识别与分类。然而,在已知矿化样本极度稀缺的地球化学数据中,平衡随机森林算法在训练每棵决策树前为了构建类别平衡的训练子集,会随机剔除过多的背景样本,导致丢失具有重要信息的背景样本。同时,已知矿化样本的稀缺容易导致决策树在学习已知矿化样本特征的过程中出现过拟合问题。因此,本研究引进 SMOTE,对地球化学数据中的已知矿化样本进行数量增强,并使用增强后的地球化学数据作为平衡随机森林算法的训练集,构建基于 SMOTE 增强的平衡随机森林金矿化预测模型(SMOTified-BRF 模型)。

本研究以汉滨—旬阳地区作为研究区,对 1:5 万水系沉积物地球化学数据分别构建 SMOTified-BRF 模型和 BRF 模型进行金矿化预测和对比研究。研究结果证明,SMOTified-BRF 模型预测的金矿化区比

BRF 模型预测的金矿化区具有更小的范围和更高的准确性。

1 地质概况

汉滨—旬阳地区位于陕西省安康市(见图 1-C),大地构造位置位于秦岭地块南缘,属扬子板块(Ⅲ)秦岭—大别山新元古代—中生代造山带(Ⅲ₁)北大巴山—西倾山早古生代裂谷带(Ⅲ₁²)舟曲—安康早古生代裂谷带(Ⅲ₁²⁻¹)(见图 1-B)。研究区主要发育断裂有北西向展布脆性断裂,近南北向次级小型脆性断裂及北西向脆韧性剪切带,由北向南划分为赵家山—斜山寨滑脱逆冲推覆体、将军山—烂木沟滑脱逆冲推覆体和安康断陷盆地 3 个次级构造单元。研究区出露地层有寒武系鲁家坪岩组(Z₂-C₁l)、箭竹坝组(Z₂-C₁j),奥陶系洞河岩组(Odh),奥陶系—志留系斑鸠关岩组(O₃-S₁b),志留系梅子垭岩组(S₁₋₂m),泥盆系大枫沟组(D₂d)、古道岭组(D₂₋₃g)、星红铺组(D₃x),以及第四系(Q)等(见图 1-A)^[22]。研究区内火山岩建造不发育,仅在图幅中部神滩河—烂木沟一带鲁家坪岩组中有少量出露;侵入岩极不发育,未见岩体出露,仅见酸性岩脉出露^[23]。研究区主体为变质岩建造区,占 70 % 以上,气-液变质岩不可见,热接触变质岩极不发育,变质岩主要为区域变质岩和动力变质岩。

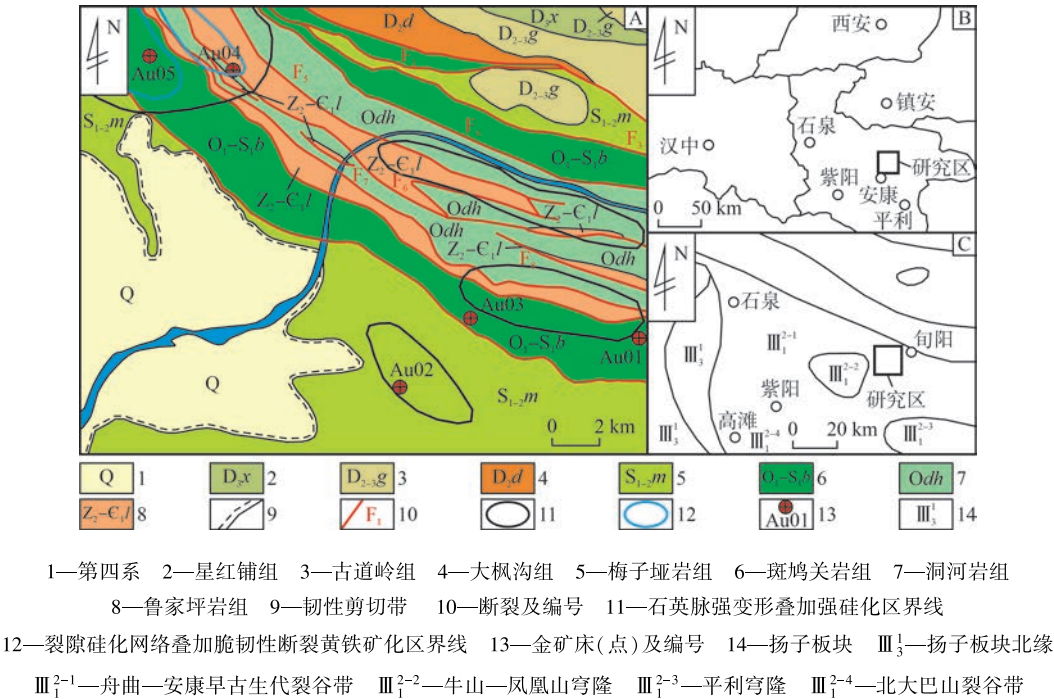


图 1 汉滨—旬阳地区地质图(A)、大地构造位置图(B)及地理位置图(C)

Fig. 1 Geological map (A), tectonic location map (B), and geographic location map (C) of the Hanbin-Xunyang area

研究区共发现典型金矿床(点)5处(见表 1),成因类型均为韧性剪切带型,主要分布于区域性韧性剪

切带两侧的次级构造带中。区域内金矿床(点)主要分布于斑鸠关岩组、洞河岩组、梅子垭岩组、鲁家坪岩

组地层中。含矿岩性主要为含碳黑云母绢云母石英片岩、绢云石英片岩、硅质板岩、碳质板岩等。研究区内与金成矿关系密切的蚀变主要有硅化、绢云母化、黄铁矿化等,表现为蚀变越强,规模越大^[24]。

