Petroleum Science 23 (2026) 4053–4064

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

# Integration of multi-source geological engineering data for fracturing parameter optimization model

Jie Li [a,b], Quan-Zhen Xiu [a], Xiao-Dong He [a,b], Gen-Sheng Li [a], Chang Li [b], Mao-Ya Xu [a], Tian-Xiang Zhou [a], Shou-Ceng Tian [a], Tian-Yu Wang [a,*]

[a] *State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China*
[b] *Xinjiang Oilfield Oil Production Technology Research Institute (Supervision Company), Karamay, 834000, Xinjiang, China*

A R T I C L E I N F O

*Article history:*
Received 19 May 2025
Received in revised form
25 March 2026
Accepted 26 March 2026
Available online 1 April 2026

Edited by Jia-Jia Fei

*Keywords:*
Fracturing parameter optimization
Fractured horizontal well
Production prediction
Deep learning
Shale oil
Logging data

A B S T R A C T

Fracturing parameter design is critical to the economic viability and effectiveness of reservoir stimulation, making multi-stage hydraulic fracturing in horizontal wells a cornerstone technology for unconventional hydrocarbon development. However, conventional design methods relying on mechanistic models and empirical expertise often fail to provide well-specific customization due to their limited adaptation to geological heterogeneity. Erratic fracturing performance often stems from the limitations of existing data-driven approaches in integrating multi-source data. To address these challenges, this study proposes a data-driven fracturing parameter design strategy based on the fusion of well-logging data. The approach utilizes a hybrid architecture combining a convolutional neural network (CNN), a multi-layer perceptron (MLP), and a comprehensive learning particle swarm optimization (CLPSO) algorithm. Specifically, the Gramian angular field (GAF) is implemented to transform logging data into visual representations, enhancing feature extraction and production prediction accuracy to inversely derive optimal operational parameters. Results demonstrate that the proposed model substantially improves production prediction accuracy and enhances single-well output through parameter optimization. Compared to the AutoGluon ensemble framework, the proposed model achieves a 20% improvement in prediction accuracy. Field applications demonstrate that the model increases single-well shale oil yields by 6.6%–18.1% compared to conventional designs. This study provides a novel framework for optimizing fracturing parameters in unconventional shale oil reservoirs.


## 1. Introduction

China's significant shale oil reserves are characterized by great burial depths, poorly developed natural fractures, and pronounced heterogeneity, posing major challenges for commercial extraction (Pang et al., 2023). Multistage hydraulic fracturing in horizontal wells has become the essential technology for developing unconventional resources, particularly shale oil (Barati and Liang, 2014; Lu et al., 2022). By creating complex fracture networks through high-pressure fluid injection, this technique significantly enhances well productivity and estimated ultimate recovery (EUR) (Li et al., 2022a; Nouri et al., 2025). Consequently, optimizing fracturing parameters—including fluid volume, proppant mass, and cluster spacing—is vital for maximizing stimulation efficiency.

Current optimization of hydraulic fracturing parameters primarily couples fracture, fluid, and rock mechanics to analyze fracture propagation behavior and fluid dynamics (Al-Attar et al., 2020; Clark, 2010). Although these methods optimize pumping rates and pressures to enhance recovery (Li et al., 2022b; Yao et al., 2025), their reliance on simplified assumptions often neglects reservoir heterogeneity and complex fracture geometries, thereby limiting predictive accuracy (Gong et al., 2024; Zhang et al., 2021). To improve fidelity, numerical simulations—often based on the Navier-Stokes equations and sophisticated physical discretizations—are employed to model complex fluid-solid coupling and failure mechanisms (Sheng et al., 2023; Zhang et al., 2025). However, the prohibitive computational overhead and execution

* Corresponding author.
*E-mail address:* wangty@cup.edu.cn (T.-Y. Wang).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2026.03.059
1995-8226/

times associated with these high-fidelity models hinder real-time field decision-making and post-fracturing optimization (Ismail and Azadbakht, 2024). Furthermore, traditional empirical designs lack consistency and frequently fail to reach the reservoir's full production potential. Consequently, balancing computational efficiency with predictive accuracy remains a fundamental challenge for robust fracturing parameter design.

The proliferation of big data in the oil industry has catalyzed the adoption of machine learning (ML) to overcome these efficiency bottlenecks, enabling advanced fracturing design and production forecasting (Li et al., 2024; Ma et al., 2025; Morozov et al., 2020). While conventional algorithms such as random forest (RF), support vector machines (SVM), and the AutoGluon framework have demonstrated efficacy (Bhattacharya et al., 2019; Chahar et al., 2022; Zhu et al., 2025), they often struggle with the "curse of dimensionality" when processing high-dimensional engineering features, potentially compromising model generalizability. To address these complexities, deep learning (DL) architectures—such as long short-term memory (LSTM), gated recurrent units (GRU), and deep feedforward neural networks (DFNN)—have been employed for their superior capacity to capture complex non-linear dependencies (Liu et al., 2021; Mahzari et al., 2022; de Oliveira Werneck et al., 2022). Furthermore, hybrid models and meta-learning approaches (e.g., CNN-GRU) have emerged to integrate multi-modal data and optimize architectural configurations (Chen et al., 2024; Yang et al., 2022; Zhao et al., 2025). However, existing research remains largely confined to time-series analysis, frequently overlooking the extraction and integration of spatial, variable-length features. This gap hinders the ability of models to characterize critical stratigraphic heterogeneity, ultimately constraining the precision of intelligent fracturing optimization.

Integrating these production proxy models allows fracturing parameter design to be framed as a constrained optimization problem aimed at maximizing reservoir performance (Hyman et al., 2016). However, the high-dimensional and non-convex nature of these engineering problems often causes traditional gradient-based methods to converge to local optima. Consequently, the research focus has shifted toward global metaheuristic and stochastic approximation strategies (Hernandez-Perez et al., 2020). Swarm intelligence algorithms—particularly particle swarm optimization (PSO) and genetic algorithms (GA)—have pioneered this domain due to their capacity for gradient-free searching (Clerc and Kennedy, 2002; Deb et al., 2002). PSO is favored for its convergence speed and robustness in continuous search spaces (Muther et al., 2022), whereas GA and its variants excel in managing discrete variables, such as perforation cluster counts (Wang et al., 2022). Despite these successes, these methods remain sensitive to initial population diversity and are prone to premature convergence. Such persistent computational inefficiencies and stability issues continue to hinder the large-scale, real-time deployment of intelligent optimization in complex field operations.

This study develops a hybrid neural network model that integrates well-logging curves with geological and engineering tabular data to predict shale oil production. The Gramian angular field (GAF) method was employed to transform 1D logging data into 2D feature maps that encapsulate critical stratigraphic information, thereby strengthening geological constraints within the model. An optimization framework for fracturing design parameters was then developed by coupling the prediction model with the comprehensive learning particle swarm optimization (CLPSO) algorithm. This framework improves the accuracy of production forecasting and addresses the operational requirements for fracturing parameter optimization, ultimately enhancing single-well shale oil productivity. The proposed approach offers a robust solution for fracturing optimization, significantly increasing both operational efficiency and predictive precision.

## 2. Experiment

### 2.1. Fracturing parameter optimization framework workflow

Fig. 1 illustrates the workflow for fracturing parameter optimization through the integration of well-logging data. The convolutional neural network (CNN) extracts high-level features from the GAF-transformed logging data, which are subsequently reshaped into a format compatible with the multi-layer perceptron (MLP). These extracted features are concatenated with geological and operational parameters to serve as joint inputs for the MLP. The MLP models the complex non-linear relationships between these multi-modal inputs and well performance to generate accurate production predictions. Finally, subject to geological constraints, the CLPSO algorithm iteratively evaluates various fracturing parameter combinations to identify the optimal configuration that maximizes predicted production, thereby guiding fracturing design optimization.

### 2.2. Gramian angular field

GAF is an advanced technique for encoding one-dimensional sequence data into two-dimensional images (Wang and Oates, 2015). In this paper, we apply it to the representation learning of spatial sequence data. Fig. 2 illustrates that the method first normalizes the one-dimensional spatial sequence $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$. Subsequently, sequence points $(t_i, x_i)$ in the Cartesian coordinate system are transformed into polar coordinates $(r_i,\ \phi_i)$. By computing trigonometric relationships between $\phi_i$ and $\phi_j$, an $n \times n$ Gram matrix is constructed to capture pairwise spatial correlations between any two positional points $(i, j)$ in the sequence. Finally, Gram angle difference fields (**GADF**) and Gram angle sum fields (**GASF**), as defined in Eqs. (1) and (2), generate the final image features shown in Fig. 3(a) and (b). This transformation preserves the positional dependency of the sequence while converting spatial structural patterns into image textures, enabling CNNs to efficiently extract high-dimensional features from spatial sequences (Hsu et al., 2024).

$$\mathbf{GADF} = \left[\sin\left(\phi_i - \phi_j\right)\right] \tag{1}$$

$$\mathbf{GASF} = \left[\cos\left(\phi_i + \phi_j\right)\right] \tag{2}$$

where $t_i$ represents the spatial position or sequence index, encoded as a radius: $r_i = \frac{t_i}{N}$ (where $N$ is the total sequence length); $x_i$ denotes the measurement value at position $t_i$, mapped to an angular cosine: $\phi_i = \arccos(x_i)$.

