Advances in Geo-Energy Research

Vol. 17, No. 2, p. 149-161, 2025

Original article

A method for enhancing well-log resolution of thin lithological heterogeneities using wavelet transform and automated machine learning

Wei Li[1,2], Lei Liu[1,2], Dali Yue[1,2]*, Jian Gao[3], Shanyan Zhang[3], Luca Colombera[4,5]

[1]*State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum, Beijing 102249, P. R. China*

[2]*College of Geosciences, China University of Petroleum, Beijing 102249, P. R. China*

[3]*Research Institute of Petroleum Exploration & Development, PetroChina, Beijing 100083, P. R. China*

[4]*Department of Earth and Environmental Sciences, University of Pavia, Pavia 27100, Italy*

[5]*School of Earth and Environment, University of Leeds, Leeds LS2 9JT, United Kingdom*

Keywords:
Well-log interpretation
permeability barrier
clastic reservoir
thin lithology
flow baffle
wavelet transform

Cited as:
Li, W., Liu, L., Yue, D., Gao, J., Zhang, S., Colombera, L. A method for enhancing well-log resolution of thin lithological heterogeneities using wavelet transform and automated machine learning. Advances in Geo-Energy Research, 2025, 17(2): 149-161. https://doi.org/10.46690/ager.2025.08.06

Abstract:
Clastic reservoirs exhibit complex and diverse lithologies. Some lithological heterogeneities, occurring as thin but effectively low-permeability units, have pronounced impact on CO_2 flooding schemes and oil recovery. Thin low-permeability units within permeable sandbodies typically exhibit weak well-log responses, and are therefore of difficult recognition using conventional well-log analysis methods. To address this challenge, a hierarchical method is proposed for interpreting thin lithological heterogeneities by integrating wavelet transform and machine learning. The discrete wavelet transform enhances well-log responses of thin heterogeneities. An automated machine-learning framework is designed, which integrates multiple algorithms and achieves automated parameter optimization. This machine-learning method is then applied to well logs to establish a nonlinear mapping model between lithology and well-log responses. Additionally, the hierarchical nature of the workflow highlights lithological contrasts, facilitating a more accurate lithological differentiation by dividing the recognition of thin heterogeneities into three levels. Benefiting from these three advantages, the proposed method offers potential to significantly enhance the accuracy of well-log interpretations. The results demonstrate that this method yields accurate identification of lithological units as thin as 0.2 m for muddy beds and 0.3 m for diagenetic units, achieving a recognition accuracy exceeding the conventional well-log interpretations. This method also shows significant potential for broader applications, including the identification of other types of geological entities of limited thickness, and determination of reservoir parameters at fine scales.

1. Introduction

Clastic reservoirs exhibit complex internal architectures, characterized by various sedimentary and diagenetic features that influence their porosity, permeability, and overall effectiveness as hydrocarbon reservoirs (Stanistreet and Stollhofen, 2002; Yue et al., 2018; Nyberg et al., 2023). Among these features, the presence of permeability barriers within sandbodies is significant as a control on reservoir behavior (Li et al., 2011), with a particularly pronounced impact on CO_2 flooding schemes and oil recovery (Wang et al., 2011). These permeability barriers can be broadly categorized into

Yandy Scientific Press

*Corresponding author.
E-mail address: wei_li@cup.edu.cn (W. Li); liulei@student.cup.edu.cn (L. Liu); yuedali@cup.edu.cn (D. Yue); gj940116@petrochina.com.cn (J. Gao); zhangshy@petrochina.com.cn (S. Zhang); luca.colombera@unipv.it (L. Colombera).

Received June 29, 2025; revised July 16, 2025; accepted August 3, 2025; available online August 6, 2025.

two types: Those of depositional origin and those resulting from diagenesis (Yue et al., 2018; Zhang et al., 2023).

Thin mudstone or siltstone layers within predominantly sandy sedimentary units commonly serve as permeability barriers of depositional origin, resulting from variations in hydrodynamic conditions of the formative flows during deposition of clastic reservoir successions (Miall, 1985; Strick et al., 2019; Li et al., 2023; Liu et al., 2024). Such thin muddy beds are commonly observed in various depositional environments and related sub-environments, such as point bars of meandering river, mid-channel bars of braided rivers, and deltaic mouth bars. These muddy beds are commonly less than 1 meter in thickness and in many cases less than 0.5 m thick (Stanistreet and Stollhofen, 2002); as such, their identification using conventional interpretation methods based on well logs is challenging (Saneifar et al., 2015; Wood, 2022b).

Additionally, diagenetic processes such as cementation can cause a marked decrease in the porosity and permeability of the primary sandstones, transforming permeable lithologies prone to cementation into barriers or baffles to fluid flow (Lynds and Hajek, 2006). Identifying these diagenetic features in well logs is particularly challenging, since these domains occur in sandstone volumes and are of limited thickness (Iraji et al., 2023; Wood, 2023). Both muddy and diagenetic permeability barriers play a pivotal role in the heterogeneity of clastic reservoirs and profoundly influence fluid-flow dynamics (Lynds and Hajek, 2006; Zecchin and Catuneanu, 2017). Therefore, understanding their distribution in clastic reservoirs is essential for reservoir characterization and prediction (Li et al., 2011). However, conventional lithologic interpretation methods based on well logs commonly struggle to identify these relatively thin features in thick sandbodies, due to resolution limitations (Lai et al., 2018; Pham et al., 2020; Wood, 2022a; Iraji et al., 2023). Therefore, enhancing the resolution of lithologic well-log interpretation is crucial for effectively characterizing permeability heterogeneity.

Recent advancements suggest that applying wavelet transforms to well logs and incorporating intelligent algorithms into well-log interpretation are effective approaches for improving the accuracy of lithology interpretation (Chandrasekhar and Rao, 2012; Zhang et al., 2018; Chen et al., 2021). The wavelet transform is a signal processing technique, which converts a signal into a different form. Using an appropriate algorithm, such as the Mallat algorithm, a complex well log can be decomposed into a set of basic signals with finite bandwidth (Mallat and Zhang, 1993; Chen et al., 2021); these basic signals can then be reconstructed. By appropriately increasing the proportion of high-frequency information during well-log reconstruction, the response of the well log to the occurrence of thin flow barriers can be enhanced (Chandrasekhar and Rao, 2012; Zhang et al., 2018; Chen et al., 2021). Machine learning with supervised algorithms can be applied to establish nonlinear mapping relationships between well-log responses and lithological types, extracting information from multiple logs fully and improving the accuracy of well-log interpretation (Feng et al., 2021; Zhang et al., 2021; Iraji et al., 2023; Wei et al., 2025). Both techniques have been considered effective in aiding the identification of thin permeability barriers in previous research (Chandrasekhar and Rao, 2012; Kuang et al., 2021; Zhang et al., 2023).

In recent years, the Amazon Web Services team has released the automated machine learning (AutoML) framework AutoGluon (Erickson et al., 2020), which automates feature engineering, model selection, and hyperparameter optimization. Therefore, this framework enables efficient end-to-end modeling while substantially reducing manual tuning costs and improving both model performance and reproducibility (Erickson et al., 2020; Zhang et al., 2024). This algorithm has demonstrated robust and accurate performance in supervised learning tasks involving structured data (Qi et al., 2021; Sun et al., 2023; Papík and Papíková, 2025; Yang et al., 2025). At present, applications of this machine learning framework in lithology interpretation from well logs remain scarce, yet it offers a novel perspective for integrating machine learning into well-log interpretation.

To address the challenges in identifying thin lithological heterogeneities below resolution limits, this study proposes a novel hierarchical interpretation method for well logs by combining wavelet transforms and AutoML. This method is applied to a subsurface dataset from the Niujuanhu Oilfield, and the results demonstrate a significant improvement in well-log resolution, enabling the accurate identification of thin low-permeability beds.

2. Dataset and methodology

2.1 Dataset

The dataset used to evaluate the proposed method comprises of 225 wells, including seven cored wells from the target interval with a total core length of approximately 400 m. These cores were systematically sampled and analyzed to determine key reservoir properties such as grain size, porosity, and permeability. Each well is equipped with a comprehensive suite of conventional well logs, including Gamma Ray (GR), Sonic (DT), Density (DEN), Deep Resistivity (RD), Shallow Resistivity (RS), and Spontaneous Potential (SP).

