2026 年第 41 卷 第 5 期
2026,41(5):2175-2195
地球物理学进展 Progress in Geophysics
http://www.progeophys.cn
ISSN 1004-2903
CN 11-2982/P

张梓童,周春栩,刘怡辰,等.2026.LogGBS:一种扩展到粒度球空间的测井岩性识别方法.地球物理学进展,41(5):2175-2195,doi:10.6038/pg2026JJ0384.

ZHANG ZiTong, ZHOU ChunXu, LIU YiChen, *et al*. 2026. LogGBS: a lithology identification method extended to granular-ball space based on well logs. *Progress in Geophysics* (in Chinese), 41(5):2175-2195,doi:10.6038/pg2026JJ0384.

LogGBS:一种扩展到粒度球空间的测井岩性识别方法

LogGBS: a lithology identification method extended to granular-ball space based on well logs

张梓童[1,2],周春栩[1,2*],刘怡辰[3],冯涵霖[4]

ZHANG ZiTong[1,2], ZHOU ChunXu[1,2*], LIU YiChen[3], FENG HanLin[4]

1. 油气资源与工程全国重点实验室,中国石油大学(北京),北京 102249
2. 地球科学学院,中国石油大学(北京),北京 102249
3. 信息与电气工程学院,中国农业大学,北京 100083
4. 计算机与信息技术学院,东北石油大学,黑龙江大庆 163318

1. *State Key Laboratory of Petroleum Resources and Engineering*, *China University of Petroleum (Beijing)*, *Beijing* 102249, *China*
2. *College of Geosciences*, *China University of Petroleum (Beijing)*, *Beijing* 102249, *China*
3. *College of Information and Electrical Engineering*, *China Agricultural University*, *Beijing* 100083, *China*
4. *School of Computer & Information Technology*, *Northeast Petroleum University*, *Daqing Heilongjiang* 163318, *China*

摘 要 为解决复杂地质环境下传统岩性识别方法精度与效率不足、现有机器学习模型泛化能力弱等问题,本研究提出基于粒度球空间的测井岩性识别方法,即LogGBS.该方法通过自适应粒度球生成器(AGBG)将原始测井数据映射至粒度球空间并生成不同的粒度球表示,以“粒度球纯度”为核心控制参数,最终结合GBkNN分类器实现准确的岩性分类.为了验证该方法的有效性,研究选取了鄂尔多斯盆地杭锦旗气田(碎屑岩储层)、苏里格气田(碳酸盐岩储层)及松辽盆地徐深气田(火山岩储层)的测井数据集,对比LogGBS与八种传统机器学习和深度学习方法的性能差异,并分析粒度球纯度对分类精度的影响.结果显示,纯度不低于0.9时,各储层平均准确率超90%,纯度为0.95时部分指标超99%.LogGBS在三类储层中均表现出更优的识别精度与泛化能力,为复杂地质条件下测井岩性识别提供新思路,助力油气勘探开发效率提升.

Abstract To address the limitations of traditional lithology identification methods in accuracy and efficiency, as well as the poor generalization of existing machine learning models under complex geological environments, this study proposes a logging lithology identification method based on granular-ball space, termed LogGBS. This method maps raw well log data into the granular-ball space and constructs multiscale granular-ball representations through an Adaptive Granular-Ball Generator (AGBG). Using “granular-ball purity” as the key control parameter, LogGBS ultimately performs lithology identification by a Granular-Ball kNN (GBkNN) classifier. To evaluate the effectiveness of the proposed method, well log data from the Hangjinqi Gas Field (clastic reservoir) and Sulige Gas Field (carbonate reservoir) in the Ordos Basin, as well as the Xushen Gas Field (volcanic reservoir) in the Songliao Basin, are selected as research objects. The performance of LogGBS is compared with that of eight traditional machine learning and deep learning methods, and the influence of granular-ball purity on classification accuracy is analyzed. The results show that when the purity is not less than 0.9, the average accuracy of each reservoir exceeds 90%; when the purity reaches 0.95, some indicators

收稿日期 2025-12-22; **修回日期** 2026-05-15. **投稿网址** http://www.progeophys.cn

基金项目 国家自然科学基金项目(42502110)、国家资助博士后研究人员计划和中国博士后科学基金(BX20250435)和中国石油大学(北京)科研基金(2462025XKBH015)联合资助.

第一作者简介 张梓童,女,1994 年生,博士后,主要从事智能测井解释、油气人工智能研究. E-mail:cugb_zzt@163.com

***通讯作者** 周春栩,男,2000 年生,硕士研究生,主要从事石油与天然气地质研究. E-mail:cxzhou89@163.com

关键词 粒度球;岩性识别;机器学习;测井数据

中图分类号 P631

文献标识码 A

doi:10.6038/pg2026JJ0384

CSTR:32085.14.pg2026JJ0384

even surpass 99%. LogGBS demonstrates superior identification accuracy and generalization ability in all three types of reservoirs, providing a new idea for lithology identification based on well logs under complex geological conditions and contributing to the improvement of oil and gas exploration and development efficiency.

Keywords Granular ball; Lithology identification; Machine learning; Well Logs

0 引 言

岩性识别作为油气勘探开发领域的核心基础任务,其准确性直接决定了储层评价、资源潜力评估及开发方案优化的可靠性.随着勘探目标向复杂地质环境延伸,常规储层与非常规储层并存、岩性组成高度异质的勘探场景日益增多,传统基于人工经验的测井曲线解释方法,因难以量化刻画测井参数与岩性类别间的非线性映射关系,在识别精度与效率上已难以满足实际需求.

机器学习技术的兴起为测井岩性分类提供了新的解决方案(王丽媛等,2025;Zhang *et al*.,2025;Feng *et al*.,2025),k 近邻(kNN)(赵彤彤等,2018;陈玉林等,2020;孙岿,2022)、支持向量机(SVM)(牟丹等,2015;苏赋等,2020)、随机森林(RF)(王光宇等,2021)及各类提升树模型(Xie *et al*.,2018,2021;宋梓豪等,2024;赵然磊等,2025)等方法被广泛应用于岩性识别任务中.然而,现有方法仍存在显著局限(韩启迪等,2019;王恒等,2021;孙岿,2022;刘国强等,2022;赵逢达等,2025):线性模型(如逻辑回归)无法适配测井响应与岩性的复杂非线性关联;基于距离的模型(如 kNN)对测井数据中的局部噪声敏感,易导致薄层岩性误判;树模型虽能处理非线性问题,但在厚层均质段易出现过拟合,且对特征重叠严重的岩性区分能力有限;BP 神经网络则依赖大量高质量标注数据,在低密度测井数据场景下鲁棒性不足.这些缺陷使得现有模型在跨地质环境(如碎屑岩、碳酸盐岩、火山岩等不同储层类型)的泛化能力上存在明显短板,难以满足多区域、多储层类型的岩性识别需求.

为突破现有机器学习方法在测井岩性分类中的瓶颈,本研究提出一种基于粒度球空间的测井岩性识别方法(LogGBS),通过自适应粒度球生成器(AGBG)将原始测井数据映射为带有岩性标签的粒度球,并结合 Granular-Ball kNN(GBkNN)分类器实现准确岩性分类.该方法创新性地引入"粒度球纯度"作为核心控制参数,可动态调整粒度球数量与决策边界,在保持测井数据全局统计特征的同时抑制局部噪声干扰,有效适配不同储层的地质复杂性.鄂尔多斯盆地与松辽盆地作为我国重要的油气资源勘探基地,其储层类型涵盖碎屑岩、碳酸盐岩及火山岩,兼具常规与非常规储层特征,且储层表现出岩性过渡特征复杂、受岩溶作用影响岩性分布不均等特点,这三类储层均面临勘探难度大、岩性识别精度要求高的挑战,成为验证新方法普适性的理想研究对象.

本研究以鄂尔多斯盆地杭锦旗气田(碎屑岩储层)、苏里格气田(碳酸盐岩储层)及松辽盆地徐家围子断陷徐深气田(火山岩储层)的测井数据为研究对象,对比实验验证 LogGBS 与八种传统机器学习和深度学习方法的性能差异,并系统分析粒度球纯度对分类精度的影响规律.研究成果旨在为复杂地质条件下的测井岩性识别提供新的范式,提升油气资源勘探开发的效率与准确性,同时为机器学习技术在地质工程领域的深度应用提供理论与实践参考.

本文提出的 LogGBS 方法包括三个主要贡献:(1)基于粒度球空间的测井岩性识别方法,能够有效将测井数据映射到粒度球空间,利用球级结构表达岩性的空间分布特征,具有较强的数据非线性拟合能力;(2)自适应粒度球生成策略,结合纯度提升与去重叠约束的动态粒度调整机制替代传统迭代聚类,从而增强不同岩性类别之间的可分性;(3)系统评估 LogGBS 在三类经典储层(碎屑岩、碳酸盐岩、火山岩储层)中的识别性能,结果表明该方法在识别精度、稳定性与泛化能力方面均优于传统机器学习和深度学习模型,具有较大的理论价值和工程应用前景.

1 方法原理

以往基于测井数据的岩性识别方法常以单一深度点的数据为输入,存在信息碎片化、噪声鲁棒性差及忽略地层连续性的缺陷,且未能充分遵循"全局优先"这一认知原则,导致未能很好地捕捉地层岩性的连续沉积序列,忽略薄互层岩性的过渡等.尽管

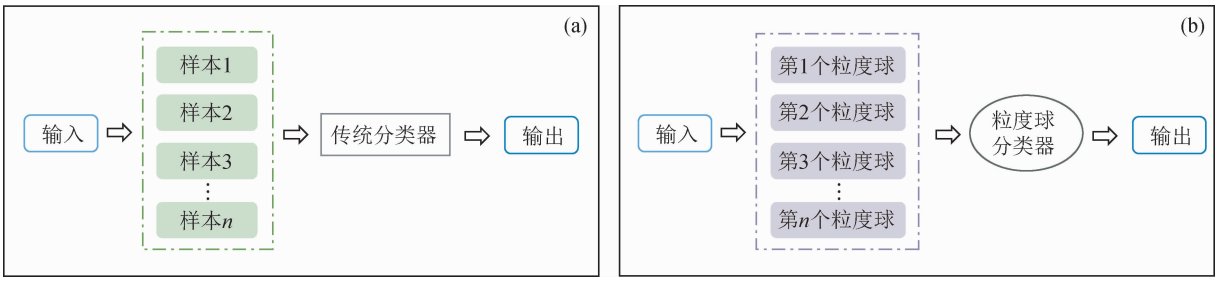


图 1　不同分类器的比较(Xia *et al.*,2022a)
(a)传统分类器;(b)粒度球分类器,由多个样本组成的基本粒度球作为输入.
Fig. 1　Comparison of different classifiers(Xia *et al.*,2022a)
(a) Traditional classifier; (b) Granular-ball classifier, the basic granular ball composed of multiple samples as input.

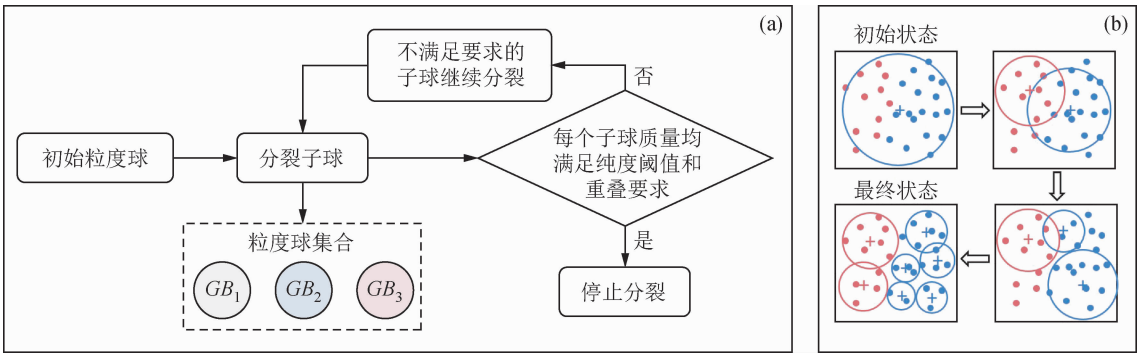


图 2　粒度球生成过程
Fig. 2　Process of the GB generation

机器学习技术有了一定进展,许多模型仍存在以上问题,如以单一深度点的测井数据为输入单元,如图 1a 所示. 此外,由于缺乏对更大范围测井信息的整合,此类方法在应对局部噪声干扰(如测井仪器误差、地层裂缝或异常测点)时表现出较低的鲁棒性. 受人类视觉认知中"大范围优先"机制的启发(Chen, 1982)——人类大脑对事物的认知遵循"从粗到细、从整体到局部"的规律,例如识别地层岩性时,先感知储层整体沉积序列,再细化至薄层岩性过渡特征. 因此,我们采用了一种基于粒度球(Granular-Ball, GB)思想的分类方法替代传统点状输入模式,以更有效地利用测井数据在深度方向上的空间相关性,如图 1b 所示. 通过对连续深度的测井数据样本进行空间范围上的划分与聚合,将多个特征相似的同类测井样本构建成粒度球,并以粒度球为基本输入单元执行后续的分类任务. 相比于传统的分类器,这种粒度级输入的分类器能够更好地捕捉储层岩性的渐变或突变信息,减少单点噪声对分类结果的干扰,从而使得分类结果更加符合实际地层的分布规律.

