

文章编号: 1671-1505(2025)04-0937-13 DOI: 10.7605/gdxb.2025.083

# 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\*

唐佰强<sup>1,2</sup> 孟庆涛<sup>1,2</sup> 杨 亮<sup>3</sup> 胡 菲<sup>1,2</sup> 谭 悦<sup>4</sup> 邢济麟<sup>3</sup>  
刘招君<sup>1,2</sup> 张恩威<sup>1,2</sup> 董秦玮<sup>1,2</sup>

1 吉林大学地球科学学院, 吉林长春 130061

2 吉林省页岩与共生能源矿产重点实验室, 吉林长春 130061

3 中国石油吉林油田公司勘探开发研究院, 吉林松原 138000

4 大庆钻探工程有限公司地质录井公司, 黑龙江大庆 163000

**摘 要** 细粒沉积岩岩相的空间展布特征是页岩油勘探的关键研究内容。由于高成本且稀缺的取心井限制了岩相的分析, 测井预测岩相的工作变得尤为重要。以松辽盆地青山口组一段(青一段)为例, 建立了“岩性+矿物+TOC+沉积构造”的岩相划分方案, 确定了 7 类岩相。结合 6 条常规测井曲线形成了 X8 井的岩相—测井数据库。综合使用机器学习中的随机森林(RF)、XGBoost 和支持向量机(SVM) 3 种模型评价岩相的预测效果, 并确定 SVM 是最优分类模型, 使用 BSMOTE 处理岩相样本的分类不平衡问题并将处理后的数据输入到 SVM 模型, 建立了 BSMOTE-SVM 的岩相预测的组合模型。BSMOTE-SVM 的预测效果最佳, 准确率(A)、精确率(P)、召回率(R)和 F1 依次分别为 86.49%、86.60%、86.49% 和 86.31%。该组合模型可快速且精准的预测多井的岩相, 并确定了松辽盆地长岭凹陷青一段的岩相分布, 为下一步页岩油有利富集区的优选提供了一定指导依据。

**关键词** 细粒沉积岩 岩相 测井预测 BSMOTE-SVM 青山口组 松辽盆地

**第一作者简介** 唐佰强, 男, 1997 年生, 博士研究生, 主要从事测井地质学、机器学习与沉积学等研究。  
E-mail: 1587185918@qq.com。

**通讯作者简介** 孟庆涛, 女, 1984 年生, 博士, 教授, 博士生导师, 主要从事沉积学、石油地质学、非常规油气勘探与开发方面的研究。E-mail: mengqt@jlu.edu.cn。

中图分类号: P618.130 文献标志码: A

## Intelligent prediction of fine-grained sedimentary lithofacies based on BSMOTE-SVM: a case study of the Member 1 of Qingshankou Formation in Songliao Basin

TANG Baiqiang<sup>1,2</sup> MENG Qingtao<sup>1,2</sup> YANG Liang<sup>3</sup> HU Fei<sup>1,2</sup> TAN Yue<sup>4</sup> XING Jilin<sup>3</sup>  
LIU Zhaojun<sup>1,2</sup> ZHANG Enwei<sup>1,2</sup> DONG Qinwei<sup>1,2</sup>

\* 吉林省自然科学基金项目(编号: 20230101081JC)和中国石油吉林油田分公司项目(编号: JS2022-W-13-JZ-78-92)联合资助。  
[Co-funded by the Project of Natural Science Foundation Jilin Province (No. 20230101081JC) and the Project of CNPC Jilin Oilfield Branch Company (No. JS2022-W-13-JZ-78-92)]

收稿日期: 2025-04-08 改回日期: 2025-05-12

1 College of Earth Sciences, Jilin University, Changchun 130061, China

2 Key-Lab for Oil Shale and Paragenetic Energy Minerals, Changchun 130061, China

3 Exploration and Development Research Institute of Jilin Oilfield Company, PetroChina, Jinlin Songyuan, 138000, China

4 Geological Logging Company of Daqing Drilling Engineering Company, Heilongjiang Daqing, 163000, China

**Abstract** The spatial distribution of lithofacies of fine-grained sedimentary rocks is a critical research focus in shale oil exploration. Due to the high cost and scarcity of core wells, which constrain direct lithofacies analysis, logging-based prediction has become increasingly essential. Taking the First Member of the Qingshankou Formation in the Songliao Basin as a case study, this research establishes a lithofacies classification scheme integrating lithology, mineral composition, total organic carbon (TOC), and sedimentary structures, resulting in the identification of seven distinct lithofacies types. A lithofacies-well log dataset was constructed for Well X8 using six conventional logging curves. Three machine learning algorithms—Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM)—were employed to evaluate classification performance, with SVM identified as the optimal model. To address class imbalance in the training data, the BSMOTE (Borderline Synthetic Minority Over-sampling Technique) algorithm was applied. The balanced dataset was then used to develop a hybrid lithofacies prediction model: BSMOTE-SVM. The BSMOTE-SVM model demonstrated the best predictive performance, achieving an accuracy of 86.49%, precision of 86.60%, recall of 86.49%, and  $F1$ -score of 86.31%. This integrated model enables rapid and accurate lithofacies prediction across multiple wells and delineates the lithofacies distribution in Member 1 of the Qingshankou Formation in the Changling sag, offering a robust foundation for selecting favorable shale oil enrichment zones in future exploration.

**Key words** fine-grained sedimentary rocks, lithofacies, logging prediction, BSMOTE-SVM, Qingshankou Formation, Songliao Basin

**About the first author** TANG Baiqiang, born in 1997, is a Ph.D. candidate. He is mainly engaged in logging geology, machine learning and sedimentology. E-mail: 1587185918@qq.com.

**About the corresponding author** MENG Qingtao, born in 1984, is a Ph.D. and professor. She is mainly engaged in sedimentology, petroleum geology and exploration and development of unconventional oil and gas. E-mail: mengqt@jlu.edu.cn.

细粒沉积岩是指粒度小于  $62.5\ \mu\text{m}$  的颗粒含量大于 50% 的碎屑沉积岩, 主要由黏土和粉砂等细粒物质组成, 并且约占全球沉积岩的 70% (姜在兴等, 2013; 朱如凯等, 2022; 林兴悦等, 2025)。随着全球能源理论和技术的不断创新, 细粒沉积岩层系中的页岩油气资源逐渐占有重要地位 (邹才能等, 2020)。全球页岩油气资源丰富, 美国在分布广泛且连续的海相细粒沉积岩地层中勘探出  $1.53 \times 10^8\ \text{t}$  的页岩油, 约占全球的五分之一 (贾承造等, 2024)。中国的多个盆地含有丰富的湖相页岩油资源, 湖相盆地对沉积环境和气候变化较为敏感, 导致了岩性接触多变、矿物类型多样和沉积结构复杂, 进而造成了细粒沉积岩的强非均质性 (朱如凯等, 2022), 这增大了页岩油勘探的难度。

因此, 以岩相为单元研究页岩油甜点分布成为 1 种重要手段, 许多学者在不同盆地或研究区提出了多种岩相划分方案。虽然岩相划分方法主要依据岩性、总有机碳 (TOC)、矿物组分和沉积构造等参数 (段宏亮和杨保良, 2024), 但是不同盆地细粒沉积岩地层的地球化学特征、储集层特征和构造背景有所差异, 以及不同学者的研究重点的不同, 共同导致这些参数有着详细的分类。如岩性按 50% 含量为界限可划分为泥岩、粉砂岩和碳酸盐岩 (孟庆涛等, 2024a), 也有学者在这 3 个岩性的基础上识别出了混合岩 (吴靖等, 2017)。在 TOC 的分类上, 李士超等 (2021) 将松辽盆地青山口组 TOC 分类为高 ( $>2\%$ )、中 ( $1\% \sim 2\%$ ) 和低 ( $<1\%$ ) 3 类, 吴靖等 (2017) 将东营凹陷的沙河

街组岩相的高、中和低的 TOC 分别定义为 >4%、2%~4% 和 <2%。在沉积构造的分类上，姜在兴等 (2023) 建议陆相页岩的构造可划分为纹层状 (<1 cm)、薄层状 (1~5 cm) 和块状 (>5 cm)。武瑾等 (2025) 将页岩构造划分为极薄纹层 (<0.1 mm)、薄纹层 (0.1~0.5 mm)、纹层 (>0.5 mm) 和块状 4 类。矿物类型的确定也有很多方法 (段宏亮和杨保良, 2024)。蕙克来等 (2015) 以陆源碎屑组分、碳酸盐组分及火山碎屑组分作为三端元进行划分。Liu 等 (2019) 以长英质矿物、碳酸盐矿物、黏土矿物相对百分含量的 1/3 和 2/3 为边界进行岩相类型划分。因此，由于不同地区的地质条件的差异，导致无法建立统一的岩相划分方案。

