

基于小样本机器学习的低渗透砂岩储集层 孔隙结构测井表征方法*

孙玉玺¹ 齐 媛¹ 陈 亮² 贺燚平³
季汉成¹ 史燕青¹ 赵博缘¹ 刘朱睿鹭⁴

1 中国石油大学(北京)地球科学学院, 北京 102249

2 中国石油大学(北京)理学院, 北京 102249

3 中国石油长庆油田公司第五采油厂, 陕西西安 710200

4 中国石油长庆油田公司油气工艺研究院, 陕西西安 710018

摘 要 近年来, 机器学习在基于测井数据的储集层参数评价中展现出显著优势, 但受限于取心实验样本数量, 模型的泛化能力往往无法保障。研究在鄂尔多斯盆地姬塬地区三叠系延长组收集了 100 余组高压压汞实验数据和 7 种常规测井数据, 采用 TabPFN 和其他常见 7 种机器学习方法, 系统地比较在小样本环境下不同方法的性能表现, 并基于 SHAP 可解释性分析理解模型决策机制。研究表明: (1) 研究区延长组砂岩储集层平均渗透率为 $0.39 \times 10^{-3} \mu\text{m}^2$, 为典型的低渗—超低渗储集层, 具有孔喉半径小、孔隙结构复杂等特点。第一主成分能解释孔隙结构参数 65.00% 的总方差, 通过生产数据、机理分析等认为其可作为表征孔隙结构的目标变量。(2) TabPFN 模型在小样本环境中表现优异, 验证集和盲井测试集中 R^2 分别为 0.81 和 0.87。由于其无需调参即可实现较高的性能, 展现出更为突出的适用性。(3) 特征重要性排名中, 密度测井、沉积相、声波测井、地质分层以及深电阻率为前 5 大特征。其中, 沉积相与地质分层在极端孔隙结构区间贡献更为突出, 提示离散地质变量对提升模型性能的关键作用。研究为在小样本环境下的低渗透砂岩储集层孔隙结构表征提供了有效的数据驱动案例。

关键词 TabPFN 机器学习 孔隙结构 低渗透砂岩 致密砂岩 延长组 鄂尔多斯盆地

第一作者简介 孙玉玺, 男, 1997 年生, 中国石油大学(北京)博士研究生, 主要从事储层地质学研究。
E-mail: 2860485642@qq.com。

通讯作者简介 季汉成, 男, 1966 年生, 中国石油大学(北京)教授, 博士生导师, 主要从事沉积学、储层地质学等研究。E-mail: jhch@cup.edu.cn。

中图分类号: TE122 **文献标志码**: A

Small-data-driven machine learning for well logging characterization of pore structure in low-permeability sandstone reservoirs

SUN Yuxi¹ QI Yuan¹ CHEN Liang² HE Yiping³
JI Hancheng¹ SHI Yanqing¹ ZHAO Boyuan¹ LIUZHU Ruizhi⁴

* 中国石油大学(北京)科研启动基金(编号: 2462023BJR011)和中国石油长庆油田分公司科研项目(编号: 2023DJ0911)联合资助。[Co-funded by the Science Foundation of China University of Petroleum (Beijing) (No. 2462023BJR011) and Research Project of PetroChina Changqing Oilfield Company (No. 2023DJ0911)]

收稿日期: 2025-08-03 改回日期: 2025-10-21

1 College of Geosciences, China University of Petroleum (Beijing), Beijing 102249, China

2 College of Science, China University of Petroleum (Beijing), Beijing 102249, China

3 The Fifth Oil Production Plant, PetroChina Changqing Oilfield Company, Xi'an 710200, China

4 Oil and Gas Technology Research Institute, PetroChina Changqing Oilfield Company, Xi'an 710018, China

Abstract In recent years, machine learning has shown significant advantages in reservoir parameter evaluation based on well logs. However, due to the limited size of core samples, the generalization ability of these models is often not guaranteed. In this study, more than 100 sets of high-pressure mercury injection data and seven types of well logs were collected from the Triassic Yanchang Formation in the Jiyuan area of the Ordos Basin. The study systematically compares the performance of TabPFN and seven other common machine learning methods under small data conditions. The SHAP interpretability analysis is also used to understand the model's decision-making mechanisms. The results show: (1) The average permeability of the Yanchang Formation sandstone reservoir in the study area is $0.39 \times 10^{-3} \mu\text{m}^2$, which is typical of low- to ultra-low permeability reservoirs. These reservoirs have small pore throat radii and complex pore structures. The first principal component explains 65.00% of the total variance of the pore structure parameters. Combined with production data and mechanism analysis, it is considered an appropriate target variable for characterizing pore structure. (2) The TabPFN model performs excellently in small-data environments, achieving R^2 values of 0.81 and 0.87 in the validation and blind well test sets, respectively. Its high performance without parameter tuning demonstrates its outstanding applicability. (3) According to feature importance ranking, density log, sedimentary facies, acoustic log, geological layering, and deep resistivity are the top five features. Sedimentary facies and geological layering make greater contributions in extreme pore structure intervals, highlighting the critical role of discrete geological variables in improving model performance. This study provides an effective data-driven case for characterizing the pore structure of low-permeability sandstone reservoirs under small data conditions.

Key words TabPFN, machine learning, pore structure, low-permeability sandstone, tight sandstone, Yanchang Formation, Ordos Basin

About the first author SUN Yuxi, born in 1997, is a Ph.D. candidate at China University of Petroleum (Beijing). He is mainly engaged in reservoir geology. E-mail: 2860485642@qq.com.

About the corresponding author JI Hancheng, born in 1966, is a professor and Ph.D. supervisor at China University of Petroleum (Beijing). He is mainly engaged in sedimentology and reservoir geology. E-mail: jhch@cup.edu.cn.

低渗透砂岩储集层作为非常规油气资源开发及其他地下能源应用的重要目标,广泛分布于陆相盆地、深水浊积体系以及深层或超深层地层(Zou *et al.*, 2018; Feng *et al.*, 2019; 吕奇奇等, 2023; 王鑫等, 2023)。这类储集层通常具有孔喉类型多样、连通性较差以及非均质性强等特征,普遍发育微米—纳米级孔喉半径(Lai *et al.*, 2018a, 2018b; Xu *et al.*, 2025)。这种复杂的孔隙结构不仅显著影响油气的储集与渗流能力,也为储集层的精细评价与表征带来了极大挑战(Pittman, 1992; Lai *et al.*, 2018b)。

目前,主流的岩石孔隙结构表征方法主要依赖实验分析和测井数据。实验手段包括铸体薄片、扫描电镜(SEM)、高压压汞(HPMI)、恒速压汞(MICP)、核磁共振(NMR)和CT扫描等(Lai *et al.*, 2018b, 2022; Xiao *et al.*, 2018)。核磁共振测井能够通过测量氢核横向弛豫时间(T_2)分布,直接量化孔隙度、孔喉分布及流体分布等信息(张宪国等, 2019; 赖锦等, 2023; Lai *et al.*, 2024b)。而常规测井则多借助机器学习技术,主要包括2大类:(1)通过非监督机器学习明确测井数据的离散型电性聚类,后基于实验获取孔隙结构参数为其赋

予岩石物理意义 (Maldar *et al.*, 2022; Iraj *et al.*, 2023; Yang *et al.*, 2025); (2) 利用监督型机器学习方法开展水力流动指数 (FZI)、储集层质量指数 (RQI) 等参数的预测 (Zhang *et al.*, 2021; Cao *et al.*, 2025; Wang and Hou, 2025)。然而, 上述方法均极度依赖获取的实验样本数量, 当样本较少时难以保障统计及建模过程的可靠性。

近年来, 小样本机器学习已在计算机视觉、自然语言处理、医学等学科领域得到了广泛应用 (He *et al.*, 2022; Ma *et al.*, 2024; Li *et al.*, 2025a)。与传统依赖大数据驱动的机器学习方法不同, 小样本学习关注如何在数据稀缺或获取成本高昂的情况下, 依然实现较高的模型性能。其核心思想是利用数据增强或先验知识等策略, 从有限的训练样本中挖掘尽可能多的信息 (赵凯琳等, 2021)。例如, Moreno-Barea 等 (2020) 基于深度生成模型显著提升了小样本数据集上的分类精度。Hollmann 等 (2025) 提出的 TabPFN 模型, 则通过 Transformer 架构在表格数据上的先验训练, 实现了小样本条件下的高效预测。相比之下, 小样本学习在油气储集层评价等领域的应用尚处于起步阶段。因此, 将小样本学习方法引入低渗透砂岩储集层孔隙结构表征, 有望减少对大规模取心实验的依赖, 为非常规储集层高效评价提供思路。

