

Reservoir fluid type identification method based on deep learning: A case study of the Chang 1 Formation in the Jiyuan oilfield of the Ordos basin, China

Wen-bo Li, Xiao-ye Wang, Lei He, Zhen-kai Zhang, Zeng-lin Hong, Ling-yi Liu, Dong-tao Li

Citation: Wen-bo Li, Xiao-ye Wang, Lei He, Zhen-kai Zhang, Zeng-lin Hong, Ling-yi Liu, Dong-tao Li, 2026. Reservoir fluid type identification method based on deep learning: A case study of the Chang 1 Formation in the Jiyuan oilfield of the Ordos basin, China, China Geology, 9, 60-74. doi: 10.31035/cg2025010.

View online: https://doi.org/10.31035/cg2025010

Related articles that may interest you

Evaluation of reservoir environment by chemical properties of reservoir waterA case study of Chang 6 reservoir in Ansai oilfield, Ordos Basin, China

China Geology. 2023, 6(3), 443 https://doi.org/10.31035/cg2022042

Soil geochemical prospecting prediction method based on deep convolutional neural networks-Taking Daqiao Gold Deposit in Gansu Province, China as an example

China Geology. 2022, 5(1), 71 https://doi.org/10.31035/cg2021044

Geological characteristics of unconventional tight oil reservoir (10^9 t): A case study of Upper Cretaceous Qingshankou Formation, northern Songliao Basin, NE China

China Geology. 2024, 7(1), 51 https://doi.org/10.31035/cg2022072

Assessment of natural gas hydrate reservoirs at Site GMGS3-W19 in the Shenhu area, South China Sea based on various well logs

China Geology. 2022, 5(3), 383 https://doi.org/10.31035/cg2022025

Application of frequency division inversion in the prediction of heterogeneous natural gas hydrates reservoirs in the Shenhu Area, South China Sea

China Geology. 2022, 5(2), 251 https://doi.org/10.31035/cg2021074

Comparative study of different machine learning models in landslide susceptibility assessment: A case study of Conghua District, Guangzhou, China

China Geology. 2024, 7(1), 104 https://doi.org/10.31035/cg2023056

China Geology 9 (2026) 60−74

China Geology

Journal homepage: http://chinageology.cgs.cn
https://www.sciencedirect.com/journal/china-geology



Reservoir fluid type identification method based on deep learning: A case study of the Chang 1 Formation in the Jiyuan oilfield of the Ordos basin, China

Wen-bo Li[a], Xiao-ye Wang[b], Lei He[b], Zhen-kai Zhang[c,*], Zeng-lin Hong[a, c], Ling-yi Liu[d], Dong-tao Li[b]

[a] *School of Earth Sciences and Resources, Chang'an University, Xi'an* 710054, *China*
[b] *The Fifth Oil Production Plant of PetroChina Changqing Oilfield Company, Xi'an* 710000, *China*
[c] *Shaanxi Satellite Application Center for Natural Resources, Xi'an* 710119, *China*
[d] *School of Computer Science, National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Shaanxi Provincial Key Laboratory of Speech Image Information Processing, Northwestern Polytechnical University, Xi'an* 710129, *China*

ARTICLE INFO

Article history:
Received 19 January 2025
Received in revised form 16 June 2025
Accepted 24 June 2025
Available online 21 August 2025

Keywords:
Low-contrast reservoirs
Fluid types
Pore structure
Clay content
LR+NB+GBDT+RF+SVM model
Machine learning
Neural networks
Loss functions
Geophysical well logging
Oil and gas reservoir prediction

ABSTRACT

With the efficient and intelligent development of computer-based big data processing, applying machine learning methods to the processing and interpretation of logging data in the field of geophysical well logging has broad potential for improving production efficiency. Currently, the Jiyuan Oilfield in the Ordos Basin relies mainly on manual reprocessing and interpretation of old well logging data to identify different fluid types in low-contrast reservoirs, guiding subsequent production work. This study uses well logging data from the Chang 1 reservoir, partitioning the dataset based on individual wells for model training and testing. A deep learning model for intelligent reservoir fluid identification was constructed by incorporating the focal loss function. Comparative validations with five other models, including logistic regression (LR), naive Bayes (NB), gradient boosting decision trees (GBDT), random forest (RF), and support vector machine (SVM), show that this model demonstrates superior identification performance and significantly improves the accuracy of identifying oil-bearing fluids. Mutual information analysis reveals the model's differential dependency on various logging parameters for reservoir fluid identification. This model provides important references and a basis for conducting regional studies and revisiting old wells, demonstrating practical value that can be widely applied.

©2026 China Geology Editorial Office.

1. Introduction

In oil exploration and development, one of the fundamental tasks of geophysical logging is the accurate identification of fluid types within the reservoir profile (Peng Z et al., 2016). Traditional reservoir identification methods primarily include rock physics templates and cross-plot techniques based on core or log responses (Li X et al., 2013; Luo SL et al., 2015; Wang DX, 2016; Das B and Chatterjee R, 2018). However, these methods rely on manual interpretation and are heavily influenced by human factors. They have low identification rates, especially in complex low-porosity, low-permeability reservoirs and low-resistivity oil layers, making it difficult to meet production requirements. Currently, data-driven machine learning methods, owing to their excellent ability to explore nonlinear relationships between well logging responses, represent a promising breakthrough in addressing this issue and have gained favor among many scholars. Cheng C et al. (2022) compared the fluid property identification results of decision trees, neural networks, eXtreme Gradient Boosting (XGBoost), and other machine learning algorithms with those of traditional identification methods for low-resistivity oil reservoirs. The machine learning algorithm results outperformed those of the graphical methods, with EXtreme Gradient Boosting showing the best performance. Bai Z et al. (2022) constructed support vector machine (SVM) classification models for reservoir fluid identification and support vector regression (SVR) models for reservoir parameter prediction. They demonstrated the feasibility of using SVM methods for logging interpretation and evaluation of low-contrast oil reservoirs, providing important references for old well reviews. Zhou

First author: *E-mail address*: 2021027014@chd.edu.cn (Wen-bo Li).

* Corresponding author: *E-mail address*: zhangzhenkai1988@foxmail.com (Zhen-kai Zhang).

Literary editor: Li-qiong Jia
doi:10.31035/cg2025010
2096-5192/© 2026 China Geology Editorial Office.

XM et al. (2023) used an SVM for fluid identification in thin reservoirs, achieving an identification accuracy of 85.7% by introducing kernel functions and establishing a thin reservoir prediction model. These traditional machine learning methods have strong interpretability and fast training speed but relatively simple model structures and few parameters, limiting their ability to characterize complex and heterogeneous subsurface environments.

Unlike traditional neural networks or other machine learning methods, deep learning methods exhibit stronger learning capabilities and more stable performance (Mohaghegh S et al., 1995, 1996; Huang ZH et al., 1996; Bhatt A and Helle HB, 1999; Helle HB et al., 2002; El-M Shokir EM, 2006; Maiti S et al., 2007; Karimpouli S et al., 2010; Nooruddin HA and Hossain ME, 2011; Alizadeh B et al., 2012; Tahmasebi P and Hezarkhani A, 2012; Anifowose FA et al., 2013; Bagheripour P, 2014; Baziar S et al., 2014). By using networks with complex structures or multiple nonlinear transformations to fit nonlinear mapping relationships from higher dimensions between multiple variables, deep learning is more suitable for exploring the relationship between multisource well logging data and reservoir fluid information in complex and heterogeneous subsurface environments. Jaikla C et al. (2019) combined bidirectional recurrent neural networks with a convolutional autoencoder to propose a FaciesNet model for reservoir and nonreservoir partitioning, achieving satisfactory results. Tan MJ et al. (2020) built a classification committee machine composed of a backpropagation neural network, a probabilistic neural network, and a decision tree classifier. The model achieved fluid identification within dense sandstone reservoirs through a voting-based classification result. He M et al. (2020) emphasized the importance of addressing the issue of imbalanced samples in predictive models. They used a combination of MAHAKIL (oversampling based on the theory of inheritance and the Mahalanobis distance) and deep learning to model well logging data for lithology and fluid identification in tight sandstone gas reservoirs. Zhou XQ et al. (2021), considering the high cost of imaging well logging and the significant influence of adjacent strata on conventional well logging curve information, constructed a bidirectional long short-term memory (Bi-LSTM) neural network, which achieved good predictive results for three types of reservoirs. Yang WB et al. (2023) developed a three-fusion algorithm that outperformed long short-term memory (LSTM) in capturing long-term dependencies in sequence data, ultimately classifying two labels: Oil reservoirs and nonoil reservoirs. However, in these studies, the training and validation datasets were mostly not divided by individual wells to ensure the ideal classification performance of the model.

Based on the limitations of previous studies, this paper constructs a deep neural network to represent the correlations between reservoir fluid types and multiple logging curves. The dataset is divided by individual wells for training the model, and the focal loss function is employed to address class imbalance issues in real-world applications. The effectiveness of the proposed reservoir fluid type recognition model is validated using real logging data from oil fields. The aim is to explore a feasible intelligent reservoir fluid type recognition model that addresses the shortcomings of manual logging interpretation and provides valuable reference results for the interpretation of geophysical data.

