



DOI:10.12404/j.issn.1671-1815.2406016

引用格式:曹脊翔,陈思源,肖柏夷,等.基于机器学习的致密砂岩储层成岩相测井识别:以川中地区沙溪庙组一段为例[J].科学技术与工程,2025,25(21):8858-8870.

Cao Jixiang, Chen Siyuan, Xiao Baiyi, et al. Machine learning based diagenetic facies logging identification: a case of Shaximiao Formation in central Sichuan Basin[J]. Science Technology and Engineering, 2025, 25(21): 8858-8870.

# 基于机器学习的致密砂岩储层成岩相测井识别: 以川中地区沙溪庙组一段为例

曹脊翔<sup>1</sup>, 陈思源<sup>2</sup>, 肖柏夷<sup>1</sup>, 杨曦冉<sup>1</sup>, 罗莹莹<sup>3</sup>, 陈宏<sup>4</sup>, 吴丰<sup>2,3\*</sup>

(1. 中石油西南油气田分公司致密油气项目部, 成都 610056; 2. 西南石油大学计算机与软件学院, 成都 610500;

3. 西南石油大学地球科学与技术学院, 成都 610500; 4. 四川兆虹油气田技术有限公司, 成都 610500)

**摘要** 川中地区金秋气田—天府气区沙溪庙组沙一段致密砂岩储层成岩相复杂,给储层评价与天然气勘探开发造成了较大困扰,但传统成岩相识别方法准确率低、对专业人员依赖性强、耗时长,急需准确率高、成本低、速度快的成岩相识别方法。首先,基于铸体薄片鉴定数据,通过组分三端元图确定了致密砂岩岩性,结合图像处理技术确定了孔隙、胶结物的类型与比例,并划分了致密砂岩成岩相。然后,对岩心划分成岩相数据对应的1 019个深度测井数据进行了分布范围、中位数、均匀性、偏斜性等特征分析,通过标准化将6条测井数据转换到了0~1范围,通过合成少数类过采样技术(synthetic minority over-sampling technique, SMOTE)处理数据不均衡问题。最后,选取传统机器学习算法和集成学习算法中的10种方法模型训练与性能对比。研究发现,集成学习算法(特别是极端随机树算法)在成岩相识别中表现最佳,其准确率和 $F_1$ 分数均高于传统机器学习算法,显著提高了识别精度与稳定性。利用构建的极端随机树算法模型对JQ8井的成岩相进行预测验证,验证了该方法的可行性,为致密砂岩成岩相的研究提供了有效的技术手段和参考。

**关键词** 成岩相; 沙溪庙组; 特征分析; 集成学习; 机器学习

中图分类号 TE19;

文献标志码 A

## Machine Learning Based Diagenetic Facies Logging Identification: A Case of Shaximiao Formation in Central Sichuan Basin

CAO Ji-xiang<sup>1</sup>, CHEN Si-yuan<sup>2</sup>, XIAO Bai-yi<sup>1</sup>, YANG Xi-ran<sup>1</sup>,

LUO Ying-ying<sup>3</sup>, CHEN Hong<sup>4</sup>, WU Feng<sup>2,3\*</sup>

(1. Tight Oil and Gas Project Department, PetroChina Southwest Oil and Gas Field Company, Chengdu 610056, China;

2. School of Computer Science and Software Engineering, Southwest Petroleum University, Chengdu 610500, China;

3. School of Geoscience and Technology, Southwest Petroleum University, Chengdu 610500, China;

4. Sichuan Rainbow Oil and Gas Field Technology Company, Chengdu 610500, China)

**[Abstract]** The complex diagenetic facies of the tight sandstone reservoir in the Shaximiao Formation, located in the Jinqu gas field to Tianfu gas area in the central Sichuan region, pose significant challenges to reservoir evaluation and natural gas exploration and development. Traditional diagenetic facies identification methods are often low in accuracy, heavily reliant on specialized personnel, and time-consuming. There is an urgent need for a diagenetic facies identification method that is highly accurate, cost-effective, and fast. Firstly, based on cast thin section identification data, the lithology of the tight sandstone was determined using a ternary plot of components. Image processing techniques were then used to identify the types and proportions of pores and cements, and the diagenetic facies of the tight sandstone were classified. Secondly, the corresponding 1 019 depth-based well log data for core-divided diagenetic facies were analyzed in terms of distribution range, median, uniformity, and skewness. These 6 types of well log data were standardized to a 0-1 range, and data imbalance was addressed using synthetic minority over-sampling technique (SMOTE). Finally, 10 traditional machine learning algorithms and ensemble learning algorithms were selected for model training and performance comparison. The study

收稿日期: 2024-08-10 修订日期: 2025-04-11

基金项目: 四川省重点研发计划(重大科技专项)(2020YFSY0039);国家自然科学基金区域创新发展联合基金(U20A20266);中国石油-西南石油大学创新联合体科技合作项目(2020CX030103)

第一作者: 曹脊翔(1991—),男,汉族,四川仁寿人,硕士,工程师。研究方向:综合地质。E-mail:caojixiang@petrochina.com.cn。

\*通信作者: 吴丰(1983—),男,汉族,湖北公安人,博士,副教授。研究方向:非常规储层评价。E-mail:wufengzh@swpu.edu.cn。

found that ensemble learning algorithms, especially the extreme randomized trees (ET) algorithm, performs best in diagenetic facies identification, achieving higher accuracy and  $F_1$  scores than traditional machine learning algorithms. This significantly improved identification accuracy and stability. The ET model was then used to predict the diagenetic facies of the JQ8 well, validating the feasibility of the method. This study provides effective technical methods and references for diagenetic facies research in tight sandstones.

[**Keywords**] diagenetic facies; Shaximiao Formation; feature analysis; ensemble learning; machine learning

致密砂岩广泛分布于鄂尔多斯、四川、准噶尔、松辽、吐哈等沉积盆地,是致密油气勘探开发的主要对象<sup>[1-4]</sup>。相比常规砂岩,致密砂岩的孔隙结构复杂、孔渗极低、非均质性强,主要原因是其成岩作用复杂且强烈。成岩相是决定致密砂岩储层是否有效的关键因素之一,尤其在构造、油气来源、盖层等成藏条件确定的前提下,成岩相的准确识别对于致密油气的寻找与评价具有重要指导作用<sup>[3-4]</sup>。

四川盆地致密气主要分布在侏罗系沙溪庙组和上三叠统须家河组,具有烃源条件优、储层分布广、通源断裂发育、古今构造有利、埋藏深度浅、纵向多层含气、天然气品质优等特点,其致密气地质资源量达  $6.9 \times 10^{12} \text{ m}^3$ <sup>[5-6]</sup>。川中—川西过渡带的金秋气田—天府气区侏罗统沙溪庙组,以其河道砂体发育、横向分布稳定被认为是致密砂岩气藏的有利勘探区<sup>[7-12]</sup>。截至 2023 年,累计提交探明储量超过千亿立方米,为目前四川盆地侏罗统沙溪庙组最大规模已探明储量气田,开发潜力巨大<sup>[13]</sup>。因此,开展致密砂岩成岩相的研究有助于更好评价与预测金秋气田—天府气区致密砂岩气藏,对后续致密气的勘探与开发具有重要指导意义。

早期的成岩相识别主要依赖于岩心铸体薄片、阴极发光、电镜扫描等岩心实验数据,但基于岩心实验的成岩相识别方法成本高、耗时久、资料有限,无法实现单井纵向深度上的成岩相连续识别,这给致密砂岩的储层评价造成了一定困扰<sup>[14]</sup>。随后,有不少学者利用测井资料开展成岩相识别工作,并取得了不错的进展<sup>[15-16]</sup>。近些年,随着人工智能技术的快速发展,机器学习方法在地学领域的应用突飞猛进<sup>[17-19]</sup>,在成岩相识别方向也展现出了巨大的潜力<sup>[20-26]</sup>。但目前基于机器学习的致密砂岩储层成岩相研究还相对较少,尤其缺少基于实验结果标定的不同机器学习方法稳定性与可行性对比。

基于金秋气田—天府气区沙溪庙组沙一段 10 余口井的岩心和测井数据,通过标准钻井作业获得岩心样本,通过测井仪器采集测井数据,所有数据经过严格质量控制,确保数据的准确性和可靠性。室内分析工作包括岩心铸体薄片鉴定,矿物组成、胶结物类型、孔隙结构等特征的识别,依据这些特征划分成岩相类型。在数据处理过程中,所有测井

数据均进行了标准化处理,使用合成少数类过采样技术 (synthetic minority over-sampling technique, SMOTE) 对少数类样本进行过采样,以平衡数据集。最后,使用多种机器学习算法对数据进行了训练,评估算法的分类性能,最终选定最佳模型,用于致密砂岩储层成岩相的快速、准确识别。结合多种机器学习算法与合成少数类过采样技术 (SMOTE),用于致密砂岩储层成岩相的自动化预测。相比传统的基于经验和手工分析的方法,不仅能够提高成岩相识别的准确性,还能有效解决数据不平衡问题,使得模型在复杂地质环境下具有较强的泛化能力。同时也为类似复杂地质环境下的成岩相识别提供了新的思路,具有较好的推广前景。

