

## Logging identification of lithology in fine-grained sedimentary rocks based on the FSSA-HKELM model: a case study of the Qingshankou Formation in the Songliao Basin (NE China)

Baiqiang Tang<sup>(a,b)</sup>, Qingtao Meng<sup>(a,b)\*</sup>, Zhaojun Liu<sup>(a,b)</sup>, Fei Hu<sup>(a,b)</sup>, Penglin Zhang<sup>(a,b)</sup>, Wei Dang<sup>(c)</sup>, Junxian Wang<sup>(d)</sup>

- (a) College of Earth Sciences, Jilin University, Changchun, Jilin 130061, China
- (b) Key-Lab for Oil Shale and Paragenetic Minerals of Jilin Province, Changchun, Jilin 130061, China
- (c) PetroChina Jilin Oilfield Company, China National Petroleum Corporation, Songyuan 138000, China
- (d) Beijing Research Institute of Uranium Geology, Beijing 100029, China

Received 18 April 2024, accepted 12 July 2024, available online 5 August 2024

**Abstract.** *The lithology distribution of fine-grained sedimentary rocks in regional space plays a crucial role in guiding the exploration of oil shale layers. Identifying lithology by conventional logging information is efficient and cost-effective. However, the strong inhomogeneity of fine-grained sedimentary rocks leads to a complex nonlinear relationship between lithology and logging data, making conventional linear methods no longer applicable. This study proposes a hybrid model for logging lithology identification of fine-grained sedimentary rocks from the first member of Qingshankou Formation in the Songliao Basin, NE China. This model is based on the hybrid kernel extreme learning machine (HKELM), and the firefly perturbation strategy is introduced into the sparrow search algorithm (FSSA) for optimization. The lithologic distribution is determined using cores, thin sections, and total organic carbon (TOC), while a total of four logging curves, the acoustic (AC), density (DEN), resistivity (RT), and natural gamma (GR) curves, were collected. FSSA-HKELM was compared with five algorithms, ELM, KELM, HKELM, PSO-HKELM, and SSA-HKELM, for lithology prediction effectiveness. The proposed hybrid method outperformed the other algorithms, achieving an accuracy of 81.78%, precision of 81.65%, recall of 87.78%, and a weighted F1 score of 82.16%. FSSA-HKELM is a very effective lithological identification method, which provides a basis for the lithological prediction of oil shale-bearing formations.*

**Keywords:** *lithology identification, sparrow search algorithm, firefly perturbation strategy, hybrid kernel extreme learning machine, fine-grained sedimentary rocks.*

\* Corresponding author, [mengqt@jlu.edu.cn](mailto:mengqt@jlu.edu.cn)

## 1. Introduction

Fine-grained sedimentary rocks are primarily composed of grains with a granularity smaller than 0.0625 mm, encompassing clay minerals, carbonates, quartz, and organic substances as their main constituents [1, 2]. As unconventional energy gains prominence, fine-grained sedimentary rocks, particularly oil shale, are increasingly becoming a focus of research in this sector [3]. The distribution of different lithologies in fine-grained sedimentary reservoirs controls hydrocarbon enrichment, which makes exploring the distribution of lithologies in regional spaces extremely important.

Drilling and coring represent the most straightforward and precise approaches for identifying lithology. However, studying the spreading pattern of lithology in two or even three dimensions using these methods is difficult due to the high cost of coring required. Currently, logging technology is widely used because of its low cost, high vertical resolution, and comprehensive geological information [4]. Common logging methods for lithology determination in conventional reservoirs include cross-plotting, curve plotting, curve superposition, and statistical analysis [5, 6]. In-depth studies of unconventional shale oil have gradually shown that fine-grained sedimentary rocks exhibit strong heterogeneity, making it difficult to express the relationship between logging and lithology as a simple linear equation. Therefore, new methods must be proposed to solve this complex nonlinear problem.

Recently, machine learning (ML) has found applications across various fields, including engineering, economics, and education. It exhibits good applicability to nonlinear problems [7–9], and many scholars have introduced it to lithology prediction in the geology field. The methods most often used for this purpose are the supervised learning algorithms, such as K-nearest neighbors (KNN), support vector machine (SVM), back propagation artificial neural network (BP-ANN), decision tree (DT), random forest (RF), AdaBoost, and XGBoost [10–13], along with the unsupervised learning versions of the K-means and SOM (self-organizing map) methods [14, 15], all of which have achieved good results in lithology prediction.

However, using the aforementioned methods to predict lithology still presents some unresolved issues. These problems can be categorized into two types:

1. Traditional algorithms exhibit slow iteration speed and poor generalizability. The extreme learning machine (ELM) can overcome these issues, thanks to its structural simplicity, rapid training speed, and high generalizability. Huang et al. [16] also proved that ELM is superior to traditional algorithms (SVM and BP-ANN). Huang et al. [17] further proposed an improved method known as the kernel extreme learning machine (KELM), which employs kernel mapping to substitute the random mapping used during the ELM training phase. This approach reduces computational complexity and enhances model

generalization, addressing issues such as the artificial setting of ELM's hidden layers and the random generation of model weights and thresholds. Wang et al. [18] also confirmed that KELM is superior to ELM.

2. ML algorithms require setting multiple parameters, and only appropriate parameters can ensure high prediction accuracy, so finding the optimal parameter combination for enabling the algorithm to build the best model is an important challenge. Lately, the swarm intelligence (SI) optimization algorithm has garnered significant interest, being successfully adopted by many scholars for ML parameter optimization [19, 20]. For instance, in the application of logging-based lithology prediction, Sun et al. [21] proposed using the Bayesian optimized parameters of XGBoost to predict the lithology of the Ordos Basin and attained an area of 0.987 under the receiver operating characteristic curve within the test dataset. Gu et al. [22] suggested employing the linear artificial fish swarm algorithm for optimizing LightGBM parameters, improving the accuracy of lithology prediction for sandstone reservoirs. SI performs a stochastic search by mimicking the relationships between individuals and groups of organisms [23]. Some of the more commonly used algorithms include ant colony optimization (ACO) [24], particle swarm optimization (PSO) [25], gray wolf optimization (GWO) [26], and the whale optimization algorithm (WOA) [27]. Inspired by the foraging and refeeding behaviors of sparrows, the sparrow search algorithm (SSA) is considered a relatively effective technique currently in use [28]. However, the initial population of this algorithm is still as easy to lack diversity as classical algorithms, such as PSO, which makes it difficult to achieve global optimization.

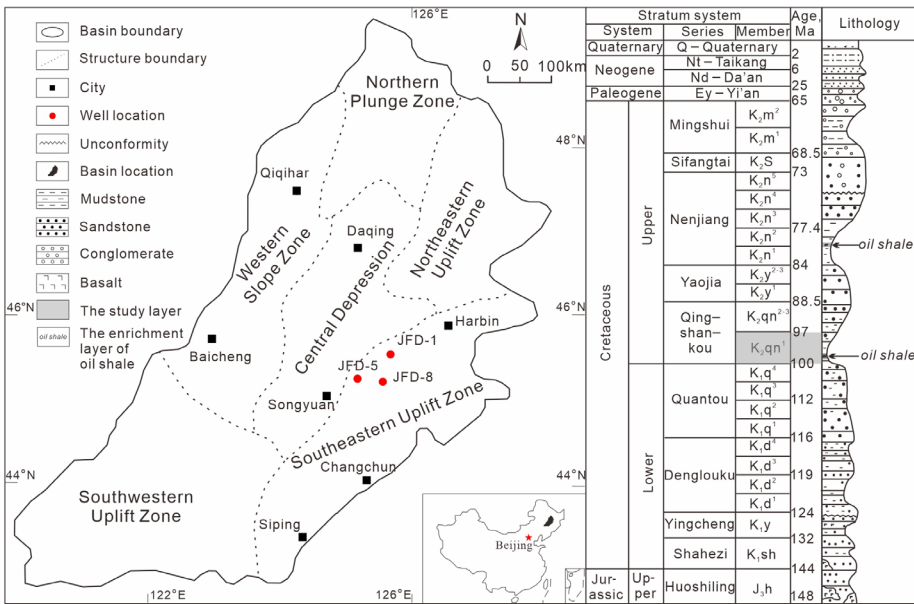
In this study, we chose the lithology and logging data of the first member of the Upper Cretaceous Qingshankou Formation (oil shale formation) in the Southeastern Uplift Zone of the Songliao Basin as an example. Using ELM as the base model, the radial basis function (RBF) and polynomial function are introduced and combined by weighting to build a hybrid kernel extreme learning machine (HKELM). This approach overcomes the limitations of KELM, which uses a single kernel function and cannot simultaneously satisfy high learning capacity and generalizability. HKELM's parameters are optimized using the SI optimization algorithm.

Given that SSA is prone to local optima and slow convergence, many scholars have enhanced SSA by integrating the firefly algorithm, thereby improving its local search capability, convergence speed, and accuracy [29, 30]. In this study, firefly perturbation is introduced into SSA for parameter optimization. A hybrid model, FSSA-HKELM, has been established to predict the lithological distribution of fine-grained sedimentary rocks. This model elevates the accuracy, stability, and training speed of lithological predictions, thereby furnishing an essential groundwork for the exploration of oil shale.