表 1 汉滨—旬阳地区典型金矿床(点)

Table 1 Typical gold deposits (points) in the Hanbin-Xunyang area

序号	名称	规模
Au01	陕西省旬阳县烂木沟金矿床	中型
Au02	陕西省汉滨区石门金矿点	矿点
Au03	陕西省汉滨区关家沟金矿点	矿点
Au04	陕西省汉滨区茨沟—早阳金矿床	中型
Au05	陕西省汉滨区将军山金矿点	矿点

2 地球化学数据预处理与指示元素选择

研究区的地球化学数据来源于安康幅 1:5 万水系沉积物调查样本,样本数量为 1 828 件。对样本测试了 14 种元素,分别为 Au、Ag、As、Bi、Cd、Co、Cr、Hg、Mo、Pb、Zn、W、Sn、Ni。对 1 828 件水系沉积物样本的 14 种元素数据进行网格化处理,为每种元素生成 379×476(180 404)个网格数据。

许多元素在空间的浓度分布与特定类型的矿化相关,其有助于指导地质专家发现特定类型矿床^[25]。本研究旨在通过分析研究区与金矿化相关的地球化学数据来指导金矿勘查。基于前人研究结果,本文使用 ROC (Receiver Operating Characteristic) 曲线来评价研究区地球化学模式与已知金矿床(点)位置之间的空间相关性^[26-28]。结合 ROC 曲线下面积(AUC)和对应的满足标准正态分布的随机变量值(Z_{AUC})选择指示研究区金矿化的元素^[14-17]。基于 14 种元素的浓度值及已知矿化点的对应组成元素浓度绘制了 ROC 曲线,结果见图 2;计算了 14 种元素 AUC 值和 Z_{AUC} 值,结果见表 2。由图 2 可知,Au、Co、Ni、Pb 和 Zn 5 种元素的 ROC 曲线更加接近 ROC 空间的左上角,表明该 5 种元素的浓度分布模式与已知金矿床(点)的位置分布具有较强的空间相关性。由表 2 可知,Au、Co、Ni、Pb 和 Zn 5 种元素的 AUC 值和 Z_{AUC} 值分别大于 0.5 和 1.96。因此,就 AUC 值和 Z_{AUC} 值而言,这 5 种元素与成矿具有正相关关系^[25]。本研究选择 Au、Co、Ni、Pb 和 Zn 5 种元素作为指示研究区金矿化的地球化学元素。

所选 5 种指示元素的浓度等值线图见图 3。由图 3 可知,除 Au 分布较为零散外,剩余 4 种元素大体呈北西向分布,与韧性剪切带展布方向一致,这 5 种元素的浓度分布与 5 个已知金矿床(点)的分布有较

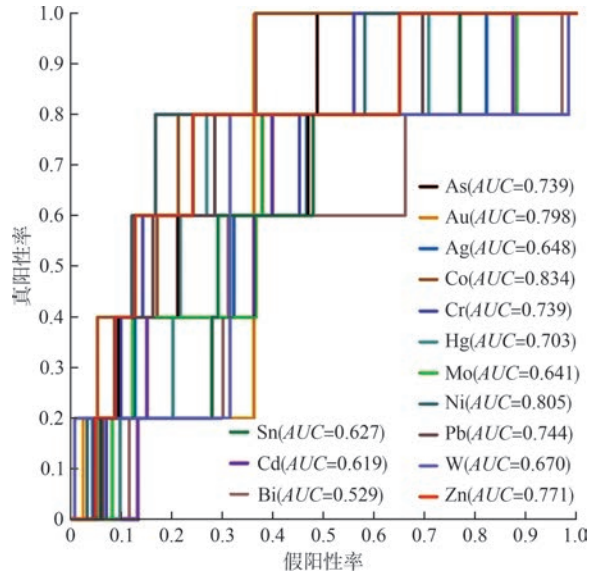


图 2 研究区 14 种地球化学元素 ROC 曲线

Fig. 2 ROC curves of 14 geochemical elements in the study area

表 2 研究区 14 种地球化学元素 AUC 值与 Z_{AUC} 值

Table 2 AUC and Z_{AUC} values of 14 geochemical elements in the study area

元素	AUC 值	Z_{AUC} 值	元素	AUC 值	Z_{AUC} 值
Ag	0.648	1.104	Hg	0.703	1.549
As	0.739	1.863	Mo	0.641	1.055
Au	0.798	2.494	Ni	0.805	2.574
Bi	0.529	0.220	Pb	0.744	1.977
Cd	0.619	0.886	Sn	0.627	0.945
Co	0.834	2.979	W	0.670	1.277
Cr	0.739	1.866	Zn	0.771	2.185

强的空间相关性。

3 基本算法

3.1 SMOTE 算法

SMOTE 算法是由 CHAWLA 等^[18]提出的一种过采样方法,旨在通过数量增强来解决数据的类不平衡问题。该算法受 K 近邻算法启发,核心思想是对每个少数类样本,在特征空间中使用欧几里得距离来确定其 k 个最近的同类样本。对于选定的少数类样本及其近邻样本,SMOTE 算法通过在它们之间的连线上进行随机插值来合成新的少数类样本。具体而言,对于每个少数类样本 $x_i \in R^d$,首先确定其同类样本中的 k 个近邻样本,然后从这 k 个近邻样本中随机选择 1 个样本($x_{ij}, j=1, 2, 3, \dots, k$)。在样本 x_i 与 x_{ij} 连接的线段上随机插值生成 1 个新的样本点 x_{new} 。其中, $x_{i,p}$ 为少数类样本 x_i 的第 p 个属性值, $x_{ij,p}$ 为近邻样本 x_{ij} 的第 p 个特征值, $p=1, 2, 3, \dots, d$ 。新样本 x_{new} 的第 p 个特征可通过式(1)计算得出:

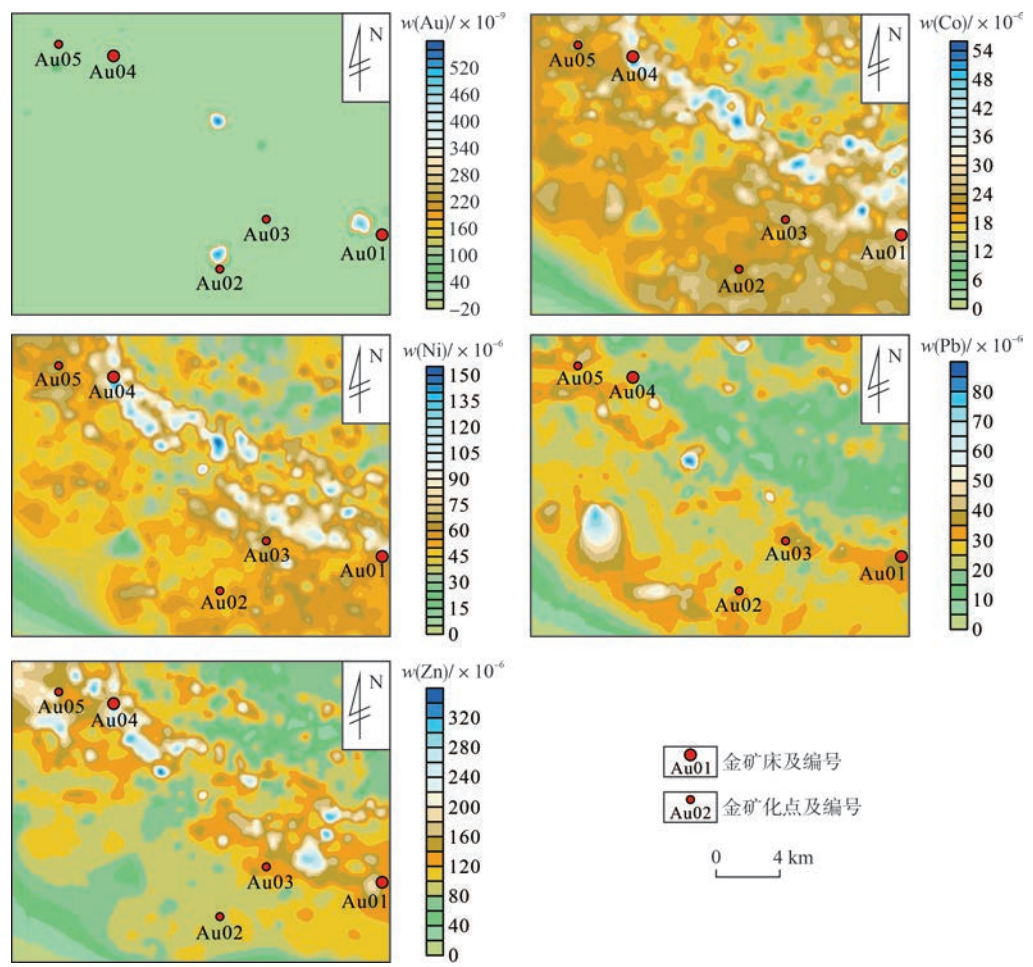


图3 所选5种指示元素的浓度等值线图

Fig.3 Contour plots of concentration of 5 selected indicator elements

$$x_{new,p} = x_{i,p} + \lambda(x_{ij,p} - x_{i,p}) \quad (1)$$
 信息点,增强了数据集中少数类样本的表达。
式中: λ 为介于0和1之间的随机数。SMOTE算法在特征空间合成少数类样本的过程见图4。
通过上述方法,SMOTE算法创造性地引入新的

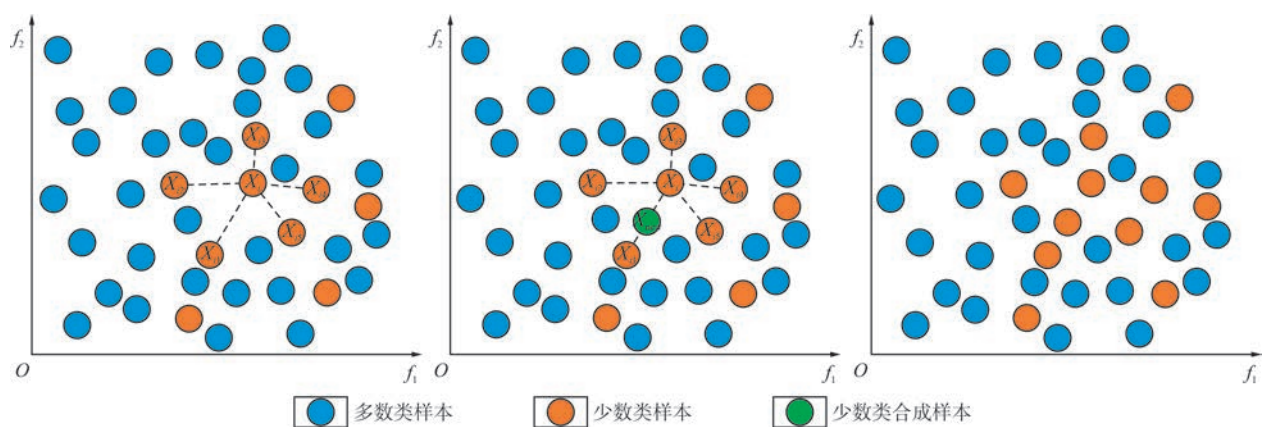


图4 SMOTE算法在特征空间合成少数类样本的过程

Fig.4 Process of synthesizing minority class samples in feature space using the SMOTE algorithm