## 1 研究区地质概况

四川盆地处于古扬子板块西缘,是一个多旋回叠合盆地,经历了复杂的构造演化历史,为中国西部最大的含油、富气盆地之一。金秋气田—天府气区位于川中平缓褶皱带和川北低平褶皱带的过渡区域,如图 1 所示。研究区侏罗系自上而下发育蓬莱镇组、遂宁组、沙溪庙组、凉高山组和自流井组 5 套地层<sup>[27]</sup>。其中,沙溪庙组为一套巨厚的陆相碎屑岩红色地层,夹中厚层块状砂岩,纵向上以区域标志层暗色“叶肢介页岩”为分界,从下至上可划分为沙一段与沙二段<sup>[28-29]</sup>。沙一段发育紫红色泥岩、泥质粉砂岩、灰色砂岩的韵律层,底部以一套灰色厚砂岩与凉高山组区分;沙二段主要由紫红色泥岩与灰色砂岩组成不等互层。地层沉积环境为三角洲—湖泊沉积体系,分流河道与河口坝叠置连片形成厚层规模网状砂组,砂体厚度为 15 ~ 30 m,为天然气的输导和储集提供重要通道和空间<sup>[30]</sup>。

## 2 成岩相识别基本流程

致密砂岩成岩相识别分为 3 个步骤,如图 2 所示。

(1) 岩心图像分析与成岩相划分。基于岩心铸体薄片照片,通过图像分析确定致密砂岩的矿物组成、胶结物类型和孔隙结构等特征,根据成岩演化过程和储集性能差异,将研究区致密砂岩划分为不同类型的成岩相。

(2) 数据预处理与样本重采样。收集各类成岩

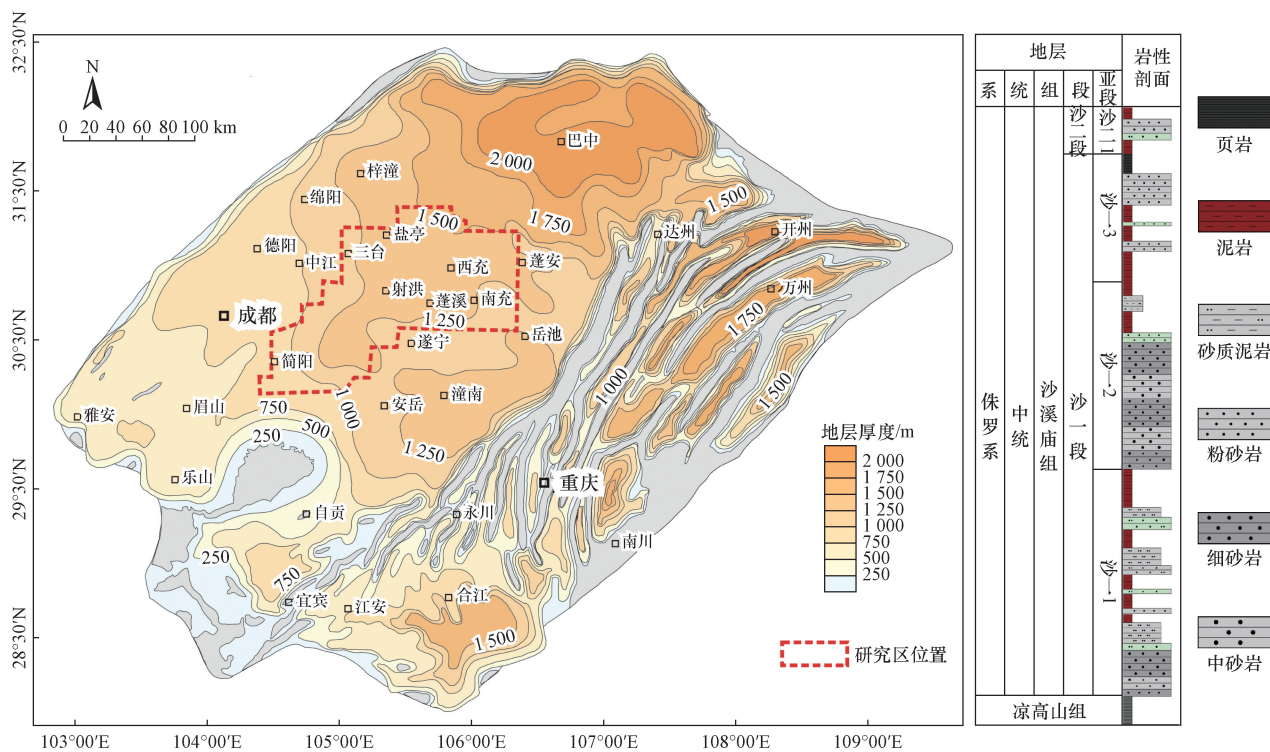


图1 金秋气田—天府气区地理位置、构造与地层特征

Fig. 1 Geographical location, structure, and stratigraphic characteristics from Jinqiu gas field to Tianfu gas area

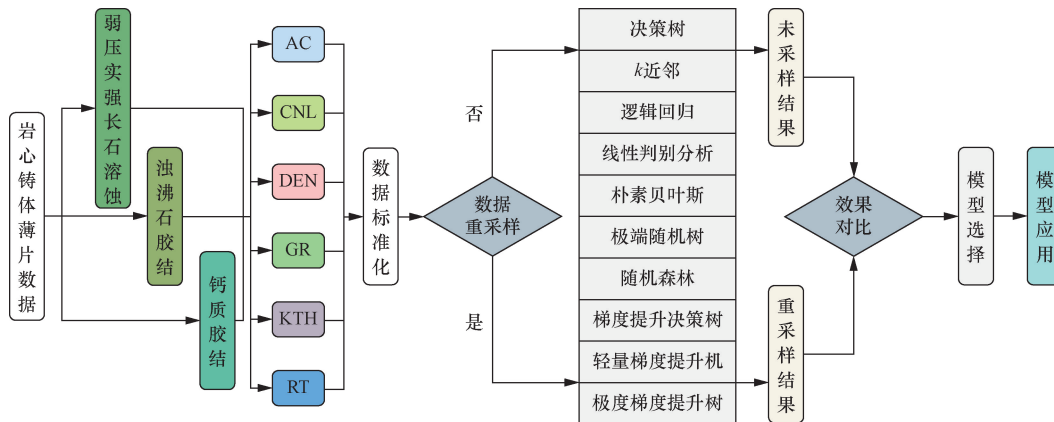


图2 成岩相识别方法研究流程

Fig. 2 Research process of diagenetic facies identification method

相对应的测井数据,对其进行数据清洗和标准化,并采用合成少数类过采样技术(SMOTE)解决样本不均衡问题。

(3)机器学习建模与性能对比。数据处理完成后,采用目前流行的10种机器学习方法开展模型训练,对比不同模型的成岩相识别效果与优缺点,从而筛选出表现最优的机器学习模型。

### 3 基于岩心的成岩相分类

#### 3.1 岩性特征

沙溪庙组沙一段砂岩的成分特征复杂多样。

岩性主要类型为岩屑长石砂岩,其次是长石岩屑砂岩和岩屑砂岩,如图3所示。砂岩分选中等,磨圆为次棱角-次圆状,砂岩的成分成熟度较低,其中石英含量介于35%~60%,长石含量介于25%~45%,岩屑含量介于9%~54%。砂岩的岩性成分直接影响其成岩过程及成岩相的发育。不同矿物组成、颗粒大小和分选性会导致成岩过程中压实、胶结和溶蚀等作用的强弱差异,从而形成不同类型的成岩相。这些岩性成分和成岩相之间的关系对于准确识别和划分成岩相,进而评估储层质量至关重要。

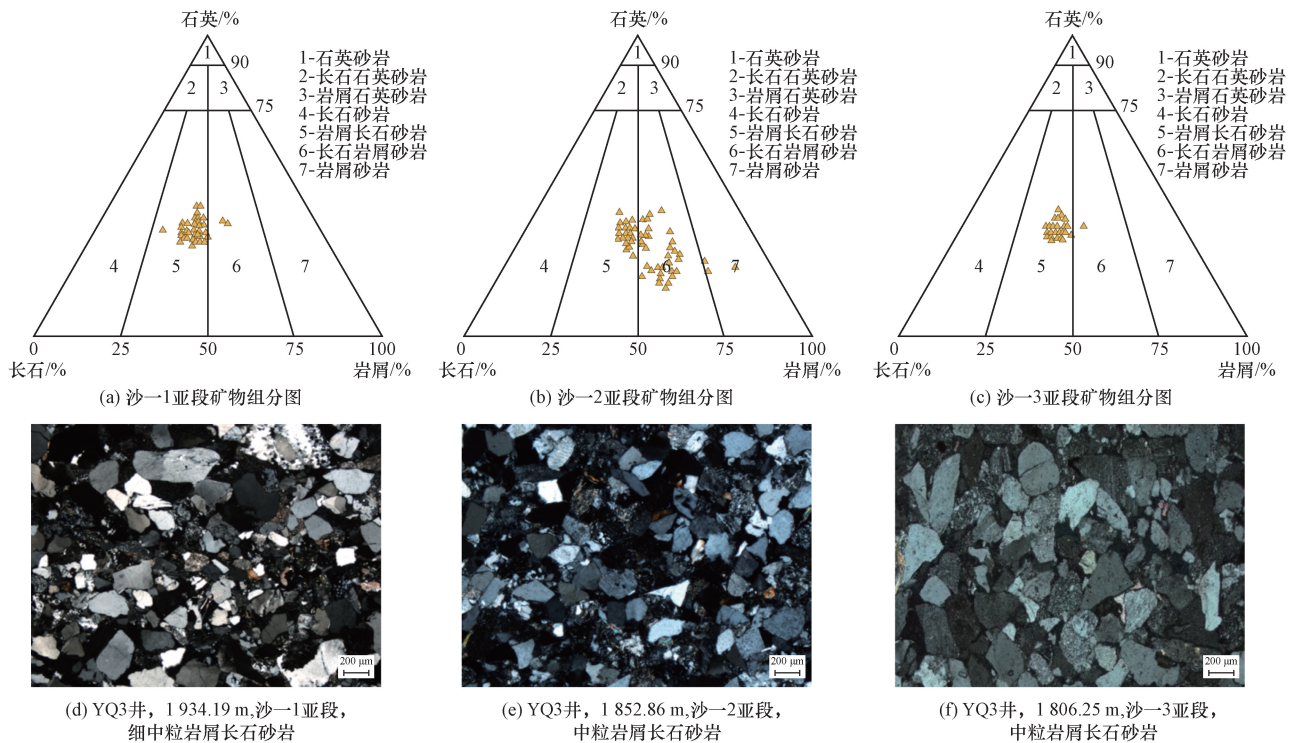


图3 沙一1、沙一2、沙一3亚段岩性三元图与铸体薄片

Fig. 3 Lithological ternary map and cast thin film of 3 subsections of the first member of Shaxi Formation