## 2. Geological setting

Situated in northeastern China, the Songliao Basin represents the country's largest terrestrial oil- and gas-bearing sedimentary basin. The basin extends up to 750 km in length and 300–370 km in width. It is segmented into six main tectonic units and has experienced four phases of tectonic development: pririfting, rifting, postrifting thermal subsidence, and tectonic inversion [31, 32]. The basin's sedimentary thickness reaches up to 10 km, with its center covering strata from the Jurassic to the Quaternary periods. Moreover, the most widely distributed Cretaceous strata are the main developmental layers (up to 7 km thick), and the sedimentary thickness gradually decreases towards the edge [33]. A detailed description is shown in Figure 1.

The Qingshankou Formation ( $K_2qn$ ) serves as a principal hydrocarbon-bearing horizon developed from the Upper Cretaceous series. It is a large-scale lake intrusion with three sets of formations, from oldest to newest as follows:  $K_2qn^1$  (10–100 m thick),  $K_2qn^2$  (310 m thick), and  $K_2qn^3$  (18–266 m thick) [34].  $K_2qn^1$  is the primary oil shale enrichment layer within  $K_2qn$  [35]. The enrichment of oil shale during this period was controlled by sea level rise and tectonic factors. A gradual increase in temperature during the Upper Cretaceous led to a rise in sea level, which increased the water depth of the lakes. At the same time, the basin was in a phase of tectonic stability and continued to decline, which further increased the water depth and the lake area. The bottom of the lake basin formed a water column stratification,



**Fig. 1.** Songliao Basin location, tectonic unit distribution, and stratigraphic map, along with the locations of boreholes for the JFD-1, JFD-5 and JFD-8 wells (modified from Jia et al. [36]).

which promoted the preservation of organic matter at the bottom of the lake. Eventually, widely distributed and thick black oil shale deposits were formed. This study focuses on the fine-grained sedimentary rock strata of  $K_2qn^1$  in the Southeastern Uplift Zone of the Songliao Basin, using it as a case study to analyze and compare lithologic identification methods based on logging.

### 3. Materials and methods

#### 3.1. Sample description

The core observation shows that oil shale, mudstone, sandstone, and dolomite are mainly developed in the study area. From wells JFD-1 (91 samples), JFD-5 (37 samples), and JFD-8 (30 samples), a total of 158 core samples were selected for evaluation, focusing on oil yield ( $\omega$ ) and total organic carbon (TOC) assays. Additionally, well logging data for JFD-1 (376.125–461.000 m in depth), JFD-5 (399.125–495.500 m), and JFD-8 (439.000–528.750 m) were gathered, with a sampling interval of 0.125 m. The data primarily consisted of four curves: resistivity ( $R$ ,  $\Omega \cdot m$ ), natural gamma (GR, API), acoustic (AC,  $\mu s/m$ ), and density curves (DEN,  $g/cm^3$ ).

#### 3.2. Geochemical methods

##### 3.2.1. Oil yield ( $\omega$ ) measurement

For the  $\omega$  test, the rock is first coarsely pulverized to 3 mm and then 50 g of rock powder is weighed into an aluminium retort. The rock powder sample is heated to 520 °C in stages according to industry standards. The water is separated from the obtained oil-water mixture, its volume is recorded, and the oil yield is calculated.

##### 3.2.2. TOC measurement

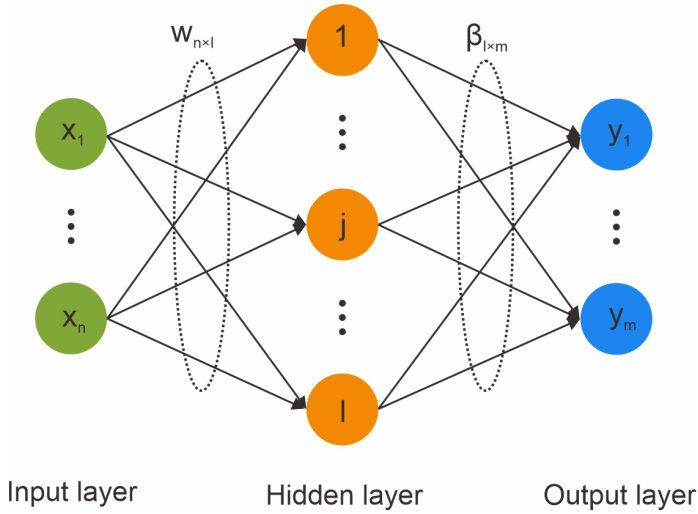
TOC is measured with a LECO analyzer. Before testing, the rock is pulverized to a particle size of 200 mesh and weighed to 70 mg in a crucible. The rock powder is then acidified using diluted hydrochloric acid at a concentration of 5% to remove inorganic carbon (carbonates) from the rock [37].

#### 3.3. Machine learning methods

##### 3.3.1. HKELM

###### *The principle of ELM*

ELM represents a straightforward yet powerful single-layer feedforward neural network. It eliminates the need for extensive parameter tuning; instead, the network randomly initializes the weights of the input layer and the biases of the hidden layer. Furthermore, its operation merely requires specifying the neuron count of the hidden layer. ELM performs well in both classification



**Fig. 2.** ELM algorithm structure.  $X_1$  to  $X_n$  represents the input vector and  $Y_1$  to  $Y_m$  the output vector of all samples.

and regression applications, and exhibits simple algorithm structures and good generalizability [38]. Figure 2 illustrates the mathematical structure of the method, depicting its input layer, hidden layer, and output layer, connected by weights and biases.

The ELM model's output is represented by Equations (1)–(3) [37]:

$$f(x) = h(x) \times \beta = H \times \beta = T, \quad (1)$$

$$\beta = H^+ \times T, \quad (2)$$

$$H^+ = H^T (HH^T)^{-1}, \quad (3)$$

where  $x$  represents the input feature vector,  $h(x)$  and  $H$  denote the hidden layer's output mapping matrices,  $\beta$  is the weight connecting the hidden layer and the output layer,  $H^+$  is the generalized inverse matrix of  $H$ , and  $T$  is the target desired output of the network.

#### *The principle of KELM*

Although ELM outperforms traditional neural networks by reducing algorithm complexity and improving training speed, computational accuracy, and generalizability, its predictive accuracy is affected by the predetermined number of neurons in the hidden layer, as well as by the random assignment of weights and thresholds. To address this, Huang et al. [17] introduced the concept of kernel functions into ELM and proposed KELM. The output function for KELM is defined as follows:

$$f(x) = h(x)\beta = h(x)H^T \left( \frac{I}{C} + HH^T \right)^{-1} T, \quad (4)$$

where  $C$  represents the penalty factor and  $I$  denotes the unit matrix. The kernel matrix for KELM is defined as follows:

$$\Omega_{KELM} = HH^T, \Omega_{KELM_{i,j}} = h(x_i)h(x_j) = K(x_i, x_j). \quad (5)$$

Therefore, based on Equations (4) and (5), KELM can be expressed as follows:

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix} \left( \frac{I}{C} + \Omega_{KELM} \right)^{-1} T. \quad (6)$$

### The principle of HKELM

Different kernel functions impact KELM's performance variably, with RBF and the polynomial function being the most commonly chosen. RBF, a local kernel function, exhibits strong learning capability but limited generalizability. Conversely, the polynomial function, as a global kernel function, has weaker learning capability but higher generalizability [39]. To capitalize on these strengths, the HKELM method combines RBF and the polynomial function, thereby enhancing the model's learning capability and generalizability simultaneously. The two kernel functions are expressed in Equations (7) and (8) [40]:

$$K_{RBF}(x, x_i) = \exp\left(\frac{-x - x_i^2}{2\sigma^2}\right), \quad (7)$$

$$K_{Poly}(x, x_i) = (x \cdot x_i + C_0)^b, \quad (8)$$

where  $\sigma$  is the coefficient of the RBF kernel function, and  $C_0$  and  $b$  are the two coefficients of the polynomial function. Therefore, HKELM is represented as follows [41]:

$$K_{hybrid} = \omega_0 \cdot \exp\left(\frac{-x - x_i^2}{2\sigma^2}\right) + (1 - \omega_0) \cdot (x \cdot x_i + C_0)^b, \omega_0 \in [0, 1], \quad (9)$$

where  $\omega_0$  is the weight coefficient of HKELM that adjusts the respective weights of the two kernel functions.

### 3.3.2. The principle of FSSA

#### SSA

SSA is a high-performing SI algorithm, developed based on the foraging and anti-trapping behavior of sparrows [28]. In terms of the convergence rate and computational precision, SSA excels beyond traditional optimization methods,

such as PSO and GWO. SSA is comprised of three roles: producer, scrounger, and vigilante. The sparrow's position is expressed as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,d} \end{bmatrix}, \quad (10)$$

where  $X$  denotes the position in space of all individuals of the population,  $d$  represents the number of parameters to be searched, and  $n$  denotes the size of the sparrow population involved in the parameter search.