2.2 Integrated workflow

Conventional well-log interpretation methods, such as the analysis of GR and SP log crossplots, have proven effective in accurately identifying thick sandstones and mudstones. Therefore, GR crossplots were initially employed to distinguish between sand and shale at first. However, this method has not proven capable of detecting thin impermeable units occurring within thick sandstone bodies. To address this problem, a novel technical workflow was further developed (Fig. 1). This proposed workflow consists of five key steps:

1) Core interpretation. Core descriptions combined with experimental analyses are used to identify the types of low-permeability features present in the sandbodies and to characterize the well-log responses of each type.
2) Well-log processing. Discrete wavelet decomposition and reconstruction techniques (Chandrasekhar and Rao, 2012) are applied to enhance the log responses of thin, low-permeability lithologies. Both the original and recon-

structed well logs are used as training data.

3) Sample equalization. The number of low-permeability unit samples is considerably smaller than that of effective reservoir samples. In machine learning, when the minority sample size is too small, these samples may be treated as outliers, making sample balancing essential. Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) is employed to balance the samples (Fig. 2).
4) Identification of thin heterogeneities. An AutoML framework is developed that integrates multiple algorithms and performs automated parameter optimization. This framework is then applied to identify thin impermeable units within sandstones.
5) Heterogeneity classification. The identified low-permeability units were further classified into three categories using the AutoML framework: Mudstone, carbonate-cemented, and "diagenetic units", i.e., units exhibiting a diagenetic imprint different from calcareous cementation. Although carbonate-cemented sandstone is essentially a subtype of diagenetic unit, it is treated as a separate category due to its distinctive well-log responses and distribution patterns.

2.3 Decomposition and reconstruction of well logs

As a signal processing method, the wavelet transform alters the representation of a signal and enables the examination of the signal at multiple scales. This makes it especially useful for revealing details concealed in well logs. The kernel function of the wavelet transform is defined as (Olkkonen, 2011):

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right) \tag{1}$$

where t is the space variable, $\varphi(t)$ is the mother wavelet, a is the scale factor that determines the wavelength, and b is a wavelet shift parameter.

Continuous wavelet transforms and discrete wavelet transforms (DWT) are common types of wavelet transforms. DWT is chosen due to its extensive application in well-log signal analysis (Chandrasekhar and Rao, 2012; Zhang et al., 2018; Yang et al., 2023). In DWT, variation of scale and shift is represented by an integer:

$$W_f(i,j) = \frac{1}{\sqrt{2^i}}\int_t f(t)\varphi\left(\frac{t-2^i j}{2^i}\right)\mathrm{d}t \tag{2}$$

where $W_f(i,j)$ is the wavelet coefficient at scale i and translation j, which represents the DWT result for the signal given $i, j \in \mathbb{Z}$; $f(t)$ is the input signal. The scale parameter i controls the dilation or compression of the wavelet. Larger i corresponds to a wider wavelet, while smaller i corresponds to a narrower wavelet. The shift parameter j determines the position of the wavelet in space for well logs.

The Mallat decomposition algorithm is employed to perform the DWT. The Mallat algorithm, is a fundamental technique for multiresolution analysis (Mallat and Zhang, 1993; Chen et al., 2021). A three-level decomposition is applied to the well logs, as illustrated in Figs. 3 and 4. cLn and cHn are defined to represent the higher-frequency and lower-frequency

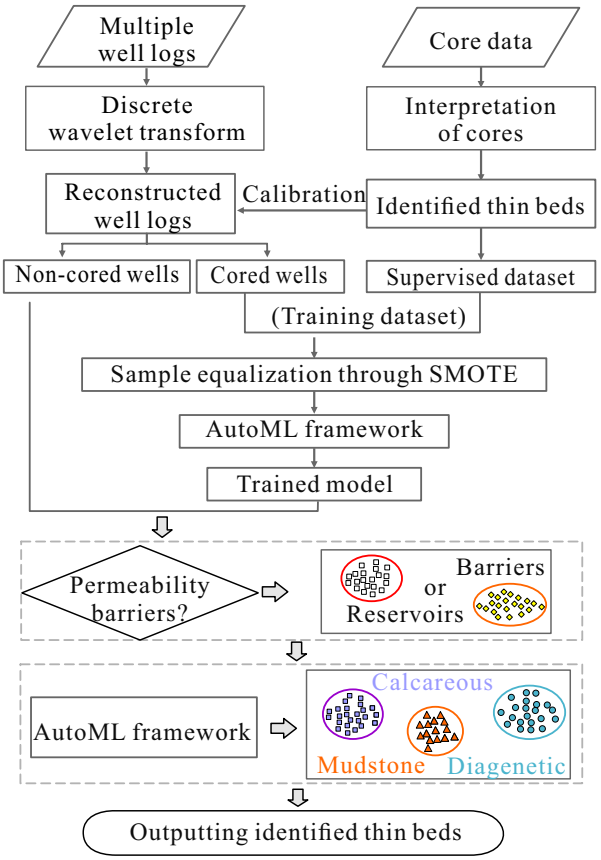

Fig. 1. Integrated workflow diagram outlining how to identify relatively thin permeability barriers in thick sandstone bodies.

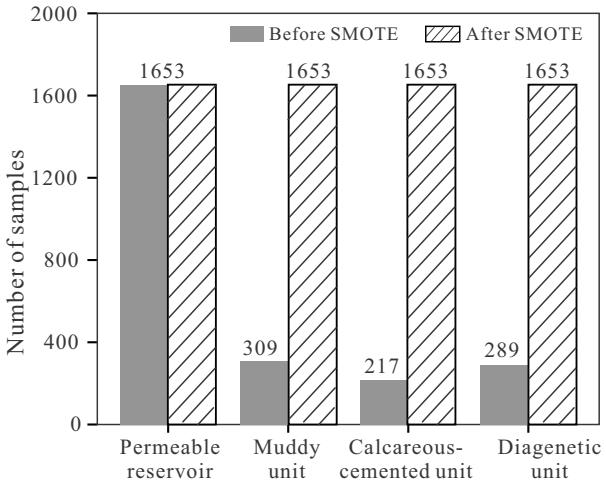

Fig. 2. Equalized samples using SMOTE.

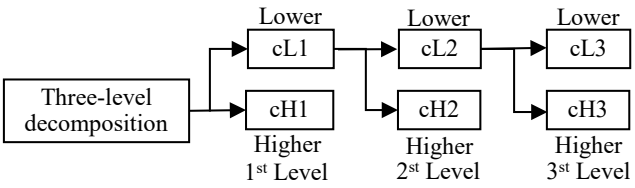

Fig. 3. Diagram outlining decomposition of well logs using the DWT algorithm.

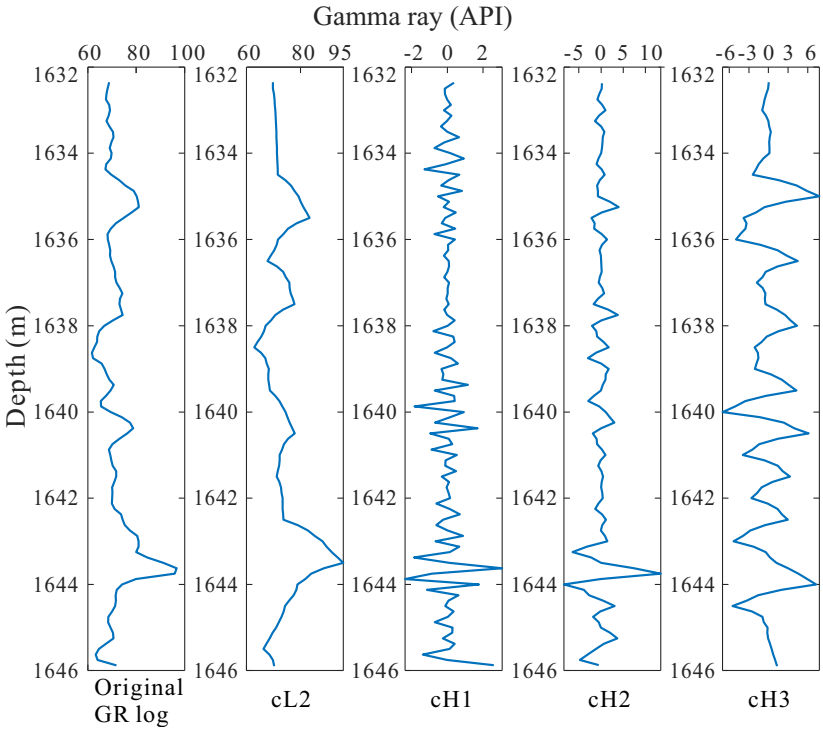

Fig. 4. GR log and its decomposed components at different dominant frequencies obtained using the DWT algorithm. This GR log is from the well CW7.

