Petroleum Science 21 (2024) 535–550

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

A hybrid machine learning optimization algorithm for multivariable pore pressure prediction

Song Deng [a], Hao-Yu Pan [a], Hai-Ge Wang [b], Shou-Kun Xu [a, *], Xiao-Peng Yan [a], Chao-Wei Li [a], Ming-Guo Peng [a], Hao-Ping Peng [a], Lin Shi [a], Meng Cui [c], Fei Zhao [c]

[a] School of Petroleum and Natural Gas Engineering, Changzhou University, Changzhou, 213164, Jiangsu, China
[b] Tubular Goods Research Institute of CNPC, Beijing, 102249, China
[c] CNPC Engineering Technology R&D Company Limited, Beijing, 102206, China

ARTICLE INFO

Article history:
Received 27 December 2022
Received in revised form
29 May 2023
Accepted 4 September 2023
Available online 9 September 2023

Edited by Jia-Jia Fei

Keywords:
Pore pressure
Grey wolf optimization
Multilayer perceptron
Effective stress
Machine learning

ABSTRACT

Pore pressure is essential data in drilling design, and its accurate prediction is necessary to ensure drilling safety and improve drilling efficiency. Traditional methods for predicting pore pressure are limited when forming particular structures and lithology. In this paper, a machine learning algorithm and effective stress theorem are used to establish the transformation model between rock physical parameters and pore pressure. This study collects data from three wells. Well 1 had 881 data sets for model training, and Wells 2 and 3 had 538 and 464 data sets for model testing. In this paper, support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB), and multilayer perceptron (MLP) are selected as the machine learning algorithms for pore pressure modeling. In addition, this paper uses the grey wolf optimization (GWO) algorithm, particle swarm optimization (PSO) algorithm, sparrow search algorithm (SSA), and bat algorithm (BA) to establish a hybrid machine learning optimization algorithm, and proposes an improved grey wolf optimization (IGWO) algorithm. The IGWO-MLP model obtained the minimum root mean square error (RMSE) by using the 5-fold cross-validation method for the training data. For the pore pressure data in Well 2 and Well 3, the coefficients of determination (R^2) of SVM, RF, XGB, and MLP are 0.9930 and 0.9446, 0.9943 and 0.9472, 0.9945 and 0.9488, 0.9949 and 0.9574. MLP achieves optimal performance on both training and test data, and the MLP model shows a high degree of generalization. It indicates that the IGWO-MLP is an excellent predictor of pore pressure and can be used to predict pore pressure.

1. Introduction

Pore pressure, also known as formation pore pressure, is the force on the pore fluid in a formation (Azadpour et al., 2015). Predicting pore pressure is critical in oil and gas development (Jorden and Shirley, 1966). Accurate pore pressure prediction results can help avoid the problems associated with drilling operations, reduce the cost of drilling operations, and improve the safety and efficiency of oil and gas development (Li et al., 2022). Traditional methods for calculating pore pressure can be divided into pre-drilling prediction, monitoring during drilling, and post-drilling evaluation. The post-drilling evaluation model usually has the highest compliance accuracy for pore pressure (Hottman and Johnson, 1965). Many scholars have established a nonlinear relationship model between rock physical parameters and effective stress and obtained pore pressure by effective stress theorem. Eaton (1972) developed a mathematical model for predicting pore pressure using mudstone properties obtained from logging data by the fitting method. This method is the most widely used method to calculate pore pressure. This is a suitable method for the calculation of pore pressure caused by under-compaction (Bektas et al., 2015). Bowers (1995) proposed the relationship between velocity and effective stress based on stress experiments, which can be divided into loading and unloading curves. This method is not only suitable for under-compaction, but also for abnormal pore pressure caused by fluid expansion. Therefore, this model has been applied in many types of research and has obtained a high accuracy in the calculation of pore pressure (Czerniak et al., 2017). Atashbari et al. (2012) derived the

* Corresponding author.
E-mail address: s21040820033@smail.cczu.edu.cn (S.-K. Xu).

https://doi.org/10.1016/j.petsci.2023.09.001

formula of pore pressure relying on compression coefficient and porosity under the condition of compaction. However, this model only considers the under-compaction case and does not solve other calculation methods for the cause of abnormal pore pressure. These models have many empirical parameters, limited ability to generalize, and a time-consuming correction process (Kuang et al., 2021). Moreover, the univariate calculation model of pore pressure does not take into account the physical properties of different rocks, and there are many restrictive conditions in use.

The traditional pore pressure prediction model generally only uses acoustic velocity as its characteristic, and does not take into account that acoustic velocity will be affected to some extent when encountering specific structures and lithology, so there are some limitations. A multivariable model is considered to be able to better reflect the variation of effective stress under sand mudstone formation, thus obtaining higher accuracy of pore pressure calculation. Eberhart-phillips et al. (1989) and Sayers et al. (2003) suggested acoustic velocity models based on previous studies and established the relationship between acoustic velocity and shale content, porosity, and effective stress. Because the multivariable model considers the influence of various rock properties, it is more suitable for predicting pore pressure. However, the establishment of a traditional multivariate nonlinear model is complicated, and it is difficult to obtain the best empirical relationship model. The machine learning algorithm is excellent in nonlinear fitting problems, so it is suitable to build pore pressure prediction models. However, there are few studies on predicting pore pressure with machine learning techniques (Xu et al., 2010). Since the machine learning algorithm has a powerful nonlinear fitting ability, this paper establishes a nonlinear relationship model between these characteristic parameters and effective stress through the machine learning algorithm. Finally, the effective stress theorem Terzaghi et al. (1996) proposed is employed to calculate the pore pressure.

