

## Research Article

# Genetic Mechanisms and Identification of Low-Resistivity Pay Zones: A Case Study of Pengyang Area, Ordos Basin, China

Peiqiang Zhao <sup>1,2</sup>, Yuting Hou,<sup>3</sup> Fengqing Ma,<sup>1,2</sup> Jixin Huang,<sup>4</sup> Xiaoyu Wang,<sup>1,2</sup> Jiarui Xie,<sup>1,2</sup> Chengxiang Deng,<sup>5</sup> and Zhiqiang Mao<sup>2</sup>

<sup>1</sup>State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China

<sup>2</sup>Beijing Key Laboratory of Earth Prospecting and Information Technology, China University of Petroleum (Beijing), Beijing 102249, China

<sup>3</sup>PetroChina Changqing Oilfield Company, Xi'an 710018, China

<sup>4</sup>Research Institute of Petroleum Exploration and Development, China National Petroleum Corporation, Beijing 100083, China

<sup>5</sup>School of Geophysics and Measurement-Control Technology, East China University of Technology, Nanchang 330013, China

Correspondence should be addressed to Peiqiang Zhao; zhaopq@cug.edu.cn

Received 11 June 2022; Accepted 20 September 2022; Published 15 October 2022

Academic Editor: Jin Qian

Copyright © 2022 Peiqiang Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Triassic Yanchang Formation in the Pengyang area of Ordos Tianhuan depression is an important oil and gas formation. However, most of the oil pays in the study formations are low-resistivity or low-contrast reservoirs with low permeability, bringing challenges to the reservoir identification and evaluation by well logs. In this paper, we first measured the nuclear magnetic resonance (NMR), phase permeability, cation exchange capacity (CEC), and X-ray diffraction for core samples. Then, the genetic types for the low-resistivity pays were analyzed based on the experiment results, water analysis, and well log data collected. It was found that large variations of formation water salinity, high irreducible water saturation, and clay conductivity are the primary genetic types. Further, the random forest (RF) algorithm with sensitive parameter inputs was used to identify the oil, oil and water, and water layers. The anomaly of spontaneous potential ( $\Delta SP$ ) that characterizes water salinity, the relative value of gamma ray log ( $\Delta GR$ ) that describes the bound water content, resistivity, density, and acoustic logs were taken as sensitive logs according to the genetic analysis. Finally, this identification method was verified by comparison with the traditional crossplot method and oil test results. The identification accuracy of the RF is 90%, far higher than that by the crossplot method.

## 1. Introduction

The oil and gas exploration in China has entered the stage of complex and unconventional oil and gas reservoirs. The geological conditions of exploration objects are complex, and exploration is becoming more and more difficult. The Mesozoic in the southern part of Ordos Tianhuan depression is an important oil and gas reservoir belt. However, the pore structure is complex, and water properties of oil and gas reservoirs are changeable [1–3], causing a large number of low-contrast or low-resistivity reservoirs to develop. The interpretation of oil and water layers with different log response characteristics is unclear, resulting in challenges to the identification and evaluation of oil reservoirs based on well logs.

In the middle of the last century, Tixier et al. [4] of Schlumberger first put forward the concept of low-resistivity reservoir, then started to investigate the definition of low-resistivity reservoir. From the perspective of oil saturation and resistivity index, Zemanek [5] thought that the resistivity index of low-resistivity reservoir should be less than three, and the oil saturation should be less than 50%. However, many scholars pay more attention to the resistivity ratio of oil and water layers to measure whether the oil layer is a low-resistivity reservoir. Ouyang et al. [6] believed that the oil layer can be determined as a low-resistivity reservoir if the resistivity of oil reservoirs is less than twice that of the water layers. The main genetic types for low resistivity of oil reservoirs are conductivity of clay, high-salinity formation

water, high irreducible water saturation, deep invasion of high-salinity filtrate, and a small amount of conductive minerals [7–12]. For the Chang 8 member, Ordos Basin, Bai et al. [13] believed that the water salinity and irreducible water saturation are the main reasons causing the resistivity of oil reservoirs to be low.

Many scholars have studied the fluid identification of low-resistivity reservoirs. The more classical methods include overlapping, crossplots, nuclear magnetic resonance logs, and mathematical statistics methods [14–17]. Among these methods, the most commonly used methods are overlapping and crossplot methods [18, 19]. Although nuclear magnetic resonance (NMR) logging plays a vital role in fluid identification [20], there is almost no NMR logging data in some old oilfields. Ren et al. [21] defined a fluid indicator factor based on the double-porosity overlap method. By constructing the crossplot of fluid identification factor with porosity, they predicted the fluid properties of low-porosity and low-permeability reservoirs in the Subei Basin with high accuracy. Sun et al. [22] established a double-porosity model of tight sandstone based on the Voigt–Reuss–Hill model and further used the elastic parameters predicted by the established model to construct a fluid identification chart suitable for low-porosity and low-permeability reservoirs. Based on the correlation coefficient for the upper Paleozoic tight sandstone reservoir in the Linxing-Shenfu area of Ordos Basin, Hou et al. [23] proposed a fluid discrimination method by using density porosity and resistivity logging data.

With the advent of the era of big data, scholars have tried to process and interpret logging data through various machine learning methods to improve production efficiency and accuracy [24–26]. The process of machine learning is to find the objective function through training data. The commonly used algorithms include decision tree, random forest algorithm, logical regression, support vector machine, neural network, and cluster analyses. These methods mine the complex non-linear relationship between data through complex transformation, which is more effective for the problems that cannot be effectively solved by traditional physical or empirical models. Combining logs with production data, Zhang et al. [27] accurately identified oil, gas, and water layers in the coexistence area of low-resistivity oil layer and high-resistivity water layer by using the support vector machine method. Based on electrical and physical properties, acoustic parameters, or gas logging parameters, Liu et al. [28] extracted oil and gas sensitive parameters and applied four mathematical algorithms of the decision tree, radial basis function, neural network, and cluster analysis to the comprehensive evaluation of fluid properties. The identification method of carbonate reservoir in the Jidong exploration area is more accurate than that of single information. Chen et al. [29] applied the machine learning AdaBoost M2 algorithm to the fluid recognition of sandy conglomerates. The K-type multifluid type is disassembled into a binary classification problem, and the decision tree is called as a weak learning algorithm to automatically obtain the classifiers for fluid discrimination. Tan et al. [30] used a committee machine method to identify the fluid types for tight sandstones from GR, resistivity, and porosity logs. Luo et al. [31] used long-term and short-term memory network (LSTM) and convolutional neural network (CNN) to characterize the timing char-

acteristics of log curves and the correlation between multiple log curves, respectively. The recognition accuracy of the oil-bearing reservoir is significantly improved by using the weighted crossentropy loss function. The multilayer fluid recognition method improves the recognition accuracy of oil, oil and water, and water layers.

Random forest (RF) is an algorithm for regression prediction and classification prediction based on multiple decision trees [32]. Numerous decision trees are established by random repeated sampling technology and node random splitting technology. The prediction results of many decision trees are combined and output as a whole. The random forest algorithm has the advantages of highly parallel training, fast training speed for large samples of big data, a small variance of the trained model, strong generalization ability, and insensitivity to the lack of some features [33, 34].

In this paper, we collected a large amount of core porosity and permeability, water analysis, and well log data. Also, we measured the NMR, phase permeability cation exchange capacity (CEC), and X-ray diffraction for core samples. Based on the well log and core analysis data, the possible reasons for the low-resistivity oil pay are analyzed and discussed. Further, the RF algorithm with sensitive parameter inputs instructed by the genesis analysis was used to identify the oil, oil and water, and water. Finally, this identification method was verified by comparison with crossplots and oil test results.

## 2. Geological Characteristics of the Chang 8 Member in the Pengyang Area

Ordos Basin is the second largest sedimentary basin in China. The basin is characterized by its large area of about 330,000 km<sup>2</sup> and broad resource distributions, with significant potential and economic reserves [35]. It is a multicycle superimposed petroliferous basin. The basin could be divided into 6 secondary structural units, including Yimeng Uplift, Weibei Uplift, Yishan Slope, Western Fold-Thrust Belt, Tianhuan depression, and Jinxi Fault Fold Belt [36]. The target area, the Pengyang area, is located in the southern Tianhuan depression, (Figure 1). The Triassic Yanchang Formation is one of the main exploration horizons.

The data were taken from the Chang 8 member of the Triassic Yanchang Formation in the Pengyang area, Ordos Basin, China. The target layers are characterized by low-porosity and low-permeability reservoir. The core porosity and permeability of Chang 8 member are shown in Figure 2. The porosity is mainly distributed between 6% and 21%, with an average of 15.59%, and the permeability is mainly distributed from 0.1 to 30 mD, with an average of 1.45 mD.

## 3. Experiments

To investigate the genetic mechanisms of the low-resistivity oil pay, we measured the NMR, phase permeability, cation exchange capacity (CEC), and X-ray diffraction (XRD) minerals of rock samples drilled from the target reservoirs. The NMR T<sub>2</sub> spectra of twenty samples were measured, while CEC and XRD were measured in ten of them, and phase permeability was measured in two.

