

引用格式: 范萱, 陈安清, 郑栋宇, 邓模, 李中超, 周凯, 王瀚, 钟瀚霆, 孙诗, 陆建林, 任强, 侯明才. 基于机器学习的碳酸盐岩测井沉积亚相智能识别——以川东北上二叠统长兴组为例[J]. 沉积学报, 2025. doi: 10.14027/j.issn.1000-0550.2025.058. CSTR: 32268.14/j.cjxb.62-1038.2025.058. [FAN Xuan, CHEN AnQing, ZHENG DongYu, DENG Mo, LI ZhongChao, ZHOU Kai, WANG Han, ZHONG HanTing, SUN Shi, LU JianLin, REN Qiang, HOU MingCai. Intelligent Identification for Carbonate Logging Sedimentary Facies Based on Machine Learning: An Example from the Late Permian Changxing Formation, Northeastern Sichuan Basin[J]. Acta Sedimentologica Sinica. DOI: 10.14027/j.issn.1000-0550.2025.058. CSTR: 32268.14/j.cjxb.62-1038.2025.058.]

基于机器学习的碳酸盐岩测井沉积亚相智能识别 ——以川东北上二叠统长兴组为例

范萱¹, 陈安清¹, 郑栋宇¹, 邓模², 李中超³, 周凯³, 王瀚¹, 钟瀚霆¹, 孙诗¹, 陆建林², 任强¹, 侯明才¹

1. 自然资源部深时地理环境重建与应用重点实验室, 成都理工大学, 成都 610059

2. 中国石化石油勘探开发研究院无锡石油地质研究所, 江苏无锡 214126

3. 中国石化中原油田勘探开发研究院, 河南濮阳 457001

摘要 【目的】精准识别碳酸盐岩沉积相带是开展岩相古地理重建的基础。然而, 沉积相特征的多解性与人工解释的主观性给识别工作带来困难, 且传统测井相解释方法存在效率低下和解释结果受经验制约的局限。本研究旨在通过机器学习算法构建可复用的智能解释模型, 改善测井沉积相解释受限的状况, 提升碳酸盐岩地层沉积相判识精度和效率。【方法】以川东北元坝地区长兴组碳酸盐岩礁滩相区的 5 口钻井为研究对象, 选取声波时差(AC)、自然伽马(GR)等测井曲线作为特征参数, 构建基于改进随机森林(Improved Random Forest, IRF)算法的碳酸盐岩测井沉积相识别模型。采用 SMOTE-Nearmiss-1 混合采样策略解决样本不均衡问题, 并通过网格搜索结合 K 折交叉验证确定最优参数组合。【结果】改进后的模型在目标井段上的盲测结果与外部专家的沉积相分析结果对比, 准确率达 87%, 相较于初始模型有所提升。有效识别出开阔台地、台缘礁滩等优势亚相, 较好反映了区域沉积演化特征。【结论】基于机器学习的沉积相识别方法能够提升识别精度, 并在实际应用中展现出有效性与实用价值。

关键词 碳酸盐岩; 随机森林算法; 沉积相智能分析模型; 长兴组; 四川盆地

第一作者简介 范萱, 女, 2002 年出生, 硕士研究生, 地质学, E-mail: fx2157381472@163.com

通讯作者 陈安清, 男, 教授, 博士生导师, 沉积学和石油地质学, E-mail: aqinth@163.com

中图分类号: P631 文献标志码: A 文章编号: 1000-0550 (2025) 00-0000-00

0 引言

碳酸盐岩储层作为全球油气资源勘探开发的核心目标层系, 其沉积相精准识别不仅是储层非均质性评价的基础, 更为揭示生、储、盖三大成藏基础要素的成因与发育分布规律提供科学指导(陈洪德等, 2004; 陈安清等, 2020)。据统计, 全球约 50% 的常规油气储量赋存于碳酸盐岩储层中(张庆龙等, 2025; 孟宪武等, 2025)。然而, 碳酸盐岩沉积环境的复杂性以及多期成岩作用的叠加效应, 使得岩相古地理重建面临严峻挑战。传统沉积相识别方

收稿日期: 2025-02-18; 收修改稿日期: 2025-10-09

基金项目: 国家自然科学基金项目(U24A20591, 42272132, 42402120); 四川省科技计划项目(2023NSFSC1986)

法主要依赖于岩心观察与测井曲线的经验性解释,但在实际应用中存在一定的局限性。首先,沉积相精确识别需要综合分析自然伽马、电阻率、密度等多组测井曲线,而人工解译过程中可能会忽略大量潜在信息(Rider, 1990; Radwan, 2021);其次,测井数据作为表征地下岩层物化性质的间接参数,易受裂缝、溶蚀孔洞等后期改造作用的影响(周进高等, 2015; 司马立强等, 2017),导致测井沉积相解释工作时间成本高、效率低(Bond *et al.*, 2012; Macrae *et al.*, 2016)。

高效、快速、高准确率的机器学习方法为测井沉积相识别提供了新的方案,研究人员能够凭借此方法提取出所需要的关键信息,并从海量的数据中获得新的发现(Jordan and Mitchell, 2015)。利用机器学习算法开展油气测井数据的沉积相解释已成为当前的研究热点(Al-Mudhafar, 2017; Qi *et al.*, 2022; Zheng *et al.*, 2022; 程希和任战利, 2024),但现有研究主要集中于碎屑岩层系。大量研究表明,机器学习模型在碎屑岩沉积相的测井解释中取得了令人满意的效果(Bhattacharya *et al.*, 2016; Zheng *et al.*, 2022; Hou *et al.*, 2023; 王民等, 2023)。例如,基于局部深度多核学习支持向量机(Local Deep Multi-Kernel Learning Support Vector Machine, LDMKL-SVM)能够结合高维、稀疏的局部特征与低维全局特征,通过自适应优化核函数参数和支持向量机参数,有效提升岩相分类精度(Liu *et al.*, 2020)。随机森林(Random Forest, RF)因其集成学习框架与特征重要性排序能力,在页岩岩相分类中展现出较高的适用性,特别是在岩相类别不均衡的情况下仍能保持良好的分类效果(王民等, 2023)。这得益于碎屑岩沉积体系中不同环境的产物往往具有较为明显的特征差异,有利于机器学习模型的特征学习。相比之下,碳酸盐岩沉积环境因矿物组分相近、成岩作用复杂、空间异质性较强,使得不同沉积相的测井参数区分度较低,且样本分布高度不均衡,进一步增加了机器学习模型的训练难度。一些研究尝试采用支持向量机(Support Vector Machine, SVM)、多分辨率聚类分析(Multi-Resolution Graph-based Clustering, MRGC)、人工神经网络(Artificial Neural Network, ANN)、 k 近邻(K-Nearest Neighbor, KNN)等传统机器学习方法进行碳酸盐岩沉积相分类(李昌等, 2021; Hou *et al.*, 2024)。此外,在缺少岩心资料的情况下,部分研究引入多模态学习方法,将常规测井和电成像测井相结合(李昌等, 2023),或利用注意力机制优化特征权重分配(曾丽丽等, 2022),但这些方法往往依赖高成本的附加数据支持,且模型的可解释性不足。如何在有限测井数据条件下提高碳酸盐岩沉积相识别的准确性和稳定性,机器学习在碳酸盐岩沉积相识别中的应用值得进一步探索。为此,本文基于四川盆地东北部元坝地区上二叠统长兴组的测井数据,构建出一种基于改进随机森林(Improved Random Forest, IRF)算法的碳酸盐岩沉积相识别模型,并采用盲

并验证评估模型的泛化能力。研究成果可为碳酸盐岩测井沉积相识别提供有效参考。

1 地质背景

二叠纪时期，华南板块位于低纬度赤道附近，古特提斯洋东缘（图 1a）。作为扬子板块西北部的重要构造单元（图 1b），四川盆地经历了多期次构造旋回叠合改造，形成了以大型碳酸盐台地为特征的含油气盆地（李雯等，2025）。中一晚二叠世期间，东吴运动、峨眉山地幔柱隆升及勉略洋扩张等多重构造事件共同作用，导致该区发生克拉通裂陷，整体呈现浅水台地—深水陆棚的沉积分异格局（杨帅等，2021；李雯等，2023；Sun *et al.*, 2025）。

