

多源数据与机器学习驱动的页岩天然裂缝含油率智能预测方法

张晨¹, 何建华^{1*}, 刘达东^{2,3*}, 贾承造^{2,4*}, 宋岩^{2,3}

1. 成都理工大学, 油气藏地质及开发工程全国重点实验室, 成都 610059

2. 中国石油大学(北京), 油气资源与工程全国重点实验室, 北京 102249

3. 中国石油大学(北京), 非常规油气科学技术研究院, 北京 102219

4. 中国石油天然气集团公司, 北京 100724

* 通讯作者, E-mail: hejianhua19@cdut.edu.cn(何建华), liudd@cup.edu.cn(刘达东), jiacz@petrochina.com.cn(贾承造)

2025-07-26收稿, 2025-11-10修回, 2025-12-02接受, 2026-01-13网络版发表

国家自然科学基金项目(42402148、42472184)和国家科技重大专项项目(2025ZD1404100-001-005、2025ZD1403900-003-002)资助

摘要 天然裂缝含油率(HSFs)是页岩油气资源评价的关键参数, 然而传统基于实验测定的方法存在连续性差、成本高及效率低等局限。本研究提出了一种融合多源地质地球物理数据与机器学习算法的页岩HSFs连续定量预测方法。基于四川盆地侏罗系页岩含油率实验测试及多源响应地球物理数据, Pearson因子分析揭示了HSFs与孔隙度、TOC含量、声波时差及密度等参数之间存在显著相关性, 结果表明并非所有裂缝发育段均具有烃类的异常富集和产能提升, 显著高的HSFs值仅出现在同时具备高TOC含量及高孔隙度的页岩储层段。鉴于此, 形成了一套融合多源数据与地质信息约束的人工智能预测模型。采用优化的极端梯度提升监督学习模型, 模型预测HSFs性能优异: 残差(RE)为0.221, 均方根误差(RMSE)为0.311, 平均绝对误差(MAE)为0.156, 决定系数(R^2)达0.959。该方法可在短时间内实现单井页岩HSFs的高效、高精度连续评价, 预测精度超过95%。为页岩油气甜点预测与资源评价提供了新的手段, 并在全油气系统各关键地质要素的智能预测及数字孪生模型构建方面展现了广阔应用场景。

关键词 天然裂缝含油率, 页岩储层, 四川盆地, 智能预测, 全油气系统

1 引言

随着全球油气需求达到前所未有的水平, 常规资源已大多进入开发后期, 因此, 需要加大对非常规资源的勘探力度(邹才能等, 2013; Aguilera, 2014)。页岩油等非常规资源已成为主要的勘探目标, 其可采资源量占全球油气资源的42%(Whitelaw等, 2019)。尽管天然裂缝被广泛认为是页岩油富集的关键, 但其作用存在显著差异(Gale等, 2014)。天然裂缝中烃类的含量为深化其对页岩油聚集的贡献认识提供了见解(Aydin, 2000; Fu等, 2023)。不同时期形成的裂缝具有不同的烃类含量, 在生排烃高峰期保持开启的裂缝更可能储集烃类(Laubach,

2003; Li等, 2019), 若裂缝形成于生烃期之后, 其对成藏的贡献则相对有限(Fu等, 2023)。且这种裂缝以高角度穿层剪切裂缝为主, 则会加剧页岩油的逸散, 不利于页岩油成藏(Du等, 2023)。裂缝的产状、连通性及充填特征直接决定了页岩油气的赋存方式与产能差异(Zeng等, 2016)。与现今地应力方向夹角较小的开放性裂缝有利于烃类沿裂缝运移与聚集(刘国平等, 2024), 而被方解石或石英充填的裂缝则会削弱渗流能力, 导致页岩油的滞留(Zeng等, 2024)。天然裂缝的位置也至关重要, 高孔隙度页岩中的裂缝往往具有更高的含油饱和度(邹才能等, 2015; Liu等, 2020)。此外, 高生烃潜力页岩中的裂缝通常比低生烃潜力页岩中的裂缝含油量更高(Liu

中文引用格式: 张晨, 何建华, 刘达东, 贾承造, 宋岩. 2026. 多源数据与机器学习驱动的页岩天然裂缝含油率智能预测方法. 中国科学: 地球科学, 56(2): 650–666, doi: 10.1360/SSTe-2025-0226

英文引用格式: Zhang C, He J, Liu D, Jia C, Song Y. 2026. Intelligent prediction of hydrocarbon situation in shale natural fractures driven by multi-source data and machine learning. Science China Earth Sciences, 69(2): 630–646, <https://doi.org/10.1007/s11430-025-1787-7>

等, 2021, 2024), 因为在后者中, 仅孔隙空间就足以储油而无需裂缝提供额外空间(Tolmachev等, 2020)。且高TOC页岩, 一般纹层更加发育, 也会使得层理缝或生烃超压缝等非构造裂缝增多, 进一步为页岩油赋存提供裂隙空间(魏永波等, 2025; Liu等, 2025)。页岩有机质成熟度相对增高, 也可增加轻烃组分等游离烃的含量, 同时伴随着大量开启生烃超压缝的形成, 也可使得页岩裂缝中油气相对富集, 尤其是游离烃含量(刘全有等, 2022; Cao等, 2024)。因此, 天然裂缝含油率(HSFs)是评价裂缝有效性的关键指标。

HSFs的直接量化仍然具有挑战性, 需要使用间接指标进行定量表征。由于页岩中天然裂缝内的烃类主要以轻质游离油的形式存在(刘全有等, 2022; Cao等, 2024), 因此, 因此通过测定基质内大孔缝中的游离油含量, 便可作为评价天然裂缝含油率的可靠指标, 进而有效表征HSFs的富集状态。而页岩含油率的准确评价难度极大, 评价方法与测试手段尚未取得共识, 应用于页岩含油率评价的技术手段较多, 但受限方法原理和页岩的特殊性, 不同方法的表征效果各有优劣。如热解-气相色谱法(PY-GC)能够同时分析烃类含量和组成, 部分弥补了轻烃损失(Zhang等, 2022)。溶剂抽提方法对溶剂选择和比例高度敏感, 且耗时, 一般萃取时间往往需要两周以上(Handwerker等, 2011)。核磁共振(NMR)是评估孔隙度和烃类饱和度的新方法, 但其准确性依赖于核磁共振含氢组分的划分图版、流体信号与质量的转换关系标定(Zhang等, 2019)。然而, 这些技术之间缺乏标准化, 阻碍了盆地间的比较并限制了其广泛应用。目前, 对于含油率等关键参数的定量评价尚未形成共识, 这导致不同区块页岩孔隙度和含油率分析结果的对比难度大, 且利用岩芯实验测试方法获取页岩含油率存在时效性差、数据结果不连续且成本过高的缺点。

近年来, 随着人工智能方法的快速发展, 基于岩芯测试结果刻度测井来评价HSFs的方法开始不断涌现, 早期主要利用电阻率-孔隙度等测井曲线叠合法或交汇图法等定性方法进行研究, 形成了 $\Delta\log R$ 或改进的 $\Delta\log R$ 方法等半定量方法(Passey等, 1990; 陈敬武等, 2017)。这些模型实现了垂向上含油率连续评价, 然而, 由于页岩的孔喉结构复杂、孔喉半径小、流体赋存状态多样, 加之非均质性强、各向异性显著(矿物、有机质呈顺层或零散分布), 使得泥页岩的电阻率不仅受控于地层水的矿化度和黏土矿物, 还受到孔喉结构、水分布、干酪根成熟度与丰度影响(朱建伟等, 2012)。随后, 学者引入了更多测井参数, 形成了多元线性回归方法来定量预测页岩含油率, 但是回归模型存在过拟合的问题, 基本上都是一

个区域, 甚至一口井一套参数, 泛化能力很差。目前, 随着机器学习技术的迅速发展, 基于数据驱动的方法在页岩含油率预测中得到了一定的推广应用, 支持向量机、BP神经网络等监督学习算法实现了垂向上含油率快速定量评价(Wu等, 2013; Chaki等, 2015; Şen, 2022; Akbari等, 2025)。且最新机器学习算法在页岩油储层其他甜点参数预测中发挥了重要作用, 主要体现在页岩油储层复杂岩相智能识别(Lu等, 2023; Li等, 2024)、TOC含量预测(秦晓艳等, 2024)、物性预测(刘姝娴等, 2024)、天然裂缝测井智能识别(Lu等, 2024, 2025)以及甜点综合智能识别(王长江等, 2024)上。然而, 最新的机器学习算法在页岩油储层含油率预测中应用相关研究鲜有报道。且目前这些方法引入的测井数据以常规测井数据为主, 未考虑更多类型的数据信息资料和地质知识模型, 针对岩相复杂, 孔隙结构多样的页岩, 其含油率的预测精度有限, 且这些算法同样存在过拟合的问题, 无法通过调整配置参数表, 达到过拟合问题的控制。

四川盆地侏罗系页岩蕴藏丰富的页岩油资源, 并在经历多期构造运动改造后其天然裂缝普遍发育(孙莎莎等, 2021; 何文渊等, 2022)。然而, 在单井尺度上, 页岩裂缝线密度与油气产量之间的关系复杂。一些裂缝密度高的井产油量低, 而另一些裂缝密度较低的井产油量却显著更高(图1), 这表明可能与HSFs存在一定联系。同时这也可能与高密度裂缝发育带使页岩油气保存条件变差有关, 部分“高裂缝密度、低产”的井, 主要位于与大巴山造山相关的NW向断裂和雪峰山造山的NE向断裂带附近, 这些区域构造活动相对强烈, 裂缝系统易于保持开启状态并与切穿顶底板断裂相沟通, 从而使得压力散失, 油气产量变低。针对目前页岩裂缝含油率预测精度不足与效率低的问题, 本研究聚焦四川盆地东北部, 融合了录井、常规及特殊测井等多源数据, 结合页岩可溶有机质、TOC和物性分析等实验。系统揭示了页岩HSFs非均质性分布的地质控制因素。在此基础上, 形成了一套融合地质分析与机器学习的页岩HSFs智能预测工作流程, 首次提出了基于地质逻辑与多源数据驱动的页岩HSFs智能预测方法。大幅提升了模型预测精度与地质可解释性, 为页岩油储层甜点评价提供了新的思路。