### 2.3. Convolutional neural network

CNN is a deep neural network architecture widely applied in computer vision, image processing, speech recognition, and other fields (Lecun et al., 1998). Although initially designed for image recognition, its ability to automatically learn spatially local features through convolutional kernels and its translation invariance make it highly suitable for processing the structured data derived from well logging sequences in this study (Fukushima, 1980).

CNN primarily consists of convolutional layers, pooling layers, and fully connected layers, with convolutional layers serving to extract features from input data through convolution operations.

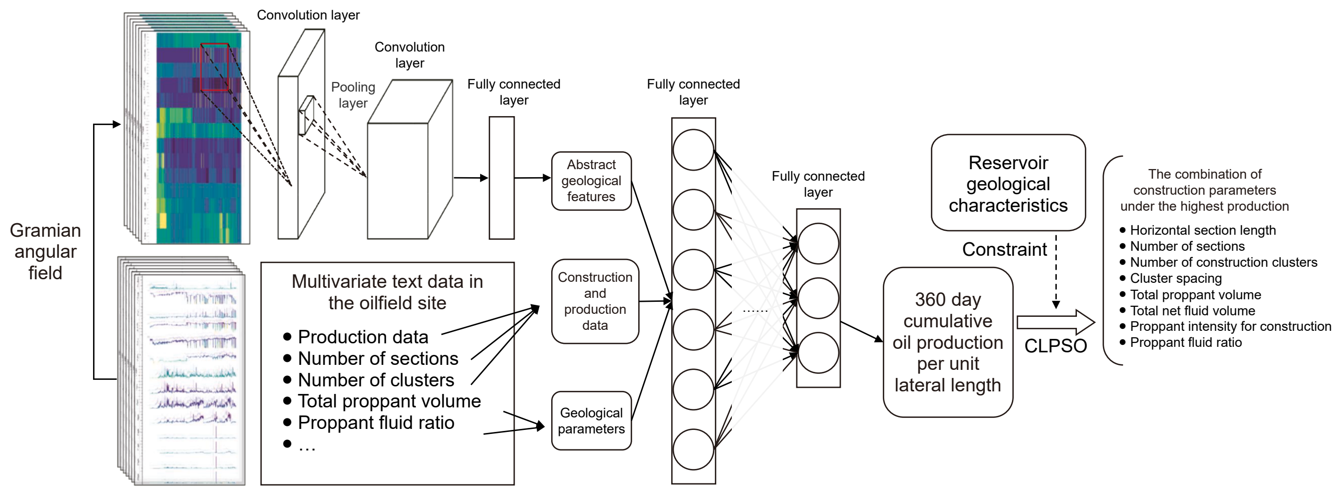


**Fig. 1.** Fracturing parameter optimization flowchart.

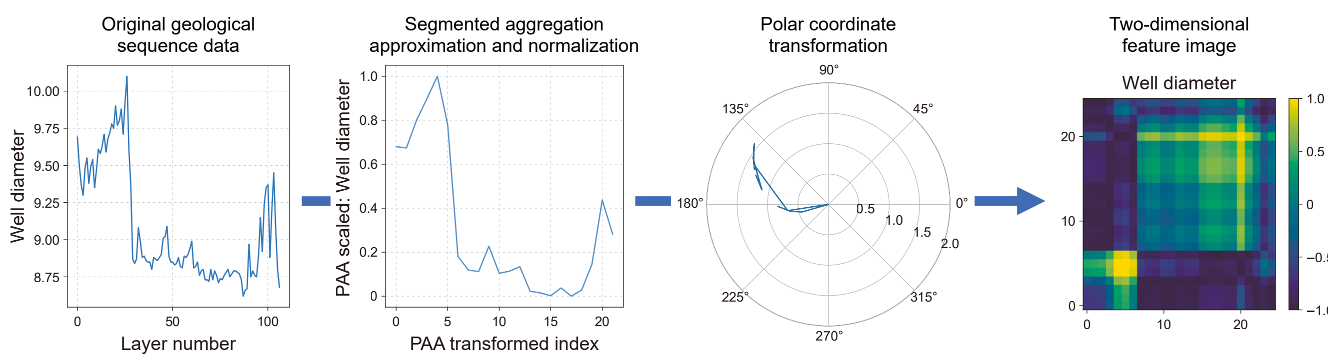


**Fig. 2.** Gram field feature image transformation process. PAA: piecewise aggregation approximation.

Convolution operation is executed through local receptive fields, where a diminutive filter (or convolutional kernel) is systematically convolved with the input image to produce feature maps. The convolutional layer produces a matrix of several feature mappings. The convolutional process is represented by Eq. (3).

$$y_{i,j}=\sum_{m=0}^{k-1}\sum_{n=0}^{k-1}x_{i+m,j+n}w_{m,n}+b \tag{3}$$

where $x$ is the input image, $w$ is the convolutional kernel, $b$ is the bias term, $k$ denotes the size of the convolutional kernel, and $y_{i,j}$ represents the output feature after convolution.

Pooling layers are generally situated after convolutional layers to diminish information redundancy, augment feature abstraction, and boost network translation invariance. The predominant pooling operations are max pooling and average pooling. This study utilizes max pooling, as delineated in Eq. (4), to extract the maximum value from each sub-region of the input feature maps. After feature extraction via convolutional and pooling layers, the network generally utilizes a fully connected layer for classification or regression purposes.

$$y_{i,j}=\max_{m,n}x_{i+m,j+n} \tag{4}$$

This study employed root mean square error (RMSE), coefficient of determination ($R^2$), mean percentage error (MPE), and mean absolute error (MAE) to thoroughly assess model performance. The MPE indicates the relative difference between expected and actual values, accurately representing prediction accuracy across different scales. The MAE measures the average absolute difference between forecasts and observations, offering a comprehensive assessment of the model's prediction error, as delineated by Eqs. (5) and (6):

$$\mathrm{MPE}=\frac{100\%}{n}\sum_{i=1}^{n}\frac{y_{\mathrm{predict}}^{(i)}-y_{\mathrm{true}}^{(i)}}{y_{\mathrm{predict}}^{(i)}} \tag{5}$$

$$\mathrm{MAE}=-\frac{1}{n}\sum_{i=1}^{n}\left|y_{\mathrm{predict}}^{(i)}-y_{\mathrm{true}}^{(i)}\right| \tag{6}$$

### 2.4. Comprehensive learning particle swarm optimization

The CLPSO algorithm builds upon the standard PSO by incorporating a comprehensive learning strategy that refines particle update rules to enhance global search capabilities (Kennedy and Eberhart, 1995). Fig. 4 illustrates the optimization workflow of the CLPSO algorithm, where candidate solutions, represented as particles, traverse the search space. Each particle represents a single candidate solution defined by two primary vectors: position and velocity. The algorithm seeks the global optimum by iteratively updating particle velocities and positions. Through the comprehensive learning mechanism, particles integrate