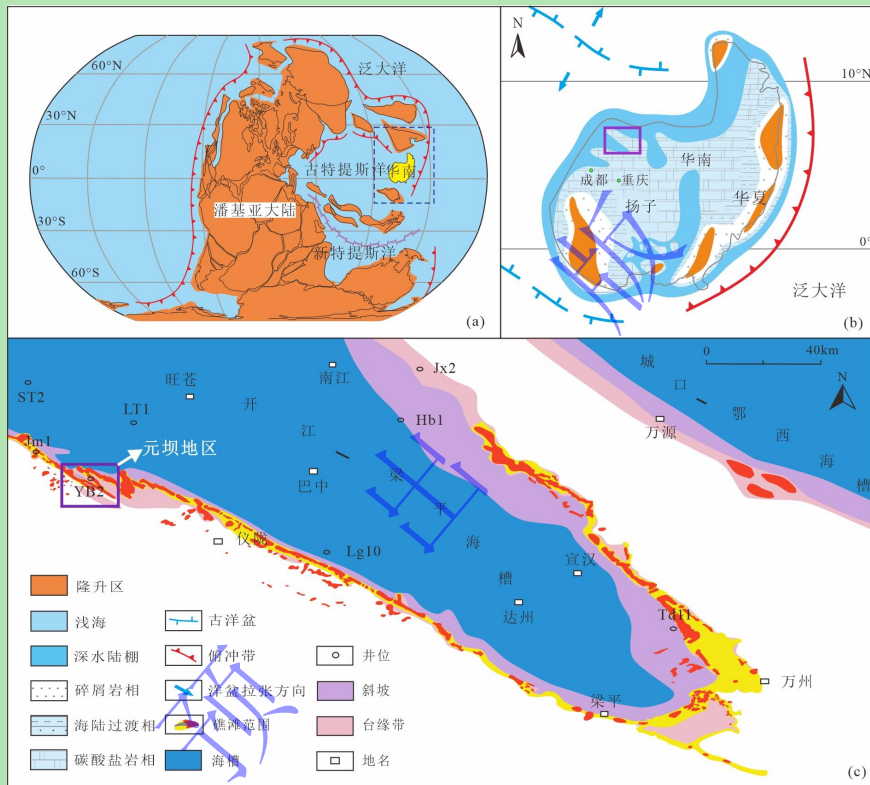


图 1 研究区地质背景图（据何登发等，2020；杨雨等，2021 修改）

（a）晚二叠世全球古地理图；（b）扬子地区二叠纪构造背景图；（c）川东地区位置及长兴组沉积格局图（图中紫色方框为研究区）

Fig.1 Geological setting of study area (modified from He *et al.*, 2020; Yang *et al.*, 2021)

(a) global paleogeographic map of the late Permian; (b) tectonic background map of the Yangtze region during the Permian; (c) location of the eastern Sichuan region and sedimentary pattern map of the Changxing Formation (purple box indicates study area)

晚二叠世长兴期，川东北地区在拉张应力作用下形成了“开江—梁平”海槽（王一刚等，1998），海槽内主要沉积深水泥页岩，环绕该海槽周缘发育多个碳酸盐镶边台地边缘礁滩复合体（图 1c）。随着对这些礁滩复合体的深入勘探，相继发现了元坝、普光、龙岗等多个中大型气田（倪新锋等，2007；马永生等，2014；郭彤楼，2019；武赛军等，2019）。其中位于川北坳陷与川中低缓构造带结合部的元坝气田是目前国内发现的最深层海相气田，含气

层为上二叠统长兴组一下三叠统飞仙关组白云岩储层（李昌峰等，2015），目前已提交探明含气面积 242.14 km²。

2 样本准备

2.1 数据集分析

本文选取了四川盆地元坝气田区 5 口钻井的测井资料作为研究对象，每隔 0.125 m 记录一次测井曲线，共获得 9357 条测井值记录，包括声波时差（AC）、自然伽马（GR）、自然电位（SP）、密度（DEN）和补偿中子（CNL）等五项测井值作为特征参数。自 2006 年气田发现以来，该地区有大量的沉积相研究（王正和等，2012；李昌峰等，2015），沉积相划分已形成共识。各沉积相的测井曲线特征明显，岩心和薄片中的相标注可靠。在此基础上，结合岩心、薄片和测井相标志，将研究区的碳酸盐台地划分为开阔台地、局限台地、台缘礁滩和台缘斜坡四类沉积亚相，为模型提供可信赖的训练标签。

进一步结合取心段不同亚相的测井响应特征，选取典型深度段并对比具有代表性的岩心和薄片图像，构建了 4 类沉积亚相的测井相模板，从而形成了不同沉积环境下的测井响应模式（图 2）。开阔台地亚相岩性以灰岩为主，夹少量白云岩。其典型 GR 曲线呈箱形，数值通常介于 15~80 API。局限台地亚相主要发育生屑灰岩、白云岩。GR 曲线呈钟形，自上而下逐渐降低，反映沉积水动力条件呈先增强后减弱的变化趋势，CNL 值波动较大。台缘礁滩亚相以生屑灰岩为主，GR 曲线值整体较低，呈指形，其余测井曲线形态波动较小，CNL 值较低，反映高能沉积环境。台缘斜坡亚相岩性以泥质灰岩为主，GR 值较高，曲线呈钟形，AC 值较高，DEN 值较低，SP 曲线变化幅度大，偶尔会夹有滑塌角砾岩（王正和等，2012），反映泥质含量高、致密性差的沉积特征。

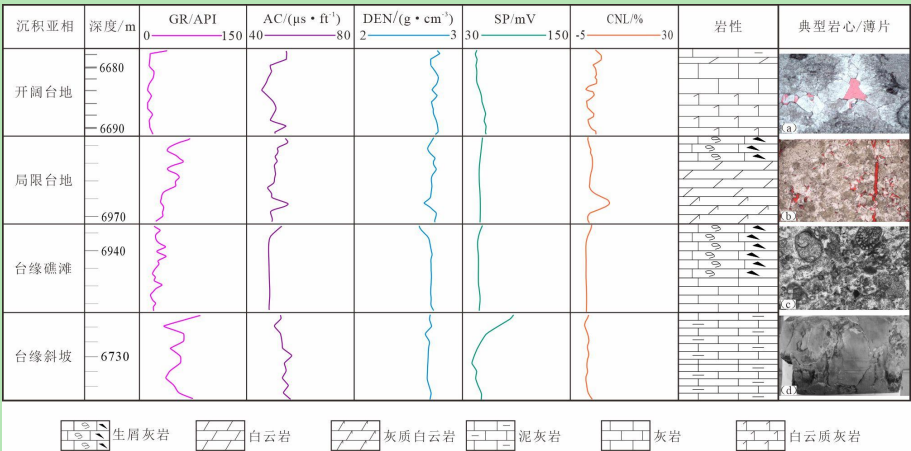


图 2 元坝地区二叠系长兴组测井相模版

(a) 亮晶生屑灰岩, Y2 井; (b) 细晶灰岩, Y16 井; (c) 藻黏结灰岩, Y11 井; (d) 台缘斜坡相滑塌角砾岩, Y10 井

Fig.2 Logging phase template of the Permian Changxing Formation in the Yuanba area(a) sparry bioclastic limestone, well Y2; (b) micritic limestone, well Y16; (c) algal-bound limestone, well Y11; (d) slump breccia of platform-margin slope facies, well Y10

为探讨测井曲线间的相关性，通过交会分析构建交叉图矩阵。图 3 的各子图呈现了 x 轴和 y 轴上两个变量间的关系，其中对角线位置为对应横轴变量的核密度估计图，其余部分为测井参数的两两关系散点图。台缘礁滩亚相在 SP 测井特征上与其他沉积亚相存在显著差异。局限台地亚相和台缘斜坡亚相在 GR 测井特征上具有一定区分度。DEN 测井值与 AC 测井值呈负相关，与 SP 测井值呈正相关。总体而言，不同沉积亚相的测井响应特征存在一定重叠，仅有少数测井参数间存在相关性。因此，交叉图难以对碳酸盐岩沉积亚相进行有效区分，需借助机器学习方法作进一步识别。

