

融合传统机器学习与深度学习的页岩气储集层岩相预测新方法：以四川盆地下寒武统筇竹寺组为例^{*}

廖 燕^{1,2} 闫建平^{1,2,3} 廖茂杰⁴ 邱小雪⁴
杨 杨⁴ 郑马嘉⁵ 闫 华⁶ 黄 毅⁷

- 1 西南石油大学地球科学与技术学院, 四川成都 610500
2 天然气地质四川省重点实验室(西南石油大学), 四川成都 610500
3 油气藏地质及开发工程全国重点实验室(西南石油大学), 四川成都 610500
4 中国石油西南油气田公司页岩气研究院, 四川成都 610051
5 中国石油西南油气田公司, 四川成都 610051
6 中国石油化工股份有限公司上海海洋油气分公司勘探开发研究院, 上海 200120
7 中国石油集团测井有限公司西南分公司, 重庆 400021

摘 要 随着页岩气勘探开发的不断深入, 精准、高效地划分与预测岩相成为评价页岩气储集层品质的关键。传统的岩相测井识别方法存在主观性强、难以应对大规模数据集以及难以有效捕捉数据中复杂的非线性关系等局限性。本研究以四川盆地下寒武统筇竹寺组页岩气储集层为例, 充分利用岩心测试分析的精确数据刻度测井曲线建立的岩相标签数据集, 提出了一种融合传统机器学习模型: 决策树、支持向量机(SVM)、K 近邻(KNN)以及深度学习模型: 一维卷积神经网络(1D-CNN)的集成学习(Voting)模型进行岩相分类预测, 且对比传统机器学习/深度学习模型与集成学习模型的预测效果, 评估其在复杂页岩岩相预测中的应用潜力。研究结果表明: 集成学习(Voting)模型在岩相预测精度上表现最佳, 预测精度达到 94.0%, 显著优于单一机器学习模型(决策树、KNN、SVM 及 1D-CNN)。形成的集成学习方法, 通过融合 4 种机器学习模型各自的优势, 有效克服了传统分类算法受岩相样本不均衡的影响引起的识别偏差, 提升了稀缺样本的岩相识别能力, 为四川盆地高—过成熟筇竹寺组深层页岩气储集层测井岩相剖面的高精度恢复及甜点预测提供了技术支撑。

关键词 筇竹寺组 岩相预测 机器学习 深度学习 集成学习

第一作者简介 廖燕, 女, 2001 年生, 西南石油大学硕士研究生, 研究方向为测井地质学、测井人工智能处理与解释。E-mail: 18283426959@163.com。

通讯作者简介 闫建平, 男, 1980 年生, 理学博士, 教授、博士生导师, 主要从事测井地质学、岩石物理及非常规油气测井评价方面的教学与研究工作。E-mail: yanjp_tj@163.com。

中图分类号: P618.130.2 文献标志码: A

A novel method for lithofacies prediction in shale gas reservoirs that integrates traditional machine learning and deep learning techniques:

^{*} 国家自然科学基金项目(编号: 42372177), 中国石油—西南石油大学创新联合体科技合作项目(编号: 2020CX020000), 四川省自然科学基金项目(编号: 2022NSFSC0287), 中石油科技部“十四五”重大专项(编号: 2021DJ1901)联合资助。[Co-funded by the National Natural Science Foundation of China (No. 42372177), the Science and Technology Cooperation Project of China Petroleum-Southwest Petroleum University Innovation Consortium (No. 2020CX020000), the Natural Science Foundation of Sichuan Province (No. 2022NSFSC0287), the 14th Five-Year Major Project of the Ministry of Science and Technology of China Petroleum (No. 2021DJ1901)]

收稿日期: 2024-12-31 改回日期: 2025-03-24

a case study of the Lower Cambrian Qiongzhusi Formation in Sichuan Basin

LIAO Yan^{1,2} YAN Jianping^{1,2,3} LIAO Maojie⁴ QIU Xiaoxue⁴

YANG Yang⁴ ZHENG Majia⁵ YAN Hua⁶ HUANG Yi⁷

1 School of Geoscience and Technology, Southwest Petroleum University, Chengdu 610500, China

2 Natural Gas Geology Key Laboratory of Sichuan Province, Southwest Petroleum University, Chengdu 610500, China

3 National Key Laboratory of Oil & Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu 610500, China

4 Research Institute of Shale gas, PetroChina Southwest Oil & Gasfield Company, Chengdu 610051, China

5 PetroChina Southwest Oil and Gas Field Company, Chengdu 610051, China

6 Institute of Exploration and Development, SINOPEC Shanghai Offshore Oil & Gas Company, Shanghai 200120, China

7 Southwest Branch, CNPC Logging Company Limited, Chongqing 400021, China

Abstract As the exploration and development of shale gas continue to deepen, the accurate and efficient division and prediction of lithofacies have become key factors in evaluating the quality of shale gas reservoirs. Traditional lithofacies logging identification methods are limited by their strong subjectivity, inability to handle large-scale datasets, and difficulty in effectively capturing the complex nonlinear relationships in the data. This study takes the Lower Cambrian Qiongzhusi Formation shale gas reservoir in the Sichuan Basin as an example. It utilizes precise data from rock core testing and logging curves to establish a lithofacies-labeled dataset. A novel ensemble learning (Voting) model is proposed, integrating traditional machine learning models (Decision Tree, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)) and a deep learning model (1D Convolutional Neural Network, 1D-CNN) for lithofacies classification and prediction. The study also compares the predictive performance of traditional machine learning and deep learning models with that of the ensemble learning model, assessing its application potential in complex shale lithofacies prediction. The results indicate that the ensemble learning (Voting) model achieves the highest lithofacies prediction accuracy of 94.0%, significantly outperforming individual base learners (Decision Tree, KNN, SVM, and 1D-CNN). The proposed ensemble learning method effectively overcomes the recognition bias caused by sample imbalance in traditional classification algorithms by leveraging the strengths of the four base learner models. This improvement enhances the ability to identify lithofacies with limited samples, providing technical support for the high-precision reconstruction of logging-based lithofacies profiles and sweet spot prediction in the deeply buried, over-mature Qiongzhusi Formation shale gas reservoirs in the Sichuan Basin.

Key words Qiongzhusi Formation, lithofacies prediction, machine learning, deep learning, ensemble learning

About the first author LIAO Yan, born in 2001, is a master's student in Southwest Petroleum University. The research direction is logging geology, logging artificial intelligence processing and interpretation. E-mail: 18283426959@163.com.

About the corresponding author YAN Jianping, born in 1980, doctor of science, professor, doctoral supervisor, mainly engaged in the teaching and research work of logging geology, rock physics and unconventional oil and gas logging evaluation. E-mail: yanjp_tj@163.com.

0 引言

四川盆地作为中国最早开展页岩气商业性勘探开发的含油气盆地（邹才能等，2023），依托其丰富的地质资源和日益成熟的勘探技术，成为国内页岩气开发的重要基地之一。下寒武统筇竹寺组页岩与上奥陶统五峰组—下志留统龙马溪组页岩同属海相沉积，具有厚度大、分布稳定、有机质含量高、成熟度高等特点（董大忠等，2016；张少龙等，2022；赵圣贤等，2022；闫建平等，2022）。尽管筇竹寺组页岩面临孔隙度低等不利因素的挑战，但通过五峰组—龙马溪组开发经验的借鉴，2023 年 4 月中石油在威远背斜东北翼的裂陷槽内部署的资 201 井取得重大突破（姜鹏飞等，2023），该井测试产气量达 $73.88 \times 10^4 \text{ m}^3/\text{d}$ （何骁等，2024），这一产量的实现与突破预示着资源潜力巨大的筇竹寺组页岩具有获得高产页岩气流及规模上产的条件。

页岩岩相划分对于深化资源评价开发具有重要的意义。不同岩相的开发潜力不同，通过精细化的岩相划分，可以更准确地理解页岩的沉积环境、岩性特征变化和有机质分布，为页岩气的富集规律、储集层特征以及开发潜力的评估提供科学依据（聂海宽等，2024）。因此，识别有利的页岩岩相段是页岩系统内寻找目的靶体层的关键所在。目前国内外专家学者提出了多种基于测井资料的划分与预测方案，能够有效地识别出页岩岩相的类型（Hall，2016；彭丽等，2019；王岚等，2023），但是也存在较大的局限性，基于岩心分析的页岩岩相划分存在取心成本高、受到个人经验和知识局限的影响，难以高效应对数据量大、变化复杂的情况，缺乏自动化和客观性（Nishitsuji and Exley，2019）。为了解决这些问题，机器学习作为一种高效的数据分析方法，逐渐在岩相分类中得到应用（李昌等，2023；王民等，2023；闫佳飞等，2025）。机器学习算法能够自动从大量测井数据中提取潜在的规律，减少人为干预，提高分类的客观性和一致性。然而单一的机器学习方法在岩相预测中可能面临过拟合、泛化能力差、对异常值敏感等问题。相较之下，集成学习方法通过融合多个基学习器的预测结果，能够有效减小单一模型的偏差和方差，提升预测的稳定性和准确性（湛丽等，2023），适用于在复杂地质背景下的岩相分类中处理具有高度不确定性和复杂性的数据集。