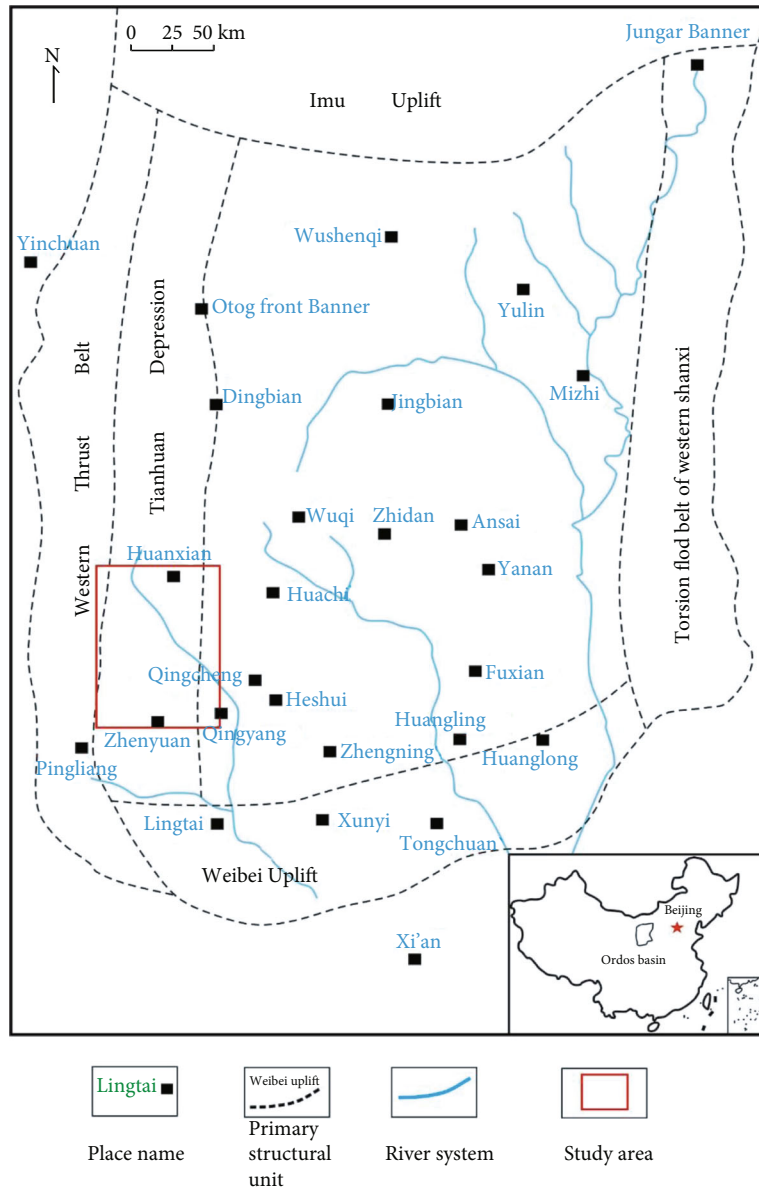


FIGURE 1: Geological map of the Pengyang area, Ordos Basin, China.

The nuclear magnetic resonance measurement adopts the nuclear magnetic resonance analysis and imaging system MesoMR23-060H-I produced by Niumet. The magnetic field strength is 21 MHz, the waiting time  $T_w$  is set to 3 s, and a value of 0.1 ms was used for the interecho spacing,  $T_E$ . And the echo number (NECH) is set to 10000, and the number of scans (NS) is 32.

For the measurement of phase permeability curve, the core samples were dried at 110°C for 24 h, cooled to room temperature in a drying dish, and weighed and measured with vernier calipers. The dry core samples were evacuated for four hours and saturated with water at pressure for 24 h, after which the imbibition of oil at atmospheric conditions was performed until there is no water flowing out. At this time, the water in the core was called the irreducible water. Then, the water flooding experiment was carried out under constant speed conditions, recording the time, pressure difference value, and oil/water volume in

the experiment. Finally, the data were substituted into the calculation formula to calculate the oil/water relative permeability value and oil and water saturation and plot the relative permeability curve.

The CEC and quantities of cation exchange ( $Q_v$ ) of reservoir rock are two critical physical parameters to characterize the conductivity of rock clay. CEC of clay minerals refers to the total number of cations that could be exchanged by clay minerals under pH value 7, that is, the number of cations that can be absorbed and exchanged by clay minerals. CEC is a measure of the number of negative charges of clay minerals. The unit of cation exchange capacity is mmol/100 g, that is, the number of millimoles of cation exchanged per 100 g dry sample.  $Q_v$  is the amount of cation exchangeable per unit pore volume of clay, in milliequivalents per liter. In the CEC measurement experiment, according to the national standard SY/T 6352-2013, the reagents used include 1 mol/L ammonium

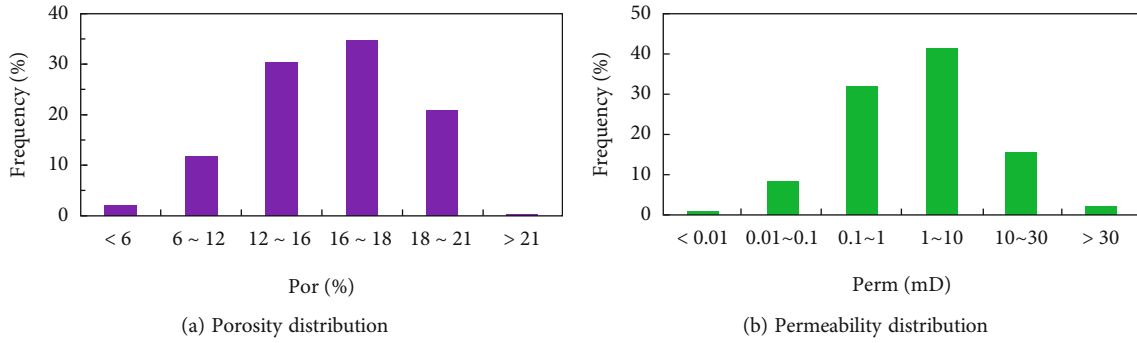


FIGURE 2: Porosity and permeability distribution histograms in Chang 8 reservoirs.

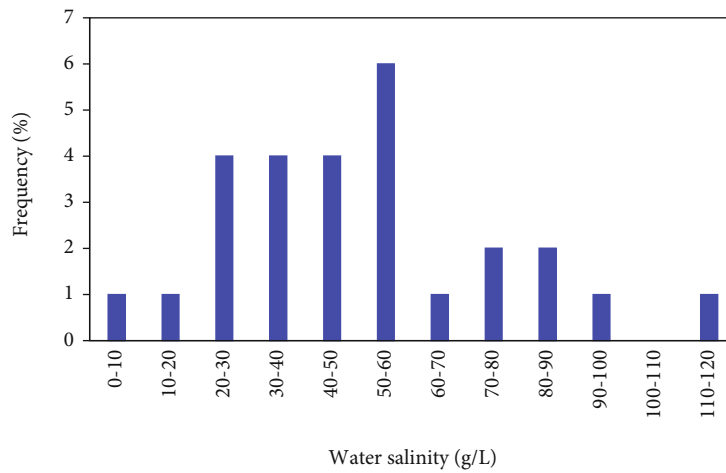


FIGURE 3: Histogram of the formation water salinity distribution.

acetate solution (pH 7.0), 77.09 g ammonium acetate solution (CH<sub>3</sub>COONH<sub>3</sub>, chemically pure), ethanol solution (for industrial use, must be free of NH<sub>4</sub><sup>+</sup>), and 0.05 mol/L hydrochloric acid standard solution.

For the XRD measurement, the core sample needs to go through the steps of separating clay particles, sample preparation, glycerol treatment, and 550°C heat treatment. The instrument used for XRD analysis is Porter Philips 00186 diffractometer. The anode of the ray tube is copper (Cu K $\alpha$  radiation) with a wavelength of 1.5406 Å. The X-ray tube was operated at 50 kV and 30 mA.

#### 4. Genetic Mechanisms of Low-Resistivity Oil Pay

**4.1. The Effect of the Formation Water Salinity.** Formation water salinity is one of the main factors affecting rock resistivity. Generally speaking, the water in the same formation is relatively stable, and the change of salinity is small. However, the water analysis data show that the formation water salinity of the target formation in the study area varied heavily. Figure 3 shows the distribution histogram of the formation water salinity collected from 27 wells. It shows that the formation water salinity of the Yanchang Formation is mainly distributed between 20 g/L and

90 g/L, but there are a few wells with salinity greater than 100 g/L. The formation water resistivity varies from about 0.1  $\Omega$ -m to 1.8  $\Omega$ -m at 25°C. The change of the formation water salinity results in low contrast between oil and water layers. And the difference in the formation water salinity weakens, conceals, or even cancels the contributions of oil content to electrical properties, resulting in a blurred boundary between oil and water layers and low coincidence rate of log interpretation.