3.2 BRF 算法

BRF算法是由YOUSRA等^[19]提出的一种为了解决训练数据中类不平衡问题而设计的改进RF算法。该算法基于bagging算法的集成思想,同时迭代

训练多棵相互独立的决策树(CART)弱分类器,然后使用训练过的每棵决策树对测试集数据进行分类预测并对预测结果通过投票或求平均值的方式进行集成。

在二分类问题中,通常将少数类样本设定为阳性样本,将多数类样本设定为阴性样本。针对二分类问题,给定训练样本集 $D = (x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$ 。其中, y_i 为样本 x_i 的类标签 ($i = 1, 2, 3, \dots, n$), 取 0 或 1 (0 代表阴性样本, 1 代表阳性样本), 样本集 D 中阳性样本数量为 m ; 测试集 $U = \{x_1, x_2, x_3, \dots, x_c\}$; 决策树弱分类器迭代次数 T , BRF 算法的步骤为:

1) 对于第 t ($t = 1, 2, 3, \dots, T$) 棵决策树而言, ①从训练集的阳性样本中随机有放回地抽取 m 个引导 (bootstrap) 样本, 同时基于随机欠采样的策略从多数类样本中随机有放回地抽取相同数量的样本; ②将抽取相同数量的 2 类样本组合成一个类平衡的训练子集 D_t ; ③从训练子集 D_t 中诱导 1 棵决策树 f_t , 使其达到指定的最大生长深度, 不进行剪枝。

2) 将步骤 (1) 重复多次, 直到训练出指定数量 (T) 的决策树弱分类器。

3) 使用 T 棵训练过的决策树对测试集数据 U 进行分类预测, 令 $f_t(x_i)$ 表示第 t 棵决策树 f_t 预测样本 x_i ($i = 1, 2, 3, \dots, c$) 为阳性的概率。

4) 对 T 棵决策树的预测结果通过求平均值得到最终预测结果。其中, T 棵决策树对测试样本 x_i ($i = 1, 2, 3, \dots, c$) 为阳性样本的概率预测结果表示为:

$$f_T(x_i) = \sum_{t=1}^T f_t(x_i) / T \quad (2)$$

上述决策树基于 CART 算法构建, 但进行了部分修改, 即在每个节点上不再搜索所有特征变量以获得最佳分割点, 而是随机选取部分特征变量进行搜索, 有助于增加决策树模型的多样性。BRF 算法流程见图 5。

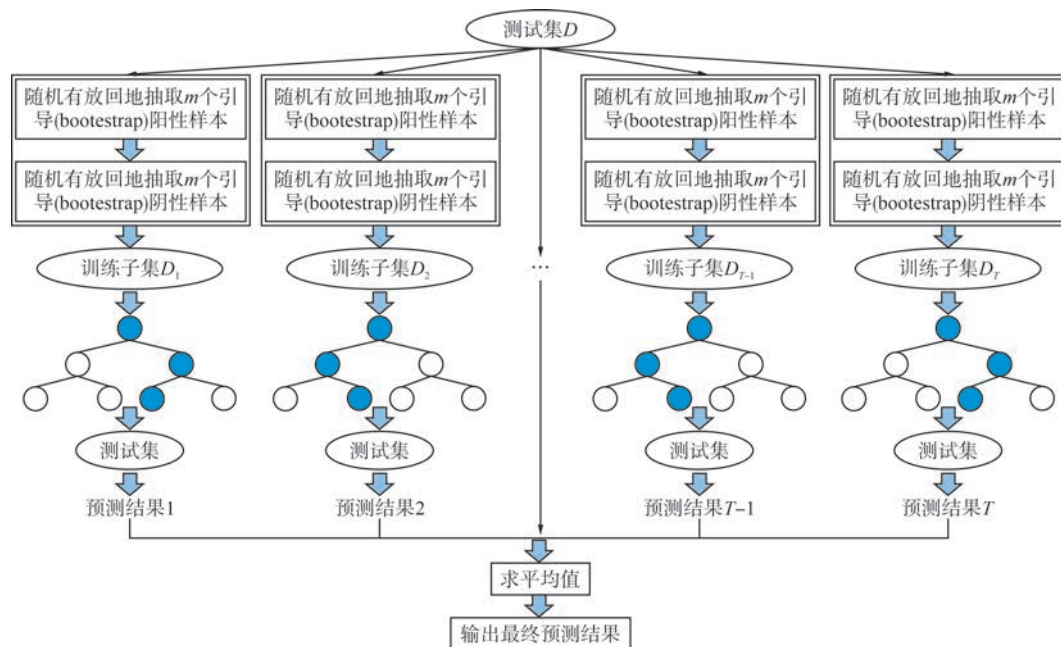


图 5 BRF 算法流程

Fig. 5 Flowchart of BRF algorithm

4 金矿化预测

4.1 SMOTified-BRF 模型与 BRF 模型

为了解决样本中存在的类比平衡问题, 首先, 使用 SMOTE 技术对研究区地球化学数据中的 5 个金矿床 (点) 样本进行过采样, 通过生成一定数量的合成矿化样本, 增加了原始地球化学数据集中的已知金矿化样本数量。其次, 将包含合成矿化样本的地球化学数据集作为 BRF 算法的训练集, 构建了 SMOTified-BRF 模型。模型的参数设置对于模型的最终性能具有决定性作用, 本研究以模型正确预测已知金矿床 (点) 的能力作为参数优化指标, 使用 GridSearchCV 进行参数优化。SMOTified-BRF 模型的参数设置见