2 方法

2.1 样品来源与实验方法

样品均来源川东北侏罗系页岩储层, 深度范围主要介于2000–3600 m, 一共开展了91项次的可溶有机质、TOC含量及孔隙度和渗透率配套测试。

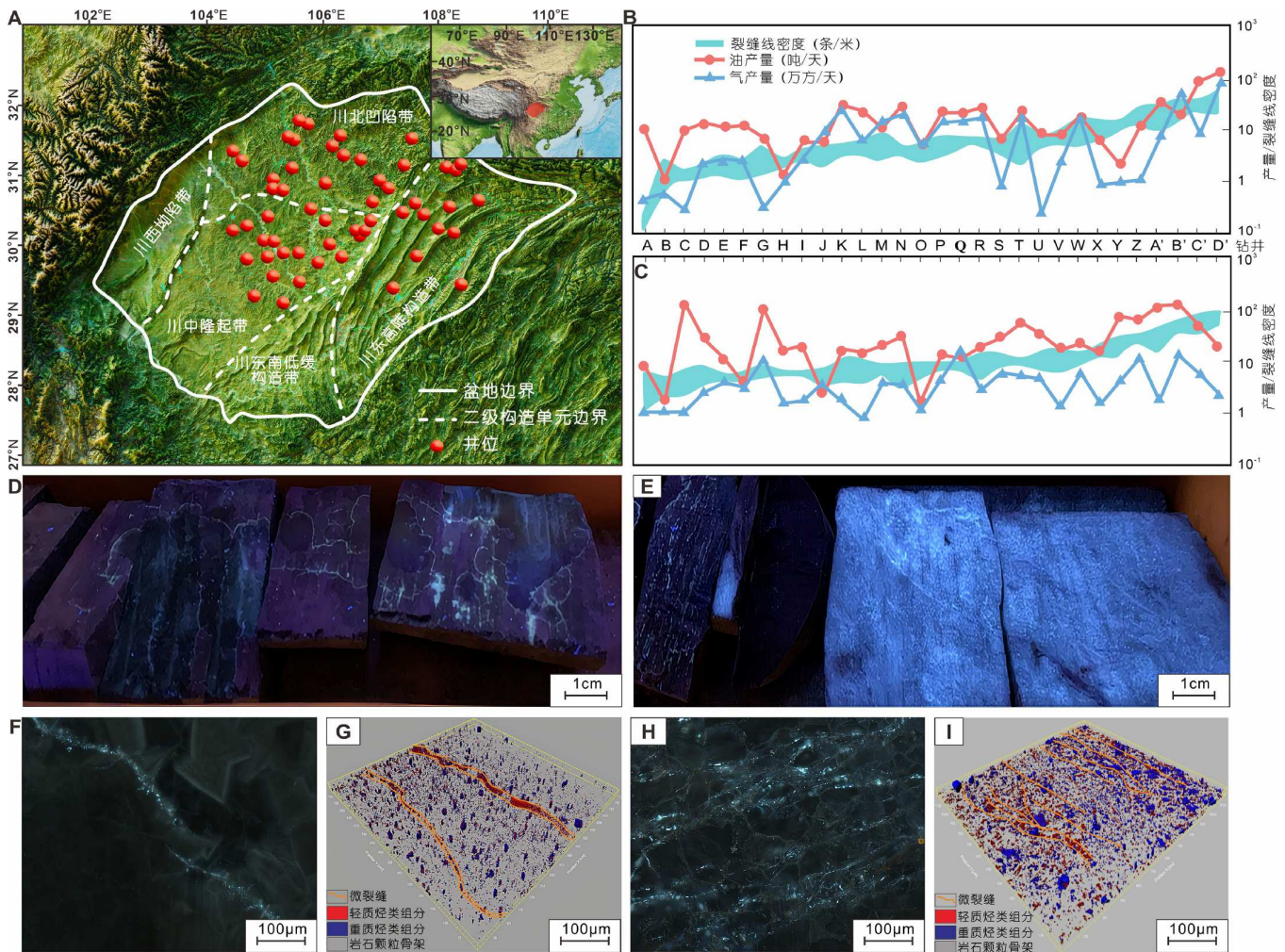


图1 四川盆地侏罗系页岩储层天然裂缝与页岩油气产能的关系图。(A) 四川盆地构造分区及侏罗系关键钻井分布图;(B) 凉高山组天然裂缝线密度与页岩油气产能的关系;(C) 自流井组大安寨段天然裂缝与页岩油气产能的关系;(D) Y1井, 1419.80–1420.10 m, 深灰色页岩中发育高角度-垂直剪切缝, 缝面可见具有分散状、发蓝色荧光特征的油斑、油迹;(E) Y1井, 1421.10–1421.40 m, 灰黑色页岩中发育垂直剪切缝, 缝面可见饱含油特征, 并显示强烈荧光;(F) P1井, 3005.6 m, 在紫外(UV)光照下, 粒间孔呈现淡黄、黄绿、黄白和蓝白的高强度荧光, 裂缝内观察到不连续的、线状的、发蓝白色荧光的烃类聚集;(G) P1井, 3005.6 m, 总含油量1.37%, TOC为0.89%, 孔隙度为2.31%, 激光共聚焦显示轻烃局部富集在裂缝中, 重质组分分散在基质孔隙内(据何文渊等, 2022修改);(H) Y1井, 1421.65 m, 在UV光照下, 方解石颗粒间出现淡黄至蓝白色荧光, 而裂缝中具有蓝白色荧光的页岩油呈现网状分布特征;(I) P1井, 3002.9 m, 总含油量3.16%, TOC为1.35%, 孔隙度为3.59%, 激光共聚焦显示重质组分呈团簇状富集于基质孔隙中, 而微裂缝和粒间孔内的轻烃呈现连续的带状富集特征(据何文渊等, 2022修改)

可溶有机质测定在成都理工大学油气藏地质及开发工程全国重点实验室完成, 首先对先前进行过扫描电镜(SEM)、光学显微镜及拉曼光谱成像分析的同一块样品表面, 实施了有机溶剂萃取。鉴于萃取有机质的数量与组分随溶剂性质而变化, 本次实验采用了非极性溶剂二氯甲烷与甲苯, 以及极性溶剂甲醇按体积比90:3:7组成的混合溶剂。样品置于装有约10 mL混合溶剂的样品瓶中, 在约50 ℃下进行30 min超声萃取, 加速溶剂-有机质相互作用。该处理过程重复多次, 直至混合溶液在常温静置一周后呈无色。萃取后的样品表面经甲醇清洗和漂洗后, 使用Baygon ArNanoFab 100系统进行5 min氩离子抛光,

以去除外部污染物。最后, 对此经溶剂萃取后的样品, 采用与原始样品相同的流程, 获取拼接的SEM图谱。

TOC含量测试在成都理工大学油气藏地质及开发工程全国重点实验室进行。样品经研磨成粉末后, 采用10%盐酸处理以去除碳酸盐及其他无机矿物。剩余物质在纯氧环境中加热, 并与溶液中的强氧化剂(CrO₃)反应。释放的CO₂气体经收集后, 使用Leco CS-230碳硫测定仪分析, 最终确定TOC含量。

孔隙度和渗透率测量在成都理工大学油气藏地质及开发工程全国重点实验室进行。样品制成直径2.5 cm的标准圆柱体样品, 去除表面杂质, 然后干燥至恒重。孔隙

度测量使用JS100007氦孔隙度仪进行, 渗透率测量使用A-10133气体渗透率仪进行。

2.2 HSFs测定与训练样本集构建

由于侏罗系中等-高成熟页岩天然裂缝中的烃类以轻质游离油为主, 且HSFs的直接量化存在困难, 故可将基质体系内大孔缝中的游离油含量(图1F和1G), 作为表征天然裂缝含油率的可靠替代指标。该方法为间接评价HSFs的富集程度提供了可行途径。本次主要采用页岩可溶有机质萃取方法得到的轻质游离油含量, 作为天然裂缝含油率的实验测试。此外, 由于本次只测定了91组页岩HSFs值, 不足以支撑大数据训练和人工智能预测, 因此, 我们也利用二维核磁测试了页岩大孔缝的游离油饱和度(图2A和2B), 测试仪器主要采用油气藏地质及开发工程国家重点实验室高性能二维核磁共振分析仪(MesoMR21), 测试原理和流程严格按照“岩样核磁共振参数实验室测量规范(SY/T 6490-2023)”标准执行, 之后建立了页岩HSFs值与大孔缝中游离油饱和度之间的定量关系模型(图2C)。由此, 基于二维核磁测井数据和上述关系模型, 我们在HSFs实验测试段, 人工生成了6000多个HSFs数据样本, 辅助HSFs大数据样本的构建(图2D)。并且针对岩芯据(共91个HSFs点)与测井曲线(超过6000个采样点)在深度分辨率上的差异, 主要采用以下三种方法进行精确对应和标定, 即(1) 深度标定与对齐: 根据取心报告及录井资料, 对岩芯深度进行系统校正, 使岩芯顶、底深度与测井曲线的层位标志(GR、DEN、AC的拐点或异常特征)严格对应; 采用核磁测井解释大孔缝轻质游离油饱和度曲线特征匹配与插值校正法实现逐段对齐; (2) 由于岩芯HSFs实验点代表一定厚度(约10–30 cm)的综合性质, 我们将其对应区间内测井曲线的参数取滑动窗口平均值(6–7 m), 要求HSFs人工生成样本数据段与实验测试段页岩岩相与有机地化特征相对一致, 这样使模型输入与输出具有相同尺度和物理意义; (3) 误差控制与敏感性评估: 匹配误差经验证一般小于 ± 0.3 m; 为验证模型稳健性, 我们进行了随机扰动试验(± 0.5 m深度偏移), 认为采样滑动窗口长度为6–7 m时, 人工生成HSFs样本的精度在0.90左右(图2D), 且其变化小于0.02, 说明模型对此滑动窗口内深度误差不敏感。

2.3 有监督学习方法

人工神经网络(ANN)是受生物神经系统结构启发的计算模型, 旨在模拟人脑处理信息的方式(Wang等, 2021; Ghanekar等, 2022), ANN由多个神经元通过连接而组成, 这些连接带有权重, 反映了信息传递的强度。

神经元通常按照层次结构排列, 包括输入层、隐藏层(层数 ≥ 1)和输出层(图3A), 本次选用了基于三层网络架构的ANN模型来预测HSFs。支持向量机(SVM)是主要用于二元分类的监督学习模型。SVM的核心概念是找到一个线性决策边界(称为最优分离超平面), 以最大化特征空间中两个类别的几何间隔, 从而区别于感知器算法(Vapnik, 1999; Nello和John, 2000)。SVM方法既可以用于分类(二/多分类), 也可用于回归和异常值检测(图3B)。且SVM具有良好的鲁棒性, 对未知数据拥有很强的泛化能力, 特别是在数据量较少的情况下, 相较其他传统机器学习算法具有更优的性能。极端梯度提升(XGBoost)是一种基于梯度提升框架构建的机器学习算法, 它迭代地构建决策树的集成以提高预测性能(Chen和Guestrin, 2016)。并通过不断地在已有模型的基础上, 拟合负梯度方向的残差(真实值与预测值的差)来构建新的弱学习器, 达到逐步优化模型的目的。XGBoost在构建决策树(CART树)时, 利用了二阶导数信息(图3C)。在损失函数的优化过程中, 不仅考虑了一阶导数(梯度), 还引入了二阶导数(海森矩阵), 这使得算法能够更精确地找到损失函数的最优解, 加速模型的收敛速度, 同时提高模型的泛化能力。轻量梯度提升机(LightGBM)是一种广泛采用的基于梯度提升决策树(GBDT)算法的机器学习框架(Ke等, 2017; 张占山等, 2021)。它通过迭代训练构建决策树的集成以优化模型性能(图3D)。

有监督机器学习模型的训练、优化及测试主要用于页岩裂缝含油率的预测, 训练数据集来源于十口井的多源数据, 包含6000个采样点。数据集包括录井、常规测井和特殊测井及测试数据。录井数据主要是气测录井, 包括气测全烃(TG)和甲烷浓度(C_1)。常规测井包括自然伽马(GR)、井径(CAL)、声波时差(AC)、中子孔隙度(CNL)、体积密度(DEN)、地层电阻率(RT)和冲洗带电阻率(RXO)。特殊测井包括光电吸收截面指数(PE)和核磁共振(NMR); 经标准化后的6000余个数据集被划分为两个部分, 利用Python软件将scikit-learn库的数据分隔功能自动生成训练集和测试/验证子集数据, 采用分层随机抽样技术划分数据集, 以降低性能偏差风险并减少变异性, 从而提升模型性能。训练集与测试/验证集分别占全数据库的80%和20%。数据集预测结果以随机50个样本的特征值随实验编号变化曲线图形式展示。

在预测HSFs之前, 建立了严格的数据预处理工作流程, 以减轻由仪器误差、环境噪声和井眼不规则引起的特征失真风险。通过减少噪声放大特征的影响, 标准化防止模型捕捉虚假相关信息, 从而, 为后续算法训练建立坚实基础。标准化公式如下:

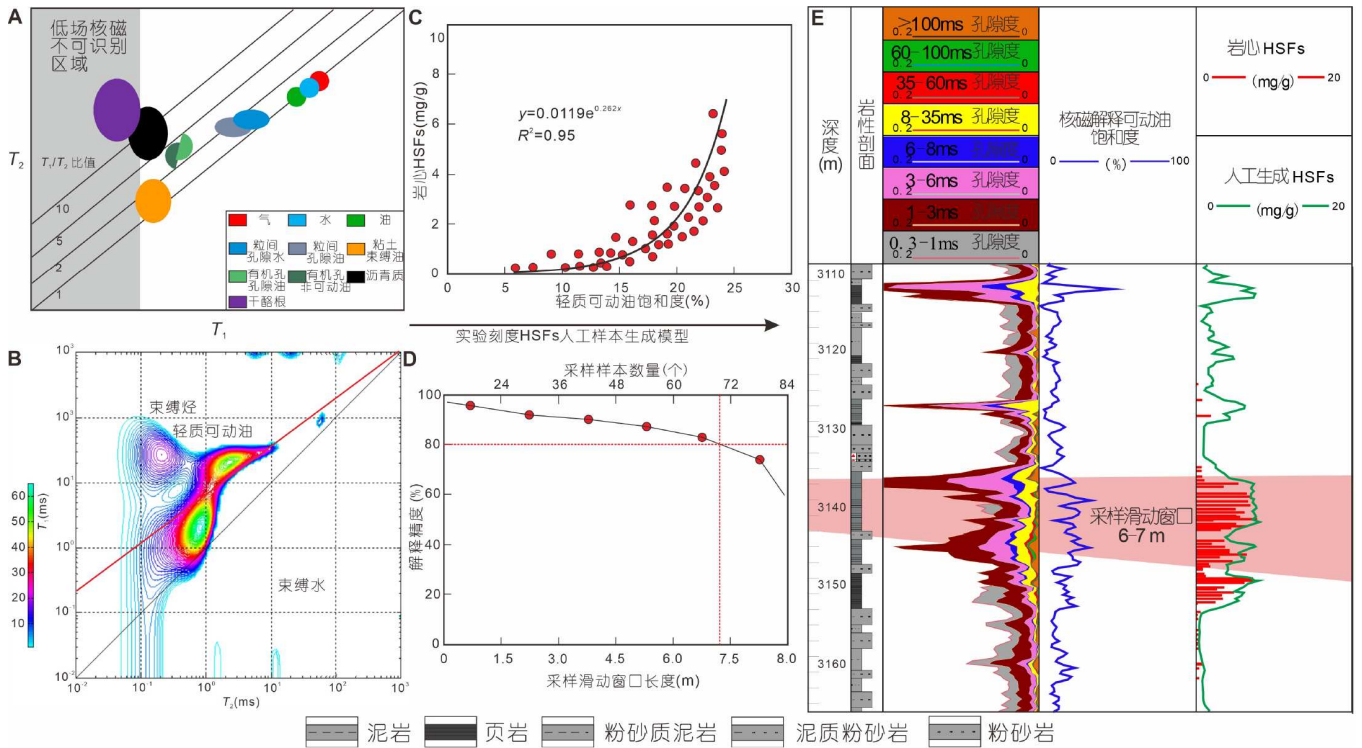


图2 基于实验刻度的人工生成HSFs训练样本集原理图。(A) 基于二维核磁测试的岩芯流体类型解释原理图(据Fleury和Romero-Sarmiento, 2016修改); (B) 基于二维核磁测试的页岩轻质可动油饱和度解释图版(据Fleury和Romero-Sarmiento, 2016修改); (C) 岩芯HSFs实验评价与核磁解释轻质可动油饱和度相关关系图; (D) 不同测井解释采样窗口下人工生成HSFs的解释精度图; (E) 基于岩芯实验刻度的人工生成HSFs训练样本示意图

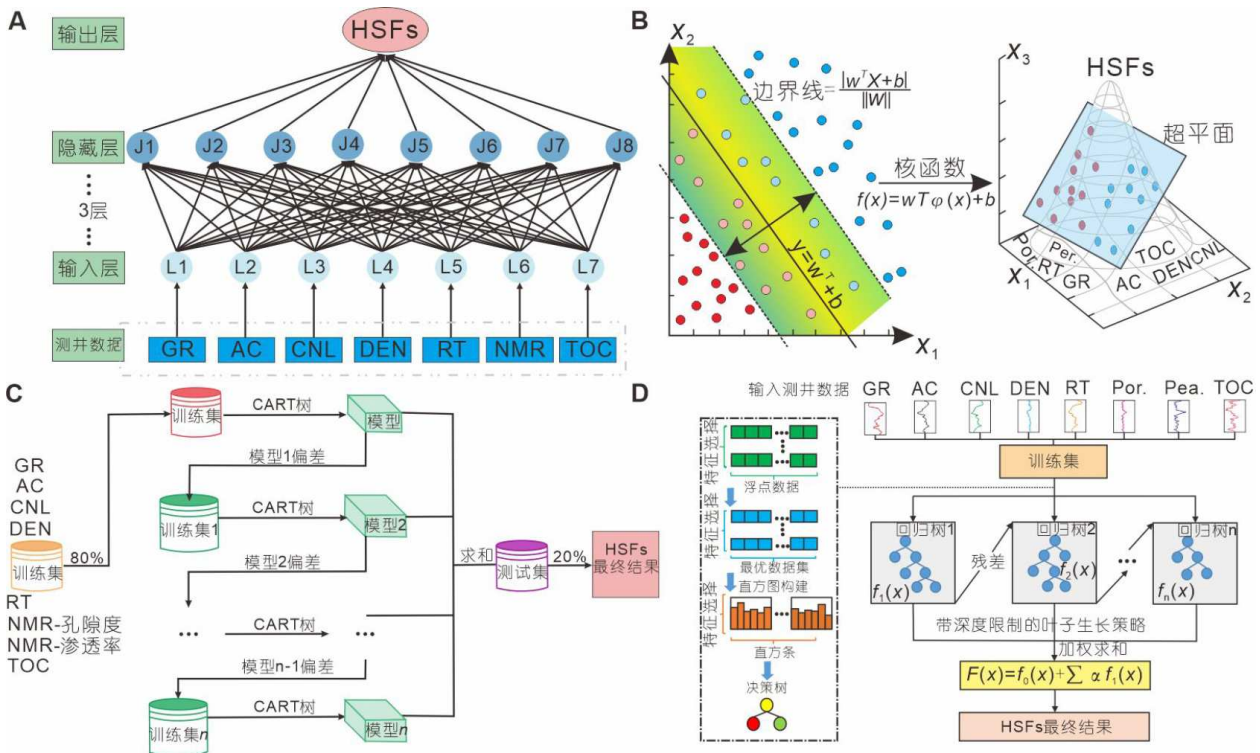


图3 四种预测算法的网络架构模型图。(A) 基于三层架构的ANN算法模型; (B) SVM回归预测算法模型; (C) XGBoost算法预测模型; (D) LightGBM算法预测模型

$$x = \frac{x - \mu}{\sigma} \quad (1)$$

其中, x 是原始值, μ 是该特征的均值, σ 是该特征的标准差。

有监督的机器学习模型超参数调优采用贝叶斯优化(王俊等, 2022)的超参数调优工具完成, 在超参数调优过程中, 通过多次训练模型以优化预测结果。采用种子函数确保相同超参数配置下可复现预测结果。本模型训练过程中同样引入了正则化参数 λ_{sparse} (稀疏性损失权重)防止过拟合, 训练参数包括 $batch_size$ (训练批量)和 $learning_rate$ (学习率), 同时训练模型引入早停机制 $patience$, 定义训练过程中容忍验证集性能未提升的连续轮次数。若超过该轮次仍无改进, 则提前终止训练, 防止过拟合。本文中的调参方法选用的是Optuna搜索超参数空间。将 n_d 与 n_a 设置在(8, 128), n_steps (1, 10), γ (1×10^{-3} , 1.0), λ_{sparse} (1×10^{-4} , 1×10^{-2}), $learning_rate$ (1×10^{-5} , 1×10^{-2}), $batch_size$ (8, 128)最后得到最优超参数。四种机器学习模型优化的超参数最终信息表见网络版附表S1, 表中包括各模型的主要参数名称、最优取值及优化方法。

为评估预测结果的可靠性, 采用了一套统计和图形度量指标。具体使用了四个关键评价指标: 残差(RE)、决定系数(R^2)、均方根误差(RMSE)和平均绝对误差(MAE)。这些指标定义如下:

$$RE = (\sigma_{\text{测试}} - \sigma_{\text{预测}}) \quad (2)$$

$$R^2 = \left[\frac{k \sum ab - (\sum a)(\sum b)}{\sqrt{k(\sum a^2) - (\sum a)^2} \sqrt{k(\sum b^2) - (\sum b)^2}} \right]^2 \quad (3)$$

$$RMSE = \sqrt{\frac{\sum (\sigma_{\text{测试}} - \sigma_{\text{预测}})^2}{N}} \quad (4)$$

$$MAE = \frac{\sum |\sigma_{\text{测试}} - \sigma_{\text{预测}}|}{N} \quad (5)$$

此处, $\sigma_{\text{测试}}$ 和 $\sigma_{\text{预测}}$ 分别代表HSFs的实验测试值和机器学习的预测值。 k 代表样本总数, 变量分别用 b 和 a 表示。 N 是数据点总数。

3 结果

3.1 页岩储层天然裂缝特征及其含油状态分析

四川盆地东北部侏罗系页岩储层中普遍发育不同成因类型的天然裂缝, 包括构造裂缝和非构造裂缝类型, 其中高角度-垂直剪切缝和水平层理缝占主导(图1D-1G)。不同类型裂缝的油气的饱和程度和分布状态存在明显差异(Saberi和Hosseini-Barzi, 2024)。在深灰色贫有机质富黏

土页岩中, 裂缝中赋存的烃类通常稀疏, 重烃比例较高, 聚集在黏土矿物周围, 轻烃含量有限且主要局限于裂缝内(图1D和1E)。相比之下, 灰黑色富有机质纹层状粉砂质页岩在基质内和裂缝中均表现出丰富的轻烃富集特征, 烃类呈现线状和带状分布(图1F和1G)。

3.2 HSFs的地质控制因素与敏感参数优选

总有机碳(TOC)含量与HSFs的相关性最强, Pearson相关系数为0.69, P 值低于0.0001(图4A), 为对HSFs最敏感的参数指标之一。此外, 孔隙度也显示出强正相关性($r=0.58$, $P<0.005$), 即孔隙度增大可以显著提升HSFs(附表S2)。相比之下, 渗透率的相关性较弱($r=0.23$, $P<0.01$), 表明其对HSFs的影响相对较小。此外, 为了分析输入特征参数与HSFs的所有非线性关系和特征间的交互作用, 本次研究也引入了基于博弈论的沙普利(SHAP)分析法(附图S1), 结果也表明了TOC和孔隙度对页岩HSFs的影响相对较大, 其SHAP影响程度值分别为0.91和0.80(图4B), 两者交互对HSFs影响也较大, 而渗透率对页岩HSFs的影响程度值为0.52, 表明了其对页岩HSFs的影响有限, 但整体地质参数SHAP影响程度值相比测井参数要高, 也说明了页岩储层地质参数在HSFs评价中发挥着重要作用。为确保HSFs测井预测的连续性, 采用了核磁测井解释孔隙度和渗透率值来替代实验测试的孔隙度和渗透率值, 采用实验刻度测井方法评价TOC值, 主要优选了对于页岩有机碳变化相对敏感的测井参数, 包括AC、RT、DEN三测井系列, 构建了实验参数刻度的TOC测井解释模型: $TOC = (15.66 \lg RT + 0.20 AC - 21.30) / DEN$, 且解释的TOC值与实验测试值相比, 精度可以达到95%以上(附图S2), 证明了该模型可用于研究区TOC的测井解释, 且孔隙度、渗透率和TOC的实验测试值也与相应的测井解释值具有较好一致性(图4C)。另外, 在页岩储层测井综合评价图中(图5), 核磁解释有效孔隙度和渗透率以及TOC解释值与岩芯值具有很好的相关性, 也能证实了核磁测井解释孔隙度和渗透率的准确性。

在常规测井响应中, 与孔隙度相关的测井曲线, 包括声波时差(AC)、中子(CNL)和密度(DEN), 对HSFs表现出最高的敏感性。AC和CNL测井曲线与HSFs呈正相关, 皮尔逊系数分别为0.55和0.52, P 值远低于0.001, 表明较高的AC和CNL值对应于较高的HSFs。相比之下, DEN测井曲线显示出强负相关性($r=-0.51$, $P<0.003$)。这与SHAP分析结果也相对一致, 声波时差(AC)、中子(CNL)和密度(DEN)的SHAP影响程度值分别为0.86、0.42及0.56, 但AC与DEN交互作用对页岩HSFs的影响更大。