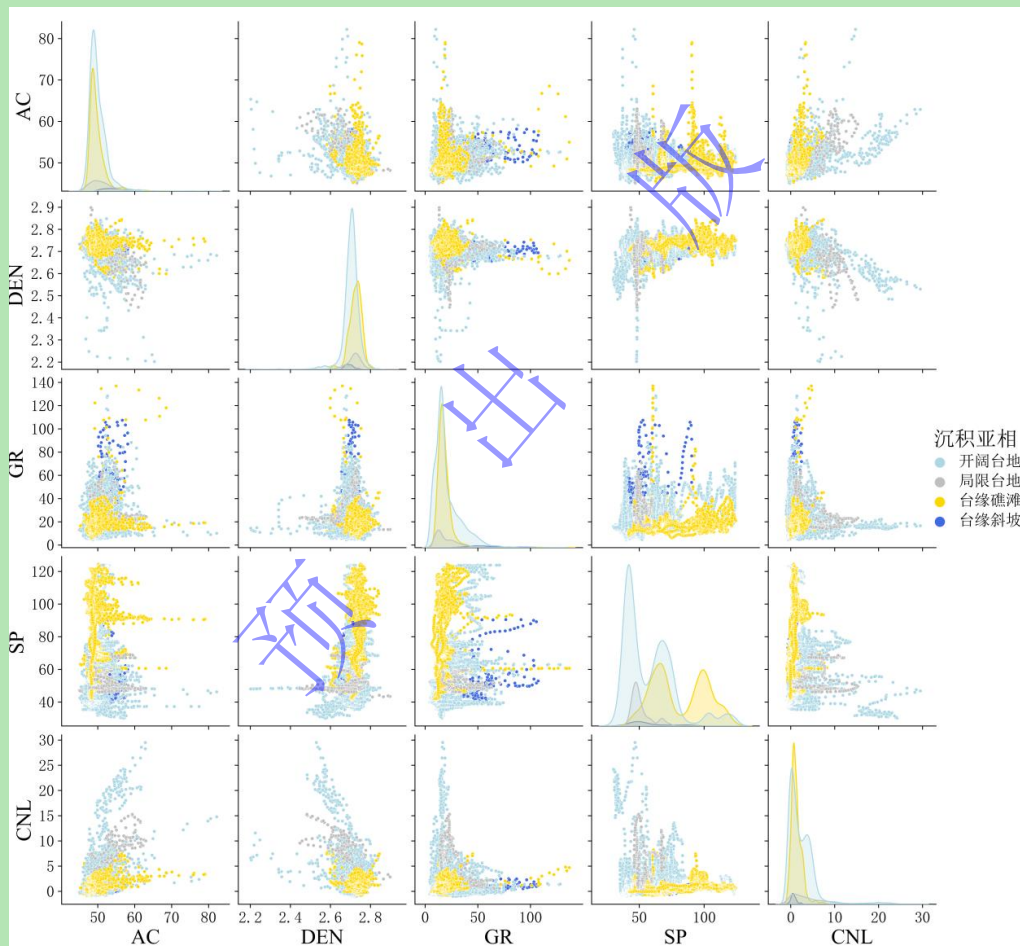


图 3 沉积亚相测井散点图矩阵

Fig.3 Scatter plot matrix of well-logging data for sedimentary subfacies

2.2 数据处理

不同类型测井曲线的量纲和数量级存在差异，若直接将测井数据作为输入引入机器学习模型，其差异性会对模型的识别精度产生影响（孙子舒等，2020）。因此，在进行模型训练前，需对数据进行预处理。本文采用 Min-Max 标准化法对数据进行归一化处理，原始数据

经标准化处理后, 各项指标处于同一数量级, 适合进行综合比较评价。归一化处理的计算公式如下:

$$x_{i,scaled} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中: x_i 为原始数据集中第 i 个测井数据, $x_{\max}$ 为该类测井数据的最大值, $x_{\min}$ 为该类测井数据最小值, $x_{i,scaled}$ 为第 i 个样本归一化后的测井值。

3 方法与应用

3.1 随机森林原理

随机森林作为一种集成学习方法, 通过整合多个决策树模型的预测结果来提高模型的准确性。在决策树的各分裂节点处, 基于基尼不纯度准则选择最优特征, 以实现分裂后的节点纯度最大化。较低的基尼不纯度表明特征更适合决策树节点分裂。此外, 每棵树的构建依赖于自助采样方法 (bootstrap sample) 随机地从原始训练数据集中抽取部分样本生成新的训练集 (Breiman, 1996)。随机森林算法具有高精度、泛化性好和鲁棒性强等优势, 在分类、回归以及异常检测等众多机器学习任务中得到了广泛应用 (Biau and Scornet, 2016; Wang *et al.*, 2020)。

随机森林算法的主要流程如下 (图 4): (1) 为每棵树随机抽样构成训练样本集; (2) 在每个样本的特征维度为 M 的条件下, 从 M 个特征中随机选取 m 个特征 ($m < M$) 组成特征子集, 并选取最优特征进行节点分裂, 生成决策树的左右子树; (3) 集成的 N 棵决策树之间相互独立, 不进行修剪。在分类阶段, 通过投票机制确定分类结果, 投票多的类别即为最终决策结果 (Salman *et al.*, 2024)。

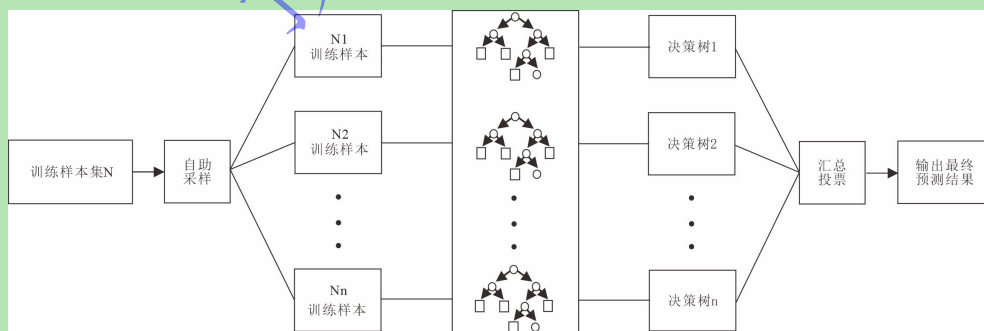


图 4 随机森林算法原理示意图 (据崔俊峰等, 2023 修改)

Fig.4 Schematic diagram of Random Forest Algorithm (modified from Cui *et al.*, 2023)

3.2 数据不平衡处理与参数优化

数据集中 4 种亚相样本数量差距较大。开阔台地亚相的数据量最大, 台缘斜坡亚相数据量最小。如果忽略数据不平衡的特征, 少数类的分类准确率可能并不高, 达不到分类要求并

且会影响机器学习的训练效果。传统采样方法如简单过采样或欠采样在提升少数类识别率的同时，可能引入过拟合风险或信息损失。为平衡样本结构并提升模型效果，采用基于混合采样策略的随机森林不平衡数据分类算法 SMOTE-NearMiss-1。

SMOTE 是一种过采样技术，通过插值的方式生成新的样本，来增加少数类样本的数量（图 5a）。假设少数类样本集为 $S=\{x_1, x_2, \dots, x_n\}$ ，其中每个 x_i 为一个特征向量。对于每个少数类样本 x_i ，计算其与样本集合内其他少数类样本 x_j 的距离：

$$d(x_i, x_j) = \|x_i - x_j\|, j \neq i \quad (2)$$

根据计算出的距离，找到 x_i 的 k 个最近邻。从 k 个最近邻样本中，随机选取一个 x_n ，作为生成新样本的参考点。在 x_i 和 x_n 之间插值生成新样本：

$$x_{new} = x_i + \delta \cdot (x_n - x_i) \quad (3)$$

式中： $\delta \sim U(0,1)$ 是一个服从 $[0,1]$ 区间的均匀分布随机数。对少数类中的每个样本 x_i ，重复上面步骤，生成预定数量的新样本。

NearMiss-1 算法是处理类别不平衡问题的一种欠采样技术（Zhang and Mani, 2003），其核心思想是从多数类样本中选取与少数类样本“平均距离最小”的子集，从而保留最具代表性的多数类样本（图 5b）。