With the development of artificial intelligence technology and the increase in the processing power of computers, machine learning technology has advanced with the times. It has significant advantages for solving complex nonlinear problems and is widely employed in oil (Moazzeni et al., 2015). Machine learning algorithms have been widely used in petroleum engineering. In petrophysical and geomechanical, machine learning can make predictions based on learning from already explored and developed reservoirs, their rock properties, and cross-water fluid flow behavior in different situations (Syed et al., 2022a, b). In artificial lift optimization systems, machine learning has been used in artificial lift selection, their predictive maintenance, and equipment malfunctioning detection (Syed et al., 2022a, b). In addition, petroleum engineering is a comprehensive discipline that relies on a variety of physical laws, and integrating physical laws into machine learning can effectively solve the shortcomings of insufficient interpretability of data-driven models. As a machine learning approach that combines physical laws, physics-informed machine learning allows the integration of physical laws in the form of partial differential equations into loss functions of machine learning, thus limiting the training of complex problems based on physical, experimental, and mathematical boundaries (Muther et al., 2022). This approach ensures that the model developed conforms to physics and may be widely used in petroleum engineering in the future.

Some scholars have established direct or indirect models of pore pressure by using machine learning algorithms based on features that can represent the pore pressure state. Hu et al. (2013) constructed a new neural network framework to use this feedforward backpropagation artificial neural network to predict pore pressure. The characteristic of this structure is that the hidden layer has two layers. The inputs to the first layer of their model are gamma rays and formation density, and the inputs to the second layer are output and acoustic time difference, formation density, and depth. In their study, this new network structure predicted pore pressure with more than twice the accuracy of conventional neural networks. Rashidi and Asadi (2018) selected mechanical-specific energy and drilling efficiency that were highly correlated with pressure difference through theoretical discussion. In this study, the artificial neural network model was used to establish the model of these two parameters and pore pressure, which obtained a good performance in the Iranian sandstone formation. However, this model does not take into account the wear and hydraulic parameters of the bit, which may lead to large errors in soft or hard formations and can be further improved. Yu et al. (2020) used acoustic velocity, shale content, and porosity as inputs to offshore exploration well data to predict effective stress. This study compared four models of RF, gradient boosting decision tree, SVM, and MLP, and found that the RF model has the highest prediction accuracy for effective stress. Their study used pore pressure data from the normal compaction interval. The trained model can be applied to high-pressure formations caused by under-compaction, but those caused by fluid expansion need to be modified by the unloading index. Huang et al. (2022) compared MLP, radial basis function neural networks, SVM, RF, and gradient boosting decision tree, and found that MLP with acoustic velocity, shale content, porosity, and density as input perform best in pore pressure prediction. Zhang et al. (2022) used the feature analysis method to select 9 features among 12 characteristic parameters to characterize pore pressure. Their study compared artificial neural networks, decision trees, RF, and SVM algorithms and found that the decision tree algorithm had the best performance in predicting pore pressure in three wells in the Middle East Oilfield. According to the authors, this is the first time anyone has applied the decision tree algorithm to the prediction of pore pressure. However, in the study of the machine learning model, they did not try to use an intelligent optimization algorithm to improve the performance of the model.

Hyperparameters are the critical parameters of the machine learning model. Grid search is a traditional approach for identifying the optimal hyperparameters of machine learning models (Fayed and Atiya, 2019). This method combines all the hyperparameters through exhaustive attempts and compares the prediction accuracy in the machine learning model. This method is extremely time-consuming. The intelligent optimization algorithm is a guided random search method, which can obtain an approximate optimal solution in a larger search space in less time (Guria et al., 2014). Therefore, an intelligent optimization algorithm improves the machine learning model (Fu and Wen, 2018). Grey wolf optimizer is an algorithm inspired by the hunting behavior of grey wolves (Mirjalili et al., 2014). The optimal hyperparameters of the model are determined quickly and efficiently, making it suitable for optimizing the machine learning model. This paper presents an IGWO algorithm and tests it with GWO, PSO, SSA, and BA on eight benchmark functions. The results show that the IGWO algorithm has the best performance on the seven test functions. In addition, the IGWO algorithm is improved compared with the GWO algorithm in all test functions.

This paper combines these intelligent optimization algorithms with machine learning algorithms to form multiple hybrid machine learning optimization algorithms. The performance of these models was trained and tested by collecting data from three wells. The results show that the IGWO algorithm has the best optimization performance among all intelligent optimization algorithms. The MLP model performed better than the SVM, RF, and XGB models in Well 2 and Well 3 for testing. The results show that the IGWO-MLP algorithm not only has higher training accuracy in pore pressure prediction but also has better generalization ability.

2. Grey wolf optimization and improvement algorithm

2.1. Grey wolf optimization

The grey wolf optimization algorithm is a global random search algorithm developed by simulating grey wolves' hunting and searching behavior during hunting (Emary et al., 2016). There are four levels of grey wolf population from high to low: α wolf, β wolf, δ wolf, and ω wolf. Target hunting is conducted strictly according to the wolf pack's hierarchy. The group hierarchy of grey wolves is shown in Fig. 1 and the search process of the algorithm for the optimal solution is shown in Fig. 2.