The producer has the duty of locating the area with food and leading the scrounger to this location. Producers are highly adaptable and usually encompass 10–20% of the population. The producer's position is calculated as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot \text{iter}_{max}}\right), & R_2 < T \\ X_{i,j}^t + Q \cdot L, & R_2 \geq T \end{cases}, \quad (11)$$

where  $t$  and  $\text{iter}_{max}$  are the current and maximum iterations of the algorithm, respectively,  $Q$  is a number randomly generated according to a normal distribution,  $L$  is a matrix consisting entirely of ones, with 1 row, and  $d$  columns,  $R_2$  and  $T$  represent the warning and safety values, with ranges of  $[0, 1]$  and  $[0.5, 1]$ , respectively.

The scrounger monitors whether the producer has found good food. Once good food is found, the scrounger immediately updates the location to obtain food for higher energy. The scrounger's location is calculated as follows:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^t - X_{i,j}^t}{i^2}\right), & i > n/2 \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \cdot A^+ \cdot L, & i \leq n/2 \end{cases}, \quad (12)$$

where  $X_{worst}^t$  is the globally worst position (lowest fitness),  $X_p^{t+1}$  is the globally best position (highest fitness) of the population at the  $t$ -th iteration,  $A$  is a matrix featuring 1 row and  $d$  columns, where each element within  $A$  is randomly assigned either 1 or  $-1$ , and  $A^+ = A^T(AA^T)^{-1}$ . When  $i$  is greater than half of  $n$ , it means that the  $i$ -th sparrow is starving due to being in a poor position (lower fitness) and urgently needs to relocate to other positions to find food and replenish its energy. When  $i$  is less than or equal to half of  $n$ , the  $i$ -th sparrow then acquires energy at a location in the vicinity of  $X_p^{t+1}$ .



To prevent predation, 10–20% of the sparrow population are set as vigilantes, warning others of danger. The position of the vigilante is calculated as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot |X_{i,j}^t - X_{best}^t|, & f_i > f_g \\ X_{i,j}^t + K \cdot \left( \frac{|X_{i,j}^t - X_{worst}^t|}{(f_i - f_w) + \varepsilon} \right), & f_i = f_g \end{cases}, \quad (13)$$

where  $X_{best}^t$  is the globally optimal position of the population at the  $t$ -th iteration,  $\beta$  is the control step when the sparrow moves, and the value is generated randomly according to a normal distribution with mean and variance of 0 and 1,  $K$  denotes a random number that determines the direction in which the sparrow searches, with values ranging from  $-1$  to  $1$ , and  $f_i$  denotes the fitness of the present sparrow individual  $i$ , with  $f_g$  being the highest and  $f_w$  the lowest fitness value. Additionally,  $\varepsilon$  is assigned the smallest possible constant to ensure the denominator does not become 0.

#### Firefly perturbation strategy

The firefly algorithm (FA) draws inspiration from the luminous behavior of fireflies, which are attracted to each other's light, and it primarily has the following three characteristics [42]:

1) Fireflies are not attracted to other fireflies because they are the opposite sex; instead, they are attracted to fireflies whose brightness is greater than their own.

2) The strength of attraction between fireflies is influenced by their light intensity; less bright individuals move towards those with more brightness. At the same time, the brightness decreases because of the increasing distance between individuals.

3) If a given firefly does not find individuals brighter than itself, it moves randomly through space.

The calculation of the firefly's light intensity value proceeds as follows:

$$I = I_0 \times e^{-\gamma r_{i,j}}, \quad (14)$$

the attraction between fireflies can be represented as follows:

$$\beta = \beta_0 \times e^{-\gamma r_{i,j}^2}, \quad (15)$$

and the location of the firefly is updated as follows:

$$x_i(t+1) = x_i + \beta \times (x_j - x_i) + \alpha \times (rand - 0.5), \quad (16)$$

where  $I_0$  represents the maximum brightness (own brightness),  $\gamma$  denotes the light absorption coefficient, inversely related to the distance between fireflies,  $r_{i,j}$  is the distance between individuals,  $\beta_0$  is the maximum attraction of the

fireflies,  $x_i$  and  $x_j$  are fireflies at positions  $i$  and  $j$ , respectively,  $\alpha$  is a parametric factor that controls the step, and  $rand$  is a random number between  $[0, 1]$ .

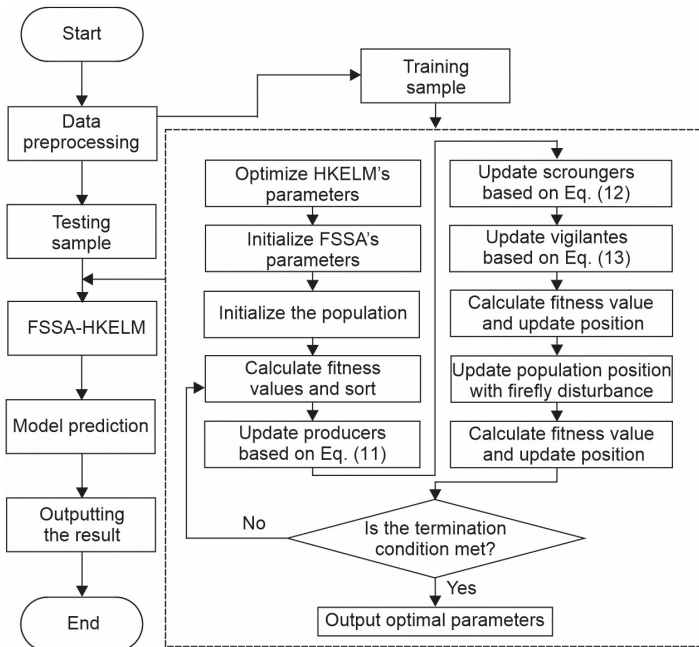
Introducing the firefly perturbation strategy updates the sparrow's search position to find the optimal individual, overcoming issues such as the sparrow population getting trapped in local optima and low search precision. This establishes the FSSA optimization algorithm framework, thereby improving the speed and accuracy of SI optimization algorithms in searching for optimal HKELM parameters.

### 3.3.3. Flow of FSSA-HKELM

The analysis of the above algorithms indicates that FSSA has better search precision and convergence speed than SSA and other classical algorithms, and HKELM has better generalizability and learning capability than KELM and ELM. Therefore, FSSA was chosen to optimize five parameters ( $C, C_0, b, \sigma, \omega_0$ ) of HKELM in this case. The flow of the algorithm is shown in Figure 3 and the text describes the flow from steps 1 to 10:

Step 1. The logging data to be input are normalized and preprocessed using Equation (1). Afterwards, wells JFD-1 and JFD-5 are set as the training set, while JFD-8 serves as the test set. The training set is then used by FSSA to optimize HKELM and obtain the optimal parameters.

Step 2. The following model parameters are initialized: number of sparrows, iteration times, ratio of producers and scroungers in the population, warning value, safety value sizes, and initial parameters of HKELM.



Step 3. The population is initialized.

Step 4. The fitness for every individual sparrow in the population is calculated, and the sparrows are ranked according to their values.

Step 5. The individuals with high fitness are selected as producers based on the set proportion of producers in the population, and their positions are updated based on Equation (11).

Step 6. The remaining sparrow individuals are used as scroungers, with their positions updated based on Equation (12).

Step 7. A certain number of sparrow individuals are chosen at random to serve as vigilantes, based on the predetermined ratio of vigilantes in the population. These individuals warn others of dangerous situations. Their positions are updated based on Equation (13) to calculate the new fitness value. If this value is superior to the optimal value in the present sparrow population, then it is updated.

Step 8. Firefly perturbation is introduced to the sparrow population based on Equation (16).

Step 9. Fitness is calculated corresponding to the updated position of the sparrow population, and optimal individuals are retained.

Step 10. If the stopping condition is satisfied, the optimal global sparrow position and the highest fitness level are simultaneously output; if not, the process reverts to step 4.

#### 3.3.4. Model evaluation parameters

To comprehensively evaluate the predictive ability of each lithology identification model and overcome the one-sidedness of evaluating models by a single indicator, the comprehensive accuracy, precision, recall, and F1 indicators were selected. The samples were divided into true negative (TN), false negative (FN), true positive (TP), and false positive (FP). Based on the predicted classification effect, the proposed four comprehensive evaluation indicators were then calculated by these four results.

Accuracy represents the proportion of correctly predicted samples to the overall sample count, though it does not evaluate the imbalance in the number of sample categories. Precision and recall can effectively overcome this. Precision is the proportion of true positive samples within the positively predicted samples by the model, while recall is the proportion of true positive samples among the actual positive samples that are correctly predicted as positive. The F1 score is calculated by averaging the reconciliation of precision and recall. The above evaluation parameters are calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (17)$$

$$Precision = \frac{TP}{TP + FP}, \quad (18)$$

$$Recall = \frac{TP}{TP + FN}, \quad (19)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (20)$$

### 3.4. Logging data normalization

Different logging curves have different units and amplitude ranges, which can affect the performance of the model when raw log data is fed directly into the training model. To prevent the effects of differences in magnitude between different logging curves, the logging data are mapped to  $[0, 1]$ , using the maximum-minimum normalization method as follows:

$$Z_n = \frac{Z - Z_{min}}{Z_{max} - Z_{min}}, \quad (21)$$

where  $Z$  represents the original logging data,  $Z_n$  denotes the normalized logging value,  $Z_{min}$  signifies the minimum value within the logging data, and  $Z_{max}$  indicates the maximum value in the logging data.