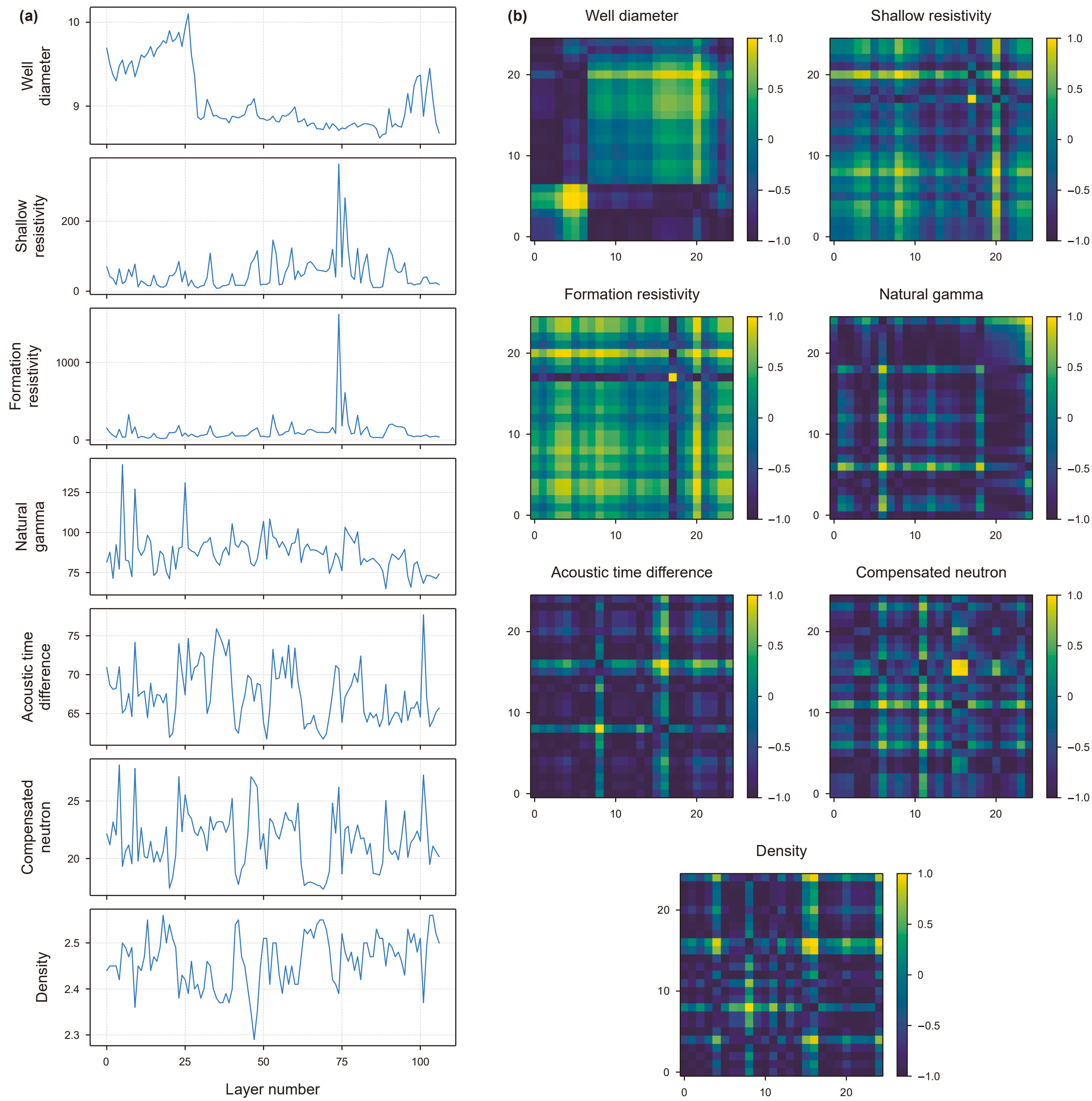


**Fig. 3.** The GAF transformation results: (**a**) logging image; (**b**) GAF-transformed image.

information from their own historical best positions and those of other particles. This multi-exemplar approach enhances population diversity and promotes robust information sharing, enabling a more thorough exploration of the search space (Liang et al., 2006).

The algorithm iteratively updates particle positions and velocities according to Eqs. (7) and (8). Specifically, Eq. (7) demonstrates that CLPSO deviates from the standard PSO by removing the centralized guidance of the global best (gbest) position. Instead, it employs a multi-exemplar approach, allowing each particle to independently construct a learning model from the historical best (pbest) positions of various individuals across each dimension of its velocity vector. This dimensional decomposition effectively mitigates the risk of premature convergence inherent in the standard PSO algorithm. The fitness values are evaluated concurrently until either the maximum iteration limit is reached or the objective function improvement falls below a predefined convergence threshold.

$$\nu_{i,d}^{(t+1)} = w \cdot \nu_{i,d}^{(t)} + c \cdot r_{i,d} \left( \mathrm{pbest}_{f_i(d),d}^{(t)} - x_{i,d}^{(t)} \right) \tag{7}$$

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \boldsymbol{\nu}_i^{(t+1)} \tag{8}$$

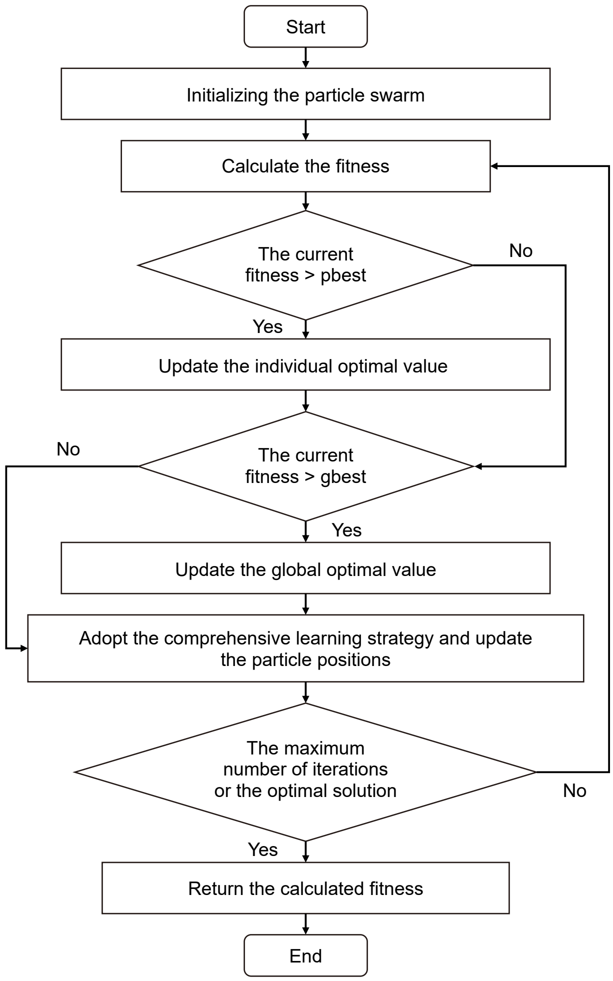


**Fig. 4.** CLPSO optimization workflow.

where $v_{i,d}^{(t)}$ and $x_{i,d}^{(t)}$ denote the velocity and position of the $i$-th particle in the $d$-th dimension at the $t$-th iteration, respectively; $\boldsymbol{v}_i^{(t+1)}$ and $\mathbf{x}_i^{(t+1)}$ represent the velocity and position vectors of particle $i$; $w$ is the inertia weight that balances exploration and exploitation; $c$ is the acceleration constant; $r_{i,d}$ is a random number within the range [0, 1] introducing stochasticity; and $\text{pbest}_{f_i(d),d}^{(t)}$ signifies the historical best position of the exemplar particle for the $d$-th dimension at iteration $t$.

## 3. Result and discussion

### 3.1. Description of the data set

The dataset for this study comprises field measurements collected from 72 horizontal wells in the Jimsar shale oilfield. To address common data quality issues—including missing values, outliers, and formatting inconsistencies—a comprehensive data preprocessing pipeline was implemented. This involved K-nearest neighbor (KNN) imputation and the $3\sigma$ (three-sigma) rule for outlier detection, ensuring a high-fidelity dataset for the training and iterative optimization of the deep learning models.

Oilfield production data encompasses a diverse range of parameters, many of which are highly interdependent and exhibit complex non-linear relationships with production performance. Given the typically limited sample sizes in oilfield applications, model performance often degrades as feature dimensionality increases. To ensure objective and comprehensive feature selection, this study employs an integrated multi-criteria feature importance evaluation approach. First, a feature importance matrix was constructed by integrating four distinct assessment techniques: the Pearson correlation coefficient, grey relational analysis (GRA), Shapley additive explanations (SHAP), and the AutoGluon feature importance module. Subsequently, the entropy weight method (EWM) was utilized to assign objective weights to the indicators within this evaluation matrix. EWM calculates weights based on the information entropy provided by each assessment method, effectively mitigating biases inherent in subjective weighting. Finally, the EWM-derived weights are applied to the importance matrix through weighted summation to generate a consolidated feature importance ranking, which informs the final dataset construction. The dataset comprises 33 parameters: geological, engineering, and production parameters. Specifically, it includes 7 logging spatial sequence data items (well diameter, shallow resistivity, formation resistivity, natural gamma, acoustic log, compensated neutron, and density data for each formation layer at various depths), 12 geological parameters, 13 engineering parameters, and 1 production parameter serving as the model label, as shown in Table 1.

Heterogeneous data distributions can introduce biases during model training, leading to insufficient feature extraction or overfitting. This study utilized stratified sampling to partition the training and test sets, ensuring representative samples across all data strata. This approach allows the model to capture a broader range of data characteristics, thereby enhancing its generalizability.

### 3.2. Model training

Hyperparameter optimization is essential for maximizing the predictive performance of neural network architectures. This study utilizes the Optuna framework to systematically determine the optimal model configuration. The optimization objective (objective function) is to minimize the mean squared error (MSE) on the test set. The defined search space comprehensively covers both architectural hyperparameters and training process parameters. Table 2 details the key hyperparameters and their respective search ranges. After 50 optimization trials, the tree-structured Parzen estimator (TPE) algorithm within Optuna efficiently explored the high-dimensional parameter space, ultimately converging on an optimal hyperparameter configuration.

Upon obtaining the optimal configuration, the final model was instantiated and retrained on the complete training dataset. During the final training phase, an early stopping mechanism was implemented to monitor validation loss, thereby preventing overfitting and ensuring the model achieved its optimal generalization performance. Fig. 5 illustrates the CNN-MLP architecture and the evolution of tensor dimensions through successive layers. The model was trained using the Adam stochastic optimizer with a predefined learning rate of 0.0003. Although the maximum training limit was set to 10,000 epochs, the process terminated at epoch 2822 due to the early stopping criterion.

Table 3 lists the specific hyperparameters utilized for the CLPSO algorithm. During the CLPSO optimization process, geological parameters are treated as fixed constraints. The search space for the engineering parameters is bounded by the 5th and 95th percentiles of the values observed in the training dataset. This approach mitigates the influence of outliers on the optimization and ensures that the particle search remains within a physically feasible solution space.