components from the n-th level decomposition, respectively. For example, the original signal is decomposed at the first level into a set of lower-frequency components (cL1) and higher-frequency components (cH1). Notably, cL1 is often considered high-frequency noise, for well logs (Zhang et al., 2018; Chen et al., 2021). At the second level, cL1 is further decomposed into cL2 and cH2. Similarly, at the third level, cL2 can be decomposed into cL3 and cH3. In brief, a signal $f(t)$ can be decomposed into cH1, cH2 and cL2 as the second level, and be decomposed into cH1, cH2 cH3 and cL3. Furthermore, the original signal can be reconstructed by summing these decomposed components, described as:

$$f(t) = \mathrm{cH1} + \mathrm{cH2} + \mathrm{cH3} \tag{3}$$

$$f(t) = \mathrm{cH1} + \mathrm{cH2} + \mathrm{cH3} + \mathrm{cL3} \tag{4}$$

where cH1, cH2 and cH3 are the higher-frequency components at first, second and third levels; cL2 and cL3 indicate the lower-frequency components at the second and the third levels, respectively.

The purpose of reconstructing well logs using the DWT algorithm and Eq. (3) is to enhance their sensitivity to thin geological units. Therefore, it is particularly important to emphasize the proportion of high-frequency components. The high-frequency components at the first level (cH1) are typically dominated by noise and should therefore be excluded, according to the principles of frequency decomposition for well logs (Zhang et al., 2018; Chen et al., 2021). The higher-frequency components at the second level (cH2) exhibit a higher frequency than those of cH3. The core-analysis results indicate that cH2 is more sensitive to the thin permeability barriers in the studied stratigraphy. Consequently, the following well-log reconstruction formula is proposed as follows:

$$f^*(t) = w_1\mathrm{CH1} + w_2\mathrm{CH2} + w_3\mathrm{cL2} \tag{5}$$

where $f^*(t)$ is the reconstructed well log, and w_1, w_2 and w_3 are set to 0, K and 1, respectively, where K commonly ranges between 2 and 4 for conventional well logs such as GR, AC and DEN.

As the value of K increases, the reconstructed well log becomes more sensitive to thin geological units; however, the number of artifacts also increases progressively. With K values of 2, 3 and 4, the identification accuracies of thin lithological beds were compared using an iterative approach. The results indicate that for the GR, RD, RS, AC, and DEN logs, the best performance is achieved when the K is set to 3. In addition, the SP curve was not subjected to DWT reconstruction. This is primarily because the SP curve lacks high-frequency information and serves mainly to distinguish thick sandstones from mudstones at the first level.

2.4 AutoML framework

The study area includes seven cored wells, named CW1 to CW7. Among them, CW3 is used as a blind-test well and excluded from the supervised learning process. The remaining six wells are used for cross-validation in turn.

The rapid advancement of machine learning has introduced novel approaches to intelligent well-log prediction. However, the proliferation of algorithms and the complexity of their parameter configurations make automated algorithm selection and parameter optimization increasingly essential (Erickson et

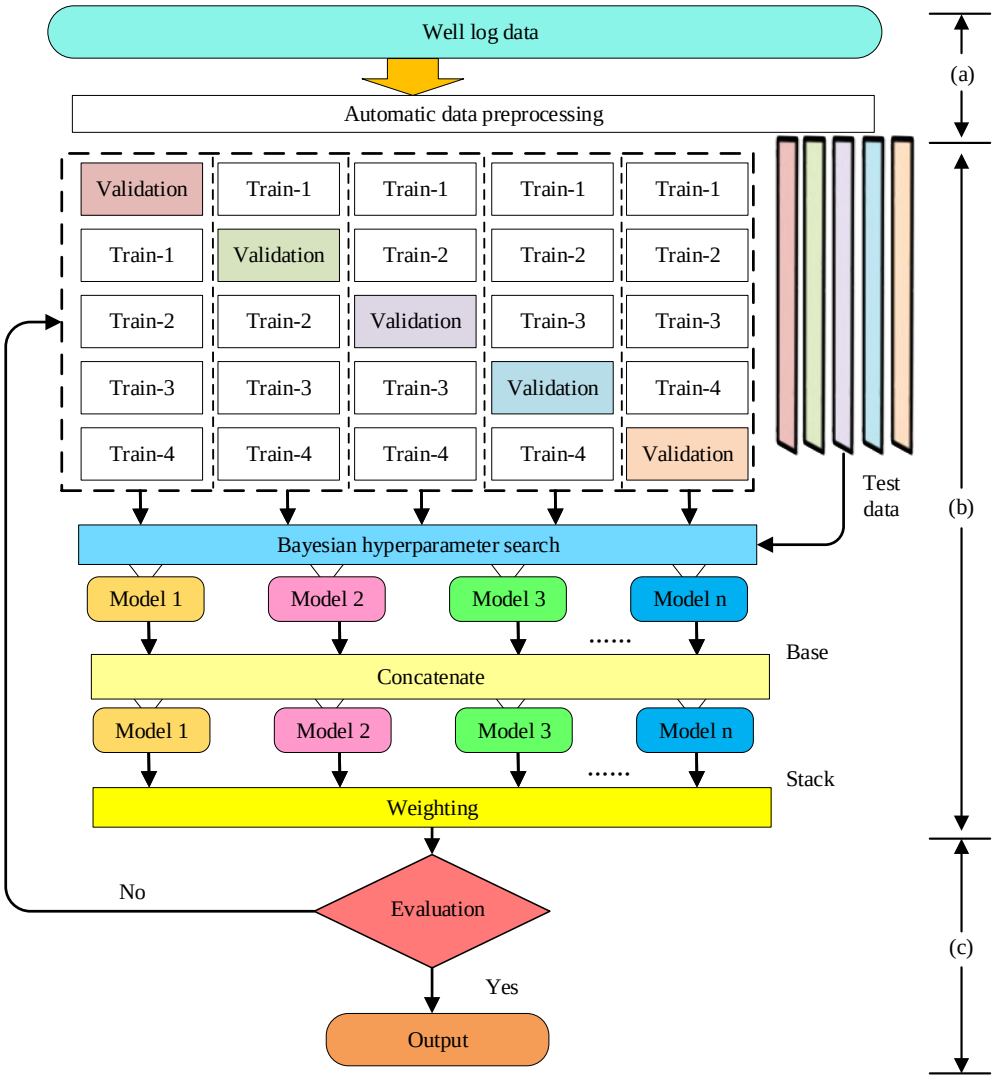

Fig. 5. Workflow diagram outlining the AutoGluon framework. Bold numbers on the right indicate the sections referred to in the text. (a) Data input and preprocessing, (b) model training and (c) model evaluation.

al., 2020; Zhang et al., 2024). In addition, low-quality reservoirs often exhibit complex spatial distributions, poor physical properties, pronounced internal interlayer heterogeneity, weak geophysical responses, and low inter-data correlations–factors that significantly constrain the predictive performance of any single algorithm.

By employing ensemble learning to integrate the advantages of multiple algorithms, prediction accuracy can be substantially improved. The open-source AutoGluon framework, developed by the Amazon Web Services team (Erickson et al., 2020), is adopted to design an integrated solution consisting of three key stages: Data preprocessing (Fig. 5(a)), model training (Fig. 5(b)), and model evaluation and output (Fig. 5(c)).

During model training, the AutoML framework automatically integrates and trains a diverse set of heterogeneous base learners–including XGBoost, Bg-KNN, Random Forest, LightGBM, CatBoost, Extra Trees, Logistic Regression, and MLP–while leveraging parallel computing to enhance training efficiency (Table 1). Bayesian optimization is applied to dynamically adjust model parameters, enabling intelligent hyperparameter search and tuning. For model integration, a 5-fold Bagging and two-layer Stacking strategy is used to combine and optimize the outputs of the base models, thereby improving generalization and significantly boosting classification accuracy.