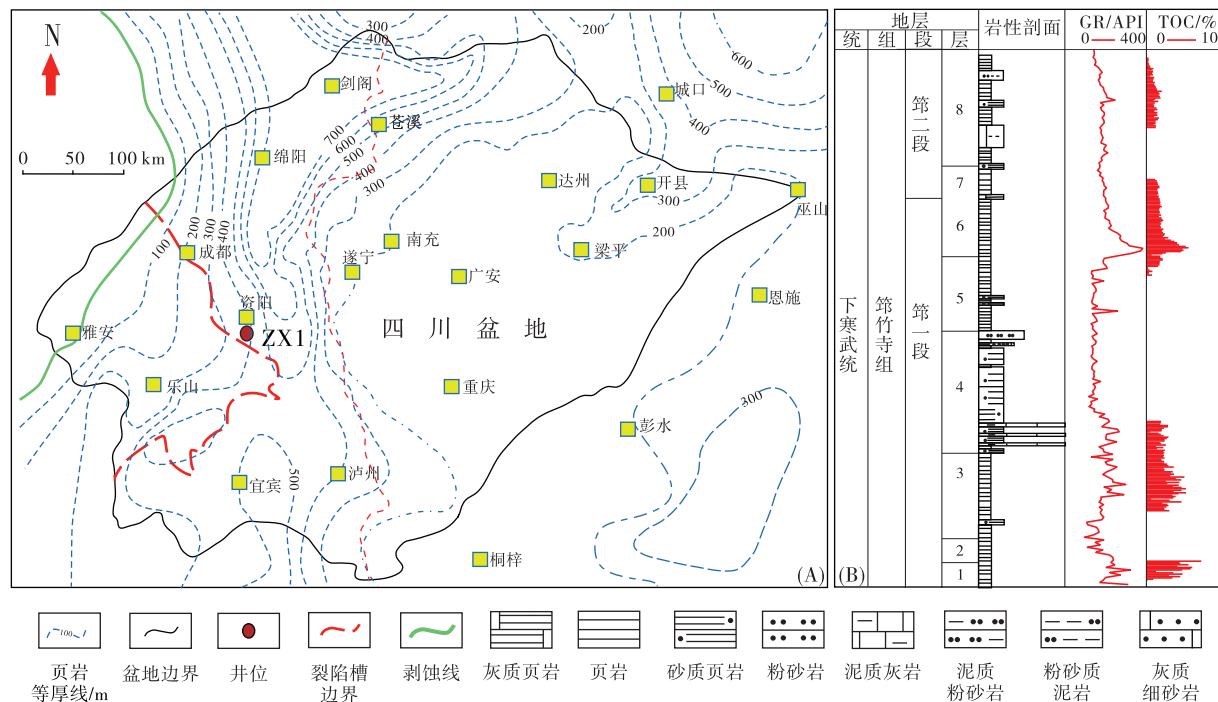
本研究基于四川盆地威远—资阳地区下寒武统筇竹寺组页岩地层岩心 XRD、有机碳及测井资料，采用了一种综合考虑“总有机碳含量（TOC）、沉积结构、岩性及脆性”信息的页岩岩相划分方法，建立了页岩岩相划分的标准。通过测井参数相关性分析，最终优选 6 种测井参数，即：自然伽马（GR）、无铀伽马（KTH）、声波时差（AC）、补偿中子（CNL）、密度（DEN）和电阻率（RT）作为模型特征输入，页岩岩相作为输出标签，构建了融合传统机器学习模型：决策树、支持向量机（SVM）、K 近邻（KNN）以及深度学习模型：一维卷积神经网络（1D-CNN）的集成学习（Voting）模型。通过参数调优，将集成学习模型应用于 ZX1 井页岩岩相预测，预测准确率达 94.0%，且召回率等各项评估指标均显著提升，克服了单一机器学习算法在页岩岩相识别时受各类别样本数量不均衡的影响引起的识别偏差，为四川盆地高—过成熟筇竹寺组页岩气储集层的资源评价与开发规划提供了支持和依据。

1 地质概况

四川盆地及其周缘在震旦纪至早寒武纪期间整体处于区域拉张构造环境。晚震旦世末期，受桐湾运动的挤压作用，四川盆地整体发生抬升，形成了区域性侵蚀谷地（武赛军等，2016）。随后，兴凯运动导致构造反转，使侵蚀谷进一步拉张沉降，最终形成了南北向的拉张槽，盆地内部则发育了德阳—安岳裂陷槽构造（邱玉超等，2019）。这些构造演化过程为筇竹寺组黑色页岩的沉积创造了良好的地质条件。

研究区位于四川盆地中南部的威远—资阳地区（图 1-A），构造上属于德阳—安岳裂陷槽，涵盖了川中平缓构造带、川西南低褶皱构造带北部以及川南低陡构造带北部。区域内筇竹寺组地层发育完整，以黑色页岩为主（梁峰等，2024），其沉积形成受控于拉张背景下的贫氧—缺氧的水体环境（赵建华等，2019），埋藏深度大，具有典型的高成熟—过成熟页岩气储集层特征。

筇竹寺组地层沉积受浅海—河口环境、潮汐作用和水深变化控制，岩性由浅海砂岩过渡至深海泥岩和灰岩，涵盖砂岩、泥岩、灰岩及粉砂岩等多种类型（Xie *et al.*，2024）（图 1-B）。岩相的变化直



A—ZX1 井位置; B—ZX1 井下寒武统综合柱状图 (据何骁等, 2024, 修改)

图 1 四川盆地 ZX1 井位置及地层综合柱状图

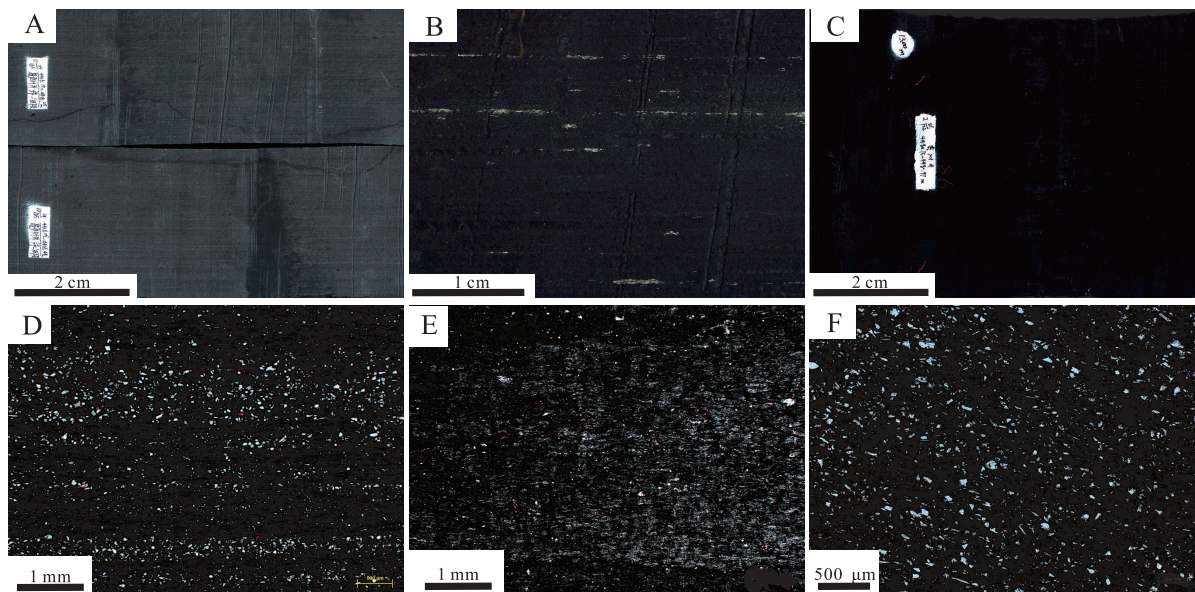
Fig. 1 Location and stratigraphic comprehensive column of Well ZX1 of Sichuan Basin

接影响页岩气储集层的物性,如孔隙度、渗透率等关键参数,这些物性特征决定了储集层的储集能力和流动性。沉积环境变化导致的岩性与岩相的频繁交替,尤其在浅海与深海的过渡区,使得岩性呈现显著的纵向差异,增加了岩相划分与预测的难度。

2 岩相分类及其划分方案

页岩岩相分类有助于识别具备勘探与开发潜力的页岩气有利层系（王心乾等，2024）。多角度的岩相分类方法能够更有效地指导勘探和开发方案。若是过于简化的方案可能无法提供足够的细节，难以有效支持沉积环境解释、甜点目标的选择。相反，过于复杂的方案则可能增加操作的难度，阻碍数据分析的效率。在岩相分类研究领域，有从基于页岩储集层品质及压裂品质 2 个要素，提出以 TOC、矿物组成为基础的“两步法”页岩岩相划分方案（蒋裕强等，2016），也有考虑从沉积、成岩作用影响岩相的要素的角度建立页岩岩相划分方案（Yan *et al.*, 2018；黄莉莎等，2022），结合前期这些关于页岩岩相分类与识别的研究进展，以及四川盆地威远—资阳地区下寒武统筇竹寺组页岩地层特

征与岩心测试、测井资料,本研究从页岩的成因关键要素和测井识别角度考虑,采用了基于“总有机碳含量(TOC)+沉积构造+岩性及脆性”信息的综合页岩岩相划分方案,对筇竹寺组页岩进行岩相类型划分。TOC含量反映了页岩中有机质的丰度,高TOC含量意味着潜在的生气潜力。通常含气量与TOC含量呈正相关,页岩气产区工业含气量下限一般为 $2.0\text{ m}^3/\text{t}$,对应TOC含量下限约为2.0%(肖贤明等,2015)。因此,以TOC值2.0%为界限,当TOC含量低于2.0%时,视为贫有机质;当TOC含量介于2.0%~3.0%之间,视为中有机介质;当TOC含量大于3.0%时,视为富有机质。页岩沉积构造可以有效地进行从微观到宏观的尺度表征,这对于理解岩性结构如何影响页岩气的富集、储集层异质性、断裂发生和力学性能至关重要(万云强等,2024)。筇竹寺组页岩沉积构造也可分为3种类型:纹层状(单层厚度小于20 mm)、层状(单层厚度大于20 mm)和块状(缺乏明显的定向组构)(图2)。另外,研究表明,筇竹寺组页岩中,长石矿物的含量较高,其次为黏土矿物,而碳酸盐矿物的含量最低(薛晓辉等,2016;雍锐等,2024)。为突出矿物组成特征



A—WY1H 井, 4416.33 m, 纹层状结构, 岩心; B—ZX1 井, 4604.01 m, 层状结构, 岩心; C—ZX1 井, 4480.40 m, 块状结构, 岩心;
D—纹层状结构, ZX1 井, 4852.58 m; E—层状结构, ZX1 井, 4864.73 m; F—块状结构, ZX1 井, 4607.33 m

图 2 四川盆地筇竹寺组页岩的岩心照片以及在偏光显微镜下观察到的沉积构造

Fig. 2 Sedimentary structures observed in core samples and under a polarizing microscope
of the Qiongzhusi Formation shale in Sichuan Basin