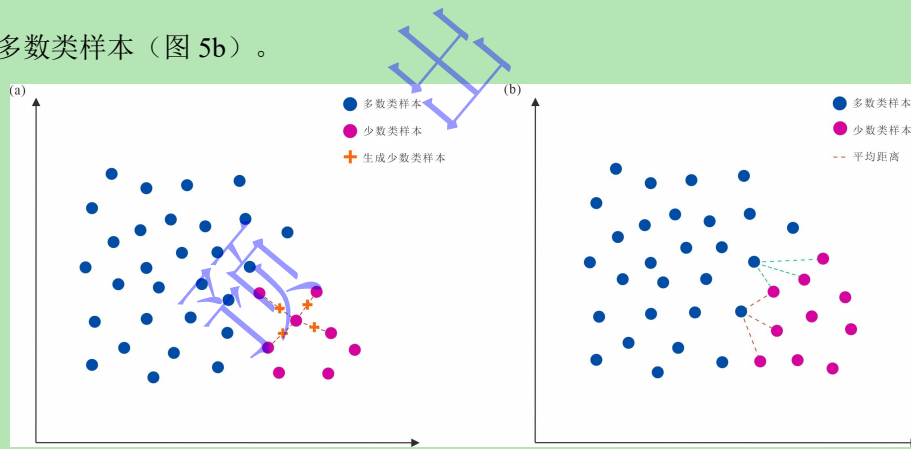


图 5 重采样原理图

(a) SMOTE 过采样技术原理；(b) NearMiss-1 欠采样技术原理

Fig.5 Resampling schematic diagram

(a) principle of the SMOTE oversampling technique; (b) principle of the NearMiss-1 undersampling technique

在数据处理后，采用随机森林算法建立用于识别碳酸盐岩沉积相的预测模型，并将数据集按 7:3 划分为训练集和测试集。对训练集进行 SMOTE- Nearmiss-1 平衡化处理后，利用该训练集进行模型训练。为了提升模型性能，使用网格搜索结合 k 折交叉验证来寻找最优参数组合。由于所采用的数据规模较小，故将 k 值取 5。在五折交叉验证的过程中，训练集被随机分成五等份，轮流取一个子样本集作为验证集，其余四份用于模型训练（Verma *et al.*,

2024)。在网格搜索过程中，对预设的参数空间进行遍历，以评估每组参数的模型性能。最终，以交叉验证评分为评估标准，确定最优参数组合（表 1），降低随机选取训练集对模型准确率的影响。在模型训练与验证过程中，使用固定随机种子以保证实验结果的可重复性。

表 1 模型最优参数组合

Table 1 Optimal parameter combination for the model		
参数	名称	最优参数
max_depth	决策树的最大深度	25
max_features	每棵决策树可使用的最大特征值	log2
min_samples_leaf	在叶子节点处需要的最小样本数	1
min_samples_split	拆分内部节点所需的最小样本数	2
n_estimators	决策树的个数	320

3.3 模型评估与预测结果

3.3.1 预测结果分析

机器学习的目的是通过学习已有样本，以适用新的样本。而过拟合是监督式机器学习中的基本问题（Ying, 2019; 张凯和张科, 2022），因此需要评估模型的泛化性能。为了综合评价算法的预测效果，选择准确率（Accuracy）、召回率（Recall）、精确率（Precision）、F1-score 等指标，从不同角度衡量了模型的性能。Accuracy 的计算公式如下：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

式中：TP 为真阳性；TN 为真阴性，FP 为假阳性，FN 为假阴性。

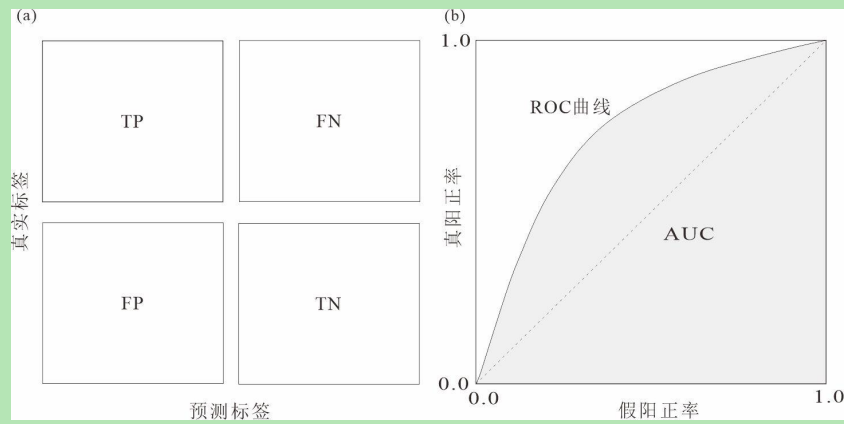
为了可视化预测结果，使用混淆矩阵展现了模型预测结果与实际结果之间的对比情况（图 6a）。召回率是衡量模型在所有实际为正类的样本中能够正确识别的比例，精确率代表对正样本结果的预测准确程度（Hall, 2016），F1 值为精确率和召回率的调和平均值。精确率、召回率和 F1 值的计算公式如下：

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (7)$$

受试者工作特征曲线（Receiver Operating Characteristic Curve, ROC）能客观反映真阳正率（True Positive Rate, TPR）和假阳正率（False Positive Rate, FPR）之间的关系（图 6b）。曲线下的面积 AUC 表示越接近 1 表示分类器越好（Zheng *et al.*, 2022）。

图6 评价指标解释（据 Zheng *et al.*, 2022 修改）

(a) 二分类混淆矩阵; (b) ROC 曲线示意图

Fig.6 Evaluation index interpretation (modified from Zheng *et al.*, 2022)

(a) binary classification confusion matrix; (b) schematic of ROC curve

3.3.2 预测结果分析

在完成数据集预处理后,将未平衡化的数据输入到默认参数下的随机森林模型中进行训练和测试,并依据分类结果构建归一化混淆矩阵。矩阵中的数字表示预测亚相数量占实际亚相数量的比例,主对角线方格对应于4种亚相各自的准确率。尽管模型能识别出大部分样本,但在某些类别上的表现仍存在差异。具体而言,初始模型对开阔台地、局限台地、台缘礁滩和台缘斜坡这4类亚相在测试集中的召回率分别为0.97、0.81、0.95和0.81(图7a)。其中模型对局限台地和台缘斜坡这2种目标亚相的识别效果较差。若直接利用这些数据训练沉积相分类模型,会更倾向于将测试集样本归类为占比更高的类别,虽然能使整体准确率较高,但这种模型并无实际应用价值。

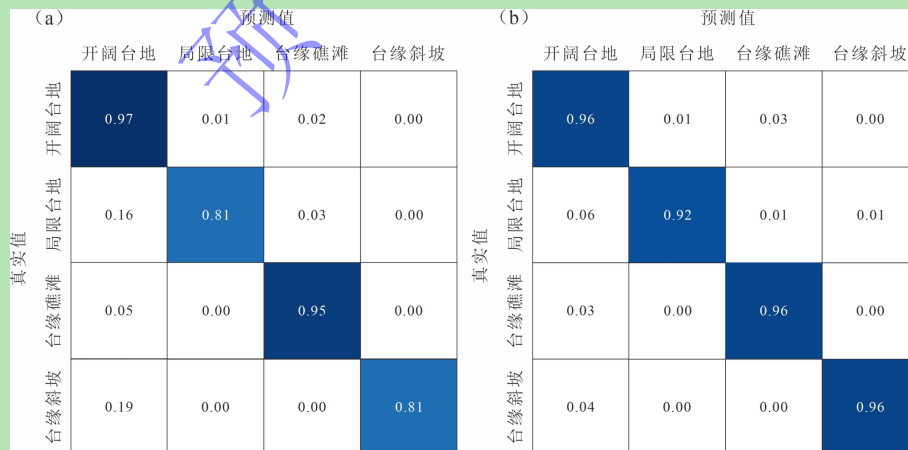


图7 模型混淆矩阵

(a) 初始模型混淆矩阵; (b) 改进模型混淆矩阵

Fig.7 Model confusion matrix

(a) initial model; (b) improved model

对训练数据集进行重采样并完成模型参数调优后,得到最终训练好的模型,并将其应用

于测试集。选用的随机森林分类器精确率、召回率、F1 值的加权平均值以及整体准确率均为 0.96（表 2）。改进后的模型对少数类样本的识别能力有所提升，局限台地、台缘礁滩、台缘斜坡这 3 类亚相的召回率分别提高了 0.11、0.01 和 0.15，整体准确率较高（图 7b）。