岩相分布特征最直接的确定方法是通过取心资料来获得的，但页岩油甜点的勘探需要确定岩相在二维甚至三维空间的分布，仅通过取心井确定岩相无法满足精细的勘探要求。测井因具有较高的纵向分辨率和更低的工程成本可被用来定量识别岩相的空间分布。由于细粒沉积岩的强非均质性导致岩相和测井之间具有非线性关系，常规的方法无法精确的表达岩相与测井的关系，机器学习是处理非线性问题的强大方法。李红斌等 (2022) 使用 Kohonen 神经网络建立了吉木萨尔坳陷二叠系的岩相测井模型，准确率达到了 87.9%。湛丽等 (2023) 基于 Bagging 预测了鄂尔多斯盆地延长组 6 类岩相。闫佳飞等 (2025) 使用了集成学习中的 XGBoost 预测了松辽盆地青山口组的 5 类岩相，准确率高达 90.03%。汪益宁等 (2016) 基于支持向量机 (SVM) 进行了致密储集层的岩相识别，并取得较好的预测效果。王民等 (2023) 基于随机森林 (RF) 驱动测井预测了泥页岩的岩相。但目前缺少综合评价常用的几种机器学习算法在岩相分类预测的对比分析。此外任何盆地的岩相类别的占比无法保证完全相同，存在着同时发育主要岩相和次要岩相的地质情况。不平衡分类的样本会导致机器学习模型对少数类岩相训练的不足 (鲁玲等, 2024)，急需 1 种方法解决岩相分类不平衡的问题。

以松辽盆地青山口组一段为研究对象，综合岩性、矿物、TOC 和沉积构造确定岩相类型。结合伽马测井 (GR)、井径测井 (CAL)、电阻率测井 (RT)、声波时差测井 (AC)、密度测井 (DEN)

和中子测井 (CNL)，建立 X8 井的岩相—测井数据库。综合评价随机森林 (RF)、XGBoost 和支持向量机 (SVM) 在研究区的预测精度，并优选出最优模型。将 BSMOTE 处理不平衡的岩相样本作为输入数据与最优模型结合建立组合模型。基于机器学习的组合模型驱动测井数据预测岩相的分布，进一步为页岩油的勘探提供一定的技术支撑。

## 1 地质背景

松辽盆地是中国东北部的超级盆地，富集着巨大的油气资源 (Liu *et al.*, 2023; 孙龙德等, 2024)。松辽盆地可划分为 6 个构造单元，其中的中央凹陷区是油气勘探的重点区域，尤其是近年来的古龙凹陷和长岭凹陷在页岩油的勘探突破进一步推进了中央凹陷区的研究 (沈华等, 2023; 朱国文等, 2023; 孟庆涛等, 2024b)。松辽盆地主要发育着上侏罗统、白垩系和古近系，沉积的最大厚度可超过 10 000 m (Wang *et al.*, 2022)。上白垩统青山口组由于沉积着厚度大且广泛分布的细粒沉积岩而成为松辽盆地的主要烃源岩层系之一 (孙龙德等, 2021)。

青山口组按岩性组合特征可分为一段、二段和三段。青山口组一段 (青一段) 时期温暖湿润的气候，大量发育的藻类及厌氧至缺氧的水体环境，共同促进了黑色富有机质页岩的大量沉积，是页岩油气勘探的重点层位之一 (姚玉来等, 2025) (图 1)。

## 2 方法原理

### 2.1 测井数据标准化

不同类型的测井数据在量纲上具有显著差异，直接输入到模型中进行训练会影响模型的预测精度。因此，对测井数据进行标准化处理后再进行模型训练是一个有效的方法。具体的方法如公式 (1) 所示：

$$y_{\log}^* = \frac{y_{\log} - \mu}{\sigma} \quad (1)$$

$y_{\log}^*$  是测井数据标准化后的值， $y_{\log}$  是原始测井值， $\mu$  和  $\sigma$  分别是测井数据的平均值和标准差。

### 2.2 BSMOTE 方法

BSMOTE 是在 SMOTE 基础上通过引入边界原

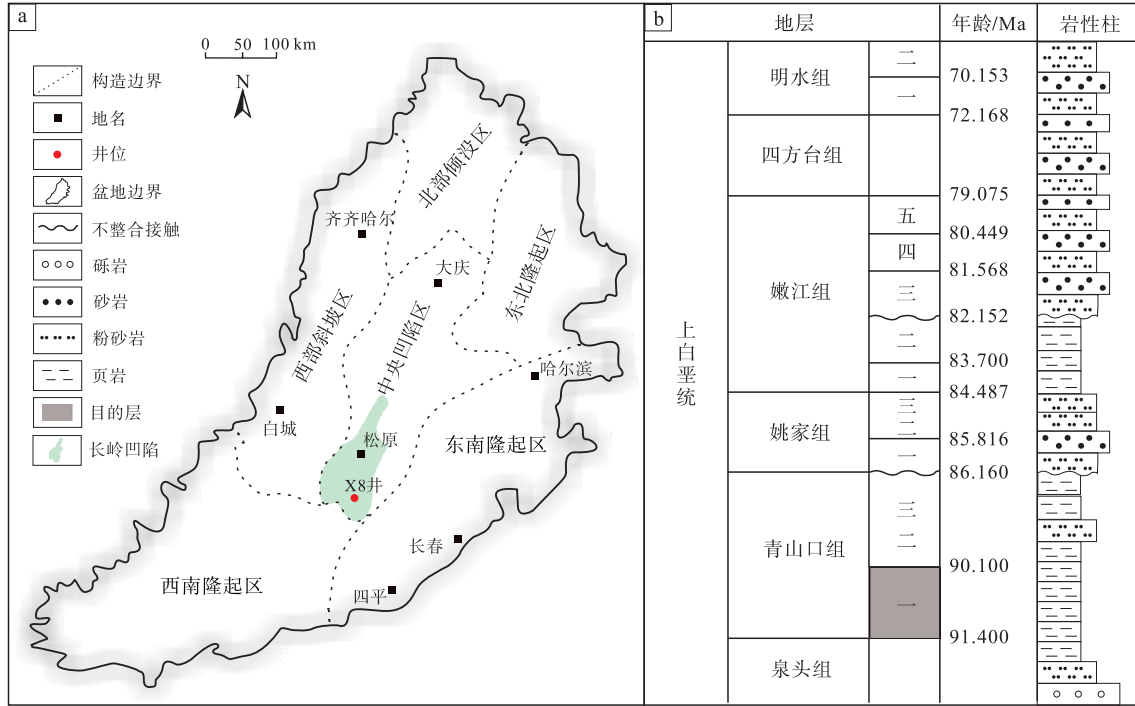


图 1 松辽盆地构造单元分布 (a) 与地层柱状图 (b) (据 Zhu et al., 2023; 有修改)

Fig. 1 Tectonic unit distribution (a) and stratigraphic column (b) in Songliao Basin (modified from Zhu et al., 2023)

则来改进的 1 种处理数据不平衡的方法 (鲁玲等, 2024)。具体步骤如下:

步骤 1: 假设多数类样本的集合为  $P$ , 少数类样本的集合为  $H$ , 通过欧氏距离来计算  $H$  的  $k$  个近邻  $S = (S_1, S_2 \dots S_n)$ 。

步骤 2: 根据  $S_i$  的近邻样本中的  $H$  类样本占  $P$  类样本的比例对每 1 个少数类样本进行分类,  $H/P = 0$  则归类为噪声样本,  $H/P = 1$  则归类为安全样本,  $0 < H/P < 1$  则归类为边界样本。

步骤 3: 设定边界样本的集合为  $A$ , 通过公式 (2) 计算新的样本集合  $A_{new}$ ,

$$A_{new} = A_j + m \times (A_k - A_j) \quad (2)$$

式中的  $A_k$  为  $A_j$  的近邻样本,  $m$  为分布在 0 和 1 之间的随机数。

步骤 4: 重复步骤 2 到步骤 3, 直至新生成的样本数量达到要求。

### 2.3 SVM 方法

SVM 是处理分类和回归问题的 1 种非常有效的监督学习方法。SVM 的数学思想是寻找 1 个超平面可以将所有数据最优的分隔开。SVM 的函数可表达为:

$$f(x) = \omega \times \varphi(x) + b \quad (3)$$

式中的  $x$  为输入的训练数据;  $f(x)$  和  $\varphi(x)$  分别是输出函数和映射函数;  $\omega$  是权重系数;  $b$  是常数。