综上, 研究通过高压压汞、测井数据、主成分分析等方法手段, 系统对比 TabPFN 和其他 7 种常见机器学习方法在小样本条件下对孔隙结构表征的适用性。并结合特征重要性排列及 Mean | SHAP | 堆叠图, 明确特征变量在不同孔隙结构区间的贡献权重, 以期低渗透砂岩储集层评价提供数据驱动决策。

1 数据与方法

1.1 数据来源与预处理

研究所用数据收集自鄂尔多斯盆地中西部姬塬地区三叠系延长组砂岩储集层。具体包括作为输入变量的地质分层、沉积相和 7 种测井数据, 以及用于孔隙结构分析的 112 组高压压汞实验数据。所选取的测井数据涵盖了声波测井、中子测井、密度测井、井径测井、伽马测井、自然电位测井和深电阻率测井。上述资料均由中国石油天然气股份有限公

司长庆油田分公司第五采油厂提供。此外, 沉积相信息参考了朱石磊 (2019) 针对研究区沉积体系开展的研究成果, 包括河流相、三角洲平原亚相、三角洲前缘亚相以及深湖相。

考虑到深电阻率、渗透率、排驱压力及中值压力等参数通常呈现对数分布的特性, 研究对相关参数进行了以 10 为底的对数 (lg) 转换, 以减少数据偏态对建模及统计分析的影响。在此基础上, 对输入特征进行了标准化处理, 以消除量纲差异和数值范围对模型训练的影响 (唐佰强等, 2025)。模型训练与评估执行了按井分组的五折交叉验证, 以避免同井样本造成的信息泄露。同时, 为客观检验模型的泛化能力, 选取了 1 口盲井作为独立测试集。

1.2 压汞参数降维

高压压汞实验能够提供丰富的孔隙结构参数信息, 但单一参数难以全面、准确地表征储集层孔隙结构的复杂性。为此, 研究采用主成分分析 (PCA), 对多维度的孔隙结构指标进行降维处理。该方法通过线性变换, 将原始多维数据投影到相互正交的主成分坐标系中 (刘爱疆等, 2013)。以第一主成分 (PC1) 最大程度地保留原始数据的方差信息, 从而有效提取具有代表性的孔隙结构特征。为避免数据泄露, 盲井测试集未参与 PCA 拟合, 仅使用相同变换矩阵进行投影。

1.3 模型构建与性能评估

研究共选用了 8 种机器学习方法, 包括 TabPFN、深度神经网络 (DNN)、K 最近邻 (KNN)、轻量梯度提升 (LightGBM)、多元线性回归 (MR)、随机森林 (RF)、支持向量机 (SVM) 和极端梯度提升 (XGBoost) (Coker, 1995; Breiman, 2001; Noble, 2006; Lecun *et al.*, 2015; Chen and Guestrin, 2016; Ke *et al.*, 2017; Taunk *et al.*, 2019; Hollmann *et al.*, 2025)。其中, TabPFN 是一种基于 Transformer 架构的机器学习模型, 专为小样本表格数据预测任务设计。TabPFN 通过在大量合成数据上预训练, 避免了传统机器学习方法繁琐的参数调优过程, 在小样本环境下表现出更优的预测精度与泛化能力。考虑到篇幅限制, 有关机器学习方法具体原理和架构可参考上述相关文献。

为了确保性能对比的公平性, 所有模型均采用了 i7-13700KF 的全部核心训练。除 TabPFN 和 MR

外,其余模型采用 Optuna 框架进行了 50 次实验的超参数优化,以确定最佳的模型参数 (Mao *et al.*, 2025)。为全面衡量模型的预测精度、稳定性及拟合效果,采用了决定系数 (R^2) 与均方根误差 (RMSE) 作为评价指标。

1.4 可解释性分析

机器学习模型通常被视为难以直接解释的“黑箱”,其决策机制和特征贡献程度不易直观揭示 (闵超等, 2024)。为深入理解模型内部的决策逻辑,研究采用了基于合作博弈论的 SHapley 加性解释 (SHAP) 方法。SHAP 通过计算每个特征在所有可能的特征组合中所产生的平均边际贡献,量化回归任务中各输入变量的重要性 (Lundberg and Lee, 2017; 周锦涛等, 2025)。此外,进一步定义了特征的相对贡献度 ($|SHAP|_{RC}$),即单个特征的 $|SHAP|$ 值占所有特征 $|SHAP|$ 值之和的百分比,以更清晰且量化的方式描述特征对模型预测结果的影响程度 (Zhang *et al.*, 2025)。其中,绝对贡献反映各特征对预测结果的总体影响强度,相对贡献则揭示各特征在不同孔隙结构区间内的重要性占比变化。

2 岩石物理特征

2.1 孔隙结构

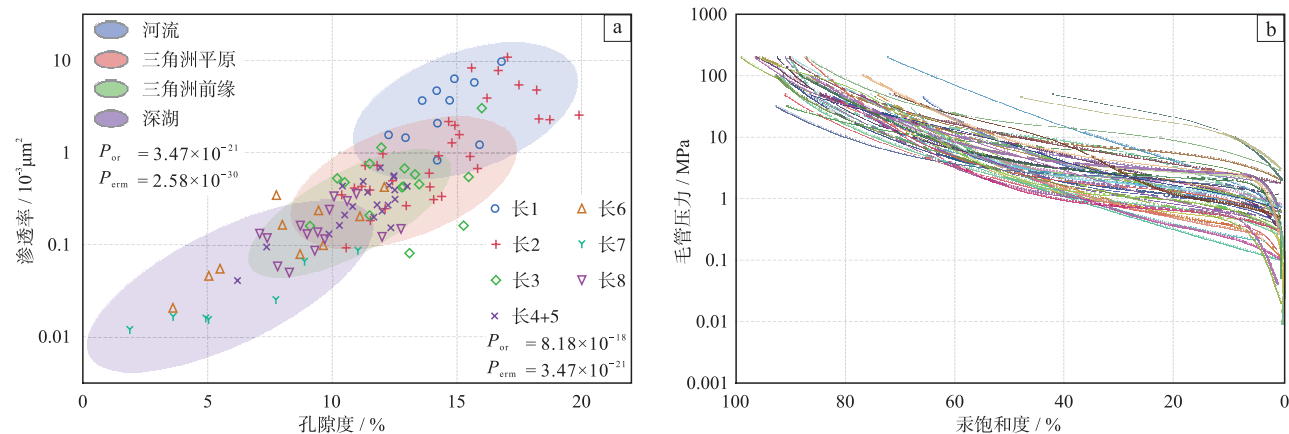
姬塬地区延长组砂岩储集层孔隙度主要分布范围在 9.94%~14.16% 之间,均值为 11.79%。渗透率集中于 $0.15 \times 10^{-3} \sim 0.92 \times 10^{-3} \mu\text{m}^2$ 之间,均值为

$0.39 \times 10^{-3} \mu\text{m}^2$ 。孔隙度与渗透率间呈显著正相关性 ($R^2=0.722$),表现为中孔、低渗—特低渗储集层特征 (图 1-a)。统计检验结果显示,不同沉积相及不同地质层位下孔隙度 (P_{or}) 和渗透率 (P_{erm}) 的 P 值均远小于 0.05,指示具有显著差异。

高压压汞实验结果显示,研究区平均孔喉半径 (D_M) 介于 $8.88 \sim 15.17 \mu\text{m}$ 之间,均值为 $11.90 \mu\text{m}$ 。最大汞饱和度 (S_{Hg}) 介于 19.80%~99.19% 之间,均值为 80.12%。排驱压力 (P_d) 主要分布在 $0.38 \sim 1.66 \text{ MPa}$ 之间,均值为 0.88 MPa 。毛细管中值压力 (P_{50}) 集中在 $2.81 \sim 9.55 \text{ MPa}$,平均为 5.23 MPa 。另有少量样品由于 S_{Hg} 未达 50%,导致 P_{50} 为空值。整体来看,姬塬地区延长组毛细管压力曲线呈现平缓的下凹型特征 (图 1-b),以细—微孔喉为主,具有排驱压力较低、中值孔喉半径较小、孔喉结构复杂且分选性较差等特征。