## 4. Results and discussion

### 4.1. Characteristics of lithologies

#### 4.1.1. Determination of oil shale

Due to the difficulty in distinguishing between organic-rich mudstone and oil shale based solely on core observations, the lack of a solid basis can lead to inaccurate training data for lithology prediction models and result in misjudgment during log-based identification. Therefore, it is proposed to use TOC to determine oil shale. Oil shale is a fine-grained sedimentary rock rich in organic matter with a  $\omega$  greater than 3.5%; oil shale can be heated to produce shale oil and gas [43–45].  $\omega$ , the mass proportion of shale oil in oil shale, is an important parameter for evaluating oil shale [44]. Therefore, the  $\omega$  parameter is selected as the basis for determining oil shale. However, due to limiting factors, such as difficult sample coring and expensive test analysis, oil shale cannot be accurately divided into the whole well segment of the target layer, so the following steps are carried out:

#### (1) Determination of TOC boundaries for oil shale

By conducting a linear fit between  $\omega$  and TOC, a good correlation ( $R^2 = 0.95$ ) is observed, as depicted in Figure 4, leading to the derivation of Equation (22):

$$TOC = 1.1845 \times \omega + 1.1086, R^2 = 0.95 \quad (22)$$

Inserting the limit value for oil shale  $\omega$  (3.5 wt%) into this equation yields a corresponding TOC of 5.25 wt%. Hence, mud shales with TOC exceeding 5.25 wt% are classified as oil shale.

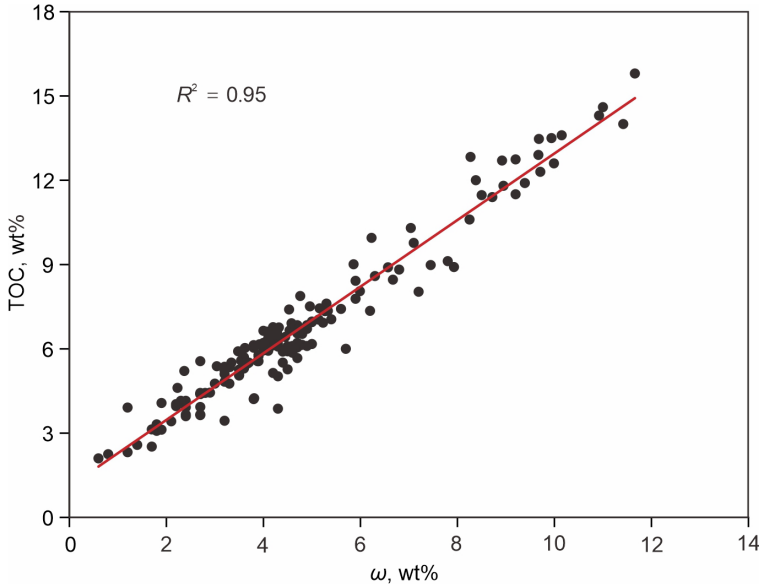


Fig. 4. Relationship between oil yield and TOC.

(2) Logging evaluation of TOC

The 158 measured TOC and corresponding logging data values of the JFD-1, JFD-5 and JFD-8 wells were used to self-predict the TOC of these three wells. The prediction method is a modified  $\Delta\log R$  method.  $\Delta\log R$  is a method proposed by Passey et al. [46] to predict the TOC of hydrocarbon source rocks. The mathematical idea is to establish an empirical formula from the difference in magnitude between the RT curve and the porosity curve (DEN, AC, and compensated neutron logging) versus TOC. The formula can be found in [46].

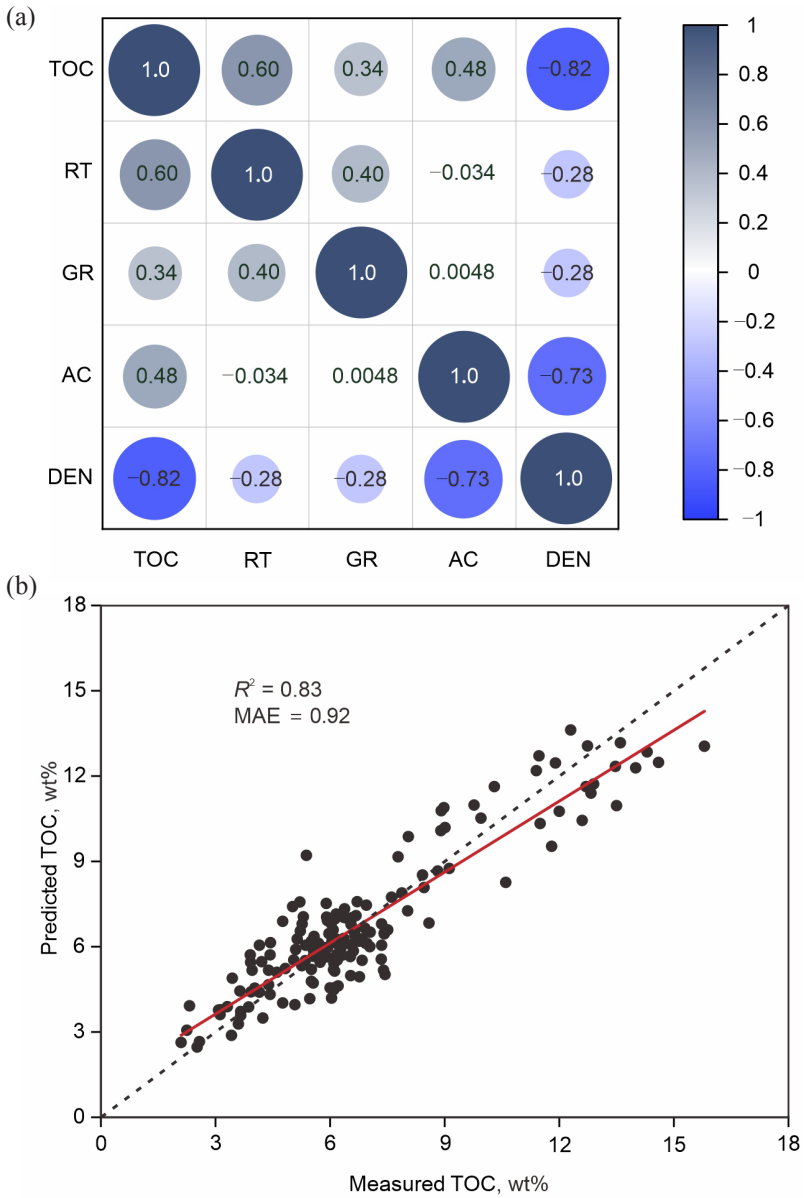
However, this method requires maturity data and artificial debugging of multiple baseline values, which is labor-intensive and time-consuming. To address these issues, the improved  $\Delta\log R$  method proposed by Hu et al. [47] is used. According to the Pearson coefficient in Figure 5a, DEN and TOC have the best correlation. Therefore, the two logging curves of RT and DEN are used for calculation, and Equations (23)–(25) are obtained:

$$\Delta\log R = \log_{10}(R / R_{min}) + \frac{\log_{10}(R_{max} / R_{min})}{\rho_{max} - \rho_{min}} \times (\rho_{min} - \rho), \quad (23)$$

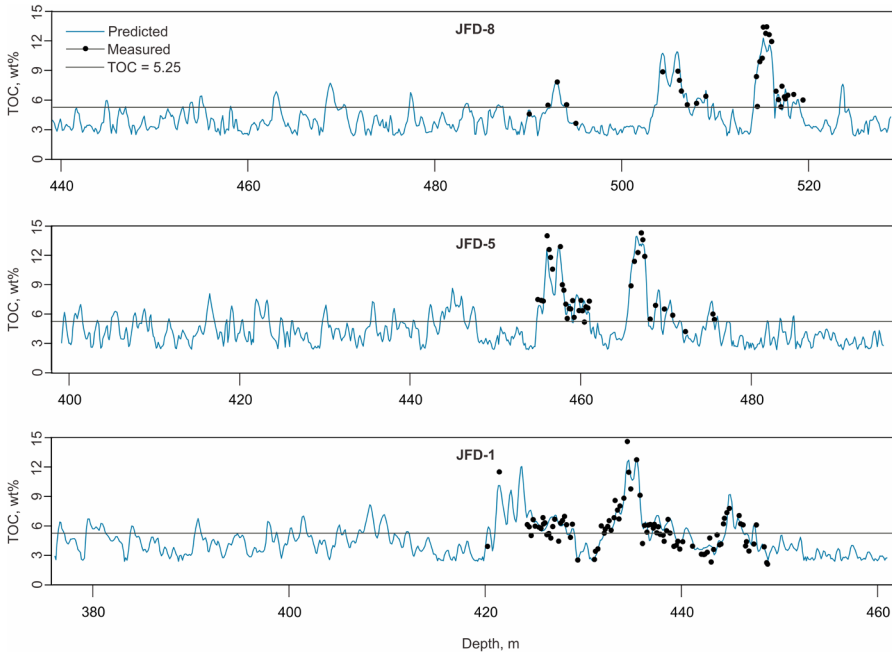
$$TOC = a\Delta\log R + b, \quad (24)$$

$$TOC = 7.2041 \times \Delta \log R + 2.411, \tag{25}$$

where  $R_{max}(\rho_{max})$  and  $R_{min}(\rho_{min})$  represent the maximum and minimum values on the corresponding scales when the curves are superimposed, respectively, and  $a$  and  $b$  are the coefficients of the linear fit of TOC and  $\Delta \log R$ .