### 3.2 成岩相分类

成岩作用主要包括压实作用、胶结作用、溶蚀作用等<sup>[31-33]</sup>。压实作用是指沉积物在埋藏过程中由于上覆地层压力增大而导致颗粒间隙减小和孔隙度降低;胶结作用是矿物质在沉积物孔隙中沉淀,填充颗粒间隙,使岩石硬化;溶蚀作用是地质流体对矿物颗粒和胶结物的化学溶解,导致孔隙度增加。研究区现今埋藏深度为1 500~3 000 m,较浅的埋藏深度使得压实作用强度整体表现为弱压实。但砂岩的胶结物种类繁多且分布广泛,包括方解石胶结、硅质胶结、自生长石胶结、浊沸石胶结和黏土胶结,这些胶结物显著影响致密砂岩的物性。根据铸体薄片鉴定,研究区主要发育弱压实强长石溶蚀、浊沸石胶结和钙质胶结3种成岩相。首先为弱压实强长石溶蚀成岩相,粒间粒内孔与粒内溶蚀孔发育,如图4(a)~图4(c)所示。其次为浊沸石胶结成岩相,粒间粒内孔发育,浊沸石胶结强非均质性,如图4(d)~图4(f)所示。少量砂体成岩相属于钙质胶结成岩相,粒间粒内孔发育,见钙质胶结,如图4(g)~图4(i)所示。

## 4 数据特征分析与标准化处理

### 4.1 数据样本特征分析

数据样本分布可以反映总体趋势、模式、均匀性、偏斜性、离散程度及极端值和异常值等信息。分析样

本分布可以识别数据的集中趋势和偏差,了解类别比例关系和不平衡情况,挖掘潜在规律和特征。为进一步的数据分析和模型构建提供依据,优化分析策略和决策。研究样本来源于金秋气田—天府气区沙溪庙组沙一段10余口井的铸体薄片鉴定结果,对应1 019个测井曲线数据。弱压实强长石溶蚀样本最多,共649个;浊沸石胶结样本占282个;钙质胶结样本数量最少,仅88个。3类成岩相样本在测井曲线上分布不均,显示在地质过程的复杂性和多样性。尽管如此,成岩相样本在测井曲线上的样本区间及中位数差异并不明显,反映了相似的成岩演化过程或相似的地质控制因素,如图5所示。

### 4.2 数据标准化与重采样

在数据预处理过程中,不同类型的测井曲线属性值差异较大,可能影响模型识别的准确度,数据标准化和重采样是数据预处理中常用的技术,数据标准化是指将数据按照一定的规则缩放,使不同特征的数据具有统一比例和范围。标准化后的数据可以消除不同特征间的量纲影响,使得特征对模型的影响权重均衡,提高模型训练的稳定性和效果。因此,对AC(声波时差)、CNL(中子)、DEN(密度)、GR(自然伽马)、KTH(无铀伽马)和RT(电阻率)这6种测井曲线进行标准化处理,如式(1)所示。通过将原始数据进行变换到均值为0,标准差为1范围内,确保不同曲线间公平比较,提高岩相识别准确性。

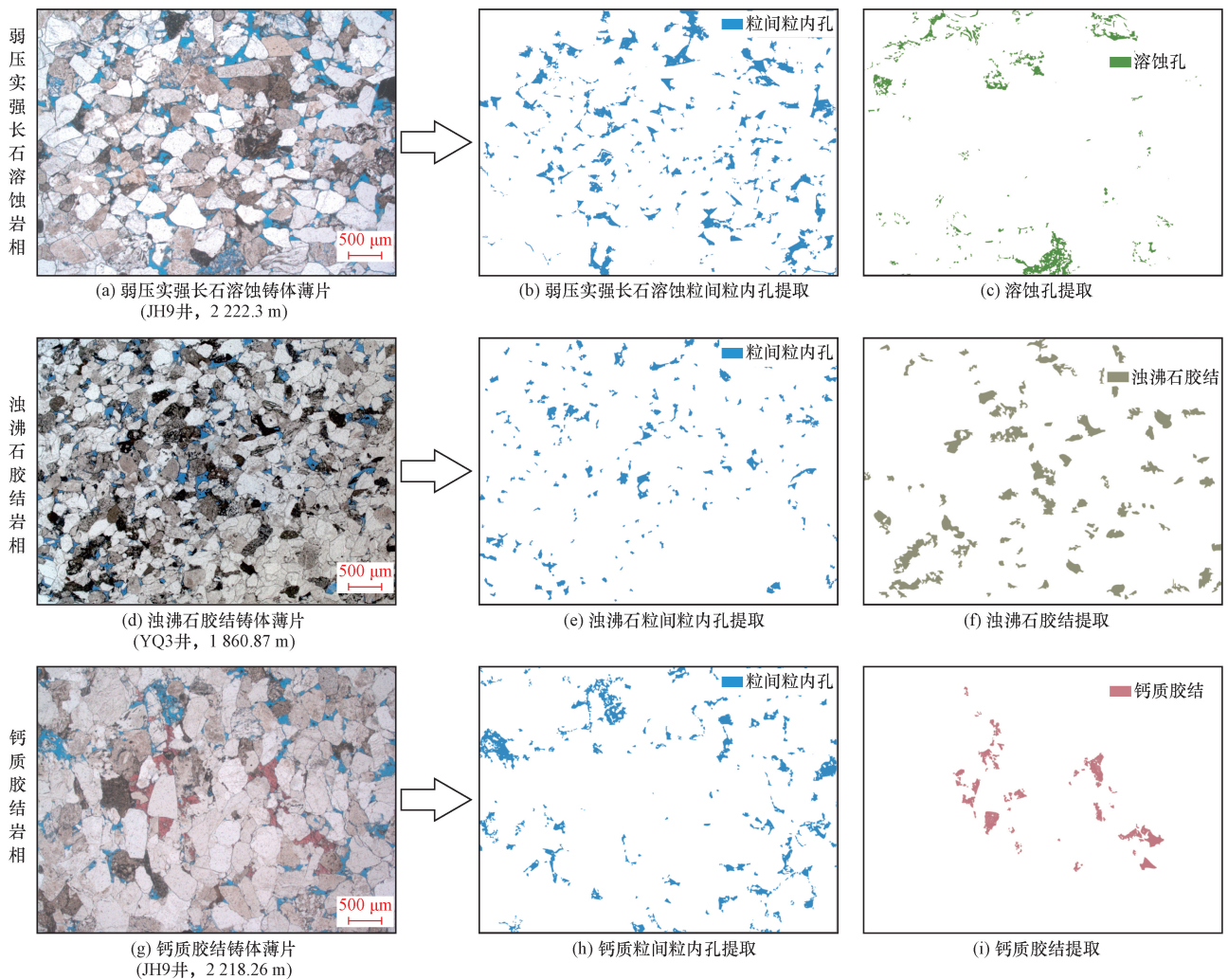


图4 沙一段3种典型成岩相的铸体薄片、孔隙结构与胶结物特征

Fig. 4 Cast thin sections, pore structures, and cement characteristics of three typical diagenetic phases

$$X' = \frac{X - \text{mean}}{\sigma} \quad (1)$$

式(1)中: $X'$ 为标准化后的测井曲线值; $X$ 为原始测井曲线值; $\text{mean}$ 为该测井曲线所在样本均值; $\sigma$ 为标准差。

数据重采样是通过调整样本数量,使得不同类别或特征的样本在训练过程中更为平衡。重采样包括欠采样和过采样:欠采样减少多数类样本来均衡数据,过采样增加少数类样本来均衡数据。使用合成少数类过采样技术 SMOTE 对少数类样本进行过采样,增加其数量,降低样本分布不均衡程度,提高对少数样本的识别准确率<sup>[34-35]</sup>,如图6所示。SMOTE是一种改进的随机过采样技术。其算法流程为:首先对于少数类样本 $x$ 计算它到临近样本 $x_i$ 之间的差值,将差值乘以 $0 \sim 1$ 的随机数,将此差值添加到样本 $x$ 中,以在特征空间中生成新的样本,如式(2)所示。将80%的样本作为训练集,20%的样本作为测试集,重采样后的训练集中,