The hunting process of the grey wolf can be divided into three parts: encircled prey, hunting, and searching and attacking prey.

2.1.1. Surround the prey

Grey wolves will gradually approach and surround prey when searching for prey, and the behavior can be expressed in Eqs. (1) and (2) as follows:

$$\boldsymbol{D} = \left|\boldsymbol{M}\cdot\boldsymbol{X}_{\mathrm{p}}(t) - \boldsymbol{X}(t)\right| \tag{1}$$

$$\boldsymbol{X}(t+1) = \boldsymbol{X}_p(t) - \boldsymbol{A}\cdot\boldsymbol{D} \tag{2}$$

where t is the number of current iterations, $\boldsymbol{X}_{\mathrm{p}}(t)$ and $\boldsymbol{X}(t)$ are the position vectors of prey and grey wolf, $\boldsymbol{X}(t+1)$ is the new location of the grey wolf, and $\boldsymbol{D}$ is the distance between the grey wolf and prey. $\boldsymbol{M}$ and $\boldsymbol{A}$ are coordination coefficient vectors that can be defined by Eqs. (3) and (4):

$$\boldsymbol{A} = 2\boldsymbol{a}\cdot\boldsymbol{r}_1 - \boldsymbol{a} \tag{3}$$

$$\boldsymbol{M} = 2\cdot\boldsymbol{r}_2 \tag{4}$$

where $\boldsymbol{a}$ is the convergence factor that linearly decreases from 2 to 0 during the iteration process, and $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are random numbers between 0 and 1.

2.1.2. Hunting the prey

Under the guidance of good wolves, wolves can identify the location of potential prey and keep approaching it. This behavior is determined by Eqs. (5)–(8):

$$\boldsymbol{D}_\alpha = \left|\boldsymbol{M}_1\cdot\boldsymbol{X}_\alpha - \boldsymbol{X}(t)\right|, \boldsymbol{X}_1 = \boldsymbol{X}_\alpha - \boldsymbol{A}_1\cdot\boldsymbol{D}_\alpha \tag{5}$$

$$\boldsymbol{D}_\beta = \left|\boldsymbol{M}_2\cdot\boldsymbol{X}_\beta - \boldsymbol{X}(t)\right|, \boldsymbol{X}_2 = \boldsymbol{X}_\beta - \boldsymbol{A}_2\cdot\boldsymbol{D}_\beta \tag{6}$$

$$\boldsymbol{D}_\delta = \left|\boldsymbol{M}_3\cdot\boldsymbol{X}_\delta - \boldsymbol{X}(t)\right|, \boldsymbol{X}_3 = \boldsymbol{X}_\delta - \boldsymbol{A}_3\cdot\boldsymbol{D}_\delta \tag{7}$$

$$\boldsymbol{X}(t+1) = \frac{\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3}{3} \tag{8}$$

where α is the closest wolf to the prey, β is the second closest wolf to the prey, and δ is the third closest wolf to the prey. The updated average of α, β, and δ wolves gives the new grey wolf location.

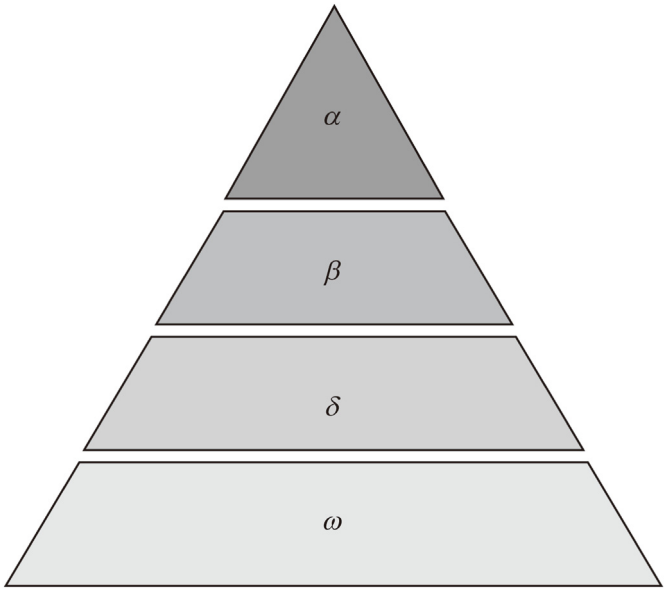

Fig. 1. Group hierarchy of grey wolves.

2.1.3. Search for and attack prey

Grey wolf groups mainly hunt according to the information of α wolf, β wolf, and δ wolf. In the mathematical description, the value of coordination coefficient vector $\boldsymbol{A}$ is employed to control whether the grey wolf is searching for prey or attacking prey. When $|\boldsymbol{A}| > 1$, the grey wolf from the prey expands its search scope to locate prey more effectively. When the $|\boldsymbol{A}| < 1$, the grey wolf narrows the search area to attack prey.