通过引入拉格朗日函数和可将数据映射到高维数据的核函数, 可将公式表达为:

$$f(x) = \text{sgn} \left[ \sum_{i=1}^l \alpha_i y_i K(x_i, x_j) + b \right] \quad (4)$$

$\alpha_i$  为拉格朗日算子;  $K(x_i, x_j)$  为核函数,  $l$  为样本函数。

核函数的不同会影响 SVM 的准确性, 高斯核函数 (RBF) 是比较有效的 1 种 (汪益宁等, 2016)。具体的公式为:

$$K(x_i, x_j) = \exp \left( \frac{-\|x_i - x_j\|^2}{2\sigma^2} \right) \quad (5)$$

式中的  $\sigma$  是核函数。

### 2.4 模型评价参数

为了综合评价模型的分类预测效果, 选取了准确率 ( $A$ )、精确率 ( $P$ ), 召回率 ( $R$ ) 和  $F1$  共 4 个指标进行评价和对比分析。

$A$  是正例和负例中所有预测正确的样本占总样本数量的比例, 可表达为:

$$A = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

$P$  是根据预测为正例的样本中实际也为正例的比例计算的，可表达为：

$$P = \frac{TP}{TP + FP} \quad (7)$$

$R$  是根据实际为正例的样本中被预测正确的占比计算的，可表达为：

$$R = \frac{TP}{TP + FN} \quad (8)$$

$F1$  是同时考虑了  $P$  和  $R$  的一种评价指标，具体为：

$$F1 = \frac{2 \times P \times R}{P + R} \quad (9)$$

式中的  $TP$  可指示为预测为正例且实际上也是正例的样本， $TN$  可指示为预测为负例且实际上为负例的样本， $FP$  可指示为预测为正例但实际上为负例的样本， $FN$  可指示为预测为负例但实际上是正例的样本。

### 3 细粒沉积岩岩相类型及特征

#### 3.1 岩相类型

岩相是岩性、有机质、矿物组分和沉积结构等因素的综合表征，岩相的分布可显著控制着页岩油气的富集情况，是油气勘探的重要研究目标。岩性和矿物组分的不同会导致储集层品质（孔隙度、裂缝）和工程品质（脆性指数）的差异，进而控制页岩油的甜点分布（王伟等，2023）。TOC 是页

岩油气潜力评价的关键参数。沉积构造可代表细粒沉积岩岩相沉积时期的水体环境（李国萃等，2023）。因此，研究选用岩性、矿物、TOC 和沉积构造综合划分松辽盆地青一段的细粒沉积岩岩相（孟庆涛等，2024b）。X8 井的岩性主要有页岩和粉砂岩 2 种，矿物根据三端元图可划分为 4 类（图 2），TOC 可划分为高（>2%）、中（1%~2%）和低（<1%），沉积构造可划分为纹层状（<1 cm）和层状（>1 cm）2 种类型（Liu *et al.*, 2019；孟庆涛等，2024a）。最终可将松辽盆地青一段划分为 7 个岩相类型（长英质指长石+石英）：（A）高有机质纹层状黏土质页岩、（B）高有机质纹层状长英质页岩、（C）中有机质纹层状黏土质页岩、（D）中有机质纹层状长英质页岩、（E）低有机质层状长英质页岩、（F）低有机质纹层状混合质页岩和（G）低有机质层状粉砂岩（表 1）。X8 井的岩相垂向分布特征如图 3 所示。

#### 3.2 岩相特征

高有机质纹层状黏土质页岩（A）和中有机质纹层状黏土质页岩（C）均以黏土矿物为主，依次分布在 50%~56% 之间和 52%~56% 之间。A（TOC：2.08%~3.11%）比 C（TOC：1.10%~1.78%）岩相的 TOC 高，这与薄片上见到 A 的富有机质纹层占比高于 C 的现象相对应。高有机质纹层状长英质页岩（B）和中有机质纹层状长英质页岩（D）均以长英质矿物为主，依次分布在 51%~58% 之间和

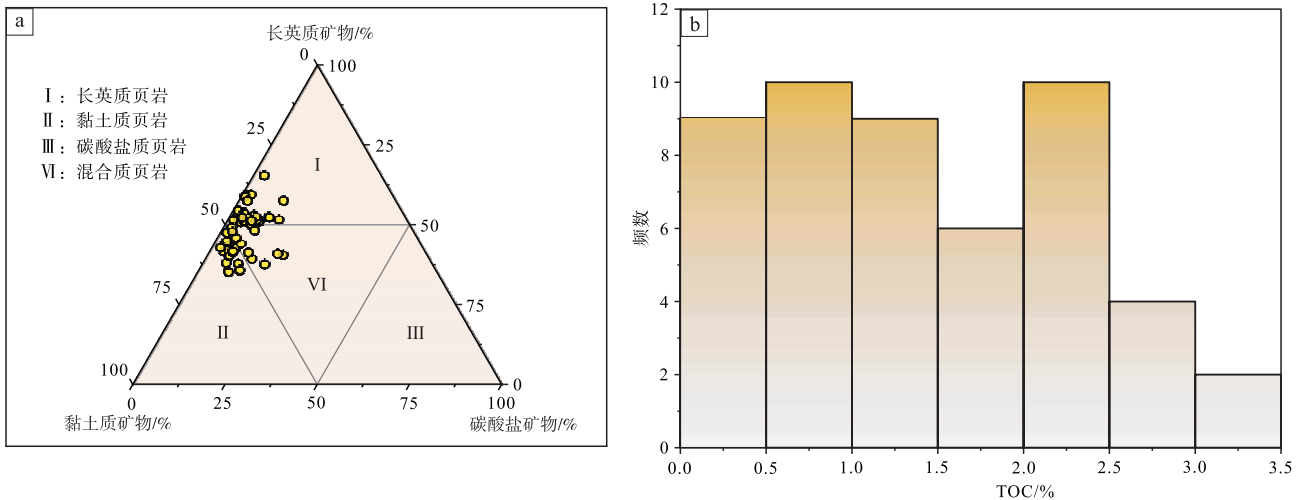

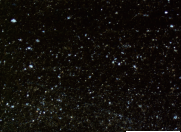

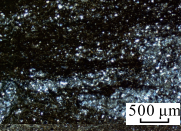
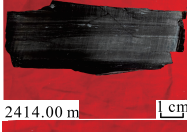
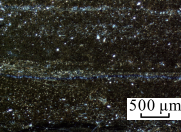
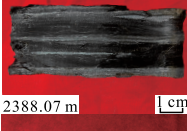
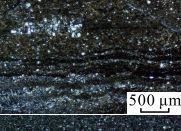
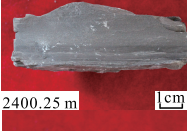
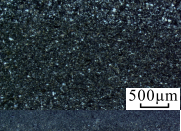
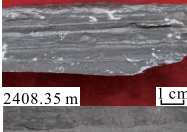
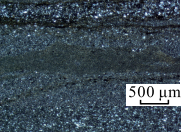
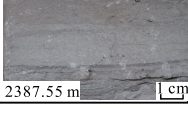
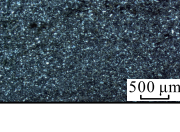


图 2 松辽盆地 X8 井青山口组一段矿物组分 (a) 与 TOC (b) 的分布  
Fig. 2 Distribution of mineral components (a) and TOC (b) of the Member 1 of Qingshankou Formation of Well X8 in Songliao Basin

表 1 松辽盆地 X8 井青山口组一段岩相类型

Table 1 Lithofacies types in the Member 1 of Qingshankou Formation of Well X8 in Songliao Basin