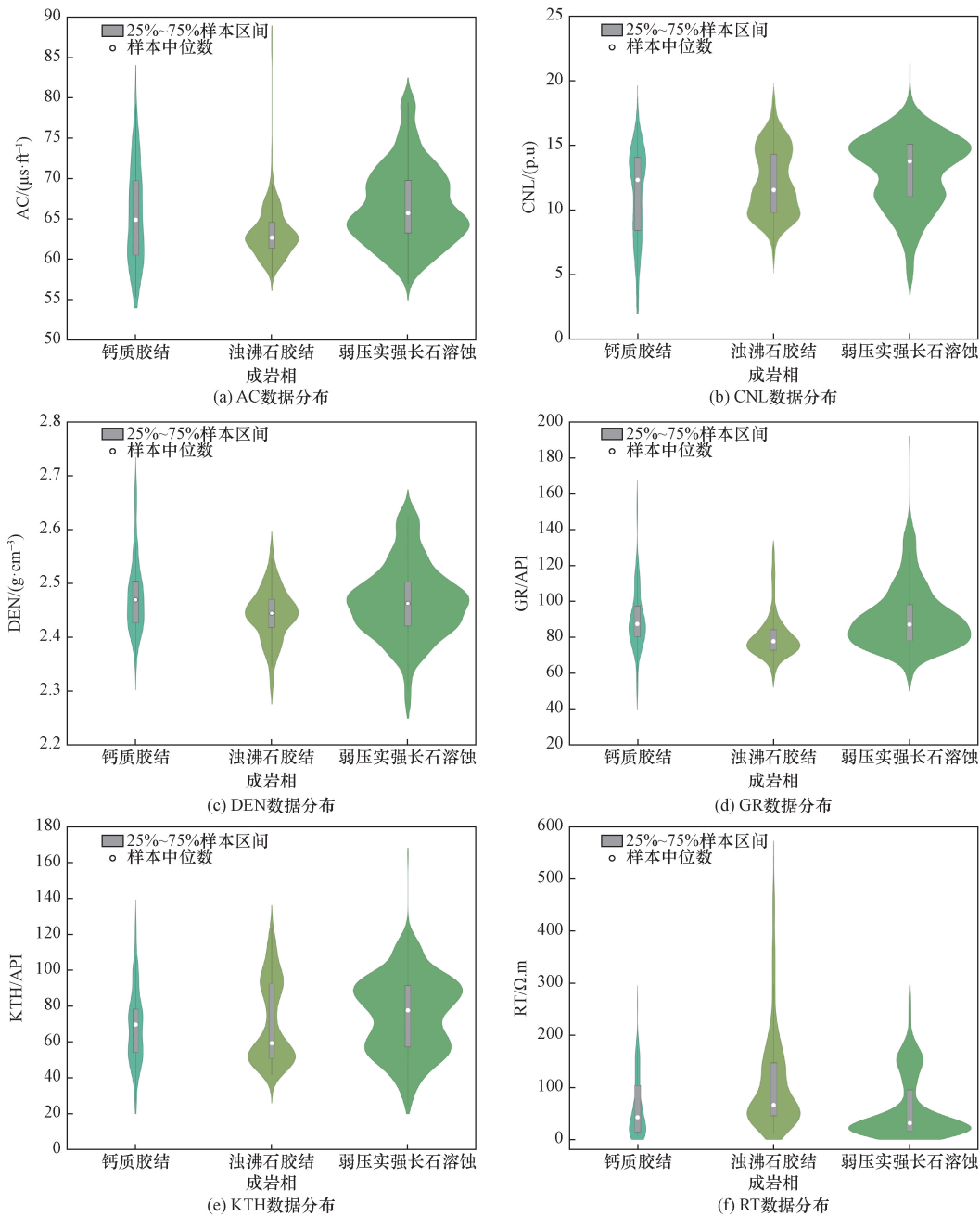
岩相样本不均衡消失,通过标准化处理和 SMOTE 算法,有效解决了样本数量不均衡问题,提高了模型对少数类成岩相的识别准确率,从而增强了模型的泛化能力和实际应用效果,如表1所示。

$$X_{\text{new}} = x + \text{rand}(0,1)(x_i - x) \quad (2)$$

式(2)中: $X_{\text{new}}$ 为构建的新样本; $x$ 为少数类中某一个样本值; $x_i$ 为 $x$ 的第 $i$ 个临近样本( $i = 1, 2, \dots, n$ ); $\text{rand}(0,1)$ 为 $0 \sim 1$ 的随机数。

表1 样本数据分布表

成岩相类型	原始数据	训练集	测试集	采样后训练集
钙质胶结	88	70	18	519
浊沸石胶结	282	226	56	519
弱压实强长石溶蚀	649	519	130	519
合计	1 019	815	204	1 557
整体百分比/%	100	79.98	20.02	—



AC 表示声波时差;CNL 表示补偿中子;DEN 表示密度;GR 表示自然伽马;KTH 表示无轴伽马;RT 表示电阻率;1  $\mu\text{s}/\text{ft} = 3.28 \mu\text{s}/\text{m}$

图5 3种成岩相测井数据特征分布

Fig. 5 Characteristics distribution of three lithofacies logging data

## 5 基于机器学习的成岩相识别

### 5.1 分类算法性能对比

在地质特征分类领域,常用的机器学习分类算法主要包括传统分类算法和集成学习算法两大类<sup>[36-44]</sup>。传统机器学习算法进一步分为决策树(decision tree, DT)、 $k$ -近邻( $k$ -nearest neighbors, KNN)、逻辑回归(logistic regression, LR)、线性判别分析(linear discriminant analysis, LDA)和朴素贝叶斯(naive Bayes, NB)等。传统机器学习算法具有理论

简单、计算效率高的优点,但在处理复杂非线性关系和高维数据时表现不够理想。集成学习算法进一步分为随机森林(random forest, RF)、极端随机树(extra trees, ET)、梯度提升决策树(CatBoost, CB)、轻量梯度提升机(light gradient boosting, LGB)和极度梯度提升树(extreme gradient boosting, EGB)等。集成学习算法通过结合多个弱分类算法,降低了单一模型容易出现的过拟合风险,显著提高了算法的泛化能力和稳定性,适用于大规模和高维数据的处理。目前,虽然已有大量学者将机器学习分类算法

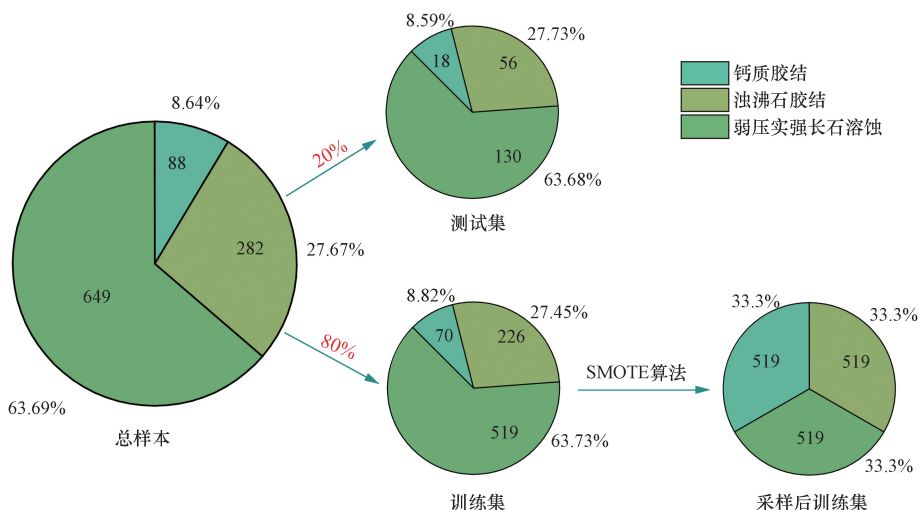


图6 3种成岩相样本数据处理划分

Fig.6 Classification of data processing for three types of lithological facies samples

用于地质现象的识别与分类,但局限于某一种或少数几种分类算法的研究。尤其在致密砂岩成岩相识别方向上,缺少不同分类算法实际应用效果的对比与分析。

采用传统分类算法和集成学习算法中的上述10种分类算法(DT、KNN、LR、LDA、NB、RF、ET、CB、LGB、EGB)开展金秋气田—天府气区沙溪庙组沙一段致密砂岩成岩相识别,并通过多个分类指标对比不同分类算法的性能。机器学习算法性能评估常用的指标包括准确率(accuracy,  $A$ )、精确率(precision,  $P$ )、召回率(recall,  $R$ )、 $F_1$ 分数( $F_1$  score,  $F_1$ )。其中,准确率( $A$ )为正确预测样本数占总样本数的比例,如式(3)所示;精确率( $P$ )为预测为正类的样本中实际为真正例的比例,如式(4)所示;召回率( $R$ )为实际为正类样本中被正确预测为正类的比例,如式(5)所示; $F_1$ 分数( $F_1$ )为精确率( $P$ )和召回率( $R$ )的调和平均,用于综合考虑模型的精确性和召回性能,比单独的精确率( $P$ )和召回率( $R$ )能更好表征算法的优劣,如式(6)所示。

$$A = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (3)$$

$$P = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

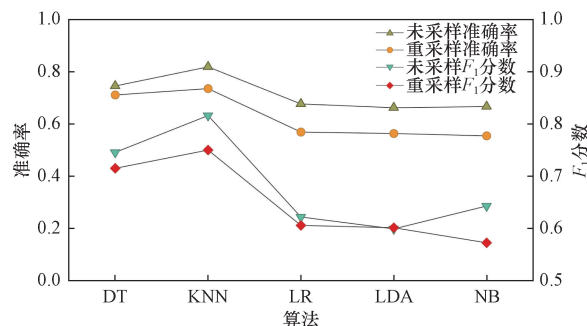
$$R = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

$$F_1 = \frac{2PR}{P + R} \times 100\% \quad (6)$$

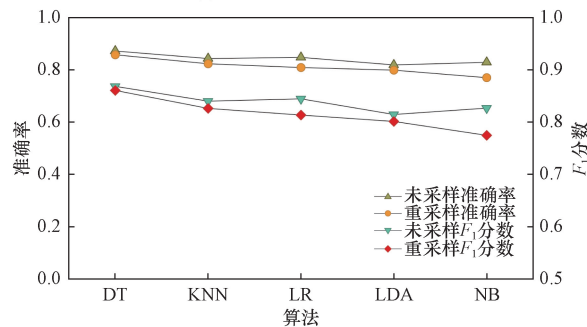
式中:TP为真正例数量,个;TN为真负例数量,个;FP为假正例数量,个;FN为假负例数量,个;TP+TN为正确预测样本数量,个;TP+TN+FP+FN为样本总数量,个。