Figure 4 shows logs of the Well M1. In this figure, the gamma-ray (GR) and caliper and spontaneous potential (SP) are displayed in track 1 (from left). The second track displays the array induction resistivity logs, including AT90, AT60, AT30, AT20, and AT10. The detection depth of AT90 is the deepest, while that of AT10 is the shallowest. Porosity logs, in terms of density (DEN), compensated neutron (CNL), and acoustic (AC) logs, are presented in track 3. Track 4 is measured depth. Tracks 5 to 7 present the porosity, permeability, and shale volume contents, respectively. The blue line in depth track indicates the oil test layer. Figure 5 shows logs of another well, Well M2. The oil test layer for Well M1 is the interval of 2439-2442 m with the resistivity of 8~10  $\Omega$ -m, and for Well M2 is the interval of 2265-2268 m with the resistivity of 5~8  $\Omega$ -m. However, the fluid type of M1 is water, and that of M2 is oil, according to oil test results. The formation water

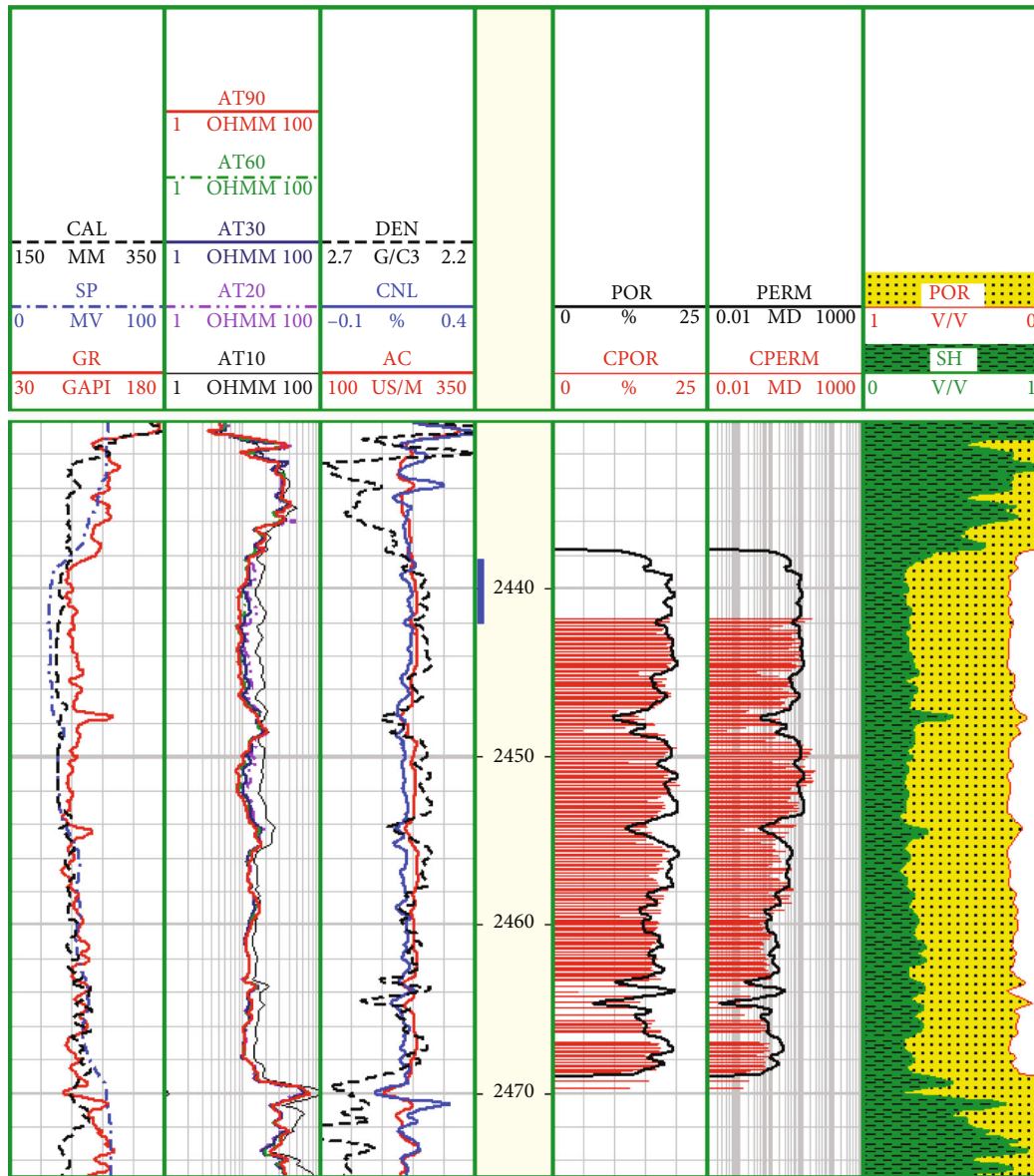


FIGURE 4: Water layers in Well M1. The resistivity is about 8~10 Ω·m.

salinity of the two wells is, respectively, about 20.5 g/L and 73.8 g/L, making it difficult to identify fluid types of target reservoirs.

The influencing factors of the formation water properties between different wells include the following: (1) the sedimentary reason; that is, the lithology of fluvial sedimentary facies varies greatly. The reservoir with heavy mud and fine lithology retains high-salinity water in the diagenetic process. (2) In fine lithologic reservoirs, free water in large pore throats is driven away during hydrocarbon migration and accumulation, while immobile water with high salinity is retained in small pore throats. (3) Frequent tectonic movement destroys complete and closed traps, and the water from the reservoir bottom or rock and mineral filtration migrate to the reservoir again. Surface water can penetrate the underground primary reservoir through open faults, changing the properties of reservoir fluid. High formation water salinity is controlled by oil accumula-

tion. With the increase of buried depth, hydrocarbon source rock is compacted. Further, the high salinity formation water and oil in the source rock pores are squeezed out, under the action of excess pressure and downward migration into the pore of the formation, displacing the original formation in pore water or mixing with the original formation water. And low salinity of formation water is controlled by tectonic setting, fracture, and fault development, which makes the upper low salinity of formation water and lower together. Low salinity of formation water along cracks, or faults to the reservoir pores, displaces the original formation in pore water or mixes the original formation water. Thus, the formation water salinity becomes lower.

4.2. *The Impact of Irreducible Water Saturation.* The high bound water saturation may also be one of the main leaders of forming low-resistivity reservoirs, so it needs to be analyzed

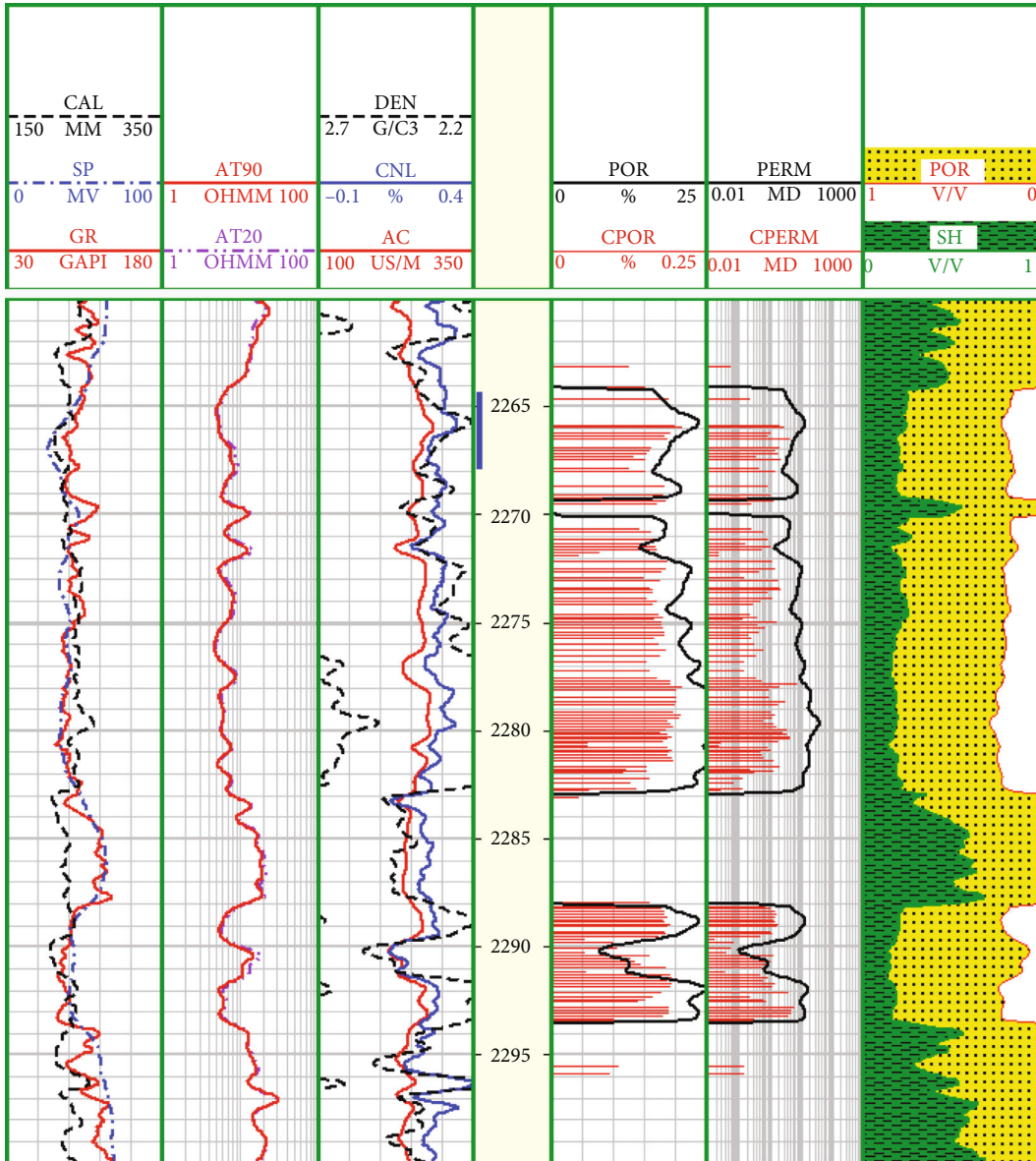


FIGURE 5: Oil layers in Well M2. The resistivity is about 5~8  $\Omega$ -m.

and studied. Reservoir bound water usually consists of three parts: (1) film water retention on the surface of rock particles (nonclay) due to wettability, (2) capillary water retention in pores, and (3) clay particle adsorption of water.