2.2 测井响应

相关系数聚类热图表明 (图 2),不同测井曲线对储集层孔隙结构的响应敏感程度存在明显差异。其中,声波测井 (AC) 与中子测井 (CNL) 在聚类中高度相似,与 P_{or} 和 P_{erm} 呈显著正相关,同时与 P_{50} 、 D_M 和 P_d 表现为中等强度的负相关关系,表明二者能够较好地反映储集层孔隙空间的发育程度。密度测井 (DEN) 对储集层孔隙结构表现出最高的敏感性,其与 P_{or} 和 P_{erm} 均呈现显著的负相关性,表明储集层密度降低通常对应孔喉半径较大、连通性较好的储集层结构特征,即更优质的储集层品质。深电阻率测井 (RILD) 与自然电位测



a—孔隙度与渗透率关系; b—毛细管压力曲线

图 1 鄂尔多斯盆地姬塬地区延长组低渗透砂岩物性特征

Fig. 1 Petrophysical characteristics of low-permeability sandstones in the Yanchang Formation, Jiyuan area, Ordos Basin

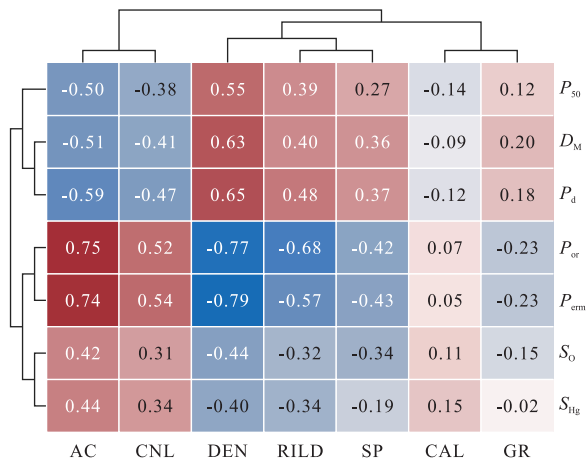


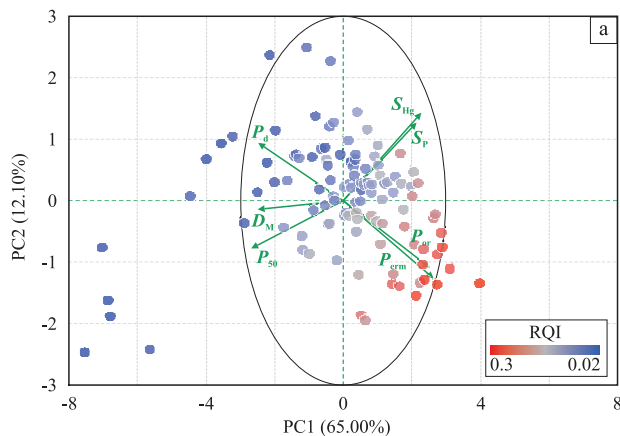
图 2 高压压汞揭示的孔隙结构参数与测井数据相关性聚类热图

Fig. 2 Clustered heatmap of the correlations between pore structure parameters revealed by HPMI and well logs

井 (SP) 整体表现出与 DEN 类似但稍弱的趋势, 能够间接指示储集层孔隙结构的变化情况。伽马测井 (GR) 整体上对孔隙结构参数的整体响应较弱, 仅与 P_{or} 和 P_{erm} 呈现微弱的相关性, 反映出泥质含量对储集层孔隙结构的控制作用。而井径测井 (CAL) 则仅对 S_{fg} 与分选系数 (S_o) 呈微弱响应。总体来看, 通过测井数据能较有效地揭示储集层孔隙结构特征, 尤其是三孔隙测井 (DEN、AC 和 CNL) 对储集层孔隙结构的表征具有显著优势。

2.3 主成分分析

主成分分析结果表明 (图 3), 第一主成分 (PC1) 可解释 65.00% 的总方差, 高度聚合了 P_{50} 、

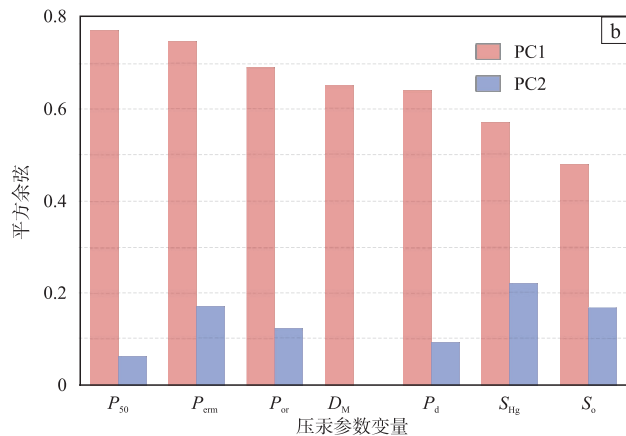


P_{erm} 、 P_{or} 和 D_M 等关键孔隙结构参数。同时, 储集层质量指数 (RQI) 沿 PC1 呈明显的正向递增趋势, 指示 PC1 能够良好的表征储集层孔隙结构。第二主成分 (PC2) 解释了 12.10% 的方差, 主要反映 S_{fg} 与 S_o 所代表的孔喉分布及连通性差异。值得注意的是, S_{fg} 与 S_o 在 PC1 上的载荷 (分别为 0.57 和 0.48) 仍高于其在 PC2 上的载荷 (分别为 0.22 和 0.17), 说明即使针对饱和度及分选性特征, PC1 的解释能力依旧更为全面。因此, PC1 更完整地体现了不同压汞参数在孔喉尺寸、孔隙连通性及孔隙形态方面的复合作用, 被确定为后续机器学习建模过程的输出变量。

3 多种机器学习模型性能比较

3.1 性能指标

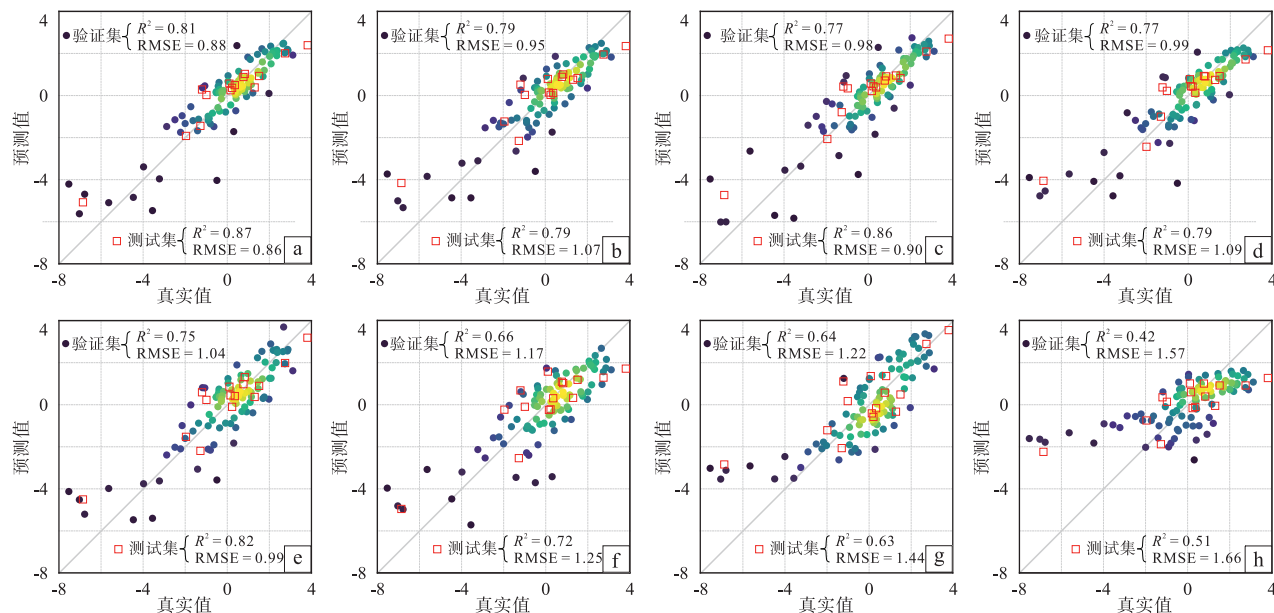
从模型性能指标来看, 基于 Transformer 架构的 TabPFN 模型表现最为突出 (验证集 R^2 为 0.81, RMSE 为 0.88; 盲井测试集 R^2 为 0.87, RMSE 为 0.86), 其预测散点最为紧凑且清晰地沿对角线分布 (图 4-a), 体现了该方法出色的预测稳健性和优异的泛化能力。3 种决策树集成模型 (RF、XGBoost 和 LightGBM) 的预测散点较为紧凑 (图 4-b, 4-c, 4-d), 验证集的 R^2 为 0.77~0.79, RMSE 为 0.95~0.99。DNN 在验证集上的表现虽稍逊于决策树模型, 但在盲井测试集中预测效果显著提高, R^2 为 0.82, RMSE 为 0.99 (图 4-e)。相比之下, SVM、MR 和 KNN 3 种模型则多表现为欠拟合, 预测散点较为离散且偏离对角线 (图 4-f, 4-g,