2. Regional geological setting

The Ordos Basin, located in the central part of China, stands as the country's second-largest inland sedimentary basin, spanning the provinces of Shaanxi, Gansu, Ningxia, Inner Mongolia, and Shanxi. This rectangular depression basin covers an area of approximately 250000 km^2 and is oriented in a north–south direction. Positioned at the western margin of the North China Block in terms of tectonic setting, the overall structural morphology of the basin exhibits a large, asymmetric north–south-trending rectangular synclinal structure characterized by a broad and gentle eastern flank and a steep and narrow western flank. Further structural division identifies six primary tectonic units within the basin, including the Yimeng uplift, Weibei uplift, Western thrust belt, Jinxi fault-fold belt, Tianhuan depression, and Yishan slope. Among these, the Yishan slope serves as the primary structural unit for the accumulation of oil and gas, with over 90% of the discovered reserves distributed in this unit. Despite undergoing multiple tectonic movements, the basin has predominantly experienced overall uplift and subsidence. Over geological time, the basin has accumulated multiple sets of stratigraphic sequences, dating back to the Paleozoic era, harboring abundant oil and gas resources. The Jiyuan Oilfield is located in the central-western part of the Ordos Basin (Fig. 1). The primary oil and gas zones are located in the vast and harsh wilderness, far from urban centers, with limited transportation access. This area is considered remote and economically disadvantaged, with challenging conditions for both production and living. The oil and gas reservoirs in this region exhibit the typical "three lows" characteristics, namely, low permeability, low pressure, and low abundance. Wells in the area generally have low or no natural productivity. In recent years, during the process of regional studies and reevaluations of old wells, several challenges have emerged. These include the strong concealment of reservoirs, making them difficult to identify, unclear patterns of reservoir enrichment, and low efficiency in the secondary interpretation of well logging data. Resolutions to these challenges are urgently needed.

The study area in the Triassic Yanchang Formation of the Mesozoic era is composed of alternating layers of dark gray, medium- to fine-grained sandstone and dark mudstone. Occasionally, thin coal layers and tuffaceous mudstone layers are interbedded. Based on sedimentary cycles, lithological compositions, electrical properties, marker beds, and

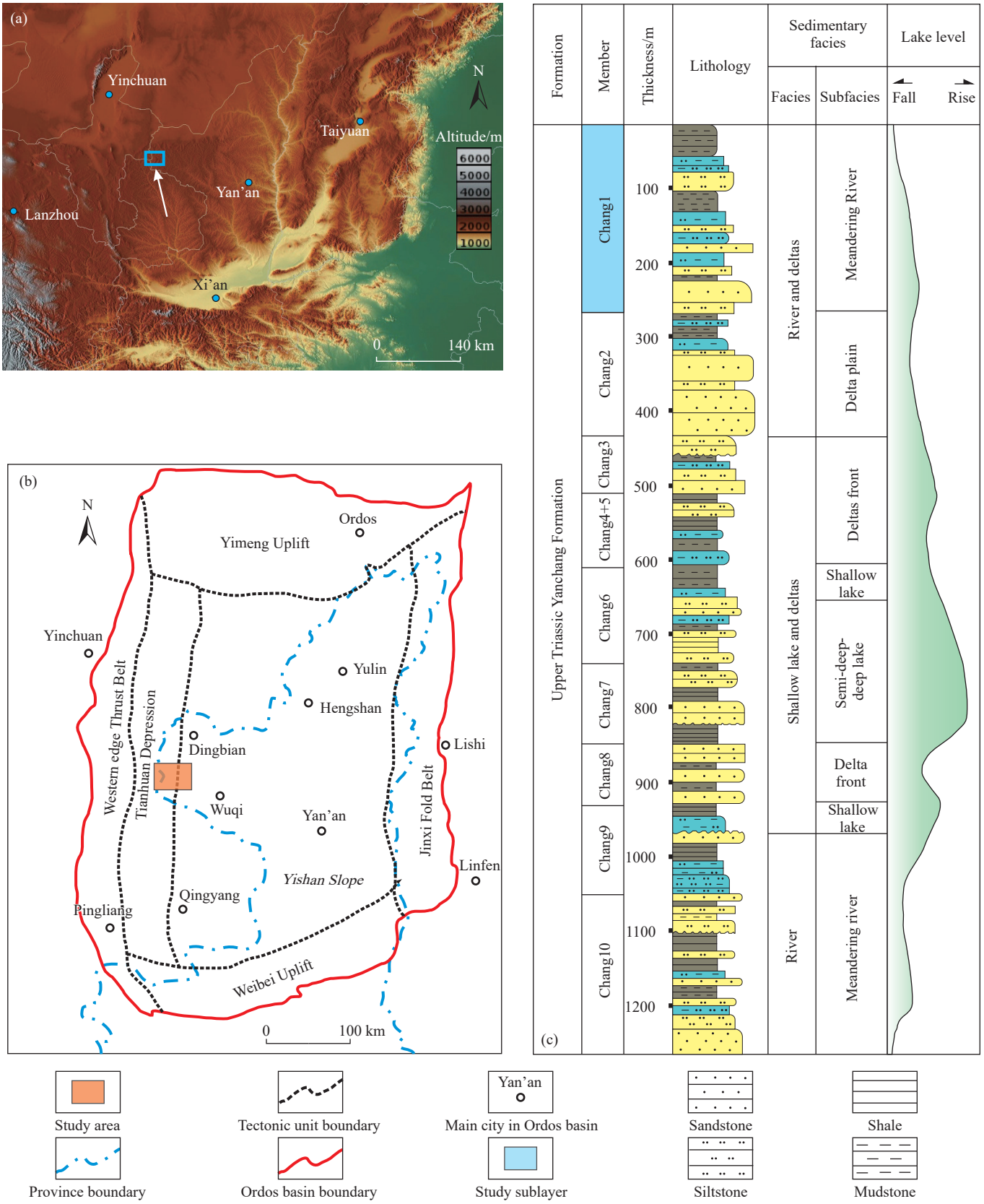


Fig. 1. a–Location of the Ordos Basin; b–substructural units and location map of the Jiyuan area in the Ordos Basin; c–comprehensive stratigraphic column (modified from He TP et al., 2024).

thickness methods, the Triassic Yanchang Formation is vertically divided into the following units from bottom to top: Chang 10, Chang 9, Chang 8, Chang 7, Chang 6, Chang 4+5, Chang 3, Chang 2, and Chang 1. Chang 1 is further subdivided into Chang 1^4, Chang 1^3, Chang 1^2, and Chang 1^1 (Table 1). The well logging data studied in this paper are

Table 1. Chang 1 stratigraphic table of the Yanchang Formation in the Jiyuan area (after Feng Q, 2018).

Stratum				Thickness/m	Lithological characteristics	Key Bed	Evolution history of the lake basin
System	Formation	Member	Bed				
Triassic	Yanchang	T_3yc_5	Chang 1^1	15–30	Sets of mudstone and siltstone with different thicknesses are interbedded. The lithology is black dark mudstone, argillaceous siltstone and silty fine sandstone. The interlayer is carbonaceous mudstone and coal line.	K9	Gentle depression, lake disappearance
			Chang 1^2	20–35			
			Chang 1^3	25–40			
			Chang 1^4	15–30			

derived from the Chang 1^2 and Chang 1^3 intervals.

3. Basic principle and method

3.1. Deep neural network

Deep learning has various network architectures, with the deep (feedforward) neural network being the most fundamental one, serving as the primary model for establishing reservoir fluid prediction models. A deep neural network refers to a standard feedforward neural network with two or more fully connected layers as hidden layers, often referred to as deep learning. A classical single-layer neural network typically has only one hidden layer and is termed shallow learning. Both the hidden layer and the output layer consist of functional neurons. Fully connected implies that neurons are connected to all neurons in the adjacent layer but not to neurons in the same layer or nonadjacent layers. The output of the previous layer serves as the input for the next layer, which can be represented by Eq. 1 and Eq. 2 (Zhang GY, 2019).

$$y^1 = f\left(W^1 x + b^1\right) \tag{1}$$

$$y^l = f\left(W^l y^{l-1} + b^l\right), l = 2, \cdots, L-1 \tag{2}$$

where x represents the input feature vector, W^l and b^l denote the weight matrix and bias vector for layer l, respectively, f represents the activation function, y^l represents the output of layer l, and L indicates the total number of layers.

For the hidden layers, a rectified linear unit (ReLU) is used as the activation function. The ReLU is a commonly used nonlinear activation function in neural networks that can mitigate the vanishing gradient problem and enhance training efficiency (Nair V and Hinton GE, 2010). The main expression is depicted as Eq. 3.

$$f(x) = max(0, x) \tag{3}$$

where x represents the input and where $f(x)$ represents the output of the ReLU activation function. In other words, when the input x is greater than zero, the ReLU activation function outputs x; when the input x is less than or equal to zero, the ReLU activation function outputs zero.

Dropout is a regularization technique used after each hidden layer to improve the generalization ability of the model. The core idea is to randomly delete a certain proportion of hidden layer neurons during the training process; the weights and deviations of the removed neurons are retained but not updated, and they can work again in the next training iteration (Srivastava N et al., 2014; Wu XM et al.,2018).

3.2. Loss Function

In traditional classification tasks, model optimization is achieved by minimizing the cross-entropy loss function (L_{ori}), enhancing the accuracy of model predictions (Andreieva V and Shvai N, 2021). The main expression is depicted as Eq. 4.

$$L_{ori} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M} y_m log\left(y'_m\right) \tag{4}$$

where the total number of samples in the training set and the number of categories in the oil and gas reservoir formations are represented by M and N, respectively. The actual value for a category and the value predicted by the model for that category are denoted as y_m and y'_m, respectively.

The traditional cross-entropy loss treats all samples equally. However, in classification tasks where class difficulty and class imbalance exist, the model tends to be dominated by easy-to-classify and majority classes. In real stratigraphic formations, different lithological layers often appear in groups, which leads to ambiguous class boundaries and increased classification difficulty.

Focal loss addresses this issue through its core design concept—adaptive balancing between easy and hard samples. It dynamically adjusts the model's focus to "hard samples" (those difficult to classify correctly) and "easy samples" (those easy to classify correctly), thereby mitigating the impact of class imbalance and improving model performance. By introducing a modulating factor (γ) and class weight (α), focal loss reduces the loss contribution of easily classified samples (mudstone layers), compelling the model to pay greater attention to hard-to-classify samples (oil-bearing layers).