表 3。训练结束后, 将研究区所有样本作为测试数据, 使用 SMOTified-BRF 模型对测试数据进行类别预测。SMOTified-BRF 模型输出的是一个向量, 表示测试样本属于背景样本的概率与金矿床 (点) 样本的概率, 2 类样本的预测概率和为 1。同时, 使用 BRF 算法对研究区原始地球化学数据进行建模, 构建 BRF 模型, 参数设置与 SMOTified-BRF 模型的 BRF 分类器相同。BRF 模型输出的是一个向量, 表示测试样本属于背景样本的概率与金矿床 (点) 样本的概率, 2 类样本的预测概率和为 1。

本研究基于 SMOTified-BRF 模型和 BRF 模型对研究区地球化学数据进行预测, 这 2 个模型的性能评估指标值见表 4。

表 3 SMOTified-BRF 模型的参数设置

Table 3 Parameter settings for the SMOTified-BRF model

SMOTified-BRF 模型组件	参数名称	值	作用描述
SMOTE 增强过程	sampling_strategy	0.01	决定了少数类样本生成的数量。可以设置为不同值来指定少数类样本与多数类样本的数量比例
	k_neighbors	4	决定了用于构建合成样本时参考邻近少数类样本的数量。此参数值越大,合成样本就越能反映整体少数类的分布,但同时可能会增加生成过度相似样本的风险
	n_estimators	50	指定了决策树弱分类器的数量
BRF 分类过程	max_depth	5	定义了决策树的最大深度。设置决策树的最大深度可以防止过拟合,但如果设置得太低,可能会导致模型欠拟合

注:其余参数设置为默认值。

表 4 模型的性能评估指标值

Table 4 Performance evaluation indicator values of the model

模型	AUC 值	Z_{AUC} 值	最大约登指数	最优阈值	召回率/%	矿化面积占比/%
SMOTified-BRF	0.987 5	13.891 4	0.980 6	0.813 6	100	1.95
BRF	0.972 6	9.223 2	0.877 7	0.678 2	100	12.23

注:召回率指模型预测的矿化概率不小于最优阈值的矿化区域所覆盖的已知金矿床(点)占总已知金矿化床(点)的百分比;矿化面积指模型预测的矿化概率不小于最优阈值的矿化区域占研究区总面积的百分比。

4.2 模型性能评估

ROC 常用于评估二分类器的分类性能,已被用于地球化学矿化预测模型的性能评估^[15,17,29]。本研究绘制了 SMOTified-BRF 模型和 BRF 模型的 ROC 曲线(见图 6)并计算了对应的 AUC 值和 Z_{AUC} 值(见表 4)。由图 6 可知,SMOTified-BRF 模型的 ROC 曲线更接近 ROC 空间的左上角。由表 4 可知,SMOTified-BRF 模型的 AUC 值为 0.987 5,高于 BRF 模型的 AUC 值(0.972 6)。同时,SMOTified-BRF 模型($Z_{AUC}=13.891 4$)和 BRF 模型($Z_{AUC}=9.223 2$)的 Z_{AUC} 值均高于 1.96,说明在 0.05 的显著性水平下,SMOTified-BRF 模型和 BRF 模型分类性能显著高于随机猜测模型。

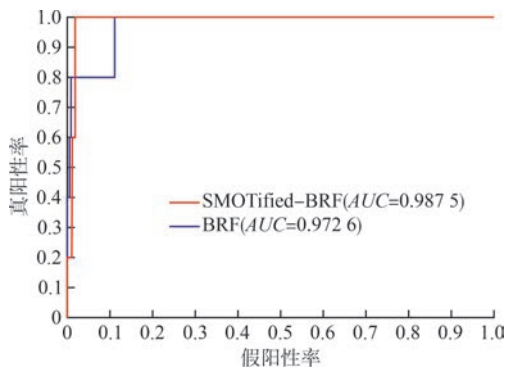


图 6 SMOTified-BRF 模型和 BRF 模型的 ROC 曲线

Fig. 6 ROC curves of SMOTified-BRF and BRF models

4.3 矿化区圈定

本文引进约登指数(Youden Index, YI)圈定模型预测最优阈值下的研究区金矿化情况^[15,29]。约登指

数是一种适用于选择最佳分类阈值的统计度量, YI 为真阳性率减去假阳性率。在本文中,真阳性率是指所有已知金矿化样本中被模型正确预测出来的已知金矿化样本的比例,假阳性率是指背景样本中被模型预测为已知金矿化样本的比例。约登指数取值为 0~1,约登指数值越大,代表模型分类性能越好。最大约登指数对应的阈值被确定为划分预测矿化区的最优阈值,结果见图 7。由图 7 可知,SMOTified-BRF 模型的最大约登指数为 0.980 6,其对应的最优阈值为 0.813 6;BRF 模型的最大约登指数为 0.877 7,其对应的最优阈值为 0.678 2。

SMOTified-BRF 模型和 BRF 模型预测的金矿化概率分布见图 8。基于图 7 中最高约登指数对应的最优阈值,在图 8-a 中圈定了金矿化概率大于 0.813 6 的矿化区,在图 8-b 中圈定了金矿化概率大于 0.678 2 的矿化区。矿化区呈北西向分布,且与已知金矿床(点)的空间分布一致。通过对比图 8-a 和图 8-b 可以发现,矿化区覆盖了 5 个已知金矿床(点)。从圈定面积来看,图 8-a 圈定的矿化面积占研究区总面积的 1.95%,小于图 8-b 的 12.23%。