**Table 1**
Experimental parameters of radial borehole fracturing in multiple layers.

| Geological parameters | Engineering parameters | Formation abstract features | Production parameter |
|---|---|---|---|
| Lateral length, m | Actual stimulated length, m | Well diameter | 360 day cumulative oil production per unit lateral length |
| Reservoir thickness, m | Stage count | Shallow resistivity | |
| Class I reservoir, m | Cluster count | Formation resistivity | |
| Class II reservoir, m | Average stage/cluster spacing, m | Natural gamma | |
| Class III reservoir, m | Total proppant volume, $m^3$ | Acoustic log | |
| Maximum porosity, % | Total fluid volume, $m^3$ | Compensated neutron | |
| Average porosity, % | Fluid–sand ratio, % | Density | |
| Average oil saturation, % | Fluid intensity, $m^3/m$ | | |
| Porosity | Proppant intensity, $m^3/m$ | | |
| Oil saturation | Pad fluid proportion, % | | |
| Permeability, mD | Average sand ratio, % | | |
| Thickness, m | Minimum injection rate, $m^3/min$ | | |
| | Maximum injection rate, $m^3/min$ | | |

**Table 2**
Model hyperparameter search range. SGD: stochastic gradient descent.

| Model hyperparameters | Search scope | Hyperparameter values |
|---|---|---|
| The number of output channels in convolutional layers | [8, 16, 24] | 16 |
| The number of features in fully connected layers | [16, 32, 64] | 64 |
| Dropout rate | (0.3–0.5) | 0.3 |
| Learning rate | (0.0001–0.01) | 0.0001 |
| Optimizer type | [Adam, AdamW, SGD] | Adam |
| Weight decay | (0.00001–0.01) | 0.0084 |

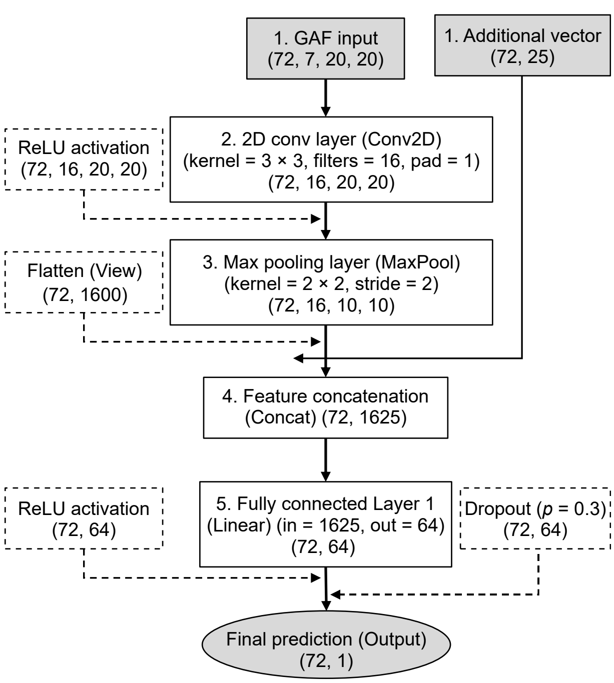


**Fig. 5.** CNN-MLP model architecture.

**Table 3**
CLPSO model parameter settings.

| Epoch | Pop_size | C_local | W_min & W_max | Max_flag |
|---|---|---|---|---|
| 100 | 35 | 15 | 0.5~1.5 | 5 |

Based on the production prediction model, the intelligent optimization framework defines the search space for fracturing parameters through technical and geological constraints, identifying the optimal parameter combination that maximizes production. The experimental results indicate that the CLPSO algorithm identified optimal fracturing configurations within an average of 100 iterations per well, requiring less than 30 min of computation time. The algorithm exhibited robust convergence and successfully avoided local optima to reach the global optimum.

### 3.3. Production prediction performance evaluation

To validate the predictive accuracy of the CNN-MLP model, an AutoGluon-based automated machine learning (AutoML) framework was implemented for benchmarking. To address the challenges associated with small-sample, high-dimensional regression, unstable neural network architectures were excluded, and the “high_quality” preset was utilized to leverage advanced base models and automated hyperparameter tuning. To mitigate overfitting, multi-layer stacking was disabled; instead, 7-fold cross-validation combined with 15-fold bagging was employed to enhance model stability and generalizability.

Fig. 6(a) and (b) illustrates the production forecast outcomes of the two models examined in this study. The precision of the CNN-MLP neural network model and the AutoGluon ensemble learning model was assessed using four metrics: RMSE, MAE, MAP, and $R^2$. The relevant metric parameters are consolidated in Table 4.

Experimental results demonstrate that the CNN-MLP model significantly outperforms the AutoGluon model across all evaluation metrics. In terms of prediction accuracy, the CNN-MLP model achieved RMSE, MAE, and MPE values of 1.23, 1.02, and 33.42 on the test set, representing reductions of 23.13%, 20.31%, and 42.02% compared to the AutoGluon model.

Analyzing goodness-of-fit and generalization capability, the CNN-MLP model achieved an $R^2$ of 0.91 on the training set, indicating strong explanatory power for the training data. Its $R^2$ on the test set was 0.68. Although this represents a decrease compared to the training set, suggesting some overfitting, it still demonstrates acceptable generalization performance. In contrast, the AutoGluon model achieved an $R^2$ of only 0.57 on the training set and a further

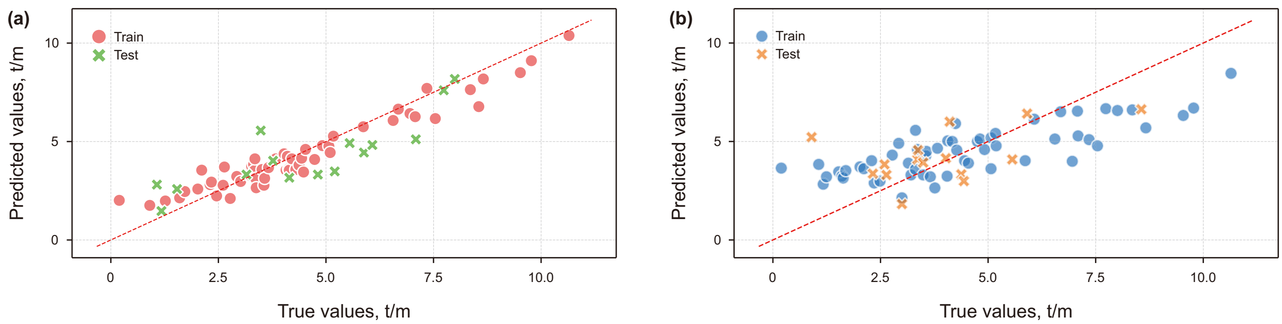


**Fig. 6.** Production prediction results of the CNN-MLP model and the ensemble framework. (**a**) CNN-MLP model; (**b**) AutoGluon ensemble framework.

**Table 4**
Parameter value evaluation.

| Model | Date set | Evaluation index | | | |
|---|---|---|---|---|---|
| | | RMSE | MAE | MPE | $R^2$ |
| CNN-MLP | Training set | 0.67 | 0.53 | 32.04 | 0.91 |
| | Testing set | 1.23 | 1.02 | 33.42 | 0.68 |
| AutoGluon | Training set | 1.58 | 1.29 | 68.97 | 0.57 |
| | Testing set | 1.60 | 1.28 | 57.64 | 0.16 |

decline to 0.16 on the test set, indicating poor fitting performance and virtually no generalization capability.

Overall, the CNN-MLP model demonstrates superior non-linear fitting capability and more stable predictive performance in this forecasting task. Despite rigorous hyperparameter optimization, the combination of a limited sample size and high feature dimensionality leads to a degree of persistent overfitting. Future research should prioritize expanding the dataset to further enhance the model's generalizability and robustness.

Experimental results indicate that the proposed convolutional neural network model, which incorporates logging data and engineering parameters, outperforms traditional ensemble methods in 360-day production forecasting, primarily due to its improved capability for multi-dimensional data co-characterization. The CNN-MLP adeptly integrates logging curve data, to accurately represent microscopic characteristics of reservoir heterogeneity and fluid states, thereby addressing the shortcomings of traditional engineering parameters in spatially continuous characterization. Furthermore, the sophisticated non-linear structure of the CNN-MLP is more proficient at capturing complex mapping associations within high-dimensional feature spaces.

### 3.4. Optimization results of fracturing parameters

This study assessed the effectiveness of parameter optimization using 10 wells from the Jimsar shale oilfield. The shale thickness in this region varies from 30 to 70 m, while the reservoir depths range from 2000 to 4000 m. Three production platforms in the oilfield were chosen, comprising six test wells (test wells 1–6) for the optimization of intelligent fracturing parameters, and four control wells (control wells 1–4) within the same block for conventional fracturing design as comparative trials. Fig. 7 illustrates the spatial arrangement of test and control wells, facilitating the assessment of shale oil production potential and extraction efficacy across various fracturing parameter design methodologies.

To examine the applicability of the intelligent fracturing parameter optimization design workflow under diverse geological settings, test groups were classified in Table 5 according to differences in well-specific geological parameters.

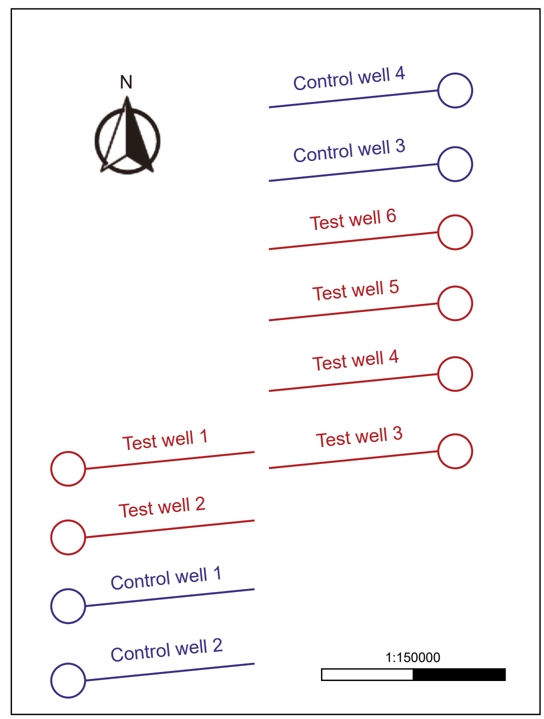


**Fig. 7.** Relative positions of test and control wells.

Fig. 8(a–d) illustrates four essential metrics of test and control wells across groups: reservoir thickness, thickness class I reservoir, porosity, and pressure coefficient. This grouping design facilitates the comparative validation of the optimization framework's adaptability to various reservoir conditions via controlled experiments.