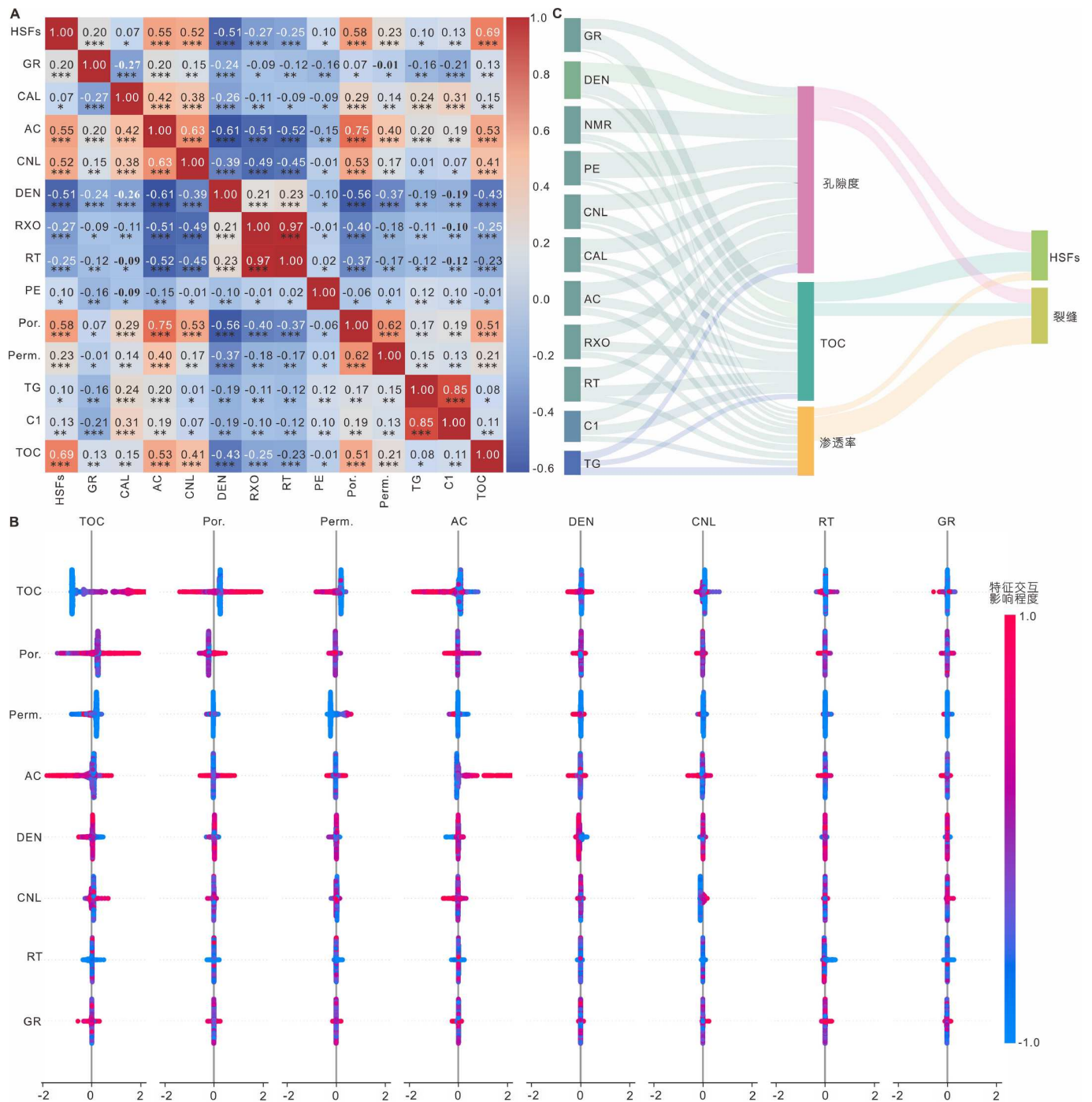


图4 预测模型输入参数与输出参数相关性分析与检验。(A) 关键输入参数的皮尔逊相关性和 t 检验结果。*表示显著性水平；***表示 $p<0.01$ ，**表示 $p<0.1$ ，*表示 $p<0.5$ ；(B) SHAP输入参数两两相互关系交互蜂群矩阵分析图；(C) 输入参数与输出参数之间的相关关系桑基分析图

岩性相关测井曲线(GR, CAL, PE)显示出较弱的相关性。自然伽马(GR)呈弱正相关($r=0.20$, $P<0.034$), 其SHAP影响程度值为0.15, 也说明对页岩HSFs对的变化不太敏感, 而PE仅表现出微弱正相关性($r=0.10$, $P>0.1$)。井径(CAL)测井曲线与HSFs无显著关系($r=0.07$, $P>0.24$)。电阻率测井曲线, 包括深电阻率(RT)和冲洗带

电阻率(RXO), 与HSFs呈中等负相关(r 分别为 -0.25 和 -0.27 ; $P<0.04$)。其SHAP影响程度值为0.28, 说明了相交于岩性, RT对页岩HSFs对的变化不太敏感。气测录井参数, 如气测全烃(TG)和甲烷(C_1), 与HSFs相关性较差(r 分别为0.10和0.13; $P>0.1$)。

总体而言, 孔隙度相关测井曲线是预测HSFs最有效

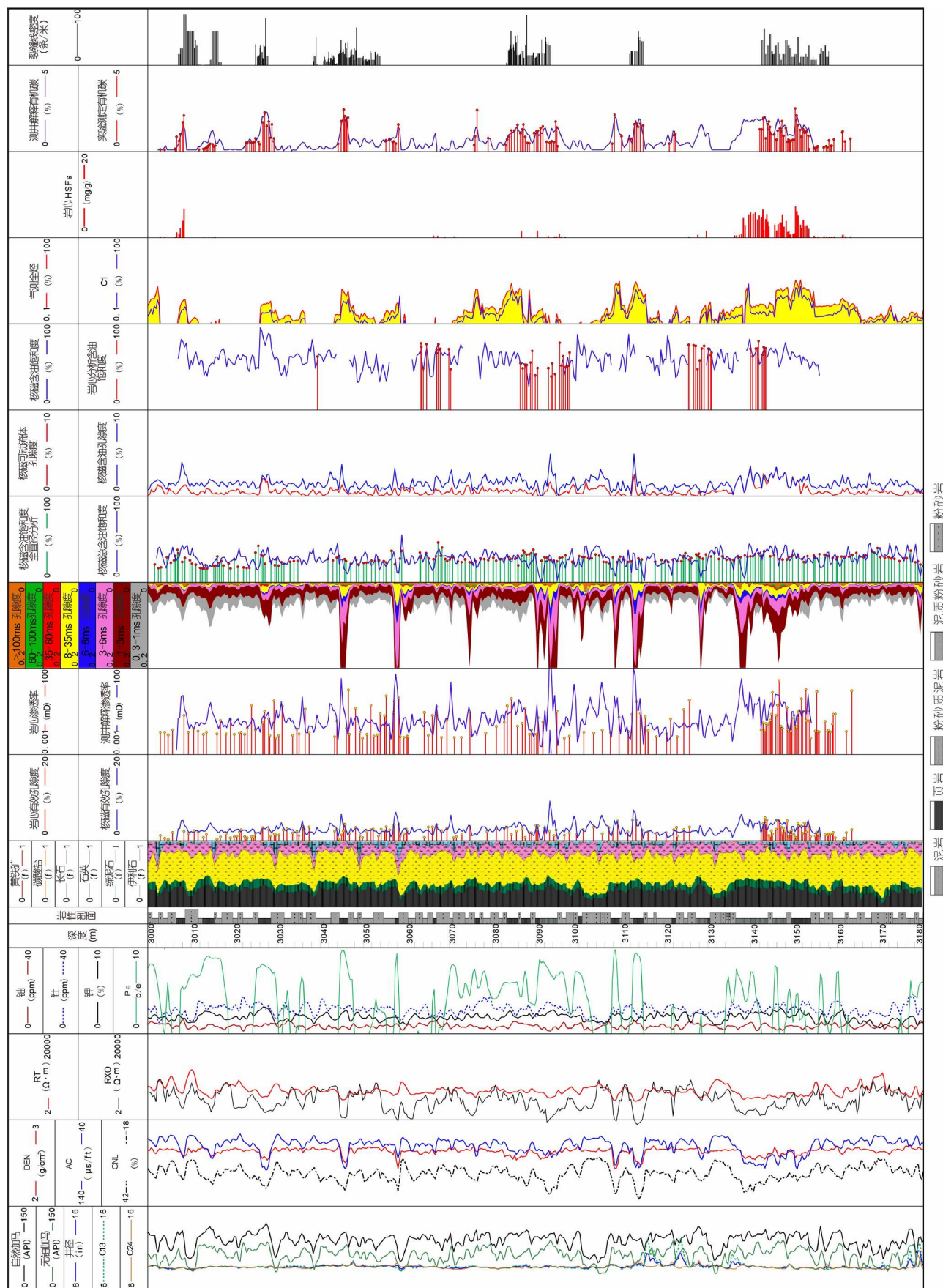


图5 P1侏罗系高山组页岩储层特征综合评价图。原始测井数据来源于大庆油田有限责任公司成都勘探开发研究院

的参数之一，其次是电阻率测井曲线，而岩性和录井参数对HSFs的敏感性相对较弱(图4C)。

在实际模型预测中，我们优选了一套Pearson相关系数大于0.20的敏感测井参数，包括GR、AC、CNL、DEN、RXO、RT、NMR解释的孔隙度和渗透率以及TOC。Pearson相关系数大于0.2反映了中等相关程度，故这些测井参数对HSFs的变化具有一定的响应(Schober等, 2018)， $r>0.2$ 常被用于识别具有一定敏感性的测井参数(Shao等, 2024)，这一经验阈值能够有效筛选出与目标变量变化趋势一致的参数。且输入和输出特征参数之间的关系也表现在两两交汇配对图中(图6)，该图是单一表示对于每对输入和输出特征之间的所有散点图。输出参数和输入参数之间的绘图反映了页岩储层孔隙度、TOC对HSFs具有显著的影响。直方图以及所有核密度估计函

数的分布特征沿对角线显示，而散点图显示在两两交汇图的下三角形中(图6)。结果表明HSFs与AC、CNL、孔隙度和TOC之间呈强正相关。相反，DEN、RXO和RT与HSFs呈负相关。

部分参数之间也存在相互依赖关系。如RXO和RT表现出强正相关，表明它们的预测信息存在冗余。因此，在实际预测模型中，仅需选择一个电阻率测量值即可获取电阻率对HSFs预测的贡献度。基于上述敏感输入参数提取，采用监督机器学习模型预测HSFs。模型中保留了八个关键输入敏感参数：GR、AC、CNL、DEN、RT、NMR解释的孔隙度和渗透率以及TOC。