Therefore, this paper employs focal loss as the loss function to address the challenges posed by extreme class imbalance in traditional cross-entropy loss (Lin TY et al.,2020). Mathematically, this can be expressed as Eq. 5.

$$FL(p_t) = -a_t(1-p_t)^{\gamma} log(p_t) \tag{5}$$

wherein,

- p_t denotes the predicted probability for the true class of a sample. For a given true class label y, if y =1, then $p_t = p$; if y=0, then $p_t = 1 - p$.
- γ is the focusing parameter (where $\gamma \geqslant 0$), which

controls the weighting between easy and hard samples. A higher value of γ downweights the loss contribution from easy examples more strongly. According to previous studies and parameter tuning, setting γ>1 increases the focus on hard examples. In this study, the authors adopt the default value γ=2.

• a_t is the class weighting factor that dynamically adjusts the weight based on the true class of the sample. This helps mitigate class imbalance by assigning higher importance to underrepresented classes during training.

By adjusting the two hyperparameters, a_t and γ, the attention level of focal loss to samples of different categories and difficulty levels can be controlled. In this way, the model focuses more on learning challenging-to-classify samples, improving training effectiveness in the presence of class imbalance. This enhances the model's robustness and generalization ability, preventing it from being dominated by the loss of many easily classifiable samples. This mechanism makes focal loss more effective than the traditional cross-entropy loss function when dealing with class imbalance and challenging-to-classify samples.

3.3. Optimization algorithm

The stochastic gradient descent (SGD) optimization algorithm is utilized to adjust the parameters of various layers in a neural network, aiming to minimize the value of the loss function (Pan Y et al., 2022). Its core idea involves iteratively updating parameters along the direction of the gradient of the loss function to gradually converge the loss function toward its minimum. First, the backpropagation algorithm is employed to calculate the gradient of each parameter with respect to the loss function. The parameters are subsequently updated in the negative direction of the gradient and multiplied by a factor known as the learning rate, which controls the step size of each update. Finally, these steps are repeated until a stopping condition is met.

Cosine annealing is a method for dynamically adjusting the learning rate during training (Jouhari H et al., 2019), allowing the learning rate to vary systematically throughout the training process to increase model performance. The main expression is depicted as Eq. 6.

$$\eta_t = \eta_{min} + \frac{1}{2}\left(\eta_{max} - \eta_{min}\right)\left(1 + cos\left(\frac{T_{cur}}{T_{max}}\pi\right)\right) \tag{6}$$

As T_{cur} increases, the learning rate linearly decreases from the initial value to the minimum value and then gradually increases back to the maximum value. This learning rate adjustment method aids the model in faster convergence during the initial stages of training while also allowing for more fine-tuned parameter adjustments later on, thereby improving overall model performance.

4. Proposed method

4.1. Dataset generation and feature selection

The research dataset originates from a genuine well logging dataset of the Chang 1 oil layer group in the Triassic Yanchang Formation (T_3y) within the Jiyuan area of the Ordos Basin (Fig. 2).

In contrast to unstructured image data, structured vector data present several advantages during the deep learning modeling process:

(i) The data possess clear and explicit features with high interpretability. Each feature typically corresponds to a dimension of the data, making the meaning and importance of features relatively easy to understand.

(ii) The dataset has relatively low data dimensionality, making it simpler to process. These data can often be directly input into deep neural networks for handling.

(iii) Handling missing data and noise is usually relatively straightforward and can be addressed using methods such as imputation and smoothing.

Therefore, in this study, structured data from well logging curves in ".wis" format are extracted to construct the dataset.

Considering that there are differences in the well logging data of different wells caused by different well logging teams, the well logging parameters common to all wells need to be selected. In addition, to eliminate interference from the physical model, well logging parameters that have not been secondarily interpreted should also be chosen. Therefore, the characteristics of the dataset selected for the model in this paper include acoustic (AC), caliper (CAL), natural gamma ray (GR), true formation resistivity (RT), spontaneous potential (SP), and spontaneous potential amplitude (SPC) data. This dataset comprises data from 9 wells, encompassing a total of 2716 samples with a sampling interval of 0.125 m. Figs. 3–8 show boxplots of the data distributions for different well logging response parameters. Table 2 presents the statistical characteristics of the different well logging response parameters.

The model outputs include three reservoir types: An oil-water layer, a water layer, and a mudstone layer. The dataset labels are derived from detailed manual interpretations. Importantly, in this regional dataset, the mudstone layer accounts for 63%, the water layer accounts for 20%, and the oil-water layer accounts for 17%. Thus, the data distribution is imbalanced (Fig. 9).

In this study, the dataset was of moderate size, and a small number of outliers were manually removed through a denoising process. When dealing with multiple features with different units, to ensure that the scale differences of various features do not adversely affect the model, feature normalization was applied to the selected well logging response parameters (Eq. 7).

$$X_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{7}$$

where X_i represents the standardized data, x_i denotes the actual well logging data, x_{min} represents the minimum value of the actual well logging curve data, and x_{max} represents the maximum value of the actual well logging curve data.

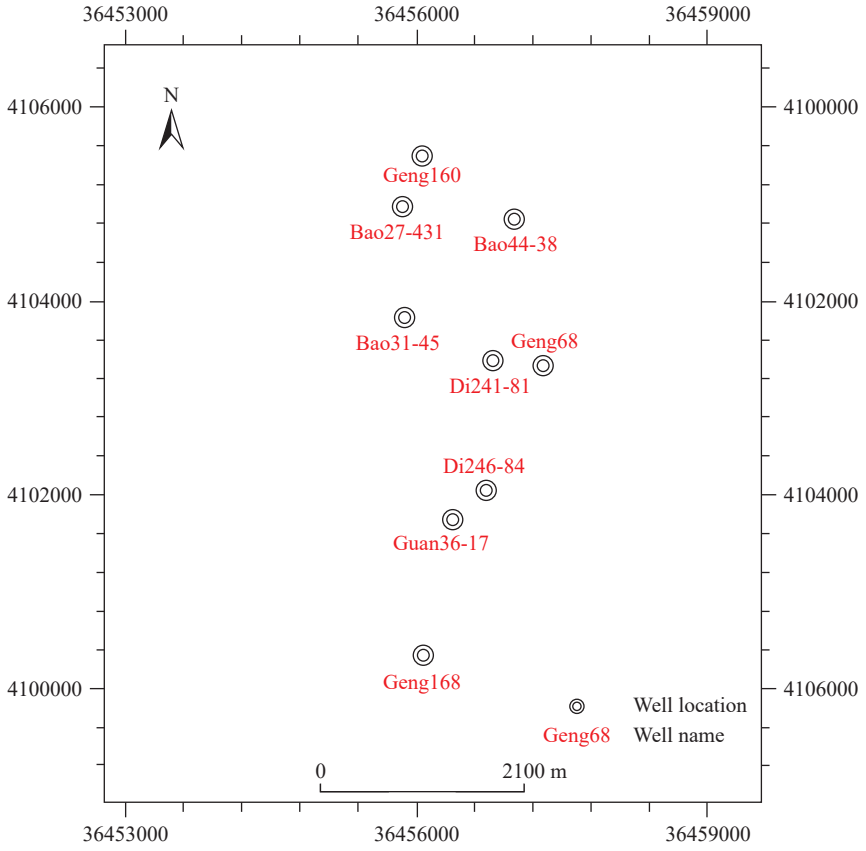


Fig. 2. Well locations in the study area [The horizontal and vertical coordinates in this map refer to the Gauss–Krüger plane rectangular coordinates (unit: meter)].

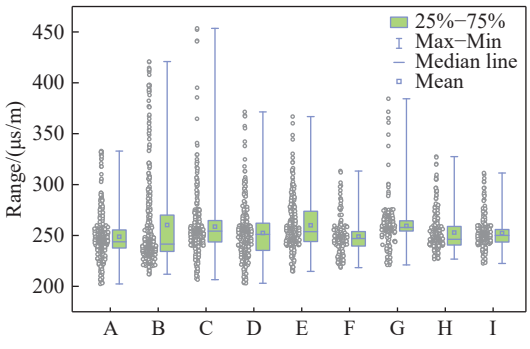


Fig. 3. Box plot of AC /(μs/m).

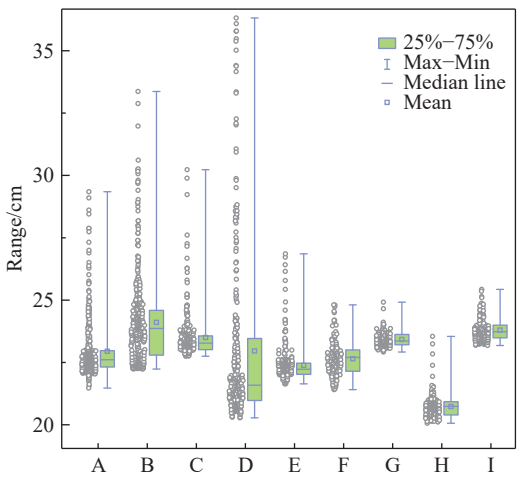


Fig. 4. Box plot of CAL/cm.

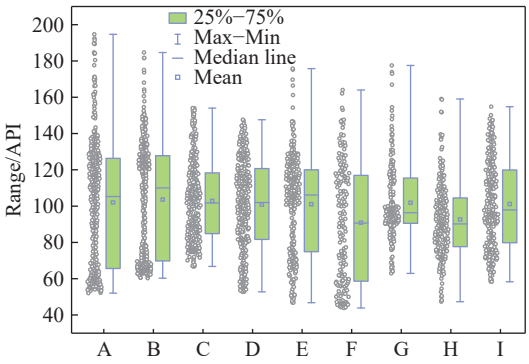


Fig. 5. Box plot of GR/API.