The fracturing stage-cluster configuration for horizontal wells in this investigation was executed with a heritage stage-cluster optimization software created from prior research (Hu et al., 2023). Table 6 displays the stage-cluster division results for test groups 1 and 2.

Under equivalent geological settings, the actual stimulated lengths of the test and control groups were consistent, illustrating the stage-cluster optimization program's capacity to precisely adjust reservoir conditions based on formation data. Test groups demonstrated 4%–8% fewer stages than control groupings, with greater average stage spacing during stimulated intervals; elevated cluster counts and 2%–4% decreased cluster spacing facilitated concentrated stimulation to improve reservoir treatment efficacy and hydrocarbon mobility.

**Table 5**
Grouping arrangements of field trials.

| Test group | Test well No. | Control well No. | Geological conditions |
|---|---|---|---|
| 1 | Test well 1 | Control well 1 | Identical |
| 2 | Test well 2 | Control well 2 | Identical |
| 3 | Test wells 3 and 6 | Control well 3 | Inferior conditions in test wells |
| 4 | Test wells 4 and 5 | Control well 4 | Inferior conditions in test wells |

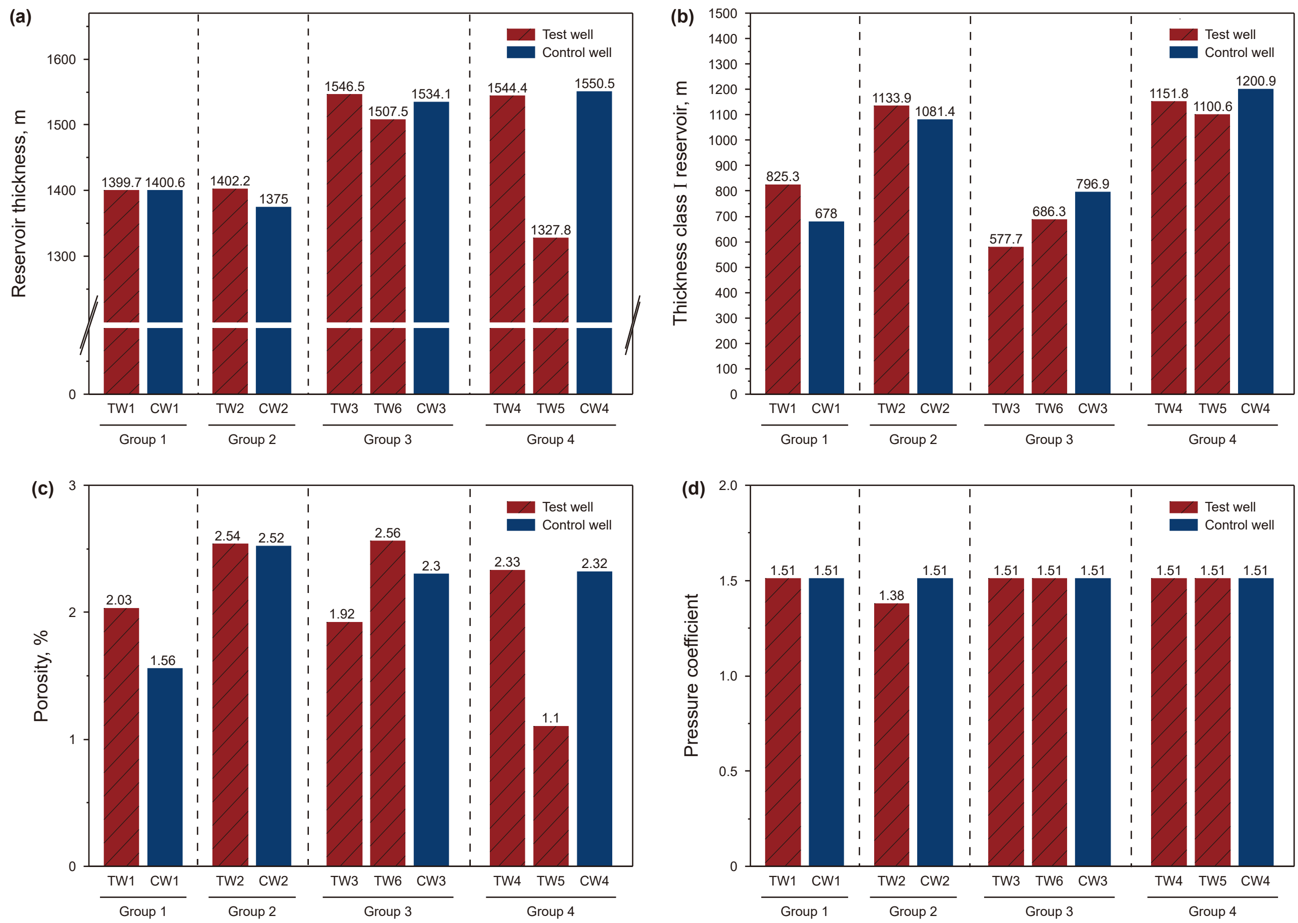


**Fig. 8.** Differences in single-well geological conditions across four test groups. (**a**) Reservoir thickness; (**b**) thickness class I reservoir; (**c**) porosity; (**d**) pressure coefficient.

**Table 6**
Stage and cluster division results for test groups 1 and 2. TW: test well; CW: control well.

| Target parameters | Group 1 | | Group 2 | |
|---|---|---|---|---|
| | TW1 | CW1 | TW2 | CW2 |
| Horizontal section length | 1401 | 1401 | 1402 | 1386 |
| Number of sections | 27 | 28 | 28 | 30 |
| Number of clusters | 159 | 159 | 180 | 174 |
| Cluster spacing | 8.6 | 8.8 | 7.7 | 8 |

The optimization of the fracturing parameter for test groups 1 and 2 was performed using the suggested framework, with the results depicted in Fig. 9(a–d). Compared to the control wells, the injected fluid volume, proppant dosage, and proppant strength parameters were all increased, while the fluid–sand ratio remained essentially unchanged. This design aims to effectively support the fracture network by increasing both fracturing fluid volume and proppant dosage, thereby enhancing fracture propagation. This approach addresses the challenges posed by the Jimsar shale reservoir's significant burial depth and high in-situ stress, promoting efficient development.

Table 7 presents the stage-cluster optimization findings for test groups 3 and 4. The test wells demonstrated real stimulated lengths that were 5%–16% longer than the control wells, despite inferior formation conditions, while maintaining similar cluster spacing, fewer phases, and lower cluster counts. This illustrates that the stage-cluster optimization algorithm emphasizes prolonged stimulated lengths to improve hydrocarbon flow paths in adverse geological conditions. The design efficiently avoids high-stress areas in the reservoir and focuses on optimal engineering locations during stage-cluster planning, enhancing fracturing operating efficiency and hydrocarbon recovery.

The optimization of the fracturing parameter for test groups 3 and 4 was executed using the proposed procedure, with the results illustrated in Fig. 10(a–d). The parameters of proppant volume, injected fluid volume, proppant intensity, and fluid–proppant ratio were markedly elevated in test wells compared to control wells,