a—孔隙结构参数在主成分空间的投影; b—孔隙结构参数主成分方差贡献图

图 3 基于高压压汞实验的孔隙结构参数主成分分析

Fig. 3 PCA of pore structure parameters revealed by HPMI



a—TabPFN 模型；b—RF 模型；c—XGBoost 模型；d—LightGBM 模型；e—DNN 模型；f—SVM 模型；g—MR 模型；h—KNN 模型

图 4 不同机器学习方法在 PC1 预测中的表现

Fig. 4 Comparative performance of various machine learning models in predicting the PC1

4-h)。其中 KNN 的预测散点甚至呈现明显的模糊发散特征，在验证集的 R^2 低至 0.42。

3.2 泰勒图评价

泰勒图通过将观测数据设为参考基准点，不同模型的预测结果以散点形式显示，可以直观展示模型输出与观测数据之间的标准差、相关系数及均方根误差 (Lai *et al.*, 2024a)。结果显示(图 5)，TabPFN 和 XGBoost 在验证集和测试集始终落在 $RMSE=1$ 的粉色内圈中，相关系数维持在 0.9~0.95，与观测数据标准差最为接近，表现出最低的误差和最强泛化。RF、LightGBM 和 DNN 投影则始终保持在 $RMSE=1$ 附近，相关系数略有降低，属于第二梯队。而 SVM、MR 和 KNN 则基本偏离 $RMSE=1$ ，相关系数处于 0.7~0.85 之间，标准差则远低于观测值，属于第三梯队。

3.3 寻优及训练时长

除预测精度外，模型构建过程中的训练与超参数寻优时间也是评估模型综合性能的重要指标之一。结果表明(表 1)，决策树集成模型 (RF、XGBoost 和 LightGBM) 及 DNN 模型相对复杂，涉及超参数较多，整体寻优耗时较长，其中 XGBoost 和 RF 的寻优时间分别达到 19.57 s 与 18.37 s。相比之下，SVM 和 KNN 涉及超参数较少，训练及寻优

时间明显更短。而 TabPFN 模型尽管单次训练时间略长 (0.11 s)，但凭借其无需调参的特性，整体运行耗时最短。

3.4 预测曲线

图 6 中展示了 3 种表现较优模型 (TabPFN、RF、XGBoost) 在盲井测试集中的预测曲线。整体来看，3 种模型曲线规律及形态相似，能够较好地反映出 PC1 的波动。其中，TabPFN 曲线最为平滑且与实测数据高度吻合，能够准确捕捉 PC1 的峰值变化。RF 与 TabPFN 的总体相似，但容易低估 PC1，高值连续条带被多次切断，表现出相对强的非均质性。而 XGBoost 预测曲线则表现出密集窄峰特征，表现出决策树模型的“过分割”倾向，易受测井数据噪声干扰。

4 讨论

4.1 主成分对孔隙刻画能力

这项研究中，通过主成分分析 (PCA) 对多维孔隙结构参数进行降维，将提取的 PC1 作为储集层孔隙结构的综合表征指标。这简化了输出指标且方便模型学习，但仍有 35.00% 的信息未被有效表征。为了验证 PC1 在表征孔隙结构的合理性，

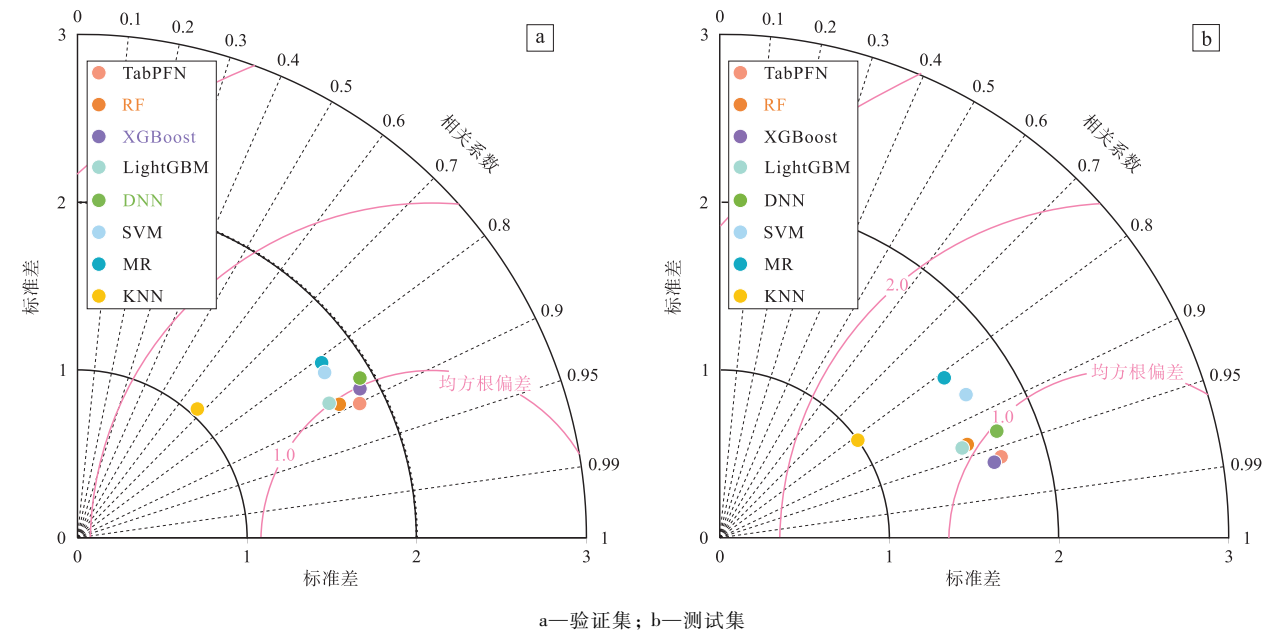


图 5 不同机器学习模型在泰勒图上的投影

Fig. 5 Taylor diagram showing the projections of various machine learning models

表 1 不同机器学习模型的寻优和训练耗时

Table 1 Model optimization and training durations for various machine learning models

模型	寻优耗时/s	训练耗时/s	主要参数	寻优范围
TabPFN	/	0.11	/	/
RF	18.37	0.08	n_estimators; max_depth; max_features; min_samples_split; bootstrap	100 ~ 600; 3 ~ 20; ['sqrt', 'log2', 0.5 ~ 1.0]; 2 ~ 20; [True, False]
XGBoost	19.57	0.07	n_estimators; max_depth; learning_rate; subsample; colsample_bytree	50 ~ 500; 3 ~ 10; 1e-3 ~ 0.3; 0.5 ~ 1.0; 0.5 ~ 1.0
LightGBM	11.09	0.03	num_leaves; learning_rate; n_estimators; feature_fraction; bagging_fraction	16 ~ 128; 1e-3 ~ 0.3; 50 ~ 500; 0.5 ~ 1.0; 0.5 ~ 1.0
DNN	11.21	0.10	n_layers; hidden; alpha; lr_init; activation; solver	1 ~ 3; 16 ~ 256; 1e-5 ~ 1e-1; 1e-4 ~ 1e-1; ['relu', 'tanh']; ['adam', 'sgd']
SVM	4.58	0.00	C; epsilon; gamma; kernel	1e-3 ~ 1e3; 1e-4 ~ 1.0; ['scale', 'auto']; ['rbf', 'poly', 'sigmoid']
MR	/	0.00	/	/
KNN	8.30	0.00	n_neighbors; weights; p	1 ~ 20; ['uniform', 'distance']; 1 ~ 2