Twenty samples from Chang 8 reservoir in the Pengyang area in 7 wells were tested for the NMR saturation experiment. The saturated NMR  $T_2$  spectra of each sample are shown in Figure 6. It shows that the  $T_2$  spectra of Chang 8 reservoir in the study area are dominated by small pores and high content of bound water. Table 1 presents the bound water saturation from NMR  $T_2$  spectra. From Table 1, it can be seen that the bound water saturation of the reservoir is mainly distributed between 40% and 80%, with an average of 73.74%.  $T_2$  logarithmic mean ( $T_{2,lm}$ ) values vary from 0.48 to 7.79 ms. The average of the  $T_{2,lm}$  is 2.03 ms. Figure 7 shows the relative permeability curves of samples T2 and T3. From this figure,

it is seen that the irreducible water saturation is about 40%. The high irreducible water saturation could reduce the resistivity of oil layers.

Figure 8 shows an example of producing different fluid types caused by different irreducible water saturation. In this figure, different from Figures 4 and 5, MSIGTA, MPHITA, and MFFI are, respectively, total porosity, free fluid porosity, and bound water porosity, obtained from NMR  $T_2$  spectra. The last track presents  $T_2$  spectra and  $T_{2,lm}$  logs. The resistivity of the upper layer (2370~2374.5 m) is about 30  $\Omega$ -m, and the oil testing shows that it is an oil and water layer. The lower layer (2384.5~2386.5 m) has resistivity values ranging from 10 to 20  $\Omega$ -m, and the oil testing result shows that it is a good oil layer. The irreducible water saturation of the lower layer is close to total water saturation; thus, they are pure oil layers. In contrast, the water saturation of the

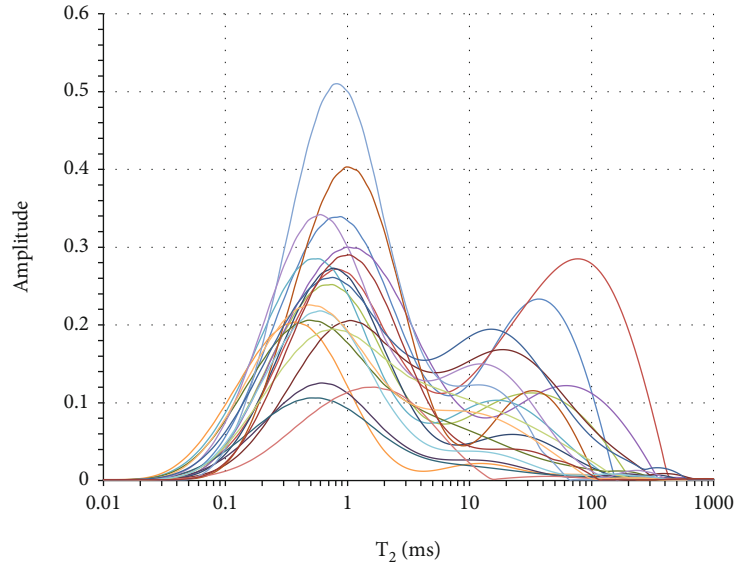


FIGURE 6: NMR  $T_2$  spectra of Chang 8 reservoir in the Pengyang area.

TABLE 1: Porosity, permeability, bound water saturation, and  $T_{2lm}$  from NMR measurements.

Sample Number	Por (%)	Permeability (mD)	Bound water saturation from NMR (%)	$T_2$ logarithmic mean (ms)
T1	20.8	11.63	54.52	2.98
T2	22.7	65.94	51.52	7.79
T3	17.8	2.29	72.30	2.29
T4	17.7	3.47	65.77	2.9
T5	17.2	1.70	84.14	1.29
T6	7.0	0.40	82.81	0.48
T7	11.5	1.28	81.59	1.39
T8	15.9	4.29	55.28	4.36
T9	10.0	0.37	83.12	1.07
T10	5.0	0.12	81.83	1
T11	7.5	0.13	83.70	0.86
T12	18.0	2.25	84.64	1.76
T13	19.2	5.38	75.21	1.3
T14	6.0	0.045	92.74	1.41
T15	13.1	0.84	66.99	1.96
T16	16.7	1.03	76.30	1.53
T17	9.4	0.13	69.71	0.92
T18	15.4	0.74	67.63	1.17
T19	15.7	1.13	56.82	2.84
T20	13.9	0.42	88.15	1.35

upper layer is greater than the irreducible water saturation, which may cause the oil and water producible at the same time.

**4.3. Conductivity of Clay.** Clay minerals in sandstone reservoirs generally have additional electrical conductivity. The electrical conductivity of clay-bearing sandstone is very different from that of pure sandstone, which is one of the important

reasons forming low-contrast or low-resistivity oil layers. Clay-bearing sandstone formations all contain a certain amount of clay, and the surface of clay particles is usually negatively charged. Under normal circumstances, the negatively charged cations adsorbed on the surface of clay particles cannot move, but the adsorption is not very tight. Under the action of an electric field, the adsorbed cations can exchange positions with other hydrated ions in the solution in the rock,

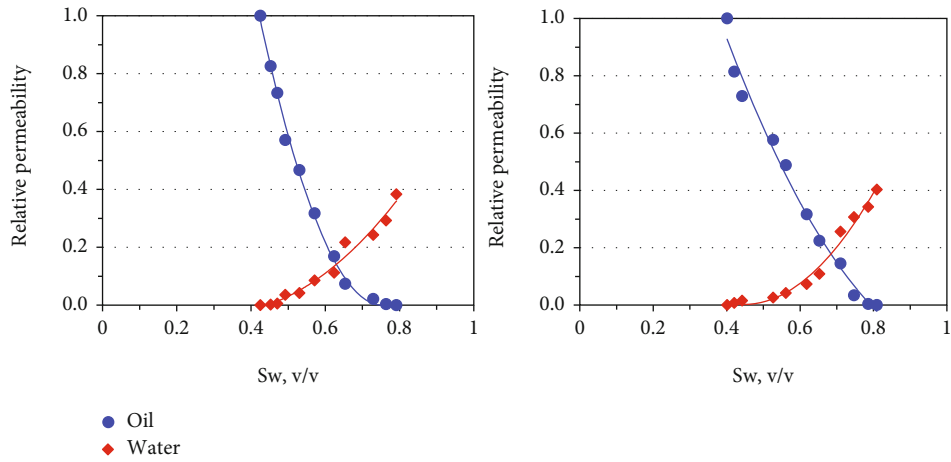


FIGURE 7: Relative permeability curves of samples T2 and T3.