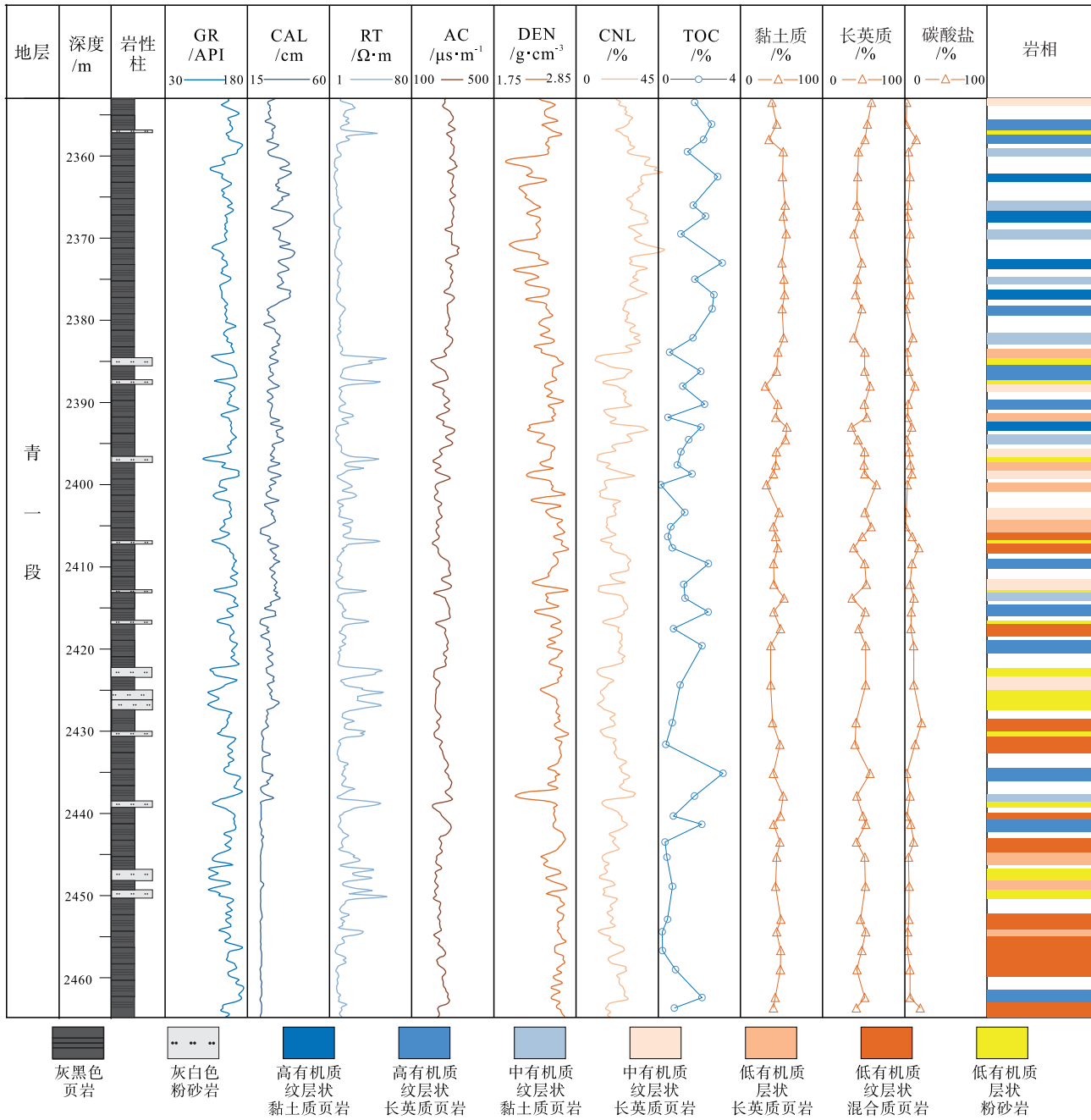
代号	岩相名称	岩心照片	薄片照片	样品测试数 /个	TOC/%	沉积构造	岩性	矿物组分
A	高有机质 纹层状 黏土质页岩	 2393.11 m	 500 μm	5	2.08~3.11	纹层状	页岩	黏土矿物:50%~56% 长英质矿物:35%~48% 碳酸盐矿物:2%~8%
B	高有机质 纹层状 长英质页岩	 2414.63 m	 500 μm	11	2.07~3.15	纹层状	页岩	长英质矿物:51%~58% 黏土矿物:35%~49% 碳酸盐矿物:1%~14%
C	中有机质 纹层状 黏土质页岩	 2414.00 m	 500 μm	8	1.10~1.78	纹层状	页岩	黏土矿物:52%~56% 长英质矿物:36%~43% 碳酸盐矿物:2%~11%
D	中有机质 纹层状 长英质页岩	 2388.07 m	 500 μm	7	1.02~1.77	纹层状	页岩	长英质矿物:51%~60% 黏土矿物:30%~47% 碳酸盐矿物:2%~12%
E	低有机质 层状 长英质页岩	 2400.25 m	 500 μm	8	0.12~0.93	层状	页岩	长英质矿物:50%~65% 黏土矿物:32%~46% 碳酸盐矿物:1%~7%
F	低有机质 纹层状 混合质页岩	 2408.35 m	 500 μm	11	0.20~0.84	纹层状	页岩	黏土矿物:39%~49% 长英质矿物:38%~49% 碳酸盐矿物:2%~20%
G	低有机质 层状粉砂岩	 2387.55 m	 500 μm		<1	层状	粉砂岩	长英质矿物>50% 黏土矿物<50% 碳酸盐矿物<50%

51%~60% 之间。同理, B (TOC: 2.07%~3.15%) 比 D (TOC: 1.02%~1.77%) 岩相的 TOC 高, B 的有机质纹层占比高于 D。低有机质层状长英质页岩 (E) 的 TOC (0.12%~0.93%) 较低, 层状构造, 以长英质矿物为主 (50%~65%)。低有机质纹层状混合质页岩 (F) 的 TOC 分布在 0.20%~0.84% 之间, 黏土、长英质和碳酸盐 3 种矿物分别分布在 39%~49%、38%~49% 和 2%~20% 之间。F 岩相在薄片上可见黏土质纹层和长英质频繁互层。低有机质层状粉砂岩 (G) 的 TOC 小于 1%, 可见石英和长石颗粒较为发育 (表 1)。

## 4 细粒沉积岩岩相测井预测

### 4.1 数据准备与特征分析

收集了 X8 井青一段的 6 条 0.125 m 采样间隔的测井曲线, 分别为 GR、CAL、RT、AC、DEN 和 CNL。建立了 X8 井的岩相—测井数据库 (图 3), 共有 615 个样本点 (表 2)。根据表 2 可明显看到不同岩相之间的岩相占比有差异, F 岩相均占比最高 (24%), A 岩相占比最低 (8%), 这会导致机器学习模型对少数类样本训练不足, 在后面部分详细论述了该问题的解决方法。



注：“岩相”一列中空白表示岩相没有被确定。

图 3 松辽盆地 X8 井青山口组一段测井数据、TOC、矿物组分和岩相分布

Fig. 3 Logging data, TOC, mineral composition and lithofacies distribution in the Member 1 of Qingshankou Formation of Well X8 in Songliao Basin

对以上收集的 6 种测井数据进行变量之间 Person 分析，可以清晰 X8 井的测井之间的相关性。过高的相关性 (Person 接近 1 或 -1) 会影响模型的训练的稳定性 (闫佳飞等, 2025)。因此进行相关性热图分析 (图 4)，发现测井变量之间没有非常强的相关性，仅有 CNL 和 AC 的呈显著的正相关

(Person=0.81)，不会因共线性问题增加模型的复杂度并影响模型的预测精度。

岩相在测井曲线的概率分布和两两组合的散点组合可以定性分析测井对岩相的响应的强弱。图 5 的对角线为 7 个岩相在 6 种测井曲线上的条件概率分布情况，可见 CAL、AC 和 CNL 可显著区分不同

表 2 松辽盆地 X8 井青山口组一段岩相样本数据

Table 2 Lithofacies sample data of Member 1 of the Qingshankou Formation of Well X8 in Songliao Basin

岩相代号	岩相名称	样本点/个	岩相占比/%
A	高有机质纹层状黏土质页岩	48	8
B	高有机质纹层状长英质页岩	122	20
C	中有机质纹层状黏土质页岩	70	11
D	中有机质纹层状长英质页岩	66	11
E	低有机质层状长英质页岩	71	12
F	低有机质纹层状混合质页岩	150	24
G	低有机质层状粉砂岩	88	14
总计		615	100

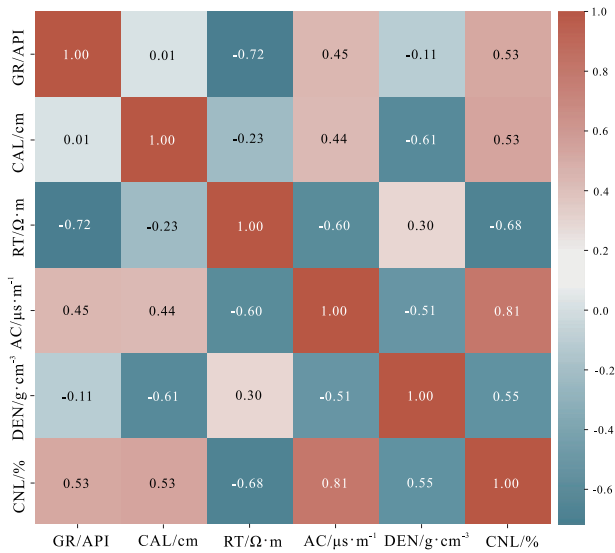


图 4 松辽盆地 X8 井青山口组一段测井数据相关性

Fig. 4 Correlation of logging data of the Member 1 of Qingshankou Formation of Well X8 in Songliao Basin