借助 ROC 曲线评估模型在整个数据集以及各个亚相上的表现情况（图 8），各亚相经 5 折交叉验证得到的 AUC 值分别为 0.99、1、0.99 和 1，说明改进后的模型在不同的数据子集上表现出了较高的稳定性，能够有效区分正负样本。且模型在整个数据集上的 AUC 值为 0.99，可对模型性能进行全面评估。改进后的模型对未知数据具备更强的泛化能力与更高的准确率。经过平衡化后，样本的合理性和适用性得到了充分验证。

表 2 随机森林分类报告

Table 2 Random Forest classification report

沉积相	精确率	召回率	F1	测试样本数
开阔台地	0.97	0.96	0.97	1 621
局限台地	0.89	0.92	0.90	192
台缘礁滩	0.95	0.96	0.96	947
台缘斜坡	0.87	0.96	0.91	48
加权平均	0.96	0.96	0.96	2 808
准确率		0.96		2 808

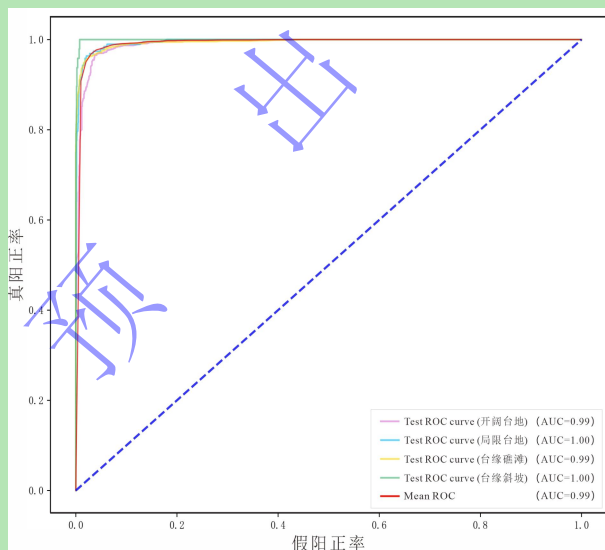


图 8 随机森林模型测试集 ROC 曲线及 AUC 值

Fig.8 ROC curve and AUC value of test set for Random Forest Model

4 讨论

4.1 单井预测验证

为验证随机森林算法在沉积相预测中的可靠性，选取研究区内 Y21 井 6 830~6 865 m 井段对模型进行外部盲测（图 9）。在取心段 6 846.45~6 848.13 m 处为典型台缘斜坡亚相，GR

呈高值，沉积速度较慢且厚度较其他沉积相最薄。运用初始模型时，该段台缘斜坡亚相样本几乎未被正确识别，均被误判为开阔台地亚相，归因于模型对非平衡数据学习不足。经重采样和参数调优后，改进后模型的预测结果与实际沉积亚相的吻合度显著提升，在该段能正确识别出 70% 的台缘斜坡亚相样本。但在 6 850~6 865 m 相变过渡带井段，模型出现识别局限性，其对开阔台地亚相与台缘礁滩亚相的区分性不理想，这与该地区礁滩相的多期次发育特征密切相关，多期生物礁垂向叠置及侧向迁移致使相带边界模糊，这时开阔台地亚相与台缘礁滩亚相的测井响应特征呈现连续性过渡趋势，从而影响了模型的区分能力。整体而言，改进后的模型对取心段的沉积相识别准确率从 83% 提升至 87%，体现了其在碳酸盐岩沉积相识别中兼具效率与可靠性。

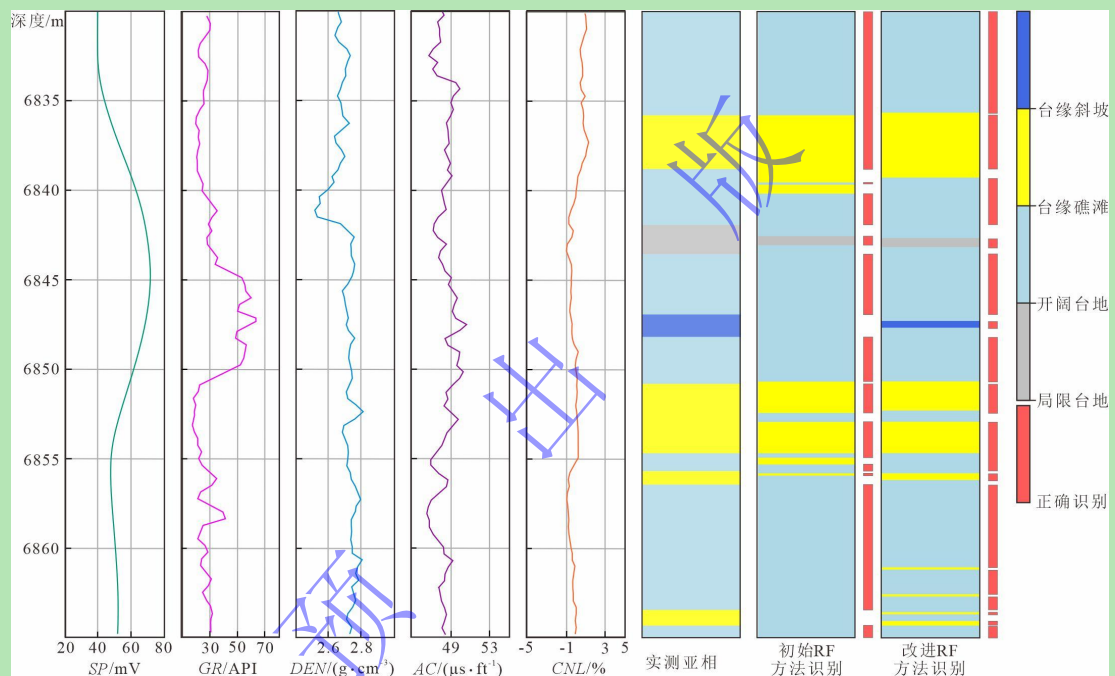


图9 沉积相识别结果

Fig.9 Sedimentary facies prediction results

川东北地区作为典型的碳酸盐台地发育区，其沉积相解释已形成共识（胡忠贵等，2014；邢凤存等，2018；黄仁春等，2019；姚鑫阳等，2024）。从沉积特征来看，二叠纪长兴期开江—梁平海槽两侧发育碳酸盐台地相沉积，呈现出“局限台地—开阔台地—台地边缘礁滩—台缘斜坡”完整有序的相带。本研究基于元坝地区测井数据，采用随机森林算法开展沉积相分类，并结合 SMOTE-NearMiss-1 采样策略优化数据样本分布。模型经参数调优后在目标井段取得 87% 的外部盲测准确率，即模型对盲井的沉积相智能预测结果与外部专家对该井沉积相分析结果对比的吻合率达 87%。单井对应的目标层段优势沉积亚相类型为开阔台地和台缘礁滩，对应于长兴组早期的海侵体系域，此时碳酸盐台地尚处于初步发育阶段，台地边

缘礁滩呈分散分布,规模较小(田成伟等,2012)。随着相对海平面的变化与碳酸盐沉积建造的持续发育,台地边缘逐渐显现,坡折地貌发育,沉积环境由开阔台地向台地边缘生屑滩过渡(程锦翔等,2010)。研究模型的分类结果较好地反映了这一沉积演化趋势,表明该方法在元坝地区碳酸盐岩沉积相分类中的有效性。