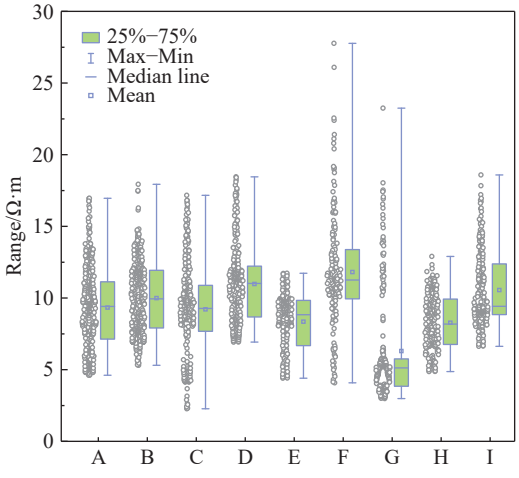


Fig. 6. Box plot of RT/Ω·m.

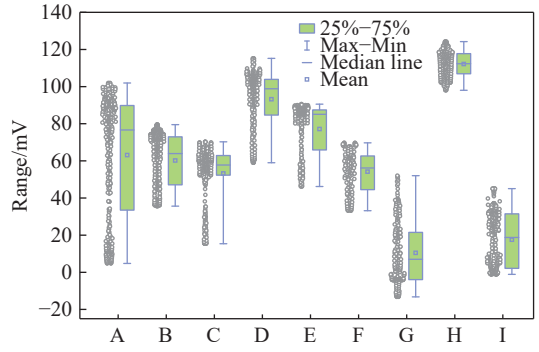


Fig. 7. Box plot of SP/mV.

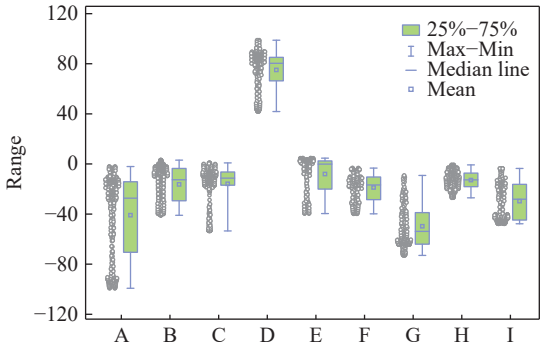


Fig. 8. Box plot of SPC.

Table 2. Statistical characteristics of logging response parameters.

Statistic	Input parameter					
	AC/(μs/m)	CAL/cm	GR/API	RT/Ω·m	SP/mV	SPC
Count	2716	2716	2716	2716	2716	2716
Mean	255.1	23.0	100.5	9.4	61.0	−12.3
Std	28.6	1.8	27.3	3.2	34.0	39.1
Min	202.4	20.1	43.9	2.3	−13.2	−99.2
25%	239.6	22.2	77.9	7.3	36.6	−33.7
50%	249.4	23.0	100.4	9.3	62.9	−14.5
75%	262.6	23.7	120.7	11.3	87.6	−3.2
Max	453.5	36.3	194.7	27.8	124.2	98.8

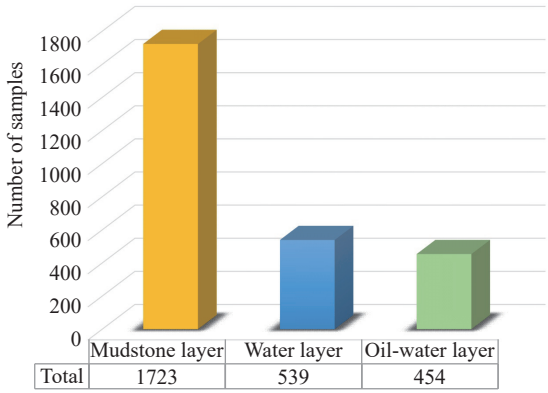


Fig. 9. Reservoir fluid type sample statistics.

4.2. Dataset splitting

To establish a predictive model with strong generalization capabilities, it is common practice to partition the dataset into training and validation data. Various methods can be employed for dataset partitioning, including random splitting based on proportions or grouping by data characteristics. In the research task of this paper, if the dataset is randomly divided in proportion, it can improve the generalization performance and recognition accuracy of the model. However, this approach may lead to the model ignoring the stratigraphic discontinuity between different wells during training, failing to fully demonstrate the model's generalization performance. When the research scope changes or expands, overfitting of the model may become more severe.

Therefore, this study chooses a single well as the unit to divide the training and validation datasets: The total data roughly follow a 7∶3 ratio, with 5 wells selected as the training set, containing a total of 1772 samples, and 4 wells selected as the validation set, containing a total of 944 samples, to train and validate the multiclass reservoir fluid type recognition model.

4.3. Model performance evaluation methods

The confusion matrix is a critical tool for evaluating the performance of classification models, particularly in binary and multiclass tasks. It presents the alignment between the predicted and actual outcomes in a matrix format, providing a comprehensive representation of the model's classification performance and its behavior across different categories. Typically, the confusion matrix is an $n \times n$ matrix, where n denotes the number of classes. In this study, there are three categories: The oil–water layer, the water layer, and the mudstone layer.

For multiclass classification problems, each row of the confusion matrix represents the distribution of samples for a specific actual class, whereas each column corresponds to the number of samples predicted as a particular class by the model. The diagonal elements indicate the number of correctly predicted samples, whereas the off-diagonal elements reflect the degree of misclassification between different categories. The core metrics include the following.

True positive (TP): Refers to the number of samples correctly classified as belonging to a specific class. For a given class, this corresponds to the diagonal element of the confusion matrix, which represents the actual samples of that class that were correctly identified.

False positive (FP): Refers to the number of samples from other classes that were incorrectly predicted as belonging to a specific class. For a given class i, FP is the sum of the off-diagonal elements in column i of the matrix.

False negative (FN): Refers to the number of samples from a specific class that were incorrectly predicted as belonging to other classes. For a given class i, FN is the sum of the off-diagonal elements in row i of the matrix.

Based on the core metrics mentioned above, several performance indicators can be calculated to evaluate the

classification model:

Precision: Measures the proportion of samples predicted as a specific class that actually belong to that class. The main expression is depicted as Eq. 8.

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{8}$$

This indicates the accuracy of the model's positive predictions for a given class.

Accuracy: Measures the proportion of correctly predicted samples to the total number of samples in a classification model. The main expression is depicted as Eq. 9.

$$Accuracy = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} TP_i + FP_i + FN_i} \tag{9}$$

This reflects the overall prediction accuracy of the model across all classes.

Recall: Measures the proportion of actual samples of a specific class that are correctly predicted by the model. The main expression is depicted as Eq. 10.

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{10}$$

This reflects the model's ability to identify positive samples for a given class.

F1 score: The harmonic mean of Precision and Recall, providing a balanced measure that considers both false positives and false negatives. It is particularly useful when the class distribution is imbalanced. The main expression is depicted as Eq. 11.

$$F1_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \tag{11}$$

The F1 score offers a comprehensive reflection of the model's performance for a given class.

In this study, Precision, Recall, and F1 score were computed based on the confusion matrix. Importantly, in this regional dataset, the mudstone layer accounts for 63%, the water layer accounts for 20%, and the oil-water layer accounts for 17%. Thus, the data distribution is imbalanced. In tasks with extremely imbalanced data, such as this one, high Recall is more meaningful than high Precision. This is because, in the present study, the class with fewer samples (oil-bearing formations) has a higher value, and predicting more instances of the minority class is crucial. In scenarios with extreme data imbalances, missing the minority class (oil-bearing formations in this case) often results in significant losses compared with misclassifying other reservoir types as oil-bearing formations. Therefore, in this study, Recall is the primary evaluation metric.

4.4. Model architecture

Faced with diverse problems and methodologies, constructing the most suitable network model forms the foundation of effective deep learning applications. A critical step in building a deep neural network lies in the determination of network hyperparameters, which significantly influence the model's performance. In this study, five major hyperparameters were considered: (1) The learning rate; (2) the number of network layers; (3) the number of neurons per hidden layer; (4) the dropout regularization parameter, and (5) the batch size.

To optimize the learning rate, dropout rate, and batch size, the authors fixed a random seed and conducted a grid search across predefined ranges within a set number of training steps. The learning rate was explored within the range [0.001–0.1], the dropout rate was varied from 0–0.9, and batch sizes of 8, 32, 64, and 96 were tested. Table 3 summarizes the tested hyperparameter configurations.

Based on the results shown in Figs. 10–11, a learning rate of 0.01 and a batch size of 32 provided the best trade-off between convergence speed and model stability on the validation set, achieving superior accuracy and lower loss. As illustrated in Fig. 12, the model's performance was positively correlated with the dropout rate in terms of accuracy and

Table 3. Network training parameters.

Parameter	Values
Learning rate	0.001, 0.005, 0.01, 0.05, 0.1
Dropout	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
Batch size	8, 32, 64, 96

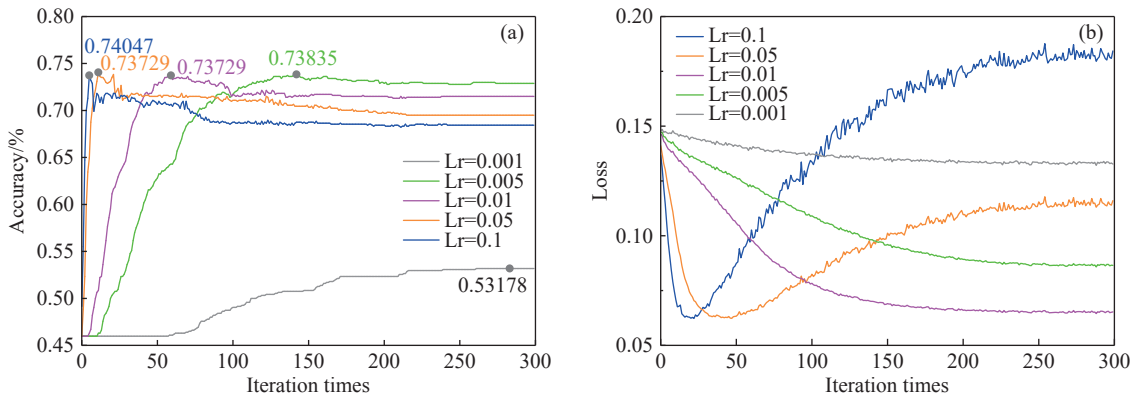


Fig. 10. Optimal learning rate simulation results. a–Variation in accuracy with the number of iterations for different lr values in the validation sets. b–variation in loss with the number of iterations for different lr values in validation sets.