3.3 HSFs的机器学习预测

以ANN机器学习预测HSFs与结果评价为例，人工神

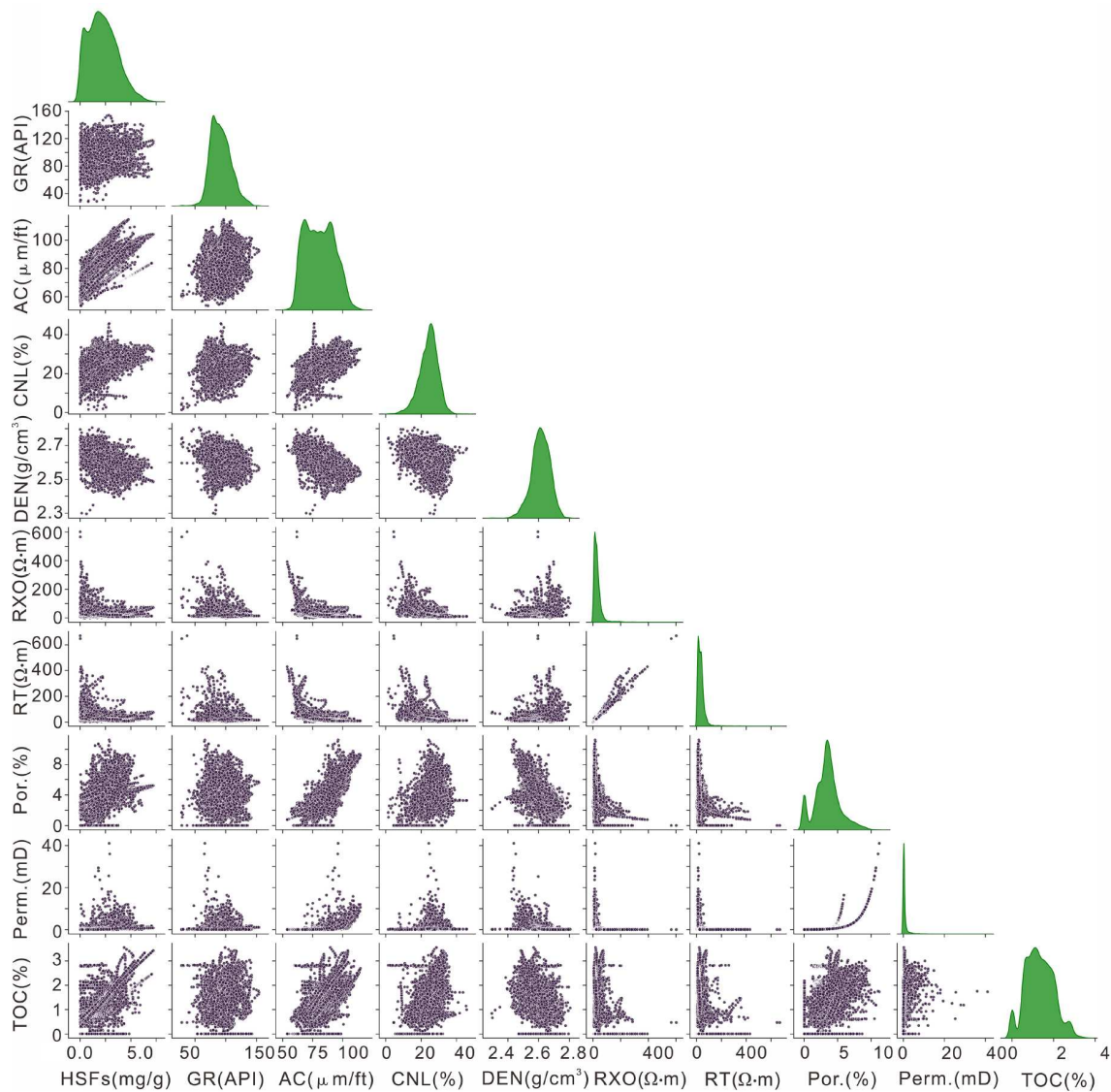


图6 输入参数与输出参数两两配对分析图

神经网络(ANN)模型中, 神经元层数与核心隐含层神经元数量构成网络基础架构。神经网络的关键组件还包括激活函数与训练算法。通过制定敏感性分析方案确定这些参数的最优组合配置, 超参数调优按敏感性分析中的操作序列实施, 形成ANN模型构建的四个阶段流程: (1) 隐含层数量优化; (2) 神经元数量评估; (3) 训练算法选择; (4) 激活函数验证; 神经元数量选择的敏感性分析如图7和图8所示, 这些图展示含油率预测模型在训练与测试/验证阶段下, 神经元数量与对应MAE、 R^2 值(见公式(3)和(5))的关联曲线。作为ANN模型开发的关键环节, 本研究执行了多次实现用以确定最佳随机种子值——该值作为本工作ANN代码开发的唯一标识符。模型共执行200–300次预测, 图7展示了HSFs在前200次实现中, 模型预测值与实验测试值之间的残差值(RE, 见公式(2))。基于上述超参数最优值与最佳实现结果的协同应用, 最终开发出针对页岩HSFs的优化ANN模型。采用均方根误差(RMSE)作为优化模型的损失函数(见公式(4)), 其量化了实验测试值与模型预测值间的偏差程度。ANN模型在预测页岩HSFs方面表现出卓越的性能, 其模型预测结果的各项评价指标显示如下: 在训练阶段中, $RE=0.325$, $R^2=0.915$, $RMSE=0.448$, $MAE=0.252$, 且在测试/验证阶段中, $RE=0.321$, $R^2=0.918$, $RMSE=0.444$, $MAE=0.251$, 图8通过实验值与预测值交汇分析, 展示ANN模型评价HSFs的预测性能。另外, 图7和图8详细对比了HSFs预测值与实验测试值, 结果均显示预测结果与实验测试的高度吻合, HSFs预测的残差频率均在低值区(近零处)出现峰值, 从而证实了ANN预测模型具有较为优异的预测能力。

除了ANN外, 本研究还基于相同的数据集对SVM、XGBoost及LightGBM等机器学习模型进行了训练与测试。通过贝叶斯优化的超参数调优工具优化各模型超参数。各模型在训练与测试/验证阶段的评估指标均进行系统对比, 图7则通过残差和差值分布图展示了所有模型在训练集、测试/验证集上的预测值与实验值对比情况。图7所示的尖峰型相对误差直方图及核密度估计(KDE)曲线表明, 除了LightGBM模型, 其他模型均表现出优异的预测性能, 且预测误差保持在较低水平。图8以交汇散点图形式呈现了所有模型在训练阶段及测试/验证阶段的HSFs预测值与实验值对比结果。

从图7和8中可以看出, 在相同的数据集上, ANN、SVM、XGBoost及LightGBM四种机器学习模型均表现出相对较好的预测HSFs性能, 其预测的各项误差与 R^2 值均处于较低的水平, 但各项预测评价结果指标显示了优化后的XGBoost模型在页岩HSFs预测中展现出最为显著的

优势, 在所有模型预测结果中, XGBoost在测试/验证阶段取得了最低的RMSE(0.311)和MAE(0.159)以及最高的 R^2 值(0.959)。因此, 我们认为优化的XGBoost机器学习模型对页岩HSFs的预测结果最准确。其次为ANN、SVM及LightGBM, 从MAE评价结果指标来看, SVM的预测结果表现比ANN相对要好, 其MAE值分别为0.249及0.251, XGBoost、ANN及SVM整体三个模型的预测结果评价指标之间差异并不是很大, 而LightGBM在各项评价指标中表现相对差一些。

4 讨论

4.1 HSFs预测模型的输入参数敏感性受控地质因素分析

为了进一步分析输入参数对预测模型输出特征重要性的影响程度, 本次基于SHAP热图和特征重要影响分析, 探讨了输入参数对XGBoost模型的特征重要性影响(图9, 附图S3), 该结果展示了模型对各输入特征各的重要性排序, 其中TOC、孔隙度(Por.)及渗透率(Peam.)在模型中具有相对最高的重要性程度, 其得分分别为0.471、0.181及0.1428, 表明这些参数对天然裂缝含油率(HSFs)的预测贡献最大; 其次为与孔隙度相关的测井输入参数, 包括了AC、DEN及CNL, 其得分为0.108、0.051及0.0169, 也表明了孔隙度对天然裂缝含油率预测影响相对较大; 而GR和RT测井参数的特征重要影响得分仅仅为0.016和0.0133, 说明了其对页岩HSFs预测的贡献有限, GR和RT在一定程度上反映了裂缝发育区的有机质丰度和流体性质, 其对模型也具有辅助作用。这些结果与地质分析及皮尔逊相关性分析结果总体一致, 但模型进一步揭示了参数间的非线性耦合效应。

在监督学习模型中使用的八个输入测井参数中, 孔隙度相关测井曲线对HSFs表现出最高的敏感性。这主要是由于较高的声波测井和中子测井值, 说明了储层裂缝相对空间越多, 裂缝开度越大, 这样容纳页岩油储存的裂隙空间就越丰富, 从而页岩裂缝含油率越大(刘国强等, 2023; 周义博等, 2025)。此外, 具有较高HSFs的页岩地层往往具有较低的体积密度。这是因为油气等烃类密度相对岩石骨架密度要低很多, 当油气等烃类组份占岩石总体积的量越大, 相应地岩石密度会大大降低。

岩性相关测井曲线表现出中等敏感性。较高的自然伽马(GR)值说明了页岩中黏土矿物和有机质等吸附介质越多, 为放射性元素和油气等提供了更多的吸附表面, 故页岩段GR值越高, 页岩含油率会有一定程度的增高, 但是黏土矿物和有机质吸附油气等分子总量是有限的,

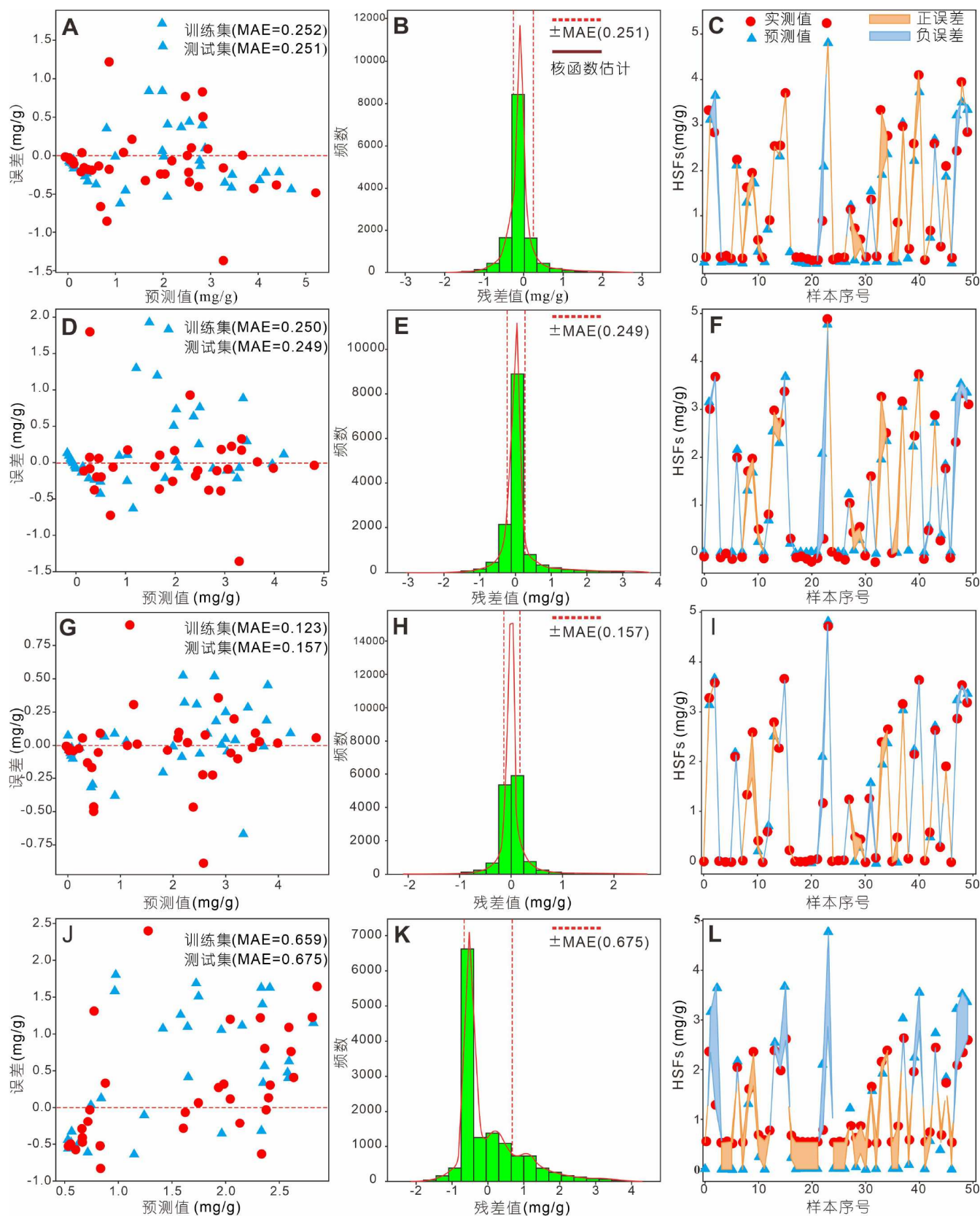


图7 基于优化后的四种监督学习模型预测HSFs的效果与残差分析图。(A-C) ANN模型结果：包括基于(A)50个数据点随机样本的残差散点图，(B)基于完整数据集的残差分布直方图，(C)预测HSFs与实测HSFs差异图；(D-F)SVM模型结果；(G-I)XGBoost模型结果；(J-L)LightGBM模型结果