基于上述分析,本研究提出了一种基于粒度球空间(Logging Granular-Ball Space, LogGBS)的测井岩性分类框架,在全局优先认知结构下,结合自适应粒度球生成器(Adaptive Granular-Ball Generator, AGBG)与多粒度特征融合策略,对测井数据进行空间范围上的分类和多层次特征提取,以有效提升复杂储层岩性识别的精度与鲁棒性. 下面将对其中的关键模块进行详细描述.

1.1　自适应粒度球生成器

AGBG 是 LogGBS 模型的核心模块,其设计灵感源于粒计算理论(Zadeh, 1996; Pedrycz and Succi, 2005; Salehi *et al.*, 2015; Xia *et al.*, 2026)与自适应优化机制(Xia *et al.*, 2022a, b, 2026). 具体而言,AGBG 通过粒度球半径和分布的自适应调整,捕捉测井数据中的全局模式和局部细节. 通过分裂过程优化粒度球分布,从而实现高效多粒度建模,粒度球生成过程示意图如图 2 所示,其具体处理步骤如下:

(1)数据粒度化映射

粒度球计算将数据空间划分为称为粒度球的球形粒子,并用于替换数据点,参与分类器的学习过程. 设原始测井数据集为 $D=\{x_i,L_i\}_{i=1}^{N}$,其中 $x_i\in\mathbb{R}^d$ 为第 i 个测井样本点的 d 维特征向量,L_i 为对应的岩性标签,N 为总样本数. 将原始测井数据 D 映射到粒度球空间,获得最终生成的总粒度球集合

$\{GB_j\}_{j=1}^{M}$,其中 $M \ll N$ 为粒度球的总个数. 每个测井样本点 x_i 均可以获得其所隶属的粒度球 GB_j,且每个粒度球 GB_j 均由“中心—半径—标签”三元组定义,其中心 C_j 和半径 r_j 的具体计算公式如下:

$$C_j = \frac{1}{|GB_j|}\sum_{x_i \in GB_j} x_i, \tag{1}$$

$$r_j = \frac{1}{|GB_j|}\sum_{x_i \in GB_j} \|x_i - C_j\|, \tag{2}$$

其中 $|GB_j|$ 为粒度球 GB_j 内包含的测井样本数量,C_j 表示此粒度球的球心,计算方式为球内所有样本的特征均值,确保表征球内样本的整体分布. r_j 则为粒度球半径,采用内部样本与球心之间的平均欧式距离定义. 在粒度球的生成过程中通过半径 r_j 来控制粒度球的大小,使其不容易受到离群样本的影响. 通过采用这种粒度球生成方法在对测井样本进行自适应分组,有效避免了传统“最大距离”定义易受测井异常值(如仪器误差、地层裂缝导致的单点跳变)对球表征的干扰,增强粒度球对局部数据分布的稳健表征.

在步骤(1)中,首先将完整测井数据集视为初始粒度球 GB_0,其纯度 T_0 为全局多数类样本占比,计算如下:

$$T_0 = \frac{|GB_0^{\text{majority}}|}{|GB_0|}, \tag{3}$$

其中 $|GB_0| = N$ 代表数据集中所有样本的个数,$|GB_0^{\text{majority}}|$ 则为初始球内占比最高的岩性样本数,$T_0 = (0.5,1]$,即初始球至少含两类岩性. 理想状态下,期望单个粒度球 GB_j 中只包含一种岩性的样本点,即 $T_j = 1$.

(2)动态粒度调整

为了有效捕捉测井数据在深度方向上的全局沉积模式和局部岩性突变特征,AGBG 在粒度球生成过程中引入了“动态粒度调整”机制,通过对粒度球的自适应分裂与优化来提升球级结构对岩性边界的刻画能力. 具体地,在纯度阈值 $T_{\text{threshold}}$ 的限制下进行迭代划分,将较大规模的粒度球分解为多个较小的粒度球. 分裂触发条件为纯度不足 $T_j < T_{\text{threshold}}$,或存在异类粒度球重叠,异类粒度球重叠的判别条件如下:

$$\exists \|C_j - C_k\| \leqslant r_j + r_k,\ \text{s.t. label}(GB_j) \neq \text{label}(GB_k), \tag{4}$$

其中,C_j 和 C_k 分别为第 j 个和第 k 个粒度球的中心,r_j 和 r_k 分别为第 j 个和第 k 个粒度球的半径(表示测井样本的分布范围),$\|C_j - C_k\|$ 为两个粒度球中心间的欧氏距离(以量化粒度球内样本的空间邻近性),label(GB_j)和 label(GB_k)为粒度球的岩性类别标签(由球内多数类样本决定). 当两个异类粒度球(标签不同)的中心距离小于或等于其半径之和时,判定为边界重叠,此时需要进一步分裂或调整粒度球,以消除模糊边界提升分类精度,避免同一测井样本点被归属于多类岩性的问题.

分裂机制采用岩性标签驱动的单轮几何划分策略,当粒度球被判定为需要分裂时,分裂数量同球内岩性类别数 m 一致,针对每个类别,提取其在球内的全部样本,重新计算其中心与半径,计算过程同公式(1)、(2)一致. 需要注意的是,在进行粒度划分的动态调整过程中,子球 GB_j' 的纯度 T_j' 大小满足 $T_j' > T_j$,代表经动态粒度调整后得到的子粒度球纯度是提升的,岩性分类的精度也相应提高. 此外,本研究采用了单次划分替代传统 k-Means 迭代划分的方式,将时间复杂度从 $O(mnN)$ 降至 $O(N)$,显著提高了粒度球生成过程的计算效率.

(3)全局优化

接着,对步骤(2)生成的粒度球集合执行“去重叠—纯度校验”的全局优化,确保岩性边界清晰且覆盖全体测井数据. 若粒度球之间出现了粒度球重叠的现象,则需要判断粒度球中的纯度大小并执行进一步的划分,以改善重叠的粒度球使得决策边界更为清晰. 在此过程中,还需确保粒度球的纯度阈值满足在自适应下界以上. 这里,我们设置自适应粒度球生成过程中的纯度阈值下界为 T_0,即初始粒度球中的纯度大小.

具体地,我们在步骤(3)中首先进行去重叠操作. 对所有异类重叠粒度球(满足公式(4)),若其中一个球的纯度已达 $T_{\text{threshold}}$,则调整另一个球的半径(缩小至与前者无重叠,保持中心不变);若两者纯度均未达标,则对纯度更低的球执行二次分裂,直至无异类重叠;接着,进行纯度校验. 计算所有粒度球的纯度,若存在 GB_j 满足 $T_0 < T_j < T_{\text{threshold}}$(其中 T_0 为初始粒度球纯度),则重新分裂该球,确保粒度球的决策边界清晰且分布均匀,同时还能避免因过度分裂导致的岩性信息碎片化. 在执行步骤(3)的过程中,在满足以下两个条件时停止迭代:①所有粒度球的纯度 $T_j \geqslant T_{\text{threshold}}$;②无任何异类粒度球满足公式(4)的重叠条件. 粒度球经过上述的全局优化处理,我们能够最终获得一个覆盖全体数据集的粒度球集合 $\{GB_j\}_{j=1}^{M}$,后续将以此粒度球集合作为不同岩性的分类依据.

值得注意的是,纯度阈值控制了粒度球分裂的敏感度,反映了模型对“岩性异质性容忍度”的地质

假设. 当粒度球纯度低于阈值而触发分裂时,意味着该空间范围内的岩性差异已超出可视为同一沉积相体或同一岩性单元的程度. 当纯度阈值取值较高(如≥0.9)时,模型倾向于生成数量更多、尺度更小、岩性更加单一的子球,这在地质意义上相当于对沉积单元进行更高分辨率的细分,适用于层内岩相刻画、薄互层识别的精细描述. 相反,当阈值设置偏低(如0.7~0.8)时,模型允许一定程度的岩性混合,其形成的粒度球通常尺度更大、类别更稳定,接近地质上可用于井间对比或储层级描述的层序单元.

因此,粒度球纯度阈值的选择应结合沉积环境复杂度、噪声水平以及任务目标综合确定. 其地质意义在于,纯度阈值决定了模型在"地质分辨率"与"统计稳定性"之间的平衡,使 LogGBS 能够从层内精细刻画到区域储层预测等不同应用场景中保持方法的灵活性和适应性.

1.2　GBkNN 分类

在获得 AGBG 生成的粒度球集合 $\{GB_j\}_{j=1}^{M}$ 的基础上,我们进一步以粒度球作为分类单元设计了适用于测井岩性分类的 GBkNN 分类器. 相比于传统以单一深度点数据为输入的方法而言,GBkNN 分类器能够有效结合局部与全局信息,增强对地层连续性和岩性变化的刻画能力. 同时,粒度球的聚合机制显著减少了噪声的影响,提高了分类的鲁棒性.

具体地,根据 1.1 节中给出的粒度球生成过程中的相关描述,我们可以总结出粒度球的相关定义及其距离度量. 对于给定的粒度球 GB_j 可表示为由粒度球中心、粒度球半径和对应的粒度球标签构成的三元组 (C_j, r_j, L_j),其中粒度球标签由粒度球内多数岩性类型样本决定(例如,若粒度球中 70% 为砂岩,则 L_j = 砂岩). 在粒度球空间上进行分类时的距离度量计算如下:

$$Dis(P, GB_j) = \max(\|P - C_j\| - r_j, 0), \tag{5}$$

其中,$Dis(P, GB_j)$ 表示查询点 P 与粒度球 GB_j 之间的距离. 这意味着若 $Dis(P, GB_j) = 0$ 表示 P 落在 GB_j 的覆盖范围内,直接继承 L_j 作为 P 的预测标签;若 $Dis(P, GB_j) > 0$ 则表示 P 位于 GB_j 外部,需计算与所有 GB_j 的距离.

那么,当采用 GBkNN 分类器进行分类时,我们需要计算测试集中的每个查询点(即单个测井样本点)P_k($k = N_{tr} + 1, \cdots, N$,$N_{tr}$ 为训练集样本数)与前述学习得到的粒度球集合 $\{GB_j\}_{j=1}^{M}$ 之间的关系. 对于查询点 P_k 预测标签由"距离最小的粒度球"决定,若存在多个粒度球距离相同(如测井数据的岩性过渡带),则取这些粒度球中多数类标签作为最终结果,以此来获得测试数据集中每个样本点对应的岩性标签. 其对应的数学表达如下:

$$L_k = L_j, \text{ where } j = \arg\min_{j \in [1, M]} Dis(P_k, GB_j), \tag{6}$$

其中 L_k 表示查询点 P_k 对应的预测标签. 对应的 GBkNN 分类算法如算法 1 所示.

算法 1　GBkNN 分类算法

输入	原始测井数据中的训练集输入数据集$\mathcal{D}$、测试集中$\mathcal{D}$位于之外的查询点集$\mathcal{P}$、由自适应粒球生成器(AGBG)模块得到的粒球集合$\{GB\}_j (j = 1, \cdots, M)$;
输出	测试集中查询点 $P_k \in \mathcal{P}(k = N_{tr} + 1, \cdots, N)$ 的岩性标签 L_k
1	计算 P_k 到每个粒度球 GB_j 的距离;
2	查询点 P_k 的岩性标签 L_k 等于与其距离最小的粒球的标签.

通过构造 GBkNN 分类器,能够将粒度球的特征输入到 kNN 分类模型中,有效提升了模型对粒度球岩性类别的预测性能,同时将结果映射到储层空间,能够生成更加符合地质学规律的岩性分布结果. 这种 GBkNN 分类方法不仅能显著提高 LogGBS 分类精度和效率,还能增强模型的地质学意义和可解释性,在复杂储层岩性分类任务中具有巨大的潜力. 提出的 GBkNN 方法的优势可总结如下:

(1)低时间复杂度:GBkNN 的时间复杂度为 $O(n)$,其中 n 为粒度球数量,远低于传统 kNN 的 $O(n^2)$. 通过自适应生成的粒度球直接确定测井样本点的标签,无需手动优化参数,避免因参数选择偏差导致的岩性分类错误.

(2)强噪声抑制与特征表示能力:GBkNN 通过纯度阈值过滤稀疏分布的异常点(如仪器噪声),同时利用粒度球的边缘分布生成隐含分类边界,结合多粒度特征(如中心、半径)量化储层宏观岩性模式与微观变异,增强对复杂结构的表征能力.

(3)稳定的分类决策:查询点的预测标签由与其欧氏距离最近的粒度球标签决定,粒度球标签由

其内部多数类样本定义.相较于传统 kNN 依赖局部近邻样本点统计分布的特点,GBkNN 通过全局粒度球映射能够实现更稳定的分类决策.

1.3 LogGBS 架构

本文提出了一种基于 LogGBS 的岩性分类模型,旨在通过多粒度建模实现测井数据表征的层次化迁移,从而提高岩性分类的准确性、鲁棒性及可解释性.所提出的 LogGBS 模型的整体处理流程如图 3 所示,由三个关键模块构成:

1)数据处理模块

在数据处理模块中,通过数据收集、数据预处理和数据划分操作,为 LogGBS 的后续建模提供了高质量、标准化的输入数据.具体而言:

(1)数据收集:收集地球物理测井数据,这些数据包含多个关键测井参数,如自然伽马测井(GR)、密度测井(DEN)、声波时差测井(AC)、补偿中子测井(CNL)、深侧向电阻率测井(LLD)、浅侧向电阻率测井(LLS),能够全面表征地下地层的物理特性.