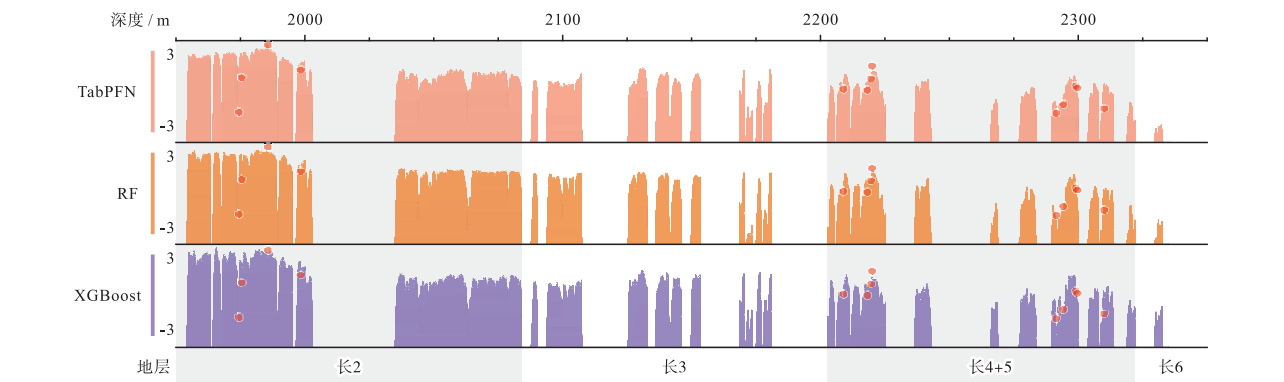


图 6 3 种表现较好的机器学习模型预测曲线

Fig. 6 Prediction curves of the three best-performing machine learning models

研究基于 TabPFN 模型对姬塬地区 P 油藏开展了孔隙结构解释, 并与钻井试油结果进行对比验证。结果表明(图 7-a), PC1 与钻井初期产能呈良好的正相关性。由于低渗透砂岩储集层存在较强的垂向与平面非均质性, 储集层孔隙结构通常直接影响油气的运聚过程 (Lai *et al.*, 2018b; Zhao *et al.*, 2024)。储集层孔隙结构较差或趋于致密时, 毛细管压力较高, 油气难以有效充注, 以束缚水为主 (Gao and Li, 2015; Li *et al.*, 2025b)。而孔隙结构较优时, 更有利于烃类的充注与储集。

进一步, 由于砂岩储集层孔隙结构通常受临近泥岩热演化影响。通过孔隙结构与临近岩性之间的联系用以评价 PC1 的合理性(图 7-b, 7-c)。结果表明, 砂岩厚度增加对应储集层孔隙结构的改善, 而泥岩较厚时相邻砂体则趋于致密。这一现象与已有研究结果基本一致 (Lu *et al.*, 2024), 泥岩在热演化过程中释放的成岩流体或促进了邻近薄层砂岩的胶结作用; 而厚层砂岩中较好的孔隙连通性可能更有利于酸性溶蚀流体的迁移。

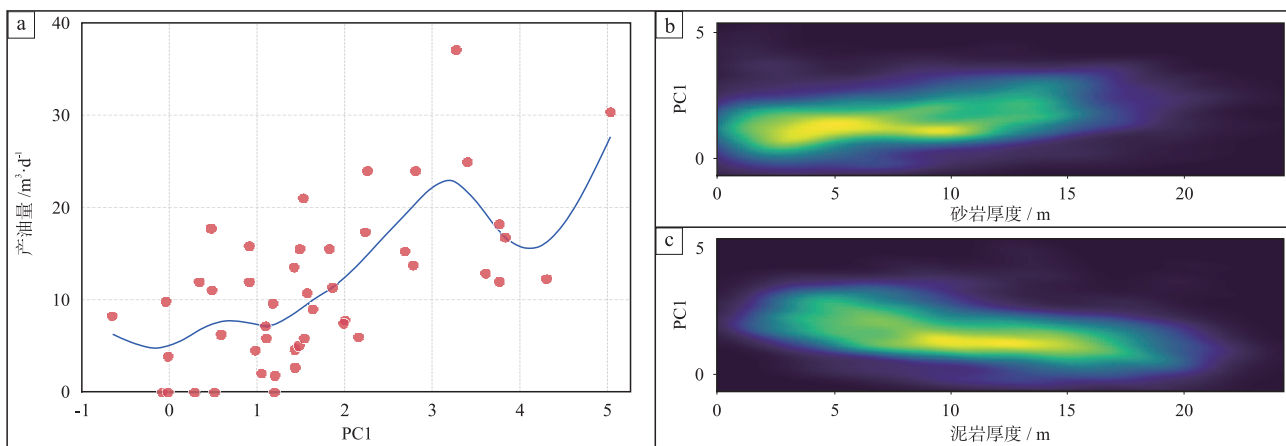
此外, 通过 PC1 和不同孔隙结构参数的散点矩阵分析(图 8)。结果表明, PC1 与孔隙结构之间整体呈现高度的线性相关性, 绝大部分样本落在 95% 置信椭圆内。但对于个别致密储集层样本, PC1 的表征能力较差。这部分样本表现为 $S_{Hg} < 50\%$, P_{50} 为空值的特殊孔隙结构, 这将一定程度上降低 PC1 能够解释的方差占比。综合而言, 通过实际生产数据验证、机理分析以及统计分析, 研究认为基于主成分分析提取的 PC1 能够有效地捕

捉并反映低渗透砂岩储集层孔隙结构参数的主要变化趋势。

4.2 特征变量对预测结果的影响

研究基于 SHAP 方法对 TabPFN 模型开展了可解释性分析, 以量化不同特征变量对储集层孔隙结构的贡献机制。特征重要性分析结果表明(图 9), 对模型预测贡献最大的前 5 个变量依次为 DEN、沉积相、AC、地层和 RILD。其中, DEN、AC 和 RILD 能够反映岩石的孔隙体积变化和孔喉连通性, 具有明确的物理解释基础。而沉积相和地层作为离散型地质变量也展现出显著的模型贡献。具体而言, 地质分层通常蕴含埋深、成岩背景等信息, 沉积相则反映了沉积环境、岩性组合等差异。对于研究区而言(图 1-a), 延长组二段埋藏较浅, 主要发育河流相—三角洲平原亚相, 经历了后期抬升和淋滤增孔作用, 孔隙结构整体较好 (尹泽等, 2019; 邓秀芹等, 2020)。而长 4+5 段和长 6 段则发育三角洲前缘亚相—深湖相细粒沉积环境, 破坏性成岩作用普遍, 储集层趋于致密化 (Ren *et al.*, 2019; Yang *et al.*, 2020)。离散型地质变量能够有效补充常规测井数据不蕴含的地质信息。