**Fig. 5.** Correlation between TOC and logging (a), improved  $\Delta \log R$  predicted TOC with measured TOC (b). MAE – mean absolute error.



**Fig. 6.** Improved  $\Delta\log R$  method for TOC prediction in the target formation of the JFD-1, JFD-5 and JFD-8 wells.

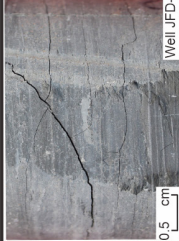
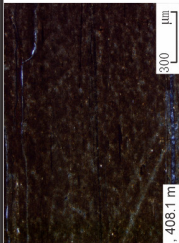
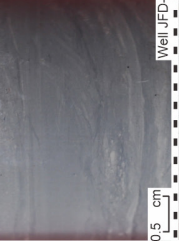
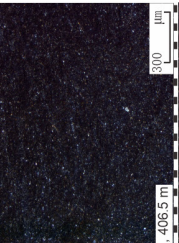

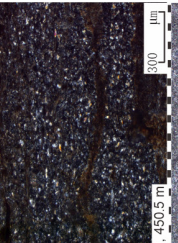
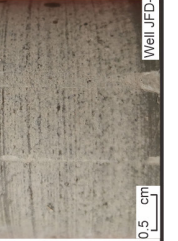
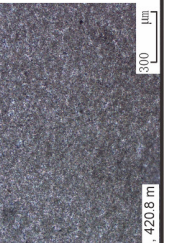
A linear fit of the measured TOC values from the experimental analysis to the TOC values predicted by the proposed method leads to Figure 5b. The correlation coefficient  $R^2 = 0.83$ , the mean absolute error MAE = 0.92, and the TOC prediction accuracy is high, as shown in Figure 6. The predicted TOC values of the three wells are very close to the measured values, which can be used to determine the presence of oil shale.

#### 4.1.2. Petrological and logging characteristics

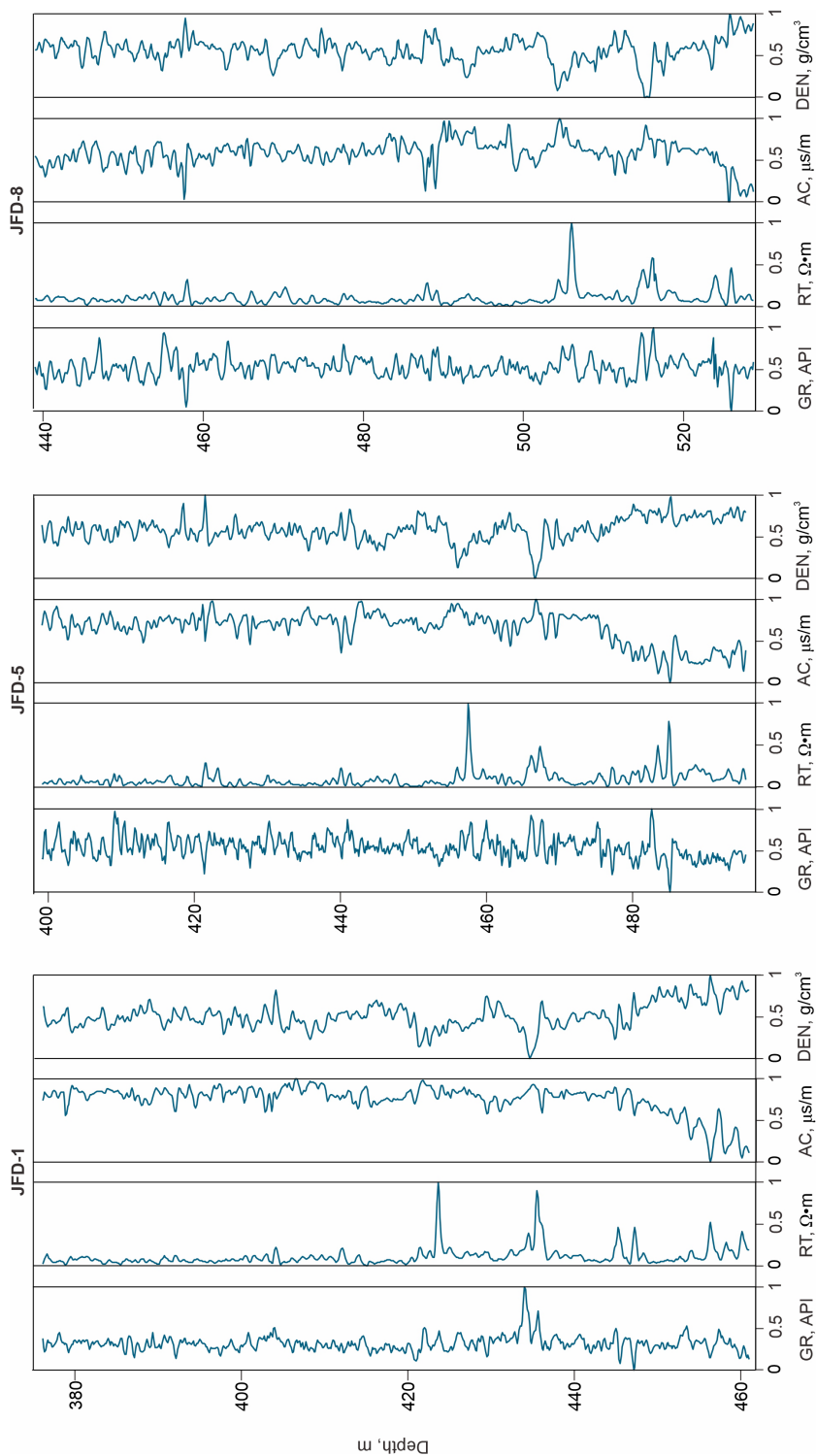
The core, thin section, and TOC were used to determine the development of the four lithologies (oil shale, mudstone, siltstone, and dolomite) in each well. The specific petrological characteristics of the lithologies are displayed in Table 1. The table also analyzes the logging response characteristics: for instance, oil shale is distinguished by high RT, AC, and GR response characteristics, alongside low DEN response features.

To mitigate the impact of differences between logging data scales on model training, the logging data are normalized during preprocessing. Figure 7 shows the normalized logging curves, while Figure 8 illustrates the distribution characteristics of the logging curves corresponding to the four lithologies. In Figure 8, the diagonal plot depicts the distribution of different lithologies on a single logging parameter, highlighting that all four lithologies have

**Table 1.** Petrological characteristics and logging data of four lithologies in the study area

Lithology	Core	Thin section	Petrological characteristics	GR, API	Logging data RT, $\Omega \cdot m$	AC, $\mu s/m$	DEN, $g/cm^3$
Oil shale			Gray black, horizontal bedding, high clay content, darker organic-rich clay laminates are developed, high TOC	98–260 146	5–49 10	380–482 432	1.94–2.39 2.24
Mudstone			Gray or dark gray massive, mainly composed of clay, contains small amounts of quartz and feldspar, low to medium TOC	71–205 140	4–40 7	255–476 416	2.22–2.58 2.36
Siltstone			Grayish white massive, mainly composed of quartz and feldspar, contains a small amount of bioclasts, low TOC	59–174 125	5–35 10	237–442 340	2.27–2.68 2.48
Dolomite			Grayish white massive, mainly composed of micrite dolomite, contains a small amount of clay, low TOC	91–181 128	4–15 6	338–450 398	2.30–2.57 2.42