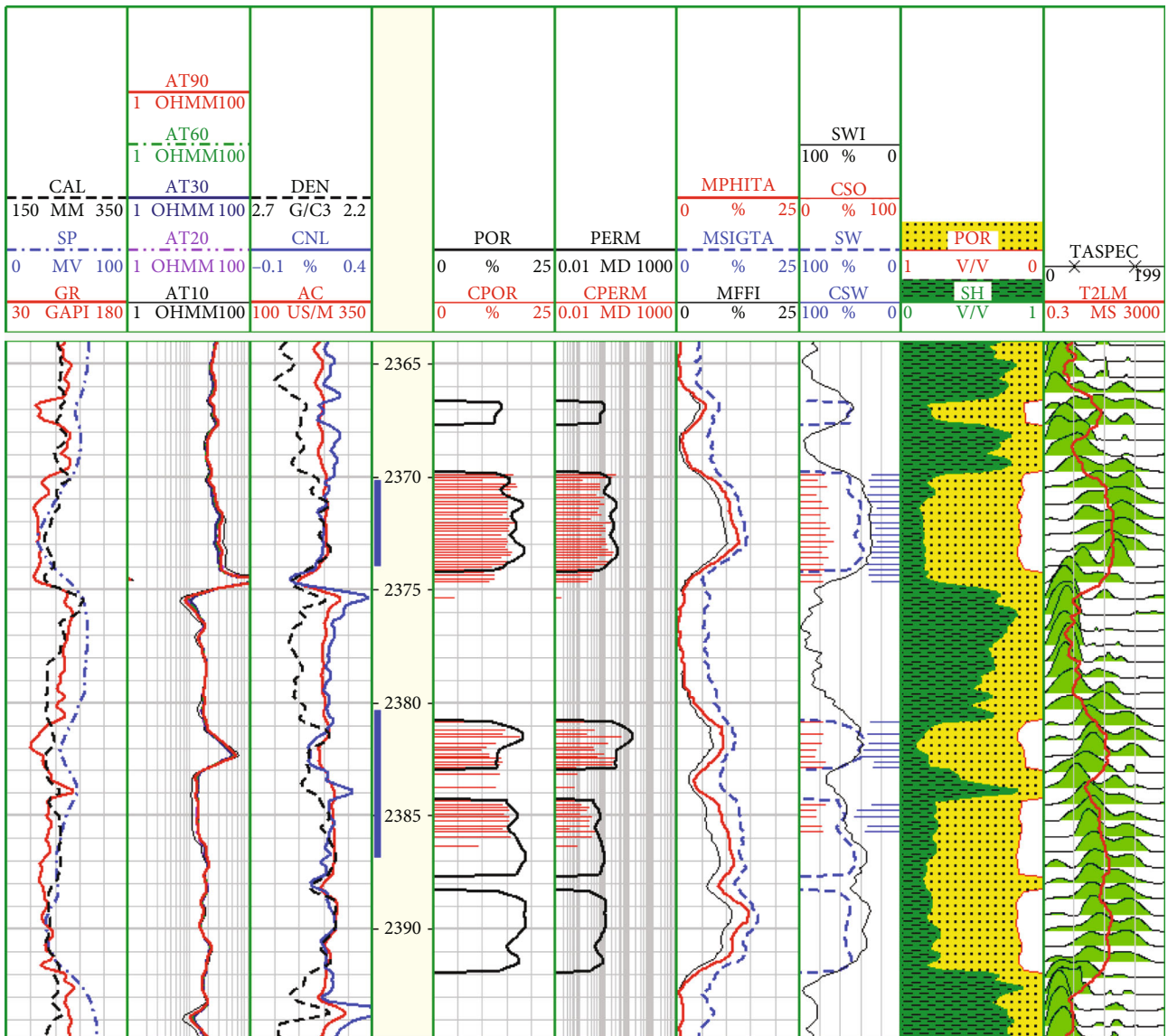


FIGURE 8: Example of producing different fluid types caused by different irreducible water saturation.



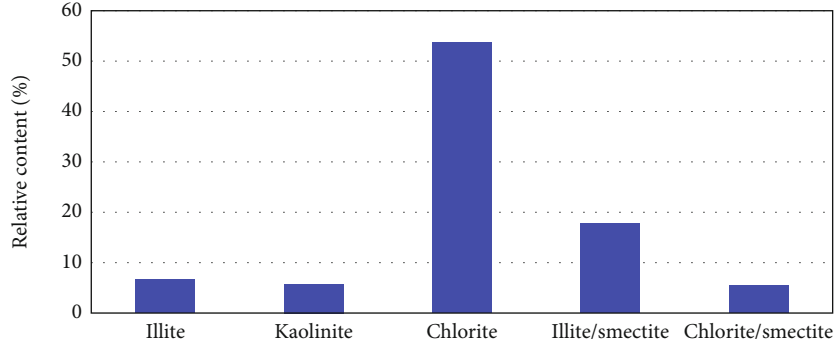
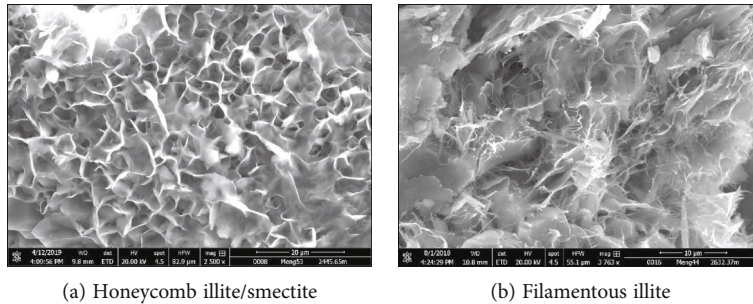


FIGURE 9: The average relative content of different clay minerals.

TABLE 2: Clay mineral analysis results of the Chang 8 member in Pengyang area.

Sample number	Illite (%)	Kaolinite (%)	Chlorite (%)	Illite/smectite (%)	Chlorite/smectite (%)
T1	0	0	22	61	0
T2	0	0	0	30	55
T3	0	0	6	29	0
T5	10	10	57	23	0
T10	13	38	44	6	0
T13	3	0	97	1	0
T14	8	0	88	4	0
T16	11	0	87	2	0
T17	6	0	80	5	0
T20	16	10	57	17	0



(a) Honeycomb illite/smectite

(b) Filamentous illite

FIGURE 10: Scanning electron microscope image for samples of Chang 8 reservoir.

causing electrical conductivity. This conductive feature produced by the cation exchange of clay minerals is called the additional conductivity of clay minerals.

The clay types and contents derived from X-ray diffraction analysis show (see Figure 9 and Table 2) that the clays in the study area are dominated by chlorite, followed by mixed layer of illite/smectite, accompanied by a small amount of illite, kaolinite, and mixed layer of chlorite/smectite. There is no independent smectite mineral. As shown in the scanning electron microscope pictures in Figure 10, the hydromica (illite) in the reservoir interstitial of Chang 8 in the study area is honeycomb or filamentous. The particle surface has a honeycomb illite/smectite clay film (Figure 10(a)). Figure 11(b) shows the pores

filled by filamentous illite clay. The chlorite film is attached to the surface of rock particles to absorb formation water, which improves the conductive network of the reservoir and reduces the resistivity of the reservoir.

Experiments worldwide indicated that the disordered illite/smectite layer clay minerals have a strong cation exchange capacity. The cation exchange capacity (CEC) and cation exchange capacity  $Q_v$  of reservoir rock are two important physical parameters that characterize the additional conductivity of clay. In this area, ten core samples were selected for CEC experimental measurement, and the measurement results are shown in Table 3. It shows that the  $Q_v$  value is small, all less than 1.0. In conventional sandstone oil and gas reservoirs, it

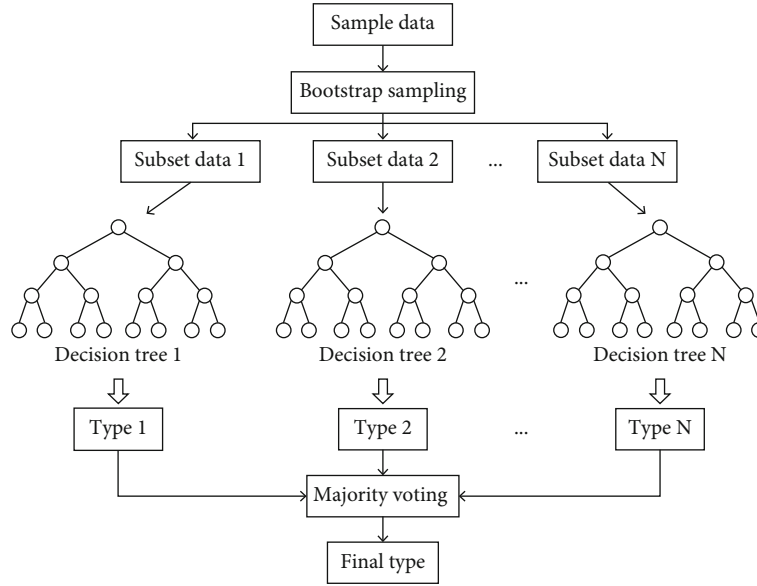


FIGURE 11: The structure of the random forest [32].

TABLE 3: Result of CEC and Qv.

Sample number	CEC (mmol/100 g)	Por (%)	Grain density (g/cm <sup>3</sup> )	Qv (meq/mL)
T2	3	0.23	2.65	0.27
T3	7.3	0.18	2.68	0.9
T5	5.1	0.17	2.71	0.66
T6	2.2	0.07	2.71	0.79
T10	1.4	0.05	2.65	0.7
T11	1.8	0.08	2.64	0.58
T15	3.1	0.13	2.64	0.54
T17	3.8	0.09	2.68	0.98
T18	2.1	0.15	2.68	0.31
T20	2.9	0.14	2.7	0.49

should not have a great impact on the conductivity of the rock. However, it has a more significant influence on Chang 8 member reservoirs with small porosity and complex pore structure. Especially in the oil layer of the target reservoirs, the equivalent cation exchange capacity  $Qv'$  ( $Qv' = Qv/Sw$ ) enhances the additional conductivity of clay. Therefore, the additional conductivity of clay minerals is one of the reasons for the formation of the low-contrast oil layer in the Chang 8 member of the Yanchang Formation.

## 5. Fluid Identification Based on RF Algorithm

**5.1. RF Algorithm.** The random forest (RF) is an ensemble learning algorithm for classification, regression, and other tasks in geological and geophysical [37]. RF utilizes multiple decision trees as base learners to build an intelligent system, which combines all prediction results from base learners by the majority voting or average approach to provide accurate results. RF, a unique form of the Bagging algorithm, applies a random selection of sample numbers and feature numbers

to generate a series of different sample subsets [32]. The hundreds of independent base decision trees are constructed based on sample subsets. In this study, the final output is obtained using majority voting due to the classification task of fluid type (Figure 11). Although a single decision tree has poor accuracy, the accuracy of the comprehensive decision could be very high since each decision tree is well trained for a specific subset.