4.2 模型的局限性分析

尽管整体识别效果良好,但在相邻相带交汇区段的识别精度仍存在局限。以 Y21 井 6 850~6 865 m 井段为例,模型对开阔台地与台缘礁滩的区分上仍存在一定误差。主要误判发生于该段台地边缘过渡带。该区段沉积环境受多期礁滩垂向叠置与侧向迁移控制,致使相带边界呈现出连续过渡、非均质与非线性的变化趋势,测井响应曲线如 GR 与 SP 在该段表现出趋势连续、界限不清,未形成典型的突变界面,削弱了基于参数统计特征的模型区分能力。此外,训练样本中相变过渡段的代表性样本比例偏低,即便经过 SMOTE 重采样策略进行补强,也难以完整还原复杂沉积过程下真实的地质变异性。尤其在类别边界模糊的情况下,重采样生成的虚拟样本可能强化了既有数据中的模式特征,却无法准确表达边界条件下的过渡性变化,从而限制了模型的泛化能力。另一方面,当前模型主要依赖常规测井参数的统计特征构建特征空间,未能充分挖掘曲线的形态学特征与沉积构造特征信息。在沉积相交界面,曲线往往存在形态复杂、变化剧烈的现象,如短周期尖峰、缓变拐点等,此类高维非线性特征尚未被现有模型体系有效利用,亦可能是识别误差的重要诱因。值得强调的是,尽管本研究在有限数据条件下取得了一定成果,仍应正视机器学习模型在沉积相分类中对训练样本规模与多样性的高度依赖性。沉积相识别任务本质上属于复杂的非线性分类问题,不同构造背景、不同沉积相序与不同测井响应机制之间存在显著差异,这使得小样本训练在某些区段的泛化能力受到显著限制。当前研究以五口井为样本,研究区集中于元坝气田,在空间覆盖范围和沉积类型的多样性方面仍存在不足,这在一定程度上限制了模型结果的区域外适用性。

未来的研究可从两个方向进一步拓展:一是扩大数据集规模,构建涵盖多个构造背景、多个地质时期的碳酸盐岩测井数据体系,以提升模型在跨区应用中的鲁棒性与适应性;二是在特征构建层面,充分挖掘测井曲线的序列特征与沉积过程的演化信息。测井数据在深度维度上具有显著的序列属性,反映了沉积环境的连续变化特征。已有研究表明,引入图卷积神经网络(Graph Convolutional Network, GCN)(Li *et al.*, 2025)、长短时记忆网络(Long Short-Term Memory, LSTM)(Haritha *et al.*, 2025)等序列模型方法,能够从测井数据的变化模式中有效识别岩相与沉积相类型。后续可在此基础上,结合曲线形态识别与沉积构造判别等高阶特征提取手段,探索沉积相之间的深层依赖关系,从而提升模型在复杂相带过渡

区的识别精度与泛化能力。最终，基于多源地质资料的融合应用，构建一个具备跨盆地、跨相区智能识别能力的沉积相分类框架，将为碳酸盐岩储层建模与油气勘探开发提供更高效、可靠的技术支撑。

5 结论

(1) 本文构建的基于混合采样策略的改进随机森林模型(Improved Random Forest, IRF)，通过结合 SMOTE 过采样与 NearMiss-1 欠采样方法，有效缓解了碳酸盐岩沉积相数据集样本不平衡问题，并经网格搜索与交叉验证确定最优参数组合，显著提升了模型的稳定性与泛化能力。

(2) 改进模型在测试集上的结果显示，局限台地、开阔台地、台缘礁滩、台缘斜坡这 4 类沉积亚相的识别准确率均达 90% 以上。在单井盲测验证中，测试准确率为 87%，模型分类结果与长兴组碳酸盐岩台地的沉积演化特征较为一致，表明其对碳酸盐岩沉积相具有较好的识别效果。

(3) 基于决策树的机器学习算法根据特征分裂机制，能从多维特征中提取有效决策规则，实现对测井数据的有效分析和分类。作为基于决策树的集成学习方法，随机森林方法的优势在于对大多数超参数的调优需求较少，但仍能提供出色的模型性能，在测井解释问题中具有一定的优势和潜力。

参考文献 (References)

- 陈安清, 侯明才, 林良彪, 等. 2020. 上扬子地区寒武纪岩相古地理: 对中国小陆块海相盆地演化特点及其控藏效应的启示[J]. 沉积与特提斯地质, 40(3): 38-47. [Chen Anqing, Hou Mingcai, Lin Liangbiao, et al. 2020. Cambrian lithofacies paleogeographic characteristics of the Upper Yangtze Block: Implications for the marine Basin evolution and hydrocarbon accumulation of small-scale tectonic blocks in China[J]. Sedimentary Geology and Tethyan Geology, 40(3): 38-47.]
- 陈洪德, 覃建雄, 田景春, 等. 2004. 中国南方古生界层序格架中的生储盖组合类型及特征[J]. 石油与天然气地质, 25(1): 62-69. [Chen Hongde, Qin Jianxiong, Tian Jingchun, et al. 2004. Types and features of source-reservoir-cap rock combinations in Paleozoic sequence in southern China[J]. Oil & Gas Geology, 25(1): 62-69.]
- 程锦翔, 谭钦银, 郭彤楼, 等. 2010. 川东北元坝地区长兴组—飞仙关组碳酸盐台地边缘沉积特征及演化[J]. 沉积与特提斯地质, 30(4): 29-37. [Cheng Jinxiang, Tan Qinyin, Guo Tonglou, et al. 2010. Sedimentary characteristics and evolution of the carbonate platform-margins in the Changxing Formation-Feixianguan Formation in Yuanba, northeastern Sichuan[J]. Sedimentary Geology and Tethyan Geology, 30(4): 29-37.]
- 程希, 任战利. 2024. 人工智能测井: 基础、原理、技术及应用[J]. 煤田地质与勘探, 52(8): 145-164. [Cheng Xi, Ren Zhanli. 2024. Artificial intelligence logging: Fundamental, principle, technique, and application[J]. Coal Geology & Exploration, 52(8): 145-164.]
- 崔俊峰, 杨金路, 王民, 等. 2023. 基于随机森林算法的泥页岩孔隙度预测[J]. 油气地质与采收率, 30(6): 13-21. [Cui Junfeng, Yang Jinlu, Wang Min, et al. 2023. Shale porosity prediction based on Random Forest Algorithm[J]. Petroleum Geology and Recovery Efficiency, 30(6): 13-21.]