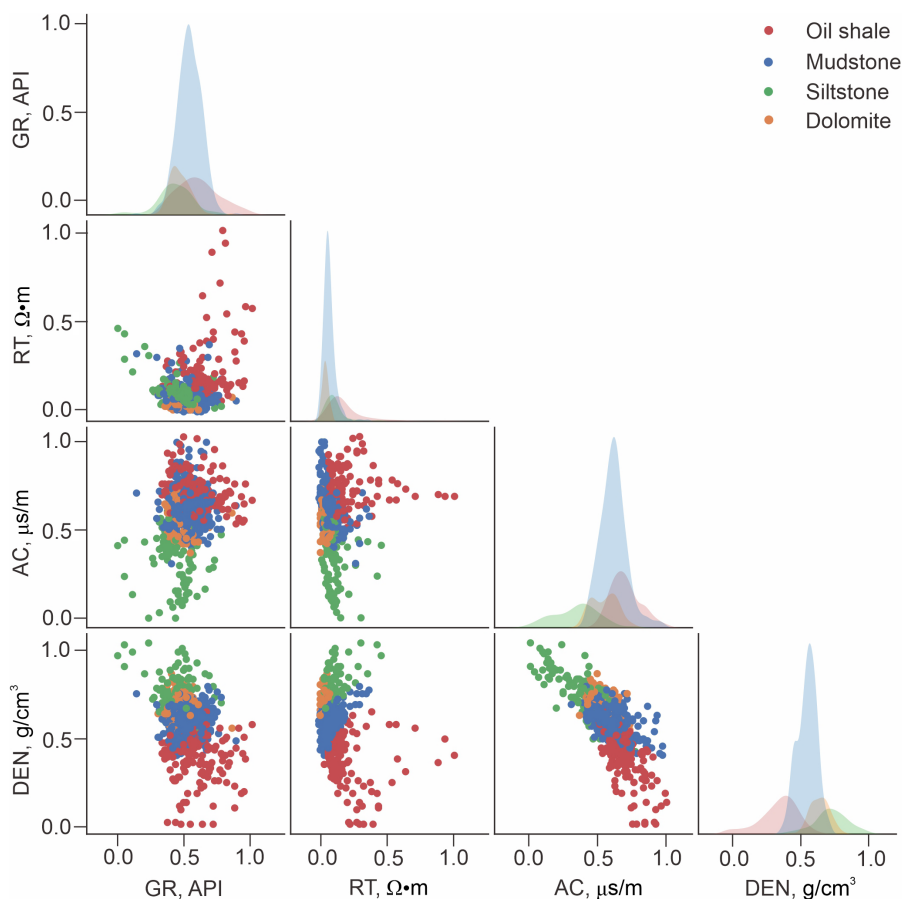


**Fig. 7.** Normalized preprocessed log curves for the JFD-1, JFD-5 and JFD-8 wells. The scatter plot under the diagonal line is the distribution of different lithologies based on the pairwise logging data.

overlapping areas on the logs, which is caused by the non-homogeneity of the fine-grained sedimentary rocks. Additionally, the lithological distribution is characterized in Table 2.

**Table 2.** Lithological distribution

Lithology	JFD-1	JFD-5	JFD-8
Oil shale	204	203	141
Mudstone	293	375	398
Siltstone	108	115	91
Dolomite	75	79	89



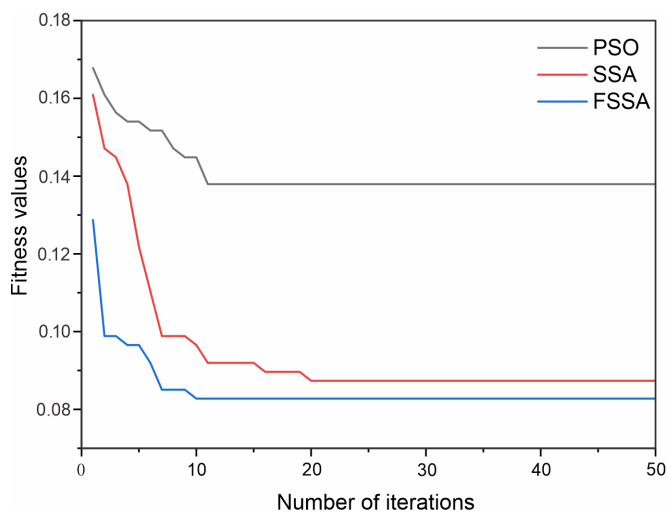
**Fig. 8.** Distribution of four types of lithologies on four logging parameters in the test well JFD-8.

### 4.2. Comparison of prediction results

To assess the performance of FSSA-HKELM, samples from a blind well (well JFD-8) were selected as the test set for lithology prediction. The classification performance of the base classifier HKELM and the parameter search speed and accuracy of FSSA were tested. Five models, namely ELM, KELM, HKELM, PSO-HKELM, and SSA-HKELM, were compared. The parameters to be optimized in HKELM were  $C$ ,  $C_0$ ,  $b$ ,  $\sigma$ , and  $\omega_0$ . To fairly evaluate the performance of the three optimization algorithms, the size of the population was established at 20, and the iteration count was capped at 50. The JFD-1 and JFD-5 wells served as training data, fed into the model to identify the optimal parameters, and the fitness value was the lithological classification error rate of the validation set ( $1 - \text{accuracy}$ ). The detailed parameters of each algorithm are shown in Table 3. Concurrently, the algorithms were executed in Matlab 2020b.

**Table 3.** Model parameters setting. Abbreviations: ST – safety thresholds, PD – producers, SD – vigilantes

Algorithm	Parameters
ELM	Hidden layer = 100, kernel = sigmod ( $1/(1+e^{-x})$ )
KELM	Kernel = RBF, $C = 10$ , $\sigma = 10$
HKELM	$C = 10$ , $C_0 = 5$ , $b = 5$ , $\sigma = 10$ , $\omega_0 = 0.5$
PSO	$\omega_0 = 0.89$ , $c_1 = 2$ , $c_2 = 2$
SSA	ST = 80%, PD = 20%, SD = 20%
FSSA	ST = 80%, PD = 20%, SD = 20%



**Fig. 9.** Comparison of the adaptation and convergence speed of optimization algorithms.

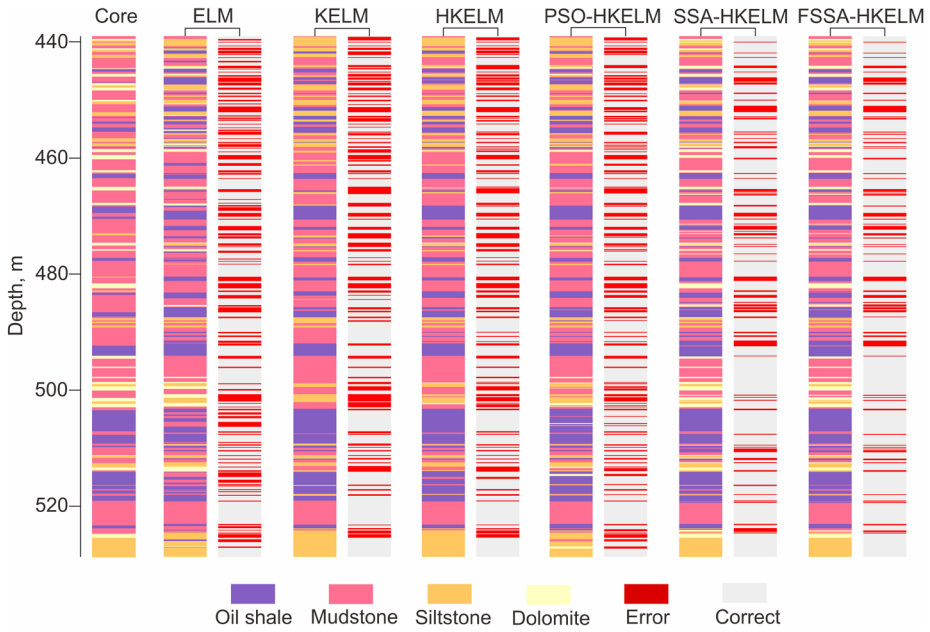
Figure 9 demonstrates that the three SI algorithms decrease in fitness and eventually converge as the number of iterations increases. Five optimal parameters for HKELM are obtained. FSSA is clearly superior to SSA and PSO, as it has the smallest fitness value (0.0828) and the fastest iteration convergence speed, converging by the 10th iteration. SSA parameter optimization is the second most effective, with a final fitness of 0.0874 and convergence starting at the 20th iteration. Although PSO (11th iteration) converges faster than SSA, it has the worst search accuracy (0.1379) among the three algorithms.

The vertical distribution of lithologies predicted by each of the six lithology identification models is shown in Figure 10. The lithology prediction results of FSSA-HKELM are better than those of the other five algorithms.

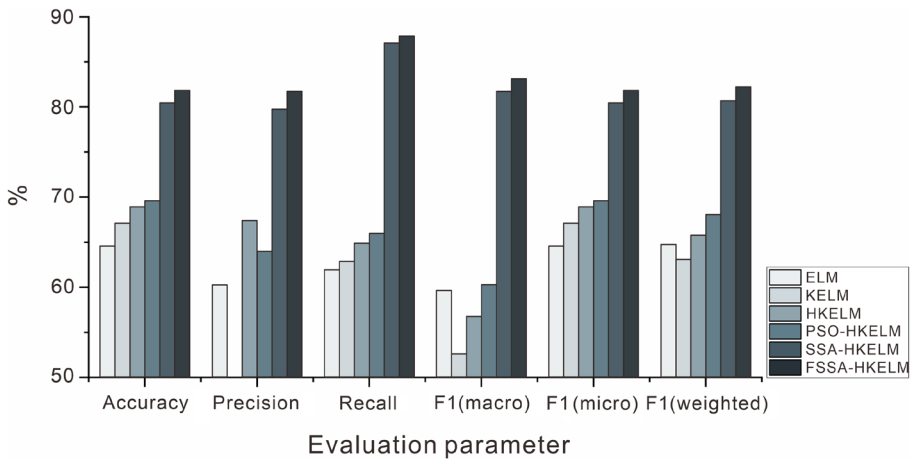
Table 4 and Figure 11 show the comprehensive evaluation of lithology prediction results for the six algorithms applied to well JFD-8. Among the three base learners, HKELM outperforms KELM and ELM in terms of accuracy, precision, recall, and F1 (both micro and weighted). Therefore, after optimizing HKELM in FSSA, the proposed model outperforms the other five models across all six evaluation metrics. This method offers considerable advantages in lithology prediction within the study area, making it particularly useful for oil shale exploration in K<sub>2</sub>qn in the Southeastern Uplift Zone of the Songliao Basin.