将数据导入10类机器学习(DT、KNN、LR、

LDA、NB、RF、ET、CB、LGB、EGB)算法进行训练,比较3类成岩相识别准确率和 $F_1$ 分数的变化趋势,如图7所示。结果对比显示,5类集成学习(RF、ET、CB、LGB、EGB)算法的准确率和 $F_1$ 分数均高于5类传统机器学习(DT、KNN、LR、LDA、NB)算法。在集成学习算法中,极端随机树(ET)表现最优,准确率和 $F_1$ 分数为87.2%和86.8%;轻量梯度提升机(LGB)的表现最差,准确率和 $F_1$ 分数为81.8%和81.4%。在传统机器学习算法中, $k$ -近邻(KNN)表现最优,准确率和 $F_1$ 分数为81.8%和81.6%;线性判别分析(LDA)



(a) 传统机器学习算法对比



(b) 集成学习算法对比

图7 3种成岩相采样前后算法准确率、 $F_1$ 分数对比  
Fig.7 Comparison of accuracy and  $F_1$  score of algorithms before and after sampling three types of diagenetic facies

表现最差,准确率和  $F_1$  分数为 66.1% 和 59.8%。但在整体上 10 类机器学习算法对重采样后的数据在准确率和  $F_1$  分数上略低于未重采样的数据,虽然重采样有助于实现数据均衡,但对于 10 类机器学习算法的性能都产生了轻微的负面影响。

### 5.2 分类算法识别结果分析

对以上 10 种分类算法性能的对比分析后发现,集成学习中极端随机树(ET)、随机森林(RF)和梯度提升决策树(CB)是成岩相分类中表现最佳的 3 种算法。使用这 3 种算法对 204 个测试样本进行成岩相识别,其中钙质胶结样本 18 个,浊沸石胶结样本 56 个,弱压实强长石溶蚀样本 130 个。图 8 展示了这 3 种算法数据重采样前后成岩相分类的混淆矩阵对比,横纵坐标分别表示预测结果与实际结果,对角线上的数字越大,其效果越好。极端随机树(ET)算法在对角线上的值最大,表明其正确识别的样本数最多;而随机森林(RF)算法和梯度提升决策树(CB)算法则在部分成岩相上表现稍逊,但总体准确率和可靠性仍然较高。分析混淆矩阵结果并进行统计,如表 2 所示。结果表明,3 种算法在采样前的整体准确率分别为 87.2% (极端随机树 ET)、84.3% (随机森林 RF)、84.8% (梯度提升决策树 CB),略高于采样后的整体准确率 85.7% (极端随机树 ET)、82.3% (随机森林 RF)、

80.8% (梯度提升决策树 CB)。具体来看,极端随机树(ET)算法采样前对钙质胶结、浊沸石胶结和弱压实强长石溶蚀的准确率分别为 55.5%、82.1%、93.8%,采样后为 88.8%、83.9%、86.1%;随机森林(RF)算法采样前分别为 55.5%、78.5%、90.7%,采样后为 83.3%、76.7%、84.6%;梯度提升决策树(CB)算法采样前分别为 61.1%、75%、92.3%,采样后为 88.8%、76.7%、81.5%。

表 2 数据采样前后模型准确率对比  
Table 2 Comparison of model accuracy before and after data sampling

模型	是否采样	极端随机树	随机森林	梯度提升决策树
总准确率/%	否	87.2	84.3	84.8
	是	85.7	82.3	80.8
浮动范围/%	—	-0.15	-2	-4
	钙质胶结准确率/%	否	55.5	55.5
	是	88.8	83.3	88.8
	浮动范围/%	—	+33.3	+27.8
浊沸石胶结准确率/%	否	82.1	78.5	75
	是	83.9	76.7	76.7
浮动范围/%	—	+0.18	-0.18	-0.17
	弱压实强长石溶蚀	否	93.8	90.7
是		86.1	84.6	81.5
浮动范围/%	—	-7.7	-6.1	-10.8

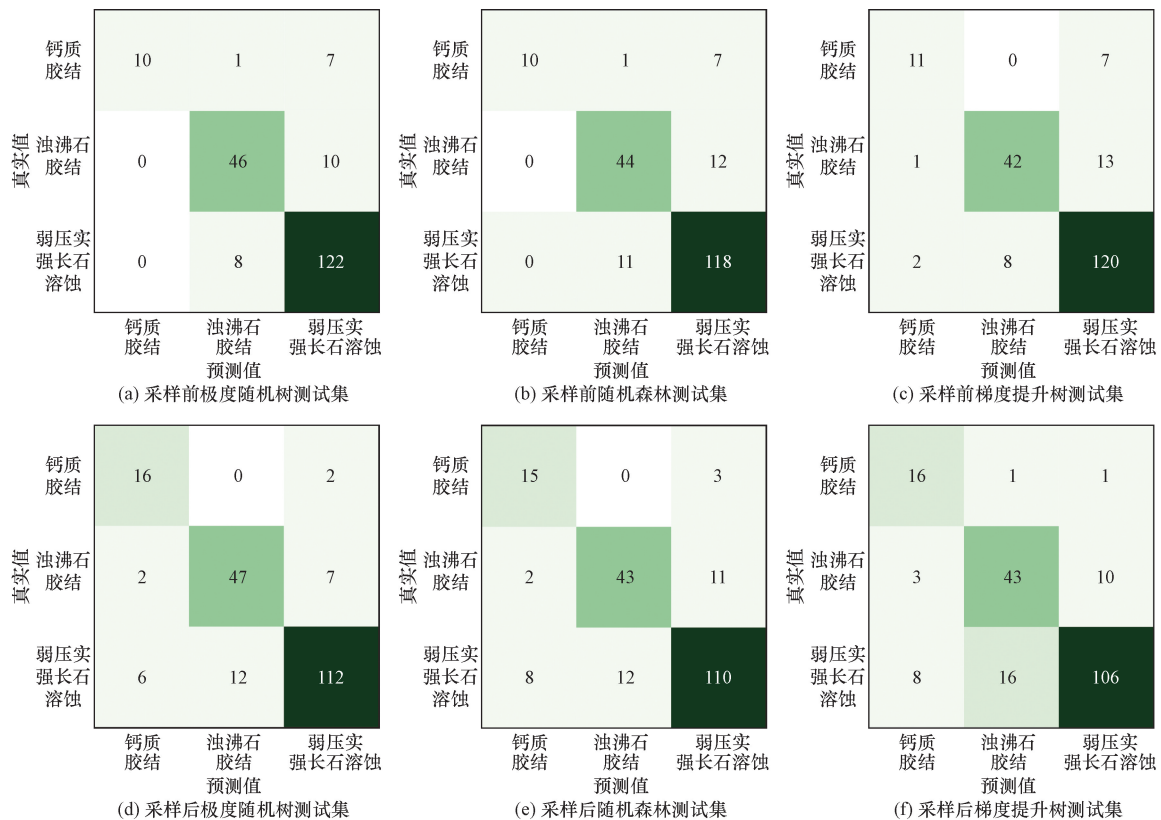


图 8 极端随机树、随机森林和梯度提升决策树算法采样前后混淆矩阵对比

Fig. 8 Comparison of confusion matrices before and after sampling for extra trees, random forest, and CatBoost algorithms

重采样处理后,3种算法对钙质胶结的识别准确率大幅提升,对浊沸石胶结的识别准确率变化不明显,但对弱压实强长石溶蚀的识别准确率有所降低。结果表明,重采样的方法能够更好地适应不平衡数据集,提升少量分布岩相的识别准确率。

观察数据重采样后,极端随机树(ET)、随机森林(RF)和梯度提升决策树(CB)算法在成岩相分类中的性能变化,图像颜色越深代表分类效果越好,如图9所示。3种分类算法中,极端随机树(ET)算法的整体效果最优。尽管由于钙质胶结样本数量少,其识别精度略低于其成岩相。根据受试者操作特征(receiver operating characteristics, ROC)曲线显示,极端随机树(ET)算法各类岩相ROC曲线靠近图像左上角,AUC(area under curve)均优于随机森林(RF)和梯度提升决策树(CB)算法,如图10所示。这表明其在金秋气田—天府气区沙一段储层成岩相分类中表现最佳。