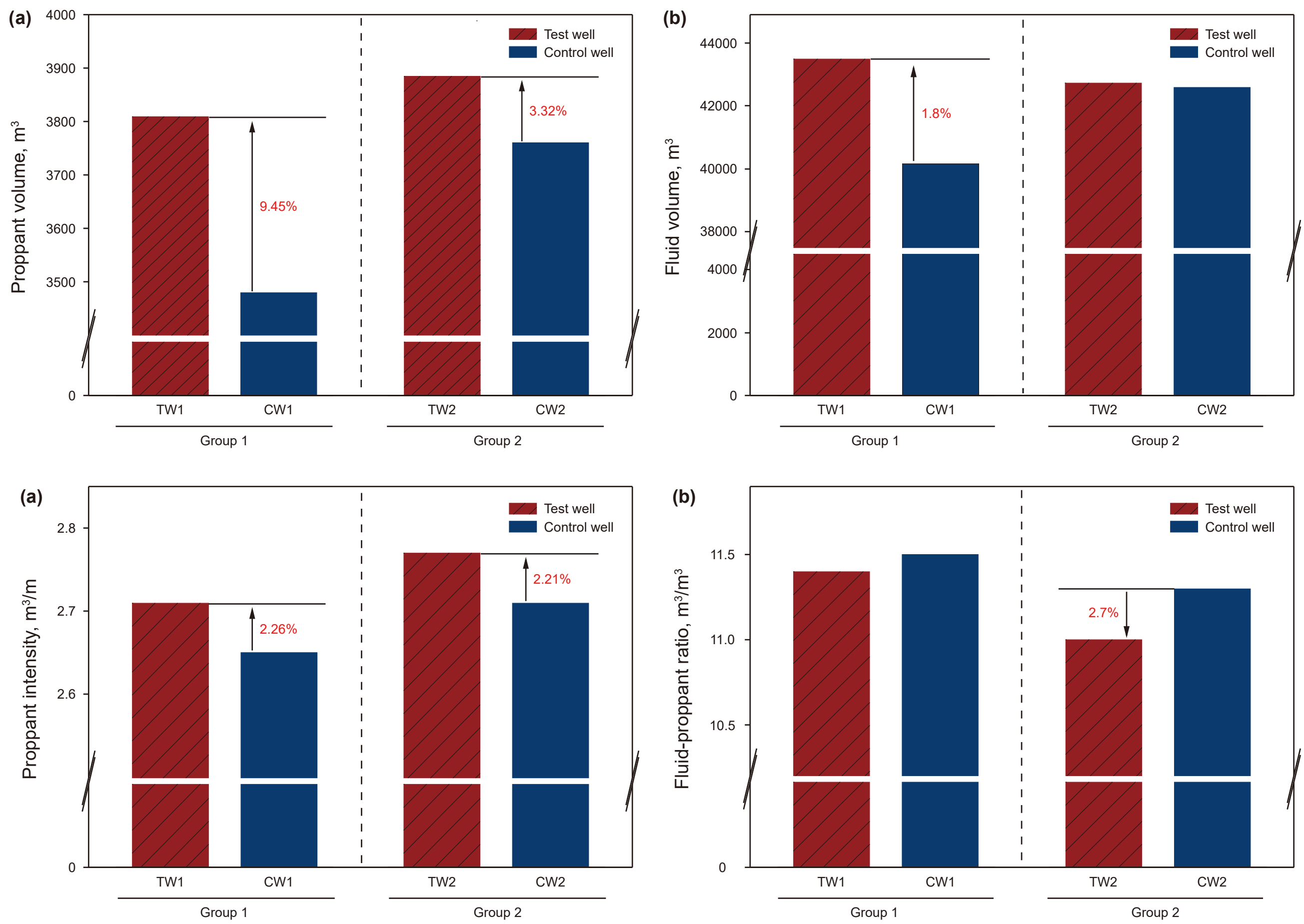


**Fig. 9.** Optimization results of operational parameters for test groups 1 and 2. (**a**) Proppant volume; (**b**) fluid volume; (**c**) proppant intensity; (**d**) fluid–proppant ratio.

**Table 7**
Stage and cluster division results for test groups 3 and 4. TW: test well; CW: control well.

| Target parameters | Group 3 | | | Group 4 | | |
|---|---|---|---|---|---|---|
| | TW3 | TW6 | CW3 | TW4 | TW5 | CW4 |
| Horizontal section length | 1656 | 1682 | 1444 | 1613 | 1610 | 1543 |
| Number of sections | 27 | 27 | 30 | 27 | 28 | 33 |
| Number of clusters | 174 | 163 | 173 | 168 | 178 | 182 |
| Cluster spacing | 8.73 | 9.96 | 8.5 | 8.61 | 8.4 | 8.5 |

consistent with findings from test groups 1 and 2 and confirming the reliability of the parameter optimization process. Considering the significant burial depth and restricted fracture formation in the Jimsar shale oilfield, the intelligent model is presumably designed to facilitate effective fracture network propagation and provide stable propping by augmenting the utilization of fracturing fluid and proppant.

### 3.5. Performance assessment of optimization

Test wells 1 and 2 and control wells 1 and 2 commenced production after 185 and 170 days post-completion, respectively. Fig. 11(a–c) illustrates that test wells demonstrated an average daily oil production and oil production per 100 m that were 6.6% and 16.2% greater, respectively, than those of control wells during the same timeframe. Test wells recorded an average daily production of 15.22 t/d, a cumulative output of 200.79 t per 100 m, and a pressure decline rate of 0.13 MPa/d, whereas control wells produced 14.28 t/d, 172.91 t, and 0.15 MPa/d, respectively.

Test wells 3–6 and control wells 3 and 4 commenced production after 138 and 129 days post-completion, respectively. Fig. 12(a–c) illustrates that test wells demonstrated an 18.1% and 19.4% increase in average daily oil production and oil production per 100 m, respectively, compared to control wells during the same timeframe. Test wells recorded an average daily production of 18.01 t/d, a cumulative output of 154.83 t per 100 m, and a pressure decline rate of 0.08 MPa/d, whereas control wells produced 15.25 t/d, 129.64 t, and 0.12 MPa/d.

Comparative trials revealed that test groups displayed considerable benefits in actual production. Increased average daily oil output and cumulative production per 100 m signify enhanced productivity from the test group parameter design, whilst reduced pressure decrease rates imply more stable production processes that may maintain long-term productivity.

The parameter optimization in test groups emphasized increased proppant volume, fluid volume, and proppant intensity. In shale oil development, when spontaneous fracture formation is constrained, artificially propped cracks are essential. Increased proppant volume and intensity facilitate the creation of broader, longer, and more stable propped fracture networks during fracturing. In thin reservoirs characterized by low porosity and high fluid viscosity, augmented proppant application effectively expands rock pores to improve flow conductivity; within the deeply

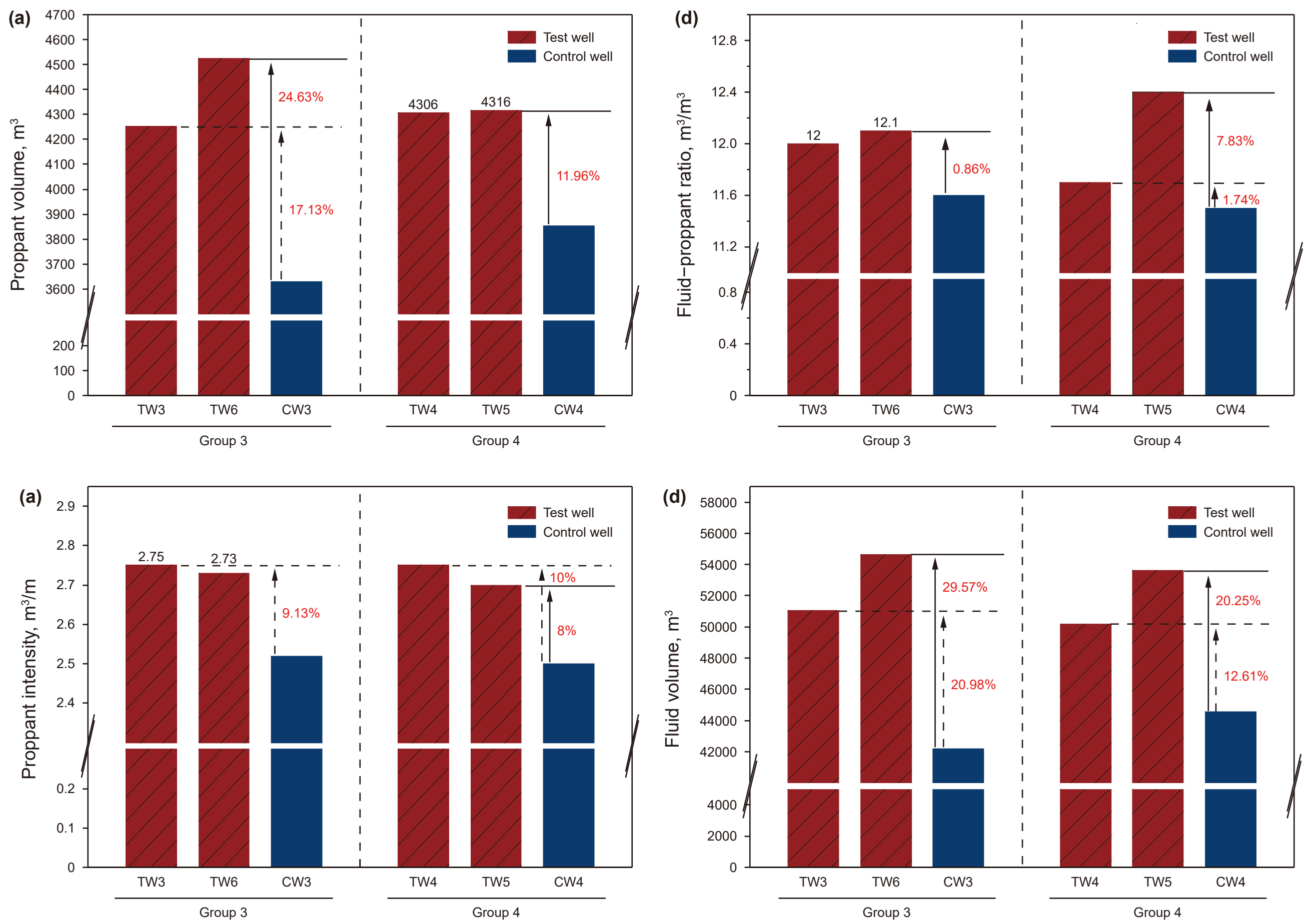


**Fig. 10.** Optimization results of operational parameters for test groups 3 and 4. (**a**) Proppant volume; (**b**) fluid volume; (**c**) proppant intensity; (**d**) fluid–proppant ratio.