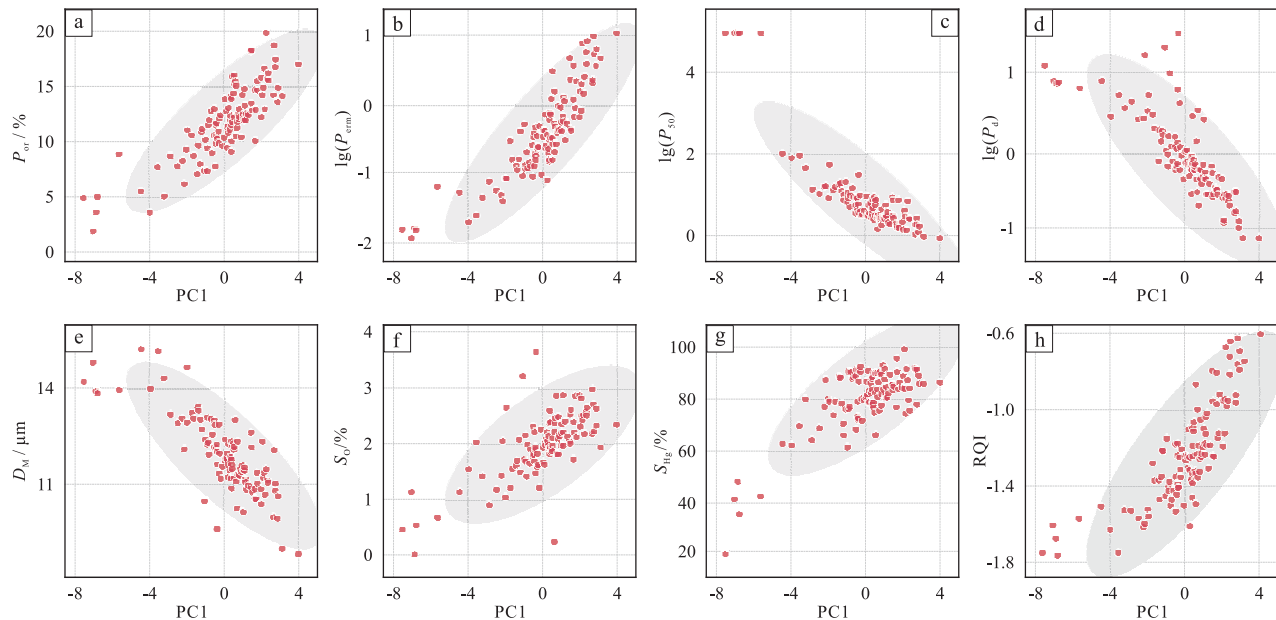
进一步, 研究通过 Mean | SHAP | 堆叠图展示了特征变量在不同孔隙结构区间的动态贡献 (图 10)。整体来看, 不同孔隙结构对特征变量的依赖呈现出非线性动态变化。当孔隙结构较差时 ($PC1 < -2$), 模型总体敏感度达到峰值, 其中沉积相的影响最为突出。进入中等孔隙结构区间时



a—投产初期产油量与 PC1 的相关性; b—砂岩厚度与 PC1 的相关性; c—泥岩厚度与 PC1 的相关性

图 7 PC1 与一些变量的相关性分析

Fig. 7 Correlation analysis between PC1 and selected variables



a—孔隙度；b—渗透率；c—中值压力；d—排驱压力；e—平均孔喉半径；f—分选系数；g—最大水饱和度；h—储集层质量指数

图 8 PC1 与高压压汞揭示的孔隙结构参数间的相关性分析

Fig. 8 Correlation analysis between PC1 and pore structure parameters revealed by HPMT

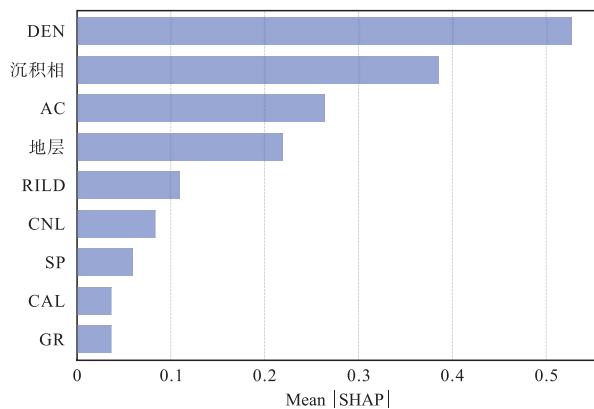


图 9 基于 SHAP 的特征重要性排列

Fig. 9 Feature importance ranking based on SHAP values

($-2 < PC1 < 2$), 总体 SHAP 值显著衰减, 测井数据的贡献权重回升, 沉积相则维持背景贡献。而孔隙结构较好时 ($PC1 > 2$), 沉积相贡献权重再次上升。这表明在极端孔隙结构条件下, 模型尤其依赖于离散型变量来进行决策。这也提示在小样本环境下基于测井数据进行储集层定量评价时, 合理引入地质分层、沉积环境或成岩背景等地质约束, 将有助于提高模型性能。

4.3 油气地质意义

低渗透砂岩储集层通常形成于低能沉积环境,

具有高泥质含量和原生孔隙有限等特征 (Lai *et al.*, 2018b)。而在成岩过程中, 机械压实作用进一步压缩并破坏原生孔隙, 硅质和碳酸盐等胶结物不断沉淀堵塞孔喉 (Yang *et al.*, 2020; Luo *et al.*, 2024)。这些因素使得测井数据与孔隙结构之间表现出复杂的非线性关系 (Zhao *et al.*, 2024)。SVM、MR 和 KNN 等简单机器学习模型在应对强非线性问题时能力有限, 容易出现欠拟合现象。而 TabPFN 具备更强的非线性建模能力, 能更好地适应低渗透砂岩复杂多变的孔隙结构特征。此外, 得益于“开箱即用”的特性, TabPFN 展现出了显著的计算效率。而随着数据集规模的扩大, 这种效率优势将更加突出。例如, 决策树集成模型在树数量增加或 DNN 网络结构复杂化时, 其训练和调参耗时会显著提升, 而 TabPFN 则能有效规避这些问题。

相较常见机器学习方法, TabPFN 擅于处理样本量 ≤ 10000 、特征数 ≤ 500 的中小型表格数据集 (Hollmann *et al.*, 2025), 这对于海上少井或老井资料稀疏等生产场景极为契合。无论是在量化孔隙度、渗透率或孔隙结构等回归任务, 还是对岩相、沉积相或储集层级别等离散标签分类, 亦或是对产量、含水率等时序数据的动态预测, TabPFN 均能够提供高效的数据驱动决策。此外, TabPFN 与 SHAP 可解释性框架的兼容, 能够帮助理解模型的

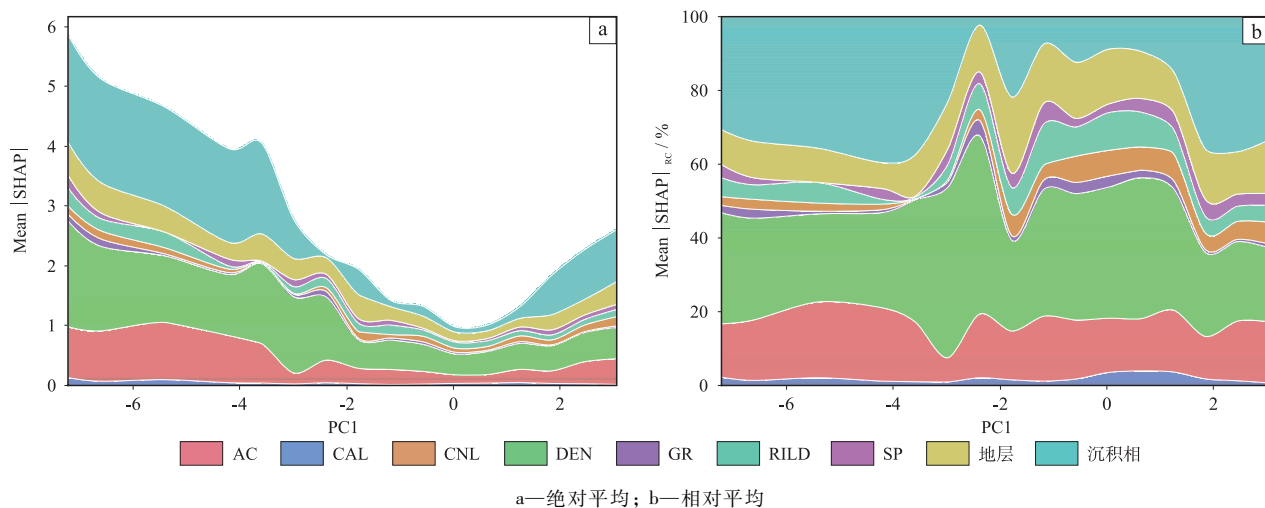


图 10 不同特征随 PC1 变化的平均 |SHAP| 贡献

Fig. 10 Mean |SHAP| contributions of different features with respect to variations in PC1

决策机制,对于烃源岩评价等需量化地球化学指标的研究具有重要意义。综合而言,TabPFN 能够在不依赖大规模取心实验的情况下,较好地兼顾预测精度、泛化能力与计算效率,展现出更加突出的适用优势。