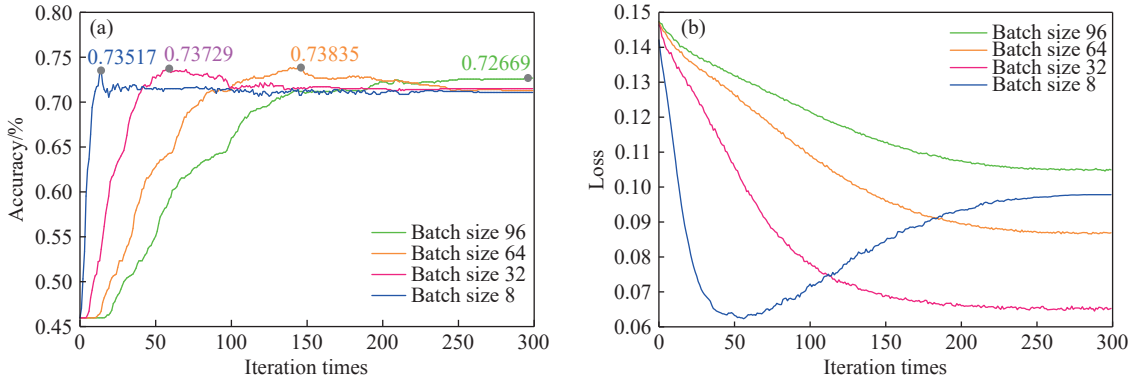


Fig. 11. Optimal batch size simulation results. a–variation in accuracy with the number of iterations for different batch sizes in the validation sets. b–variation in loss with the number of iterations for different batch sizes in validation sets.

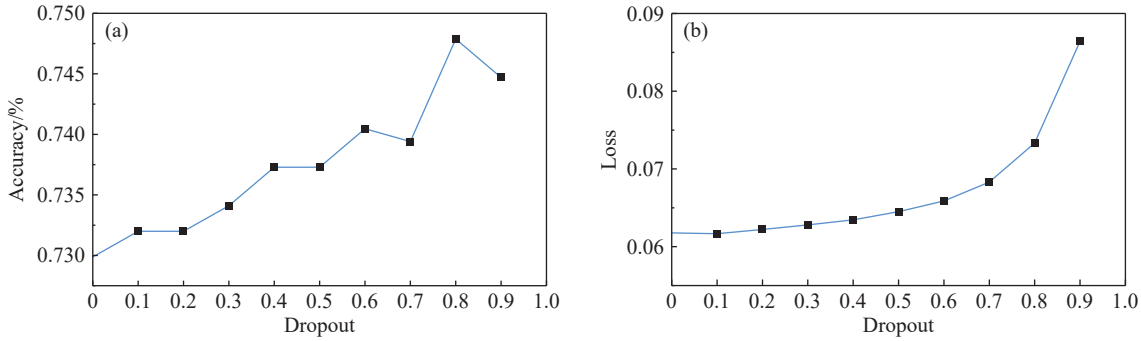


Fig. 12. Optimal dropout simulation results. a–Variation in accuracy with the number of iterations for different dropouts in the validation sets. b–variation in loss with the number of iterations for different dropouts in the validation sets.

negatively correlated with respect to the loss value. However, excessively high dropout values disrupt the learning process, resulting in significant accuracy degradation and increased loss. Therefore, after careful evaluation, a moderate dropout value of 0.5 was selected as the optimal setting, balancing regularization effectiveness and training performance.

To determine the optimal number of hidden layers and neurons per layer, the authors fixed a random seed and designed 12 ablation experiments across different network configurations in terms of width, depth, and combined width-depth structures. All other variables and parameters were held constant to isolate the impact of these architectural choices. The objective was to identify the most effective network structure by comprehensively evaluating model accuracy on the validation set, as well as the Recall specifically for the oil-water layer, which is of particular importance in this study.

The experimental results are summarized in Table 4. Among the tested configurations, the network architecture with three hidden layers and neuron sizes of 100–100–50 (denoted as 6–100–100–50–3) demonstrated the best performance. It not only maintained a high overall classification accuracy but also achieved the highest Recall for the oil–water layer, validating it as the optimal structure for this task.

After the optimal network parameters are determined, the neural network model adopts the structure (6–100–100–50–3) shown in Fig. 13, which consists of four fully connected

Table 4. Statistics of the ablation experiment results.

Model	Accuracy	Parameters	Recall (Oil-water layer)
Proposed model (6–100–100–50–3)	0.715	16.0 K	0.5798
Depth Ablation			
2 hidden layers (6–100–100–3)	0.7648	11.1 K	0.5378
1 hidden layer (6–100–3)	0.7606	1.0 K	0.5042
Width Ablation			
First layer halved (6–50–100–50–3)	0.6737	10.7 K	0.2185
Second layer halved (6–100–50–50–3)	0.7468	8.5 K	0.2605
Third layer halved (6–100–100–25–3)	0.7828	13.4 K	0.4958
First layer doubled (6–200–100–50–3)	0.7161	26.7 K	0.3782
Second layer doubled (6–100–200–50–3)	0.6886	31.1 K	0.4958
Third layer doubled (6–100–100–100–3)	0.7765	21.2 K	0.458
All layers halved (6–50–50–25–3)	0.6186	4.3 K	0.1471
All layers doubled (6–200–200–100–3)	0.7479	62.0 K	0.458
Mixed Architecture Ablation (Depth + Width)			
Reduced depth and modified width (6–100–50–3)	0.75	5.9 K	0.437
Increased depth (6–100–100–50–25–3)	0.6981	17.2 K	0.4538

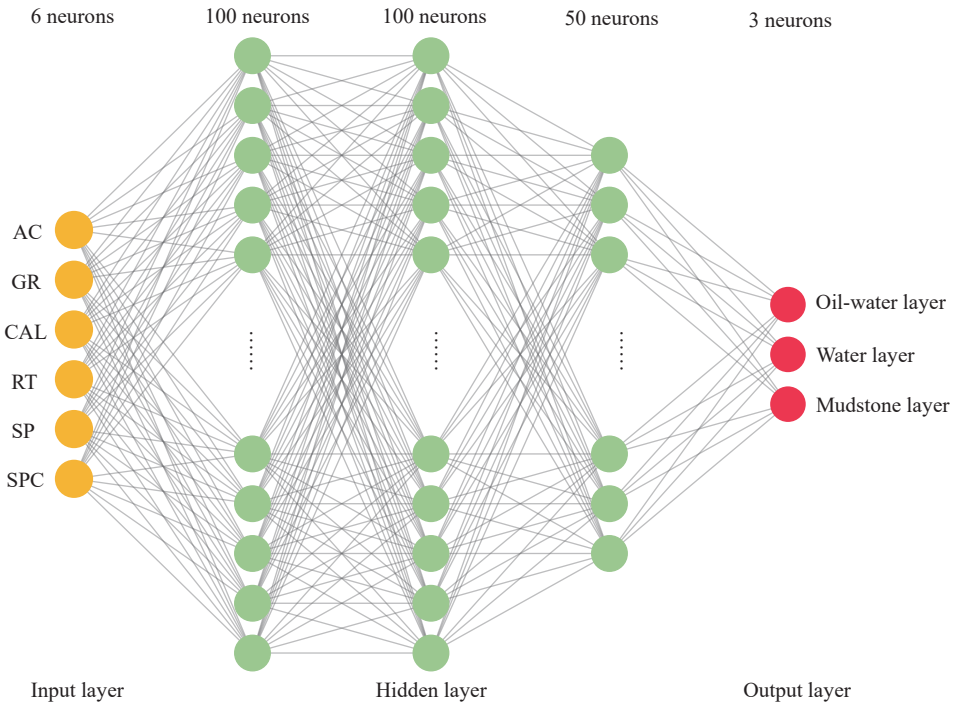


Fig. 13. Deep neural network model for reservoir fluid identification.

layers. ReLU activation functions are applied after the first three fully connected layers, and dropout is applied after the third fully connected layer. The last layer is the output layer. To increase the training efficiency of the neural network, mini-batch training was employed with a batch size set to 32, indicating that 32 sets of training data were randomly selected for each training iteration. Focal loss was used as the loss function to address the issue of class imbalance. The stochastic gradient descent (SGD) optimizer was chosen to update the model parameters, and a cosine annealing decay strategy was applied to adjust the learning rate, with the initial learning rate set to 0.01.

5. Results

5.1. Identification performance of different machine learning models

To assess the accuracy of the deep learning model, the authors conducted validation on the constructed dataset using our network architecture. the authors compared the predictive performance of the model with those of five machine learning methods, namely, linear regression (LR), naive Bayes (NB), gradient boosting decision tree (GBDT), random forest (RF), and support vector machine (SVM).

The confusion matrices of different models on the validation set after training are shown in Fig. 14.

Based on the confusion matrix, the remaining evaluation metrics are calculated for all the models (Tables 5–8).

The results shown in the table indicate that, compared with all the other machine learning models, the deep neural network-based reservoir fluid prediction model achieves the highest Recall and the best prediction results for the oil-water layer.