(2)数据预处理:①深度匹配:对不同测井曲线进行深度对齐,确保数据的一致性,为后续分析奠定基础;②数据清洗:剔除测井数据中的异常值(如因仪器故障导致的极端值),采用插值等方法填补缺失值,有效提升数据质量;③特征提取与选择:保留对岩性识别具有重要意义的关键参数,去除冗余特征,降低数据维度,减少计算量;④数据归一化:消除不同测井参数量纲差异带来的影响,为后续建模提供标准化输入,避免因参数量级差异导致模型偏差.

(3)数据划分:将数据集划分为训练集(80%)和测试集(20%),以保证模型训练和评估的独立性,避免因数据重叠导致的评估偏差.

2)基于粒度球的测井建模

本文引入自适应粒度球生成器(AGBG),结合测井曲线特征及地质背景信息,实现对测井数据的动态分割与多粒度建模(详见 1.1 节).具体而言,AGBG 自适应地生成粒度球集合,在地层连续性较高的区域,生成较大的粒度球,以捕捉全局模式;在岩性变化剧烈的区域,生成较小的粒度球,以刻画局部细节.这一操作使得模型能够在不同尺度上提取储层的全局模式与局部细节信息,进而提高模型对复杂岩性过渡带的识别能力.

此外,在粒度球生成完成后,我们在粒度层面上设计了 GBkNN 分类器来进行后续的岩性识别.相比传统方法,该分类器能够综合利用不同尺度的测井特征,学习储层岩性的全局模式和局部细节,提升模型的噪声鲁棒性,从而使得分类结果的鲁棒性和地质学解释能力得以有效提升.

为了验证 LogGBS 的有效性,本文在多个油藏测井数据集上进行了实验,并与八种传统机器学习和深度学习方法(基线模型)进行对比,包括 k 近邻(kNN)、支持向量机(SVM)、随机森林(RF)、BP 神经网络(BP)、提升树模型(XGBoost、LGBM、CatBoost)、卷积神经网络(CNN).所有对比模型与 LogGBS 模型使用相同的训练集和测试集,且训练过程保持一致,以确保对比实验的公平性,从而准确评估 LogGBS 模型的性能优势.

3)模型评估

最后,通过定量指标与对比实验验证 LogGBS 模型在复杂储层中的分类有效性和泛化性.具体而言,我们采用了准确率和混淆矩阵等评价指标来论证分析 LogGBS 模型的整体分类性能和在不同岩性上的分类精度.同时,通过在不同类型油藏数据上评估模型的表现,验证 LogGBS 模型对多样化储层地质特征的适应性,确保其在实际复杂地质条件下的应用价值.

2 地质背景

为验证 LogGBS 方法在基于测井数据识别岩性方面具备跨地质环境的普适性,本研究选取鄂尔多斯盆地与松辽盆地内三类典型油气储层——碎屑岩、碳酸盐岩和火山岩储层作为研究对象.上述研究对象兼具常规与非常规储层,且分布广泛、岩性组成复杂、勘探难度较高,因此实现精准的岩性识别对验证 LogGBS 方法在不同地质背景下的可迁移性、提升油气资源开发效率与准确性具有重要意义.

碎屑岩储层测井数据来自鄂尔多斯盆地杭锦旗气田,石炭系太原组(C_3t)、二叠系山西组(P_1s)下石河子组(P_1x)的开源数据集(Xia *et al.*, 2019).杭锦旗地区位于鄂尔多斯盆地北缘过渡带,横跨伊盟隆起和伊陕斜坡两个构造单元(图 4a),面积约 9000 km^2,是盆地内上古生界天然气勘探的重要靶区,该地区上古生界主要烃源岩为太原—山西组煤系烃源岩(主要是Ⅲ型干酪根),呈现南厚北薄的平面分布特征(纪文明等,2013;倪春华等,2018;Qin *et al.*, 2025).上古生界碎屑岩系气田主力储层发育于石炭系太原组、二叠系山西组及下石盒子组.石河子组下段砂岩是该地区的主要储层,下伏的山西和太原组是次生组(图 4b).

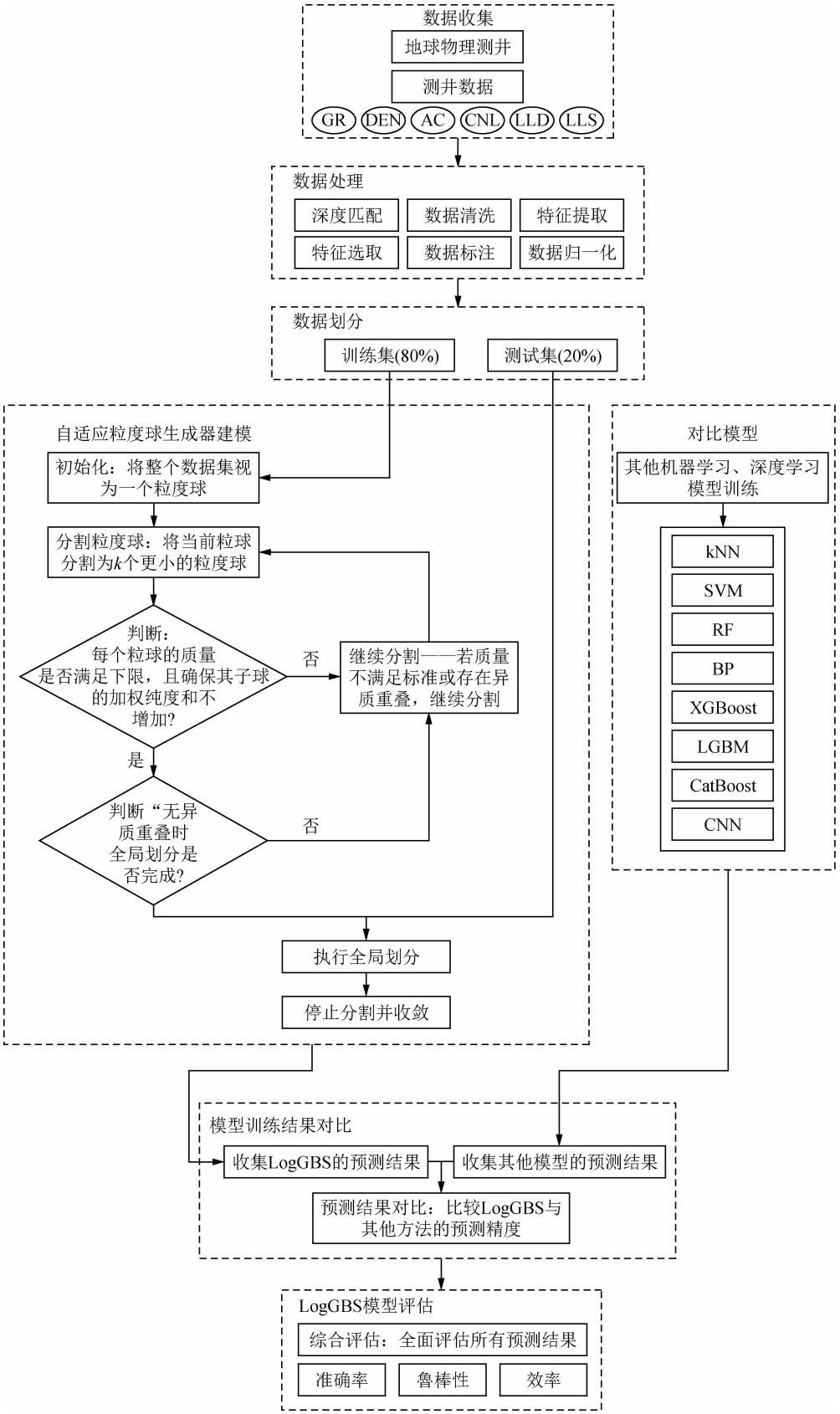


图 3　LogGBS 模型流程图

Fig. 3　LogGBS model flow chart

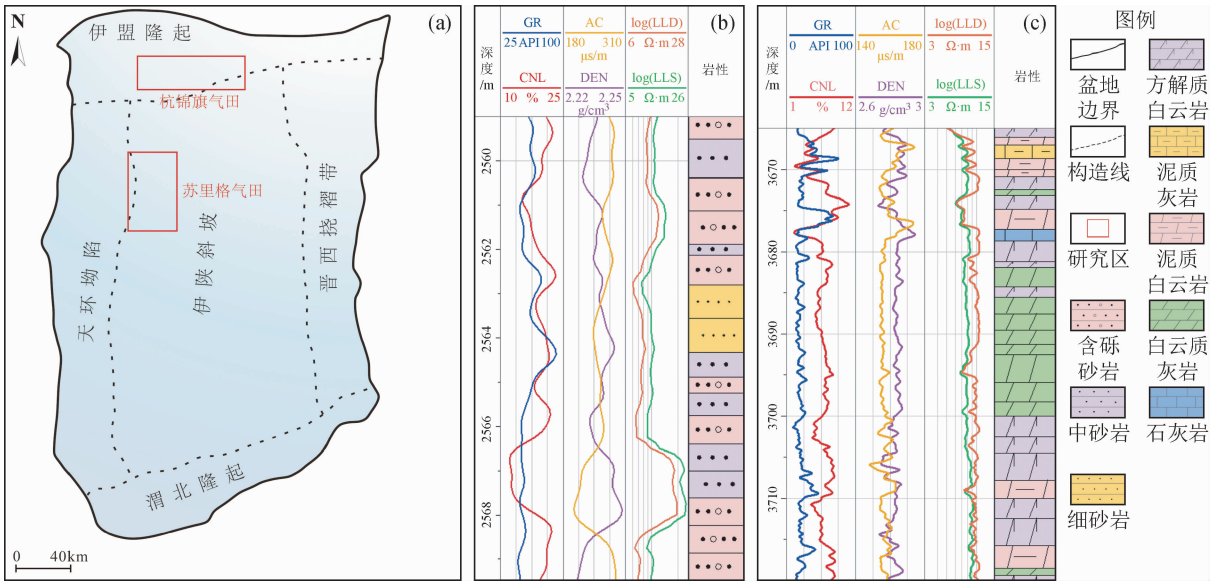


图 4 碎屑岩储层与碳酸盐岩储层研究区位置及测井曲线

(a) 研究区位置;(b) 碎屑岩储层测井曲线;(c) 碳酸盐岩储层测井曲线.

Fig. 4 The location and logging profile of the study area of clastic and carbonate reservoirs

(a) The location of the study area; (b) Logging profile of the clastic reservoir; (c) Logging profile of the carbonate reservoir.

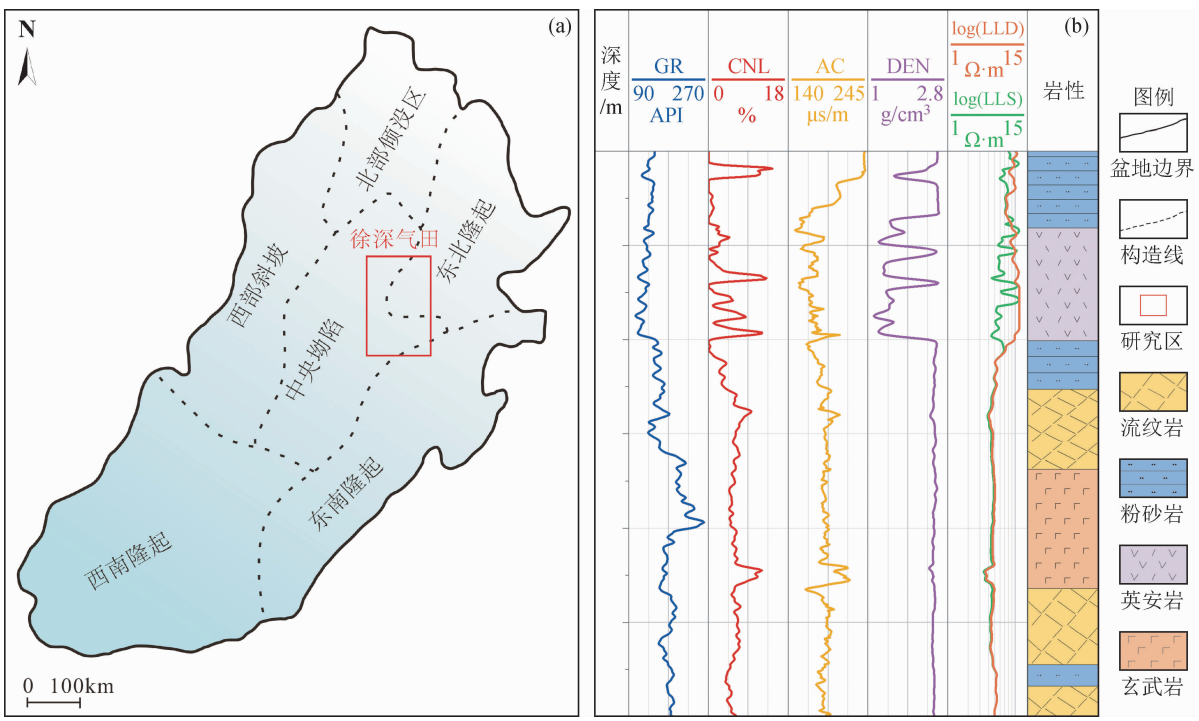


图 5 火山岩储层研究区位置及测井剖面

Fig. 5 Study area location and logging profile of volcanic reservoir

碳酸盐岩储层测井数据来自鄂尔多斯盆地中部苏里格气田马家沟组,主要集中在马五段(图 4c).苏里格气田位于鄂尔多斯盆地伊陕斜坡中北部,上古生界自上而下沉积太原组、本溪组和马家沟组(Xie *et al.*,2023).马家沟组地层自下而上分为马一段至马六段,其中马五段为主要含气层段,属于蒸发台地—局限台地相沉积,马五段发育典型的岩溶碳酸盐岩储层,主要为白云岩,夹少量泥质岩、灰岩及蒸发岩,是苏里格气田碳酸盐岩气藏的核心发育区(周进高等,2011;沈安江等,2015;Shen *et al.*,2015;Bai *et al.*,2016).