同时,为避免由于选取不同最优阈值导致的圈定面积差异问题,选取 SMOTified-BRF 模型的最优阈值 0.813 6 同时作为 BRF 模型的高矿化概率边界值,预测的矿化概率不小于 0.813 6 的金矿化区分布见图 9。由图 9 可知,SMOTified-BRF 模型圈定的矿化区覆盖了 5 个已知金矿床(点),BRF 模型圈定的矿化区仅覆盖了 4 个已知金矿床(点);在准确性存在差异的情况下,SMOTified-BRF 模型圈定的矿化区面积占

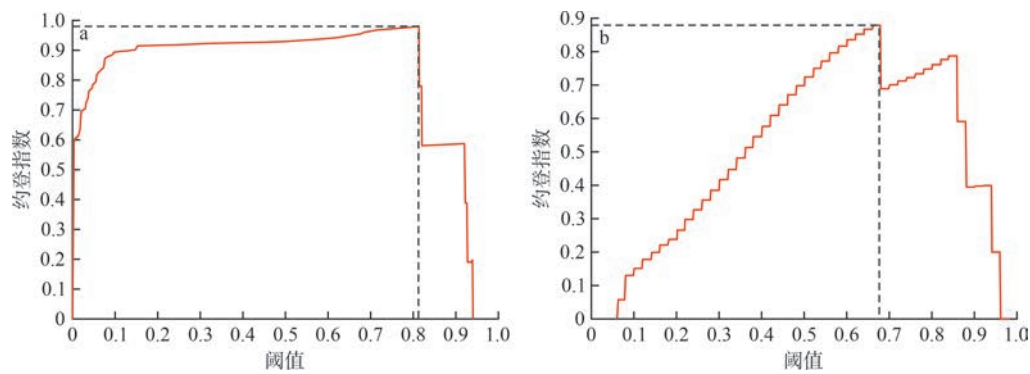


图 7 SMOTified-BRF 模型(a) 和 BRF 模型(b) 的约登指数随阈值变化曲线

Fig. 7 Youden index versus threshold curves for SMOTified-BRF model (a) and BRF model (b)

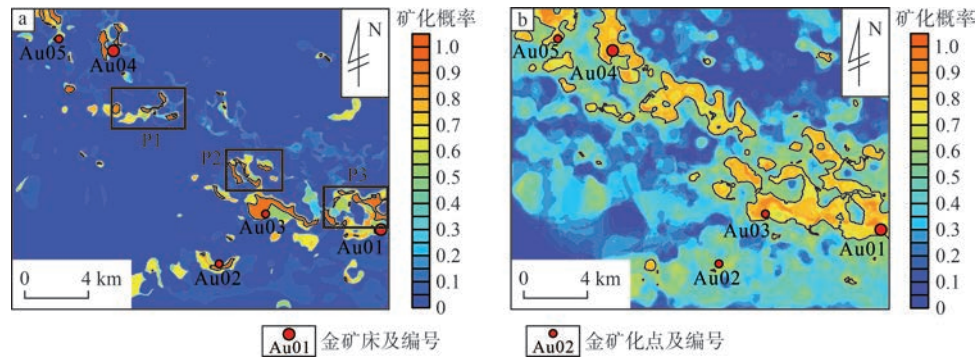


图 8 SMOTified-BRF 模型(a) 和 BRF 模型(b) 预测的金矿化概率分布

Fig. 8 Probability distribution of gold mineralization predicted by SMOTified-BRF model (a) and BRF model (b)

比(1.95 %)仍小于 BRF 模型中圈定的面积占比(3.96 %)。由此可见,SMOTified-BRF 模型具有更

准确的金矿化相关地球化学金矿化预测性能。

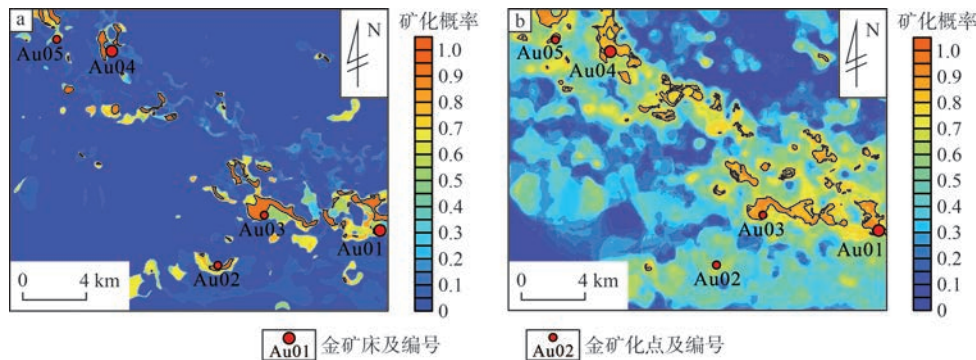


图 9 SMOTified-BRF 模型(a) 和 BRF 模型(b) 预测的矿化概率不小于 0.813 6 的金矿化区分布

Fig. 9 Distribution of gold mineralization zones with a probability of mineralization not less than 0.813 6 predicted by the SMOTified-BRF model (a) and the BRF model (b)

从圈定区域的合理性考虑,图 8-a 圈定的矿化范围覆盖地层为斑鸠关岩组、洞河岩组、梅子垭岩组;图 8-b 圈定的矿化范围覆盖地层为斑鸠关岩组、洞河岩组、梅子垭岩组、鲁家坪岩组;图 9-b 圈定的矿化范围覆盖的地层为斑鸠关岩组、洞河岩组、梅子垭岩组、鲁家坪岩组。从圈定地层与含矿地层情况对比可知,SMOTified-BRF 模型和 BRF 模型所圈定的金矿化预测结果与含矿地层吻合程度较好;从圈定区域与构造的对比情况来看,SMOTified-BRF 模型和 BRF 模型圈定的矿化区整体上沿北西向分布,与韧