岩相类别，GR、RT 和 DEN 对岩相的响应较弱一些。上三角和下三角分别为 7 个岩相在两两组合的测井曲线的分类情况，CAL-CNL，GR-RT，RT-DEN，RT-CNL，AC-CNL 和 DEN-CNL 可对岩相有一定的区分能力。但概率分布和两两组合散点图均展示不同岩相在测井上有重叠的区间，常规的线性方法无法准确识别岩相，这是岩相的强非均质性导致的。

### 4.2 模型建立

松辽盆地湖相细粒沉积岩岩相由于对气候和沉积环境的敏感而具有强的非均质性特征，机器学习是解决该类问题的有效方法，提出了被广泛使用且效果较好的 RF、XGBoost 和 SVM 的分类模型，并

用 A、P、R 和 F1 综合评价分类的最优模型。同时面对不平衡分布的岩相样本数据，提出 BSMOTE 法增加少数类样本数量来平衡样本类别的分布，将 BSMOTE 与最优模型组合，来建立松辽盆地青一段的最优岩相预测组合模型。将 X8 井岩相一测井样本按 7:3 定义训练集和测试集。模型的参数是通过 5 折交叉验证的网格搜索法确定的，最优参数值如表 3 所示。

### 4.3 模型预测结果对比

通过对比 RF、XGBoost 和 SVM 的 3 个基础分类器在测试集的预测效果，发现 SVM 为最优模型，A、P、R 和 F1 依次分别为 82.16%、82.64%、82.16% 和 81.73%（表 4），均是 3 个模型的最高，其次为 XGBoost 和 RF 模型。因此选用 BSMOTE-SVM 组合模型进一步对比，分析 BSMOTE 处理后对模型的预测精度的贡献。BSMOTE-SVM 比 3 个基础分类模型拥有更高的分类精度，A、P、R 和 F1 依次分别为 86.49%、86.60%、86.49% 和 86.31%（表 4），这证明了 BSMOTE 是处理不平衡样本的 1 种有效算法。图 6 展示了 X8 井定义的真实岩相和 4 种方法的预测岩相的情况，红色为模型判断错误的标记，可清晰看到 BSMOTE-SVM 的预测效果为最优。图 7 直观的展示了 4 种模型在测试集的综合指标的结果对比，进一步证实了 BSMOTE-SVM 的分类精度。可以有效的解决岩相的非均质性和样本分类不平衡的问题。

### 4.4 模型误差分析

BSMOTE 很好地处理了岩相的分类不平衡的问题，提高了模型的训练精度。同时，SVM 在处理类似本次预测岩相的小样本数据和岩相与测井之间的非线性关系具有很好的优势（薛凯隆等，2024）。以上分析可能是 BSMOTE-SVM 优于 RF、XGBoost 和 SVM 模型的原因。但尽管 BSMOTE-SVM 的预测岩相的 A 达到了 86.49%，该模型在图 6 仍可直观地发现预测错误的岩相样本，造成这种误差的原因可能有以下 2 种：（1）测井的 0.125 m 的分辨率可能太低，当纵向上的矿物、TOC 和沉积构造以厘米尺度快速变化导致岩相频繁变化时，测井对于岩相的精确表达可能受到限制（赖锦等，2024）。（2）地层中的黄铁矿含量也会对测井曲线造成影响，并可能一定程度上掩盖了岩

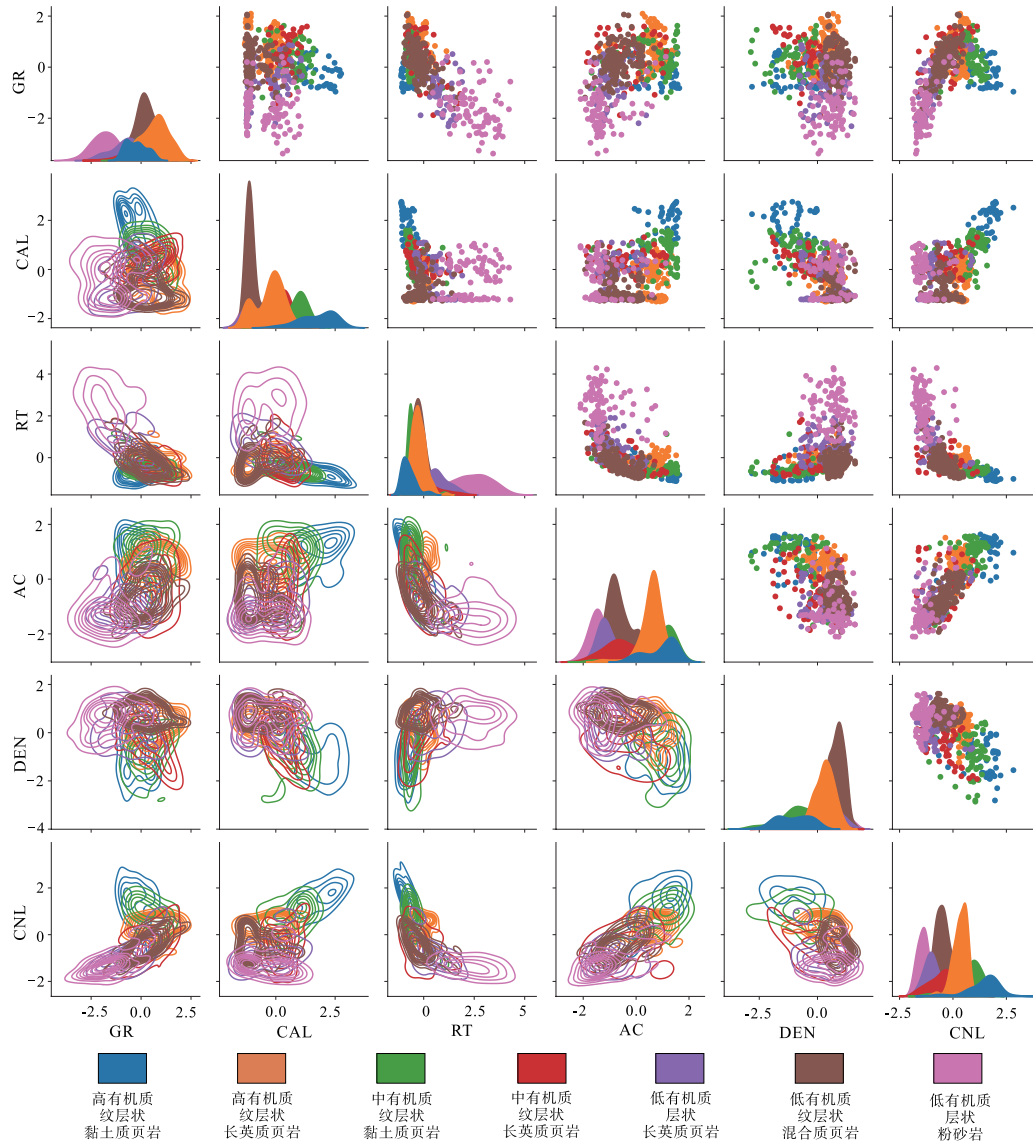


图 5 松辽盆地 X8 井 7 个岩相在 6 种测井曲线的分布特征

Fig. 5 Distribution characteristics of seven lithofacies in six type logging curves of Well X8 in Songliao Basin

表 3 松辽盆地 4 个岩相预测模型参数设置

Table 3 Parameter settings for predictive models of four lithofacies in Songliao Basin

模型	参数	最优参数值	参数搜索范围
RF	树最大树深	5	5 ~ 50
	树数量	60	50 ~ 1000
	节点划分最小样本数	4	2 ~ 20
XGBoost	树最大树深	5	5 ~ 50
	树数量	80	50 ~ 1000
	学习率	0.050	0.010 ~ 0.400
	正则化系数	2.300	0.100 ~ 10.000
SVM/BSMOTE-SVM	正则化参数	1.200	0.001 ~ 10.000
	核函数系数	0.500	0.001 ~ 10.000
	核函数类型	RBF	RBF