Following training, prediction performance is evaluated on the test set. If the results meet the required thresholds, the model is saved; otherwise, retraining is performed. Overall, this framework markedly enhances both the stability and accuracy of prediction results, providing a robust technical foundation for intelligent well-log applications.

3. Example application to a subsurface dataset

3.1 Geological settings

The proposed method was applied to a braided-river delta reservoir within the Niujuanhu Oilfield of the Santanghu Basin. The Santanghu Basin, located in the northeastern

Table 1. Parameter settings for the used models in AutoGluon framework.

Model	Main model parameters
XGBoost	n_estimators = 50, max_depth = 6, learning_rate = 0.1, subsample = 0.8, colsample_bytree = 0.8
Bg-KNN	n_neighbors = 3, weights = 'uniform', algorithm = 'auto' n_neighbors = 5, weights = 'distance', algorithm = 'auto' n_neighbors = 7, weights = 'uniform', algorithm = 'ball_tree' n_neighbors=9, weights ='distance', algorithm = 'kd_tree'
Random forest	n_estimators = 50, max_depth = 10
LightGBM	n_estimators = 50, max_depth = 6, learning_rate = 0.1
CatBoost	iterations = 50, depth = 6, learning_rate = 0.1
Extra trees	n_estimators = 50, max_depth = 10
Logistic regression	C = 1.0
MLP	epochs = 50, batch_size = 32, learning_rate = 0.001
Bagging	num_bag_folds = 5, num_bag_sets = 3
Stacking	num_stack_levels = 2

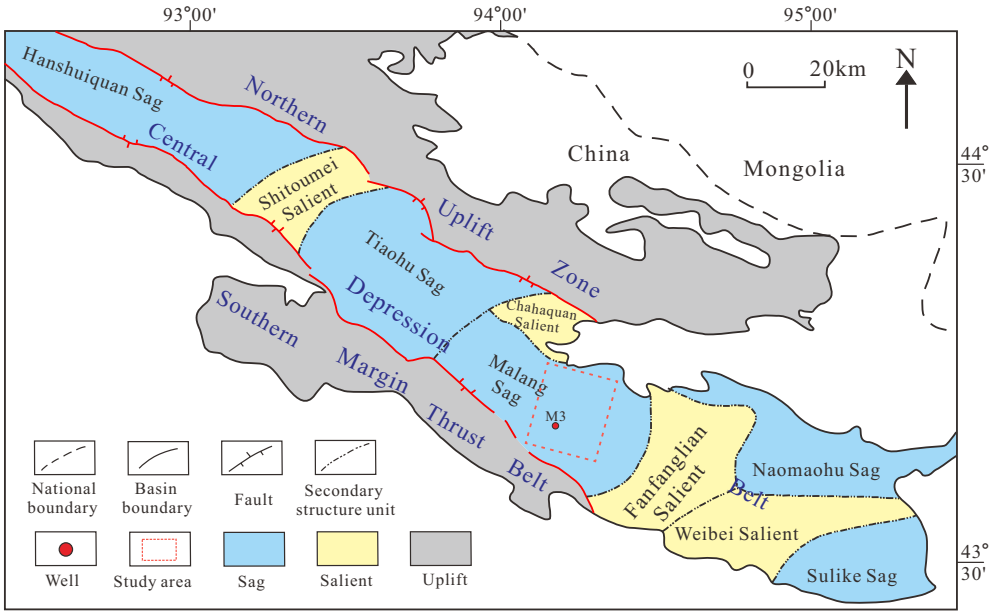

Fig. 6. Structural features of the Santanghu Basin and location of the study area.

Xinjiang Uygur Autonomous Region in northwestern China (Zhang et al., 2021). This intermontane basin exhibits a complex structural configuration and is divided into three principal sectors: The Northeast Fold-and-Thrust Belt, the Central Depression Belt, and the Southwest Fold-and-Thrust Belt (Fig. 6) (Wang et al., 2022).

The Niujuanhu Oilfield is located in the northwestern part of Central Depression Belt. The main reservoir target is the Jurassic Xishanyao Formation, ranging in depth from 1,400 to 2,000 m. This formation is interpreted as having originated from clastic deposition in a braided-river deltaic setting. These sand bodies exhibit low porosity and permeability values, averaging approximately 12.4% and 3.24 mD, respectively (Fig. 7). Core sample analysis reveals that the target formation consists predominantly of mudstone and sandstone, with occasional conglomeratic intervals (Fig. 7). The sand-to-mud ratio exceeds 70%. These extensive sand bodies contain many thin flow barriers, which pose significant challenges for CO_2 flooding projects–some reservoirs show no response, while others experience severe gas channeling.

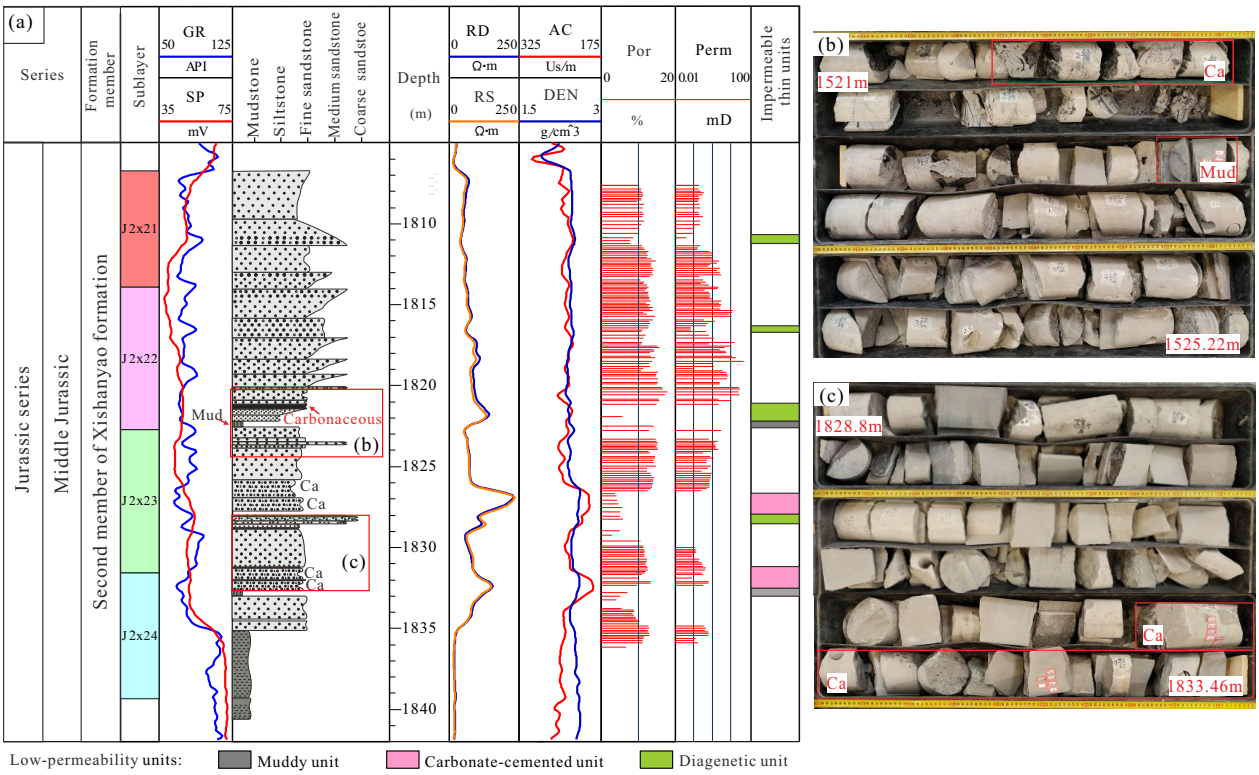

Fig. 7. (a) Lithological features of corehole CW1 in the studied Middle Jurassic formation, depicting the lithological section alongside log facies of GR, SP, RD, RS, DT, and DEN, as well as permeability (perm) and porosity (por) of cores. (b), (c) Photographs of typical core sections, highlighted in part (a). Annotation in part (b) and (c) Ca–Carbonate-cemented.