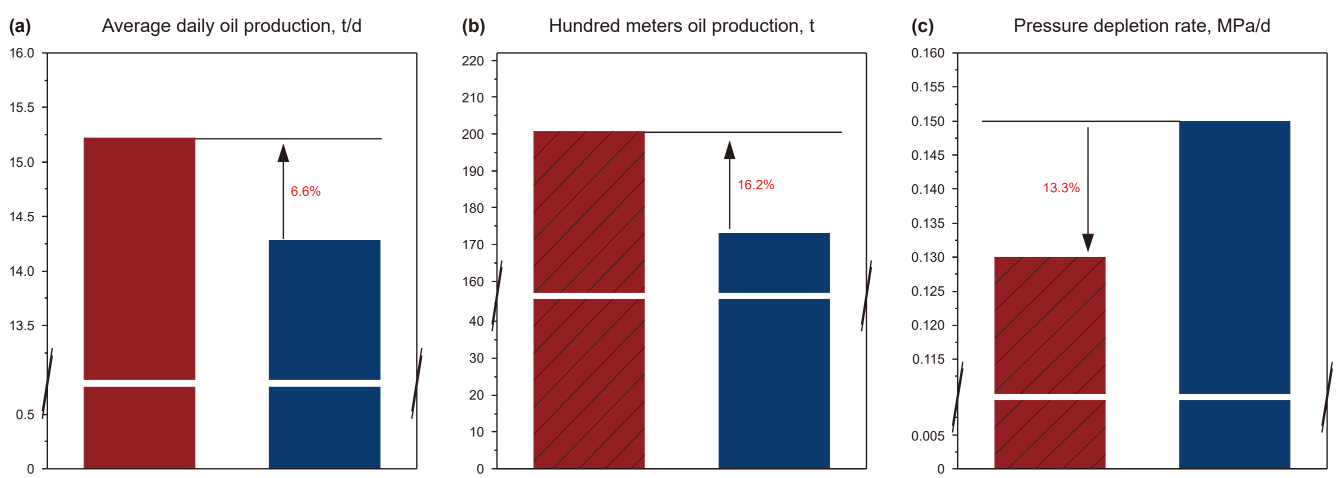


**Fig. 11.** Production performance of test and control wells in groups 1 and 2. (**a**) Average daily oil production; (**b**) oil production per 100 m; (**c**) pressure depletion rate.

buried Jimsar shale oil reservoir, increased fluid volume aids in overcoming in-situ stress to initiate fractures in deeper strata, transmits fracturing energy to deeper formations for prolonged fracture propagation and stimulated reservoir volume, and links additional reservoir areas. Moreover, it boosts fracture network connectivity by cleansing fracture surfaces, diminishes hydrocarbon flow resistance, and improves extraction efficiency.

## 4. Limitations and future work

While this study offers novel insights into the intelligent optimization of fracturing parameters through multi-source data fusion, several limitations must be acknowledged. A primary constraint is the limited sample size. Although 72 samples are sufficient for initial model development, the CNN-MLP

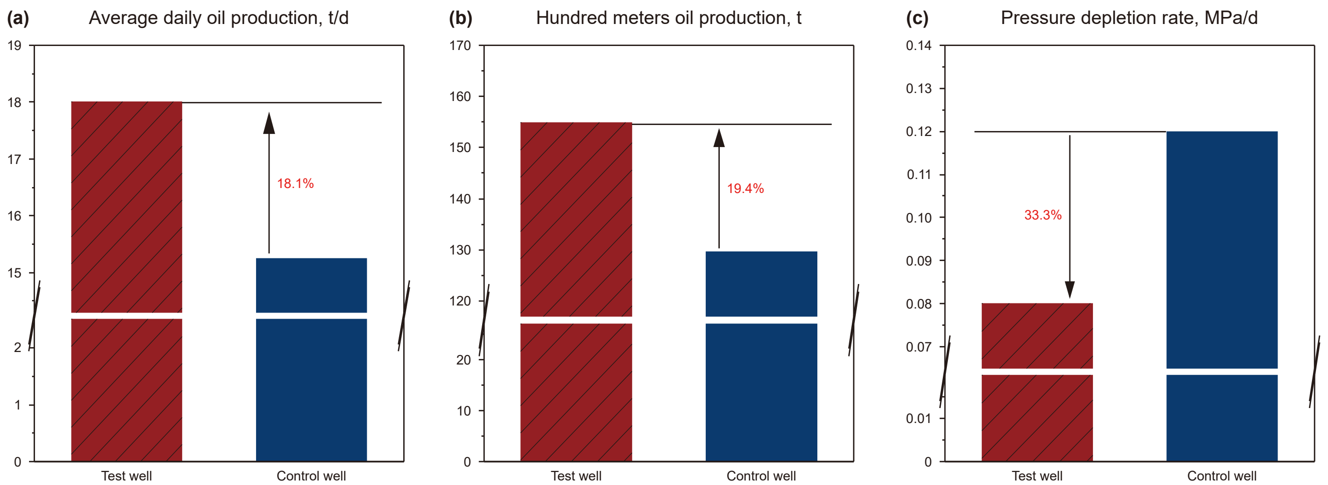


**Fig. 12.** Production performance of test and control wells in groups 3 and 4. (**a**) Average daily oil production; (**b**) oil production per 100 m; (**c**) pressure depletion rate.

architecture exhibits minor overfitting and a systematic "central tendency bias"—specifically, overestimating low-yield wells and underestimating high-yield ones. This likely stems from high model complexity relative to the available data, which can impede generalization performance. Furthermore, the sparse representation of extreme values (high- and low-yield samples) prevents the model from capturing sufficient variance, causing predictions to regress toward the mean. Additionally, the model was trained and validated exclusively on data from the Jimsar shale oilfield; thus, its transferability to other shale reservoirs with distinct geological characteristics requires further verification. Future research should prioritize the acquisition of multi-basin data to enhance the model's predictive precision, generalizability, and cross-field applicability.

Furthermore, the current field trials were conducted across six test wells and four control wells. While preliminary findings are promising, the limited sample size constrains the statistical significance of the resulting conclusions. Consequently, broader field validation is necessary to further substantiate the robustness and industrial scalability of the optimization framework.

## 5. Conclusions

This paper presents an optimization approach for fracturing design parameters that integrates a CNN (convolutional neural network)-MLP (multi-layer perceptron) neural network model with the CLPSO (comprehensive learning particle swarm optimization) algorithm. The neural network utilizes the GAF (Gramian angular field) technique to convert logging data, discern abstract formation characteristics, and integrates geological and operational variables for precise production forecasting, surpassing the predictive accuracy of machine learning ensemble models. The CLPSO algorithm modifies operational parameters within geological limits to forecast optimal production and identify the most effective parameter combinations. Field applications in the Jimsar shale oil reservoir exhibit considerable efficacy. The primary conclusions are as follows.

(1) In comparison to the AutoGluon machine learning ensemble framework, the CNN-MLP model exhibits enhanced performance in predicting unconventional hydrocarbon single-well production by proficiently recognizing formation characteristics and adjusting to mapping relationships within high-dimensional features, thereby increasing prediction accuracy by 20%.

(2) Within the geological context of the Jimsar shale reservoir, a fracturing design employing high proppant volume, high fracturing fluid volume, and high proppant strength facilitates the propagation of primary and secondary fractures, forming a complex fracture network. This ensures effective proppant support and sufficient hydrocarbon flow pathways. Consequently, it enhances flow capacity near the wellbore, enabling sustained and stable production while boosting output.

(3) Field test results from 138 to 185 days at the Jimsar shale oilfield demonstrate that test wells utilizing intelligent fracturing design optimization technology achieved a 6.6% to 18.1% increase in average daily oil production compared to control wells employing conventional fracturing designs during the same period. Cumulative production per 100 m increased by 16.2% to 19.4%, while pressure decline rates decreased by 13% to 38.5%. This approach significantly enhances short-term production capacity per well, while the reduced pressure decline rate also improves the long-term sustainable production potential of individual wells. However, this model has only been applied to the Jimsar shale oilfield, and its applicability to other shale oil reservoirs requires further validation.

The proposed optimization framework for fracturing design parameters, which integrates neural network models with the CLPSO method, exhibits improved adaptation to the unfavorable geological conditions of the Jimsar shale oil reservoir, characterized by restricted natural fracture development and thin oil layers. This technique tackles unconventional hydrocarbon fracturing design issues, enhancing practical engineering decision-making through concrete implementation strategies.

### CRediT authorship contribution statement

**Jie Li:** Writing – review & editing, Writing – original draft, Methodology. **Quan-Zhen Xiu:** Writing – review & editing, Writing – original draft, Investigation. **Xiao-Dong He:** Writing – review & editing, Writing – original draft. **Gen-Sheng Li:** Validation, Resources, Project administration, Funding acquisition. **Chang Li:** Project administration, Investigation, Funding acquisition, Data curation. **Mao-Ya Xu:** Writing – review & editing, Writing – original draft, Methodology. **Tian-Xiang Zhou:** Writing – review & editing, Writing – original draft. **Shou-Ceng Tian:**

Supervision, Project administration, Funding acquisition. **Tian-Yu Wang:** Supervision, Resources, Project administration.

## Conflict of interest

No conflict of interest exists in the submission of this manuscript, and the manuscript is approved by all authors for publication.

## Acknowledgements

The authors gratefully acknowledge the financial support received from the International (Regional) Cooperation and Exchange Projects of China (No. 52320105002).

## References

Al-Attar, H., Alshadafan, H., Al Kaabi, M., et al., 2020. Integrated optimum design of hydraulic fracturing for tight hydrocarbon-bearing reservoirs. J. Petrol. Explor. Prod. Technol. 10 (8), 3347–3361. https://doi.org/10.1007/s13202-020-00990-6.