对页岩脆性的影响, 将脆性矿物的含量纳入岩性分类标准。一般认为, 页岩气储集层的脆性矿物含量应大于 40.0%, 才能具备增产改造的商业开发价值。因此, 以脆性矿物含量 40.0% 和 60.0% 为界限分为低脆性、中等脆性和高脆性, 当石英+长石含量 60.0% 时则划分为强脆性 (李丹龙等, 2023)。同时, 将黏土矿物含量、长英质矿物含量 (石英+长石) 和碳酸盐矿物含量 (方解石+白云石) 以 40.0% 和 60.0% 为界绘制矿物含量三端元图版 (图 3), 结合三端元图发现筇竹寺组主要发育强脆性硅质 (粉砂质) 页岩 (I), 中等脆性硅质/黏土混合页岩 (VI) 以及高脆性中硅质页岩 (V), 将此岩性作为岩相综合命名后缀。基于以上划分方案, 筇竹寺组页岩岩相可划分为: 中有机质层状硅质页岩 (MB I)、贫有机质粉砂质页岩 (P I)、贫有机质块状硅质页岩 (PM V)、贫有机质层状硅质/黏土混合页岩 (PL VI) 以及富有机质层状硅质页岩 (RL I) 共计 5 种岩相类型。

3 测井参数特征分析与数据预处理

3.1 数据准备

研究数据来源于四川盆地德阳—安岳裂陷槽内

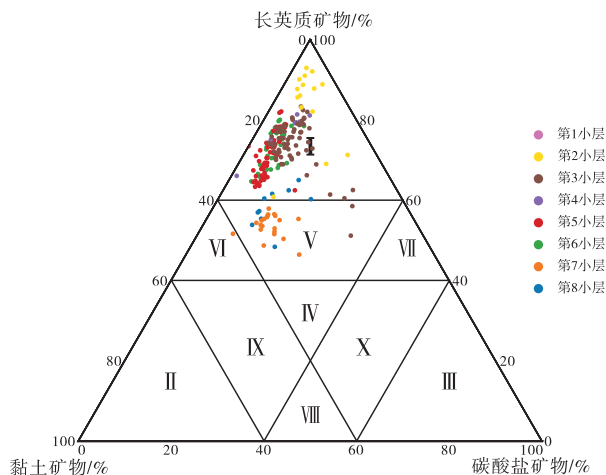


图 3 四川盆地筇竹寺组页岩矿物组分三端元投点图

Fig. 3 Three-end-member plot of shale mineral composition
in the Qiongzhusi Formation, Sichuan Basin

的关键取心井 (ZX1 井) 下寒武统筇竹寺组页岩, 其中 525 个岩相样本经过数据匹配、人工标注并分配了分类标签。为了训练和验证模型的性能, 这些带标签的数据集按照 7:3 的比例随机划分为训练集和测试集。在岩相识别过程中, 采用 LabelEncoder 算法对岩相标签进行编码转换, 使用 0、1、2、3 和 4 分别表示以下 5 种不同的岩相类型。其中, 0: 中有机质层状硅质页岩 (MB I)、1: 贫有机质粉

砂质页岩 (P I)、2: 贫有机质块状硅质页岩 (PM V)、3: 贫有机质层状硅质/黏土混合页岩 (PLVI) 和 4: 富有机质层状硅质页岩 (RL I)。各岩相类型的样本数量及占总标签样本的百分比见表 1 所示。

3.2 测井响应分析

鉴于测井基本方法原理, 自然伽马 (GR)、无铀伽马 (KTH)、铀 (U)、声波时差 (AC)、补偿中子 (CNL)、密度 (DEN) 和电阻率 (RT) 这 7 种测井参数对页岩 TOC 及岩性判识有一定的敏感性, 因此选取了以上 7 种测井参数进行响应特征分析。7 种测井参数的散点矩阵图(图 4) 对角线上的子图为每一种测井参数的核密度分布图, 它直观地展示了 5 种页岩岩相测井参数的数据分布情况。横轴代表了测井参数的取值, 密度曲线宽, 表明数据分散程度高, 分布范围更广。例如富有机质层状硅质页岩 (RL I)、中有机质层状硅质页岩 (MB I) 的 GR 密度曲线更宽, 表明 2 种页岩岩相的 GR 总体数值范围更广, 核密度曲线峰值对应的 x 轴数值代表了数据分布中密度最高的区域。结合图 4 和表 2 可知, 中有机质层状硅质页岩 (MB I) 呈较高的 GR 和 KTH 值, 表明该岩相富含有机质和硅质矿物, U 值中高。贫有机质粉砂质页岩 (P I) 呈较低的 GR 值, 矿物成分主要由陆源粉砂质矿物 (长英质) 组成, U 值较低。贫有机质块状硅质页岩 (PM V) 也表现为较低的 GR 值, 富含陆源硅质矿物, RT 值较高, U 值较低。贫有机质层状硅质/黏土混合页岩 (PLVI) GR 值也较低, 硅质矿物与黏土矿物的比例较为平衡。富有机质层状硅质页岩 (RL I) 具有较高的 GR 值, 富含有机质, 硅质矿物丰富, U 值适中偏高, RT 值较低。

3.3 测井参数相关性分析

在模型构建过程中, 通过合理的特征选择, 可

以有效降低特征维度, 从而减少计算时间并减轻过拟合风险。某些特征可能包含噪声, 或特征间存在高度相关性, 这些冗余信息会影响模型的计算效率并增加复杂度。因此, 采用特征选择方法去除这些干扰因素 (王民等, 2023), 保留对模型性能至关重要的特征, 有助于提升模型的训练效果和预测能力。本研究通过相关性分析法对 7 种测井参数进行取舍, 保留具有高信息量的测井参数。通过观察散点矩阵图(图 4) 初步判断变量之间的相关性。若两个变量之间的散点图呈现明显的线性趋势, 则二者存在较强的相关性 (闫佳飞等, 2025)。图 4 中点的密集程度和分布情况可以初步判断变量间的相关性和数据分布情况: GR (自然伽马) 和 U (铀), AC (声波时差) 和 CNL (补偿中子) 2 组变量的数据点分别在散点矩阵图中呈现明显的线性趋势, 通过变量相关性热力图分析 (皮尔逊相关系数) 设定阈值, 当一组测井参数之间相关性大于 0.90 的时候, 为避免冗余, 选择保留其中一个测井参数。由图 5 可知, GR (自然伽马) 和 U (铀) 之间相关性高达 0.92, 表明 2 个测井参数在页岩岩相分类任务中提供了大量相同的信息, 选择保留 GR (自然伽马), 最终确定模型的特征参数为 6 条测井曲线变量: 自然伽马 (GR)、无铀伽马 (KTH)、声波时差 (AC)、补偿中子 (CNL)、密度 (DEN) 和电阻率 (RT)。

3.4 数据标准化

鉴于上述测井参数相关性的分析, 将 GR、KTH、AC、CNL、DEN 和 RT 6 种测井参数作为模型的特征值输入。每个样本由一个 7 维向量组成, 其中包含 6 个特征参数值和 1 个对应的岩相标签。考虑到不同测井参数之间存在量纲和数值范围的差异, 实施标准化处理至关重要 (杨洋等, 2021)。标准化可以将多维特征映射到统一的度量空间, 消

表 1 四川盆地 ZX1 井筇竹寺组页岩相标签数据

Table 1 Data on shale facies labels for the Qiongzhusi Formation in Well ZX1, Sichuan Basin

岩相标签	岩相	样本数/个	占比/%
0	中有机质层状硅质页岩 (MB I)	126	24.00
1	贫有机质粉砂质页岩 (P I)	226	43.00
2	贫有机质块状硅质页岩 (PM V)	65	12.38
3	贫有机质层状硅质/黏土混合页岩 (PLVI)	2	0.38
4	富有机质层状硅质页岩 (RL I)	106	20.19
统计		525	

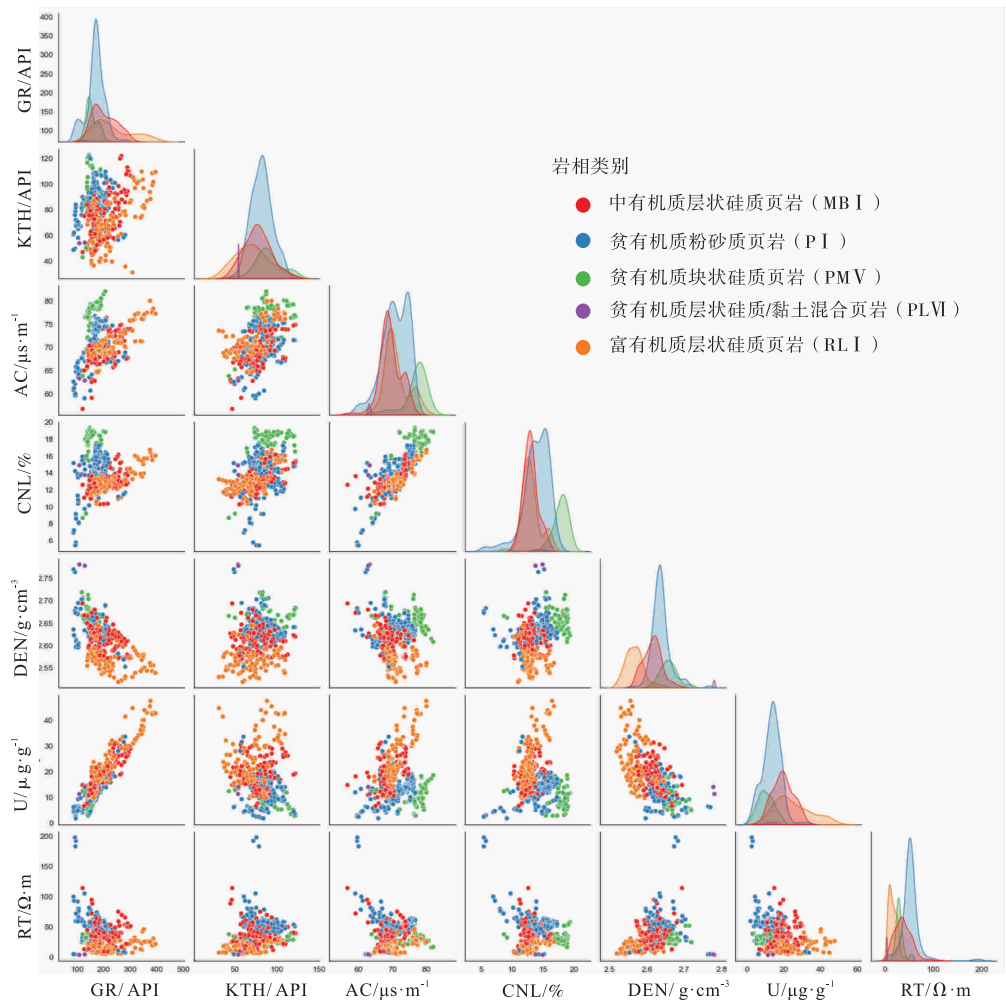


图 4 四川盆地 ZX1 井筇竹寺组页岩地层 7 种测井参数变量的散点矩阵图
Fig. 4 Scatter matrix of seven log parameter variables for the Qiongzhusi Formation shale in Well ZX1, Sichuan Basin