3.2 Well-log responses of permeability barriers

3.2.1 Features of permeability barriers

Five types of relatively thin low-permeability units are primarily identified in the reservoir volumes in the study area: Thin mudstone units, siltstone units, carbonaceous sandstone units, carbonate-cemented sandstone units, and gravelly or muddy-gravelly sandstone units (Figs. 8(a)-8(e)). Nonetheless, petroleum geologists working on this oilfield tend to group these features into three categories based on their genetic significance and well-log response: Mudstone units, carbonate-cemented units, and other diagenetically modified sandstone units.

The thickness of different flow barriers in the study area exhibits significant variability. Core analysis from seven cored wells reveals that: Mudstone units occurring within thick sandstones range from 0.2 to 0.8 m in thickness; carbonaceous sandstone and conglomerate units are between 0.3 and 1.5 m thick; siltstone units are between 0.5 to 2.5 m thick; calcareous sandstone units typically exceed 1 m in thickness.

3.2.2 Well-log responses

Based on core calibration results, sandstones typically exhibit low GR and high SP responses, while mudstones show low GR and high SP (Fig. 8). Due to the significantly lower resolution of SP compared to GR in the dataset, the identification of thick sandstone units relies primarily on the GR log, with 95 API serving as the tentative cutoff between sandstone and mudstone (shown as "shale line" in Fig. 8). Siltstone units exhibit responses that are intermediate between those of sandstone and mudstone units, with slightly higher GR readings in sandstone-dominated packages. Carbonaceous sandstones are characterized by low GR, higher DT, and lower DEN values. Calcareous sandstones display anomalously high resistivity, high DEN, and low DT. Low-permeability conglomerates are identified by low GR, relatively high DEN, lower DT, and fairly elevated RT values.

In practice, the log response amplitude of thin units is closely related to their thickness. For example, in Fig. 8(f), a 0.85 m mudstone bed occurring in a thick sandstone body shows a GR value exceeding 95 API, i.e., beyond the "shale line" threshold. On the contrary, the stratigraphic level corresponding to the position of a 0.33 m mudstone bed maps onto a comparatively reduced GR value, which is well below the "shale line" threshold. Thus, it is apparent that identifying low-permeability units that may be only a few decimeters thick presents a considerable challenge. As shown in Fig. 8, potential flow barriers or baffles with thicknesses lower than 1 m display variable log responses, such that conventional well-log interpretation methods are unsuitable for their identification.

3.2.3 Well-log resolution analyses

The resolution of well logs is influenced by many factors, including the design of the logging tool, the frequency of the electrical current used, the depth and physical characteristics

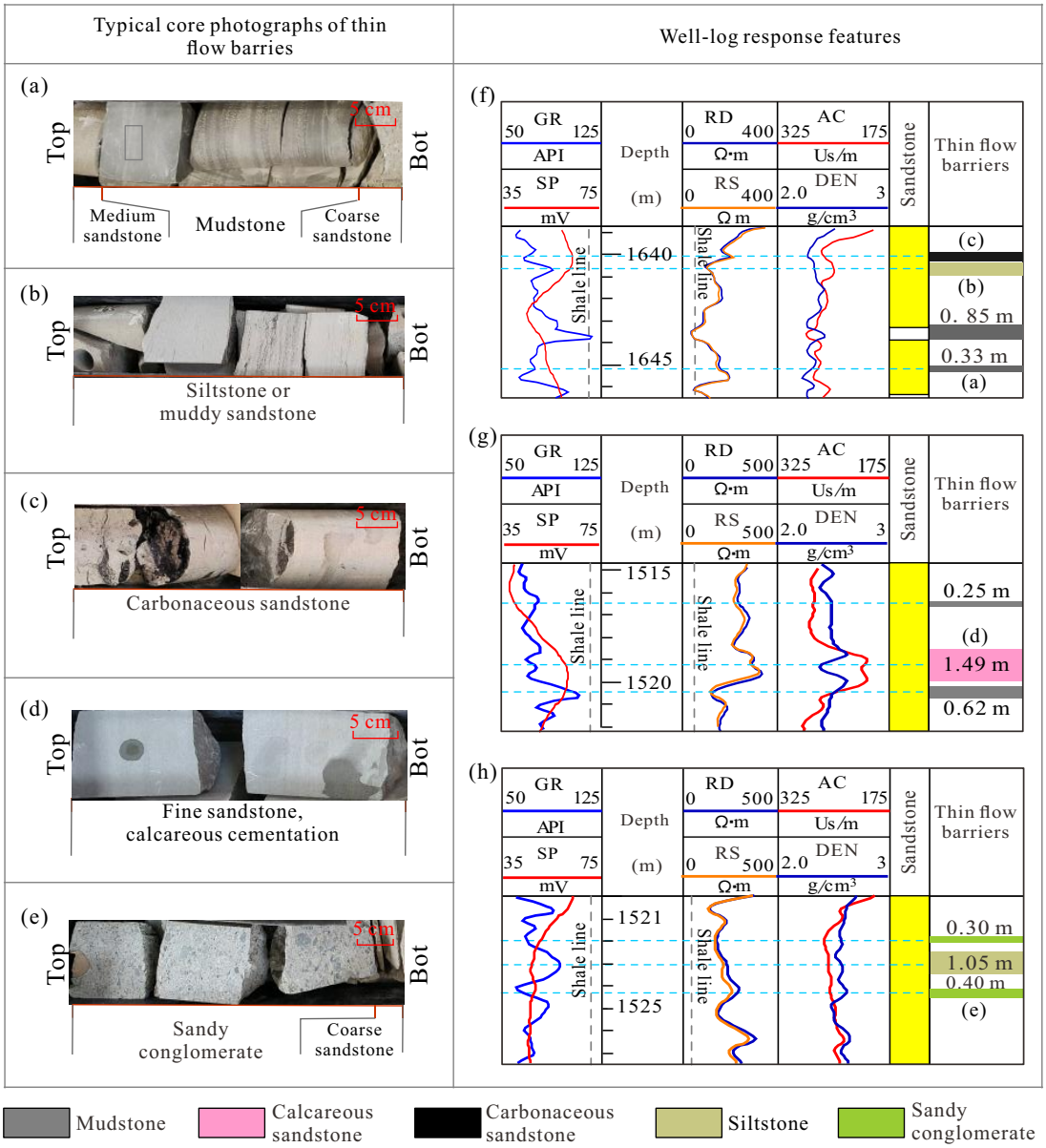

Fig. 8. Core photographs of relatively thin sedimentary heterogeneities and their well-log responses. (a)-(e) Core samples of mudstone beds, siltstone beds, carbonaceous sandstone beds, calcareous sandstone beds, and conglomerate beds, respectively and (f)-(h) well-log responses of typical low-permeability units.

of the rock formation, and the type of drilling fluid, among others. Nevertheless, the resolutions and investigation depths of the six considered well-log types have been analyzed. This analysis incorporates statistical data on resolution from published studies (Saneifar et al., 2015; Wood, 2022a) and evaluations conducted by geophysicists based on core samples from the Niujuanhu Oilfield (Table 2). The resolution of these six well-logs for lithologies with typical sensitivity ranges from 0.4 to 1.5 m. However, the lithologies that may act as low-permeability units do not belong to these typical cases. For instance, the resolution of the GR log in the study area can reach up to 0.4 m for pure mudstone beds under ideal conditions, whereas the actual resolution of the GR log is often larger than 1 m for siltstone units. In summary, the difficulty in identifying thin low-permeability units varies as a function of their type and nature, in addition to their thickness. Mudstone units exhibit more distinct, higher resolution well-log responses, whereas other thin units–such as siltstone, carbonaceous-coaly sandstone, and conglomerate–display less distinct log responses, expressed at lower resolution, making their identification particularly challenging.

3.3 Reconstruction of well logs using DWT

Using the method described in Section 2.2, five types of well logs–GR, DT, DEN, RD, and RS–were decomposed and reconstructed. By way of example, a GR log is shown in Fig. 9(a) that illustrates the reconstruction results for values of the parameter K set to 1, 2, 3 and 4, respectively. As the value of K

Table 2. Parameter settings for the used models in AutoGluon framework.