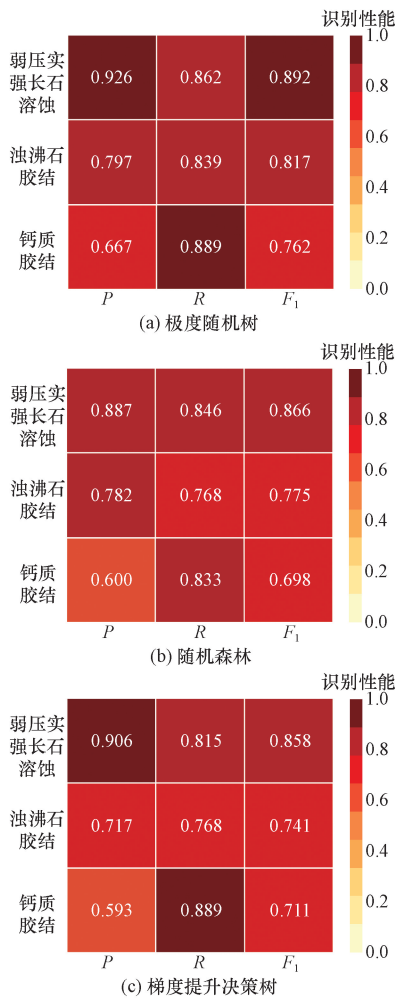
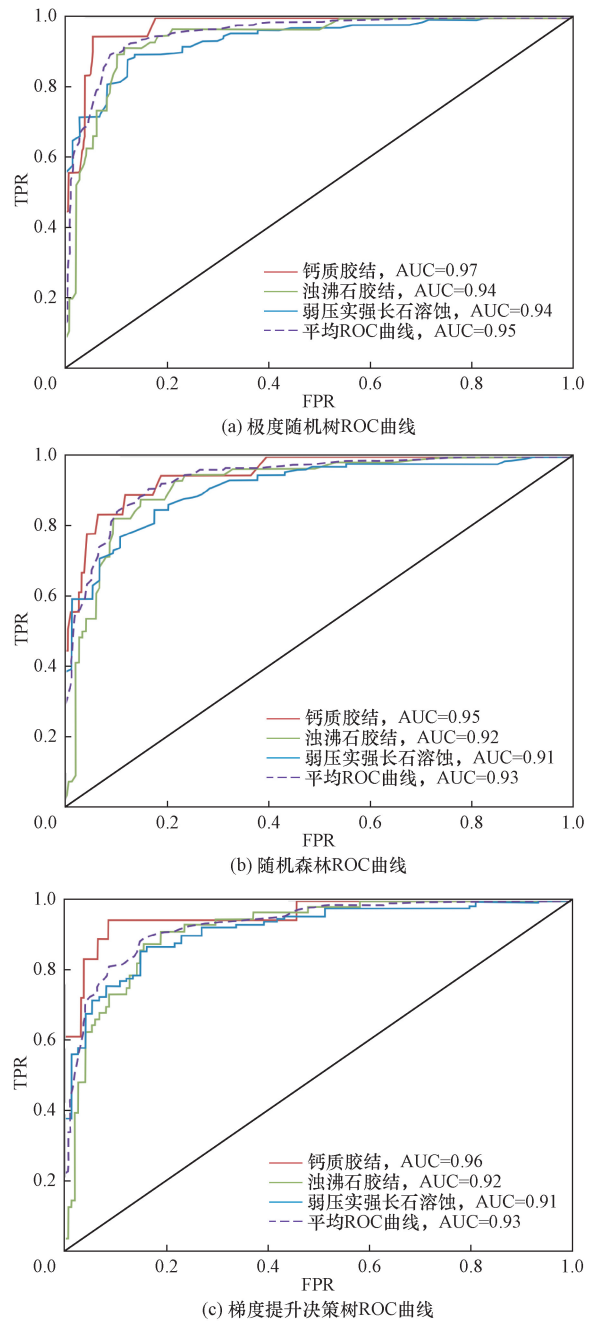


图9 极端随机树、随机森林和梯度提升决策树算法分类性能热力图

Fig. 9 Thermal map of classification performance of extra trees, random forest, and CatBoost algorithms



FPR 为多数类被错误分类的比率;TPR 为召回率

图10 极端随机树、随机森林和梯度提升决策树算法ROC曲线

Fig. 10 ROC curves of extra trees, random forest, and CatBoost algorithms

## 6 应用效果

使用上述3种机器学习模型对JQ8井的85个数据样本进行了预测验证,结果如图11所示。第五道为成岩相解释结论,其中第一列为岩心铸体薄片鉴定结果,第二列为极端随机树(ET)预测结果,第三列为随机森林(RF)预测结果,第四列为梯度提升决策树(CB)预测结果。

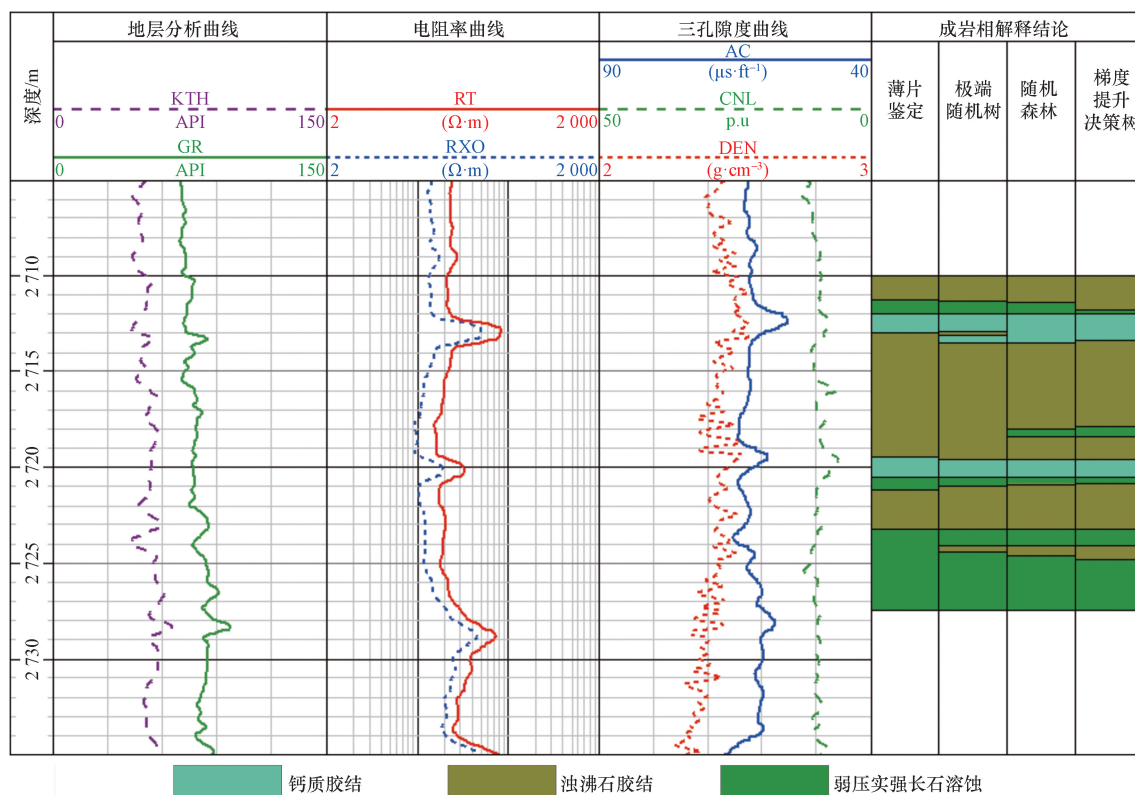


图 11 JQ8 井成岩相识别效果图

Fig. 11 Identification effect diagram of diagenetic facies in JQ8 well

结果表明,极端随机树(ET)算法在成岩相识别上表现最佳,预测结果与薄片鉴定结果的符合率较高。极端随机树(ET)算法正确识别了75个样本,准确率为88.23%;随机森林(RF)算法正确识别了70个样本,准确率为82.35%;梯度提升决策树(CB)算法正确识别了69个样本,准确率为81.17%。进一步分析发现,极端随机树(ET)算法在各类型成岩相的识别上均表现出色,识别准确率更高。相比之下,随机森林(RF)和梯度提升决策树(CB)算法识别准确率稍低。这表明,极端随机树(ET)算法在处理成岩相数据时具有更好的泛化能力和鲁棒性,有效地应对样本数据的复杂性和不均衡性,并能保持较高的预测性能。

综上所述,极端随机树(ET)算法在成岩相识别中具有明显的优势,高准确率和强分类能力使其成为最优模型。而随机森林(RF)算法和梯度提升决策树(CB)算法也表现良好,能够为辅助模型提供可靠的识别结果。这3种模型的综合应用,有助于提高成岩相识别的准确性和效率,为相关地质研究提供支持。

## 7 结论

(1)处理样本分布不均衡的成岩相数据集时,重采样是一种有效的手段。采用了合成少数类过采样技术(SMOTE),显著提高了对少量类样本的识

别效果。增加少数类样本的数量,不仅改善了模型的分类性能,还增强了其稳定性和可靠性。重采样后的模型在识别钙质胶结岩相时尤为突出,有效解决了因样本不均衡导致的识别偏差问题。

(2)在天府气田—金秋气区沙一段储层成岩相识别中,集成学习类算法准确率和 $F_1$ 分数明显优于传统机器学习算法,集成学习通过结合多个弱分类算法来提升整体模型的准确性、稳定性和泛化能力。

(3)在集成学习类算法中,极端随机树(ET)算法表现最佳,无论在测试集还是实例应用中,其整体识别准确率均超过85%。能够有效地识别3种成岩相类型,表现出较高的识别准确率和鲁棒性,即使在复杂地质条件下也能保持良好性能。

## 参 考 文 献

- [1] 贾爱林,位云生,郭智,等. 中国致密砂岩气开发现状与前景展望[J]. 天然气工业, 2022, 42(1): 83-92.  
Jia Ailin, Wei Yunsheng, Guo Zhi, et al. Development status and prospect of tight sandstone gas in China[J]. Natural Gas Industry, 2022, 42(1): 83-92.
- [2] 朱如凯,邹才能,吴松涛,等. 中国陆相致密油形成机理与富集规律[J]. 石油与天然气地质, 2019, 40(6): 1168-1184.  
Zhu Rukai, Zou Caineng, Wu Songtao, et al. Formation mechanism and enrichment law of tight oil in continental China[J]. Oil & Gas Geology, 2019, 40(6): 1168-1184.