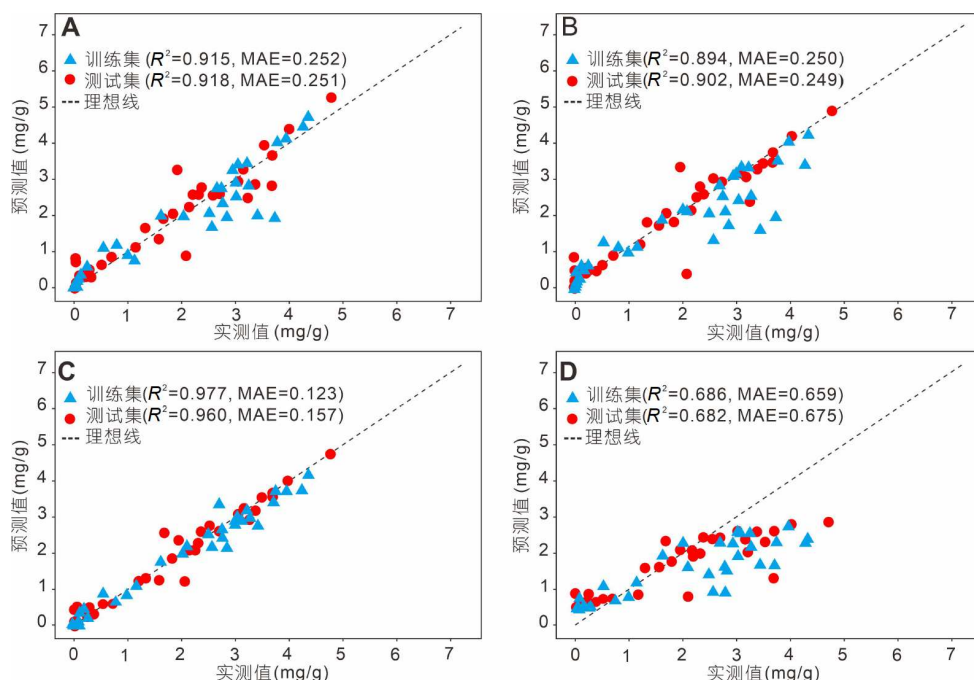


图8 基于四种不同监督机器学习模型预测的HSFs与其实测值对比图。散点图显示了50个随机选取样本的预测与实测HSFs值的相关分析性：(A) ANN、(B)SVM、(C)XGBoost、(D)LightGBM

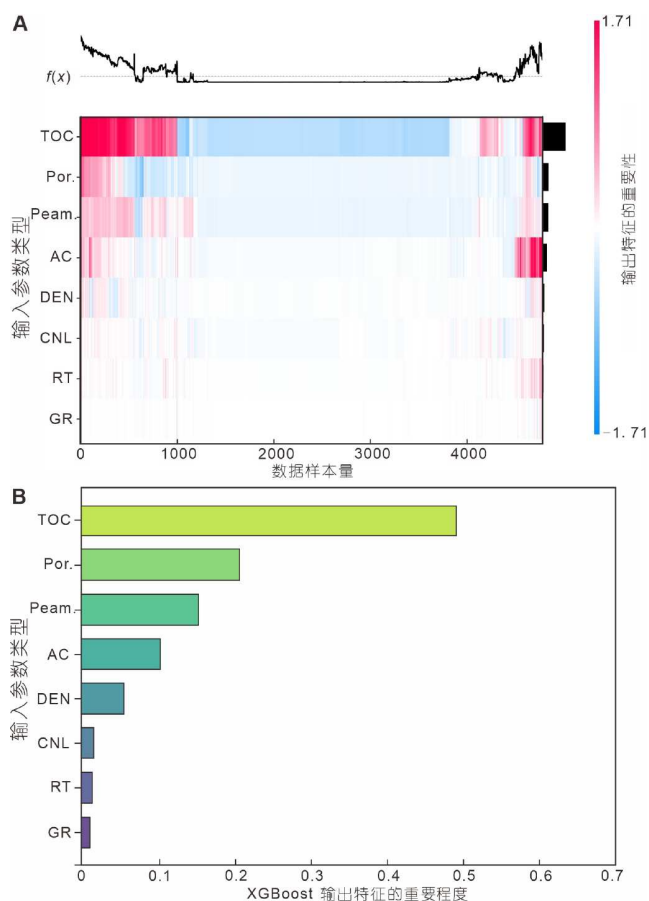


图9 基于XGBoost模型输出的特征重要性分析图。(A) 模型输出特征重要性的SHAP热图；(B) 模型输出特征重要程度统计图

且吸附在黏土矿物和有机质表面的油分子量比较大，流动性相对较差。因此，页岩段GR测井值的增高，并不能使页岩油含量迅速增加，且主要是影响了以吸附态形式存在的页岩油含量(李军等，2024；钱诗友等，2025)。其他岩性相关测井曲线，如井径(CAL)和光电吸收截面指数(PE)对页岩HSFs并不是很敏感，原因在于页岩中高原子序数矿物(黄铁矿/黏土)的光电吸收主导性彻底掩盖了烃类的微弱贡献，同时有机质自身演化和纳米孔隙流体的混合效应进一步削弱了Pe与含油率的相关性。因此，Pe测井相对于岩性特征区分度较好，而对烃类敏感程度相对较低(刘国强，2021；李宁等，2023)。而页岩段的CAL测井值与岩性具有较大的关系，当页岩中黏土矿物含量相对较高，容易造成页岩段吸水膨胀，使得井径扩大，此外，页岩段裂缝发育或地应力相对集中也会使井径扩大，可见页岩CAL值影响地质力学因素相对较多。

电阻率测井系列与页岩含油率具有一定的负相关性，这与常规砂岩储层含油率与电阻率具有正相关性存在一定矛盾性，一般而言，油的导电性相对地层水导电性低很多，如果油含量增大，电阻率一般是增大的，但页岩储层矿物组成与孔隙结构和常规砂岩储层具有很大的差异性，页岩储层中黏土矿物的离子导电性和黄铁矿的电子导电性共同构成了强大的背景导电网络。即使含油率增加，油的绝缘效应也无法有效阻断这一网络，导致电阻率不升反降。这一现象的本质是页岩的复杂矿物

组成与纳米孔隙结构削弱了烃类的电学贡献，使导电矿物的影响占据主导地位(李宁等, 2020; 姚东华等, 2022)。相比之下，气测录井与HSFs的相关性最弱。这主要是由于四川盆地侏罗系页岩储层的热成熟度相对较低($R_o < 1.12\%$)，其烃类组成以重质分子组分为主，且甲烷浓度较低，平均值不到15%(徐碧兰等, 2025)。因此，HSFs的增加虽然可提高裂缝储集空间与流体运移能力，但由于裂缝中以液态烃为主，未能显著增强气相组分的释放，因此气测录井响应变化不明显。

总体而言，孔隙度相关测井参数是预测HSFs最为敏感的参数，其次是电阻率测井曲线。岩性相关和气测录井参数表现出相对较低的敏感性。

4.2 各监督学习模型预测性能对比分析与验证

在本次研究中，将训练好的四种模型分别对验证集的测井数据集进行预测，将得到的预测含油率与实测含油率利用RMSE、MAE及 R^2 三种评价指标进行了对比分析，如表1所示，也可以看出XGBoost模型总体的各项预测结果指标均优于其他三种监督学习模型，根据三类预测结果评价指标分析的预测精度从高到低依次是XGBoost、ANN、SVM及LightGBM，同时，我们选取了川东北侏罗系7口天然裂缝相对发育的钻井，分析了侏罗系这些层段各模型预测页岩HSFs的表现情况，同样，XGBoost模型预测页岩HSFs值与实验测试值相比，其精度更高，且尤其是在油气富集程度相对较高的页岩层段表现更为出色。

在页岩HSFs非均质性分布的主控地质因素分析和敏感测井参数提取过程中，认为页岩HSFs受多种地质、地球物理及地球化学因素(如TOC、孔隙度、成熟度、矿物组成及测井响应等)非线性因素交互影响，HSFs预测的不确定性大大增强，而优化后的XGBoost相比其他模型具有决策树集成结构(如梯度提升、正则化及特征重要性方面)，能够有效获取这些复杂的非线性关系和特征交互信息，同时，地质数据往往维度高且可能存在冗杂或噪声，而XGBoost内置的特征重要性评估机制(Patience机制)可以帮助我们有效捕获复杂地质信息中的关键特征参数，且其正则化项(L1/L2)可以进一步防止过拟合，提高模型的泛化能力(张兵等, 2025; 陈家鑫等, 2025)，另外，与ANN和SVM相比，XGBoost提供了更高的透明度、效率和鲁棒性(田仁飞等, 2024)。虽然ANN模型功能强大，但它们通常被视为缺乏可解释性的“黑箱”模型，并且需要大量的参数调优和计算资源(Kraiem等, 2024; Vásquez和Kumral, 2025)。SVM虽然对某些任务有效，但在处理高维稀疏特征时可能遇到困难，并且在应用于大型

表1 ANN、SVM、XGBoost和LightGBM模型在HSFs预测的训练和测试阶段的评价指标表

ML model	RMSE	MAE	R^2
ANN	0.444	0.251	0.9180
SVM	0.486	0.249	0.9017
XGBoost	0.311	0.156	0.9590
LightGBM	0.575	0.378	0.8310

地质数据集时计算效率低下(Li等, 2025)。XGBoost在处理测井数据中常见的缺失值和噪声输入时也表现出更强的鲁棒性。虽然XGBoost和LightGBM都基于梯度提升决策树(GBDT)，但XGBoost精确的贪婪算法在预测复杂页岩储层结构方面提供了更高的精度(Yang等, 2025)。相比之下，LightGBM的leaf-wise算法虽然在计算上高效，但在富有机质页岩层段更容易出现过拟合问题(Liu等, 2023)。

4.3 井剖面HSFs预测与空间泛化外推能力

为评估优化的模型在井尺度纵向和空间上的预测性能，选取了川东地区其他的四口天然裂缝相对发育的钻井。在所有模型中，XGBoost始终表现出最准确的HSFs预测能力，尤其是在高孔隙度页岩层段内(图10A-D)。这一结果突显了优化XGBoost模型的鲁棒性和适应性，表明了其在相似构造背景下预测页岩储层HSFs具有更广泛的适用性。

优化的XGBoost模型为侏罗系陆相复杂页岩储层HSFs高效、精准量化评价提供了强有力的手段，对页岩储层各甜点参数预测和勘探开发决策具有重要的实际应用价值。然而，本研究也依赖于实验测试结果的标定、数据质量及数据集的代表性，未来研究工作将进一步聚焦在，(1) 扩充样本量并涵盖更广泛地质条件，用以提升模型的普适性，增强模型应用的泛化能力；(2) 融合更多的地质要素信息(如地质力学或成藏富集过程关键要素)，实现地质知识与数据信息驱动的页岩含油率高精度预测；(3) 融合多种模型的预测结果，实现在复杂地质背景下页岩HSFs预测的泛化能力大幅提升；(4) 将训练模型应用于更多钻井HSFs的测井解释中，并结合地震数据，实现页岩HSFs的空间分布高分辨率预测(图10E-H)。