性剪切带分布一致。以上结果证明,圈定的金矿化预测区具有一定的可靠性。

以 SMOTified-BRF 模型圈定的金矿化预测区中(见图 8-a),除原有金矿床(点)外,另圈定出 3 处不含已知金矿床(点)的找矿远景区,分别为 P1、P2 与 P3。

找矿远景区 P1 所处地层为斑鸠关岩组、洞河岩组和鲁家坪岩组,岩性主要为黑色碳质板岩、黑色板岩、绢云石英片岩等,在空间分布上大体沿韧性剪切带分布,且靠近 1 处石英脉强变形叠加强硅化区界

线;找矿远景区 P2 所处地层为鲁家坪岩组和洞河岩组,岩性主要为绢云石英片岩、黑色板岩等,在空间分布上大体沿韧性剪切带分布,且靠近 1 处石英脉强变形叠加强硅化区界线;找矿远景区 P3 所处地层为鲁家坪岩组和洞河岩组,岩性主要为绢云石英片岩、黑色板岩等,在空间分布上,沿韧性剪切带分布,且靠近 2 处石英脉强变形叠加强硅化区界线。

通过对比地层、岩性、分布及蚀变等情况,圈定的 3 处找矿远景区具有较好的合理性,可以作为下一步找矿依据。

5 结 论

1) SMOTified-BRF 模型与 BRF 模型地球化学金矿化预测均能取得较好效果,且预测金矿化区的地层情况与含矿地层一致,分布区域与韧性剪切带吻合,说明具有较好的可靠性,为矿产勘查方面新质生产力的发展提供了新思路。

2) 在本次预测中,SMOTified-BRF 模型表现优于 BRF 模型,以 AUC 值作为评价指标的情况下,SMOTified-BRF 模型(0.987 5)的表现优于 BRF 模型(0.972 6);基于约登指数指示的最优阈值的情况下,SMOTified-BRF 模型圈定出了更为精确的矿化面积占比(1.95 %),相较于 BRF 模型(12.23 %)更为准确。

3) 基于 SMOTified-BRF 模型在汉滨—旬阳地区圈定了 3 处未包含已知金矿床点的找矿远景区,为进一步找矿提供了新方向。

[参 考 文 献]

- [1] 习近平主持召开新时代推动东北全面振兴座谈会强调:牢牢把握东北的重要使命 奋力谱写东北全面振兴新篇章[N].人民日报,2023-09-10(01).
- [2] 杨建锋,余韵,姚晓峰,等.矿产勘查推动新质生产力发展路径初探[J].中国矿业,2024,33(5):39-45.
- [3] 王登红.试论稀散金属矿产与新质生产力[J].中国矿业,2024,33(4):2-12,1.
- [4] 朱清.发展 AI 找矿这种新质生产力[N].中国矿业报,2024-03-09.
- [5] 张七道,肖长源,李致伟,等.黔北普宜地区水系沉积物地球化学特征及成矿预测[J].地质与勘探,2021,57(5):1 040-1 052.
- [6] 安朝,李德彪,柴云,等.青海省东昆仑纳赤台地区正、负异常特征及找矿方向[J].黄金,2019,40(3):23-29.
- [7] 张凯伦,温守钦,汪徽,等.辽宁阜新排山楼金矿床深穿透地球化学找矿方法的应用研究[J].黄金,2023,44(9):111-116.
- [8] ZUO R G.Exploration geochemical data mining and weak geochemical anomalies identification[J].Earth Science Frontiers,2019,26(4):65-75.
- [9] ZUO R G,XIONG Y H.Big data analytics of identifying geochemical anomalies supported by machine learning methods[J].Natural Resources Research,2018,27:5-13.
- [10] ZUO R G,CARRANZA E J M.Support vector machine:A tool for mapping mineral prospectivity[J].Computers and Geosciences,2011,37:1 967-1 975.
- [11] ABEDI M,NOROUZI G H,BAHROUDI A.Support vector machine for multi-classification of mineral prospectivity areas[J].Computers and Geosciences,2012,46:272-283.
- [12] RODRIGUEZ-GALIANO V F,CHICA-OLMO M,CHICA-RIVAS M.Predictive modelling of gold potential with the integration of multi-source information based on random forest:A case study on the Rodalquilar Area,Southern Spain[J].International Journal of Geographical Information Science,2014,28:1 336-1 354.
- [13] CARRANZA E J M,LABORTE A G.Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines)[J].Computer and Geosciences,2015,74:60-70.
- [14] CHEN Y L.Mineral potential mapping with a restricted boltzmann machine[J].Ore Geology Reviews,2015,71:749760.
- [15] SHAYILAN A,CHEN Y L.A smotified extreme learning machine for identifying mineralization anomalies from geochemical exploration data:A case study from the Yeniugou Area,Xinjiang,China[J].Earth Science Informatics,2024,17:1 329-1 343.
- [16] CHEN Y L,DU X D,GUO M.Self-paced ensemble for constructing an efficient robust high-performance classification model for detecting mineralization anomalies from geochemical exploration data[J].Ore Geology Reviews,2023,157:105418.
- [17] GUO M,CHEN Y L.High-performance imbalanced learning ensembles of decision trees for detecting mineralization anomalies from geochemical exploration data[J].Journal of Geochemical Exploration,2024,259:107443.
- [18] CHAWLA N V,BOWYER K W,HALL L O,et al.SMOTE:Synthetic minority over-sampling technique[J].Journal of Artificial Intelligence Research,2002,16:321-357.
- [19] YOUSRA A,AHMAD K M,BASIT R,et al.Predicting influential blogger's by a novel,hybrid and optimized case based reasoning approach with balanced random forest using imbalanced data[J].IEEE Access,2021,9:6 836-6 854.
- [20] 姜万录,马歆宇,岳毅,等.类间数据不均衡条件下基于平衡随机森林的轴向柱塞泵故障诊断方法[J].液压与气动,2022,46(3):45-54.
- [21] 胡学敏,曾晟,宋良灵.基于灰狼算法改进随机森林算法的爆破振动速度预测研究[J].黄金,2024,45(1):12-16.
- [22] 廖华.陕西旬阳烂木沟金矿控矿构造与矿床成因[D].北京:中国地质大学(北京),2020.
- [23] 许锋,李卫波,宋公社,等.陕西石泉—旬阳地区金矿床控矿因素与成矿模式[J].矿产勘查,2018,9(1):70-78.
- [24] 邹海洋,陈松岭,胡祥昭.陕西旬阳淋湘金矿床成矿机制[J].大地构造与成矿学,1997,21(3):221-227.
- [25] ZUO R G.Selection of an elemental association related to mineralization using spatial analysis[J].Journal of Geochemical Exploration,2018,184:150-157.
- [26] CHEN Y L.Mineral potential mapping with a restricted boltzmann machine[J].Ore Geology Reviews,2015,71:749-760.
- [27] CHEN Y L,WU W.Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data[J].Geochemistry: Exploration, Environment, Analysis,2017,