2.2. Improve methods and validation

The grey wolf optimization algorithm is a metaheuristic swarm intelligence algorithm based on the grey wolf hierarchy. Although it is faster than the traditional algorithm, it still has the problem that it is easy to fall into local optimal solutions to complex problems (Dhargupta et al., 2020). Hence, this paper improves the grey wolf optimization algorithm by three methods and verifies the superiority of the improved algorithm by eight typical benchmark functions.

2.2.1. Method of decreasing cosine law of convergence factor

One way to enhance the performance of the meta-heuristic algorithm is to balance the global search ability and local development ability effectively to improve the algorithm's optimization ability. The cooperation coefficient vector $\boldsymbol{A}$ value of the grey wolf optimization algorithm affects the behavior of the grey wolf and determines whether the algorithm is in the global search or local development process. In the GWO algorithm, the value of $\boldsymbol{A}$ decreases linearly from 2 to 0 by the convergence factor a. However, this linear decreasing method cannot meet the balance of the global search and local search ability of the algorithm. Because this method leads to the reduction of the search space equilibrium in the whole iterative process, it cannot reflect the emphasis on the optimization process of nonlinear problems in different iterative processes. Therefore, this paper proposes a nonlinear convergence factor based on cosine law. The method can be described in Eq. (9):

$$a = \cos\left(\frac{\pi t}{n}\right) + 1 \tag{9}$$

where n is the total number of iterations, as shown in Fig. 3. The convergence factor under the cosine law decreases nonlinearly. The improved algorithm has a larger convergence factor value in the early stages of iteration, which means that the search range is expanded early. It improves the global search ability of the algorithm and reduces the possibility of the algorithm falling into the local optimum. At the same time, the enhanced convergence factor decreases faster in the later stages of the algorithm, which

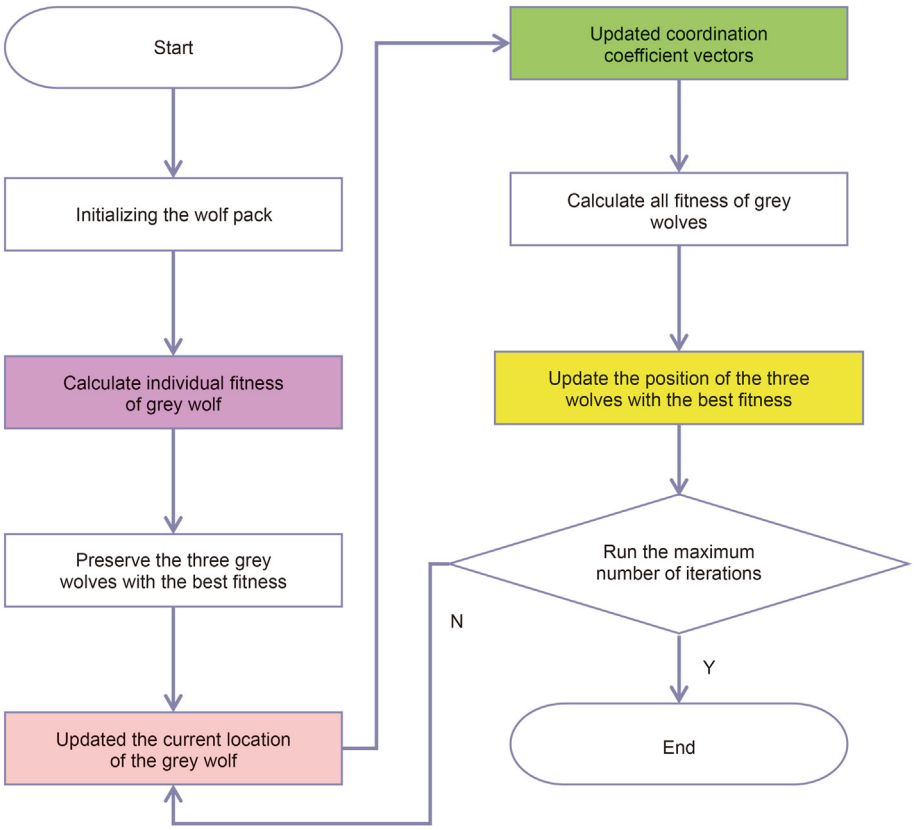

Fig. 2. Search process of the algorithm for the optimal solution.

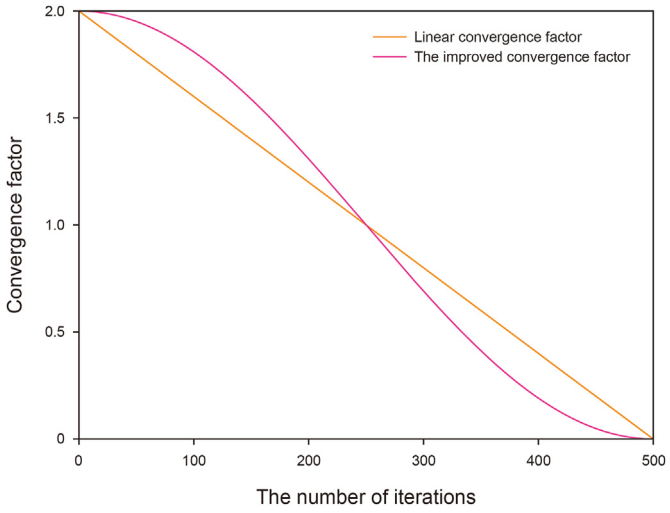

Fig. 3. Comparison of decreasing law of convergence factor.

accelerates the convergence speed of the algorithm. This can make the algorithm pay more attention to the local development in the final stage and improve the algorithm's accuracy.