表 2 四川盆地 ZX1 井筇竹寺组页岩不同岩相类型其测井响应值范围及特征
Table 2 Range and characteristics of well logging response values for different lithofacies types of
the Qiongzhusi Formation shale in Well ZX1, Sichuan Basin

岩相	数量	GR/ API	KTH/ API	AC/ $\mu\text{s}\cdot\text{m}^{-1}$	CNL/%	DEN/ $\text{g}\cdot\text{cm}^{-3}$	U/ $\mu\text{g}\cdot\text{g}^{-1}$	RT/ $\Omega\cdot\text{m}$
		范围/平均值	范围/平均值	范围/平均值	范围/平均值	范围/平均值	范围/平均值	范围/平均值
中有机质层状硅质页岩(0)	126	120.22~301.17/ 202.73	44.85~122.04/ 77.07	56.73~76.39/ 69.30	10.30~16.16/ 12.93	2.57~2.69/ 2.61	6.73~31.28/ 19.41	9.06~114.94/ 37.70
贫有机质粉砂质页岩(1)	226	85.77~280.99/ 167.88	48.80~121.54/ 82.03	56.02~76.758/ 70.80	5.45~17.25/ 13.86	2.55~2.77/ 2.64	1.89~33.56/ 13.29	4.61~197.73/ 52.84
贫有机质块状硅质页岩(2)	65	117.09~205.85/ 167.88	42.93~122.81/ 89.55	63.83~82.06/ 76.80	8.69~19.32/ 17.59	2.61~2.72/ 2.66	2.89~18.61/ 10.21	16.01~57.91/ 30.49
贫有机质层状硅质/黏土混合页岩(3)	2	107.64~128.25/ 177.95	53.89~54.35/ 54.12	62.85~63.29/ 63.07	13.77~14.90/ 14.33	2.77~2.78/ 2.77	11.36~14.09/ 12.73	3.72~4.32/ 4.02
富有机质层状硅质页岩(4)	106	137.73~395.59/ 232.42	31.09~109.94/ 70.21	64.65~80.05/ 71.15	10.39~16.74/ 13.05	2.52~2.65/ 2.57	9.83~47.46/ 24.76	4.84~45.94/ 16.37

除量纲效应，确保各特征能够公平比较并合理加权，从而提升预测精度和模型的泛化能力。通过公

式（1）最大最小归一化函数，将输入测井参数值映射到 $[0, 1]$ 区间：

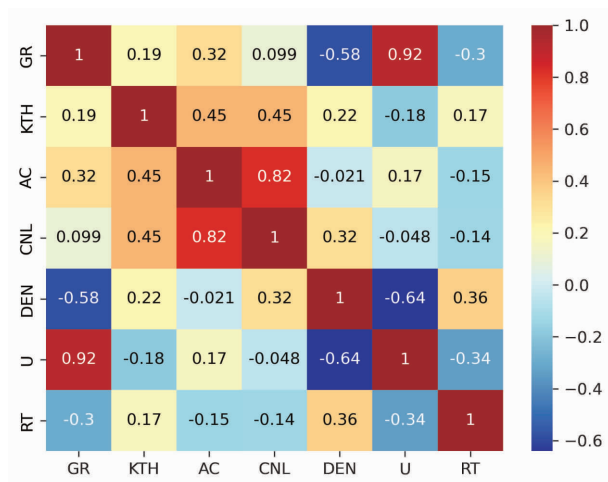


图 5 四川盆地 ZX1 井筇竹寺组页岩地层 7 种测井参数变量的相关性热力图

Fig. 5 Correlation heatmap of seven log parameter variables for the Qiongzhusi Formation shale in Well ZX1, Sichuan Basin

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

其中: X 为原始样本数据, $X_{\min}$ 为样本数据最小值, $X_{\max}$ 为样本数据最大值, X^* 为归一化后的数据, 通过归一化处理有助于提高对数据尺度敏感的机器学习算法的性能(如支持向量机、线性回归等)。

4 页岩气储集层岩相识别方法优选

4.1 集成学习

集成学习的基本原理是通过结合多个基学习器来提升预测精度(谭勇等, 2024)。各基学习器在不同数据子集上的优异表现能有效降低过拟合风险, 并在特征捕捉上互为补充, 从而增强模型的鲁棒性。基学习器作为集成学习框架的核心, 其性能直接决定了集成模型的性能表现。因此, 基学习器的选择至关重要, 应充分考虑其多样性、独立性和性能。理想的基学习器能够不同角度学习和预测数据, 从而提高集成效果, 减少基学习器间预测错误的相关性, 有效提升整体性能, 并增强集成模型在未知数据上的泛化能力。

4.2 基学习器选择

支持向量机(SVM)是一种广泛应用于多分类任务的有效监督学习算法(高飞等, 2024)。其基本原理是构建一个最优分界面(即超平面), 以

区分不同的岩相类型。对于线性可分的情况, SVM能够在特征空间中直接找到一个最佳的“分隔面”, 从而将不同岩相的样本有效地分隔开。然而, 在实际应用中, 岩相特征数据通常是非线性可分的。为解决这一问题, SVM采用核函数将数据从低维空间映射到更高维空间, 进而使得在新空间中原本无法分开的数据变得线性可分, 以此来解决复杂的岩相分类问题。

决策树(Decision Tree)模型在分类任务中通过树状结构将复杂的分类问题逐步分解为一系列简单的决策过程(何小龙等, 2024)。模型利用测井参数(如自然伽马、密度、声波时差等)作为输入特征, 从根节点开始, 通过一系列基于特征值的“是”或“否”判断, 将数据逐步划分成不同的分支, 直至达到叶节点, 最终叶节点对应不同的岩相类型。在每个节点上, 模型根据一定的准则(如信息增益、基尼指数或熵)选择最优的特征及其分割点, 使得划分后的子节点内部数据更为纯净, 即属于同一页岩岩相的样本占比更高。这一过程类似于根据多个条件逐层筛选数据, 直至找到最匹配的岩相类别。决策树结构简单直观, 易于理解, 能够有效捕捉测井参数与岩相类型之间的非线性关系, 且对数据预处理要求较低, 也适用于页岩岩相分类任务。

K最近邻算法(KNN)是一种基于距离度量的分类方法(李昌等, 2023)。“距离”是衡量不同样本之间相似性的重要指标, 通常使用欧几里得距离来计算。对于每个待分类样本, KNN算法通过计算其与其他样本在多个测井参数(如声波时差、铀含量、自然伽马等)上的差异, 来确定样本间的相似度。距离越小, 表示样本越相似。具体来说, 欧几里得距离是通过各个测井参数之间的差值平方和开方得出的, 用于量化样本间的整体差异。根据最相似的K个样本的类别信息, KNN算法预测待分类样本的岩相类别。

一维卷积神经网络(1D-CNN)是一种专门处理序列数据的深度学习模型(袁士博等, 2023)。在页岩岩相分类中, 1D-CNN通过模拟生物神经系统的工作原理来自动提取数据中的特征。模型通过多个卷积层对输入的测井参数数据进行处理, 对于页岩岩相分类, 网络会自动学习并识别出哪些数据模式(如特定的声波时差和铀含量的变化趋势)

对于区分不同岩相类型最为重要。这些模式可能是数据中反映岩性变化、孔隙度、岩相界限等关键因素的“信号”。然后经过池化层（用来降低数据的复杂度）和全连接层，最终输出一个预测结果，确定样本属于哪种岩相类型。1D-CNN 的优势在于它能够自动从原始数据中学习有效的特征，而不需要手动提取特征，提升了岩相分类的准确性和效率。

上述 3 种机器学习算法和 1 种深度学习算法均可用于分类算法中的基学习器，但各自均有不同的优缺点和适用场景，见表 3 所示。

4.3 传统机器学习与深度学习算法融合的模式

4.3.1 Voting 投票机制

本研究采用了集成学习中的 Voting（投票法）算法，旨在提高岩相分类任务的精度和稳定性。Voting 算法结合多个基学习器的预测结果，从而提升整体模型的表现。其原理基于集成学习的核心思想，即通过集成多个基模型的优势，弥补单一模型的不足，增强分类系统的泛化能力。

在 Voting 算法中，每个基学习器根据样本给出预测结果，输出为类别标签或类别概率。Voting 算法主要分为 2 种类型：硬投票（Hard Voting）和软投票（Soft Voting）（黄富幸和韩文花，2023）。在硬投票机制中，每个基学习器为样本进行分类，最终的预测结果由所有模型预测的类别标签中出现频次最多的标签决定，适用于大多数分类任务。而在软投票机制中，最终的预测结果基于所有基学习器预测的类别概率的加权平均值，每个基学习器的输出概率被赋予与其性能相关的权重，最终选择概率值最高的类别作为预测结果。软投票方法通过充

分利用各模型的输出信息，能够在模型性能差异较大时提供更稳定和准确的分类结果。

在页岩岩相分类任务中，由于数据的复杂性及不同模型在特征提取能力上的差异，集成 Voting 算法通过聚合多种模型的预测结果，预期能够有效提升分类的鲁棒性、准确性和泛化能力。决策树（Decision Tree）、支持向量机（SVM）、一维卷积神经网络（1D-CNN）及 K 近邻（KNN）算法各自具有不同的优势，SVM 擅长处理非线性数据关系，决策树抗数据干扰能力强，1D-CNN 能够有效学习时序特征，而 KNN 模型简单，降低了训练难度。通过 Voting 机制，这些模型能够在互补性基础上优化决定最终分类结果，从而提升融合的集成学习模型在页岩岩相识别中的精度。