5 结论

1) 通过对高压压汞揭示的孔隙结构参数进行主成分分析降维,所提取的第一主成分(PC1)能够解释 65.00% 的总方差,高度聚合了 P_{50} 、 P_{erm} 、 P_{or} 和 D_{M} 等关键孔隙结构参数。通过生产数据、机理分析以及统计分析认为,PC1 能够作为表征储集层孔隙结构的有效目标变量。

2) 在小样本环境下,TabPFN 相较其他机器学习方法表现较为突出。验证集和盲井测试集中, R^2 分别为 0.81 和 0.87, RMSE 分别为 0.88 和 0.86。预测曲线平滑且能够准确捕捉 PC1 的峰值变化。得益于其无需超参数寻优的特性,整体训练耗时仅 0.11 s,展现出卓越的泛化能力和计算效率。

3) SHAP 可解释分析中,沉积相与地质分层对模型的贡献分别位列第 2 位和第 4 位。二者的相对贡献权重随 PC1 的变化呈现非线性动态波动,在极端孔隙结构条件下,模型尤其依赖于离散型变量来进行决策。表明离散型地质变量可有效补充测井数据中不具备的地质信息,进一步提升模型预测性能。

致谢 中国石油天然气股份有限公司长庆油田分公司第五采油厂为本研究提供宝贵的数据支持,以及编辑和审稿人提出的宝贵意见和建议,在此一并致以衷心感谢。

参考文献 (References)

- 邓秀芹,程党性,周新平,时孜伟,郭懿萱. 2020. 鄂尔多斯盆地地下侏罗统地层水化学特征及成因分析. 沉积学报, 38(5): 1099–1110. [Deng X Q, Cheng D X, Zhou X P, Shi Z W, Guo Y X. 2020. Formation hydrochemical characteristics and genesis of the lower Jurassic, Ordos Basin. Acta Sedimentologica Sinica, 38(5): 1099–1110]
- 赖锦,李红斌,张梅,白梅梅,赵仪迪,范旗轩,庞小娇,王贵文. 2023. 非常规油气时代测井地质学研究进展. 古地理论, 25(5): 1118–1138. [Lai J, Li H B, Zhang M, Bai M M, Zhao Y D, Fan Q X, Pang X J, Wang G W. 2023. Advances in well logging geology in the era of unconventional hydrocarbon resources. Journal of Palaeogeography (Chinese Edition), 25(5): 1118–1138]
- 刘爱疆,左烈,李景景,李瑞,张玮. 2013. 主成分分析法在碳酸盐岩岩性识别中的应用: 以地区寒武系碳酸盐岩储层为例. 石油与天然气地质, 34(2): 192–196. [Liu A J, Zuo L, Li J J, Li R, Zhang W. 2013. Application of principal component analysis in carbonate lithology identification; a case study of the Cambrian carbonate reservoir in YH field. Oil & Gas Geology, 34(2): 192–196]
- 吕奇奇,辛红刚,王林,罗顺社,谈卫东,冯胜斌. 2023. 鄂尔多斯盆地宁县地区三叠系延长组 7 段湖盆细粒重力流沉积类型、特征及模式. 古地理论, 25(4): 823–840. [Lü Q Q, Xin H G, Wang L, Luo S S, Dan W D, Feng S B. 2023. Sedimentary types, characteristics and model of lacustrine fine-grained gravity flow in the Member 7 of Triassic Yanchang Formation in Ningxian area, Ordos Basin. Journal of Palaeogeography (Chinese Edition), 25(4): 823–840]

- 闵超,文国权,李小刚,赵大志,李昆成. 2024. 可解释机器学习在油气领域人工智能中的研究进展与应用展望. 天然气工业, 44(9): 114–126. [Min C, Wen G Q, Li X G, Zhao D Z, Li K C. 2024. Research progress and application prospect of interpretable machine learning in artificial intelligence in oil and gas industry. Natural Gas Industry, 44(9): 114–126]
- 唐佰强,孟庆涛,杨亮,胡菲,谭悦,邢济麟,刘招君,张恩威,董秦玮. 2025. 基于BSMOTE-SVM的细粒沉积岩岩相智能预测:以松辽盆地青山口组一段为例. 古地理学报, 27(4): 937–949. [Tang B Q, Meng Q T, Yang L, Hu F, Tan Y, Xing J L, Liu Z J, Zhang E W, Dong Q W. 2025. Intelligent prediction of fine-grained sedimentary lithofacies based on BSMOTE-SVM: a case study of the Member 1 of Qingshankou Formation in Songliao Basin. Journal of Palaeogeography (Chinese Edition), 27(4): 937–949]
- 王鑫,曾灏辉,贾昆昆,王伟庆,李博,安丛,赵文. 2023. 成岩作用控制下低渗透砂岩润湿性演化过程及机制:以渤海湾盆地东营凹陷为例. 石油与天然气地质, 44(5): 1308–1320. [Wang X, Zeng J H, Jia K K, Wang Q W, Li B, An C, Zhao W. 2023. Evolutionary process of the wet tability of low-permeability sandstone reservoirs under the control of diagenesis and its mechanism: a case study of the Dongying sag, Bohai Bay Basin. Oil & Gas Geology, 44(5): 1308–1320]
- 尹泽,刘自亮,彭楠,沈芳,谭梦琪,王玉冲,梁雨晨,王红伟. 2019. 鄂尔多斯盆地西缘上三叠统延长组沉积相特征研究. 沉积学报, 37(1): 163–176. [Yin Z, Liu Z L, Peng N, Shen F, Tan M Q, Wang Y C, Liang Y C, Wang H W. 2019. Study on sedimentary facies features of the upper Triassic Yanchang Formation, in the western margin, Ordos Basin. Acta Sedimentologica Sinica, 37(1): 163–176]
- 张宪国,刘玉从,林承焰,张涛,黄鑫,段冬平. 2019. 低渗—致密气层渗透率核磁测井解释方法. 中国矿业大学学报, 48(6): 1266–1275. [Zhang X G, Liu Y C, Lin C Y, Zhang T, Huang X, Duan D P. 2019. Permeability interpretation of low permeability and tight gas reservoirs with nuclear magnetic resonance logging. Journal of China University of Mining & Technology, 48(6): 1266–1275]
- 赵凯琳,靳小龙,王元卓. 2021. 小样本学习研究综述. 软件学报, 32(2): 349–369. [Zhao K L, Jin X L, Wang Y Z. 2021. Survey on few-shot learning. Journal of Software, 32(2): 349–369]
- 周锦涛,余文超,杜远生,邓旭升,翁申富,雷志远. 2025. 中国含铝岩系中稀土元素的赋存形式: 主量元素数据驱动的集成机器学习分析. 古地理学报, 27(2): 307–320. [Zhou J T, Yu W C, Du Y S, Deng X S, Weng S F, Lei Z Y. 2025. Occurrence forms of rare earth elements in aluminum-bearing rock series in China: ensemble machine learning analysis driven by major element data. Journal of Palaeogeography (Chinese Edition), 27(2): 307–320]
- 朱石磊. 2019. 鄂尔多斯盆地姬源地区三叠系延长组低孔低渗透砂岩储层特征及控制因素研究. 中国地质大学(北京)博士学位论文: 67–83. [Zhu S L. 2019. Reservoir characteristics and controlling factors analyzing of low porosity, low permeability sand reservoir, Jiyuan in Ordos. Doctoral dissertation of China University of Geosciences (Beijing): 67–83]
- Breiman L. 2001. Random forests. Machine Learning, 45(1): 5–32.
- Cao L, Jiang F J, Chen Z X, Gao Y, Huo L, Chen D. 2025. Data-driven interpretable machine learning for prediction of porosity and permeability of tight sandstone reservoir. Advances in Geo-Energy Research, 16(1): 21–35.
- Chen T Q, Guestrin C. 2016. XGBoost: a scalable tree boosting system. In: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 785–794.
- Coker A K. 1995. Fortran Programs for Chemical Process Design, Analysis, and Simulation. Houston: Gulf Professional Publishing, 1–102.
- Feng R M, Chen S N, Bryant S, Liu J. 2019. Stress-dependent permeability measurement techniques for unconventional gas reservoirs: review, evaluation, and application. Fuel, 256: 115987.
- Gao H, Li H Z. 2015. Determination of movable fluid percentage and movable fluid porosity in ultra-low permeability sandstone using nuclear magnetic resonance (NMR) technique. Journal of Petroleum Science and Engineering, 133: 258–267.
- He T, An L J, Chen P S, Chen J Z, Feng J S, Bzdok D, Holmes A J, Eickhoff S B, Yeo B T T. 2022. Meta-matching as a simple framework to translate phenotypic predictive models from big to small data. Nature Neuroscience, 25(6): 795–804.
- Hollmann N, Müller S, Purucker L, Krishnakumar A, Körfer M, Hoo S B, Schirmer R T, Hutter F. 2025. Accurate predictions on small data with a tabular foundation model. Nature, 637(8045): 319–326.
- Iraji S, Soltanmohammadi R, Matheus G F, Basso M, Vidal A C. 2023. Application of unsupervised learning and deep learning for rock type prediction and petrophysical characterization using multi-scale data. Geoenery Science and Engineering, 230: 212241.
- Ke G L, Meng Q, Finley T, Wang T F, Chen W, Ma W D, Ye Q W, Liu T Y. 2017. LightGBM: a highly efficient gradient boosting decision tree. In: The 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 3149–3157.
- Lai J, Wang G W, Wang S, Cao J T, Li M, Pang X J, Zhou Z L, Fan X Q, Dai Q Q, Yang L, He Z B, Qin Z Q. 2018a. Review of diagenetic facies in tight sandstones: diagenesis, diagenetic minerals, and prediction via well logs. Earth-Science Reviews, 185: 234–258.
- Lai J, Wang G W, Wang Z Y, Chen J, Pang X J, Wang S C, Zhou Z L, He Z B, Qin Z Q, Fan X Q. 2018b. A review on pore structure characterization in tight sandstones. Earth-Science Reviews, 177: 436–457.
- Lai J, Wang G W, Fan Q X, Pang X J, Li H B, Zhao F, Li Y H, Zhao X, Zhao Y D, Huang Y Y, Bao M, Qin Z Q, Wang Q Q. 2022. Geophysical well-log evaluation in the era of unconventional hydrocarbon resources: a review on current status and prospects. Surveys in Geophysics, 43(3): 913–957.
- Lai D D, Demartino C, Xiao Y. 2024a. Probabilistic machine leaning models for predicting the maximum displacements of concrete-filled