2.2.2. Dynamic weight update

In the GWO algorithm, α wolf, β wolf, and δ wolf give the same guiding weight to ω wolf, which does not conform to the class concept of wolves and cannot reflect the leadership role of excellent wolves. At the same time, equal weight will slow the convergence speed of the algorithm, making it easy to fall into local optimization. Hence, this paper proposes an improved dynamic weight position update method. The method can be expressed as Eq. (10) through (13):

$$\boldsymbol{w}_1 = \frac{|\boldsymbol{X}_1|}{|\boldsymbol{X}_1| + |\boldsymbol{X}_2| + |\boldsymbol{X}_3|} \tag{10}$$

$$\boldsymbol{w}_2 = \frac{|\boldsymbol{X}_2|}{|\boldsymbol{X}_1| + |\boldsymbol{X}_2| + |\boldsymbol{X}_3|} \tag{11}$$

$$\boldsymbol{w}_3 = \frac{|\boldsymbol{X}_3|}{|\boldsymbol{X}_1| + |\boldsymbol{X}_2| + |\boldsymbol{X}_3|} \tag{12}$$

$$\boldsymbol{X}(t+1) = (\boldsymbol{w}_{11}\cdot\boldsymbol{X}_1 + \boldsymbol{w}_{22}\cdot\boldsymbol{X}_2 + \boldsymbol{w}_{33}\cdot\boldsymbol{X}_3)\left(1 - \frac{t}{T}\right) + [(1 - \boldsymbol{n} - \boldsymbol{l})\cdot\boldsymbol{X}_1 + \boldsymbol{n}\cdot\boldsymbol{X}_2 + \boldsymbol{l}\cdot\boldsymbol{X}_3]\frac{t}{T} \tag{13}$$

where $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, and $\boldsymbol{w}_3$ are the new positions after the updated positions of α wolf, β wolf, and δ wolf, $\boldsymbol{w}_{11}$, $\boldsymbol{w}_{22}$, and $\boldsymbol{w}_{33}$ are the new values of $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, and $\boldsymbol{w}_3$ in descending order, $\boldsymbol{n}$ and $\boldsymbol{l}$ are random values between 0 and 0.3. The improved dynamic weight updating method can better balance the learning and searching ability of the algorithm. At an early stage of iteration, the enhanced method can highlight the leadership ability of α wolf more effectively. In the final iteration, the leadership of α wolf is guaranteed, and the randomness of the leadership of β wolf and δ wolf is added to prevent local optimum from forming.

2.2.3. Local random walk method

The dynamic weight updating method can ensure the guidance of superior wolves and accelerate the convergence speed of the algorithm. However, the problem of insufficient search interval can occur in the early stage of iteration. Therefore, to improve the algorithm's ability to avoid the local optimum, a mutation operation similar to the GA algorithm is introduced to make the grey wolf perform a random walk near the current position. This method can be represented by Eq. (14):

$$\boldsymbol{X}(t+1) = \boldsymbol{X}(t) + \boldsymbol{r}\cdot(\boldsymbol{X}_{\mathrm{a}} - \boldsymbol{X}_{\mathrm{b}}) \tag{14}$$

where $\boldsymbol{r}$ is a random value between 0 and 1, $\boldsymbol{X}_{\mathrm{a}}$ is a random individual wolf in the wolf pack between the current wolf and the position after the dynamic weight update, and $\boldsymbol{X}_{\mathrm{b}}$ is a random individual wolf in the wolf pack. The purpose of local search is to add a random disturbance near each wolf in the wolf pack to obtain a new solution near these potential optimal solutions and avoid local optimization.

2.3. Algorithm performance test

To verify the performance of the improved algorithm, five unimodal test functions, *F*1 to *F*5, and three multimodal test functions, *F*6 to *F*8, are selected to test the algorithm's performance. The unimodal test function can effectively reflect the accuracy and convergence speed of the algorithm, while the multimodal test function can represent the global search ability and the capability to avoid the local optimum of the algorithm. Table 1 contains the mathematical expression of the benchmark function, and Fig. 4 depicts the two-dimensional search space of the test function.

In addition to IGWO and GWO algorithms, this paper selected PSO of the simulated bird population, SSA of the simulated sparrow population, and BA of the simulated bat population for comparison. The test parameters selected by the five algorithms are as follows: the initial population number is 30, the maximum number of iterations is 500, and each function is solved 30 times. The accuracy and stability of the algorithm are evaluated comprehensively by worst values, best values, means, and standard deviations. Table 2 includes the test results.

Table 2 indicates that IGWO achieves better performance in each test function compared with the GWO algorithm. The convergence ability and global optimization ability of the IGWO algorithm are improved. Except for the *F*4 test function, SSA performs better than the other four algorithms. IGWO and GWO algorithms have better performance in seven test functions than the other three algorithms. In the *F*1, *F*2, and *F*7 test functions, the performance of IGWO and GWO algorithms is much better than the other three algorithms. On the *F*6 and *F*8 test functions, both the IGWO algorithm and GWO algorithm find the real optimal solution 0. In conclusion, compared with other intelligent optimization algorithms, the population guided by the grey wolf is better than the other three animal populations in the ability to find the optimal solution.