- [3] 刘阳, 刘彦成, 林利明, 等. 鄂尔多斯盆地临兴—神府地区上古生界致密砂岩成岩作用与成储关系研究[J]. 科学技术与工程, 2024, 24(4): 1378-1391.  
Liu Yang, Liu Yancheng, Lin Liming, et al. Upper Paleozoic tight sandstone diagenesis and its relationship with reservoir formation in Linxing-Shenfu area of Ordos Basin[J]. Science Technology and Engineering, 2024, 24(4): 1378-1391.
- [4] 李新虎, 丁佳萌, 程光艺, 等. 鄂尔多斯盆地中东部石炭纪本溪组致密砂岩储层含气性测井综合评价[J]. 科学技术与工程, 2024, 24(13): 5364-5371.  
Li Xinhui, Ding Jiameng, Cheng Guangyi, et al. Comprehensive evaluation of gas bearing properties of carboniferous Benxi Formation tight sandstone reservoir in central and eastern Ordos Basin[J]. Science Technology and Engineering, 2024, 24(13): 5364-5371.
- [5] 吴双, 商晓飞, 李蒙, 等. 裂缝性致密砂岩气藏剩余气分布预测及开发优化建议: 以川西坳陷新场地区须家河组二段为例[J]. 科学技术与工程, 2023, 23(29): 12484-12493.  
Wu Shuang, Shang Xiaofei, Li Meng, et al. Prediction of residual gas distribution and suggestion of development optimization for fracture-developed tight sandstone gas reservoir: a case study of the 2nd Member of Xujiahe Formation in Xinchang area, western Sichuan Depression[J]. Science Technology and Engineering, 2023, 23(29): 12484-12493.
- [6] 陈贤良, 纪友亮, 杨克明. 川西中段上沙溪庙组层序格架下成岩相及储层评价[J]. 断块油气田, 2019, 26(5): 550-554.  
Chen Xianliang, Ji Youliang, Yang Keming. Diagenetic facies and reservoir evaluation in sequence framework of Upper Shaximiao Formation in middle area of western Sichuan Depression[J]. Fault-Block Oil & Gas Field, 2019, 26(5): 550-554.
- [7] 马宝全. 低渗透储层测井成岩相及关键参数研究——以鄂尔多斯盆地演武地区延长组 8 段为例[D]. 青岛: 中国石油大学(华东), 2018.  
Ma Baoquan. Study on well-logging diagenetic facies and key parameters in low permeability reservoir: taking Chang 8 member of Yan-chang Formation in Yanwu area, Ordos Basin as an example[D]. Qingdao: China University of Petroleum(East China), 2018.
- [8] 张道伟, 杨雨. 四川盆地陆相致密砂岩气勘探潜力与发展方向[J]. 天然气工业, 2022, 42(1): 1-11.  
Zhang Daowei, Yang Yu. Exploration potential and development direction of continental tight sandstone gas in the Sichuan Basin[J]. Natural Gas Industry, 2022, 42(1): 1-11.
- [9] 戴金星, 倪云燕, 刘全有, 等. 四川超级气盆地[J]. 石油勘探与开发, 2021, 48(6): 1081-1088.  
Dai Jinxing, Ni Yunyan, Liu Quanyou, et al. Sichuan super gas basin in southwest China[J]. Petroleum Exploration and Development, 2021, 48(6): 1081-1088.
- [10] 黄东, 杨跃明, 杨光, 等. 四川盆地侏罗系致密油勘探开发进展与对策[J]. 石油实验地质, 2017, 39(3): 304-310.  
Huang Dong, Yang Yueming, Yang Guang, et al. Countermeasure and progress of exploration and development of Jurassic tight oil in the Sichuan Basin[J]. Petroleum Geology & Experiment, 2017, 39(3): 304-310.
- [11] 肖富森, 韦腾强, 王小娟, 等. 四川盆地川中—川西地区沙溪庙组层序地层特征[J]. 天然气地球科学, 2020, 31(9): 1216-1224.  
Xiao Fusen, Wei Tengqiang, Wang Xiaojuan, et al. Research on the sequence stratigraphy of the Shaximiao Formation in Chuanzhong-Chuanxi area, Sichuan Basin[J]. Natural Gas Geoscience, 2020, 31(9): 1216-1224.
- [12] 杨春龙, 谢增业, 李剑, 等. 四川盆地中侏罗统沙溪庙组天然气地球化学特征及成因[J]. 天然气地球科学, 2021, 32(8): 1117-1126.  
Yang Chunlong, Xie Zengye, Li Jian, et al. Geochemical characteristics and genesis of natural gas in Shaximiao Formation of Middle Jurassic in Sichuan Basin[J]. Natural Gas Geoscience, 2021, 32(8): 1117-1126.
- [13] 黄东, 李育聪, 刘敏, 等. 川中地区中侏罗统沙溪庙组一段油气藏特征及勘探潜力评价[J]. 中国石油勘探, 2017, 22(2): 44-49.  
Huang Dong, Li Yucong, Liu Min, et al. Reservoir features and exploration potential of the 1st member of Shaximiao Formation of Middle Jurassic in central Sichuan Basin[J]. China Petroleum Exploration, 2017, 22(2): 44-49.
- [14] 冉冶, 王贵文, 赖锦, 等. 利用测井交会图法定量表征致密油储层成岩相——以鄂尔多斯盆地华池地区长 7 致密油储层为例[J]. 沉积学报, 2016, 34(4): 694-706.  
Ran Ye, Wang Guiwen, Lai Jin, et al. Quantitative characterization of diagenetic facies in tight oil reservoirs using logging cross plots: a case study of the Chang 7 tight oil reservoir in the Huachi area of the Ordos Basin[J]. Acta Sedimentologica Sinica, 2016, 34(4): 694-706.
- [15] 李明强, 张立强, 李政宏, 等. 塔里木盆地地下侏罗统阿合组下砂砾岩段致密砂岩成岩相划分及测井识别——以库车坳陷奇克里克地区为例[J]. 天然气地球科学, 2021, 32(10): 1559-1570.  
Li Mingqiang, Zhang Liqiang, Li Zhenghong, et al. Diagenetic facies division and logging identification of tight sandstone in the Lower Conglomerate Member of Lower Jurassic Ahe Formation in Tarim Basin: case study of Yiqikelike area in Kuqa Depression [J]. Natural Gas Geoscience, 2021, 32(10): 1559-1570.
- [16] 彭晓勇, 刘国利, 王兵, 等. 陇东地区 HQ 区块东部长 8 致密砂岩储集层成岩相划分[J]. 新疆石油地质, 2023, 44(4): 383-391.  
Peng Xiaoyong, Liu Guoli, Wang Bing, et al. Diagenetic facies division of Chang 8 tight sandstone reservoirs in eastern HQ Block, Longdong area[J]. Xinjiang Petroleum Geology, 2023, 44(4): 383-391.
- [17] 李天平, 李功权. 融合多种深度学习模型的岩石铸体薄片孔隙自动分割方法[J]. 科学技术与工程, 2023, 23(21): 9168-9175.  
Li Tianping, Li Gongquan. Automatic pore segmentation method for rock cast thin section based on multiple deep learning models [J]. Science Technology and Engineering, 2023, 23(21): 9168-9175.
- [18] 苗鹏宇, 王自法, 位栋梁, 等. 基于机器学习方法的地震破坏预测[J]. 科学技术与工程, 2023, 23(14): 5903-5913.  
Miao Pengyu, Wang Zifa, Wei Dongliang, et al. Earthquake damage prediction based on machine learning[J]. Science Technology