Log type	Investigation depth (m)	Theoretical resolution (m)	Evaluated resolution in study area (m)	Major sensitive lithologies
GR	0.4-0.8	0.3	0.4-1.0	Sandstone or mudstone
SP	< 1	0.7	1.5	Sandstone or mudstone
DEN	0.15-0.4	0.2	0.4-1.0	Calcareous sandstone
DT	0.2-0.3	0.3-1.0	0.7-1.5	Calcareous sandstone
RD	N/A	N/A	0.4-1.0	Calcareous and oil-bearing sandstone
RS	N/A	N/A	0.4-1.0	Calcareous and oil-bearing sandstone

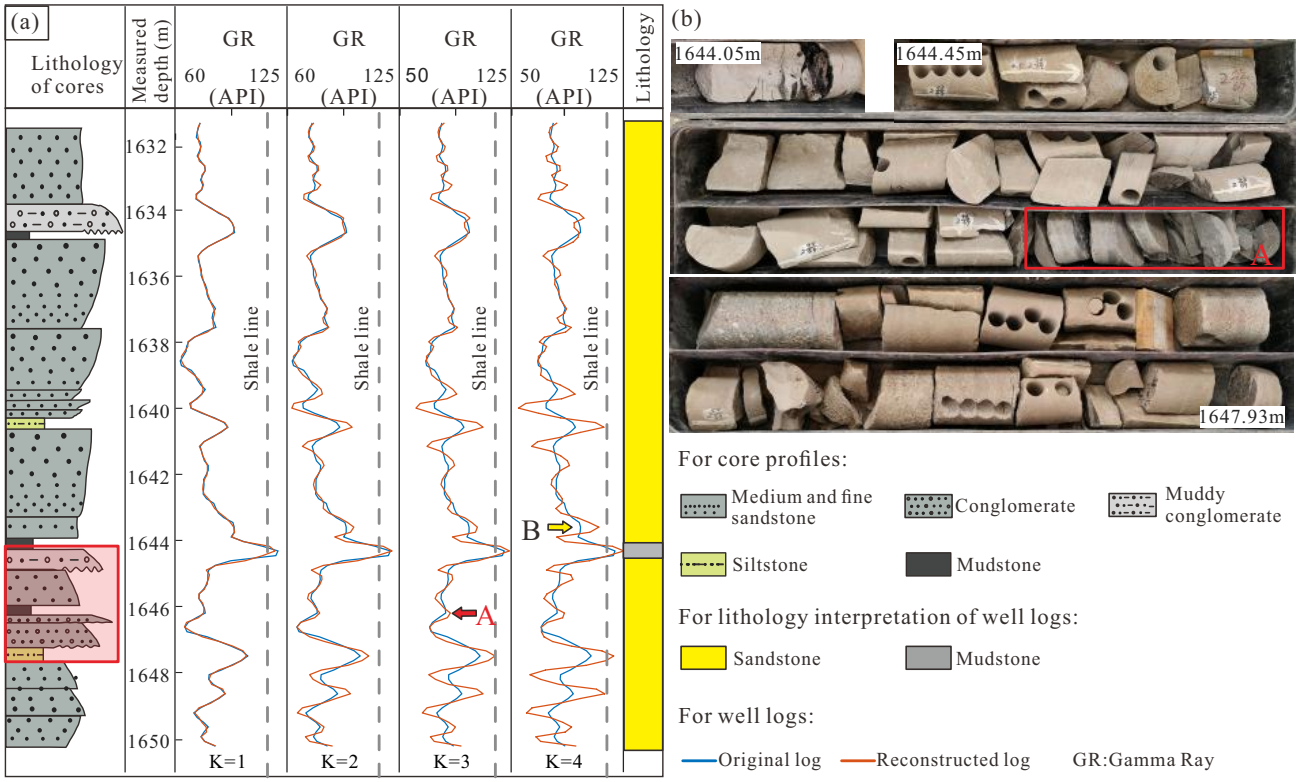

Fig. 9. (a) Reconstructed GR logs using the DWT with different K values from 1 to 4 and (b) core photographs, corresponding to the red box in part (a). The mudstone unit shown in the red box is indicated by the red arrow (A label) in part (a), for $K = 3$.

increases, the magnitude anomaly of a relatively thin mudstone bed becomes more pronounced (red arrow in Fig. 9(a)). When K is set to 3, a tradeoff is achieved whereby the magnitude anomaly of thin mudstone units in the reconstructed log is significantly enhanced and yet their thickness is not exaggerated. Specifically, the well-log response of a 38-cm-thick mudstone unit (Fig. 9(b)) was highlighted in the reconstructed log for $K = 3$ (Fig. 9(a)), but the thickness of this unit was markedly overestimated for $K = 4$ (yellow arrow in Fig. 9(a). On this basis, a K value of 3 was considered optimal and adopted. As detailed in Section 2.2, with K values of 2, 3 and 4, the identification accuracies of thin lithological beds are compared using an iterative approach, from which the K value with the highest accuracy is selected.

3.4 Recognition of thin heterogeneities

Using the method detailed in Section 3.2, thin heterogeneities are identified, including thin mudstone units, carbonate-cemented units, and other diagenetically modified sandstone units (Fig. 10). The results from the six cross-validation wells (Figs. 10(a) and 10(b)) and the blind-test well indicate that most low-permeability units are accurately interpreted, although discrepancies exist between their predicted thickness and their actual thickness measured in core. It should be emphasized that cored well CW3 was excluded from the machine learning process and thus serves as an independent test well (Fig. 10(c)). The interpretation results obtained from this well are considered broadly representative of the method's performance in non-cored wells. Finally, the interpretation