The corresponding basic steps of the algorithm are as follows [38]: (1) bootstrap sampling with the return is used to generate multiple sample subsets from the sample data. Each subset generates a decision tree through training. (2) The optimal feature is determined from the sample subset when the decision tree divides nodes. The decision tree makes the branches of the optimal feature grow until they cannot regenerate. (3) The eventual result is obtained by voting for each prediction result from the basic decision tree.

The main advantage of the RF is that each decision tree uses part of the sample data and extracts several features for modeling. The multiple independent decision trees provide

TABLE 4: The statistical information of the input data for building the model.

	AT90 ( $\Omega\cdot\text{m}$ )	DEN ( $\text{g}/\text{cm}^3$ )	AC ( $\mu\text{s}/\text{m}$ )	$\Delta\text{SP}$ (mV)	$\Delta\text{GR}$ (API)
Maximum value	42.000	2.500	289.700	78.000	0.583
Minimum value	3.400	2.230	219.400	7.0	0.052
Average value	12.350	2.360	256.350	33.510	0.200
Standard deviation	8.082	0.056	14.239	14.222	0.099

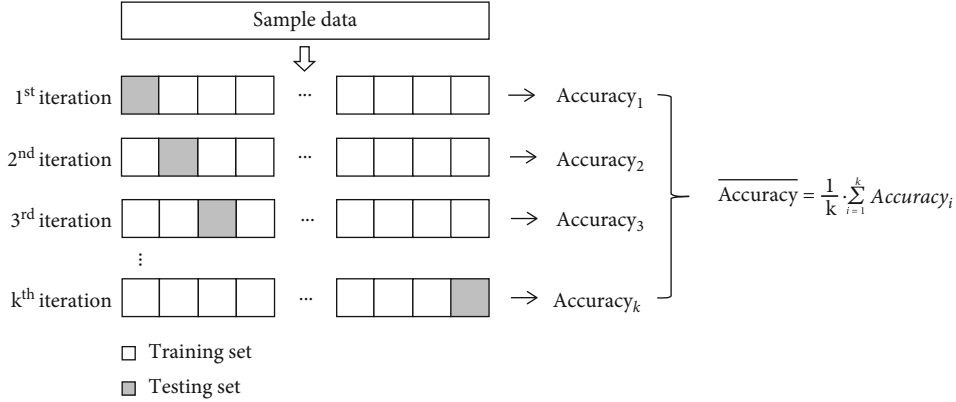


FIGURE 12: The schematic of the cross-validation method.

TABLE 5: Optimized hyperparameter settings of the random forest algorithm.

	n_estimators	Max_depth	Min_sample_split	Min_samples_leaf	Max_features	Max_samples
Optimum value	200	3	2	1	60%	2/3

RF with higher accuracy, better generation, and superior stability. Especially for a high-dimensional small sample classification problem like fluid identification of low-resistivity pay zones, the stability and generation of the model are very significant. RF has been successfully applied in many aspects, such as lithology identification [39], hydrocarbon source rock evaluation [40], and reservoir parameter prediction [41].

**5.2. Data Description and Preprocessing.** In order to illustrate the fluid identification capability of tight sandstone reservoirs of random forest, we collected data regarding the Chang 8 formation of the Triassic Yanchang Formation from the Pengyang area, Ordos Basin, China. Oil testing data of sixty layers are obtained as sample data in this study. Considering the analysis above of genetic mechanisms of low-resistivity oil pay, the complicated water salinity distribution, the impact of irreducible water saturation, and the additional conductivity of clay mainly lead to low-resistivity oil reservoirs. The distinct oil test layers may have different causes of low resistivity. The array induction resistivity log with a detection depth of 90 inch can well indicate oil and gas bearing property to some extent. The density log (DEN) and sonic time log (AC) represent the porosity of rocks. The anomaly of spontaneous potential ( $\Delta\text{SP}$ ) that equals to the difference between the SP and shale baseline could characterize water salinity. The relative value of the natural gamma ray log ( $\Delta\text{GR}$ ) describes the bound water

content and CEC. Therefore, according to the petrophysics analysis of reservoirs, the RT, DEN, AC,  $\Delta\text{SP}$ , and  $\Delta\text{GR}$  are determined as input parameters of random forest.

Sample data includes feature data from well logs and label data from fluid types of reservoirs. Table 4 shows the statistics of sample data, including the maximum value, minimum value, average value, and standard deviation of well logs. The label data includes three types of layers in the study region based on the oil testing results, namely, oil layer (OL), oil and water layer (OWL), and water layer (WL). The corresponding label data is determined using a unique vector. For example, the oil layer is (1, 0, 0), which indicates that the probability of an oil layer is 1, and the probability of other types is 0. The oil and water layer is (0, 1, 0), and the water layer is (0, 0, 1).

Due to the difference in dimension and order of magnitude of each parameter, the min-max normalization method is utilized to scale these sensitive data. Data normalization can eliminate the unit difference between different logs and improve the convergence speed of the prediction process [42]. In this study, data is normalized in the range [0,1] based on the following equation:

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}, \quad (1)$$

TABLE 6: Confusion matrix of the classification identification.

Fluid types	Predicted			Recall
	Oil layer	Oil and water layer	Water layer	
OL	10	0	0	1.0
OWL	0	8	4	0.67
WL	0	2	36	0.95
Precision	1.00	0.80	0.90	0.90

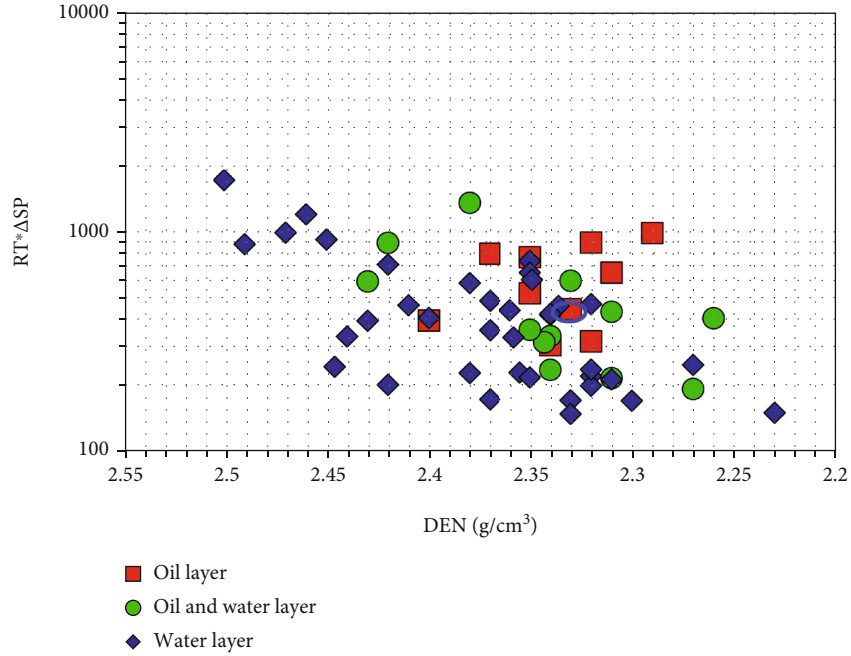


FIGURE 13: Cross plot of DEN-AT90 \* ΔSP crossplot.

where  $x_{\text{norm}}$  is the normalization processing result,  $x$  is the raw data,  $x_{\text{min}}$  is the minimum value from the original data, and  $x_{\text{max}}$  is the maximum value of the raw data.

**5.3. Hyperparameter Selection and Model Establishment.** To boost the generalization ability of the random forest, it is necessary to use the cross-validation method with grid search to optimize the hyperparameters (Figure 12). The cross-validation method divides sample data into  $k$  portions. Each portion is called a fold. The data of  $(k - 1)$  folds are exploited to the train model, and the remaining data is employed to the test model. The above steps are implemented  $k$  times until each fold is used as once testing fold. We arrange all parameters into a network of parameter combinations. Each parameter combination is input into the cross-validation method in turn for performance evaluation. The optimum parameter with the highest accuracy is determined.

In this study, hyperparameters such as the number of trees in the random forest ( $n_{\text{estimators}}$ ), the maximum depth of the tree ( $\text{max\_depth}$ ), the minimum number of samples required to split the internal nodes ( $\text{min\_sample\_split}$ ), the minimum number of samples required to be at a

leaf node ( $\text{min\_samples\_leaf}$ ), the maximum number of features to be considered when looking for the best split ( $\text{max\_features}$ ), and the number of samples for training each basic estimator ( $\text{max\_samples}$ ) are optimized using the 10-fold cross-validation method with grid search. For the random forest algorithm, those hyperparameters have a great effect on the prediction result. The  $n_{\text{estimators}}$  control the strength of random forest. The  $\text{max\_features}$  and  $\text{max\_samples}$  determine the diversity of the base decision trees. The  $\text{max\_depth}$ ,  $\text{min\_sample\_split}$ , and  $\text{min\_samples\_leaf}$  represent the complexity of the base decision trees. In our practices, the optimal values of  $n_{\text{estimators}}$  are 50, 100, 200, 300, and 500, and the optional values of  $\text{max\_depth}$  are ranged from 1 to 10, while the optional values of  $\text{min\_samples\_leaf}$  are 1, 2, 5, 10, 15, 20, 30, 40, and 60. The  $\text{min\_sample\_split}$ ,  $\text{max\_features}$ , and  $\text{max\_samples}$  are set as 2, 60%, and 2/3, respectively. According to the optimization result of 10-fold cross-validation method, the optimized hyperparameter settings are listed in Table 5.