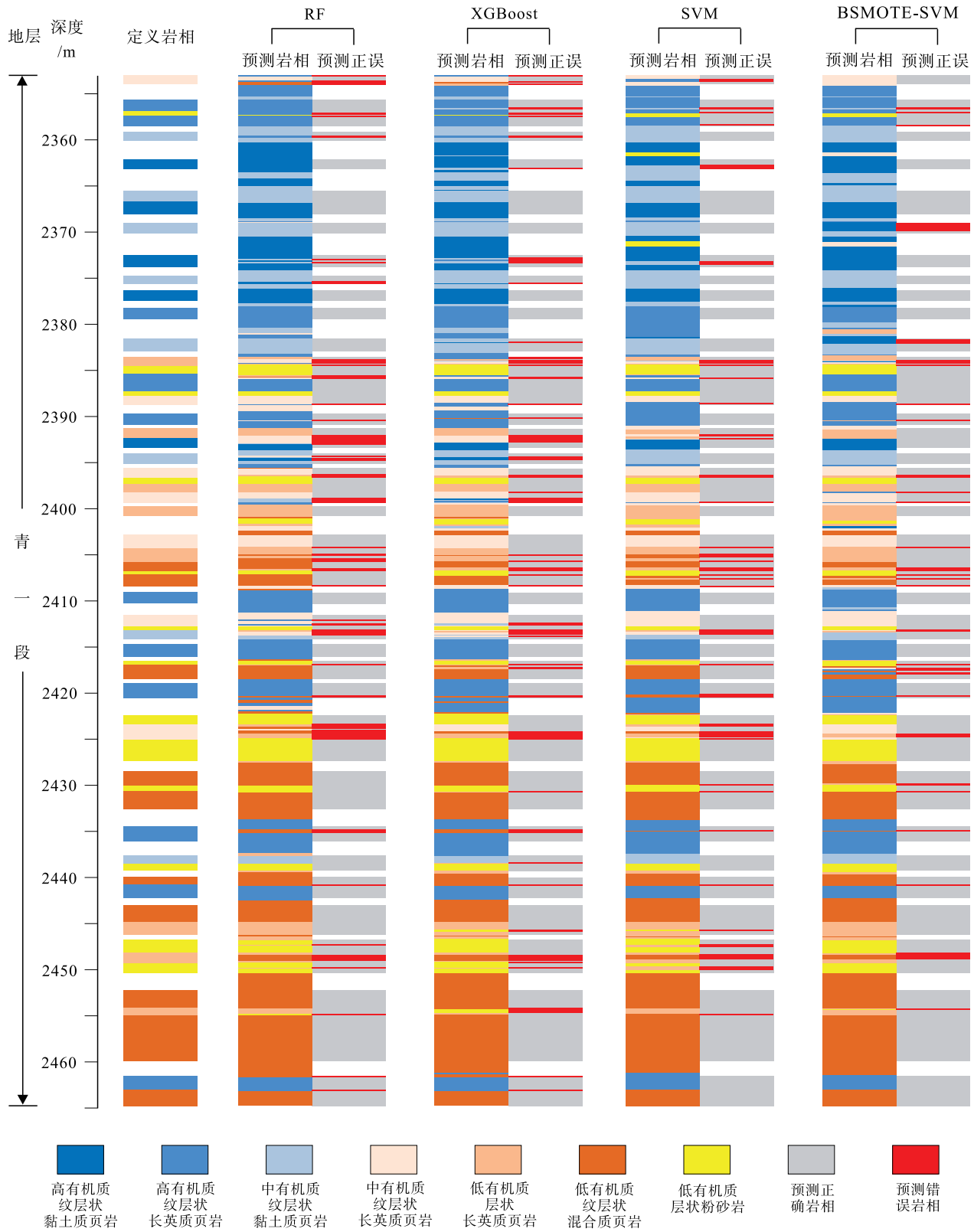


图 6 松辽盆地 X8 井 4 个模型的岩相预测效果

Fig. 6 Lithofacies prediction performance of four models of Well X8 in Songliao Basin

表 4 松辽盆地 4 个模型在测试集的分类预测效果

Table 4 Classification performance of four models on test set in Songliao Basin

岩相预测模型	准确率/%	精确率/%	召回率/%	F1/%
RF	75.68	75.59	75.68	74.58
XGBoost	75.14	74.91	75.14	74.56
SVM	82.16	82.64	82.16	81.73
BSMOTE-SVM	86.49	86.60	86.49	86.31

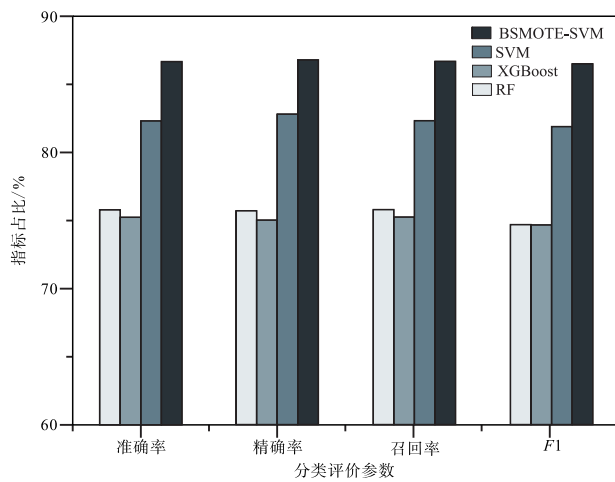


图 7 松辽盆地 X8 井 4 个模型在测试集的分类评价指标  
Fig. 7 Classification evaluation metrics of four models in Well X8 on the test set in Songliao Basin

相的真实测井响应信号 (赖锦等, 2023)。如黄铁矿会增加地层的导电能力, 导致 RT 减小, 并增大岩石密度, 进而增大 DEN 值。

#### 4.5 岩相平面分布的预测

基于前文研究确认了 BSMOTE-SVM 为岩相最优预测模型。将 BSMOTE-SVM 推广应用在另外的 29 口井, 在平面上确定了岩相的分布规律 (图 8)。松辽盆地长岭凹陷岩相的平面分布受西南方向物源的影响。从西南物源向凹陷沉积中心延伸的过程中, 砂体厚度逐渐减小、水动力条件逐步减弱, 同时水体深度不断增加。这 3 者的共同作用, 使得沿此方向的黏土矿物含量和 TOC 呈现出递增趋势。靠近物源区域主要发育低有机质层状粉砂岩。在三角洲外前缘主要发育低有机质层状长英质页岩和低有机质纹层状混合质页岩。乾安附近为半深湖相, 发育中有机质纹层状黏土质页岩和中有机质纹层状长英质页岩。继续向大安方向则为深湖的沉积相, 沉积着高有机质纹层状黏土质页岩和高有机质纹层

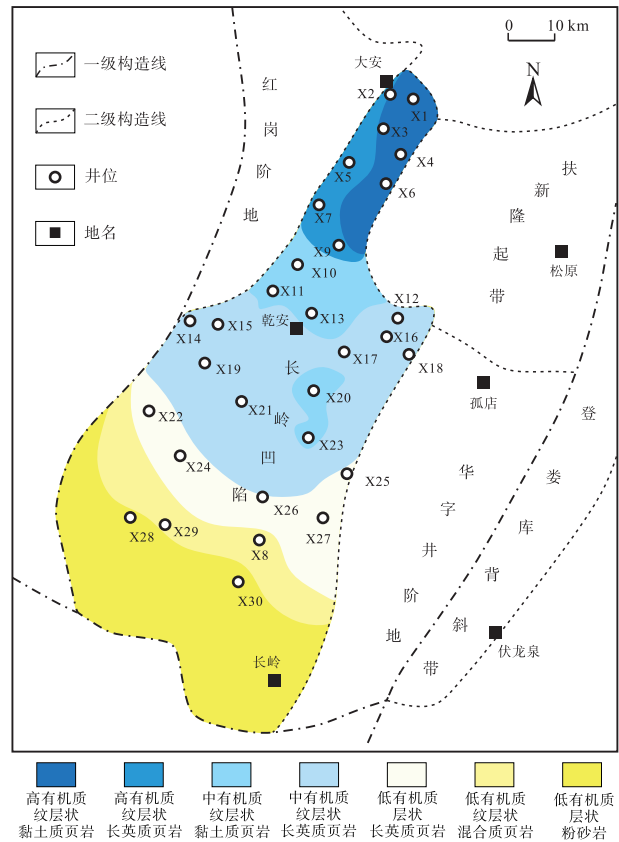


图 8 松辽盆地长岭凹陷青一段岩相分布  
Fig. 8 Lithofacies distribution of the Member 1 of Qingshankou Formation in Changling Sag of Songliao Basin