4.3.2 模型搭建

研究中运用了 Voting 软投票方法对经过训练与优化的 4 类基学习器，即支持向量机（SVM）、决策树、K-近邻（KNN）以及一维卷积神经网络（1D-CNN）模型进行集成操作。构建出一个在准确率方面表现更优且泛化能力更强的融合强学习器，搭建流程图见图 6。

模型构建过程如下：

- 1) 对原始数据集进行数据预处理，包括降维处理和数据标准化，以提高模型的训练效率和预测性能。
- 2) 将处理后的数据样本集作为 SVM、决策树、KNN、1D-CNN 模型的特征输入，对每个模型进行训练和参数调优，从而得到 4 个训练好的基学习器。在参数调优的过程中，SVM 模型需根据数据特点选择合适的核函数、惩罚参数等；决策树模型要确定树的结构参数，如深度、节点最小样本数

表 3 4 类基学习器（模型）算法的优缺点分析
Table 3 Advantages and disadvantages of four types of machine learning algorithms

特征	模型			
	SVM	决策树	KNN	1D-CNN
优点	在未知数据集上表现优秀（即优秀的泛化能力）、适用于解决测井参数之间非线性问题、适用于小样本学习	对输入的特征值要求不高、抗数据干扰能力（即鲁棒性）优秀、决策过程直观，可以清晰地看到模型如何根据输入特征做出决策（即可解释强）	模型简单易懂，不需要复杂的训练过程	能够从数据中自动提取重要的特征适用于测井参数之间复杂的非线性关系适用于大规模数据处理
缺点	不适用于大型数据集、模型抗数据抗干扰能力（即鲁棒性）较差	容易过拟合（即泛化能力差）、对各种方案的出现概率的确定有时主观性较大，可能导致决策失误	计算开销大对数据噪声敏感	对模型参数调优要求高

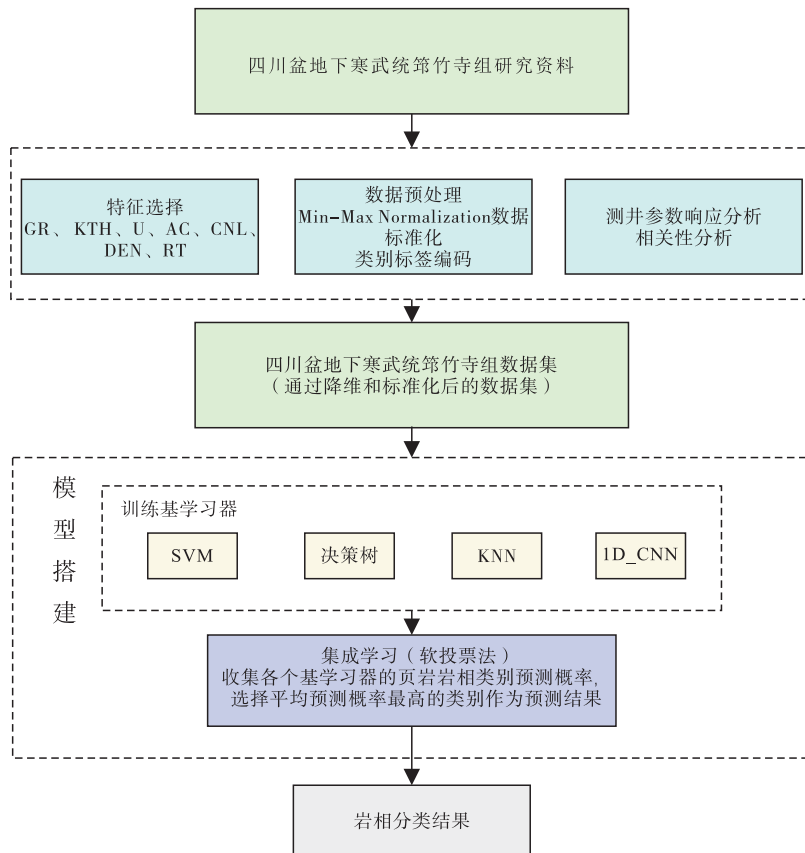


图 6 集成学习模型搭建流程图

Fig. 6 Flow chart of model building principle

等; KNN 模型要确定合适的 K 值; 1D-CNN 模型要设计好网络结构, 包括层数、卷积核大小等参数, 并完成参数初始化和训练过程。

3) 将训练好的这 4 个基学习器采用软投票的方式集成为一个融合强学习器。对于每一个测试样本, 各个模型都会输出其预测结果, 将这些结果按照软投票规则进行整合。软投票是对每个模型预测的类别概率进行加权平均, 得到综合的类别概率, 最终取概率最高的类别作为该样本的预测类别。

软投票法的公式为:

设基模型的集合为 $\{M_1, M_2, \dots, M_n\}$, 对于类别集合 $\{c_1, c_2, \dots, c_k\}$ 和一个测试样本 x , 第 i 个基学习器 M_i 对样本 x 预测属于类别 c_j 的概率为 p_{ij} 。

软投票法计算样本 x 属于类别 c_j 的综合概率 P_j 的公式为:

$$P_j = \sum_{i=1}^n w_i p_{ij} \quad (2)$$

其中 w_i 是第 i 个基学习器的权重。通常情况

下, 如果没有特殊说明, 可以假设所有基学习器的权重相等, 即 $w_i = \frac{1}{n}$, 此时公式变为 $P_j = \frac{1}{n} \sum_{i=1}^n p_{ij}$ 。

最终预测的类别 C 是综合概率最大的类别, 即 $C = \arg \max_{1 \leq j \leq k} P_j$ 。这个公式表示在所有的类别 c_j ($1 \leq j \leq k$) 中, 选择使得综合概率 P_j 最大的那个类别作为样本的最终预测类别。

4.4 评价指标

由于训练数据中各个类型的页岩岩相样本数量分布存在不均衡现象。例如, 贫有机质层状硅质/黏土混合页岩 (PL VI) 样本仅占总样本数的 0.38%, 这可能导致模型在训练过程中偏向于学习多数类样本的特征, 而忽视少数类样本的特征 (Saito and Rehmsmeier, 2015)。单纯依赖模型的准确率作为唯一评估指标往往无法全面反映模型在页岩岩相分类预测任务中的表现。因此, 采用精确率 (Precision)、召回率 (Recall)、准确率 (Accuracy)、F1 值 (F1 Score) 综合评价分类模型的优劣。

混淆矩阵是一种有效的工具，用于比较模型的预测结果与实际标签之间的差异。在表 4 中，TP、FP、FN 和 TN 分别表示真正例（True Positive）、假正例（False Positive）、假负例（False Negative）和真负例（True Negative）样本。尽管表 4 展示的是二分类问题，但可以通过将每个类别视为一个特定的正例（如 0、1、2、3 或 4），并将其余类别视为负例，从而将多分类问题转化为二分类问题。基于上述，可得出模型评估指标精确率（PRE）、召回率（REC）、F1-score（F1）和准确率（ACC）计算公式如下：

$$\begin{aligned} \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \\ \text{PRE} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{REC} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{F1} &= 2 \times \frac{\text{PRE} \times \text{REC}}{\text{PRE} + \text{REC}} \end{aligned} \quad (3)$$

表 4 岩相分类结果混淆矩阵

Table 4 Lithofacies classification confusion matrix

真实结果	预测结果	
	正例	负例
正例	TP(真正例)	FN(假负例)
负例	FP(假正例)	TN(真负例)

4.5 预测岩相与标签岩相的比较

利用上述 4 种基学习器及集成学习（Voting）算法对研究区 ZX1 井页岩岩相测试集数据进行岩相分类。通过各个模型预测准确率柱状图比较（图 7），以及分类性能热力图（图 8），观察这 5 种算法在页岩岩相分类任务中的性能。热力图中颜色越深代表分类算法性能越差，从而直观地反映各算法的优劣。从结果来看，决策树和一维卷积神经网络对标签 3—贫有机质层状硅质/黏土混合页岩（PLVI）分类性能较差，且前 4 种单一基学习器对标签 0—中有机质层状硅质页岩（MBI）类别的精确率（PRE）和召回率（REC）表现均不理想，根据以上数据结果显示单一的机器/深度学习模型在页岩岩相分类任务中对于复杂情况下页岩岩相识别的精度不够，且受样本比例影响明显。相较之下，测试集预测结果显示集成学习（Voting）方法

不仅准确率（ACC）提升至 94.0%，且对每一类岩相的精确率、召回率和 F1 值均有明显的提升，表明了集成学习（Voting）方法能够通过加权投票机制，整合多个弱分类器的预测结果，融合成一个强分类器，实现模型间的互补，提升算法的整体性能。