3. Methodology

3.1. Data collection

Well 1 and Well 2 selected in this study are located in the southern margin of Junggar Basin. Well 3 is located in the Dongping area of the Qaidam Basin. The southern margin is one of the most abundant oil and gas resources in the Junggar Basin. The stratigraphic lithology distribution in this area is relatively regular. Fourth, Tertiary strata from top to bottom to sand conglomerate, sand mudstone. The Paleogene is dominated by mudstone. The Cretaceous strata are mainly composed of sandy mudstone and argillaceous siltstone. Jurassic strata are mainly siltstone and argillaceous siltstone. The bedrock in the Dongping area is rich in natural gas resources and is an important area for future exploration in Qaidam Basin. The bedrock reservoirs in this area are lithologically complex, with both magmatic rocks and metamorphic rocks. The pore structure is diverse and has the characteristics of a double pore structure of fracture dissolution pore.

Data preprocessing removes invalid and vacant values and noise points in logging data. The acoustic velocity is calculated by taking the reciprocal of the acoustic time difference, the shale content is calculated by natural gamma rays, and the porosity is calculated by density logging data. Eq. (15) describes the calculation method of shale content. Eq. (16) describes the calculation method of porosity.

$$V_{\mathrm{sh}} = 0.33\left(2^{2\left(\frac{GR-GR_{\min}}{GR_{\max}-GR_{\min}}\right)} - 1\right) \tag{15}$$

$$\varphi = \frac{\rho_{\mathrm{m}} - \rho}{\rho_{\mathrm{m}} - \rho_{\mathrm{f}}} \tag{16}$$

where V_{sh} is shale content; φ is porosity; GR is natural gamma value, API; $GR_{\min}$ is natural gamma value corresponding to pure mudstone formation, API; $GR_{\max}$ is natural gamma value corresponding to pure sandstone formation, API; ρ is rock density, g/cm^3; ρ_{m} is rock skeleton density, g/cm^3; ρ_{f} is rock fluid density, g/cm^3.

In this paper, the Bowers method is used to obtain the actual

Table 1
The parameters of the benchmark function test.

Name	Function	Range	Dimension
*F*1	$f_1(x) = \sum_{i=1}^{D} x_i^2$	[−100,100]	30
*F*2	$f_2(x) = \sum_{i=1}^{D} \lvert x_i\rvert + \prod_{i=1}^{D} \lvert x_i\rvert$	[−10,10]	30
*F*3	$f_3(x) = \sum_{i=1}^{D-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	[−30,30]	30
*F*4	$f_4(x) = \sum_{i=1}^{D} (\lvert x_i + 0.5\rvert)^2$	[−100,100]	30
*F*5	$f_5(x) = \sum_{i=1}^{n} i x_i^4 + \mathrm{random}[0,1]$	[−128,128]	30
*F*6	$f_6(x) = \sum_{i=1}^{D} [x_i^2 - 10\cos(2\pi x_i) + 10]$	[−5.12,5.12]	30
*F*7	$f_7(x) = -20\exp\left(-0.2\sqrt{\frac{1}{D}\sum_{i=1}^{D} x_i^2}\right) - \exp\left[\frac{1}{D}\sum_{i=1}^{D}\cos(2\pi x_i)\right] + 20 + \mathrm{e}$	[−32,32]	30
*F*8	$f_8(x) = \frac{1}{4000}\sum_{i=1}^{D} x_i^2 - \prod_{i=1}^{D}\cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	[−600,600]	30