- 郭彤楼. 2019. 元坝气田成藏条件及勘探开发关键技术[J]. 石油学报, 40(6): 748-760. [Guo Tonglou. 2019. Gas accumulation conditions and key exploration & development technologies in Yuanba gas field[J]. Acta Petrolei Sinica, 40(6): 748-760.]
- 何登发, 李英强, 黄涵宇, 等. 2020. 四川多旋回叠合盆地的形成演化与油气聚集[M]. 北京: 科学出版社, 1-569. [He Dengfa, Li Yingqiang, Huang Hanyu, et al. 2020. Formation and Evolution of Polycyclic Superimposed Basins in Sichuan and Their Hydrocarbon Accumulation[M]. Beijing: Science Press, 1-569.]
- 胡忠贵, 胡明毅, 廖军, 等. 2014. 鄂西建南地区长兴组沉积相及生物礁沉积演化模式[J]. 天然气地球科学, 25(7): 980-990. [Hu Zhonggui, Hu Mingyi, Liao Jun, et al. 2014. Sedimentary facies and evolution model of reef deposition in Jiannan area of western Hubei[J]. Natural Gas Geoscience, 25(7): 980-990.]
- 黄仁春, 邢凤存, 范小军, 等. 2019. 四川盆地元坝地区长兴组—飞仙关组高精度层序地层格架的建立及礁滩储集层预测[J]. 古地纪, 21(2): 369-378. [Huang Renchun, Xing Fengcun, Fan Xiaojun, et al. 2019. Establishment of high-precision sequence stratigraphic framework of the Changxing-Feixianguan Formations and prediction of reef-shoal reservoirs in Yuanba area, Sichuan Basin[J]. Journal of Palaeogeography (Chinese Edition), 21(2): 369-378.]
- 李昌, 沈安江, 常少英, 等. 2021. 机器学习法在碳酸盐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, 11(4): 586-596. [Li Chang, Shen Anjiang, Chang Shaoying, et al. 2021. 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, 11(4): 586-596.]
- 李昌, 王鑫, 冯周, 等. 2023. 基于 KNN 分类算法的微生物白云岩岩相测井综合识别: 以四川盆地 GM 地区灯四段为例[J]. 海相油气地质, 28(4): 433-440. [Li Chang, Wang Xin, Feng Zhou, et al. 2023. Comprehensive logging identification of microbial carbonate lithofacies based on KNN classification algorithm: A case study of Dengying Formation in GM area, Sichuan Basin[J]. Marine Origin Petroleum Geology, 28(4): 433-440.]
- 李昌峰, 侯明才, 张小青, 等. 2015. 川东北元坝地区上二叠统长兴组优质储层形成机制[J]. 地质力学学报, 21(2): 278-289. [Li Changfeng, Hou Mingcai, Zhang Xiaoqing, et al. 2015. Formation mechanism of high-quality reservoirs in the Upper Permian Changxing Formation in Yuanba gas field, northeastern Sichuan Basin[J]. Journal of Geomechanics, 21(2): 278-289.]
- 李雯, 杨帅, 陈安清, 等. 2023. 川西北深层中二叠统茅口组岩相古地理及勘探意义[J]. 地球科学, 48(2): 609-620. [Li Wen, Yang Shuai, Chen Anqing, et al. 2023. The lithofacies paleogeographic of the Middle Permian deep Maokou Formation and its significance for hydrocarbon exploration, northwestern Sichuan Basin[J]. Earth Science, 48(2): 609-620.]
- 李雯, 陈安清, 张玺华, 等. 2025. 四川盆地二叠纪两类型碳酸盐台地边缘及古地理分异格局[J/OL]. 地质学报. <https://doi.org/10.19762/j.cnki.dizhixuebao.2024441>. [Li Wen, Chen Anqing, Zhang Xihua, et al. 2025. Permian paleogeographic differentiation evolution and platform margin model comparison in Sichuan Basin[J]. Acta Geologica Sinica. <https://doi.org/10.19762/j.cnki.dizhixuebao.2024441>.]
- 马永生, 蔡勋育, 赵培荣. 2014. 元坝气田长兴组—飞仙关组礁滩相储层特征和形成机理[J]. 石油学报, 35(6): 1001-1011. [Ma Yongsheng, Cai Xunyu, Zhao Peirong. 2014. Characteristics and Formation mechanisms of reef-shoal carbonate reservoirs of Changxing-Feixianguan Formations, Yuanba gas field[J]. Acta Petrolei Sinica, 35(6): 1001-1011.]
- 孟宪武, 陈安清, 李雯, 等. 2025. 四川盆地南缘二叠系茅口组构造—沉积分异及控储效应[J]. 成都理工大学学报(自然科学版), 52(1): 94-107. [Meng Xianwu, Chen Anqing, Li Wen, et al. 2025. Platform margin tectono-sedimentary differentiation and reservoir control effects of the Permian Maokou Formation, southern margin of Sichuan Basin[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 52(1): 94-107.]
- 倪新锋, 陈洪德, 田景春, 等. 2007. 川东北地区长兴组—飞仙关组沉积格局及成藏控制意义[J]. 石油与天然气地质, 28(4): 458-465. [Ni Xinfeng, Chen Hongde, Tian Jingchun, et al. 2007. Sedimentary framework of Changxing-Feixianguan Formations and its control on reservoiring in northeastern Sichuan Basin[J]. Oil & Gas Geology, 28(4): 458-465.]
- 司马立强, 陈志强, 王亮, 等. 2017. 基于滩控岩溶型白云岩储层分类的渗透率建模方法研究: 以川中磨溪—高石梯地区龙王庙组为例[J]. 岩性油气藏, 29(3): 92-102. [Sima Liqiang, Chen Zhiqiang, Wang Liang, et al. 2017. Permeability modeling based on the classification of beach-controlled karst dolomite reservoirs: A case from Longwangmiao Formation in Moxi-Gaoshiti area, central Sichuan Basin[J]. Lithologic Reservoirs, 29(3): 92-102.]

- 孙予舒, 黄芸, 梁婷, 等. 2020. 基于 XGBoost 算法的复杂碳酸盐岩岩性测井识别[J]. 岩性油气藏, 32(4): 98-106. [Sun Yushu, Huang Yun, Liang Ting, et al. 2020. Identification of complex carbonate lithology by logging based on XGBoost algorithm[J]. Lithologic Reservoirs, 32(4): 98-106.]
- 田成伟, 安显银, 罗清园, 等. 2012. 川东北元坝地区长兴组层序地层及沉积相分析[J]. 地球科学与环境学报, 34(4): 36-42. [Tian Chengwei, An Xianyin, Luo Qingyuan, et al. 2012. Analysis of sequence stratigraphy and sedimentary facies of Changxing Formation in Yuanba area of northeast Sichuan[J]. Journal of Earth Sciences and Environment, 34(4): 36-42.]
- 王民, 杨金路, 王鑫, 等. 2023. 基于随机森林算法的泥页岩岩相测井识别[J]. 地球科学, 48(1): 130-142. [Wang Min, Yang Jinlu, Wang Xin, et al. 2023. Identification of shale lithofacies by well logs based on Random Forest Algorithm[J]. Earth Science, 48(1): 130-142.]
- 王一刚, 文应初, 张帆, 等. 1998. 川东地区上二叠统长兴组生物礁分布规律[J]. 天然气工业, 18(6): 10-15. [Wang Yigang, Wen Yingchu, Zhang Fan, et al. 1998. Distribution law of the organic reefs in Changxing Formation of Upper Permian in East Sichuan[J]. Natural Gas Industry, 18(6): 10-15.]
- 王正和, 邓剑, 谭钦银, 等. 2012. 元坝地区长兴组典型沉积相及各相带物性特征[J]. 矿物岩石, 32(2): 86-96. [Wang Zhenghe, Deng Jian, Tan Qinyin, et al. 2012. Characteristics of typical sedimentary facies and theirs petrophysics of Changxing Formation in Yuanba district[J]. Journal of Mineralogy and Petrology, 32(2): 86-96.]
- 武赛军, 魏国齐, 杨威, 等. 2019. 开江—梁平海槽东侧长兴组台缘生物礁发育特征及油气地质勘探意义[J]. 中国石油勘探, 24(4): 457-465. [Wu Saijun, Wei Guoqi, Yang Wei, et al. 2019. Development characteristics of reefs on the platform margin of Changxing Formation in eastern Kaijiang-Liangping ocean trough and its significance for petroleum geological exploration[J]. China Petroleum Exploration, 24(4): 457-465.]
- 邢凤存, 胡华蕊, 侯明才, 等. 2018. 构造和古地理控制下的碳酸盐岩储集体旋回和集群性探讨: 以四川盆地为例[J]. 地球科学, 43(10): 3540-3552. [Xing Fengcun, Hu Huarui, Hou Mingcai, et al. 2018. Carbonate reservoirs cycles and assemblages under the tectonic and palaeogeography control: A case study from Sichuan Basin[J]. Earth Science, 43(10): 3540-3552.]
- 杨帅, 陈安清, 张玺华, 等. 2021. 四川盆地二叠纪栖霞—茅口期古地理格局转换及勘探启示[J]. 沉积学报, 39(6): 1466-1477. [Yang Shuai, Chen Anqing, Zhang Xihua, et al. 2021. Paleogeographic transition of the Permian Chihhsia-Maokou Period in the Sichuan Basin and indications for oil-gas exploration[J]. Acta Sedimentologica Sinica, 39(6): 1466-1477.]
- 杨雨, 钟原, 李林娟. 2021. 开江—梁平海槽周缘二叠系—三叠系礁滩组合分布与成因[J]. 成都理工大学学报(自然科学版), 48(6): 683-690. [Yang Yu, Zhong Yuan, Li Linjuan. 2021. Distribution and genesis of Permian-Triassic reef-shoal combination around Kaijiang-Liangping trough in Sichuan Basin, China[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 48(6): 683-690.]
- 姚鑫阳, 胡忠贵, 武赛军, 等. 2024. 川东北地区开江—梁平海槽东侧长兴组沉积特征及有利区预测[J]. 沉积学报, 42(6): 2102-2118. [Yao Xinyang, Hu Zhonggui, Wu Saijun, et al. 2024. Sedimentary characteristics and prediction of favorable areas of the changxing Formation on the eastern side of the Kaijiang-Liangping trough in northeastern Sichuan Basin[J]. Acta Sedimentologica Sinica, 42(6): 2102-2118.]
- 曾丽丽, 孟凡月, 汤华贝, 等. 2022. 基于注意力机制的碳酸盐岩储层岩相识别方法[J]. 测井技术, 46(3): 294-303. [Zeng Lili, Meng Fanyue, Tang Huabei, et al. 2022. Lithofacies identification method of carbonate reservoir based on attention mechanism[J]. Well Logging Technology, 46(3): 294-303.]
- 张凯, 张科. 2022. 基于 LightGBM 算法的边坡稳定性预测研究[J]. 中国安全科学学报, 32(7): 113-120. [Zhang Kai, Zhang Ke. 2022. Prediction study on slope stability based on LightGBM algorithm[J]. China Safety Science Journal, 32(7): 113-120.]
- 张庆龙, 毛元元, 冯建松, 等. 2025. 裂缝—孔隙型碳酸盐岩油藏加密井井位部署新方法: 以渤海湾盆地黄骅坳陷唐海油田寒武系油藏为例[J]. 岩性油气藏, 37(3): 165-175. [Zhang Qinglong, Mao Yuanyuan, Feng Jiansong, et al. 2025. Quantitative selection and application of well locations for encrypted wells in fractured-porosity carbonate rocks: A case study of Cambrian oil reservoirs in Tanghai oilfield, Huanghua Depression, Bohai Bay Basin[J]. Lithologic Reservoirs, 37(3): 165-175.]