火山岩储层测井数据来源于松辽盆地北部徐家

围子断陷徐深气田(图 5),营城组,主要是营一段的数据. 徐家围子断陷位于松辽盆地中央坳陷区,形成于晚侏罗—早白垩世早期,是盆地内最大的中生代陆相火山—沉积盆地,并且是其中规模最大的含气断陷之一(张居和等,2009). 徐家围子断陷的构造演化可分为 5 个阶段,初始断陷期(火石岭期)、强烈断陷期(沙河子期)、挤压反转期(沙河子期末)、断坳转化期(营城期)和坳陷期(登娄库期)(刘国平等,2016). 深层由下而上依次发育火石岭组、沙河子组、营城组(营二段发生了区域缺失)、登娄库组、泉头组地层,其中有三套火山岩气藏储集层,即营城组一段、营城组三段和火石岭组(Cai *et al.*,2012). 该岩系主要由玄武岩和安山岩构成,其次发育火山集块岩、熔结流纹岩等岩石类型.

3　数据集概况及实验设置

在本节中,我们将所提出的 LogGBS 方法应用于上述研究区的测井数据集,以验证其有效性与实用性. 实验流程主要包括以下四个部分:第一,利用 AGBG 将训练集中的原始测井数据点映射为一系列 GB 球形粒子. 在预设纯度阈值 $T_{\text{threshold}}$ 的约束下,通过迭代划分形成带有岩性标签的粒度球;第二,采用 GBkNN 分类器,输入测试集在已形成的粒度球空间中实现岩性分类;第三,从 0.7 开始,以 0.05 为间隔设置不同的粒度球纯度阈值,系统比较 LogGBS 不同设置下的预测精度与鲁棒性;第四,将 LogGBS 与 kNN、SVM、CNN 等八种主流机器学习和深度学习方法进行对比实验,在三类不同储层的测井数据集中开展测试,以验证该方法在实际应用中的显著优势.

3.1　数据集介绍

本研究选取鄂尔多斯盆地与松辽盆地的三类典型储层测井数据作为研究对象,具体包括:(1)碎屑岩储层测井数据集(HGF),来源于鄂尔多斯盆地杭锦旗气田;(2)碳酸盐岩储层测井数据集(SGF),取自鄂尔多斯盆地中部苏里格气田马家沟组;(3)火山岩储层测井数据集(XGF),数据源自松辽盆地北部徐家围子断陷营城组,以营一段为主. 在构建用于 LogGBS 岩性预测的数据集时,共采用了六类测井曲线:GR、DEN、CNL、AC、LLD、LLS. 数据采样间隔为 0.125 m,测井数据集的详细统计信息见表 1,统计指标包括:平均值(Mean)、标准差(Std)、最小值(Min)、50% 分位数(50%)和最大值(Max).

表 1　测井数据统计信息

Table 1　Logging data statistics information

数据集	样本数	统计指标	测井参数					
			GR (API)	DEN (g/cm^3)	CNL (%)	AC ($\mu s/m$)	LLD ($\Omega\cdot m$)	LLS ($\Omega\cdot m$)
HGF	1238	Mean	80.80	2.45	22.11	245.70	47.47	55.41
		Std	35.85	0.17	9.55	28.81	178.53	241.34
		Min	29.88	1.38	8.24	189.28	5.35	5.54
		50%	71.30	2.49	18.90	243.06	17.46	18.10
		Max	231.87	2.78	63.85	513.01	2322.73	3279.91
SGF	2881	Mean	31.50	2.82	6.69	162.73	11554.89	2114.52
		Std	21.85	0.06	2.88	8.64	13325.32	2147.50
		Min	9.30	2.39	0.03	146.43	16.03	18.20
		50%	22.42	2.83	6.70	160.49	5142.99	1338.50
		Max	179.71	2.98	19.69	215.53	40000.00	14816.15
XGF	3534	Mean	155.73	2.47	3.97	188.24	16393.40	2267.61
		Std	24.27	0.18	4.50	20.82	32213.24	5501.05
		Min	77.15	1.13	0.01	148.85	17.86	12.82
		50%	150.46	2.53	2.72	185.38	428.76	316.67
		Max	261.24	2.69	18.41	262.97	104989.49	92130.85

每个测井数据样本均标注有对应的岩性类别标签,三类储层样本的岩性标签如表 2 所示. 这些标签是基于岩石学特征、沉积构造、岩心观察、微观分析以及成像测井等多源信息综合判识的结果. 图 6 展示了 HGF、SGF 和 XGF 数据集中不同测井参数的交会图. 在该图中,不同岩性类别的测井概率密度曲线显示于对角线位置,其分离程度反映了各类别之间的区分能力(Zhao *et al.*,2023).

表 2 岩性分布统计

Table 2 Lithology distribution statistics

类别	HGF			SGF			XGF		
	岩性	样本数	占比(%)	岩性	样本数	占比(%)	岩性	样本数	占比(%)
C1	煤	14	1.13	石灰岩	358	12.43	粉砂岩	1926	54.50
C2	粗砂岩	207	16.72	白云质灰岩	122	4.23	砾岩	572	16.19
C3	细砂岩	146	11.79	泥质灰岩	386	13.40	流纹岩	835	23.63
C4	泥岩	248	20.03	白云岩	383	13.29	玄武岩	103	2.91
C5	中砂岩	206	16.64	灰质白云岩	814	28.25	英安岩	98	2.77
C6	含砾砂岩	370	29.89	泥质白云岩	818	28.39			
C7	粉砂岩	47	3.80						
	总计	1238		总计	2881		总计	3534	

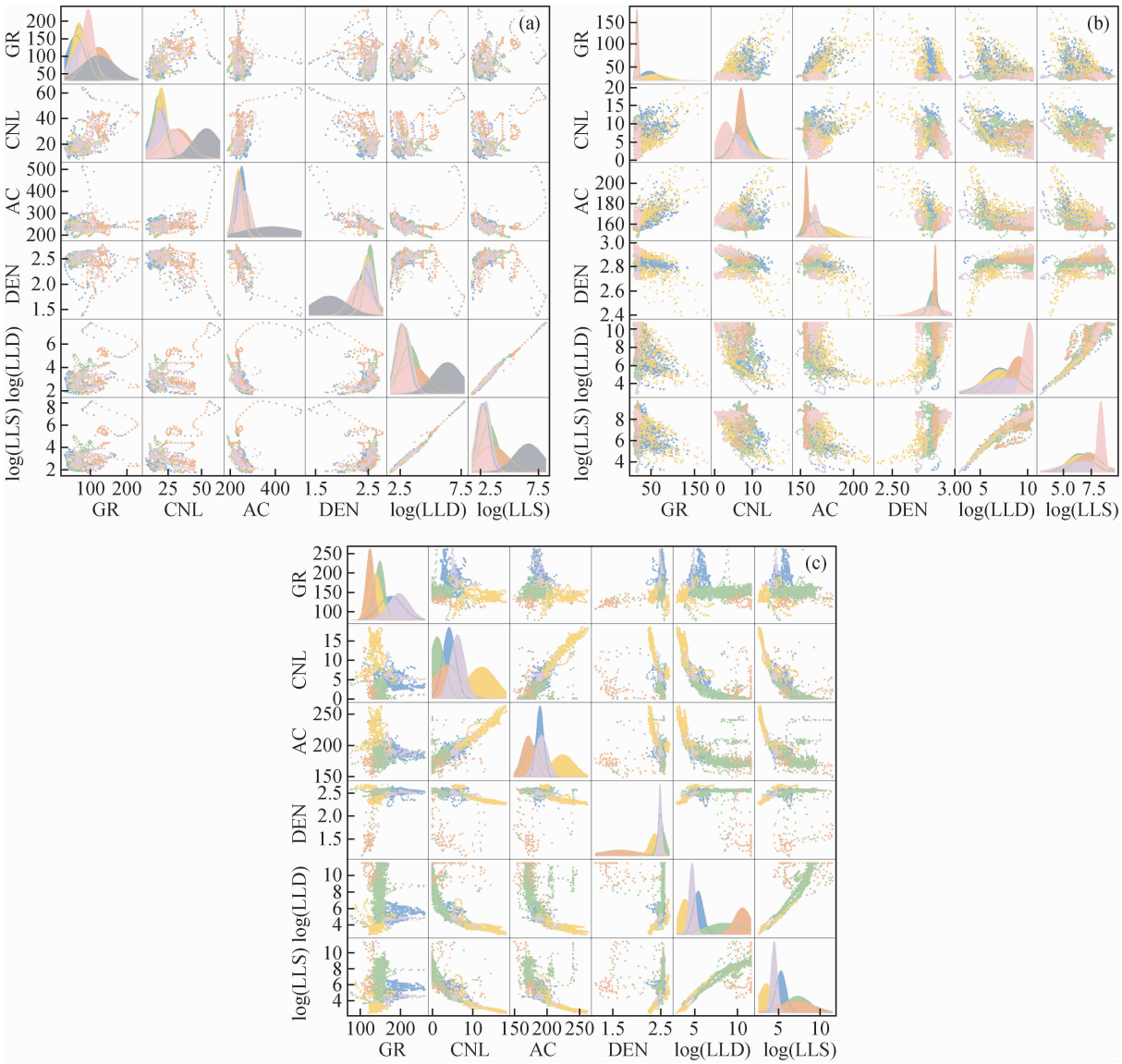


图 6 测井参数交会图

(a) HGF 数据集;(b) SGF 数据集;(c) XGF 数据集.

Fig. 6 Cross plots of different logging parameters

(a) HGF dataset; (b) SGF dataset; (c) XGF dataset.

以 HGF 数据集为例,煤表现出较高的 AC、LLD、LLS 和 CNL,以及较低的 DEN;泥岩则具有较高的 GR 和 AC.尽管这些特征使煤和泥岩在一定程度上区别于其他岩性类别,其测井概率密度曲线仍存在显著重叠,这些复杂特征给岩性识别带来挑战.

在不同类型储层中,准确识别岩性对沉积相分析与储层表征至关重要.然而,测井曲线特征之间的重叠表明测井参数与岩性之间存在复杂的非线性关系.本文提出的 LogGBS 方法以多粒度球作为输入,并通过平均距离定义半径,能够在保持测井数据全局统计特征的同时,天然抑制局部噪声(如个别异常测井值)的干扰.

3.2　实验设置及评价指标

本研究的模型基于 Python3.10 构建,所有实验均在以下硬件环境中进行:IntelCorei7-11700K CPU、128GB RAM 以及 NVIDIA GeForce RTX3090 24GB GPU.为在保证机器学习模型准确性的同时避免过拟合,我们将三个储层数据集按 8∶2 的比例划分为训练集和测试集.

为评估并比较不同模型的性能,采用混淆矩阵及其衍生的多项指标进行综合评价,具体包括总体准确率(OA)、平均准确率(AA)和 Kappa 系数.其中,OA 表示正确分类样本占总样本的比例;AA 为各类别分类准确率的平均值;Kappa 系数用于评估分类结果与真实标签之间的一致性,可作为分类可靠性的度量.以上三项指标的数值越高,表明分类性能越好,结果越可靠.

4　结果与讨论

AGBG 是 LogGBS 方法中的核心组成部分,其核心操作在于对粒度球进行划分.为衡量划分后粒度球的质量,本研究采用纯度这一指标,即一个粒度球中多数类样本所占的比例.该参数是控制粒度球划分过程的关键,对模型的识别性能与计算效率具有重要影响.随着纯度取值的提高,所生成的粒度球数量逐步增多,且整个过程具有良好的收敛性,可保证不同类别的粒度球决策边界互不重叠.

在 HGF、SGF 和 XGF 三类储层的测井数据集中,通过在 0.7 至 0.95 范围内以 0.05 为步长设置不同纯度,分别评估其分类性能及最大粒度球数(表 3~5).三类储层在不同纯度下的性能结果如图 7 所示.可以看出,随着纯度的提高,分类精度总体呈上升趋势.当纯度不低于 0.9 时,所有数据集的 AA 均达到 90% 以上,说明高纯度条件下生成的粒度球能够更充分捕捉岩性的全局模式与局部差异,也反映出 LogGBS 方法在高纯度下的稳定性和泛化能力.