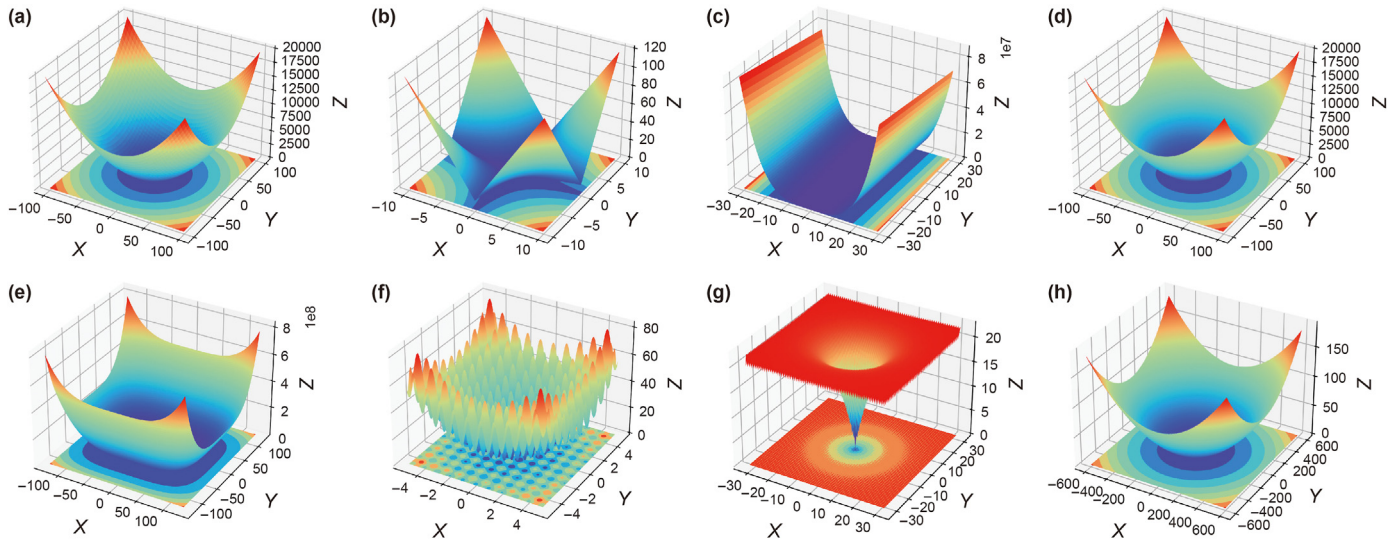

Fig. 4. 2D search space visualization of benchmark functions: (**a**) *F*1; (**b**) *F*2; (**c**) *F*3; (**d**) *F*4; (**e**) *F*5; (**f**) *F*6; (**g**) *F*7; (**h**) *F*8.

Table 2
Test results of five algorithms in F1 to F8.

Name	Algorithm	Worst value	Best value	Mean	Standard deviation
*F*1	IGWO	8.12E-34	2.94E-37	1.45E-34	2.46E-34
	GWO	4.52E-27	5.14E-29	8.61E-28	1.13E-27
	SSA	1.85E-07	3.00E-08	7.78E-08	4.33E-08
	PSO	1.18E+04	1.34E+03	4.34E+03	3.69E+03
	BA	1.07E+04	4.26E-02	2.99E+03	3.89E+03
*F*2	IGWO	7.03E-22	3.93E-23	2.99E-22	2.37E-22
	GWO	3.48E-16	3.20E-17	1.11E-16	7.52E-17
	SSA	4.08E+00	8.60E-01	2.12E+00	8.97E-01
	PSO	4.83E+02	2.01E+01	8.62E+01	1.41E+02
	BA	5.09E+04	5.84E-01	5.11E+03	1.61E+04
*F*3	IGWO	2.51E+01	2.42E+01	2.46E+01	2.88E-01
	GWO	2.85E+01	2.57E+01	2.78E+01	6.08E-01
	SSA	8.54E+02	2.77E+01	2.24E+02	2.60E+02
	PSO	6.49E+05	5.18E+03	3.01E+05	2.01E+05
	BA	2.57E+02	2.45E+01	1.11E+02	9.83E+01
*F*4	IGWO	2.53E-01.	1.19E-06	5.05E-02	1.06E-01
	GWO	1.49E+00	8.21E-05	7.52E-01	3.56E-01
	SSA	1.99E-07	4.29E-08	8.48E-08	5.31E-08
	PSO	2.43E+04	8.64E+02	7.84E+03	8.65E+03
	BA	1.12E+04	2.69E-02	2.99E+03	4.13E+03
*F*5	IGWO	4.62E-03	8.72E-04	2.20E-03	1.20E-03
	GWO	5.11E-03	7.83E-04	2.46E-03	1.26E-03
	SSA	6.45E-01	2.33E-01	4.42E-01	1.03E-01
	PSO	8.03E+08	5.32E+05	1.11E+08	2.46E+08
	BA	1.65E+09	3.21E-01	3.66E+08	6.52E+08
*F*6	IGWO	2.01E+00	0	2.01E-01	6.35E-01
	GWO	1.91E+01	0	4.26E+00	4.97E+00
	SSA	8.01E+01	2.78E+01	4.86E+01	1.75E+01
	PSO	2.88E+02	1.63E+02	2.37E+02	3.47E+01
	BA	3.61E+02	2.91E+02	3.31E+02	2.37E+01
*F*7	IGWO	1.42E-14	1.06E-14	1.35E-14	1.50E-15
	GWO	1.46E-13	7.46E-14	1.06E-13	2.08E-14
	SSA	3.88E+00	9.31E-01	2.63E+00	9.13E-01
	PSO	1.99E+01	6.69E+00	1.69E+01	5.38E+00
	BA	1.99E+01	1.99E+01	1.99E+01	9.72E-04
*F*8	IGWO	1.48E-02	0	2.23E-03	5.01E-03
	GWO	2.60E-02	0	6.21E-03	8.69E-03
	SSA	2.79E-02	1.16E-04	9.89E-03	8.38E-03
	PSO	9.99E+01	1.01E+01	3.75E+01	3.30E+01
	BA	3.61E+02	1.16E+02	2.19E+02	7.96E+01

value of pore pressure, and the actual value of effective stress is obtained by subtracting the pore pressure from the overburden pressure. Moreover, observing the relationship between acoustic velocity and effective stress removes the noise points that deviate from the overall trend. Besides, deleting noise points can ensure the training data's rationality and the trained model's reliability. After data processing, Well 1 has 881 data sets for model training. Well 2 and Well 3 have 538 and 464 data sets for model testing. The data for Well 1, Well 2, and Well 3 are shown in Fig. 5, Fig. 6, and Fig. 7 respectively.

3.2. Machine learning algorithm

3.2.1. Support vector machine

The support vector machine is a machine learning method developed from statistical learning theory. It was initially proposed to study linear separability problems. In the process of inversion using SVM, the selection of kernel function is crucial. The application of the kernel function can transform it from low-dimensional to high-dimensional space without changing the original data, allowing the high-dimensional dot product to be calculated in low-dimensional space (Wang et al., 2008).