- steel tubular columns subjected to lateral impact loading. *Engineering Applications of Artificial Intelligence*, 135: 108704.
- Lai J, Su Y, Xiao L, Zhao F, Bai T Y, Li Y H, Li H B, Huang Y Y, Wang G W, Qin Z Q. 2024b. Application of geophysical well logs in solving geologic issues: past, present and future prospect. *Geoscience Frontiers*, 15(3): 101779.
- Lecun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436–444.
- Li P K, Tong X R, Wang Y J, Zhang Q. 2025a. Meta doubly robust: debiasing CVR prediction via meta-learning with a small amount of unbiased data. *Knowledge-Based Systems*, 310: 112898.
- Li Y R, Liu B, Hu Z M, Cai C H, Zhang Q X, Guo J S, Zeng S T, Li F. 2025b. Conductive characterization of shale reservoir induced by pore fluids: a case study of longmaxi formation in Southern Sichuan region. *Fuel*, 387: 134412.
- Lu H, Yue D L, Jones S J, Li S X, Wang W R, Bai B, Hou X L, Li Z, Wu S G, Li Q. 2024. Lithofacies assemblage and effects on diagenesis in lacustrine tight sandstone reservoirs: samples from Upper Triassic Yanchang Formation, Ordos Basin, China. *Marine and Petroleum Geology*, 167: 107001.
- Lundberg S M, Lee S I. 2017. A unified approach to interpreting model predictions. In: *The 31st International Conference on Neural Information Processing Systems*. New York: Curran Associates Inc.: 4768–4777.
- Luo J C, Yan J P, Liao M J, Wang M, Geng B, Hu Q H. 2024. Evaluation of pore structure characteristics of deep clastic rocks in the Huangliu Formation of LD-X area, Yinggehai Basin. *Marine and Petroleum Geology*, 167: 106969.
- Ma Y Y, Qiao Y H, Chen M X, Rui D N, Zhang X X, Liu W J, Ye L. 2024. How small is big enough? big data-driven machine learning predictions for a full-scale wastewater treatment plant. *Water Research*, 274: 123041.
- Maldar R, Ranjbar-karami R, Behdad A, Bagherzadeh S. 2022. Reservoir rock typing and electrofacies characterization by integrating petrophysical properties and core data in the Bangestan reservoir of the Gachsaran oilfield, the Zagros basin, Iran. *Journal of Petroleum Science and Engineering*, 210: 110080.
- Mao J Y, Chen J, Deng Y J. 2025. Optimization strategy for ensemble learning models based on fusing resampling, adaptive dimensionality reduction, and Optuna in intelligent flight technology evaluation. *Aerospace Science and Technology*, 162: 110251.
- Moreno-barea F J, Jerez J M, Franco L. 2020. Improving classification accuracy using data augmentation on small data sets. *Expert Systems with Applications*, 161: 113696.
- Noble W S. 2006. What is a support vector machine? *Nature Biotechnology*, 24(12): 1565–1567.
- Pittman E D. 1992. Relationship of porosity and permeability to various parameters derived from mercury injection-capillary pressure curves for Sandstone. *AAPG Bulletin*, 76(2): 191–198.
- Ren D Z, Zhou D S, Liu D K, Dong F J, Ma S W, Huang H. 2019. Formation mechanism of the Upper Triassic Yanchang Formation tight sandstone reservoir in Ordos Basin: take Chang 6 reservoir in Jiyuan Oil Field as an example. *Journal of Petroleum Science and Engineering*, 178: 497–505.
- Taunk K, De S, Verma S, Swetapadma A. 2019. A brief review of nearest neighbor algorithm for learning and classification. In: *The 2019 International Conference on Intelligent Computing and Control Systems (ICCS)*. New York: IEEE, 1255–1260.
- Wang F Y, Hou X M. 2025. Machine learning-based prediction of physical parameters in heterogeneous carbonate reservoirs using well log data. *Energy Geoscience*, 6(2): 100383.
- Xiao D S, Jiang S, Thul D, Lu S F, Zhang L C. 2018. Impacts of clay on pore structure, storage and percolation of tight sandstones from the Songliao Basin, China: implications for genetic classification of tight sandstone reservoirs. *Fuel*, 211: 390–404.
- Xu X T, Zeng L B, Dong S Q, Li H M, Liu J Z, Ji C Q. 2025. The characteristics and controlling factors of high-quality reservoirs of ultra-deep tight sandstone: a case study of the Dabai Gas Field, Tarim Basin, China. *Petroleum Science*, 22(9): 3473–3496.
- Yang S F, Bao Z D, Wang N, Qu X F, Lin Y B, Shen J J, Awan R S. 2020. Diagenetic evolution and its impact on reservoir quality of tight sandstones: a case study of the Triassic Chang 6 Member, Ordos Basin, northwest China. *Marine and Petroleum Geology*, 117: 104360.
- Yang T, Tang H M, Dai J W, Wang H, Wen X, Wang M, He J, Wang M N. 2025. Quantitative classification and prediction of pore structure in low porosity and low permeability sandstone: a machine learning approach. *Geoenergy Science and Engineering*, 247: 213708.
- Zhang S B, Wang S S, Huo J T, Gong C L, Li Z Q, Liu J Q, Xue R B, Yan Y H, Li B H, Shi Y H, Zhou B. 2025. Exploring aerosol vertical distributions and their influencing factors: insight from MAX-DOAS and machine learning. *Environmental Science & Technology*, 59(23): 11616–11627.
- Zhang Z, Zhang H, Li J, Cai Z X. 2021. Permeability and porosity prediction using logging data in a heterogeneous dolomite reservoir: an integrated approach. *Journal of Natural Gas Science and Engineering*, 86: 103743.
- Zhao W W, Zhang Z H, Liao J B, Zhang J W, Zhang W T. 2024. Prediction method for the porosity of tight sandstone constrained by lithofacies and logging resolution. *Marine and Petroleum Geology*, 170: 107114.
- Zou C N, Yang Z, He D B, Wei Y S, Li J, Jia A L, Chen J J, Zhao Q, Li Y L, Li J, Yang S. 2018. Theory, technology and prospects of conventional and unconventional natural gas. *Petroleum Exploration and Development*, 45(4): 604–618.

(责任编辑 郑秀娟)