表 3　在不同粒度球纯度阈值设置下 LogGBS 算法对 HGF 数据集的分类效果,最佳结果以粗体标出

Table 3　The classification performance of the LogGBS algorithm on the HGF dataset under different purity threshold settings of granular ball, the best results are shown in bold

粒度球纯度	0.7	0.75	0.8	0.85	0.9	0.95
C1	64.29	57.14	50.00	85.71	92.86	**100.0**
C2	78.26	85.51	86.96	93.72	89.37	**100.0**
C3	83.56	87.67	89.73	93.15	93.84	**100.0**
C4	89.92	96.37	95.16	97.98	97.18	**98.75**
C5	83.98	83.01	88.84	94.66	93.20	**100.0**
C6	89.46	92.16	95.68	**97.57**	96.49	97.21
C7	87.23	95.74	**100.00**	**100.00**	**100.00**	97.87
OA(%)	85.70	89.58	91.92	94.67	95.96	**98.85**
AA(%)	82.39	85.37	86.62	94.69	94.70	**99.12**
Kappa(×100)	82.02	86.92	89.86	93.32	94.94	**98.57**
最大粒度球数	300	344	404	449	483	498

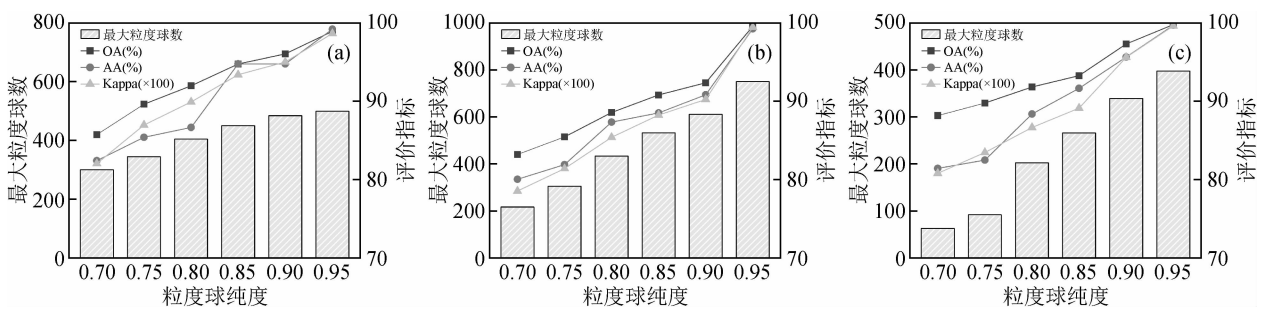


图 7　实验数据集评估指标分布图
(a)HGF 数据集;(b)SGF 数据集;(c)XGF 数据集.
Fig.7　Experimental dataset evaluation index distribution maps
(a) HGF dataset; (b) SGF dataset; (c) XGF dataset.

表 4 在不同粒度球纯度阈值设置下 LogGBS 算法对 SGF 数据集的分类效果,最佳结果以粗体标出

Table 4 The classification performance of the LogGBS algorithm on the SGF dataset under different purity threshold settings of granular ball, the best results are shown in bold

粒度球纯度	0.7	0.75	0.8	0.85	0.9	0.95
C1	92.72	93.00	94.40	94.96	96.36	**99.72**
C2	58.68	61.98	79.34	76.03	82.64	**97.52**
C3	85.19	82.60	83.64	85.71	87.53	**98.44**
C4	74.87	77.98	88.34	89.38	91.19	**100.00**
C5	81.97	87.94	88.79	91.96	93.79	**99.76**
C6	86.79	87.88	89.44	92.92	93.28	**99.76**
OA(%)	83.19	85.43	88.53	90.77	92.32	**99.52**
AA(%)	80.04	81.90	87.32	88.49	90.80	**99.20**
Kappa(×100)	78.54	81.40	85.38	88.21	90.20	**99.39**
最大粒度球数	217	305	433	532	610	**749**

表 5 在不同粒度球纯度阈值设置下 LogGBS 算法对 XGF 数据集的分类效果,最佳结果以粗体标出

Table 5 The classification performance of the LogGBS algorithm on the XGF dataset under different purity threshold settings of granular ball, the best results are shown in bold

粒度球纯度	0.7	0.75	0.8	0.85	0.9	0.95
C1	92.37	92.42	94.69	94.90	98.14	**99.69**
C2	92.47	93.35	95.97	95.10	97.90	**99.47**
C3	79.76	85.88	84.35	89.06	95.76	**100.00**
C4	51.96	46.08	83.33	85.29	89.22	**99.02**
C5	90.72	94.85	83.51	93.81	96.91	**100.00**
OA(%)	88.17	89.75	91.80	93.23	97.25	**99.72**
AA(%)	81.46	82.51	88.37	91.63	95.59	**99.64**
Kappa(×100)	80.83	83.48	86.65	89.11	95.56	**99.55**
最大粒度球数	64	93	203	266	339	**397**

具体而言,在 HGF 数据集中,当粒度球纯度从提升至 0.8 到 0.85 时,AA 与 Kappa 系数随纯度提升的幅度较大,分别从 86.62% 提升至 94.69%,以及从 89.86 提升至 93.32. 在 SGF 和 XGF 数据集中,LogGBS 所得的 OA 和 AA 与纯度呈现显著正相关,纯度升至 0.95 时,两类指标均超过 99%. 由此可见,LogGBS 在处理火山岩和碳酸盐岩储层时表现出优越的岩性识别能力;而在碎屑岩储层中,因岩性过渡特征复杂,单一纯度下的完全区分存在一定难度,但借助多粒度特征融合,该方法仍可实现有效的岩性预测.

为验证本研究提出的 LogGBS 模型在储层岩性识别中的优越性能,在固定训练集与测试集划分的前提下,引入了八种传统机器学习和深度学习方法作为基线模型进行对比实验. 同时,根据前文中关于纯度与分类精度正相关性的结论,将 LogGBS 模型的粒度球纯度阈值设定为 0.95,以充分发挥其识别能力. 基于混淆矩阵及三项性能评价指标,对各模型在 HGF、SGF 和 XGF 三类储层测井数据集上的岩性预测结果进行了定量比较与分析.

如图 8 和表 6 所示,展示了九个模型在 HGF 数据集上测试结果的混淆矩阵及实验结果对比表. 从整体分类性能来看,与其他模型相比,SVM 表现较差,kNN 的性能有所改善,但对细砂岩(C3)和中砂岩(C5)的识别准确率仅约为 82%. RF 的整体识别能力显著优于前述两者,然而对煤(C1)的分类准确率仍仅为 92.86%,BP、XGBoost、LGBM 和 CatBoost 模型则表现出相近的综合性能. 深度学习模型 CNN 由于其较为优异的非线性特征提取能力取得了次优的 OA(98.84%),相比之下,LogGBS 模型以微弱优势取得了最高的 OA(98.85%),并且在多数岩性类别上表现出优异的识别能力,表明该方法能够有效处理复杂储层岩性识别任务.

表 6 HGF 数据集中不同模型的实验结果对比,最佳结果以粗体标出

Table 6 Comparison of experimental results using different models in the HGF dataset, the best results are shown in bold

岩性标签	kNN	SVM	RF	BP	XGBoost	LGBM	CatBoost	CNN	LogGBS
C1	92.86	78.57	92.86	**100.00**	**100.00**	**92.86**	**100.00**	92.86	**100.0**
C2	85.51	42.51	98.07	95.65	94.69	97.58	98.07	99.52	**100.0**
C3	82.19	33.56	95.89	94.52	93.84	95.21	95.89	97.95	**100.0**
C4	90.32	79.03	96.77	95.97	96.77	97.18	97.18	**99.58**	98.75
C5	82.52	56.80	96.60	95.63	97.57	98.06	96.60	98.54	**100.0**
C6	85.95	86.76	99.46	98.92	**99.73**	98.92	99.19	98.88	97.21
C7	85.11	53.19	97.87	**100.00**	**100.00**	97.87	97.87	97.87	97.87
OA(%)	85.78	65.19	97.66	96.77	97.25	97.66	97.74	98.84	**98.85**
AA(%)	86.35	61.49	96.79	97.24	97.51	96.81	97.83	97.89	**99.12**
Kappa(×100)	82.25	55.31	97.06	95.95	96.56	97.06	97.17	**98.57**	**98.57**
训练时长(s)	**0.001**	0.048	0.147	0.100	1.329	0.097	1.964	816.721	0.142

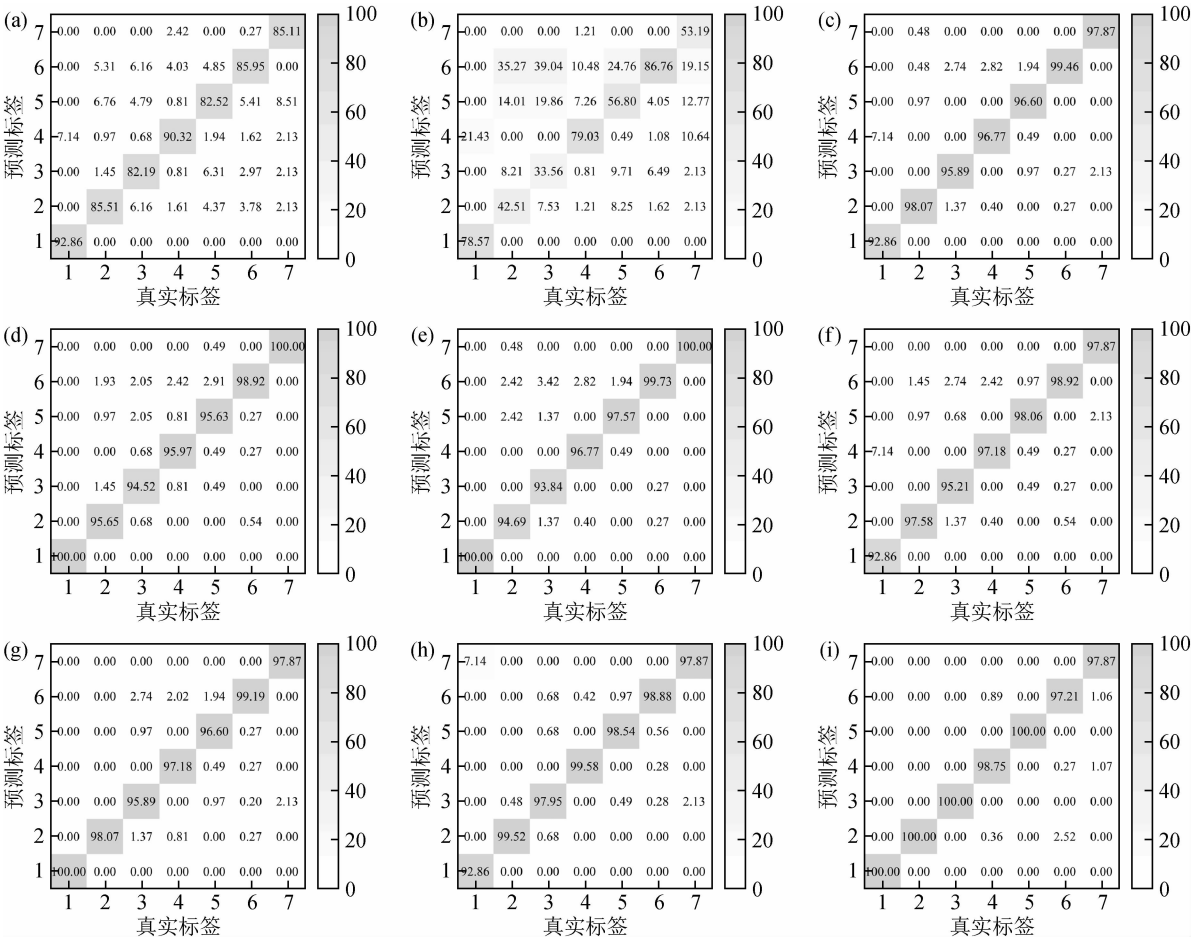


图 8　HGF 数据集中不同模型的混淆矩阵
(a) kNN; (b) SVM; (c) RF; (d) BP; (e) XGBoost; (f) LGBM; (g) CatBoost; (h) CNN; (i) LogGBS.
Fig. 8　Confusion matrices of different models in the HGF dataset

如图 9 和表 7 所示,展示了九种模型在 SGF 数据集上测试结果的混淆矩阵和实验结果对比表.整体而言,模型分类性能较 HGF 数据集有所提升.kNN 和 SVM 在白云质灰岩(C2)和泥质灰岩(C3)的识别中表现较差,RF、BP、XGBoost、LGBM、CatBoost 和 CNN 六个模型的分类性能显著优于前述方法,大多数岩性类别的识别准确率均超过 95%.然而,在面对白云质灰岩(C2)时,这些模型的判别能力仍存在一定局限,准确率均维持在 95% 左右.相比之下,本研究提出的 LogGBS 模型展现出全面而优越的岩性识别能力:不仅在石灰岩(C1)和白云岩(C4)等类别中表现突出,对白云质灰岩(C2)也达到了 97.52% 的识别准确率,显著优于所有对比模型.