5.2. Performance of models using different loss functions

The confusion matrices for the deep learning model with different loss functions on the validation set are shown in Fig. 15.

Based on the confusion matrices, the remaining evaluation metrics for all the models were calculated (Tables 9–11).

The table shows that the deep neural network model for reservoir fluid type identification, which uses the focal loss function, achieves the highest recall and the best prediction results for the oil-water layer.

5.3. Dependence of model classification results on different features

To identify key logging parameters, we calculated mutual information scores between each feature and the target (fluid type), and the calculation results are shown in Fig. 16.

6. Discussion

6.1. Comparison with traditional machine learning models

This study utilizes real logging data from the Jiyuan oilfield in the Ordos Basin to develop an intelligent identification model aimed at assisting in the interpretation of logging information. By comparing the Recall of the deep learning model built in this study on the validation set with those of the identification models constructed using classic machine learning methods such as RF, SVM, LR, NB, and GBDT, the results reveal that the deep learning identification model achieves the highest Recall for the oil–water layer. This indicates that the deep learning model has the highest

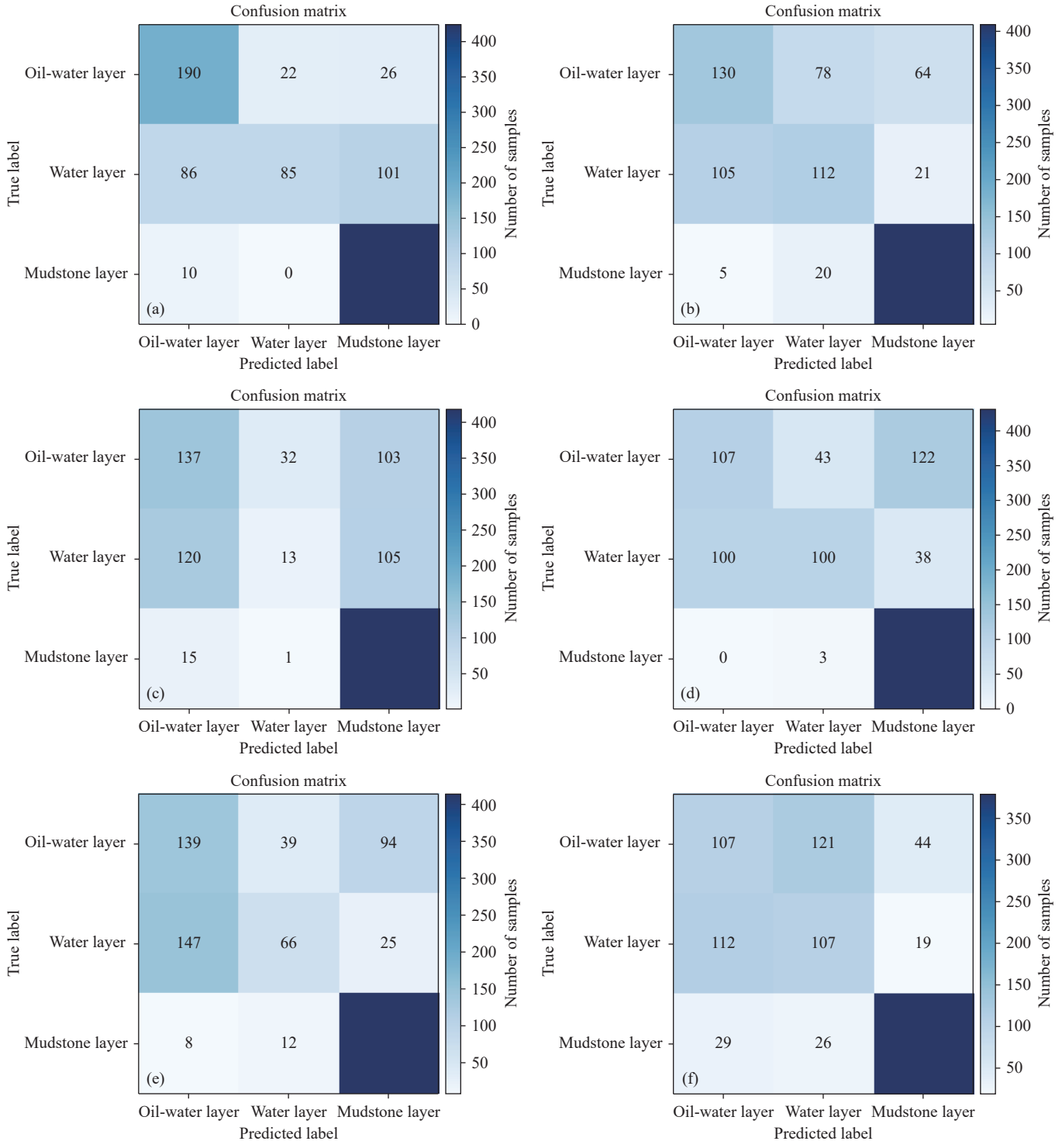


Fig. 14. Model validation confusion matrices. a–Deep neural network (DNN); b–random forest (RF); c–support vector machine (SVM); d–logistic regression (LR); e–naive Bayes (NB); f–gradient boosting decision trees (GBDT).

probability of capturing the oil-water layer, which is of greater concern in this study, effectively reducing economic losses from missing oil-bearing layers in the strata. This demonstrates that deep learning models can more fully explore the intrinsic relationships between heterogeneous oil-water layers and various logging attributes, achieving accurate and intelligent extraction of key logging features and fulfilling the objectives of this study.

6.2. Impact of loss function design

When applying different loss functions to the deep learning model, the results show that using the traditional cross-entropy loss function yields better prediction performance for the mudstone layer, which has relatively high Recall. However, the prediction performance for the oil-water layer and water layer methods is poorer. This discrepancy is due to the overrepresentation of the mudstone layer in the

Table 5. Support for each class label.

	Oil–water Layer	Water Layer	Mudstone Layer
Support	238	272	434

Table 6. Precision of different models.

	RF	SVM	LR	NB	GBDT	Method in this study
Oil–water Layer	0.53	0.28	0.68	0.56	0.42	0.66
Water Layer	0.54	0.50	0.52	0.47	0.43	0.79
Mudstone Layer	0.83	0.67	0.73	0.78	0.86	0.77
Macro avg	0.63	0.48	0.64	0.60	0.57	0.74
Weighted avg	0.67	0.52	0.66	0.64	0.62	0.75

Table 7. Recall of different models.

	RF	SVM	LR	NB	GBDT	Method in this study
Oil–water Layer	0.47	0.05	0.42	0.28	0.45	0.80
Water Layer	0.48	0.50	0.39	0.51	0.39	0.31
Mudstone Layer	0.94	0.96	0.99	0.95	0.87	0.98
Macro avg	0.63	0.51	0.60	0.58	0.57	0.70
Weighted avg	0.69	0.60	0.68	0.66	0.63	0.74

Table 8. F1-Scores of different models.

	RF	SVM	LR	NB	GBDT	Method in this study
Oil–water Layer	0.50	0.09	0.52	0.37	0.43	0.73
Water Layer	0.51	0.50	0.45	0.49	0.41	0.45
Mudstone Layer	0.88	0.79	0.84	0.86	0.87	0.86
Macro avg	0.63	0.46	0.60	0.57	0.57	0.68
Weighted avg	0.67	0.53	0.65	0.63	0.63	0.71

training samples, where sample imbalance significantly impacts the model's ability to focus on and learn from minority class features. To address sample imbalance, the authors introduced a class-weighted cross-entropy loss function, which improved the Recall for the water layer by 9%. However, it did not significantly enhance the Recall for the oil-water layer, which is the primary focus. By adopting the focal loss function, the recognition performance for the water layer showed only marginal improvement, but the Recall for the oil–water layer increased significantly to 80%. These results demonstrate that improving the model's loss function can effectively overcome the issue of class imbalance in reservoir fluid identification tasks, further enhancing the recognition performance of oil-bearing layers.

6.3. Dependence of fluid identification on key logging parameters

Mutual information analysis reveals the importance ranking of logging parameters in fluid identification within low-permeability reservoirs: SPC, SP, and GR are core parameters, whereas RT, CAL, and AC serve as auxiliary parameters.

SP typically reflects differences in formation water salinity, whereas SPC may further capture dynamic variations in fluid mobility (Rider MH, 2018). In low-permeability reservoirs, the distribution of formation water and hydrocarbons is often controlled by micropore structures, and the high sensitivity of SPC and SP makes them key indicators for distinguishing oil-bearing layers. GR effectively identifies mudstone layers, which serve as boundaries between reservoir and nonreservoir formations. The mudstone content directly affects the effective porosity. In low-permeability reservoirs, even slight variations in clay content can significantly impact reservoir quality; thus, GR plays an important role in eliminating nonreservoir interference. RT is a key indicator of hydrocarbon saturation in conventional reservoirs; however, in low-permeability reservoirs, its response may be influenced by multiple factors (e.g., complex pore structures and clay cementation), reducing its direct ability to indicate fluid type. CAL mainly reflects borehole stability, and AC is related to the elastic modulus of the rock. These parameters are less directly associated with fluid types but may indirectly indicate fracture development or rock mechanical properties in certain scenarios and can be interpreted in conjunction with other parameters.

6.4. Practical implications of misclassification

Considering the economic significance of oil-bearing layers, the evaluation metrics were designed to prioritize the Recall for oil-water layers, thereby minimizing the risk of undetected hydrocarbon-bearing zones. However, this prioritization may lead to misclassification errors in other categories, such as increased false positives for water layers. Nevertheless, such errors are operationally manageable through multistage validation workflows:

(i) Failure to detect oil-bearing layers may result in irreversible production loss. Undetected hydrocarbon zones could lead to long-term declines in productivity, directly impacting the economic efficiency of oilfields. In contrast, false-positive predictions for water layers (misclassified as oil-bearing layers) can be rectified through secondary logging or production testing, with subsequent verification costs remaining operationally manageable.