4.4 全油气系统各地质要素数字孪生模型构建的潜在应用

在本研究中，我们通过融合多源数据与地质信息，成功构建了页岩HSFs的智能预测模型。在未来，该思路方法将在预测全油气系统各关键地质要素方面具有重大

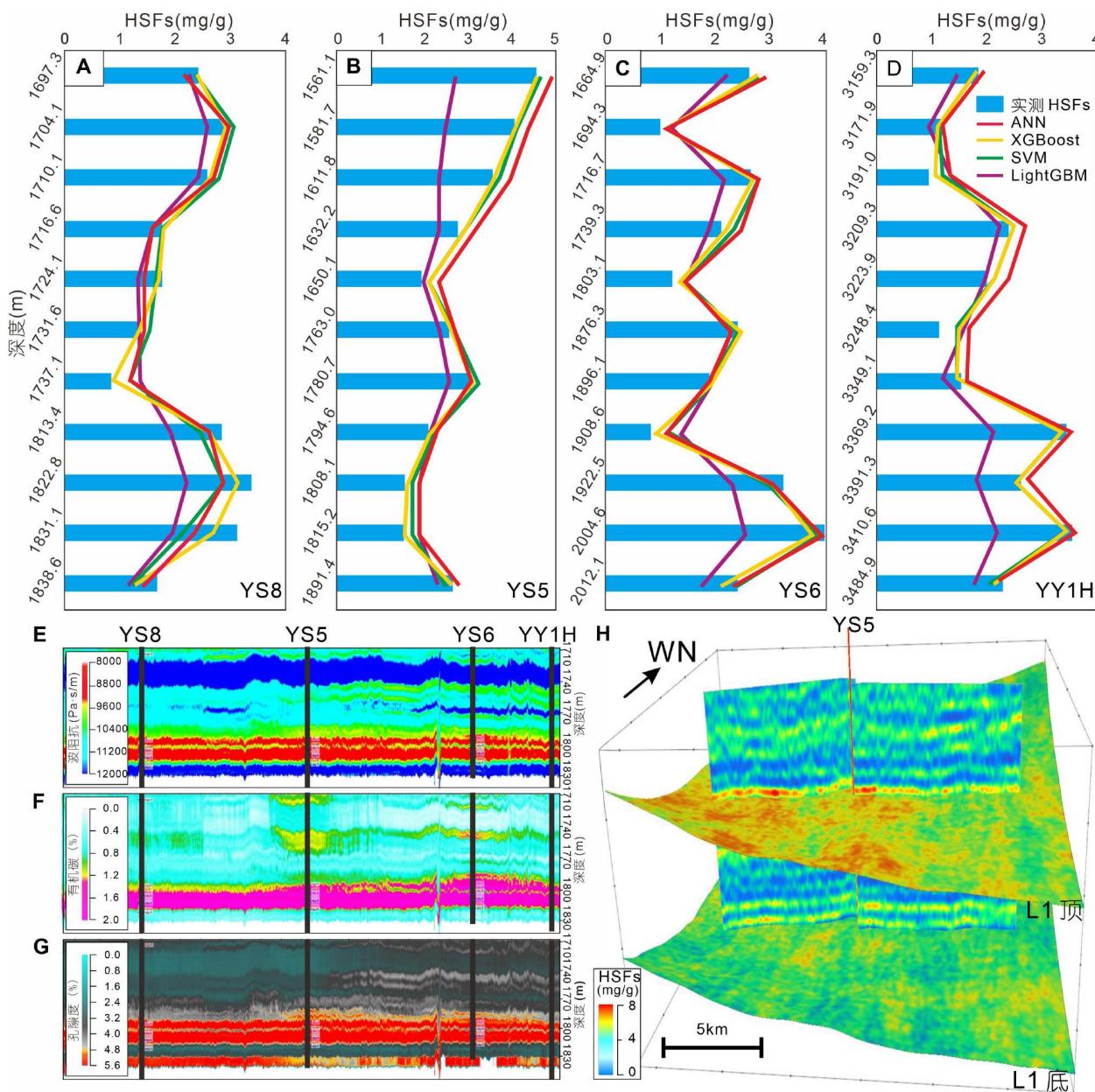


图 10 侏罗系钻井不同模型预测HSFs值与其实测值对比以及HSFs的空间分布预测图。(A–D) 四口关键井侏罗系凉高山组上段HSFs预测值与实测值对比图：(A)YS8井，(B)YS5井，(C)YS6井，(D)YY1H井；(E) 过YS8-YS5-YS6-YY1H井的地震波阻抗反演连井剖面图；(F) 过YS8-YS5-YS6-YY1H井的总有机碳(TOC)地震预测剖面图；(G) 过YS8-YS5-YS6-YY1H井的孔隙度地震预测剖面图；(H) 凉高山组上段HSFs空间分布预测图，L1代表凉高山组凉1段。原始地震数据体来源大庆油田有限责任公司成都勘探开发研究院

应用前景。全油气系统是由烃源岩、储层、油气水和流体动力场等各种地质要素构成的一套完整的流体系统(贾承造等, 2024)。对此类系统进行准确表征和预测仍然是油气勘探开发的核心挑战之一。

未来的研究应聚焦于利用深度学习技术(如多任务学习和图神经网络(GNNs))的强大能力进行非线性映射和模式识别(Cao等, 2022; Tang和Durlifsky, 2024)。这些方法提供了一种有效工具, 可将地质概念模型(包括生烃

动力学、流体运移模型和油气系统演化模型)与油气勘探开发全生命周期中产生的大量、异构、多尺度的数据集进行系统整合。这些数据集涵盖钻井、录井、测井、试井、生产动态和地震(图11)。关键技术突破包含了复杂知识(如地质过程规则、物理化学约束)的有效嵌入机制(如约束损失函数、先验引导初始化、知识图谱协同推理, Deswarte等, 2019)、多源异构数据的智能融合方法以及预测结果的不确定性量化技术(如贝叶斯深度学习、

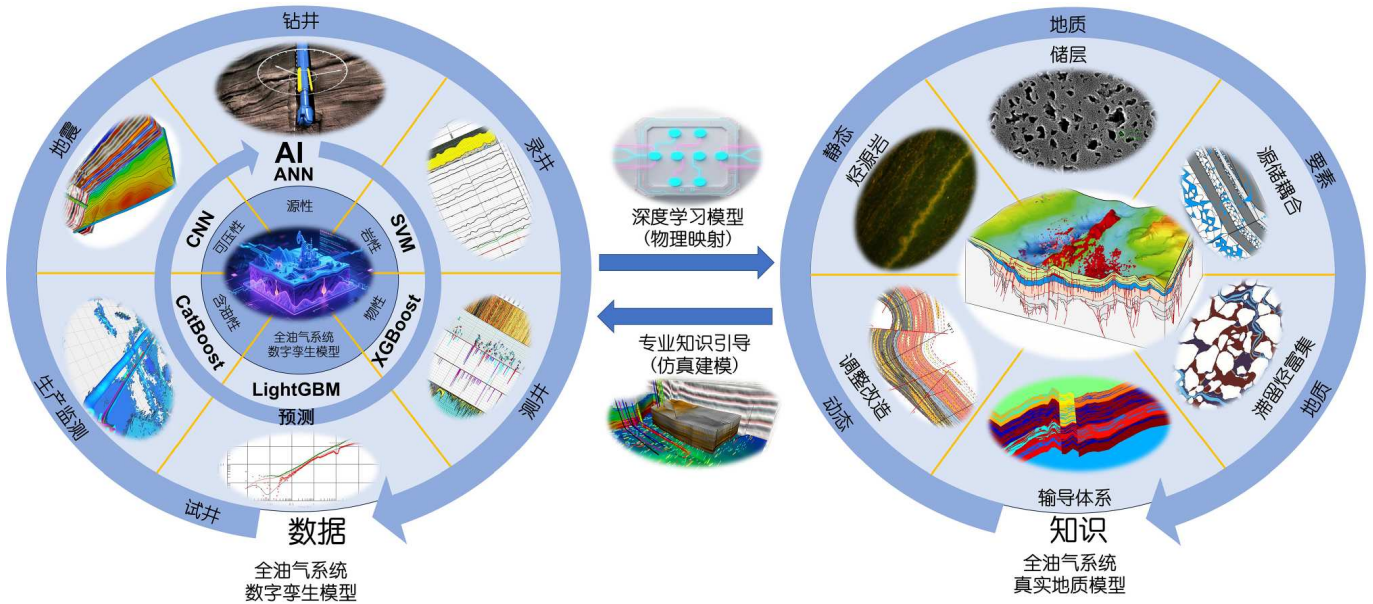


图 11 由数据和专业知识驱动的全油气系统各地质要素智能预测与数字孪生建模概念图

集成学习, Tadjer等, 2022, Fan等, 2024), 以构建更加鲁棒、可解释的数据与知识双驱动预测模型。此类模型可以实现全油气系统中关键地质要素的协同预测, 包括烃源岩生排烃潜力、储层品质和含油气性、源-储连通性及滞留烃分布。

基于上述智能模型对全油气系统各地质要素的预测结果, 我们可以构建一个高保真的、动态演化的全油气系统数字孪生体。该数字孪生体将集成地质、地球物理、油藏工程等多学科知识, 并能持续同化实时更新的生产监测数据。与静态数字地质模型不同, 它可以用来模拟和预测全油气系统的全生命周期过程, 涵盖烃源岩生烃、运移、聚集、后期调整改造直至开采枯竭, 实现对地下流体赋存状态与流动规律的“透明化”分析。

本方向的发展是推动油气工业向智能化、数字化转型迈进的战略性里程碑。其成功运用有可能通过将决策从基于经验的方法转向由数据和地质知识模型共同驱动的方法, 从根本上重塑传统的油气勘探工作流程。实现从经验单向驱动向数据与知识双驱动决策的根本性转变, 最终达到大幅降本增效、提高资源探明率与油气采收率的核心目标。同时, 通过构建和迭代优化数字孪生体, 也将极大地深化人类对复杂油气系统形成演化机制这一基础地质科学问题的认知, 形成技术突破与科学认知相互促进的良性循环。未来研究将积极探索人机协同智能决策框架, 确保地质学家专业智慧与模型强大计算预测能力的有机结合, 共同应对能源勘探开发的复杂挑战。

5 结论

通过多源测井数据与机器学习融合方法, 可以用于有效提取页岩裂缝含油率的关键地质控制因素, 并由此构建页岩裂缝含油率的高精度预测模型, 页岩较高的总有机碳含量(TOC)与孔隙度提升了天然裂缝中的含油率, 二者呈显著正相关关系, 协同促进页岩油富集。且TOC含量的影响尤为显著; 而页岩渗透率与孔喉特征对裂缝含油率的控制作用相对有限。裂缝发育密度及含烃品质直接控制页岩油储层甜点分布。常规测井参数(GR、AC、CNL、DEN、RT)及核磁共振解释的孔隙度/渗透率对页岩裂缝含油率的响应相对敏感, 为机器学习模型优化提供了有效的输入特征。对比ANN、SVM、XGBoost和LightGBM算法, 优化后的XGBoost模型预测性能最优($R^2=0.959$, RMSE=0.311)。基于贝叶斯优化的超参数调优策略, 进一步提升了模型的泛化能力。该项研究构建了一套“地质分析-敏感参数优选-机器学习建模-盲井验证”的完整工作流程, 实现了地质约束与数据驱动方法的有机融合, 首次将XGBoost模型应用于四川盆地侏罗系页岩层系的天然裂缝含油率预测, 实现了多源测井参数的高精度非线性建模, 使得单井HSFs连续、快速评价精度可达95%以上。为页岩油甜点预测提供了数据与机器学习驱动的新范式, 可降低其对昂贵实验测试的依赖, 未来可以进一步用于全油气系统各地质关键要素的智能预测, 构建全油气系统数字孪生模型, 从而实现全油气系统的高精度地质建模与全生命周期数值模拟。