- and Engineering, 2023, 23(14): 5903-5913.
- [19] 张哲, 廖广志, 肖立志, 等. 基于机器学习算法预测核磁共振  $T_2$  谱[J]. 科学技术与工程, 2023, 23(17): 7282-7292.  
Zhang Zhe, Liao Guangzhi, Xiao Lizhi, et al. NMR  $T_2$  spectrum prediction method based on machine learning algorithm[J]. Science Technology and Engineering, 2023, 23(17): 7282-7292.
- [20] 岳中文, 闫逸飞, 王煦, 等. 基于随钻数据的岩性识别机器学习算法研究进展[J]. 科学技术与工程, 2023, 23(10): 4044-4057.  
Yue Zhongwen, Yan Yifei, Wang Xu. Research progress of machine learning algorithm for lithology identification based on data while drilling[J]. Science Technology and Engineering, 2023, 23(10): 4044-4057.
- [21] 李昌, 沈安江, 常少英, 等. 机器学习法在碳酸盐岩岩相测井识别中应用及对比——以四川盆地 MX 地区龙王庙组地层为例[J]. 油气藏评价与开发, 2021, 11(4): 586-596.  
Li Chang, Shen Anjiang, Chang Shaoying, et al. Application and contrast of machine learning in carbonate lithofacies log identification: a case study of Longwangmiao Formation of MX area in Sichuan Basin[J]. Petroleum Reservoir Evaluation and Development, 2021, 11(4): 586-596.
- [22] 谌丽, 王才志, 宁从前, 等. 基于机器学习的鄂尔多斯盆地陇东地区区长 7 段岩相测井识别方法[J]. 油气藏评价与开发, 2023, 13(4): 525-536.  
Shen Li, Wang Caizhi, Ning Congqian, et al. Well-log lithofacies classification based on machine learning for Chang-7 Member in Longdong area of Ordos Basin[J]. Petroleum Reservoir Evaluation and Development, 2023, 13(4): 525-536.
- [23] 张驰, 潘懋, 胡水清, 等. 融合储层纵向信息的机器学习岩性识别方法[J]. 地质科技通报, 2023, 42(3): 289-299.  
Zhang Chi, Pan Mao, Hu Shuiqing, et al. A machine learning lithologic identification method combined with vertical reservoir information[J]. Bulletin of Geological Science and Technology, 2023, 42(3): 289-299.
- [24] 刘凯, 邹正银, 王志章, 等. 基于机器学习的火山岩岩性智能识别及预测[J]. 特种油气藏, 2022, 29(1): 38-45.  
Liu Kai, Zou Zhengyin, Wang Zhizhang, et al. Intelligent identification and prediction of lithology of volcanic reservoirs based on machine learning[J]. Special Oil & Gas Reservoirs, 2022, 29(1): 38-45.
- [25] Nishitsuji Y, Exley R. Elastic impedance based facies classification using support vector machine and deep learning[J]. Geophysical Prospecting, 2019, 67(4): 1040-1054.
- [26] Simoes V, Maniar H, Zhao T, et al. A comparative study for machine-learning-based methods for log prediction[J]. Society of Petrophysicists and Well Log Analysts, 2023, 64(2): 192-212.
- [27] 唐大海, 王旭雨, 曾琪, 等. 四川盆地西南部沙溪庙组致密砂岩低渗储层特征及形成机理[J]. 科学技术与工程, 2023, 23(1): 143-154.  
Tang Dahai, Wang Xuli, Zeng Qi, et al. Reservoir characteristics and formation mechanism of low permeability tight sandstone of Shaximiao Formation in southwest Sichuan Basin[J]. Science Technology and Engineering, 2023, 23(1): 143-154.
- [28] 郑有成, 韩旭, 曾冀, 等. 川中地区秋林区块沙溪庙组致密砂岩气藏储层高强度体积压裂之路[J]. 天然气工业, 2021, 41(2): 92-99.  
Zheng Youcheng, Han Xu, Zeng Ji, et al. Practice of high-intensity volume fracturing in the Shaximiao Formation tight sandstone gas reservoirs of the Qiulin Block, central Sichuan Basin[J]. Natural Gas Industry, 2021, 41(2): 92-99.
- [29] 刘忠群, 徐士林, 刘君龙, 等. 四川盆地川西坳陷深层致密砂岩气藏富集规律[J]. 天然气工业, 2020, 40(2): 31-40.  
Liu Zhongqun, Xu Shilin, Liu Junlong, et al. Enrichment laws of deep tight sandstone gas reservoirs in the western Sichuan Depression, Sichuan Basin[J]. Natural Gas Industry, 2020, 40(2): 31-40.
- [30] 韦腾强, 张本健, 王小娟, 等. 四川盆地秋林地区侏罗系沙溪庙组二段河流相沉积特征及储集差异性分析[J]. 科学技术与工程, 2021, 21(29): 12438-12446.  
Wei Tengqiang, Zhang Benjian, Wang Xiaojuan, et al. Sedimentary characteristics of fluvial facies and analysis of reservoir differences in the second member of Jurassic Shaximiao Formation in Qiulin area, Sichuan Basin[J]. Science Technology and Engineering, 2021, 21(29): 12438-12446.
- [31] 陈小虎, 杨申谷, 王青青. 鄂尔多斯盆地坪北地区区长 6 段储层主控成岩作用类型及平面分布[J]. 科学技术与工程, 2022, 22(27): 11801-11811.  
Chen Xiaohu, Yang Shengu, Wang Qingqing. Main diagenesis type and plane distribution of Chang 6 reservoir in Pingbei area, Ordos Basin[J]. Science Technology and Engineering, 2022, 22(27): 11801-11811.
- [32] 于海跃, 肖冬生, 王波. 吐哈盆地深层、超深层致密砂岩成岩作用及成岩相[J]. 断块油气田, 2024, 31(3): 369-378.  
Yu Haiyue, Xiao Dongsheng, Wang Bo. Diagenesis and diagenetic facies of deep and ultra deep tight sandstone in Turpan-Hami Basin[J]. Fault-Block Oil & Gas Field, 2024, 31(3): 369-378.
- [33] 魏真真, 朱善瑜, 王学武. 鄂尔多斯盆地陇东地区区长 8 段储层特征及成岩相定量研究[J]. 科学技术与工程, 2021, 21(29): 12519-12528.  
Wei Zhenzhen, Zhu Shanyu, Wang Xuewu. Research on reservoir characteristics of Chang 8 Oil-Bearing Formation and quantitative analysis of diagenetic facies in Longdong area of Ordos Basin[J]. Science Technology and Engineering, 2021, 21(29): 12519-12528.
- [34] 王晓霞, 李雷孝, 林浩. SMOTE 类算法研究综述[J]. 计算机科学与探索, 2024, 18(5): 1135-1159.  
Wang Xiaoxia, Li Leixiao, Lin Hao. Survey of research on SMOTE type algorithms[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(5): 1135-1159.
- [35] 周玉, 孙红玉, 房倩, 等. 不平衡数据集分类方法研究综述[J]. 计算机应用研究, 2022, 39(6): 1615-1621.  
Zhou Yu, Sun Hongyu, Fang Qian, et al. Review of imbalanced data classification methods[J]. Application Research of Computers, 2022, 39(6): 1615-1621.
- [36] 马陇飞, 萧汉敏, 陶敬伟, 等. 基于深度学习岩性分类的研究与应用[J]. 科学技术与工程, 2022, 22(7): 2609-2617.  
Ma Longfei, Xiao Hanmin, Tao Jingwei, et al. Research and ap-

- plication of lithology classification based on deep learning[J]. Science Technology and Engineering, 2022, 22(7): 2609-2617.
- [37] 赵军, 汪峻宇, 赖强, 等. 基于XGBoost算法的走滑断裂内部特征带的精细识别[J]. 地质科技通报, 2025, 44(2): 182-192.
- Zhao Jun, Wang Junyu, Lai Qiang, et al. Fine-grained identification of internal characteristic zones within the strike-slip fault using the XGBoost algorithm[J]. Bulletin of Geological Science and Technology, 2025, 44(2): 182-192.
- [38] 陈玉林, 李戈理, 杨智新, 等. 基于KNN算法识别合水区长7储层岩性岩相[J]. 测井技术, 2020, 44(2): 182-185.
- Chen Yulin, Li Geli, Yang Zhixin, et al. Identification of lithology and lithofacies of Chang 7 reservoir in Heshui area by KNN algorithm[J]. Well Logging Technology, 2020, 44(2): 182-185.
- [39] 曹文庚, 潘登, 徐邳杰, 等. 河南省滑坡灾害易发性制图研究: 多种机器学习模型的对比[J]. 地质科技通报, 2025, 44(1): 101-111.
- Cao Wengeng, Pan Deng, Xu Zhijie, et al. Landslide disaster vulnerability mapping study in Henan Province: comparison of different machine learning models[J]. Bulletin of Geological Science and Technology, 2025, 44(1): 101-111.
- [40] 谷宇峰, 张道勇, 鲍志东, 等. 利用GS-LightGBM机器学习模型识别致密砂岩地层岩性[J]. 地质科技通报, 2021, 40(4): 224-234.
- Gu Yufeng, Zhang Daoyong, Bao Zhidong, et al. Lithology prediction of tight sandstone formation using GS-LightGBM hybrid machine learning model[J]. Bulletin of Geological Science and Technology, 2021, 40(4): 224-234.
- [41] 陈跃, 王丽雅, 李国富, 等. 基于随机森林算法的低煤阶煤层气开发选区预测[J]. 油气藏评价与开发, 2022, 12(4): 596-603, 616.
- Chen Yue, Wang Liya, Li Guofu, et al. Prediction of favorable areas for low-rank coalbed methane based on random forest algorithm[J]. Petroleum Reservoir Evaluation and Development, 2022, 12(4): 596-603, 616.
- [42] 钱玉贵. 机器深度学习技术在致密砂岩储层预测中的应用——以川西坳陷新场须家河组为例[J]. 油气藏评价与开发, 2023, 13(5): 600-607.
- Qian Yugui. Application of machine deep learning technology in tight sandstones reservoir prediction: a case study of Xujiache Formation in Xinchang, western Sichuan Depression[J]. Petroleum Reservoir Evaluation and Development, 2023, 13(5): 600-607.
- [43] Hamid R O, Mohammad A R, Mohammad M. The effect of supervised feature extraction techniques on the facies classification using machine learning[J]. Journal of Seismic Exploration, 2022, 31(6): 563-577.
- [44] Hall B. Facies classification using machine learning[J]. Leading Edge, 2016, 35(10): 906-909.