状长英质页岩。

### 5 结论

1) 在松辽盆地青山口组一段建立了“岩性+矿物+TOC+沉积构造”的岩相分类方案, 确定了高有机质纹层状黏土质页岩、高有机质纹层状长英质页岩、中有机质纹层状黏土质页岩、中有机质纹层状长英质页岩、低有机质层状长英质页岩、低有机质层状混合质页岩和低有机质层状粉砂岩 7 类细粒沉积岩岩相。

2) 使用 BSMOTE 法处理岩相分类不平衡的样本, 并与最优分类器 SVM 组合建立了 BSMOTE-SVM 模型, 岩相预测效果明显优于 RF、XGBoost 和 SVM 模型, 取得了较高的精度, A、P、R 和 F1 依次分别为 86.49%、86.60%、86.49% 和 86.31%。

3) BSMOTE-SVM 可以快速的预测多井的岩相, 进而确定了松辽盆地长岭凹陷青一段的岩相分

布。岩相的分布主要受西南方向的物源影响,长岭凹陷西南至东北方向依次发育低有机质层状粉砂岩、低有机质纹层状混合质页岩、低有机质层状长英质页岩、中有机质纹层状长英质页岩、中有机质纹层状黏土质页岩、高有机质纹层状长英质页岩和高有机质纹层状黏土质页岩。

### 参考文献 (References)

- 段宏亮,杨保良. 2024. 陆相细粒沉积岩岩相组合类型划分研究现状及其油气地质意义. 复杂油气藏, 17(1): 1-10. [Duan H L, Yang B L. 2024. Research status of lithofacies assemblage types classification of continental fine-grained sedimentary rocks and their hydrocarbon geological significance. *Complex Hydrocarbon Reservoirs*, 17(1): 1-10]
- 贾承造,王祖纲,姜林,赵文. 2024. 中国页岩油勘探开发研究进展与科学技术问题. 世界石油工业, 31(4): 1-11, 13. [Jia C Z, Wang Z G, Jiang L, Zhao W. 2024. Progress and key scientific and technological problems of shale oil exploration and development in China. *World Petroleum Industry*, 31(4): 1-11, 13]
- 李士超,杨建国,柳波,姚玉来,肖飞,白龙辉,黄一鸣,李昂,张丽艳. 2021. 松辽盆地三肇凹陷青山口组一段泥页岩岩石学特征及岩相划分: 以松页油3井为例. 地质与资源, 30(3): 317-324, 295. [Li S C, Yang J G, Liu B, Xiao F, Bai L H, Huang Y M, Li A, Zhang L Y. 2021. Petrology and lithofacies of shale from the first member of Qingshankou Formation in Sanzhao Sag, Songliao Basin: a case study of SYY-3 well. *Geology and Resources*, 30(3): 317-324, 295]
- 姜在兴,梁超,吴靖,张建国,张文昭,王永诗,刘惠民,陈祥. 2013. 含油气细粒沉积岩研究的几个问题. 石油学报, 34(6): 1031-1039. [Jiang Z X, Liang C, Wu J, Zhang J G, Zhang W Z, Wang Y S, Liu H M, Chen X. 2013. Several issues in sedimentological studies on hydrocarbon-bearing fine-grained sedimentary rocks. *Acta Petrolei Sinica*, 34(6): 1031-1039]
- 姜在兴,张建国,孔祥鑫,谢环羽,程浩,王力. 2023. 中国陆相页岩油气沉积层研究进展及发展方向. 石油学报, 44(1): 45-71. [Jiang Z X, Zhang J G, Kong X X, Xie H Y, Cheng H, Wang L. 2023. Research progress and development direction of continental shale oil and gas deposition and reservoirs in China. *Acta Petrolei Sinica*, 44(1): 45-71]
- 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. 古地理学报, 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]
- 赖锦,白天宇,苏洋,赵飞,李玲,黎雨航,李红斌,王贵文,肖承文. 2024. 烃源岩测井识别与评价方法研究进展. 地质论评, 70(2): 721-741. [Lai J, Bai T Y, Su Y, Zhao F, Li L, Li Y H, Li H B, Wang G W, Xiao C W. 2024. Researches progress in well log recognition and evaluation of source rocks. *Geological Review*, 70(2): 721-741]
- 李国萃,石巨业,樊太亮,胡德胜,游君君,李一凡,高志前,邓成昆,范家豪,周刚. 2023. 天文周期约束下始新统湖相地层页岩岩相组合类型及其发育模式: 以北部湾盆地涠西南凹陷流沙港组为例. 第四纪研究, 43(6): 1614-1629. [Li G C, Shi J Y, Fan T L, Hu D S, You J J, Li Y F, Gao Z Q, Deng C K, Fan J H, Zhou G. 2023. Shale lithofacies association types and development models in eocene lacustrine facies strata constrained by astronomical cycles: a case study of the Liushagang Formation in the Weixinan sag, Beibuwan Basin. *Quaternary Sciences*, 43(6): 1614-1629]
- 李红斌,王贵文,王松,庞小娇,刘士琛,包萌,彭寿昌,赖锦. 2022. 基于 Kohonen 神经网络的页岩油岩相测井识别方法: 以吉木萨尔凹陷二叠系芦草沟组为例. 沉积学报, 40(3): 626-640. [Li H B, Wang G W, Wang S, Pang X J, Liu S C, Bao M, Peng S C, Lai J. 2022. Shale oil lithofacies identification by Kohonen neural network method: the case of the Permian Lucaogou Formation in Jimusar Sag. *Acta Sedimentologica Sinica*, 40(3): 626-640]
- 林兴悦,朱筱敏,王晓琳,张美洲. 2025. 细粒沉积岩优质储集层发育主控因素分析: 以渤海湾盆地沾化凹陷为例. 古地理学报, 27(1): 55-71. [Lin X Y, Zhu X M, Wang X L, Zhang M Z. 2025. Reservoir characteristics and distribution of lacustrine fine-grained sedimentary rocks: a case study from Zhanhua sag, Bohai Bay Basin, China. *Journal of Palaeogeography (Chinese Edition)*, 27(1): 55-71]
- 鲁玲,高诚,熊威,龚康,马辉,张鑫. 2024. 基于 CD-BSMOTE 的 D-S 证据融合变压器故障诊断. 水电能源科学, 42(5): 192-196. [Lu L, Gao C, Xing W, Gong K, Ma H, Zhang X. 2024. CD-BSMOTE based D-S evidence fusion transformer fault diagnosis. *Water Resources and Power*, 42(5): 192-196]
- 孟庆涛,胡菲,刘招君,孙平昌,柳蓉. 2024a. 陆相拗陷湖盆细粒沉积岩岩相类型及成因: 以松辽盆地晚白垩世青山口组为例. 吉林大学学报(地球科学版), 54(1): 20-37. [Meng Q T, Hu F, Liu Z J, Sun P C, Liu R. 2024. Lithofacies types and genesis of fine-grained sediments in Terrestrial Depression Lake Basin: taking Upper Cretaceous Qingshankou Formation in Songliao Basin as an example. *Journal of Jilin University (Earth Science Edition)*, 54(1): 20-37]
- 孟庆涛,张训,杨亮,高家俊,刘招君,胡菲,邢济麟,张成铭,康嘉楠,崔博,董秦玮,张恩威. 2024b. 陆相拗陷湖盆细粒沉积有机质富集特征及控制因素研究: 以松辽盆地南部长岭凹陷青山口组一段为例. 古地理学报, 26(2): 401-415. [Meng Q T, Zhang X, Yang L, Gao J J, Liu Z J, Hu F, Xing J L, Zhang C M, Kang J N, Cui B, Dong Q W, Zhang E W. 2024. Characterization and controlling factors of organic matter enrichment in fine-grained sediments in continental depression lacustrine basin: a case study of the Member 1 of Qingshankou Formation in Changling sag, southern Songliao Basin. *Journal of Palaeogeography (Chinese Edition)*, 26(2): 401-415]
- 沈华,杨亮,韩昊天,王颖,邢济麟,薛松,刘红超. 2023. 松辽盆地南部油气勘探新领域、新类型及资源潜力. 石油学报, 44(12): 2104-2121. [Shen H, Yang L, Han H T, Wang Y, Xing J L, Xue S, Liu H C. 2023. New fields, new types and resource potentials of oil-gas exploration in southern Songliao Basin. *Acta Petrolei Sinica*, 44(12): 2104-2121]
- 谌丽,王才志,宁从前,刘英明,王浩. 2023. 基于机器学习的鄂尔多斯盆地陇东地区长7段岩相测井识别方法. 油气藏评价与开发, 13(4): 525-536. [Shen L, Wang C Z, Ning C Q, Liu Y M,