(ii) In practical field operations, model predictions are typically integrated with manual verification and fluid sampling analysis. This workflow ensures enhanced decision-making confidence through cross-validation with multisource data prior to high-cost operations (e.g., hydraulic fracturing, well completion), thereby preventing resource misallocation caused by single-model errors.

6.5. Model application

Based on the model trained in this study, a secondary interpretation was conducted on selected old wells in the Chang 1 layer of this area. In 2024, well Di246–84 in the northeastern part of the block was selected for perforation adjustment and recompletion in the Chang 1^2 layer. The well was completed on September 27, and its current productivity was 3.0 t/day, with a water cut of 14.9%. The new layer has shown good oil production performance, providing a positive indication for regional reservoir evaluation. Further well deployment is needed to expand exploration. Given the low

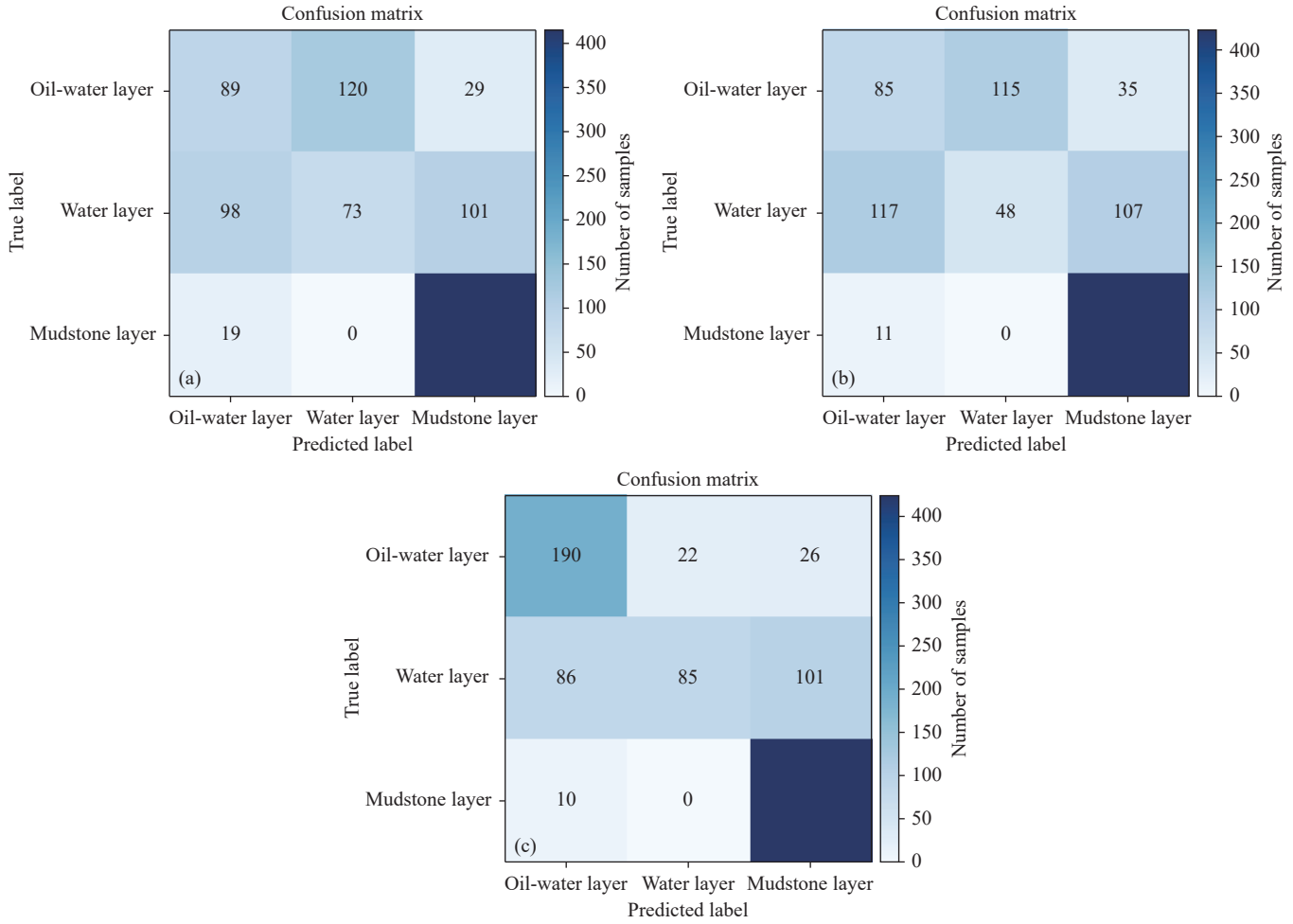


Fig. 15. Model validation confusion matrices. a–Class-weighted cross-entropy loss function; b–traditional cross-entropy loss function; c–focal loss function.

Table 9. Precision of models with different loss functions.

	Traditional cross-entropy loss function	Class-weighted cross-entropy loss function	Focal loss function
Oil-water Layer	0.40	0.43	0.66
Water Layer	0.29	0.38	0.79
Mudstone Layer	0.75	0.76	0.77
Macro avg	0.48	0.52	0.74
Weighted avg	0.53	0.57	0.75

Table 10. Recall of models with different loss functions.

	Traditional cross-entropy loss function	Class-weighted cross-entropy loss function	Focal loss function
Oil-water Layer	0.36	0.37	0.80
Water Layer	0.18	0.27	0.31
Mudstone Layer	0.97	0.96	0.98
Macro avg	0.50	0.53	0.70
Weighted avg	0.59	0.61	0.74

degree of well control in the southern part of the region, one appraisal well has already been deployed to further delineate the reservoir extent.

Owing to the heterogeneity of the strata in the study area and the confidentiality of oilfield geophysical data, the training and validation datasets for the identification model constructed in this study are limited. In the future, the authors' goal is to actively expand data access to increase the volume and diversity of data, addressing these issues while exploring the potential of leveraging more advanced neural network architectures and solving this task under more complex tectonic conditions, thereby enhancing the comprehensiveness and depth of our research.

7. Conclusions

This study utilized real well-logging data from the Jiyuan area in the Ordos Basin to construct a deep neural network model for reservoir fluid type prediction. The model takes multiple well-logging response data from the strata as inputs and predicts the reservoir fluid type as the output. Through experimentation, it has been demonstrated that:

(i) Compared with traditional machine learning models,

Table 11. F1-Scores of models with different loss functions.

	Traditional cross-entropy loss function	Class-weighted cross-entropy loss function	Focal loss function
Oil-water Layer	0.38	0.40	0.73
Water Layer	0.22	0.31	0.45
Mudstone Layer	0.85	0.85	0.86
Macro avg	0.48	0.52	0.68
Weighted avg	0.55	0.58	0.71

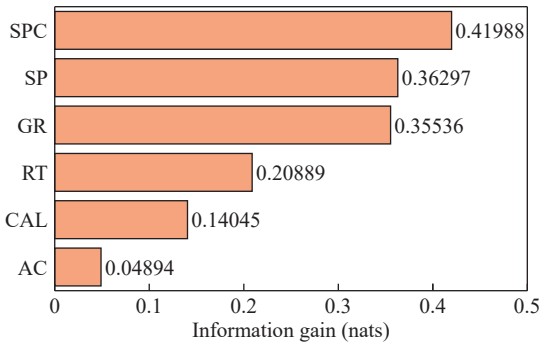


Fig. 16. Mutual information ranking for logging parameters. AC−Acoustic; CAL−Caliper; GR−Natural gamma ray; RT−True formation resistivity; SP−Spontaneous potential; SPC−Spontaneous potential amplitude.

reservoir type prediction models built on deep neural networks are more effective at revealing the inherent connections between oil-bearing reservoirs and various raw well-logging responses, resulting in an ideal predictive performance. Addressing the issue of sample imbalance during modeling, the introduction of focal loss, along with an increased emphasis on minority class weights during model training, significantly enhances the model's ability to predict oil-bearing reservoir categories.

(ii) The deep learning prediction model, trained with single-well unit data division, performs well in predicting the oil-bearing layers in the Chang 1 Formation of the Jiyuan area in the Ordos Basin. This validates the feasibility of the proposed method and expands the application scenarios of using artificial intelligence for well-logging data to predict oil-bearing reservoirs intelligently. In the future, continuous optimization of the model can increase its efficiency and accuracy in recognizing all labels through the interpretation of massive amounts of well-logging data.

(iii) Among all the features influencing the fluid identification model, SPC, SP, and GR are core features, whereas RT, CAL, and AC serve as auxiliary features. This reflects that, in low-permeability reservoirs, the microscopic control mechanisms of fluid distribution–such as pore structure and clay content–have a dominant influence on well logging responses. By prioritizing high mutual information parameters and validating with multisource data, the practicality and reliability of the fluid identification model can be enhanced.

In subsequent research, the authors will consider expanding the scope and depth of the research, collecting more types of sample data to optimize the model and further improving the accuracy and applicability of the machine learning model for reservoir fluid identification.

CRediT authorship contribution statement

Wen-bo Li, Zhen-kai Zhang and Dong-tao Li conceived of the presented idea. Wen-bo Li, Xiao-ye Wang and Lei He collected and processed the dataset for this study. Wen-bo Li and Ling-yi Liu constructed the deep learning model. Wen-bo Li wrote the manuscript. Zen-lin Hong reviewed and proofread the manuscript. All authors discussed the results and contributed to the final manuscript.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgment

This study was supported by a project of the Shaanxi Youth Science and Technology Star (2021KJXX-87), and public welfare geological survey projects of Shaanxi Institute of Geologic Survey (20180301, 201918 and 202103). The authors would like to extend their profound gratitude to the editors and anonymous reviewers of this manuscript for their constructive advice and corrections.