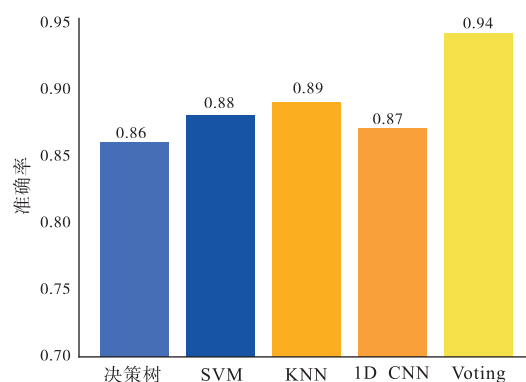


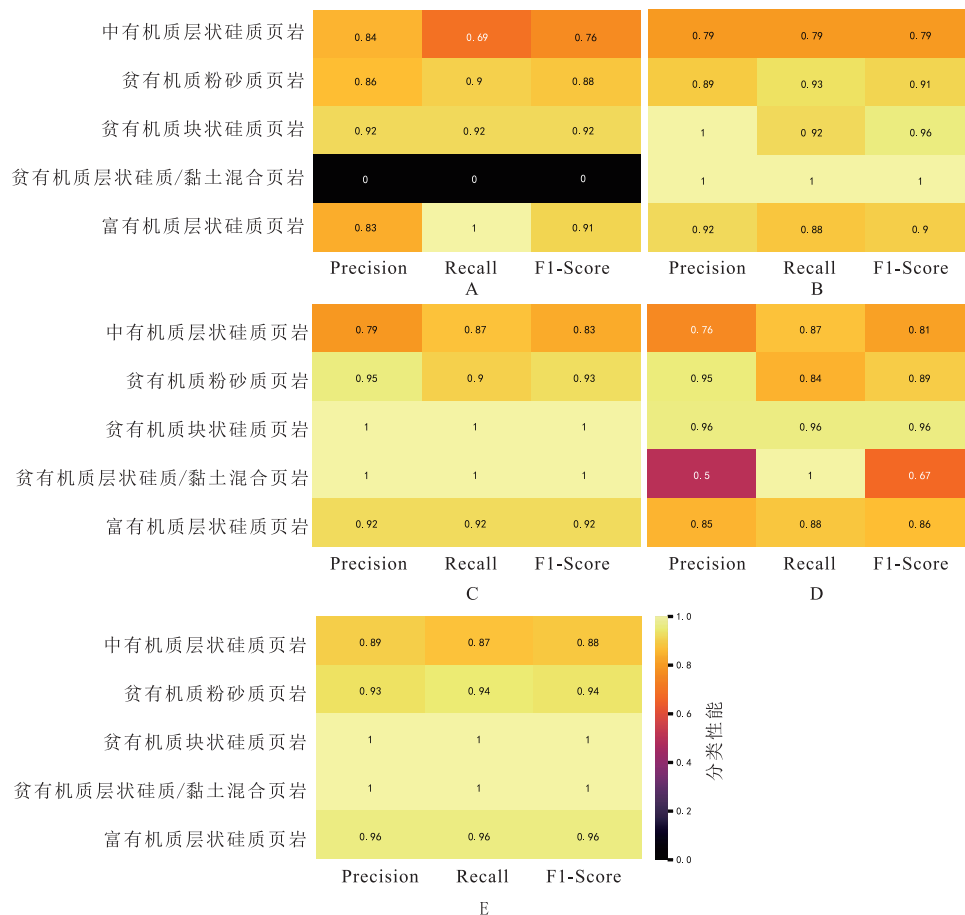
图 7 4 种基学习器（模型）及集成学习模型（Voting）准确率对比柱状图

Fig. 7 Bar chart comparing accuracy of four base learners (models) and ensemble learning model (Voting)

图 9 显示了 ZX1 井取心段岩心测试资料标定的真实岩相（即标签岩相）与 4 种基学习器模型（SVM、1D-CNN、决策树、KNN）、集成学习（Voting）模型预测页岩岩相的对比结果。此外，由于在 4444 ~ 4452 m、4508 ~ 4549 m、4624 ~ 4739 m 和 4811 ~ 4851 m 区间为非取心段，无法提供岩心刻度下的岩相标签，因此，基于上述已构建的 4 类基学习器模型与集成学习模型，利用测井曲线数据预测了这些空白区段的页岩岩相类型。

本研究针对划分的 5 种岩相进行分类并利用软件成图，并标注成不同的颜色以示区分：富有机质层状硅质页岩（RLI：红色）、贫有机质块状硅质页岩（PMV：蓝色）、贫有机质粉砂质页岩（PI：黄色）、中有机质层状硅质页岩（MBI：橙色）、贫有机质层状硅质/黏土混合页岩（PLVI：绿色）。从图 9 可以看出，各模型对页岩岩相类型的预测效果存在一定的差异。

KNN 模型对贫有机质粉砂质页岩（PI：黄色）的分类预测较为稳定，但在复杂岩相过渡区域（如 4500 ~ 4520 m 和 4600 ~ 4620 m）表现略逊，无法稳定识别。这与 KNN 模型的特点有关，其对于噪声数据较为敏感，对局部噪声无平滑能力，易



A—决策树算法分类性能热力图；B—支持向量机（SVM）算法分类性能热力图；C—K 最近邻（KNN）算法分类性能热力图；

D—一维卷积神经网络（1D-CNN）算法分类性能热力图；E—决策树算法分类性能热力图

图 8 四川盆地 ZX1 井筇竹寺组页岩 4 种机器学习器和集成学习算法岩相分类性能热力图

Fig. 8 Heatmap of lithofacies classification performance for four base learners and ensemble learning algorithm for the Qiongzhusi Formation in Well ZX1, Sichuan Basin

将噪声点误判为邻近类，导致页岩岩相识别类型变化快。

SVM 模型对富有机质层状硅质页岩（RL I：红色）和贫有机质块状硅质页岩（PM V：蓝色）的识别效果较强，这是由于 2 类页岩岩相矿物组成差异显著（RL I 高 TOC，PM V 高石英含量），测井特征在高维空间线性可分，SVM 通过最大间隔超平面划分类别，对线性可分问题具有理论最优性，尤其适合矿物成分主导的页岩岩相分类。但在细小的岩相区域如中有机质层状硅质页岩（MB I：橙色）的分类上存在一定误差，可能是由于 SVM 模型在处理非线性问题时，对复杂页岩岩相的识别能力有限。

1D-CNN 模型的整体预测效果较均衡，但在复杂过渡区域的边界处理上有一定不足，原因可能是

由于过渡区岩相变化伴随测井曲线的渐变或突变，而 1D-CNN 层数较浅（3 层），高层特征抽象能力有限，以及池化操作导致信息损失：最大池化（pool_size=2）削弱了薄互层（如纹层页岩）的细节特征，导致边界附近预测结果模糊化。

决策树模型则在连续区域的岩相预测表现均衡，在复杂模式识别上可能不如其他模型，决策树按局部最优策略逐层分裂，难以建模多参数协同效应（如 TOC 与脆性指数的非线性耦合关系），且树深度受限（max_depth=8）时，无法生成足够细粒度的规则描述过渡区岩相混合特征。但是由于决策树模型对数据的完整性要求相对较低，在数据可能存在误差或者缺失测井参数的情况下具有一定的优势。

Voting 模型的预测结果与岩心标定的岩相（标

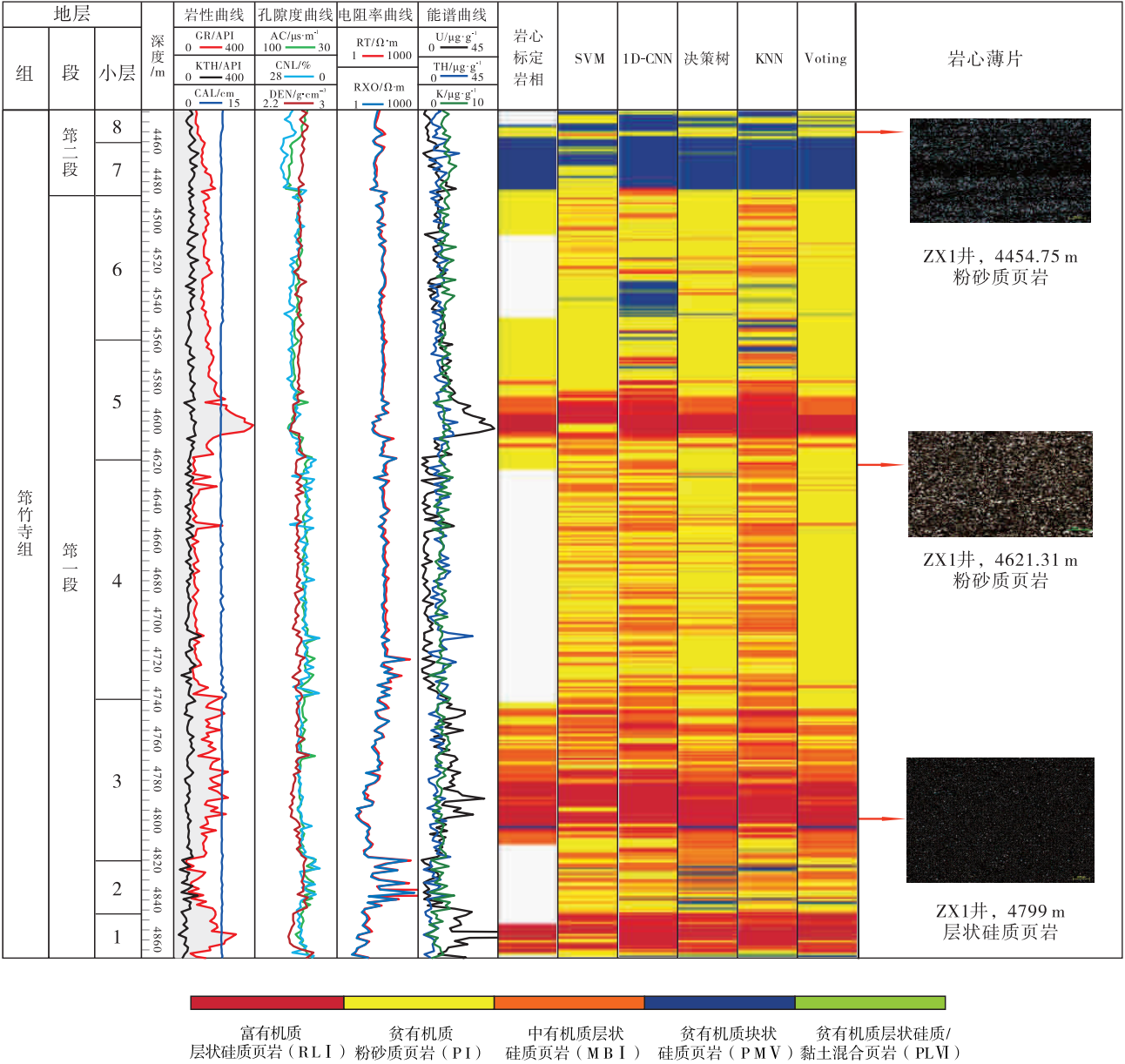


图 9 四川盆地 ZX1 井筇竹寺组页岩岩相测井识别结果

Fig. 9 Lithofacies logging identification results for the Qiongzhusi Formation shale in Well ZX1, Sichuan Basin