- 周进高, 徐春春, 姚根顺, 等. 2015. 四川盆地寒武统龙王庙组储集层形成与演化[J]. 石油勘探与开发, 42(2): 158-166. [Zhou Jingao, Xu Chunchun, Yao Genshun, et al. 2015. Genesis and evolution of Lower Cambrian longwangmiao Formation reservoirs, Sichuan Basin, SW China[J]. Petroleum Exploration and Development, 42(2): 158-166.]
- Al-Mudhafar W J. 2017. Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms[J]. Journal of Petroleum Exploration and Production Technology, 7(4): 1023-1033.
- Bhattacharya S, Carr T R, Pal M. 2016. Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA[J]. Journal of Natural Gas Science and Engineering, 33: 1119-1133.
- Biau G, Scornet E. 2016. A Random Forest guided tour[J]. Test, 25(2): 197-227.
- Bond C E, Lunn R J, Shipton Z K, et al. 2012. What makes an expert effective at interpreting seismic images?[J]. Geology, 40(1): 75-78.
- Breiman L. 1996. Bagging predictors[J]. Machine Learning, 24(2): 123-140.
- Hall B. 2016. Facies classification using machine learning[J]. The Leading Edge, 35(10): 906-909.
- Haritha D, Satyavani N, Ramesh A. 2025. Generation of missing well log data with deep learning: CNN-Bi-LSTM approach[J]. Journal of Applied Geophysics, 233: 105628.
- Hou M Q, Xiao Y X, Lei Z D, et al. 2023. Machine learning algorithms for lithofacies classification of the gulong shale from the Songliao Basin, China[J]. Energies, 16(6): 2581.
- Hou X M, Lian P Q, Zhao J Y, et al. 2024. Identification of carbonate sedimentary facies from well logs with machine learning[J]. Petroleum Research, 9(2): 165-175.
- Jordan M I, Mitchell T M. 2015. Machine learning: Trends, perspectives, and prospects[J]. Science, 349(6245): 255-260.
- Li K R, Song J M, Wang H, et al. 2025. Spectral graph convolution networks for microbialite lithology identification based on conventional well logs[J]. Petroleum Science, 22(4): 1513-1533.
- Liu X Y, Zhou L, Chen X H, et al. 2020. Lithofacies identification using support vector machine based on local deep multi-kernel learning[J]. Petroleum Science, 17(4): 954-966.
- Macrae E J, Bond C E, Shipton Z K, et al. 2016. Increasing the quality of seismic interpretation[J]. Interpretation, 4(3): T395-T402.
- Qi M, Han C C, Ma C F, et al. 2022. Identification of diagenetic facies logging of tight oil reservoirs based on deep learning: A case study in the Permian lucaogou Formation of the Jimsar Sag, Junggar Basin[J]. Minerals, 12(7): 913.
- Radwan A E. 2021. Modeling the depositional environment of the sandstone reservoir in the Middle Miocene Sidri member, Badri Field, gulf of Suez Basin, Egypt: Integration of Gamma-Ray Log patterns and petrographic characteristics of lithology[J]. Natural Resources Research, 30(1): 431-449.
- Rider M H. 1990. Gamma-ray log shape used as a facies indicator: Critical analysis of an oversimplified methodology[J]. Geological Society, London, Special Publications, 48(1): 27-37.
- Salman H A, Kalakech A, Steiti A. 2024. Random Forest Algorithm overview[J]. Babylonian Journal of Machine Learning, 2024: 69-79.
- Sun S, Chen A Q, Hou M C, et al. 2025. A decrease in low-latitude weathering flux facilitated the demise of the Late Paleozoic Ice Age[J]. Global and Planetary Change, 247: 104730.
- Verma V K, Saxena K, Banodha U. 2024. Analysis effect of K values used in K fold cross validation for enhancing performance of machine learning model with decision tree[C]//Proceedings of the 13th international conference on advanced computing. Kolhapur: Springer, 374-396.
- Wang P, Chen X H, Wang B F, et al. 2020. An improved method for lithology identification based on a hidden Markov model and Random Forests[J]. Geophysics, 85(6): IM27-IM36.
- Ying X. 2019. An overview of overfitting and its solutions[J]. Journal of Physics: Conference Series, 1168(2): 022022.
- Zhang J P, Mani I. 2003. KNN approach to unbalanced data distributions: A case study involving information extraction[C]//Proceeding of international conference on machine learning. Washington: ICML.
- Zheng D Y, Hou M C, Chen A Q, et al. 2022. Application of machine learning in the identification of fluvial-lacustrine lithofacies from well logs: A case study from Sichuan Basin, China[J]. Journal of Petroleum Science and Engineering, 215: 110610.

Intelligent Identification for Carbonate Logging Sedimentary Facies Based on Machine Learning: An example from the Late Permian Changxing Formation, Northeastern Sichuan Basin

FAN Xuan¹, CHEN AnQing¹, ZHENG DongYu¹, DENG Mo², LI ZhongChao³, ZHOU Kai³, WANG Han¹, ZHONG HanTing¹, SUN Shi¹, LU JianLin², REN Qiang¹, HOU MingCai¹

1. Key Laboratory of Deep-Time Geography and Environment Reconstruction and Applications of Ministry of Natural Resources, Chengdu University of Technology, Chengdu 610059, China

2. Wuxi Research Institute of Petroleum Geology, Petroleum Exploration and Production Research Institute, SINOPEC, Wuxi, Jiangsu 214126, China

3. Institute of Exploration and Development, Zhongyuan Oilfield, SINOPEC, Puyang, Henan 457001, China

Abstract: [Objective] Accurate identification of carbonate rock sedimentary facies is basic to conduct paleogeographic reconstruction. However, the inherently ambiguous facies characteristics coupled with the subjectivity of manual interpretation pose significant challenges in facies identification. Traditional well-log facies interpretation methods are further constrained by low efficiency and experience-dependent limitations. To address these challenges, in this study a reusable and intelligent interpretation model was developed based on machine learning algorithms with the aim of improving the precision and efficiency of sedimentary facies identification in carbonate formations. [Methods] Focusing on the reef-shoal facies of the late Permian Changxing Formation in the Yuanba area of the northeastern Sichuan Basin, well-log data from five representative boreholes was analyzed. Various well-log curves (e.g., acoustic travel time (AC) and natural gamma ray (GR)) were selected as feature parameters to construct an Improved Random Forest (IRF)-based facies recognition model. A hybrid sampling strategy combining the Synthetic Minority Oversampling Technique (SMOTE) and the NearMiss-1 algorithm was implemented to address the problem of classification imbalance, while grid search combined with K-fold cross-validation ensured robust hyperparameter tuning. [Results] Blind test results for the target well interval using the proposed improved model showed an accuracy of 87% compared to sedimentary facies interpretations by external experts, indicating a significant improvement over the initial model. The proposed model effectively identifies favorable subfacies such as open platform and platform-margin reef-shoal, accurately reflecting regional sedimentary evolution characteristics. [Conclusions] Machine-learning-based approaches for sedimentary facies recognition not only enhance identification accuracy but also demonstrate robustness and applicability in practical geological interpretation.

Key words: carbonate rocks; Random Forest Algorithm; intelligent analysis model of sedimentary facies; Changxing Formation; Sichuan Basin