Data availability statement

Owing to confidentiality regulations of the oilfield company regarding logging data, the original well logging data cannot be publicly shared. However, a synthetic dataset generated through quantile-matching methodology is available at Figshare (https://doi.org/10.6084/m9.figshare.29126513.v2), with all sensitive information removed. The synthetic dataset preserves the quantile characteristics of numerical columns and the categorical distribution ratios of nonnumerical columns from the original data, maintaining consistency in extreme values and distribution patterns.

References

Alizadeh B, Najjari S, Kadkhodaie-Ilkhchi A. 2012. Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the South Pars Gas Field, Persian Gulf, Iran. Computers & Geosciences, 45, 261–269. doi: 10.1016/j.cageo.2011.11.024.

Andreieva V, Shvai N. 2021. Generalization of cross-entropy loss function for image classification. Mohyla Mathematical Journal, 3, 3–10. doi: 10.18523/2617-7080320203-10.

Anifowose FA, Labadin J, Abdulraheem A. 2013. Prediction of petroleum reservoir properties using different versions of adaptive Neuro-Fuzzy inference system hybrid models. International Journal of Computer Information Systems and Industrial Management, Applications 5: 413–426.

Bagheripour P. 2014. Committee neural network model for rock permeability prediction. Journal of Applied Geophysics, 104,

142–148. doi: 10.1016/j.jappgeo.2014.03.001.
Bai Z, Tan MJ, Shi YJ, Guan XN, Wu HB, Huang YH. 2022. Log interpretation method of resistivity low-contrast oil pays in Chang 8 tight sandstone of Huanxian area, Ordos Basin by support vector machine. Scientific Reports, 12, 1046. doi: 10.1038/s41598-022-04962-0.
Baziar S, Tadayoni M, Nabi-Bidhendi M, Khalili M. 2014. Prediction of permeability in a tight gas reservoir by using three soft computing approaches: A comparative study. Journal of Natural Gas Science and Engineering, 21, 718–724. doi: 10.1016/j.jngse.2014.09.037.
Bhatt A, Helle HB. 1999. Porosity, Permeability and TOC Prediction from Well Logs Using a Neural Network Approach. In 61st EAGE Conference and Exhibition, Helsinki, Finland: European Association of Geoscientists & Engineers. doi:10.3997/2214-4609.201407849.
Cheng C, Li PY, Chen Y, Ye Y, Gao Y, Zhang L. 2022. Research progress of reservoir logging evaluation based on machine learning. Progress in Geophysics, 37(1), 164–177 (in Chinese with English abstract). doi: 10.6038/pg2022FF0031.
Das B, Chatterjee R. 2018. Well log data analysis for lithology and fluid identification in Krishna-Godavari Basin, India. Arabian Journal of Geosciences, 11(10), 231. doi: 10.1007/s12517-018-3587-2.
El-M Shokir EM. 2006. A novel model for permeability prediction in uncored wells. SPE Reservoir Evaluation & Engineering, 9(3), 266–273. doi: 10.2118/87038-pa.
Feng Q. 2018. Yan 9, Chang 1reservoir “four” and the logging interpretation model extension ZJ Ding bian Oilfield. Xi’an, Northwest University, Master thesis, 1–83 (in Chinese with English abstract)
Helle HB, Bhatt A, Ursin B. 2002. Fluid Saturation from Well Logs Using Modular Neural Networks. In 64th EAGE Conference & Exhibition, Florence, Italy: European Association of Geoscientists & Engineers. doi:10.3997/2214-4609-pdb.5.B045.
He M, Gu HM, Wan H. 2020. Log interpretation for lithology and fluid identification using deep neural network combined with MAHAKIL in a tight sandstone reservoir. Journal of Petroleum Science and Engineering, 194, 107498. doi: 10.1016/j.petrol.2020.107498.
He TP, Zhou YQ, Li YH, Xie HY, Shang YH, Chen TT, Zhang ZW. 2024. Research on the microscopic pore-throat structure and reservoir quality of tight sandstone using fractal dimensions. Scientific Reports, 14, 22825. doi: 10.1038/s41598-024-74101-4.
Huang ZH, Shimeld J, Williamson M, Katsube J. 1996. Permeability prediction with artificial neural network modeling in the Venture gas field, offshore eastern Canada. Geophysics, 61(2), 422–436. doi: 10.1190/1.1443970.
Jaikla C, Devarakota P, Auchter N, Sidahmed M, Espejo I. 2019. FaciesNet: Machine Learning Applications for Facies Classification in Well Logs. In Machine Learning and the Physical Sciences Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS).
Jouhari H, Lei DM, Al-qaness MA, Elaziz MA, Ewees AA, Farouk O. 2019. Sine-Cosine Algorithm to Enhance Simulated Annealing for Unrelated Parallel Machine Scheduling with Setup Times. Mathematics, 7(11), 1120. doi:10.3390/math7111120.
Karimpouli S, Fathianpour N, Roohi J. 2010. A new approach to improve neural networks' algorithm in permeability prediction of petroleum reservoirs using supervised committee machine neural network (SCMNN). Journal of Petroleum Science and Engineering, 73(3–4), 227–232. doi: 10.1016/j.petrol.2010.07.003.
Lin TY, Goyal P, Girshick R, He KM, Dollár P. 2020. Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 318–327. doi: 10.1109/TPAMI.2018.2858826.
Li X, Shi YJ, Wang L, Hu S. 2013. Logging identification and evaluation technique of tight sandstone gas reservoirs: Taking Sulige gas field as an example. Natural Gas Geoscience, 24(1), 62–68 (in Chinese with English abstract). doi: 10.11764/j.issn.1672-1926.2013.01.62.
Luo SL, Hu GM, Li LX, Wang JW, Li XT. 2015. Genetic analysis and well-log evaluation of the productivity simulation for unconventional gas reservoirs of tight sanstone: A case from B gas reservoirs in A sag. Progress in Geophysics, 30(6), 2714–2722 (in Chinese with Englishabstract). doi: 10.6038/pg20150632.
Maiti S, Krishna Tiwari R, Kümpel HJ. 2007. Neural network modelling and classification of lithofacies using well log data: A case study from KTB borehole site. Geophysical Journal International, 169(2), 733–746. doi: 10.1111/j.1365-246x.2007.03342.x.
Mohaghegh S, Arefi R, Bilgesu I, Ameri S, Rose D. 1995. Design and Development of An Artificial Neural Network for Estimation of Formation Permeability. SPE Computer Applications 7(6): 151–54. doi:10.2118/28237-PA.
Mohaghegh S, Arefi R, Ameri S, Aminiand K, Nutter R. 1996. Petroleum reservoir characterization with the aid of artificial neural networks. Journal of Petroleum Science and Engineering, 16(4), 263–274. doi: 10.1016/s0920-4105(96)00028-9.
Nair V, Hinton GE. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. Proceedings of ICML, 27: 807–814.
Nooruddin HA, Hossain ME. 2011. Modified Kozeny–Carmen correlation for enhanced hydraulic flow unit characterization. Journal of Petroleum Science and Engineering, 80(1), 107–115. doi: 10.1016/j.petrol.2011.11.003.
Pan Y, Wang J, Wu Y, Peng HH, Yang H, Chen C. 2022. Reconstructed quality improvement with a stochastic gradient descent optimization algorithm for a spherical hologram. Applied Optics, 61(17), 5341–5349. doi: 10.1364/AO.462161.
Peng Z, Yang HJ, Pan HP, Ji Y. 2016. Identification of low resistivity oil and gas reservoirs with multiple linear regression model. 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). August 13-15, 2016, Changsha, China. IEEE, 529–533. doi: 10.1109/FSKD.2016.7603229.
Rider MH. 2018. The Geological Interpretation of Well Logs. Cambridge, Cambridge University Press, 207–220.
Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014. Dropout: Prevent NN from overfitting. Journal of Machine Learning Research, 15, 1929–1958. doi: 10.5555/2627435.2670313.
Tahmasebi P, Hezarkhani A. 2012. A fast and independent architecture of artificial neural network for permeability prediction. Journal of Petroleum Science and Engineering, 86–87, 118–126. doi: 10.1016/j.petrol.2012.03.019.
Tan MJ, Bai Y, Zhang HT, Li GR, Wei XP, Wang AD. 2020. Fluid typing in tight sandstone from wireline logs using classification committee machine. Fuel, 271, 117601. doi: 10.1016/j.fuel.2020.117601.
Wang DX. 2016. Study on the rock physics model of gas reservoirs in tight sandstone. Chinese Journal of Geophysics, 59, 4603–22. doi:10.6038/cjg20161222.
Wu XM, Shi YZ, Fomel S, Liang LM. 2018. Convolutional Neural Networks for Fault Interpretation in Seismic Images. In SEG Technical Program Expanded Abstracts 2018, Anaheim, California: Society of Exploration Geophysicists, 1946–50. doi:10.1190/segam2018-2995341.1.
Yang WB, Xia KW, Fan SR. 2023. Oil logging reservoir recognition based on TCN and SA-BiLSTM deep learning method. Engineering Applications of Artificial Intelligence, 121, 105950. doi:10.1016/j.engappai.2023.105950.
Zhang GY. 2019. Recognition and Prediction Methods of High-quality Reservoirs of Tight Sandstones in Shaximiao Formation in Chuanxi Depression. Beijing, China University of Petroleum, Ph. D thesis, 1-151 (in Chinese with English abstract).
Zhou XM, Li YW, Song XD, Jin LX, Wang XX. 2023. Thin reservoir identification based on logging interpretation by using the support vector machine method. Energies, 16(4), 1638. doi:10.3390/en16041638.
Zhou XQ, Zhang ZS, Zhang CM. 2021. Bi-LSTM deep neural network reservoir classification model based on the innovative input of logging curve response sequences. IEEE Access, 9, 19902–15. doi:10.1109/ACCESS.2021.3053289.