Barati, R., Liang, J., 2014. A review of fracturing fluid systems used for hydraulic fracturing of oil and gas wells. J. Appl. Polym. Sci. 131 (16), 40735. https://doi.org/10.1002/app.40735.

Bhattacharya, S., Ghahfarokhi, P.K., Carr, T.R., et al., 2019. Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiber-optic, completions, and surface data: A case study from the Marcellus shale, North America. J. Pet. Sci. Eng. 176, 702–715. https://doi.org/10.1016/j.petrol.2019.01.013.

Chahar, J., Verma, J., Vyas, D., et al., 2022. Data-driven approach for hydrocarbon production forecasting using machine learning techniques. J. Pet. Sci. Eng. 217, 110757. https://doi.org/10.1016/j.petrol.2022.110757.

Chen, G., Tian, H., Xiao, T., et al., 2024. Time series forecasting of oil production in enhanced oil recovery system based on a novel CNN-GRU neural network. Geoenergy Sci. Eng. 233, 212528. https://doi.org/10.1016/j.geoen.2023.212528.

Clark, P.E., 2010. Analysis of fluid loss data II: models for dynamic fluid loss. J. Pet. Sci. Eng. 70 (3), 191–197. https://doi.org/10.1016/j.petrol.2009.11.010.

Clerc, M., Kennedy, J., 2002. The particle swarm-explosion, stability, and convergence in a multidimensional complex space. IEEE Trans. Evol. Comput. 6 (1), 58–73. https://doi.org/10.1109/4235.985692.

de Oliveira Werneck, R., Prates, R., Moura, R., et al., 2022. Data-driven deep-learning forecasting for oil production and pressure. J. Pet. Sci. Eng. 210, 109937. https://doi.org/10.1016/j.petrol.2021.109937.

Deb, K., Pratap, A., Agarwal, S., et al., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6 (2), 182–197. https://doi.org/10.1109/4235.996017.

Fukushima, K., 1980. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36 (4), 193–202. https://doi.org/10.1007/BF00344251.

Gong, X., Ma, X., Liu, Y., 2024. Analysis of geological factors affecting propagation behavior of fracture during hydraulic fracturing shale formation. Geomech. Geophys. Geo-energ Geo-resour. 10 (1), 102. https://doi.org/10.1007/s40948-024-00819-0.

Hernandez-Perez, L.G., Lira-Barragan, L.F., Ponce-Ortega, J.M., 2020. Hybrid multiobjective optimization using deterministic and metaheuristic techniques for flowback water reusing in hydraulic fracturing processes. Ind. Eng. Chem. Res. 59 (34), 15298–15308. https://doi.org/10.1021/acs.iecr.0c02867.

Hsu, M., Ma, Z., Tian, S., et al., 2024. Geological feature extraction and data-driven production prediction based on well logging data. In: ARMA 58th U.S. Rock Mechanics/Geomechanics Symposium. In: https://onepetro.org/ARMAUSRMS/proceedings/ARMA24/ARMA24/D042S058R010/549423.

Hu, S., Sheng, M., Shi, S., et al., 2023. Optimization of fracturing stages/clusters in horizontal well based on unsupervised clustering of bottomhole mechanical specific energy on the bit. Nat. Gas. Ind. B. 10 (6), 583–590. https://doi.org/10.3787/j.issn.10000976.2023.07.006.

Hyman, J.D., Jimenez-Martinez, J., Viswanathan, H.S., et al., 2016. Understanding hydraulic fracturing: A multi-scale problem. Philos. Trans. R. Soc. A. 374 (2078), 20150426. https://doi.org/10.1098/rsta.2015.0426.

Ismail, A., Azadbakht, S., 2024. A comprehensive review of numerical simulation methods for hydraulic fracturing. Int. J. Numer. Anal. Methods GeoMech. 48 (5), 1433–1459. https://doi.org/10.1002/nag.3689.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. Proc. ICNN'95-Int. Conf. Neural Netw. 4, 1942–1948. https://doi.org/10.1109/ICNN.1995.488968.

Lecun, Y., Bottou, L., Bengio, Y., et al., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324. https://doi.org/10.1109/5.726791.

Li, G., Song, X., Tian, S., et al., 2022a. Intelligent drilling and completion: A review. Engineering 18, 33–48. https://doi.org/10.1016/j.eng.2022.07.014.

Li, Y., Huang, Y., Ma, H., et al., 2022b. Study on the development options of tight sandstone oil reservoirs and their influencing factors. Front. Energy Res. 10, 1007224. https://doi.org/10.3389/fenrg.2022.1007224.

Li, W., Zhang, T., Liu, X., et al., 2024. Machine learning-based fracturing parameter optimization for horizontal wells in Panke field shale oil. Sci. Rep. 14 (1), 6046. https://doi.org/10.1038/s41598-024-56660-8.

Liang, J.J., Qin, A.K., Suganthan, P.N., et al., 2006. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. IEEE Trans. Evol. Comput. 10 (3), 281–295. https://doi.org/10.1109/TEVC.2005.857610.

Liu, Y.Y., Ma, X.H., Zhang, X.W., et al., 2021. A deep-learning-based prediction method of the estimated ultimate recovery (EUR) of shale gas wells. Pet. Sci. 18 (5), 1450–1464. https://doi.org/10.1016/j.petsci.2021.08.007.

Lu, C., Jiang, H., Qu, S., et al., 2022. Hydraulic fracturing design for shale oils based on sweet spot mapping: A case study of the Jimusar formation in China. J. Pet. Sci. Eng. 214, 110568. https://doi.org/10.1016/j.petrol.2022.110568.

Ma, Z., Weng, J., Zhang, J., et al., 2025. Intelligent prediction of rate of penetration through meta-learning and data augmentation synergy under limited sample. Geoenergy Sci. Eng. 250, 213818. https://doi.org/10.1016/j.geoen.2025.213818.

Mahzari, P., Emambakhsh, M., Temizel, C., et al., 2022. Oil production forecasting using deep learning for shale oil wells under variable gas-oil and water-oil ratios. Petrol. Sci. Technol. 40 (4), 445–468. https://doi.org/10.1080/10916466.2021.2001526.

Morozov, A.D., Popkov, D.O., Duplyakov, V.M., et al., 2020. Data-driven model for hydraulic fracturing design optimization: Focus on building digital database and production forecast. J. Pet. Sci. Eng. 194, 107504. https://doi.org/10.1016/j.petrol.2020.107504.

Muther, T., Syed, F.I., Dahaghi, A.K., et al., 2022. Socio-inspired multi-cohort intelligence and teaching-learning-based optimization for hydraulic fracturing parameters design in tight formations. J. Energy Resour. Technol. 144 (7), 073201. https://doi.org/10.1115/1.4052182.

Nouri, D.E., Benyounes, K., Smail, F., et al., 2025. A comprehensive analysis of innovations to overcome challenges in horizontal well hydraulic fracturing in tight reservoirs: a case study on the lower Devonian in Berkine Basin, Algeria. J. Petrol. Explor. Prod. Technol. 15 (10), 159. https://doi.org/10.1007/s13202-025-02024-5.

Pang, X., Li, M., Li, B., et al., 2023. Main controlling factors and movability evaluation of continental shale oil. Earth Sci. Rev. 243, 104472. https://doi.org/10.1016/j.earscirev.2023.104472.

Sheng, G.L., Zhao, H., Ma, J.L., et al., 2023. A new approach for flow simulation in complex hydraulic fracture morphology and its application: Fracture connection element method. Pet. Sci. 20 (5), 3002–3012. https://doi.org/10.1016/j.petsci.2023.03.026.

Wang, Z., Oates, T., 2015. Imaging time-series to improve classification and imputation. arXiv preprint arXiv:1506.00327. https://doi.org/10.48550/arXiv.1506.00327.

Wang, H., Qiao, L., Zhang, J., et al., 2022. An effective integration optimization algorithm for regional fracturing design and drilling placement. J. Nat. Gas Sci. Eng. 101, 104505. https://doi.org/10.1016/j.jngse.2022.104505.

Yang, R., Qin, X., Liu, W., et al., 2022. A physics-constrained data-driven workflow for predicting coalbed methane well production using artificial neural network. SPE J. 27 (3), 1531–1552. https://doi.org/10.2118/205903-PA.

Yao, M., Yang, H., Xia, Y., et al., 2025. Optimization of multiple hydraulic fracture initiation and propagation in deep tight reservoirs considering fracture macro/micro characteristics. Int. J. Rock Mech. Min. Sci. 191, 106126. https://doi.org/10.1016/j.ijrmms.2025.106126.

Zhang, S., Chen, M., Ma, X., et al., 2021. Research progress and development direction of hydrofracturing design models. Xinjiang Oil Gas 17 (3), 67–73 (in Chinese).

Zhang, D., Zhou, Z., Gao, C., et al., 2025. A novel hybrid PD-FEM-FVM approach for simulating hydraulic fracture propagation in saturated porous media. Comput. Geotech. 177, 106821. https://doi.org/10.1016/j.compgeo.2024.106821.

Zhao, Y.L., Xiao, Q.Y., Zeng, X.J., et al., 2025. Unconventional oil production forecasting based on PiAM meta-learning. Pet. Sci. https://doi.org/10.1016/j.petsci.2025.12.015.

Zhu, R., Li, N., Duan, Y., et al., 2025. Well-production forecasting using machine learning with feature selection and automatic hyperparameter optimization. Energies 18 (1), 99. https://doi.org/10.3390/en18010099.