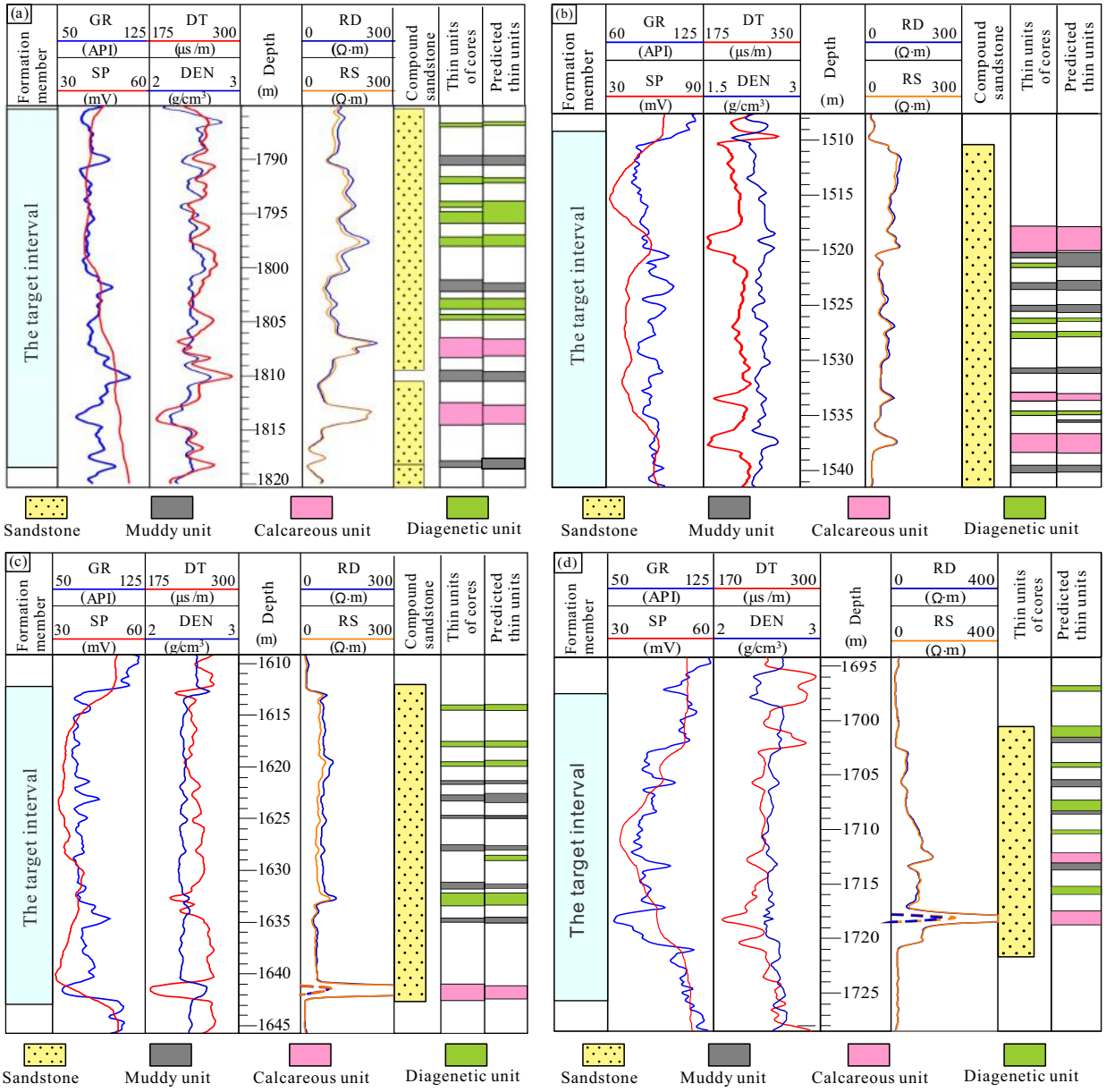

Fig. 10. Results of application of the machine learning model to predict the occurrence of thin sedimentary heterogeneities using the proposed method. (a)-(b) Two cross-validation cored wells, (c) blind-test cored well CW3, and (d) application to a non-cored well.

model was applied to 118 non-cored wells across the study area, with the results from one representative well presented in Fig. 10(d).

4. Discussion

4.1 Identification accuracy and reliability

The model performance of thin lithological heterogeneities is statistically summarized in Table 3 and Fig. 10, based on six cross-validation wells and a blind-test well. Two accuracy evaluation methods were applied: One based on general accuracy and a second one based on regular well-log sampling. The former method estimates the probability of correctly identifying thin impermeable lithological beds within thick sandstones, disregarding the difference between interpreted and actual thickness. As this study targets permeability barriers that tend to be thinner than 1 m, petroleum geologists are chiefly concerned with their presence rather than their exact thickness, making this approach both objective and reasonable. The latter method assesses the reliability of identification results by sampling the well logs at a rate of eight sampling points per meter.

For mudstone units with thicknesses between 0.2 and 0.5 m, the identification accuracy reaches 85% when evaluated using the general evaluation method; the accuracy decreases to 56% when assessed using the well-log sampling method. These results suggest that the proposed method is capable of qualitatively detecting thin mudstone units within the thickness

Table 3. Identification accuracy of the different types of heterogeneities based on six cross-validation wells and a blind-test well.

Heterogeneity unit	Thickness (m)	Accuracy of qualitative evaluation (%)	Accuracy of quantitative evaluation (%)	
			Sub-item	Total accuracy
Thin mudstone	0.2-0.5	85	56	86
	0.5-0.8	95	82	86
	> 0.8	100	91	86
Diagenetic sandstone	0.3-0.5	81	53	84
	0.5-0.8	87	80	84
	> 0.8	95	89	84
Carbonate-cemented	> 1.0	100	86	86

sandstones, but lacks the capability to quantitatively resolve their precise thickness. For reservoir units that exceed 0.5 m in thickness, accuracy values exceed 95% with the general evaluation method and are higher than 82% with the well-log sampling evaluation method, indicating that the proposed method can provide quantitative interpretations in such cases.

The identification performance for thin diagenetic sandstone units is similar to those for mudstone units, though with slightly reduced accuracy, with values of 82% and 56% for thicknesses of 0.3-0.5 m based on the general and well-log sampling evaluation methods respectively, and $>95\%$ and $>80\%$ for thicknesses >0.5 m. The carbonate-cemented units, a specific type of diagenetic sandstone, are always readily identified, presumably because their thickness is consistently larger than 1 m in the study area.

4.2 Strengths of the proposed approach

Based on the test and cross-validation wells, a high degree of match is observed between the sedimentary heterogeneities predicted via the proposed method and those identified visually in core, indicating the reliability and potential of the proposed method (Fig. 10 and Table 3). Due to their weak well-log responses, thin sedimentary heterogeneities are below the detectability of conventional interpretation methods based on well logs; the resolution of well logs prevents identification of mudstone units thinner than 0.5 m and of diagenetic sandstone units thinner than 1 m, in the study area. However, the proposed method has enabled the identification of sedimentary heterogeneities with thicknesses ranging from 0.3 to 0.5 m (Fig. 10), thereby significantly improving the breadth of geological information that can be extracted from well logs.

The advantages of this proposed method rest primarily on three key methodological aspects. First, the hierarchical approach in lithological interpretation. Conventional methods are initially used to distinguish thicker sand-prone and mud-prone units. Subsequently, lower-order, smaller-scale heterogeneities that are present within the thick sand-prone packages are identified and then classified. Second, the steps of well-log decomposition and reconstruction using DWT. The reconstruction process effectively enhances the well-log responses of thin sedimentary heterogeneities within sand-prone units (Fig. 9), improving well-log resolution (Zhang et al., 2018; Chen et al., 2021). Third, the application of intelligent algorithms. Compared to traditional lithological interpretation approaches applied to well logs, the application of intelligent algorithms allows users to extract information from multiple well logs more comprehensively (Iraji et al., 2023; Wei et al., 2025). The established nonlinear models can better represent the complex mapping relationships between lithology and well logs (Shan et al., 2021; Zhang et al., 2021; Iraji et al., 2023). Collectively, these advantages enable a fuller recognition of the lithological heterogeneity of reservoir units, thereby enhancing the interpretation accuracy of well logs.

4.3 Potential applications and limitations

The proposed method enables the accurate identification of thin sedimentary heterogeneity and supports the refined interpretation of lithology, lithofacies, and reservoir-unit types. Consequently, it can be also applied to the following research area:

1) Sedimentological characterization of well-log data, since facies sequences can be interpreted in greater detail and this can help support the process of conceptual reservoir modeling (definition of facies associations as elements/reservoir building blocks; development of depositional models);
2) Improved net-to-gross quantification, which together with improved porosity estimations supports refinement of volume computations;
3) Improved recognition of permeability heterogeneity of reservoir units, which can be incorporated in upscaling workflows.

As currently developed, the method is subject to some notable limitations. The technique currently requires an extensive collection of core samples for model calibration, due to the inherent data dependency of machine learning algorithms which require substantial training datasets to achieve optimal predictive accuracy. To address severe sample imbalance, this

study employs SMOTE for data balancing, which enhances the identification accuracy of thin lithological heterogeneities. It should be noted, however, that SMOTE interpolation may disrupt the Markovian property inherent to geological sequences, potentially introducing spurious artifacts when analyzing well-log curve morphology. Consequently, comparative validation must be performed between pre- and post-SMOTE implementations in practical applications.

Additionally, the predictive power of the machine-learning models is highly dependent on the variety and original resolution of the well logs in the target field to which they may be applied. The trained model for recognizing thin lithological heterogeneities is specific to the study area. When applied to other regions, it needs to be retrained using a sample set from the target area to ensure accuracy and reliability.

5. Conclusions

This study introduces a hierarchical well-log interpretation method that integrates DWT and AutoML framework to improve the recognition of thin lithological heterogeneities. Applied to a subsurface dataset, the method proves especially effective in identifying thin permeability barriers in thick compound sandbodies.

1) A proposed method consists of five key steps: Core interpretation, well-log processing, sample equalization, identification of thin heterogeneities, and heterogeneity classification. The AutoML framework integrates multiple algorithms with automated parameter optimization, demonstrating excellent robustness and generalization capability.
2) There are three strengths in the proposed method. The DWT enhances the well-log responses of the considered lithologies, facilitating the identification of thin units. The hierarchical interpretation of lithologies at different scales highlights lithological contrasts, facilitating a more accurate lithological differentiation. Machine learning leverages multiple well logs to capture complex lithology-response relationships.
3) The results demonstrate that this new method can lead to the accurate identification of thin lithological heterogeneities thicker than 0.2 m for thin mudstone units and 0.3 m for diagenetically modified sandstone units.
4) The proposed method shows potential for broader applications in reservoir studies, such as interpretations of thin sedimentary heterogeneities, lithologies, lithofacies, reservoir-unit types and net-to-gross quantification.