表 7　SGF 数据集中不同模型的实验结果对比,最佳结果以粗体标出

Table 7　Comparison of experimental results using different models in the SGF dataset, the best results are shown in bold

岩性标签	kNN	SVM	RF	BP	XGBoost	LGBM	CatBoost	CNN	LogGBS
C1	95.24	97.48	99.44	99.16	99.16	99.44	**99.72**	**99.72**	**99.72**
C2	77.69	76.86	95.87	95.87	95.04	95.87	95.04	94.21	97.52
C3	85.97	91.17	97.92	98.18	98.44	98.18	**98.96**	97.66	98.44
C4	93.26	93.78	99.74	99.22	99.22	99.74	**100.00**	96.60	**100.00**
C5	92.69	94.15	99.51	99.51	99.63	**99.76**	**99.76**	98.54	**99.76**
C6	91.96	91.60	99.52	99.52	**99.88**	99.64	99.64	98.53	99.76
OA(%)	91.35	92.66	99.17	99.10	99.24	99.31	99.45	98.13	**99.52**
AA(%)	89.47	90.84	98.67	98.58	98.56	98.77	98.85	97.54	**99.20**
Kappa(×100)	88.99	90.67	98.95	98.86	99.03	99.12	99.30	97.61	**99.39**
训练时长(s)	**0.002**	0.100	0.245	0.009	2.589	0.139	0.125	795.265	2.065

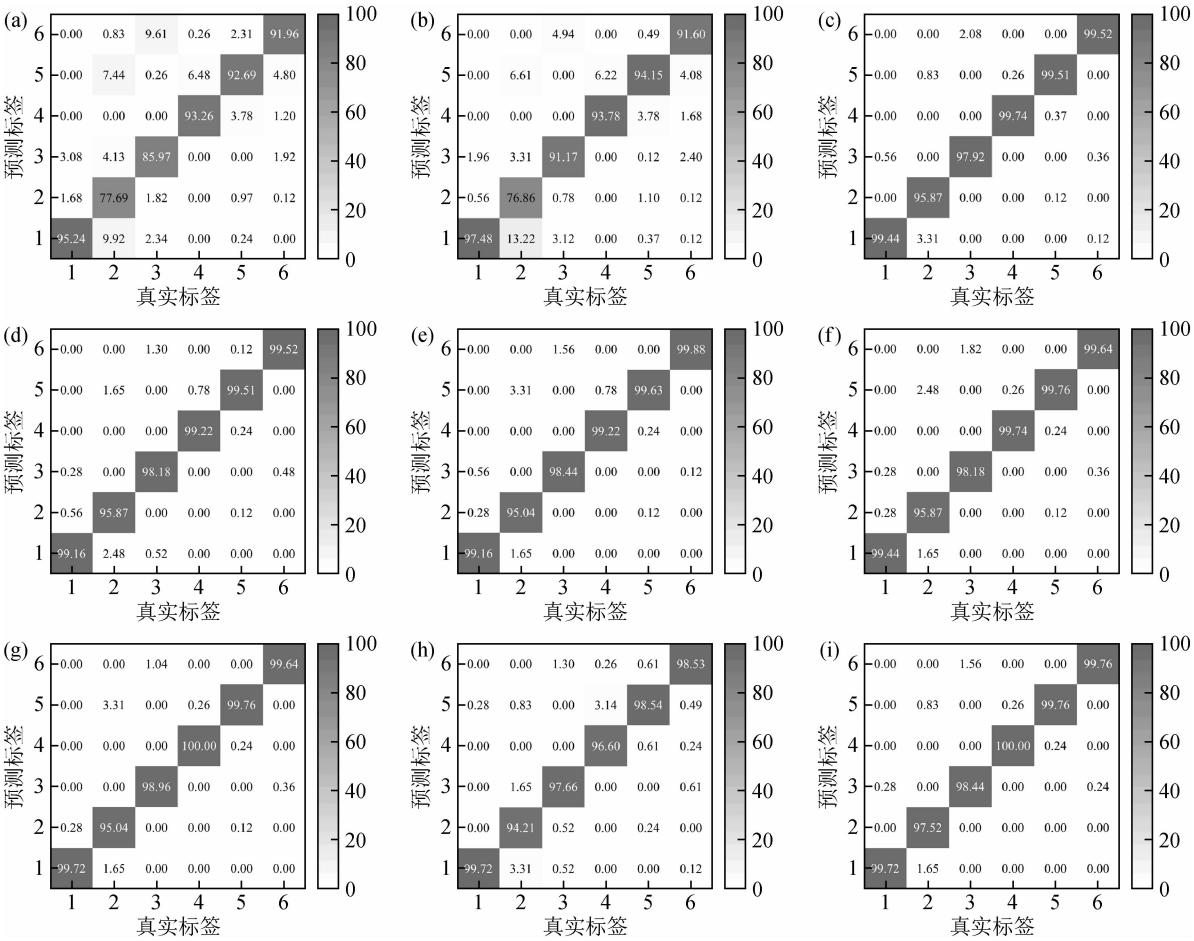


图 9 SGF 数据集中不同模型的混淆矩阵
(a)kNN;(b)SVM;(c)RF;(d)BP;(e)XGBoost;(f)LGBM;(g)CatBoost;(h)CNN;(i)LogGBS.
Fig. 9 Confusion matrices of different models in the SGF dataset

如图 10 和表 8 所示,展示了九种模型在 XGF 数据集上测试结果的混淆矩阵和实验结果对比表. 总体来看,SVM 在粉砂岩(C1)、砾岩(C2)和英安岩(C5)上表现出较好的识别能力,各类准确率均超过 90%. 然而,在流纹岩(C3)和玄武岩(C4)的分类中存在明显困难,表现出较大程度的误判. kNN、RF 和 BP 的整体分类性能显著提升,多数岩性类别准确率均达到 97% 至 99%. XGBoost、LGBM、CatBoost、CNN 及本研究提出的 LogGBS 模型表现整体更优异,整体识别准确率均超过 99%. 特别地,LogGBS 在流纹岩(C3)和英安岩(C5)的预测中实现了 100% 的准确率,展现出卓越的岩性识别能力.

表 8 XGF 数据集中不同模型的实验结果对比,最佳结果以粗体标出

Table 8 Comparison of experimental results using different models in the XGF dataset, the best results are shown in bold

岩性标签	kNN	SVM	RF	BP	XGBoost	LGBM	CatBoost	CNN	LogGBS
C1	98.40	97.06	99.38	99.07	99.64	99.64	99.74	98.96	**99.69**
C2	98.95	94.05	99.30	99.47	99.30	99.47	99.47	**99.65**	99.47
C3	97.18	79.76	98.24	98.00	**100.00**	**100.00**	99.76	99.29	**100.00**
C4	99.02	54.90	96.08	96.08	**99.02**	**99.02**	**99.02**	**99.02**	**99.02**
C5	97.94	97.94	100.00	98.97	98.97	**100.00**	**100.00**	97.94	**100.00**
OA(%)	98.20	91.26	99.02	98.79	99.63	99.69	99.69	99.12	**99.72**
AA(%)	98.30	84.74	98.60	98.32	99.39	99.63	99.60	98.97	**99.64**
Kappa(×100)	97.10	85.59	98.41	98.05	99.41	99.50	99.50	98.59	**99.55**
训练时长(s)	**0.013**	0.152	4.172	0.020	0.408	0.800	0.265	634.230	3.190

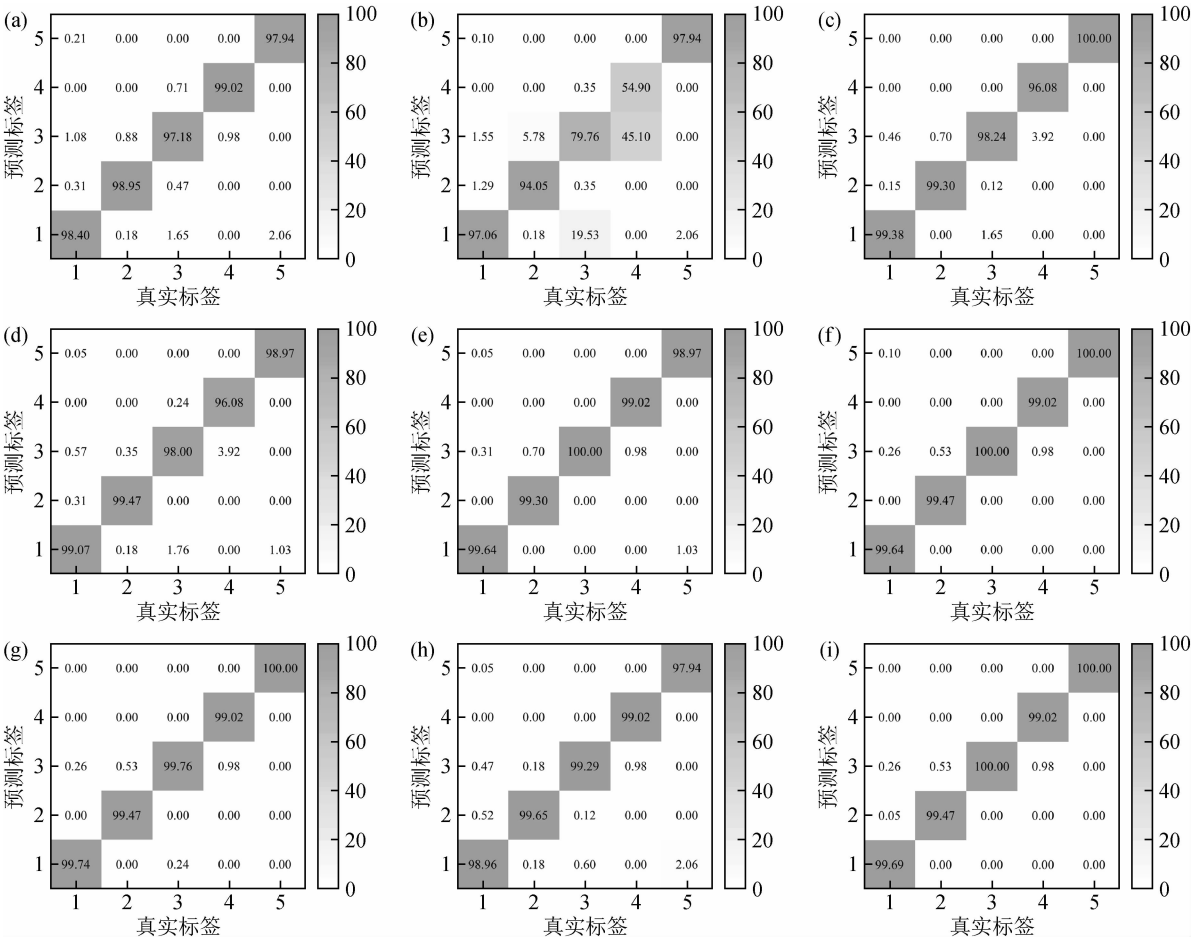


图 10　XGF 数据集中不同模型的混淆矩阵

(a) kNN; (b) SVM; (c) RF; (d) BP; (e) XGBoost; (f) LGBM; (g) CatBoost; (h) CNN; (i) LogGBS.

Fig. 10　Confusion matrices of different models in the XGF dataset

从 OA、AA 和 Kappa 系数三项评价指标综合来看,LogGBS 模型在所有三类储层数据集上均表现出卓越且稳定的性能,其指标数值均高于 97%,显示出该模型在面对不同岩性及储层类型时具有较强的泛化能力(表 6 ~ 8). 相比之下,SVM 模型的综合性能较差,而在碳酸盐岩和火山岩中,指标略有提升. kNN 模型的表现波动较大:在火山岩储层中指标可达 98.20%,但在另两类储层中的指标位于 82% 至 91% 左右. RF、BP、XGBoost、LGBM、CatBoost 和 CNN 等模型整体表现良好,三类储层中的评价指标普遍高于 95%,而 CNN 由于深度学习架构的复杂性导致训练时长明显落后于常规机器学习模型,同时深度学习训练需要大量的样本,这导致在测井岩性识别任务中性能受限. 而本文所提出的 LogGBS 在 OA、AA 和 Kappa 三项指标上总体优于其他对比模型,并且其训练时长也令人满意,在计算效率和分类精度中取得了最优平衡.

图 11 ~ 13 可视化了三类储层测井数据集中验证井的岩性识别结果图,综合对比可以看出,所有对比模型的岩性分类表现均逊于 LogGBS,主要受算法局限性和对地质复杂性适应能力不足的影响. SVM 模型虽具备核函数映射和非线性建模能力,但仍难以处理特征重叠严重的岩性,树模型(RF、XGBoost、LGBM、CatBoost)倾向于对总体特征进行分割,在厚层均质段易过拟合,可见图 13,在 3983.75 ~ 3996.5 m 处易将玄武岩误判为流纹岩. 而 CNN 依赖大量高质量标注数据,其训练成本高且在岩性过渡带处的判别效果劣于 LogGBS. 相比而言,LogGBS 模型通过全局优先级认知结构与局部球级结构建模,自适应地融合多尺度测井特征,有效缓解了均质厚层中的过拟合问题,同时增强了对薄层变化的敏感性,能够精准捕捉岩性与测井响应间的复杂关系,显著优于传统机器学习和深度学习方法.