**Table 4.** Comparison of evaluation parameters of the six lithology logging identification models

Models	Accuracy	Precision	Recall	F1 (macro)	F1 (micro)	F1 (weighted)
ELM	64.53%	60.21%	61.88%	59.59%	64.53%	64.69%
KELM	67.04%	47.51%	62.83%	52.54%	67.04%	63.03%
HKELM	68.85%	67.36%	64.83%	56.71%	68.85%	65.73%
PSO-HKELM	69.54%	63.93%	65.92%	60.23%	69.54%	68.01%
SSA-HKELM	80.39%	79.71%	87.04%	81.66%	80.39%	80.63%
FSSA-HKELM	81.78%	81.65%	87.78%	83.08%	81.78%	82.16%



**Fig. 10.** Lithology prediction results of the six lithology logging identification models in the test well JFD-8. Beneath each algorithm's name, there are two columns: the left column shows the predicted lithology sequence, and the right column displays the sequence indicating whether the lithology predictions are correct or not.



**Fig. 11.** Accuracy, precision, recall and F1 of the six classification algorithms.

## 5. Conclusions

In this study, a linear fitting based on the  $\omega$  and TOC data from the wells JFD-1, JFD-5, and JFD-8 established that mud shale with TOC greater than 5.25 wt% is defined as oil shale. Using the measured TOC and logging data from these three wells, combined with the improved  $\Delta\log$  method, TOC self-prediction was utilized to define the oil shales in these wells. The core, thin section, and TOC were used to determine the vertical distribution of the four lithologies (oil shale, mudstone, siltstone, and dolomite) in each well. After obtaining accurate lithology samples, an optimization algorithm and classifier were integrated to establish a model of lithology identification for the fine-grained sedimentary rocks of the Qingshankou Formation in the Southeastern Uplift Zone of the Songliao Basin, based on FSSA-HKELM, leading to the following insights:

1. The base classifier HKELM, integrating RBF and the polynomial function, has enhanced model generalization and learning capability, outperforming KELM and ELM in comprehensive evaluation metrics.

2. FSSA, leveraging the firefly perturbation strategy from the SSA algorithm, has improved algorithm performance. Comparing the fitness curves (error rate in lithology classification) during parameter optimization for HKELM among PSO, SSA, and FSSA, it was found that FSSA has a faster iteration speed, converging by the 10th iteration with the lowest fitness, thus indicating higher search precision.

3. Comparing FSSA-HKELM with five other models (ELM, KELM, HKELM, PSO-HKELM, SSA-HKELM) in the blind well (JFD-8), it is evident that the proposed algorithm significantly outperforms the others in classification evaluation metrics, achieving an accuracy rate of 81.78%, and demonstrating a strong capability in lithology identification.

## Acknowledgments

This study was financially supported by the Program of Natural Science Foundation of Jilin Province (No. 20230101081JC), the Program for Jilin University Science and Technology Innovative Research Team (2021TD-05), and the Program of China Geological Survey (grants No. DD2019139-YQ19JJ04 and No. DD20189606). The publication costs of this article were partially covered by the Estonian Academy of Sciences.

## References

1. Aplin, A. C., Macquaker, J. H. S. Mudstone diversity: origin and implications for source, seal, and reservoir properties in petroleum systems. *AAPG Bull.*, 2011, **95**(12), 2031–2059. <https://doi.org/10.1306/03281110162>
2. Jiang, Z. X., Liang, C., Wu, J., Zhang, J. G., Zhang, W. Z., Wang, Y. S., Liu, H. M., Chen, X. Several issues in sedimentological studies on hydrocarbon-bearing fine-grained sedimentary rocks. *Acta Pet. Sin.*, 2013, **34**(6), 1031–1039. <https://doi.org/10.7623/syxb201306001>
3. Zou, C. N., Zhu, R. K., Chen, Z. Q., Ogg, J. G., Wu, S. T., Dong, D. Z., Qiu, Z., Wang, Y. M., Wang, L., Lin, S. H., Cui, J. W., Su, L., Yang, Z. Organic-matter-rich shales of China. *Earth-Sci. Rev.*, 2019, **189**, 51–78. <https://doi.org/10.1016/j.earscirev.2018.12.002>
4. He, J. H., Ding, W. L., Jiang, Z. X., Li, A., Wang, R. Y., Sun, Y. X. Logging identification and characteristic analysis of the lacustrine organic-rich shale lithofacies: a case study from the Es<sub>3</sub><sup>1</sup> shale in the Jiyang Depression, Bohai Bay Basin, Eastern China. *J. Pet. Sci. Eng.*, 2016, **145**, 238–255. <https://doi.org/10.1016/j.petrol.2016.05.017>
5. Bartetzko, A., Delius, H., Pechnig, R. Effect of compositional and structural variations on log responses of igneous and metamorphic rocks. I: mafic rocks. *Geol. Soc. Spec. Publ.*, 2005, **240**, 255–278. <https://doi.org/10.1144/gsl.sp.2005.240.01.19>
6. Zhao, X. Z., Pu, X. G., Han, W. Z., Zhou, L. H., Shi, Z. N., Chen, S. Y., Xiao D. Q. A new method for lithology identification of fine grained deposits and reservoir sweet spot analysis: a case study of Kong 2 Member in Cangdong sag, Bohai Bay Basin, China. *Pet. Explor. Dev.*, 2017, **44**(4), 524–534. [https://doi.org/10.1016/S1876-3804\(17\)30061-7](https://doi.org/10.1016/S1876-3804(17)30061-7)
7. Gregori, E. B., Zhang, J. J., Galván-Fernández, C., Fernández-Navarro, F. Learner support in MOOCs: identifying variables linked to completion. *Comput. Educ.*, 2018, **122**, 153–168. <https://doi.org/10.1016/j.compedu.2018.03.014>
8. Tran, V. L., Nguyen, D.-D. Novel hybrid WOA-GBM model for patch loading resistance prediction of longitudinally stiffened steel plate girders. *Thin-Walled Struct.*, 2022, **177**, 109424. <https://doi.org/10.1016/j.tws.2022.109424>
9. Wang, X., Gao, S., Guo, Y. B., Zhou, S. Y., Duan, Y. H., Wu, D. Q. A combined prediction model for hog futures prices based on WOA-LightGBM-CEEMDAN. *Complexity*, 2022. <https://doi.org/10.1155/2022/3216036>
10. Dev, V. A., Eden, M. R. Formation lithology classification using scalable gradient boosted decision trees. *Comput. Chem. Eng.*, 2019, **128**, 392–404. <https://doi.org/10.1016/j.compchemeng.2019.06.001>
11. Xie, Y. X., Zhu, C. Y., Zhou, W., Li, Z. D., Liu, X., Tu, M. Evaluation of machine learning methods for formation lithology identification: a comparison of tuning processes and model performances. *J. Pet. Sci. Eng.*, 2018, **160**, 182–193. <https://doi.org/10.1016/j.petrol.2017.10.028>
12. Ye, Z. H., Guo, S. Y., Chen, D., Wang, H., Li, S. D. Drilling formation perception