- Wang H. 2023. Well-log lithofacies classification based on machine learning for Chang-7 Member in Longdong area of Ordos Basin. *Petroleum Reservoir Evaluation and Development*, 13(4): 525-536]
- 孙龙德, 刘合, 何文渊, 李国欣, 张水昌, 朱如凯, 金旭, 孟思炜, 江航. 2021. 大庆古龙页岩油重大科学问题与研究路径探析. *石油勘探与开发*, 48(3): 453-463. [Sun L D, Liu H, He W Y, Li G X, Zhang S C, Zhu R K, Jin X, Meng S W, Jiang H. 2021. An analysis of major scientific problems and research paths of Gulong shale oil in Daqing Oilfield, NE China. *Petroleum Exploration and Development*, 48(3): 453-463]
- 孙龙德, 贾承堃, 张君峰, 崔宝文, 白静, 霍秋立, 徐兴友, 刘卫彬, 曾花森, 刘伟. 2024. 松辽盆地古龙页岩油重点地区资源潜力. *石油学报*, 45(12): 1699-1714. [Sun L D, Jia C Z, Zhang J F, Cui B Q, Bai J, Huo Q L, Xu X Y, Liu W B, Zeng H S, Liu W. 2024. Resource potential of Gulong shale oil in the key areas of Songliao Basin. *Acta Petrolei Sinica*, 45(12): 1699-1714]
- 汪益宁, 闫荣堃, 罗佳洁, 欧阳静芸, 段秋红, 徐涛. 2016. 基于支持向量机的致密储层岩相识别: 以徐家围子断陷下白垩统沙河子组为例. *长江大学学报(自科版)*, 13(29): 33-38, 5. [Wang Y N, Yan R K, Luo J J, Ouyang J Y, Duan Q H, Xu T. 2016. Tight reservoir lithofacies identification based on Support Vector Machine: by taking Shahezi Formation of Xujiaweizi Faulted Depression of Lower Cretaceous for example. *Journal of Yangtze University(Natural Science Edition)*, 13(29): 33-38, 5]
- 王民, 杨金路, 王鑫, 李进步, 徐亮, 言语. 2023. 基于随机森林算法的泥页岩岩相测井识别. *地球科学*, 48(1): 130-142. [Wang M, Yang J L, Wang X, Li J B, Xu L, Yan Y. 2023. Identification of shale lithofacies by well logs based on Random Forest algorithm. *Earth Science*, 48(1): 130-142]
- 王伟, 王振林, 刘财广, 郑孟林, 张融, 郑国庆, 余佩蓉. 2023. 页岩油甜点评价关键技术及甜点类型划分: 以玛湖凹陷二叠系风城组为例. *地球科学*, 48(1): 223-234. [Wang W, Wang Z L, Liu C G, Zheng M L, Zhang R, Zheng G Q, Yu P R. 2023. Key technology of shale oil sweet spot evaluation and sweet spot type division in Fengcheng Formation of Mahu sag. *Earth Science*, 48(1): 223-234]
- 吴靖, 姜在兴, 梁超. 2017. 东营凹陷沙河街组四段上亚段细粒沉积岩岩相特征及与沉积环境的关系. *石油学报*, 38(10): 1110-1122. [Wu J, Jiang Z X, Liang C. 2017. Lithofacies characteristics of fine-grained sedimentary rocks in the upper submember of Member 4 of Shahejie Formation, Dongying sag and their relationship with sedimentary environment. *Acta Petrolei Sinica*, 38(10): 1110-1122]
- 武瑾, 曾凡成, 唐晴, 杨立民, 邱振, 赵圣贤, 邹晓品, 李伟华. 2025. 川南地区五峰组—龙马溪组页岩岩相特征及沉积环境演化. *古地理学报*, 27: 1-17. DOI:10.7605/gdxb.2025.07. [Wu J, Zeng F C, Tang Q, Yang L M, Qiu Z, Zhao S X, Zou X Q, Li W H. 2025. Lithofacies characteristics and sedimentary environment evolution of the Wufeng-Longmaxi Formation shale in southern Sichuan Basin. *Journal of Palaeogeography(Chinese Edition)*, 27: 1-17. DOI:10.7605/gdxb.2025.07]
- 蒯克来, 操应长, 朱如凯, 邵雨, 薛秀杰, 王小军, 高阳, 张景. 2015. 吉木萨尔凹陷二叠系芦草沟组致密油储层岩石类型及特征. *石油学报*, 36(12): 1495-1507. [Xi K L, Cao Y C, Zhu R K, Shao Y, Xue X J, Wang X J, Gao Y, Zhang J. 2015. Rock types and characteristics of tight oil reservoir in Permian Lucaogou Formation, Jimsar sag. *Acta Petrolei Sinica*, 36(12): 1495-1507]
- 薛凯隆, 崔欣超, 祁云, 齐庆杰. 2024. 基于 DBO-SVM 的采空区煤自燃危险性预测. *沈阳理工大学学报*, 43(6): 85-90. [Xue K L, Cui X C, Qi Y, Qi Q J. 2024. Prediction of spontaneous combustion fire in goaf based on DBO-SVM. *Journal of Shenyang Ligong University*, 43(6): 85-90]
- 闫佳飞, 李胜利, 魏泽德, 吴忠宝, 陈建阳. 2025. 基于 XGBoost 算法的页岩岩相测井预测方法. *古地理学报*, 27(3): 763-776. [Yan J F, Li S L, Wei Z D, Wu Z B, Chen J Y. 2025. Shale lithofacies prediction method with well-logging data based on XGBoost algorithm. *Journal of Palaeogeography(Chinese Edition)*, 27(3): 763-776]
- 姚玉来, 肖飞, 李士超, 杨建国, 公繁浩, 王杰. 2025. 松辽盆地齐家凹陷南部青山口组一段古环境及其对有机质富集的控制作用. *中国石油大学学报(自然科学版)*, 49(1): 59-71. [Yao Y L, Xiao F, Li S C, Yang J G, Gong F H, Wang J. 2025. Paleoenvironment of the first member of Qingshankou Formation in southern Qijia Sag, Songliao Basin and its control on enrichment of organic matter. *Journal of China University of Petroleum(Edition of Natural Science)*, 49(1): 59-71]
- 朱国文, 王小军, 张金友, 刘召, 白云风, 赵莹, 付秀丽, 曾花森. 2023. 松辽盆地陆相页岩油富集条件及勘探开发有利区. *石油学报*, 44(1): 110-124. [Zhu G W, Wang X J, Zhang J Y, Liu Z, Bai Y F, Zhao Y, Fu X L, Zeng H S. 2023. Enrichment conditions and favorable zones for exploration and development of continental shale oil in Songliao Basin. *Acta Petrolei Sinica*, 44(1): 110-124]
- 朱如凯, 李梦莹, 杨静儒, 张素荣, 蔡毅, 曹琰, 康缘. 2022. 细粒沉积学研究进展与发展方向. *石油与天然气地质*, 43(2): 251-264. [Zhu R K, Li M Y, Yang J R, Zhang S R, Cai Y, Cao Y, Kang Y. 2022. Advances and trends of fine-grained sedimentology. *Oil & Gas Geology*, 43(2): 251-264]
- 邹才能, 潘松圻, 荆振华, 高金亮, 杨智, 吴松涛, 赵群. 2020. 页岩油气革命及影响. *石油学报*, 41(1): 1-12. [Zou C N, Pan S Q, Jing Z H, Gao J L, Yang Z. 2020. Shale oil and gas revolution and its impact. *Acta Petrolei Sinica*, 41(1): 1-12]
- Liu B, Liu L B, Fu J, Lin T F, He J L, Liu X Z, Liu Y C, Fu X F. 2023. The Songliao Super Basin in northeastern China. *AAPG Bulletin*, 107(8): 41.
- Liu B, Wang H L, Fu X F, Bai Y F, Bai L H, Jia M C, He B. 2019. Lithofacies and depositional setting of a highly prospective lacustrine shale oil succession from the Upper Cretaceous Qingshankou Formation in the Gulong sag, northern Songliao Basin, northeast China. *AAPG Bulletin*, 103(2): 405-432.
- Wang J X, Xu Y B, Sun P C, Liu Z J, Zhang J Q, Meng Q T, Zhang P L, Tang B Q. 2022. Prediction of organic carbon content in oil shale based on logging: a case study in the Songliao Basin, Northeast China. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 8(2): 1-23.
- Zhu X M, Cao J, Xia L W, Bian L Z, Liu J C, Zhang R J. 2023. Links between marine incursions, lacustrine anoxia and organic matter enrichment in the Upper Cretaceous Qingshankou Formation, Songliao Basin, China. *Marine and Petroleum Geology*, 158(Part A): 106536.