签岩相) 契合度最高, 尤其是对富有机质层状硅质页岩 (RL I: 红色) 和贫有机质粉砂质页岩 (PI: 黄色) 的边界区域预测精度最佳。这是由于算法综合了不同模型的优势: KNN 捕捉局部相似性, SVM 强化大类间隔, CNN 提取深层特征, 决策树鲁棒性强, 集成后方差显著降低。且单一模型在边界区域的误判方向不同 (如 KNN 因噪声误判, SVM 因线性核误判), 而投票机制通过多数表决纠正了个体偏差。

5 结论

- 1) 依据“有机质+纹层构造+岩性及脆性”3重信息, 四川盆地寒武统筇竹寺组页岩岩相可划分为5种岩相类型: 中有机质层状硅质页岩 (MB I)、贫有机质粉砂质页岩 (PI)、贫有机质块状硅质页岩 (PM V)、贫有机质层状硅质/黏土混合页岩 (PL VI) 以及富有机质层状硅质页岩 (RL I)。
- 2) 鉴于测井曲线人工分类、单一机器/深度

学习识别岩相方法存在的局限性,提出一种融合传统机器学习模型——决策树、支持向量机(SVM)、K近邻(KNN)以及深度学习模型——一维卷积神经网络(1D-CNN)的集成学习(Voting)模型,实现了4个弱基学习器模型集成于一个强学习器模型的搭建。

3) 集成学习(Voting)模型在筇竹寺组页岩岩相预测精度上表现最佳,对比标签岩相显示预测精度达到了94.0%,显著优于单一基学习器模型,有效克服了传统分类算法受岩相样本不均衡影响引起的识别偏差,提升了稀缺样本的岩相识别能力,具有较好的推广应用价值。

参考文献 (References)

- 董大忠,王玉满,李新景,邹才能,管全中,张晨晨,黄金亮,王淑芳,王红岩,刘洪林,拜文华,梁峰,吝文,赵群,刘德勋,邱振. 2016. 中国页岩气勘探开发新突破及发展前景思考. 天然气工业, 36(1): 19-32. [Dong D Z, Wang Y M, Li X J, Zou C N, Guan Q Z, Zhang C C, Huang J L, Wang S F, Wang H Y, Liu H L, Bai W H, Liang F, Lin W, Zhao Q, Liu D X, Qiu Z. 2016. Breakthrough and prospect of shale gas exploration and development in China. Natural Gas Industry, 36(1): 19-32]
- 高飞,曲志鹏,魏震,朱剑兵,程远锋. 2024. 基于机器学习方法的测井岩相分类研究. 地球物理学进展, 39(3): 1173-1192. [Gao F, Qu Z P, Wei Z, Zhu J B, Cheng Y F. 2024. Study on well-log lithofacies classification based on machine learning methods. Progress in Geophysics, 39(3): 1173-1192]
- 何晓,梁峰,李海,郑马嘉,赵群,刘勇,刘文平. 2024. 四川盆地寒武统筇竹寺组海相页岩气高产井突破与富集模式. 中国石油勘探, 29(1): 142-155. [He X, Liang F, Li H, Zheng M J, Zhao Q, Liu Y, Liu W P. 2024. Breakthrough and enrichment mode of marine shale gas in the Lower Cambrian Qiongzhusi Formation in high-yield wells in Sichuan Basin. China Petroleum Exploration, 29(1): 142-155]
- 何小龙,张兵,杨凯,何一帆,李琢. 2024. 基于沉积微相特征挖掘的随机森林岩石相测井识别方法:以新场地区须家河组二段致密砂岩为例. 物探与化探, 48(5): 1337-1347. [He X L, Zhang B, Yang K, He Y F, Li Z. 2024. A log-based lithofacies identification method based on random forest and sedimentary microfacies characteristics: a case study of tight sandstones in the second member of the Xujiahe Formation in the Xinchang area. Geophysical and Geochemical Exploration, 48(5): 1337-1347]
- 黄富幸,韩文花. 2023. 基于Voting机制的IMA-BP不平衡数据分类算法. 科学技术与工程, 23(27): 11698-11705. [Huang F X, Han W H. 2023. Classification algorithm of IMA-BP for unbalanced data based on voting mechanism. Science Technology and Engineering, 23(27): 11698-11705]
- 黄莉莎,闫建平,刘明洁,张庄,叶素娟,张帆,钟光海,王敏,王军,耿斌. 2022. 深层致密砂岩气储层成岩相测井识别及应用:以川西坳陷大邑须家河组须三段为例. 中国矿业大学学报, 51(1): 107-123. [Huang L S, Yan J P, Liu M J, Zhang Z, Ye S J, Zhang F, Zhong G H, Wang M, Wang J, Geng B. 2022. Diagenetic facies logging identification and application of deep tight sandstone gas reservoir: a case study of the third member of Xujiahe formation in Dayi area of western Sichuan depression. Journal of China University of Mining & Technology, 51(1): 107-123]
- 姜鹏飞,吴建发,朱逸青,张德宽,吴伟,张芮,吴喆,王青,杨雨然,杨雪,伍秋姿,陈丽清,何一凡,张娟. 2023. 四川盆地海相页岩气富集条件及勘探开发有利区. 石油学报, 44(1): 91-109. [Jiang P F, Wu J F, Zhu Y Q, Zhang D K, Wu W, Zhang R, Wu Z, Wang Q, Yang Y R, Yang X, Wu Q Z, Chen L Q, He Y F, Zhang J. 2023. Enrichment conditions and favorable areas for exploration and development of marine shale gas in Sichuan Basin. Acta Petrolei Sinica, 44(1): 91-109]
- 蒋裕强,宋益滔,漆麟,陈雷,陶艳忠,甘辉,吴佩津,叶子亿. 2016. 中国海相页岩岩相精细划分及测井预测:以四川盆地南部威远地区龙马溪组为例. 地学前缘, 23(1): 107-118. [Jiang Y Q, Song Y T, Qi L, Chen L, Tao Y Z, Gan H, Wu P J, Ye Z Y. 2016. Fine lithofacies of China's marine shale and its logging prediction: a case study of the Lower Silurian Longmaxi marine shale in Weiyuan area, southern Sichuan Basin, China. Earth Science Frontiers, 23(1): 107-118]
- 李昌,王鑫,冯周,宋连腾. 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. 海相油气地质, 28(4): 433-440. [Li C, Wang X, Feng Z, Song L T. 2023. Comprehensive logging identification of microbial carbonate lithofacies based on KNN classification algorithm: a case study of Dengying Formation in GM area, Sichuan Basin. Marine Origin Petroleum Geology, 28(4): 433-440]
- 李丹龙,伏美燕,邓虎成,胥旺,刘四兵,吴冬. 2023. 上扬子地区下寒武统牛蹄塘组富有机质页岩岩相及沉积环境分析:以贵州温水村剖面为例. 天然气地球科学, 34(3): 445-459. [Li D L, Fu M Y, Deng H C, Xu W, Liu S B, Wu D. 2023. Analysis of lithofacies and sedimentary environment of shale deposited in shelf facies: a case study of the Wenshuicun section in Guizhou Province, South China. Natural Gas Geoscience, 34(3): 445-459]
- 梁峰,吴伟,张琴,罗超,王玉满,刘宇,姜巍,卢斌. 2024. 四川盆地南部下寒武统筇竹寺组页岩孔隙结构特征与页岩气赋存模式. 天然气工业, 44(3): 131-142. [Liang F, Wu W, Zhang Q, Luo C, Wang Y M, Liu Y, Jiang W, Lu B. 2024. Shale pore structure characteristics and shale gas occurrence pattern of the Lower Cambrian Qiongzhusi Formation in the southern Sichuan Basin. Natural Gas Industry, 44(3): 131-142]
- 聂海宽,党伟,张珂,苏海琨,丁江辉,李东晖,刘喜武,李沛,李鹏,杨升宇,赵建华,刘秘,陈前,孙川翔,王鹏威,刘子驿. 2024. 中国页岩气研究与发展20年:回顾与展望. 天然气工业, 44(3): 20-52. [Nie H K, Dang W, Zhang K, Su H K, Ding J H, Li D H, Liu