- by supervised learning: model evaluation and parameter analysis. *J. Nat. Gas Sci. Eng.*, 2021, **90**, 103923. <https://doi.org/10.1016/j.jngse.2021.103923>
13. Zhang, J. L., He, Y. B., Zhang, Y., Li, W. F., Zhang, J. J. Well-logging-based lithology classification using machine learning methods for high-quality reservoir identification: a case study of Baikouquan Formation in Mahu area of Junggar Basin, NW China. *Energies*, 2022, **15(10)**, 3675. <https://doi.org/10.3390/en15103675>
  14. Jadoon, Q. K., Roberts, E. M., Henderson, B., Blenkinsop, T. G., Wüst, R. A. J., Mtelega, C. Lithological and facies analysis of the Roseneath and Murteree shales, Cooper Basin, Australia. *J. Nat. Gas Sci. Eng.*, 2016, **37**, 138–168. <https://doi.org/10.1016/j.jngse.2016.10.047>
  15. Woo, J., Rhee, C. W., Jeung, T. J., Jang, S. Application of multi-resolution graph-based clustering for electrofacies prediction: a case study from the Horn River Shale, British Columbia, Canada. *Geosci. J.*, 2020, **24(5)**, 507–518. <https://doi.org/10.1007/s12303-019-0046-3>
  16. Huang, G. B., Zhu, Q. Y., Siew, C. K. Extreme learning machine: theory and applications. *Neurocomputing*, 2006, **70(1–3)**, 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
  17. Huang, G. B., Zhou, H. M., Ding, X. J., Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. B Cybern.*, 2012, **42(2)**, 513–529. <https://doi.org/10.1109/TSMCB.2011.2168604>
  18. Wang, Z. G., Chen, H. Y., Wang, M., Zhang, X., Dou, Y. H. Solid particle erosion prediction in elbows based on machine learning and swarm intelligence algorithm. *J. Pet. Sci. Eng.*, 2022, **218**, 111042. <https://doi.org/10.1016/j.petrol.2022.111042>
  19. Yan, S. C., Wu, L. F., Fan, J. L., Zhang, F. C., Zou, Y. F., Wu, Y. A novel hybrid WOA-XGB model for estimating daily reference evapotranspiration using local and external meteorological data: applications in arid and humid regions of China. *Agric. Water Manag.*, 2021, **244(1)**, 106594. <https://doi.org/10.1016/j.agwat.2020.106594>
  20. Yu, C. X., Koopialipoor, M., Murlidhar, B. R., Mohammed, A. S., Armaghani, D. J., Mohamad, E. T., Wang, Z. L. Optimal ELM–Harris Hawks optimization and ELM–Grasshopper optimization models to forecast peak particle velocity resulting from mine blasting. *Nat. Resour. Res.*, 2021, **30(4)**, 2647–2662. <https://doi.org/10.1007/s11053-021-09826-4>
  21. Sun, Z. X., Jiang, B. S., Li, X. L., Li, J. K., Xiao, K. A data-driven approach for lithology identification based on parameter-optimized ensemble learning. *Energies*, 2020, **13(15)**, 3903. <https://doi.org/10.3390/en13153903>
  22. Gu, Y. F., Zhang, D. Y., Lin, Y. B., Ruan, J. F., Bao, Z. D. Data-driven lithology prediction for tight sandstone reservoirs based on new ensemble learning of conventional logs: a demonstration of a Yanchang member, Ordos Basin. *J. Pet. Sci. Eng.*, 2021, **207**, 109292. <https://doi.org/10.1016/j.petrol.2021.109292>
  23. Tang, J., Liu, G., Pan, Q. T. A review on representative swarm intelligence algorithms for solving optimization problems: applications and trends. *IEEE-CAA J. Autom. Sin.*, 2021, **8(10)**, 1627–1643. <https://doi.org/10.1109/JAS.2021.1004129>



24. Dorigo, M., Maniezzo, V., Colorni, A. Ant system: optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybern. B Cybern.*, 1991, **26**(1), 29–41. <https://doi.org/10.1109/3477.484436>
25. Kennedy, J., Eberhart, R. Particle swarm optimization. In: *Proceedings of ICNN'95 – International Conference on Neural Networks*, November 27–December 1, 1995, Perth, Australia. IEEE, 2017, 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
26. Mirjalili, S., Mirjalili, S. M., Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.*, 2014, **69**, 46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
27. Mirjalili, S., Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.*, 2016, **95**, 51–67. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
28. Xue, J. K., Shen, B. A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst. Sci. Control Eng.*, 2020, **8**(1), 22–34. <https://doi.org/10.1080/21642583.2019.1708830>
29. Nguyen, T. T., Ngo, T. G., Dao, T. K., Nguyen, T. T. T. Microgrid operations planning based on improving the flying sparrow search algorithm. *Symmetry*, 2022, **14**(1), 168. <https://doi.org/10.3390/sym14010168>
30. Qiao, L., Jia, Z. N., Cui, Y., Xiao, K., Su, H. N. Shear sonic prediction based on DELM optimized by improved sparrow search algorithm. *Appl. Sci.*, 2022, **12**(16), 8260. <https://doi.org/10.3390/app12168260>
31. Feng, Z. Q., Jia, C. Z., Xie, X. N., Zhang, S., Feng, Z. H., Cross, T. A. Tectono-stratigraphic units and stratigraphic sequences of the nonmarine Songliao basin. *Basin Res.*, 2010, **22**(1), 79–95. <https://doi.org/10.1111/j.1365-2117.2009.00445.x>
32. Xu, J. J., Bechtel, A., Sachsenhofer, R. F., Liu, Z. J., Gratzer, R., Meng, Q. T., Song, Y. High resolution geochemical analysis of organic matter accumulation in the Qingshankou Formation, Upper Cretaceous, Songliao Basin (NE China). *Int. J. Coal Geol.*, 2015, **141–142**, 23–32. <https://doi.org/10.1016/j.coal.2015.03.003>
33. Gao, R. Q., Cai, X. Y. *Formation Conditions and the Distribution of Oil and Gas Fields in Songliao Basin*. Petroleum Industry Press, Beijing, 1997.
34. Chen, C. Non-marine setting of petroleum in the Songliao basin of Northeast China. *J. Pet. Geol.*, 1980, **2**(3), 233–264. <https://doi.org/10.1111/j.1747-5457.1980.tb00706.x>
35. Bechtel, A., Jia, J. L., Strobl, S. A. I., Sachsenhofer, R. F., Liu, Z. J., Gratzer, R., Püttmann, W. Palaeoenvironmental conditions during deposition of the Upper Cretaceous oil shale sequences in the Songliao Basin (NE China): implications from geochemical analysis. *Org. Geochem.*, 2012, **46**, 76–95. <https://doi.org/10.1016/j.orggeochem.2012.02.003>
36. Jia, J. L., Bechtel, A., Liu, Z. J., Strobl, S. A. I., Sun, P. C., Sachsenhofer, R. F. Oil shale formation in the Upper Cretaceous Nenjiang Formation of the Songliao Basin (NE China): implications from organic and inorganic geochemical analyses. *Int. J. Coal Geol.*, 2013, **113**, 11–26. <https://doi.org/10.1016/j.coal.2013.03.004>
37. Jackson, L. L., Roof, S. R. Determination of the forms of carbon in geologic materials. *Geostand. Geoanalytical Res.*, 1992, **16**(2), 317–323. <https://doi.org/10.1111/j.1751-908X.1992.tb00495.x>

38. Ding, S. F., Zhao, H., Zhang, Y. N., Xu, X. Z., Nie, R. Extreme learning machine: algorithm, theory and applications. *Artif. Intell. Rev.*, 2015, **44**(1), 103–115. <https://doi.org/10.1007/s10462-013-9405-z>
39. Tian, Z. D., Li, S. J., Wang, Y. H., Wang, X. D. Wind power prediction method based on hybrid kernel function support vector machine. *Wind Eng.*, 2018, **42**(3), 252–264. <https://doi.org/10.1177/0309524X17737337>
40. Ahuja, B., Vishwakarma, V. P. Deterministic multikernel extreme learning machine with fuzzy feature extraction for pattern classification. *Multimed. Tools Appl.*, 2021, **80**(21), 32423–32447. <https://doi.org/10.1007/s11042-021-11097-3>
41. Lv, L., Wang, W. H., Zhang, Z. Y., Liu, X. G. A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. *Knowl.-Based Syst.*, 2020, **195**, 105648. <https://doi.org/10.1016/j.knosys.2020.105648>
42. Gandomi, A. H., Yang, X. S., Talatahari, S., Alavi, A. H. Firefly algorithm with chaos. *Commun. Nonlinear Sci.*, 2013, **18**(1), 89–98. <https://doi.org/10.1016/j.cnsns.2012.06.009>
43. Hutton, A. C. Petrographic classification of oil shales. *Int. J. Coal Geol.*, 1987, **8**(3), 203–231. [https://doi.org/10.1016/0166-5162\(87\)90032-2](https://doi.org/10.1016/0166-5162(87)90032-2)
44. Liu, Z. J., Yang, H. L., Dong, Q. S., Zhu, J. W., Guo, W., Ye, Q. S., Liu, R., Meng, Q. T., Zhang, H. L., Gan, S. C. *Oil Shale in China*. Petroleum Industry Press, Beijing, 2009.
45. Zhao, W. Z., Hu, S. Y., Hou, L. H. Connotation and strategic role of in-situ conversion processing of shale oil underground in the onshore China. *Pet. Explor. Dev.*, 2018, **45**(4), 563–572. [https://doi.org/10.1016/S1876-3804\(18\)30063-6](https://doi.org/10.1016/S1876-3804(18)30063-6)
46. Passey, Q. R., Creaney, S., Kulla, J. B., Moretti, F. J., Stroud, J. D. A practical model for organic richness from porosity and resistivity logs. *AAPG Bull.*, 1990, **74**(12), 1777–1794. <https://doi.org/10.1306/0C9B25C9-1710-11D7-8645000102C1865D>
47. Hu, H. T., Lu, S. F., Liu, C., Wang, W. M., Wang, M., Li, J. J., Sang, J. H. Models for calculating organic carbon content from logging information: comparison and analysis. *Acta Sedimentol. Sin.*, 2011, **29**(6), 1199–1205. <https://doi.org/10.14027/j.cnki.cjxb.2011.06.012>