LogGBS 的优势主要源于粒度球的空间聚合特

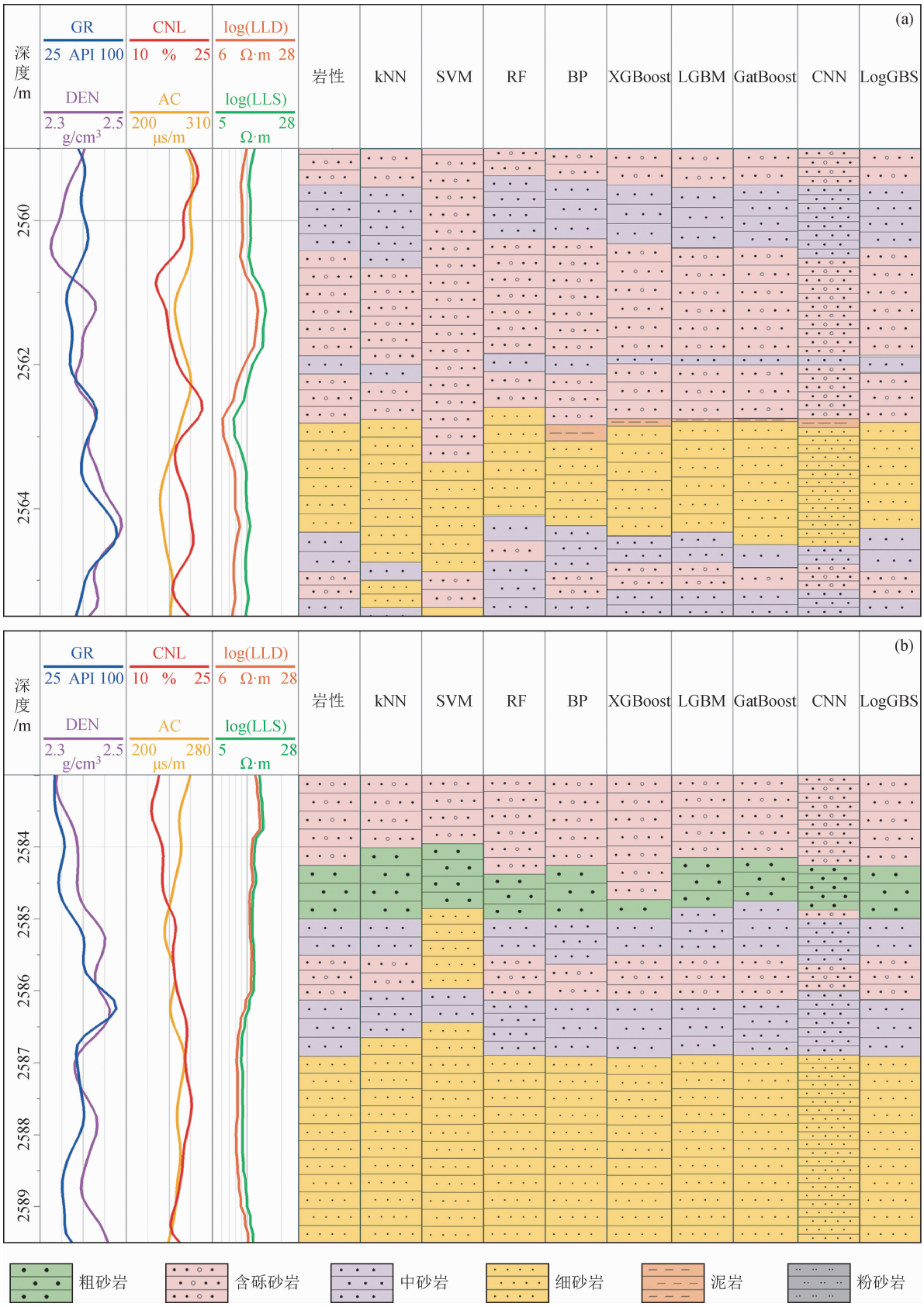


图 11 HGF 数据集中 J66 井的岩性识别结果

Fig. 11 Lithology identification results of Well J66 in the HGF dataset

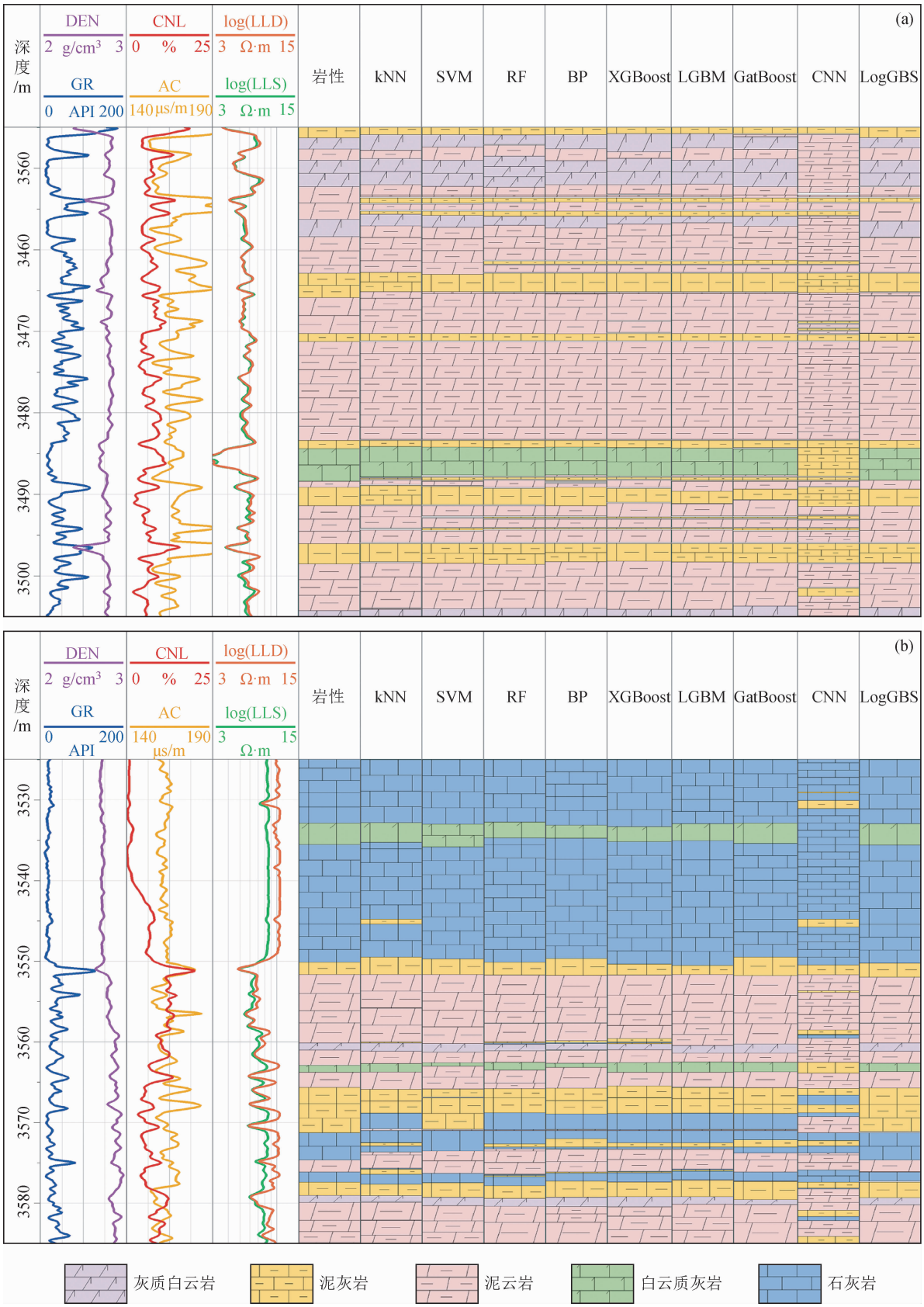


图 12　SGF 数据集中 J1 井的岩性识别结果

Fig. 12　Lithology identification results of Well J1 in the SGF dataset

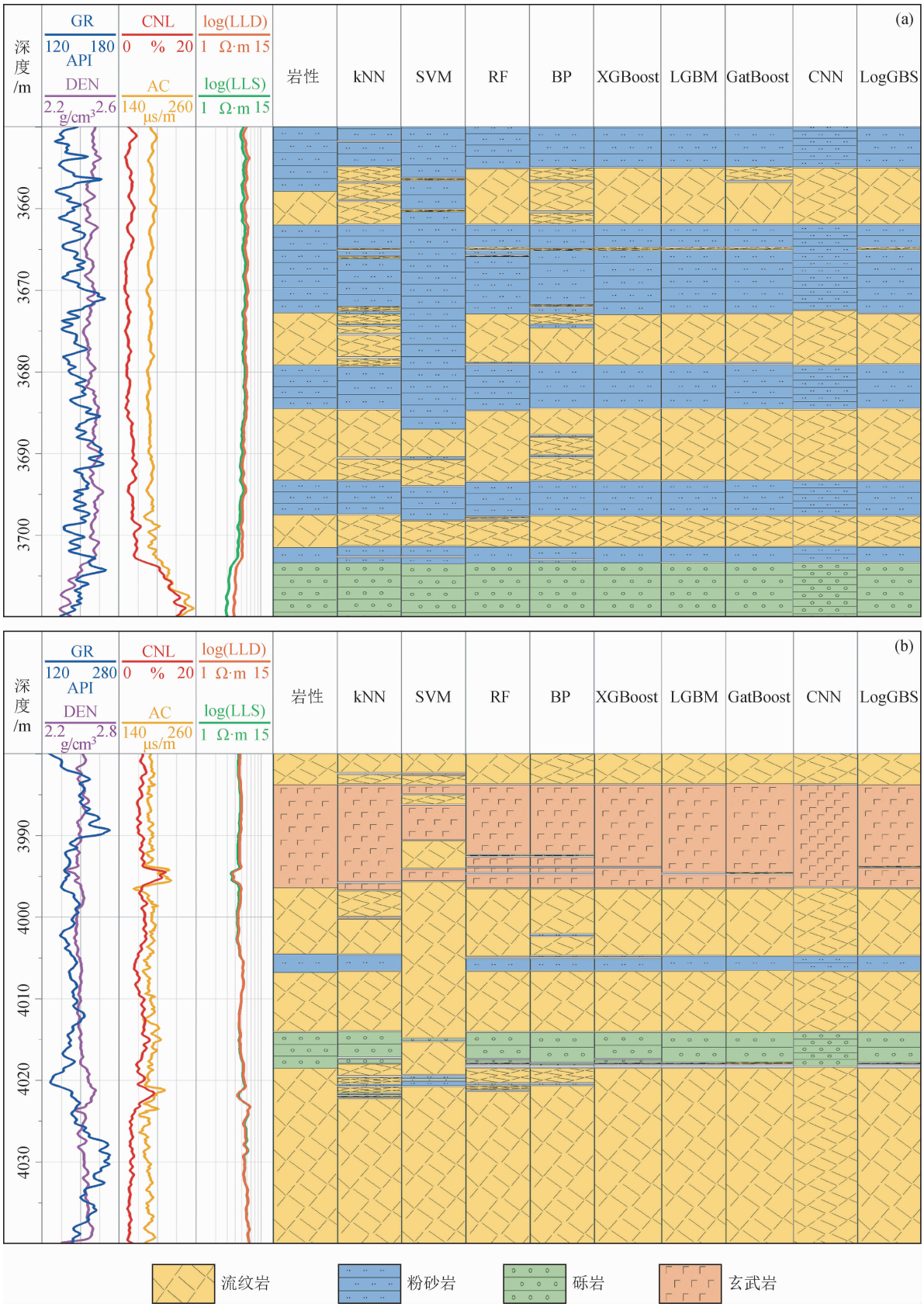


图 13 XGF 数据集中 X8 井的岩性识别结果

Fig. 13 Lithology identification results of Well X8 in the XGF dataset

性,粒度球以球心和半径表征测井数据的局部统计模式,通过空间聚合抑制噪声,基于纯度驱动分裂增强薄层岩性突变位置的识别,通过去重叠化优化构建清晰的岩性边界,从而在井剖面上呈现更为准确的岩性分布. 相比传统机器学习方法,LogGBS 能准确刻画岩性渐变带及快速变化带,实现对复杂地质环境的高精度岩性识别,表现出显著优越性.

综上所述,LogGBS 在多种储层的岩性分类任务中均表现出优越的精度和泛化性,实现了基于常规测井数据的高可靠性岩性识别.

5　结　论

本研究针对复杂地质环境下传统测井岩性识别方法精度不足、现有机器学习模型泛化能力弱等问题,提出基于粒度球空间的测井岩性识别方法(LogGBS),通过多区域、多储层类型的实验验证与对比分析,得出以下核心结论:

(1) LogGBS 方法构建了测井岩性分类的新范式

LogGBS 创新性融合自适应粒度球生成器(AGBG)与 GBkNN 分类器,以"粒度球纯度"为核心控制参数,实现了测井数据从"点状输入"到"粒度球聚合"的表征升级. 通过改进的距离度量公式实现查询点与粒度球的快速匹配,既保留测井数据的全局统计特征,又有效抑制局部噪声(如仪器误差、异常测点)干扰,解决了传统方法对薄层岩性误判、厚层过拟合等问题.

(2)粒度球纯度对分类精度具有显著调控作用

在碎屑岩(杭锦旗气田)、碳酸盐岩(苏里格气田)与火山岩(徐深气田)三类储层中,LogGBS 分类精度随粒度球纯度提升呈稳定上升趋势:当纯度≥0.9 时,三类储层的平均准确率均超过 90%,表明高纯度粒度球能更精准捕捉岩性类别的局部差异特征,尤其适用于岩性过渡复杂、非均质性强的储层场景.