17(3):231–238.

- [28] PARSA M, MAGHSOUDI A, YOUSEFI M. A receiver operating characteristics-based geochemical data fusion technique for targeting undiscovered mineral deposits[J]. *Natural Resources Research*,

2017,27(1):15–28.

- [29] CHEN Y L, AN A. Application of ant colony algorithm to geochemical anomaly detection [J]. *Journal of Geochemical Exploration*, 2016,164:75–85.

Research on geochemical gold mineralization anomaly prediction based on the SMOTified–BRF model

—A case study of the Hanbin–Xunyang area, Shaanxi Province

Xu Zhenglin¹, Wang Xi¹, Xue Linfu¹, Ran Xiangjin¹, Yan Qun¹, Li Yongsheng^{2,3}, Yu Xiaofei^{2,3}

(1. *College of Earth Sciences, Jilin University*; 2. *Development and Research Center, China Geological Survey*; 3. *Mineral Exploration Technology Guidance Center, Ministry of Natural Resources*)

Abstract: Mineral exploration is not only a critical component of new quality productive forces but also a driving force for their development. To address the challenges of sample imbalance caused by the limited number of known mineralization samples and the scarcity of mineralization information in gold mineralization prediction during mineral exploration, this study proposes the SMOTified–BRF model. This model applies the SMOTE method to augment the extremely limited known mineralization samples and employs the balanced random forest (BRF) method for prediction. Using the Hanbin–Xunyang area as the study area, geochemical data from stream sediments were analyzed using the SMOTified–BRF and BRF models and compared the gold mineralization prediction outcome and model performance. The results show that the SMOTified–BRF model achieves a higher *AUC* value (0.987 5) compared to the BRF model (0.972 6). Additionally, at the optimal threshold indicated by the Youden index, the predicted mineralized area ratio of the SMOTified–BRF model (1.95 %) is significantly smaller than that of the BRF model (12.23 %), demonstrating that the SMOTified–BRF model offers more accurate and efficient performance in predicting gold mineralization anomalies.

Keywords: gold mineralization; anomaly; prediction; balanced random forest; SMOTE; geochemistry; Hanbin–Xunyang area

(上接第 5 页)

- [27] DIONISIO R, RUBIO E M, TORRES P. Smart gateways for IOT–factory integration: Trends and use case [M] // FERREIRA L, LOPES N, SILVA J, et al. *Technological developments in industry 4.0 for business applications*. Hershey: IGI global, 2019: 149–170.
- [28] ALLAHLOH A S, SARFRAZ M, GHALEB A M, et al. Revolutionizing IC genset operations with IIoT and AI: A study on fuel savings and predictive maintenance [J]. *Sustainability*, 2023, 15(11): 8808.

- [29] MOURTZIS D, ANGELOPOULOS J, PANOPOULOS N. Design and development of an edge-computing platform towards 5G technology adoption for improving equipment predictive maintenance [J]. *Procedia Computer Science*, 2022, 200: 611–619.
- [30] 王庆凯, 杨天皓, 莫雪磊, 等. 黄金矿山智能选矿厂建设关键技术研究与应用 [J]. *黄金*, 2023, 44(9): 69–74.

Application of artificial intelligence in predictive maintenance of mining equipment

Li Tao, Sun Yan, Hou Jianshuo, Zhou Bin
(*Wanbao Mining Co., Ltd.*)

Abstract: Predictive maintenance (PdM) leverages data and analytics to anticipate potential failures of system components, enabling preemptive maintenance measures to prevent damage. This approach aims to address predictive challenges in mining equipment maintenance, enhancing equipment reliability and production efficiency. The research process encompasses stages such as data collection, preprocessing, model training, and prediction, as well as decision support and execution. Challenges in implementing PdM with artificial intelligence are analyzed from 3 perspectives: data sources, model transparency and interpretability, and system integration. The findings indicate that AI-based PdM significantly reduces equipment downtime, improves maintenance efficiency, and lowers operating costs. Additionally, the study outlines the application prospects of technologies such as machine learning, IoT, cloud computing, and digital twins in PdM, offering directions for future research.

Keywords: predictive maintenance; artificial intelligence; deep learning; machine learning; digital twin; blockchain technology