Finally, the built random forest model is obtained to predict the fluid property of tight sandstone reservoirs. In order to analyze the prediction result, a confusion matrix is calculated from the actual and predicted types of samples. The

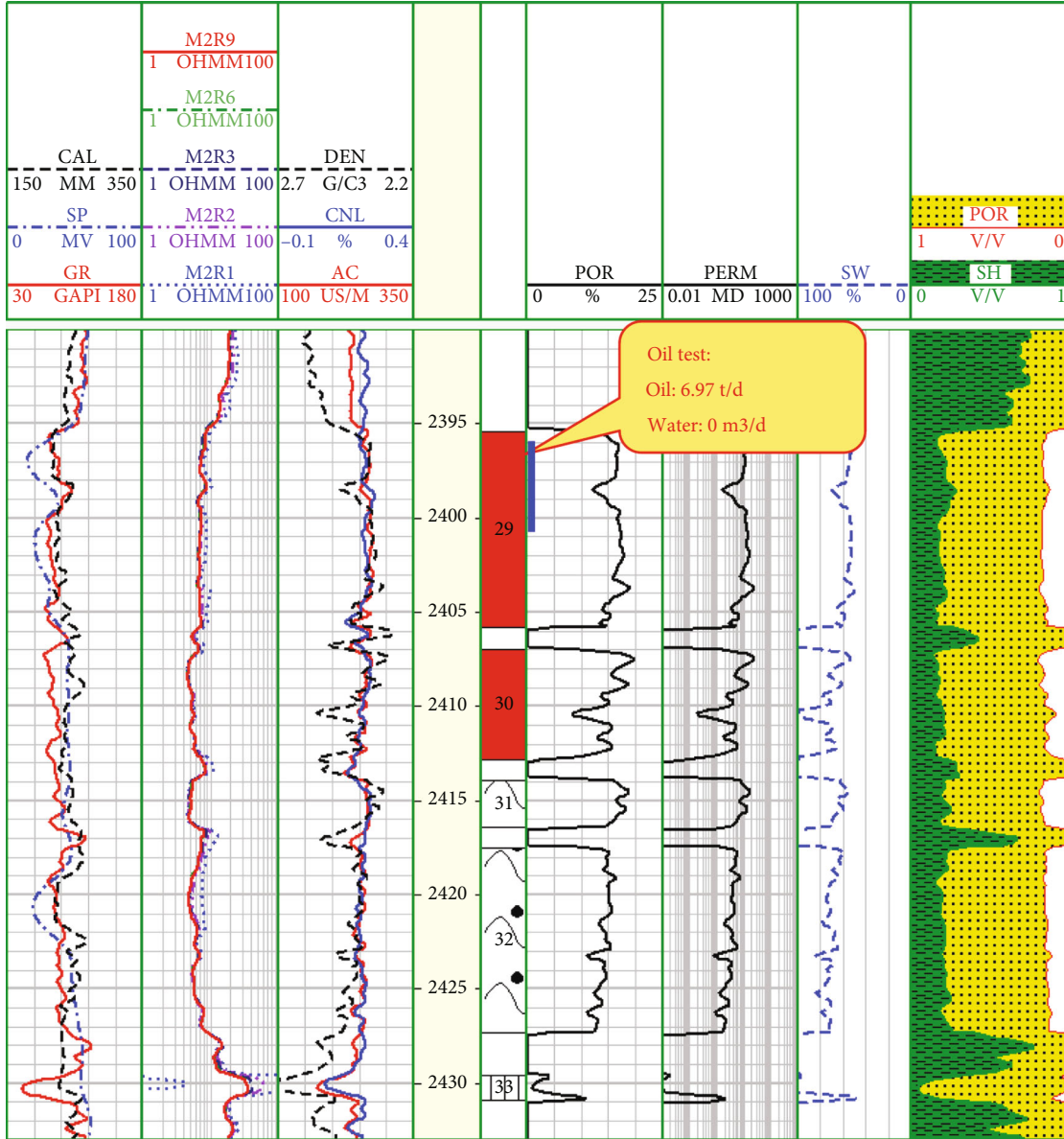


FIGURE 14: Well logs, oil test, and predicted results of Well J.

precision and recall of each sample type can be calculated based on the confusion matrix, and their equations are as follows:

$$\text{Precision} = \frac{T_p}{T_p + F_p}, \quad (2)$$

$$\text{Recall} = \frac{T_p}{T_p + F_N}, \quad (3)$$

where  $T_p$  is the number of true-positive samples,  $F_p$  is the number of false-positive samples, and  $F_N$  is the number of false-negative samples. Table 6 presents the confusion matrix of the classification identification. Taking the oil and water layer in Table 6 as an example, the  $T_p$  of the water layer is 8, the  $F_p$  of the oil and water layer is 2, and the  $F_N$  of the oil and water layer is 4.

From Table 6, the overall classification accuracy of the random forest is approximately 90.0%, which is a general coincidence rate, namely, the ratio of the correct number of layers to the total number of layers. The recall value of the oil and water layer is relatively low, because some oil and water layer samples are predicted as water layers due to the similar well log responses. Figure 13 presents a crossplot of  $\Delta SP$  with DEN, which is utilized to identify the fluid type of reservoirs in the study region. The red, green, and blue symbols represent the dots of oil, oil and water, and water layers, respectively. However, different types of fluids overlap, making it difficult to identify the fluid types of target reservoirs based on this crossplot. The random forest algorithm provides a better prediction performance than the traditional crossplot method.

To intuitively indicate the effectiveness of the random forest algorithm in the study region, we exhibit the well logs,

oil testing layers, and predicted results of Well J (Figure 14). In this figure, the array induction resistivity logs are M2R9, M2R6, M2R3, M2R2, and M2R1, which are presented at track 2. The detection depth of M2R9 is the deepest, while that of M2R1 is the shallowest. The predicted results by the RF algorithm is displayed in Track 5. Tracks 6 to 9 present the porosity, permeability, water saturation, and shale volume contents, respectively. The predicted fluid types of layers 29 and 30 are oil, and of layers 31 and 32 are water. The oil test of layer 29 is shown inside the yellow box. The blue oval in Figure 13 represents the dot of layer 29. From Figure 13, layer 29 is located in the mixed area of oil layer and water layer, making it hard to distinguish the oil from the water layers. However, the fluid of this layer predicted by the RF algorithm is oil, which verifies the effectiveness of the RF algorithm with sensitivity parameters based on the genetic mechanism of low-resistivity pay.

## 6. Conclusions

- (1) The porosity of the Chang 8 member of Yanchang Formation in the Pengyang area of Ordos Basin, China, is mainly distributed between 6% and 21%, with an average of 15.59%. And the permeability varied from 0.1 mD to 30 mD, with an average of 1.45 mD
- (2) The genetic types of the low-resistivity pay zones of the target formation are a large variations of formation water salinity, high irreducible water saturation, and clay conductivity. The water salinity mainly varied from 20 g/L to 90 g/L. The irreducible water saturation obtained from NMR and phase permeability is mainly distributed between 40% and 80%. The  $Q_v$  values varied from 0.27 to 0.98, which is relatively high in the low-porosity oil reservoirs
- (3) The RF algorithm with sensitive parameter inputs instructed by the genesis analysis was used to identify the oil, oil and water, and water. The anomaly of spontaneous potential ( $\Delta SP$ ) that characterizes water salinity, the relative value of gamma ray log ( $\Delta GR$ ) that describes the bound water content and CEC, resistivity, density, and acoustic logs were taken as sensitive logs
- (4) The identification results by RF were verified by comparison with oil test results. The accuracy of the identification is 90%, far higher than that by the crossplot method

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgments

This paper is supported by the National Natural Science Foundation of China (42004087); the Strategic Cooperation Technology Projects of China National Petroleum Corporation and China University of Petroleum, Beijing (ZLZX2020-03); the Science Foundation of China University of Petroleum, Beijing (246202BJRC001); and Natural Science Foundation of Jiangxi Province (20202BABL211020).