(3)为复杂地质条件下油气勘探提供技术支撑

LogGBS 方法无需依赖大量高质量标注数据,且对测井数据中的噪声具有强鲁棒性,可适配碎屑岩、碳酸盐岩、火山岩等不同储层类型的地质复杂性,为非常规储层、薄互层储层等难识别区域的岩性判识提供了高效解决方案. 其建立的"粒度球纯度—分类精度"调控关系,可根据实际勘探需求灵活调整参数,在提升岩性识别效率的同时,为储层评价、资源潜力评估及开发方案优化提供更精准的基础数据,对推动油气勘探开发智能化具有重要实践意义.

致　谢　感谢审稿专家提出的修改意见和编辑部的大力支持!

References

Bai X L, Zhang S N, Huang Q Y, *et al*. 2016. Origin of dolomite in the Middle Ordovician peritidal platform carbonates in the northern Ordos Basin, western China. Petroleum Science, 13(3): 434-449, doi: 10.1007/s12182-016-0114-5.

Cai Z R, Huang Q T, Xia B, *et al*. 2012. Development features of volcanic rocks of the Yingcheng Formation and their relationship with fault structure in the Xujiaweizi Fault Depression, Songliao Basin, China. Petroleum Science, 9(4): 436-443, doi: 10.1007/s12182-012-0227-4.

Chen L. 1982. Topological structure in visual perception. Science, 218(4573): 699-700, doi: 10.1126/science.7134969.

Chen Y L, Li G L, Yang Z X, *et al*. 2020. Identification of lithology and lithofacies of Chang7 reservoir in Heshui area by KNN algorithm. Well Logging Technology (in Chinese), 44(2): 182-185.

Feng H L, Zhang Z T, Zhang C L, *et al*. 2025. Sequence-variable attention temporal convolutional network for volcanic lithology identification based on well logs. Journal of Petroleum Exploration and Production Technology, 15(1): 6, doi: 10.1007/s13202-024-01887-4.

Han Q D, Zhang X T, Shen W. 2019. Application of support vector machine based on decision tree feature extraction in lithology classification. Journal of Jilin University (Earth Science Edition) (in Chinese), 49(2): 611-620.

Ji W M, Li W L, Liu Z, *et al*. 2013. Research on the Upper Paleozoic gas source of the Hangjinqi block in the northern Ordos Basin. Natural Gas Geoscience (in Chinese), 24(5): 905-914.

Liu G P, Zeng L B, Lei M S, *et al*. 2016. Fracture development characteristics and main controlling factors of the volcanic reservoir in Xujiaweizi fault depression. Geology in China (in Chinese), 43(1): 329-337.

Liu G Q, Gong R B, Shi Y J, *et al*. 2022. Construction of well logging knowledge graph and intelligent identification method of hydrocarbon-bearing formation. Petroleum Exploration and Development (in Chinese), 49(3): 502-512.

Mou D, Wang Z W, Huang Y L, *et al*. 2015. Lithological identification of volcanic rocks from SVM well logging data: case study in the eastern depression of Liaohe Basin. Chinese Journal of Geophysics (in Chinese), 58(5): 1785-1793, doi: 10.6038/cjg20150528.

Ni C H, Liu G X, Zhu J H, *et al*. 2018. Origin and source of natural gas in the Upper Paleozoic in Hangjinqi area, Ordos Basin. Petroleum Geology and Experiment (in Chinese), 40(2): 193-199.

Pedrycz W, Succi G. 2005. Genetic granular classifiers in modeling software quality. Journal of Systems and Software, 76(3): 277-285, doi: 10.1016/j.jss.2004.06.018.

Qin Y, Liu C Y, Zhao J F, *et al*. 2025. Unravelling the origin of gas in tight sandstones of the Hangjinqi gas field, Ordos Basin, China: New insights from natural gas geochemistry data. Organic Geochemistry, 206: 105012, doi: 10.1016/j.orggeochem.2025.105012.

Salehi S, Selamat A, Mashinchi M R, *et al*. 2015. The synergistic combination of particle swarm optimization and fuzzy sets to design granular classifier. Knowledge-Based Systems, 76: 200-218, doi:

10.1016/j.knosys.2014.12.017.

Shen A J, Zhao W Z, Hu A P, *et al*. 2015. Major factors controlling the development of marine carbonate reservoirs. Petroleum Exploration and Development (in Chinese), 42(5): 545-554.

Song Z H, Gong H Y, Ran A H, *et al*. 2024. Lithology logging identification of volcanic rock based on ADASYN-GS-XGBOOST hybrid model. Marine Origin Petroleum Geology (in Chinese), 29(2): 188-196.

Su F, Ma L, Luo R Z, *et al*. 2020. Research and application of logging lithology identification based on improve multi-class twin support vector machine. Progress in Geophysics (in Chinese), 35(1): 174-180, doi: 10.6038/pg2020DD0185.

Sun K. 2022. Logging identification method of complex lithology in buried hill based on the improved KNN algorithm. Special Oil & Gas Reservoirs (in Chinese), 29(3): 18-27.

Wang G Y, Song J G, Xu F, *et al*. 2021. Random forests lithology prediction method for imbalanced data sets. Oil Geophysical Prospecting (in Chinese), 56(4): 679-687.

Wang H, Jiang Y N, Zhang X, *et al*. 2021. Lithology identification method based on gradient boosting algorithm. Journal of Jilin University (Earth Science Edition) (in Chinese), 51(3): 940-950.

Wang L Y, Liu H Q, Cheng C, *et al*. 2025. Research progress on intelligent identification methods for well logging lithofacies based on deep learning. Progress in Geophysics (in Chinese), 40(4): 1773-1787, doi: 10.6038/pg2025II0394.

Xia S Y, Liu Y S, Ding X, *et al*. 2019. Granular ball computing classifiers for efficient, scalable and robust learning. Information Sciences, 483: 136-152, doi: 10.1016/j.ins.2019.01.010.

Xia S Y, Zhang H, Li W H, *et al*. 2022a. GBNRS: a novel rough s*et al*gorithm for fast adaptive attribute reduction in classification. IEEE Transactions on Knowledge and Data Engineering, 34(3): 1231-1242, doi: 10.1109/TKDE.2020.2997039.

Xia S Y, Peng D W, Meng D Y, *et al*. 2022b. Ball k-means: fast adaptive clustering with no bounds. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1): 87-99, doi: 10.1109/TPAMI.2020.3008694.

Xia S Y, Wang G Y, Gao X B, *et al*. 2026. Granular-ball computing: an efficient, robust, and interpretable adaptive multi-granularity representation and computation method. arXiv: 2304.11171, doi: 10.48550/arXiv.2304.11171.

Xie J, Hao X F, Zhang Y P, *et al*. 2023. Fracture identification and characteristics of carbonate underground gas storage: an example from the eastern area of Sulige gas field, Ordos Basin, China. Scientific Reports, 13(1): 22446, doi: 10.1038/s41598-023-50072-w.

Xie Y X, Zhu C Y, Zhou W, *et al*. 2018. Evaluation of machine learning methods for formation lithology identification: a comparison of tuning processes and model performances. Journal of Petroleum Science and Engineering, 160: 182-193, doi: 10.1016/j.petrol.2017.10.028.

Xie Y X, Zhu C Y, Hu R S, *et al*. 2021. A coarse-to-fine approach for intelligent logging lithology identification with extremely randomized trees. Mathematical Geosciences, 53(5): 859-876, doi: 10.1007/s11004-020-09885-y.

Zadeh L A. 1996. Fuzzy sets and information granularity. //Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh. River Edge, NJ: World Scientific Publishing Co., Inc., 433-448.

Zhang J H, Fang W, Li J K, *et al*. 2009. Deep gases and their genetic types of the Xujiaweizi fault depression zone, Songliao Basin and their contribution. Acta Geologica Sinica (in Chinese), 83(4): 579-589.

Zhang Z T, Zhang X, Zhang C L, *et al*. 2025. Category-wise sequence correlation filter for lithology identification with well logs. IEEE Transactions on Geoscience and Remote Sensing, 63: 4503814, doi: 10.1109/TGRS.2025.3547940.

Zhao F D, Han Z M, Fu X F, *et al*. 2025. LogDiffusion: a method of lithology identification based on diffusion probability model. Progress in Geophysics (in Chinese), 40(1): 106-120, doi: 10.6038/pg2025II0075.

Zhao J L, Hu L, Huang L S, *et al*. 2023. MSRA-G: combination of multi-scale residual attention network and generative adversarial networks for hyperspectral image classification. Engineering Applications of Artificial Intelligence, 121: 106017, doi: 10.1016/j.engappai.2023.106017.

Zhao R L, Yang L S, Xu X, *et al*. 2025. Lithology identification method and research of volcanic rock based on XGBoost algorithm. Progress in Geophysics (in Chinese), 40(2): 646-657, doi: 10.6038/pg2025II0019.

Zhao T T, Zhang C L, Zhang C Y, *et al*. 2018. Application of KNN classification model based on fuzzy entropy in lithology recognition. Computer Engineering and Applications (in Chinese), 54(24): 260-265.

Zhou J G, Zhang F, Guo Q X, *et al*. 2011. Barrier-lagoon sedimentary model and reservoir distribution regularity of Lower-Ordovician Majiagou Formation in Ordos Basin. Acta Sedimentologica Sinica (in Chinese), 29(1): 64-71, doi: 10.14027/j.cnki.cjxb.2011.01.013.

附中文参考文献

陈玉林, 李戈理, 杨智新, 等. 2020. 基于 KNN 算法识别合水地区长 7 储层岩性岩相. 测井技术, 44(2): 182-185.

韩启迪, 张小桐, 申维. 2019. 基于决策树特征提取的支持向量机在岩性分类中的应用. 吉林大学学报(地球科学版), 49(2): 611-620.

纪文明, 李潍莲, 刘震, 等. 2013. 鄂尔多斯盆地北部杭锦旗地区上古生界气源岩分析. 天然气地球科学, 24(5): 905-914.

刘国平, 曾联波, 雷茂盛, 等. 2016. 徐家围子断陷火山岩储层裂缝发育特征及主控因素. 中国地质, 43(1): 329-337.

刘国强, 龚仁彬, 石玉江, 等. 2022. 油气层测井知识图谱构建及其智能识别方法. 石油勘探与开发, 49(3): 502-512.

牟丹, 王祝文, 黄玉龙, 等. 2015. 基于 SVM 测井数据的火山岩岩性识别——以辽河盆地东部坳陷为例. 地球物理学报, 58(5): 1785-1793, doi: 10.6038/cjg20150528.

倪春华, 刘光祥, 朱建辉, 等. 2018. 鄂尔多斯盆地杭锦旗地区上古生界天然气成因及来源. 石油实验地质, 40(2): 193-199.

沈安江, 赵文智, 胡安平, 等. 2015. 海相碳酸盐岩储集层发育主控因素. 石油勘探与开发, 42(5): 545-554.

宋梓豪, 巩红雨, 冉爱华, 等. 2024. 基于 ADASYN-GS-XGBOOST 混合模型的火山岩测井岩性识别. 海相油气地质, 29(2): 188-196.

苏斌, 马磊, 罗仁泽, 等. 2020. 基于改进多分类孪生支持向量机的测井岩性识别方法研究与应用. 地球物理学进展, 35(1): 174-180, doi: 10.6038/pg2020DD0185.

孙岿. 2022. 基于改进 KNN 算法的潜山复杂岩性测井识别方法. 特种油气藏, 29(3): 18-27.

王光宇, 宋建国, 徐飞, 等. 2021. 不平衡样本集随机森林岩性预测方法. 石油地球物理勘探, 56(4): 679-687.

王恒, 姜亚楠, 张欣, 等. 2021. 基于梯度提升算法的岩性识别方法. 吉林大学学报(地球科学版), 51(3): 940-950.

王丽媛, 刘红岐, 程超, 等. 2025. 基于深度学习的测井岩相智能识别方法研究进展. 地球物理学进展, 40(4): 1773-1787, doi: 10.6038/pg2025II0394.

张居和, 方伟, 李景坤, 等. 2009. 松辽盆地徐家围子断陷深层天然气成因类型及各种成因气贡献. 地质学报, 83(4): 579-589.

赵逢达, 韩滋民, 付晓飞, 等. 2025. LogDiffusion: 一种基于扩散概率模型的岩性识别方法. 地球物理学进展, 40(1): 106-120, doi: 10.6038/pg2025II0075.

赵然磊, 杨留栓, 徐晓, 等. 2025. 基于 XGBoost 算法的火山岩岩性识别方法与研究. 地球物理学进展, 40(2): 646-657, doi: 10.6038/pg2025II0019.

赵彤彤, 张春雷, 张春雨, 等. 2018. 基于模糊熵的 KNN 分类模型在岩性识别中的应用. 计算机工程与应用, 54(24): 260-265.

周进高, 张帆, 郭庆新, 等. 2011. 鄂尔多斯盆地下奥陶统马家沟组障壁潟湖沉积相模式及有利储层分布规律. 沉积学报, 29(1): 64-71, doi: 10.14027/j.cnki.cjxb.2011.01.013.