致谢 本研究所用计算资源由成都理工大学所属的“油气藏地质及开发工程全国重点实验室”超级计算平台提供。

补充材料 补充材料为作者提供的原始数据，作者对其学术质量和内容负责，详见网络版(<http://earthcn.scichina.com>)。

参考文献

- 陈家鑫, 赵军龙, 崔文洁, 等. 2025. 基于GWO-XGBoost算法的流体识别——以陇东油田CX区长2储层为例. *地球物理学进展*, 40: 1115–1124
- 陈敬武, 朱建伟, 孙平昌. 2017. 采用支持向量回归从测井曲线定量计算油页岩含油率. *地质与资源*, 26: 157–161
- 何文渊, 白雪峰, 蒙启安, 等. 2022. 四川盆地陆相页岩油成藏地质特征与重大发现. *石油学报*, 43: 885–898
- 贾承造, 庞雄奇, 宋岩. 2024. 全油气系统理论基本原理. *石油勘探与开发*, 51: 679–691
- 李军, 邹友龙, 路菁. 2024. 陆相页岩油储层可动油含量测井评价方法——以苏北盆地古近系阜宁组二段页岩油为例. *石油与天然气地质*, 45: 816–826
- 李宁, 闫伟林, 武宏亮, 等. 2020. 松辽盆地古龙页岩油测井评价技术现状、问题及对策. *大庆石油地质与开发*, 39: 117–128
- 李宁, 冯周, 武宏亮, 等. 2023. 中国陆相页岩油测井评价技术方法新进展. *石油学报*, 44: 28–44
- 刘国平, 金之钧, 曾联波, 等. 2024. 玛湖凹陷二叠系风城组深层陆相页岩储层天然裂缝及其有效性. *地球科学*, 49: 2346–2358
- 刘国强. 2021. 非常规油气勘探测井评价技术的挑战与对策. *石油勘探与开发*, 48: 891–902
- 刘国强, 赵先然, 袁超, 等. 2023. 陆相页岩油宏观结构测井评价及其甜点优选. *中国石油勘探*, 28: 120–134
- 刘全有, 李鹏, 金之钧, 等. 2022. 湖相泥页岩层系富有机质形成与烃类富集——以长7为例. *中国科学: 地球科学*, 52: 270–290
- 刘姝娴, 李晓明, 李勇根, 等. 2024. 裂缝-孔隙型页岩气储层岩石物理建模及物性参数预测. *地球物理学进展*, 39: 594–605
- 钱诗友, 杨志强, 徐晨. 2025. 苏北盆地低有机质断块型页岩油测井评价方法及应用. *油气藏评价与开发*, 15: 19–27, 39
- 秦晓艳, 王震亮, 程昊, 等. 2024. 基于多测井参数的陆相页岩储层总有机碳含量预测: 以和尚岭地区延长组长7段为例. *新疆大学学报(自然科学版中英文)*, 41: 620–628
- 孙莎莎, 董大忠, 李育聪, 等. 2021. 四川盆地侏罗系自流井组大安寨段陆相页岩油气地质特征及成藏控制因素. *石油与天然气地质*, 42: 124–135
- 田仁飞, 李山, 刘涛, 等. 2024. 基于XGBoost算法的 V_p/V_s 预测及其在储层检测中的应用. *石油地球物理勘探*, 59: 653–663
- 王长江, 颜世翠, 张娟, 等. 2024. 基于模糊融合预测的页岩地质甜点识别技术——以胜利油田渤南洼陷为例. *油气地质与采收率*, 31: 73–83
- 王俊, 曹俊兴, 赵爽, 等. 2022. 基于深度混合神经网络的横波速度反演预测方法. *中国科学: 地球科学*, 52: 1151–1169
- 魏永波, 刘全有, 卢双舫, 等. 2025. 中国陆相页岩油富集机制: 以渤海湾盆地饶阳凹陷沙河街组为例. *中国科学: 地球科学*, 55: 2268–2289
- 徐碧兰, 何建华, 吴炎峰, 等. 2025. 多源数据与地质逻辑融合的深层页岩储层构造裂缝测井智能识别. *成都理工大学学报(自然科学版)*, 52: 1–20
- 姚东华, 周立宏, 王文革, 等. 2022. 页岩油综合甜点测井评价——以沧东凹陷孔店组二段为例. *石油学报*, 43: 912–924
- 张兵, 王晓婷, 徐福颖, 等. 2025. 基于XGBoost的孔隙压力预测方法研究. *地球物理学进展*, 40: 541–555
- 张占山, 姚鑫, 刘兆庚, 等. 2021. 基于LightGBM的特征选择算法. *东北大学学报(自然科学版)*, 42: 1688–1695
- 周义博, 肖苏芸, 曾芙蓉, 等. 2025. 川东复兴地区东岳庙段陆相页岩测井相划分及储层分类方法. *地球物理学进展*, 40: 1032–1044
- 朱建伟, 赵刚, 刘博, 等. 2012. 油页岩测井识别技术及应用. *吉林大学学报(地球科学版)*, 42: 289–295
- 邹才能, 杨智, 崔景伟, 等. 2013. 页岩油形成机制、地质特征及发展对策. *石油勘探与开发*, 40: 15–27
- 邹才能, 翟光明, 张光亚, 等. 2015. 全球常规-非常规油气形成分布、资源潜力及趋势预测. *石油勘探与开发*, 42: 14–28
- Aguilera R. 2014. Flow units: From conventional to tight-gas to shale-gas to tight-oil to shale-oil reservoirs. *SPE Reservoir Evaluation Eng*, 17: 190–208
- Akbari A, Ranjbar A, Kazemzadeh Y, et al. 2025. Enhanced water saturation estimation in hydrocarbon reservoirs using machine learning. *Sci Rep*, 15: 29846
- Aydin A. 2000. Fractures, faults, and hydrocarbon entrapment, migration and flow. *Mar Pet Geol*, 17: 797–814
- Cao D P, Ji S Q, Cui R G, et al. 2022. Multi-task learning for digital rock segmentation and characteristic parameters computation. *J Pet Sci Eng*, 208: 109202
- Cao Y, Jin Z J, Zhu R K, et al. 2024. Comprehensive evaluation of the organic-rich saline lacustrine shale in the Lucaogou Formation, Jimusar sag, Junggar Basin, NW China. *Energy*, 294: 130786
- Chaki S, Routray A, Mohanty W K, et al. 2015. Development of a hybrid learning system based on SVM, ANFIS and domain knowledge. In: *Proceedings of DKFIS 2015 Annual IEEE India Conference*. New Delhi. 1–6
- Chen T, Guestrin C. 2016. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, 785–794
- Deswarthe R, Gervais V, Stoltz G, et al. 2019. Sequential model aggregation for production forecasting. *Comput Geosci*, 23: 1107–1124
- Du X Y, Jin Z J, Zeng L B, et al. 2023. Formation of natural fractures and their impact on shale oil accumulation in the Mahu Sag, Junggar Basin, NW China. *Int J Coal Geol*, 279: 104385
- Fan Z, Liu X S, Wang Z Q, et al. 2024. A novel ensemble machine learning model for oil production prediction with two-stage data preprocessing. *Processes*, 12: 587
- Fleury M, Romero-Sarmiento M. 2016. Characterization of shales using T1–T2 NMR maps. *J Pet Sci Eng*, 137: 55–62
- Fu X F, Su X C, Gong L, et al. 2023. Control of faults and fractures on shale oil enrichment. *Geoenergy Sci Eng*, 228: 212080
- Gale J F W, Laubach S E, Olson J E, et al. 2014. Natural fractures in shale: A review and new observations. *AAPG Bull*, 98: 2165–2216
- Ghanekar P G, Deshpande S, Greeley J. 2022. Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis. *Nat Commun*, 13: 5788
- Handwerker D A, Suarez-Rivera R, Vaughn K I, et al. 2011. Improved petrophysical core measurements on tight shale reservoirs using retort and crushed samples. *SPE*, 147456
- Ke G L, Meng Q, Finley T, et al. 2017. LightGBM: A highly efficient gradient boosting decision tree. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Montreal. 3146–3154
- Kraiem Z, Zouari K, Trabelsi R. 2024. Harnessing machine learning tools for water quality assessment in the Kebili shallow aquifers, Southwestern Tunisia. *Acta Geochim*, 43: 1065–1086
- Laubach S E. 2003. Practical approaches to identifying sealed and open fractures. *AAPG Bull*, 87: 561–579
- Li H, Chen M, Zhang X K, et al. 2025. Reservoir fluid identification based on bayesian-optimized SVM model. *Processes*, 13: 369
- Li M W, Chen Z H, Ma X X, et al. 2019. Shale oil resource potential and oil mobility characteristics of the eocene-oligocene Shahejie Formation, Jiyang super-depression, Bohai Bay Basin of China. *Int J Coal Geol*,

- 204: 130–143
- Li Z J, Deng S G, Hong Y Z, et al. 2024. A novel hybrid CNN–SVM method for lithology identification in shale reservoirs based on logging measurements. *J Appl Geophys*, 223: 105346
- Liu D D, Zhang C, Pan Z K, et al. 2020. Natural fractures in carbonate-rich tight oil reservoirs from the Permian Lucaogou Formation, southern Junggar Basin, NW China: Insights from fluid inclusion microthermometry and isotopic geochemistry. *Mar Pet Geol*, 119: 104500
- Liu G P, Jin Z J, Zeng L B, et al. 2025. Laminar controls on bedding-parallel fractures in Permian lacustrine shales, Junggar Basin, northwestern China. *GSA Bull*, 137: 3512–3526
- Liu Q Y, Li P, Jin Z J, et al. 2021. Preservation of organic matter in shale linked to bacterial sulfate reduction (BSR) and volcanic activity under marine and lacustrine depositional environments. *Mar Pet Geol*, 127: 104950
- Liu Q Y, Li P, Jiang L, et al. 2024. Distinctive volcanic ash-rich lacustrine shale deposition related to chemical weathering intensity during the Late Triassic: Evidence from lithium contents and isotopes. *Sci Adv*, 10: eadi6594
- Liu Y, Zhu R, Zhai S, et al. 2023. Lithofacies identification of shale formation based on mineral content regression using LightGBM algorithm: A case study in the Luzhou block, South Sichuan Basin, China. *Energy Sci Eng*, 11: 4256–4272
- Lu G Q, Zeng L B, Dong S Q, et al. 2023. Lithology identification using graph neural network in continental shale oil reservoirs: A case study in Mahu Sag, Junggar Basin, Western China. *Mar Pet Geol*, 150: 106168
- Lu G Q, Zeng L B, Liu G P, et al. 2024. Bedding-parallel fracture density prediction using graph convolutional network in continental shale oil reservoirs: A case study in Mahu Sag, Junggar basin, China. *Mar Pet Geol*, 167: 106992
- Lu G Q, Zeng L B, Liu G P, et al. 2025. Subsurface natural fracture identification using an integrated ensemble learning method. *IEEE T Geosci Remote*, 63: 1–16
- Nello C, John S T. 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. London: Cambridge University Press. 33–41
- Passey Q R, Moretti F J, Kulla J B, et al. 1990. A practical model for organic richness from porosity and resistivity logs. *AAPG Bull*, 74: 1777
- Schober P, Boer C, Schwarte L A. 2018. Correlation coefficients: Appropriate use and interpretation. *Anesth Analg*, 126: 1763–1768
- Saberi F, Hosseini-Barzi M. 2024. Effect of thermal maturation and organic matter content on oil shale fracturing. *Int J Coal Sci Technol*, 11: 16
- Şen Ş. 2022. Prediction of fluid oil and gas volumes of shales with a deep learning model and its application to the Bakken and Marcellus shales. *Sci Rep*, 12: 20842
- Shao R B, Wang H, Xiao L Z. 2024. Reservoir evaluation using petrophysics informed machine learning: A case study. *Artif Intelligence Geosci*, 5: 100070
- Tadjer A, Hong A, Bratvold R. 2022. Bayesian deep decline curve analysis: A new approach for well oil production modeling and forecasting. *SPE Reservoir Evaluation Eng*, 25: 568–582
- Tang H Y, Durlofsky L J. 2024. Graph network surrogate model for subsurface flow optimization. *J Comput Phys*, 512: 113312
- Tolmachev O, Urunov A, Muminova S, et al. 2020. Review of unconventional hydrocarbon resources: Production technologies and opportunities for development. *Min Miner Deposits*, 14: 1–15
- Vapnik V N. 1999. The Nature of Statistical Learning Theory. New York: Springer-Verlag Press, 13–25
- Vásquez J C G, Kumral M. 2025. Artificial neural networks and support vector machines for more accurate cost estimation in underground mining: A contractor's viewpoint. *Machine Learn Appl*, 21: 100689
- Wang S H, Pillai H S, Wang S W, et al. 2021. Infusing theory into deep learning for interpretable reactivity prediction. *Nat Commun*, 12: 5288
- Whitelaw P, Uguna C N, Stevens L A, et al. 2019. Shale gas reserve evaluation by laboratory pyrolysis and gas holding capacity consistent with field data. *Nat Commun*, 10: 3659
- Wu H L, Li N, Lan C L, et al. 2013. Standard spectrum measurement and simulation of elemental capture spectroscopy log. *Appl Geophys*, 10: 109–116
- Yang T, Tang H, Dai J W, et al. 2025. Quantitative classification and prediction of pore structure in low porosity and low permeability sandstone: A machine learning approach. *Geoenergy Sci Eng*, 247: 213708
- Zeng L B, Lyu W Y, Li J, et al. 2016. Natural fractures and their influence on shale gas enrichment in Sichuan Basin, China. *J Nat Gas Sci Eng*, 30: 1–9
- Zeng X, Yang T, Feng J W, et al. 2024. Development characteristics and controlling factors of fractures in lacustrine shale and their geological significance for evaluating shale oil sweet spots in the third member of the Shahejie Formation in the Qikou Sag, Bohai Bay Basin. *Pet Sci*, 21: 791–805
- Zhang C, Liu D D, Jiang Z X, et al. 2022. Mechanism for the formation of natural fractures and their effects on shale oil accumulation in Junggar Basin, NW China. *Int J Coal Geol*, 254: 103973
- Zhang S, Cao Y C, Liu K Y, et al. 2019. Characterization of lacustrine mixed fine-grained sedimentary rocks using coupled chemostratigraphic-petrographic analysis: A case study from a tight oil reservoir in the Jimusar Sag, Junggar Basin. *Mar Pet Geol*, 99: 453–472

(责任编辑: 刘全有)