Acknowledgements

This work was supported by the National Major Science and Technology Project titled as “Key Technologies for CO_2-Enhanced Miscibility and Transition Control” (No. 2024ZD1406601), the National Natural Science Foundation Project of China (Nos. 42472179 and 42302128).

Conflict of interest

The authors declare no competing interest.

References

Chandrasekhar, E., Rao, V. E. Wavelet analysis of geophysical well-log data of Bombay Offshore Basin, India. Mathematical Geosciences, 2012, 44(8): 901-928.

Chawla, N. V., Bowyer, K. W., Hall, L. O., et al. SMOTE: Synthetic minority over-sampling technique nitesh. Journal of Artificial Intelligence Research, 2002, 16: 321-357.

Chen, S., Liu, P., Tang, D., et al. Identification of thin-layer coal texture using geophysical logging data: Investigation by wavelet transform and linear discrimination analysis. International Journal of Coal Geology, 2021, 239: 103727.

Erickson, N., Mueller, J., Shirkov, A., et al. AutoGluon-Tabular: Robust and accurate AutoML for structured data. ArXiv Preprint ArXiv: 2003.06505, 2020.

Feng, R., Grana, D., Balling, N. Variational inference in Bayesian neural network for well-log prediction. Geophysics, 2021, 86(3): M91-M99.

Iraji, S., Soltanmohammadi, R., Matheus, G. F., et al. Application of unsupervised learning and deep learning for rock type prediction and petrophysical characterization using multi-scale data. Geoenergy Science and Engineering, 2023, 230: 212241.

Kuang, L., Liu, H., Ren, Y., et al. Application and development trend of artificial intelligence in petroleum exploration and development. Petroleum Exploration and Development, 2021, 48(1): 1-14.

Lai, J., Wang, G., Wang, S., et al. Review of diagenetic facies in tight sandstones: Diagenesis, diagenetic minerals, and prediction via well logs. Earth-Science Reviews, 2018, 185: 234-258.

Li, W., Mamora, D., Li, Y., et al. Numerical investigation of potential injection strategies to reduce shale barrier impacts on SAGD process. Journal of Canadian Petroleum Technology, 2011: 50(3): 57-64.

Li, W., Yue, D., Du, Y., et al. Controls of accommodation to sediment-supply ratio on sedimentary architecture of continental fluvial successions. Petroleum Science, 2023, 20(4): 1961-1977.

Liu, R., Yue, D., Li, W., et al. Characterization of tight sandstone and sedimentary facies using well logs and seismic inversion in lacustrine gravity-flow deposits. Journal of Asian Earth Sciences, 2024, 259: 105897.

Lynds, R., Hajek, E. Conceptual model for predicting mudstone dimensions in sandy braided-river reservoirs. AAPG Bulletin, 2006, 90(8): 1273-1288.

Mallat, S. G., Zhang, Z. Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing, 1993, 41(12): 3397-3415.

Miall, A. D. Architectural-element analysis: A new method of facies analysis applied to fluvial deposits. Earth-Science Reviews, 1985, 22(4): 261-308.

Nyberg, B., Henstra, G., Gawthorpe, R. L. et al. Global scale

analysis on the extent of river channel belts. Nature Communications, 2023, 14(1): 2163.

Olkkonen, J. O. Discrete Wavelet Transforms-Theory and Applications. Rijeka, Croatia, InTech, 2011.

Papík, M., Papíková, L. The possibilities of using AutoML in bankruptcy prediction: Case of Slovakia. Technological Forecasting and Social Change, 2025, 215: 124098.

Pham, N., Wu, X., Naeini, E. Z. Missing well log prediction using convolutional long short-term memory network. Geophysics, 2020, 85(4): WA159-WA171.

Qi, W., Xu, C., Xu, X. AutoGluon: A revolutionary framework for landslide hazard analysis. Natural Hazards Research, 2021, 1(3): 103-108.

Rafieepour, S., Zheng, D., Miska, S., et al. Combined experimental and well log evaluation of anisotropic mechanical properties of shales: An application to wellbore stability in bakken formation. Paper SPE 201334 Presented at the SPE Annual Technical Conference and Exhibition, Virtual, 26-29 October, 2020.

Saneifar, M., Heidari, Z., Hill, A. D. Application of conventional well logs to characterize spatial heterogeneity in carbonate formations required for prediction of acid-fracture conductivity. SPE Production and Operations, 2015, 30(3): 243-256.

Shan, L., Liu, Y., Tang, M., et al. CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction. Journal of Petroleum Science and Engineering, 2021, 205: 108838.

Stanistreet, I. G., Stollhofen, H. Hoanib River flood deposits of Namib Desert interdunes as analoques for thin permeability barrier mudstone layers in aeolianite reservoirs. Sedimentology, 2002, 49(4): 719-736.

Strick, R. J. P., Ashworth, P. J., Sambrook Smith, G. H., et al. Quantification of bedform dynamics and bedload sediment flux in sandy braided rivers from airborne and satellite imagery. Earth Surface Processes and Landforms, 2019, 44(4): 953-972.

Sun, B., Cui, W., Liu, G., et al. A hybrid strategy of AutoML and SHAP for automated and explainable concrete strength prediction. Case Studies in Construction Materials, 2023, 19: e02405.

Wang, R., Lv, C., Lun, Z., et al. Study on gas channeling characteristics and suppression methods in CO_2 flooding for low permeability reservoirs. Paper SPE 142306 Presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Jakarta, Indonesia, 20-22 September, 2011.

Wang, W., Lyu, Q., Fan, T., et al. Characteristics of the tuffaceous shale oil reservoir and its sweet spots: A case study of the Tiaohu depression in the Santanghu Basin. Unconventional Resources, 2022, 2: 192-199.

Wei, T., Xu, J., Song, L., et al. Reservoir porosity interpretation method and application based on intelligent algorithms. Geoenergy Science and Engineering, 2025, 247: 213650.

Wood, D. A. Gamma-ray log derivative and volatility attributes assist facies characterization in clastic sedimentary sequences for formulaic and machine learning analysis. Advances in Geo-Energy Research, 2022a, 6(1): 69-85.

Wood, D. A. Extracting useful information from sparsely logged wellbores for improved rock typing of heterogeneous reservoir characterization using well-log attributes, feature influence and optimization. Petroleum Science, 2022b, 22: 2307-2311.

Wood, D. A. Well-log attributes assist in the determination of reservoir formation tops in wells with sparse well-log data. Advances in Geo-Energy Research, 2023, 8(1): 45-60.

Yang, L., Wang, S., Chen, X., et al. Deep-learning missing well-log prediction via long short-term memory network with attention-period mechanism. Geophysics, 2023, 88(1): D31-D48.

Yang, X., Wang, R., Li, K., et al. Meta-Black-Box optimization for evolutionary algorithms: Review and perspective. Swarm and Evolutionary Computation, 2025, 93: 101838.

Yue, D., Wu, S., Xu, Z., et al. Reservoir quality, natural fractures, and gas productivity of upper Triassic Xujiahe tight gas sandstones in western Sichuan Basin, China. Marine and Petroleum Geology, 2018, 89: 370-386.

Zecchin, M, Catuneanu, O. High-resolution sequence stratigraphy of clastic shelves VI: Mixed siliciclastic-carbonate systems. Marine and Petroleum Geology, 2017, 88: 712-723.

Zhang, X., Li, Q., Li, L., et al. Combination of sonic wave velocity, density and electrical resistivity for joint estimation of gas-hydrate reservoir parameters and their uncertainties. Advances in Geo-Energy Research, 2023, 10(2): 133-140.

Zhang, S., Liu, C., Liang, H., et al. Mineralogical composition and organic matter characteristics of lacustrine fine-grained volcanic-hydrothermal sedimentary rocks: A data-driven analytics for the second member of Permian Lucaogou Formation, Santanghu Basin, NW China. Marine and Petroleum Geology, 2021, 126: 104920.

Zhang, X., Liu, M., Liu, Y. Battery state of health estimation using an AutoGluon-tabular model incorporating uncertainty quantification. Journal of Energy Storage, 2024, 101: 113920.

Zhang, Q., Zhang, F., Liu, J., et al. A method for identifying the thin layer using the wavelet transform of density logging data. Journal of Petroleum Science and Engineering, 2018, 160: 433-441.