- X W, Li P, Li P, Yang S Y, Zhao J H, Liu M, Chen Q, Sun C X, Wang P W, Liu Z Y. 2024. Two decades of shale gas research & development in China: review and prospects. *Natural Gas Industry*, 44(3): 20–52]
- 彭丽, 伍轶鸣, 练章贵, 彭鹏, 王剑, 苏洲, 易珍丽. 2019. 陆相断陷湖盆高频层序特征及其沉积演化: 以渤海湾盆地济阳坳陷沙三下亚段为例. *石油与天然气地质*, 40(4): 789–798. [Peng L, Wu Y M, Lian Z G, Peng P, Wang J, Su Z, Yi Z L. 2019. Features and sedimentary evolution of high-frequency sequence in continental lacustrine rift basin: Example of the lower Shahejie member 3 in Jiyang Depression, Bohai Bay Basin. *Oil & Gas Geology*, 40(4): 789–798]
- 邱玉超, 罗冰, 夏茂龙, 陈友莲, 戴鑫, 李亚丁, 周红飞, 和源. 2019. 四川盆地长宁地区震旦系—寒武系裂陷槽的发现及其地质意义. *天然气勘探与开发*, 42(2): 22–28. [Qiu Y C, Luo B, Xia M L, Chen Y L, Dai X, Li Y D, Zhou H F, He Y. 2019. Discovery and geological significance of Sinian-Cambrian rift trough, Changning area, Sichuan Basin. *Natural Gas Exploration and Development*, 42(2): 22–28]
- 湛丽, 王才志, 宁从前, 刘英明, 王浩. 2023. 基于机器学习的鄂尔多斯盆地陇东地区长 7 段岩相测井识别方法. *油气藏评价与开发*, 13(4): 525–536. [Shen L, Wang C Z, Ning C Q, Liu Y M, Wang H. 2023. Well-log lithofacies classification based on machine learning for Chang-7 member in Longdong area of Ordos Basin. *Petroleum Reservoir Evaluation and Development*, 13(4): 525–536]
- 谭勇, 陈记, 杨忠民, 张化进. 2024. 基于 CatBoost 集成学习的边坡稳定性预测方法. *科学技术与工程*, 24(30): 13153–13160. [Tan Y, Chen J, Yang Z M, Zhang H J. 2024. Slope stability prediction method based on CatBoost ensemble learning. *Science Technology and Engineering*, 24(30): 13153–13160]
- 万云强, 陈践发, 陈聪, 刘超, 马勇, 梁佳宝. 2024. 四川盆地涪陵气田白马向斜五峰组—龙马溪组页岩气成藏地质条件及成藏模式. *天然气地球科学*, 35(9): 1656–1670. [Wan Y Q, Chen J F, Chen C, Liu C, Ma Y, Liang J B. 2024. Geological conditions and accumulation patterns of shale gas accumulation in Wufeng-Longmaxi Formations of Baima Syncline, Fuling Gas Field, Sichuan Basin. *Natural Gas Geoscience*, 35(9): 1656–1670]
- 王岚, 李文厚, 刘群, 王大兴, 张盟勃, 白斌. 2023. 鄂尔多斯盆地延长组长 7 黑色页岩岩相分类与沉积环境恢复. *古地理学报*, 25(3): 598–613. [Wang L, Li W H, Liu Q, Wang D X, Zhang M B, Bai B. 2023. Lithofacies characteristics and sedimentary environment of Chang 7 black shale in the Yanchang Formation, Ordos Basin. *Journal of Palaeogeography (Chinese Edition)*, 25(3): 598–613]
- 王民, 杨金路, 王鑫, 李进步, 徐亮, 言语. 2023. 基于随机森林算法的泥页岩岩相测井识别. *地球科学*, 48(1): 130–142. [Wang M, Yang J L, Wang X, Li J B, Xu L, Yan Y. 2023. Identification of shale lithofacies by well logs based on random forest algorithm. *Earth Science*, 48(1): 130–142]
- 王心乾, 余文端, 马晓东, 周韬, 邵浩, 崔钦宇, 邓空, 陆永潮, 刘占红. 2024. 基于常规测井曲线的页岩岩相识别与应用: 以苏北盆地溱潼凹陷阜宁组二段为例. *油气藏评价与开发*, 14(5): 699–706, 713. [Wang X Q, Yu W D, Ma X D, Zhou T, Tai H, Cui Q Y, Deng K, Lu Y C, Liu Z H. 2024. Identification and application of shale lithofacies based on conventional logging curves: a case study of the second member of Funing Formation in Qintong Sag, Subei Basin. *Petroleum Reservoir Evaluation and Development*, 14(5): 699–706, 713]
- 武赛军, 魏国齐, 杨威, 谢武仁, 曾富英. 2016. 四川盆地桐湾运动及其油气地质意义. *天然气地球科学*, 27(1): 60–70. [Wu S J, Wei G Q, Yang W, Xie W R, Zeng F Y. 2016. Tongwan movement and its geologic significances in Sichuan basin. *Natural Gas Geoscience*, 27(1): 60–70]
- 肖贤明, 王茂林, 魏强, 田辉, 潘磊, 李腾飞. 2015. 中国南方下古生界页岩气远景区评价. *天然气地球科学*, 26(8): 1433–1445. [Xiao X M, Wang M L, Wei Q, Tian H, Pan L, Li T F. 2015. Evaluation of lower Paleozoic shale with shale gas prospect in South China. *Natural Gas Geoscience*, 26(8): 1433–1445]
- 薛晓辉, 胡琳, 张郁云. 2016. 滇东下寒武统筇竹寺页岩气储层特征. *中国煤炭地质*, 28(3): 35–39, 44. [Xue X H, Hu L, Zhang Y Y. 2016. Shale gas reservoir features in lower Cambrian qiongzhusi formation, eastern Yunnan. *Coal Geology of China*, 28(3): 35–39, 44]
- 闫佳飞, 李胜利, 魏泽德, 吴忠宝, 陈建阳. 2025. 基于 XGBoost 算法的页岩岩相测井预测方法. *古地理学报*, 27(3): 763–776. [Yan J F, Li S L, Wei Z D, Wu Z B, Chen J Y. 2025. Shale lithofacies prediction method with well-logging data based on XGBoost algorithm. *Journal of Palaeogeography (Chinese Edition)*, 27(3): 763–776]
- 闫建平, 罗静超, 石学文, 钟光海, 郑马嘉, 黄毅, 唐洪明, 胡钦红. 2022. 川南泸州地区奥陶系五峰组—志留系龙马溪组页岩裂缝发育模式及意义. *岩性油气藏*, 34(6): 60–71. [Yan J P, Luo J C, Shi X W, Zhong G H, Zheng M J, Huang Y, Tang H M, Hu Q H. 2022. Fracture development models and significance of Ordovician Wufeng-Silurian Longmaxi shale in Luzhou area, southern Sichuan Basin. *Lithologic Reservoirs*, 34(6): 60–71]
- 杨洋, 石万忠, 张晓明, 王任, 徐笑丰, 刘俞佐, 白卢恒, 曹沈厅, 冯芊. 2021. 页岩岩相的测井曲线识别方法: 以焦石坝地区五峰组—龙马溪组为例. *岩性油气藏*, 33(2): 135–146. [Yang Y, Shi W Z, Zhang X M, Wang R, Xu X F, Liu Y Z, Bai L H, Cao S T, Feng Q. 2021. Identification method of shale lithofacies by logging curves: a case study from Wufeng-Longmaxi Formation in Jiaoshiba area, SW China. *Lithologic Reservoirs*, 33(2): 135–146]
- 雍锐, 吴建发, 吴伟, 杨雨然, 徐亮, 罗超, 刘佳, 何一凡, 钟可塑, 李彦佑, 朱逸青, 陈丽清. 2024. 四川盆地寒武系筇竹寺组页岩气勘探发现及其意义. *石油学报*, 45(9): 1309–1323. [Yong R, Wu J F, Wu W, Yang Y R, Xu L, Luo C, Liu J, He Y F, Zhong K S, Li Y Y, Zhu Y Q, Chen L Q. 2024. Exploration discovery of shale gas in the Cambrian Qiongzhusi Formation of Sichuan Basin and its significance. *Acta Petrolei Sinica*, 45(9): 1309–1323]
- 袁士博, 赵耀东, 李鹏, 荆泽寰. 2023. 基于多尺度一维卷积神经网络

- 的雷达辐射源脉内调制类型分类方法. 电子信息对抗技术, 38(5): 21–28. [Yuan S B, Zhao Y D, Li P, Jing Z H. 2023. Radar emitter intra-pulse modulation classification based on multi-scale one-dimensional convolutional neural network. *Electronic Information Warfare Technology*, 38(5): 21–28]
- 赵建华, 金之钧, 林畅松, 刘光祥, 刘可禹, 刘忠宝, 张钰莹. 2019. 上扬子地区下寒武统筇竹寺组页岩沉积环境. 石油与天然气地质, 40(4): 701–715. [Zhao J H, Jin Z J, Lin C S, Liu G X, Liu K Y, Liu Z B, Zhang Y Y. 2019. Sedimentary environment of the lower Cambrian qiongzhusi formation shale in the Upper Yangtze Region. *Oil & Gas Geology*, 40(4): 701–715]
- 赵圣贤, 夏自强, 刘文平, 张成林, 张鉴, 文莉, 何沅翰, 王重善. 2022. 四川盆地南部泸 203 井区五峰组: 龙马溪组页岩裂缝特征及形成演化. 油气地质与采收率, 29(5): 28–38. [Zhao S X, Xia Z Q, Liu W P, Zhang C L, Zhang J, Wen L, He Y H, Wang C S. 2022. Fracture characteristics and evolution of Wufeng–Longmaxi Formation shale in Lu203 well area in southern Sichuan Basin. *Petroleum Geology and Recovery Efficiency*, 29(5): 28–38]
- 张少龙, 闫建平, 石学文, 钟光海, 郑马嘉, 黄毅, 蔡进功, 唐洪明, 胡钦红, 梁强. 2022. 深层页岩气甜点分类的地质—工程评价指标体系及应用: 以四川盆地 LZ 地区五峰组—龙马溪组为例. 中南大学学报(自然科学版), 53(9): 3666–3680. [Zhang S L, Yan J P, Shi X W, Zhong G H, Zheng M J, Huang Y, Cai J G, Tang H M, Hu Q H, Liang Q. 2022. Geological and engineering evaluation index system for deep shale gas sweet spots classification and its application: a case of Wufeng: Longmaxi Formations in LZ area, Sichuan Basin. *Journal of Central South University (Science and Technology)*, 53(9): 3666–3680]
- 邹才能, 杨智, 张国生, 朱如凯, 陶士振, 袁选俊, 侯连华, 董大忠, 郭秋麟, 宋岩, 冉启全, 邱振, 吴松涛, 马锋, 白斌, 王岚, 熊波, 潘松圻, 刘翰林, 王小妮. 2023. 非常规油气地质学理论技术及实践. 地球科学, 48(6): 2376–2397. [Zou C N, Yang Z, Zhang G S, Zhu R K, Tao S Z, Yuan X J, Hou L H, Dong D Z, Guo Q L, Song Y, Ran Q Q, Qiu Z, Wu S T, Ma F, Bai B, Wang L, Xiong B, Pan S Q, Liu H L, Wang X N. 2023. Theory, technology and practice of unconventional petroleum geology. *Earth Science*, 48(6): 2376–2397]
- Hall B. 2016. Facies classification using machine learning. *The Leading Edge*, 35(10): 906–909.
- Nishitsuji Y, Exley R. 2019. Elastic impedance based facies classification using support vector machine and deep learning. *Geophysical Prospecting*, 67: 1040–1054.
- Saito T, Rehmsmeier M. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, 10(3): e0118432.
- Xie L, Yang X F, Zhang Y H, Xu Y, Zeng D M, Tang R F, Du Y, Wang X Z. 2024. Sedimentary characteristics and lithofacies paleogeography of the Cambrian in Sichuan basin, Southwest China. *Petroleum*, 10(2): 224–242.
- Yan J P, He X, Hu Q H, Liang Q, Tang H M, Feng C Z, Geng B. 2018. Lower Es3 in Zhanhua Sag, Jiyang depression: a case study for lithofacies classification in lacustrine mud shale. *Applied Geophysics*, 15(2): 151–164.

(责任编辑 李新坡)