## References

- [1] Y. Guo, J. Liu, H. Yang, Z. Liu, and Y. Zhang, "Hydrocarbon accumulation mechanism of low permeable tight lithologic oil fields in the Yanchang Formation, Ordos Basin, China," *Petroleum Exploration and Development*, vol. 39, no. 4, pp. 447–456, 2012.
- [2] C. Feng, H. Fan, Y. Shi, X. Chen, G. Li, and Z. Mao, "Prediction for the formation water salinity in low-permeability reservoirs with complex wettability: a case study of the Triassic Chang 81 member, Longdong area, Ordos Basin," *Oil & Gas Geology*, vol. 41, no. 2, p. 7, 2020.
- [3] B. Li, X. Nie, J. Cai, X. Zhou, C. Wang, and D. Han, "U-Net model for multi-component digital rock modeling of shales based on CT and QEMSCAN images," *Journal of Petroleum Science and Engineering*, vol. 216, p. 110734, 2022.
- [4] M. P. Tixier, R. L. Morris, and J. G. Connell, "Log-Evaluation Of Low-Resistivity Pay Sands In The Gulf Coast," in *Paper presented at the SPWLA 9th Annual Logging Symposium*, Louisiana, New Orleans, 1968.
- [5] J. Zemanek, "Low-resistivity hydrocarbon-bearing sand reservoirs," *SPE Formation Evaluation*, vol. 4, no. 4, pp. 515–521, 1989.
- [6] J. Ouyang, Z. Mao, L. Xiu, Y. Shi, C. Li, and S. Li, *Genetic Mechanism for Low Contrast Hydrocarbon Bearing Reservoirs and Well Log Evaluation Method*, Petroleum Industry Press, Beijing, 2009.
- [7] O. A. de Lima and M. M. Sharma, "A grain conductivity approach to shaly sandstones," *Geophysics*, vol. 55, no. 10, pp. 1347–1356, 1990.
- [8] A. Boyd, H. Darling, J. Tabanou et al., "The lowdown on low resistivity pay," *Oilfield Rev*, vol. 7, no. 3, pp. 4–18, 1995.
- [9] Z. Mao, F. Gong, C. Liu, D. Zhu, and Z. Zuo, "Experimental study on the genesis of low resistivity pay zone in northern Tarim basin (1)," *Well Logging Technology*, vol. 23, no. 4, pp. 243–245, 1999.
- [10] Z. Mao, W. Zhu, and R. Wang, "Experimental study on the genesis of low resistivity pay zone in northern Tarim basin (2)," *Well Logging Technology*, vol. 23, no. 6, pp. 404–410, 1999.
- [11] Z. Mao, L. Kuang, C. Xiao, G. Li, and C. Zhou, "Identification and evaluation of low resistivity pay zones by well logs and the petrophysical research in China," *Petroleum Science*, vol. 4, no. 1, pp. 41–48, 2007.
- [12] L. Zhu, Y. Ma, J. Cai, C. Zhang, S. Wu, and X. Zhou, "Key factors of marine shale conductivity in southern China—part II: the influence of pore system and the development direction of shale gas saturation models," *Journal of Petroleum Science and Engineering*, vol. 209, p. 109516, 2022.
- [13] Z. Bai, M. Tan, G. Li, and Y. Shi, "Analysis of low-resistivity oil pay and fluid typing method of Chang 8<sub>1</sub> Member, Yanchang

- Formation in Huanxian area, Ordos Basin, China,” *Journal of Petroleum Science and Engineering*, vol. 175, pp. 1099–1111, 2019.
- [14] H. A. Kelleher, E. M. Braun, B. E. Milligan et al., “Low-resistivity pay evaluation using multidimensional and high-resolution magnetic resonance profiling 1,” *Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description*, vol. 49, no. 4, 2008.
- [15] Z. Huang, X. Huang, F. He, and Y. Fan, “The application of multi-parameters identification technology to low resistivity payzones in Shengli Oilfield,” *Well Logging Technol*, vol. 37, no. 6, pp. 633–637, 2013.
- [16] M. Tan, Y. Shi, W. Zhao, and G. Xie, “Joint inversion method for NMR dual-Tw logging data and fluid typing,” *Chinese Journal of Geophysics*, vol. 51, no. 5, pp. 1582–1590, 2008.
- [17] S. Stephan and J. J. Stuart, “Enhanced porosity preservation by pore fluid overpressure and chlorite grain coatings in the Triassic Skagerrak, Central Graben, North Sea, UK,” *Geological Society of London, Special Publication*, vol. 435, no. 1, pp. 321–341, 2018.
- [18] J. G. Connell, R. L. Morris, and M. P. Tixier, “Log evaluation of low-resistivity pay sands in the Gulf Coast,” *The Log Analysts*, vol. 9, no. 6, 1968.
- [19] B. Das and R. Chatterjee, “Well log data analysis for lithology and fluid identification in Krishna-Godavari Basin, India,” *Arabian Journal of Geosciences*, vol. 11, no. 10, p. 231, 2018.
- [20] G. M. Hamada and M. N. J. Al-Awad, “Petrophysical evaluation of low resistivity sandstone reservoirs,” *Journal of Canadian Petroleum Technology*, vol. 39, no. 7, pp. 7–14, 2000.
- [21] P. Ren, Z. Shi, B. Hou, Y. Zheng, and S. Cao, “Comprehensive logging evaluation of low porosity and permeability reservoir based on dual porosity model,” *Special Oil & Gas Reservoirs*, vol. 20, no. 5, p. 62-67+153-154, 2013.
- [22] J. Sun, X. Wei, and X. Chen, “Fluid identification in tight sandstone reservoirs based on a new rock physics model,” *Journal of Geophysics and Engineering*, vol. 13, no. 4, pp. 526–535, 2016.
- [23] Z. Hou, B. Li, W. Tang, Y. Li, D. Xu, and W. Wang, “New method for quantitative evaluation fluid properties of tight sandstone-correlation coefficient method,” *Progress in Geophysics*, vol. 32, no. 5, pp. 1984–1991, 2017.
- [24] B. Xu, H. Wu, Z. Feng, Y. Li, K. Wang, and P. Liu, “Application status and prospects of artificial intelligence in well logging and formation evaluation,” *Acta Petrolei Sinica*, vol. 42, no. 4, pp. 508–522, 2021.
- [25] L. Gao, R. Xie, L. Xiao, S. Wang, and C. Xu, “Identification of low-resistivity-low-contrast pay zones in the feature space with a multi-layer perceptron based on conventional well log data,” *Petroleum Science*, vol. 19, no. 2, pp. 570–580, 2022.
- [26] J. Liu and J. Liu, “Permeability predictions for tight sandstone reservoir using explainable machine learning and particle swarm optimization,” *Geofluids*, vol. 2022, Article ID 2263329, 15 pages, 2022.
- [27] Y. Zhang, K. Tong, J. Zheng, and D. Dao, “Application of support vector machine method for identifying fluid in low-resistivity oil layers,” *Geophysical Prospecting for Petroleum*, vol. 47, no. 3, pp. 306–310, 2008.
- [28] D. Liu, F. Lin, and Z. Liang, “An integrated method to determine fluid properties in carbonate reservoirs,” *Special Oil & Gas Reservoirs*, vol. 22, no. 4, pp. 104–107, 2015.
- [29] G. Chen, S. Liang, J. Wang, S. Di, Y. Zhuge, and Y. Liu, “Fluid identification in glutenites with the machine learning Ada-Boost.M2 algorithm,” *Oil Geophysical Prospecting*, vol. 54, no. 6, pp. 1357–1362, 2019.
- [30] M. Tan, Y. Bai, H. Zhang, G. Li, and A. Wang, “Fluid typing in tight sandstone from wireline logs using classification committee machine,” *Fuel*, vol. 271, p. 117601, 2020.
- [31] G. Luo, L. Xiao, X. Shi, and R. Sao, “Machine learning for reservoir fluid identification with logs,” *Petroleum Science Bulletin*, vol. 7, no. 1, pp. 24–33, 2022.
- [32] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [33] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, “Random forests for land cover classification,” *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.
- [34] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, “Customer churn prediction using improved balanced random forests,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, 2009.
- [35] R. Qin, H. Pan, P. Zhao et al., “Petrophysical parameters prediction and uncertainty analysis in tight sandstone reservoirs using Bayesian inversion method,” *Journal of Natural Gas Science and Engineering*, vol. 55, pp. 431–443, 2018.
- [36] P. Zhao, J. Fu, Y. Shi et al., “Hydrocarbon saturation in shale oil reservoirs by inversion of dielectric dispersion logs,” *Fuel*, vol. 266, p. 116934, 2020.
- [37] M. Feng, W. Yan, X. Ge, and L. Zhu, “Predicting total organic carbon content by random forest regression algorithm,” *Bulletin of Mineralogy, Petrology and Geochemistry*, vol. 37, no. 3, pp. 475–481, 2018.
- [38] Y. Ao, H. Li, L. Zhu, S. Ali, and Z. Yang, “Logging lithology discrimination in the prototype similarity space with random forest,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 687–691, 2019.
- [39] J. Cracknell and A. M. Reading, “The upside of uncertainty: identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines,” *Geophysics*, vol. 78, no. 3, pp. WB113–WB126, 2013.
- [40] L. Zhao, J. Liu, Y. Yao et al., “Quantitative seismic characterization of source rocks in lacustrine depositional setting using the random forest method: an example from the Changjiang Sag in East China Sea basin,” *Chinese Journal of Geophysics*, vol. 64, no. 2, pp. 700–715, 2021.
- [41] X. Shi, J. Wang, G. Liu, L. Yang, X. Ge, and S. Jiang, “Application of extreme learning machine and neural networks in total organic carbon content prediction in organic shale with wire line logs,” *Journal of Natural Gas Science and Engineering*, vol. 33, pp. 687–702, 2016.
- [42] J. Song, Q. Gao, and Z. Li, “Application of random forests for regression to seismic reservoir prediction,” *Oil Geophysical Prospectin*, vol. 51, no. 6, pp. 1202–1211, 2016.