The commonly used kernel functions are linear kernel function,

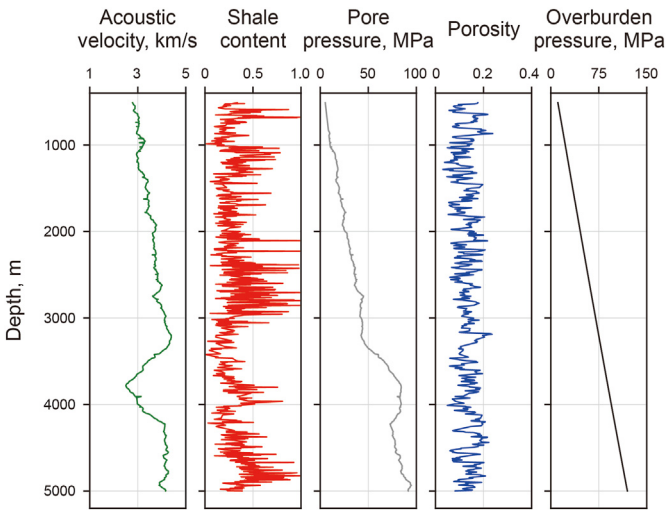

Fig. 5. Well 1 data used to train the model.

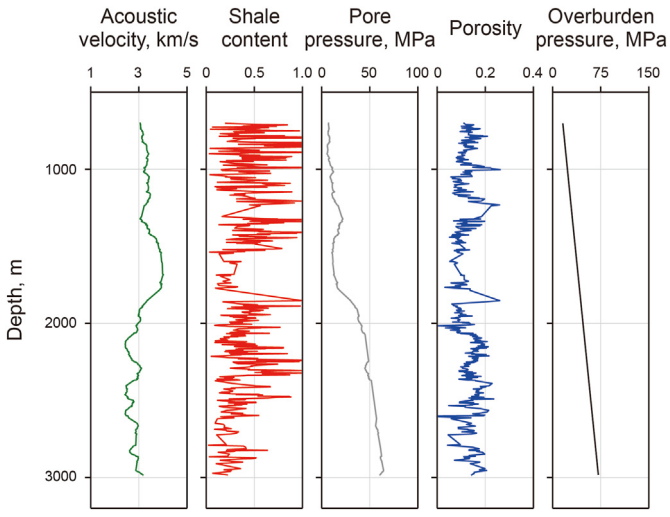

Fig. 6. Well 2 data used to test the model.

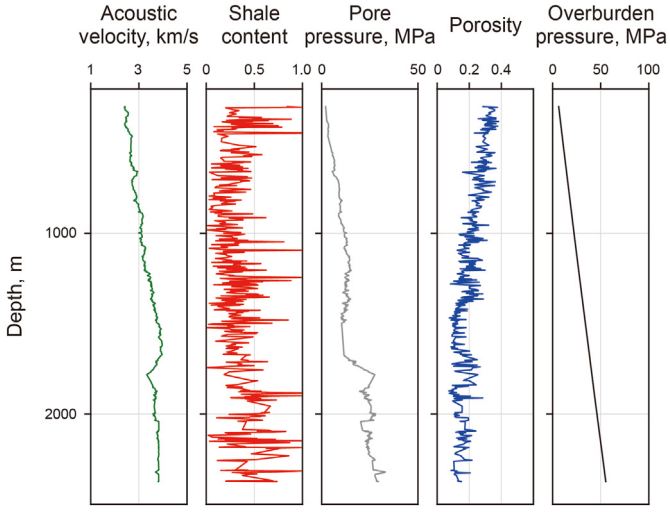

Fig. 7. Well 3 data used to test the model.

polynomial kernel function, Sigmoid kernel function, and Gaussian radial basis kernel function. Because the Gaussian radial basis kernel function can realize nonlinear mapping and has fewer parameters, which reduces the model's complexity and the calculation difficulty, this paper adopts the Gaussian radial basis kernel function as the kernel function used in the algorithm. The hyperparameters in SVM are the penalty factor C and the Gaussian kernel parameter gamma. C is the tolerance of the model for error, and gamma reflects the distribution of the data after mapping to the high-dimensional feature space (Rebentrost et al., 2014). The combination results of C and gamma affect the SVM fitting accuracy. To avoid overfitting or underfitting, it is necessary to optimize the selection of these two kinds of parameters.

3.2.2. Random forest

RF algorithm is a machine learning algorithm that combines the Bagging ensemble learning theory with the random subspace method. Its prediction principle is to combine multiple decision tree algorithms into a forest and rely on the voting results of each tree to determine the final predicted value. The structure of this voting system allows each decision tree to grow naturally and be independent of each other. Therefore, compared with a single decision tree algorithm, it can effectively improve the prediction accuracy of the model. At the same time, because of the random growth of each decision tree, the overfitting phenomenon of the model can be reduced (Fawagreh et al., 2014).

The hyperparameters of the random forest are the number of decision trees and the maximum depth of the decision tree. Normally, the number of base models for RF algorithms is directly proportional to the prediction accuracy. However, too many decision trees and too much depth will lead to reduced generalization ability of the model and overfitting phenomenon. Therefore, it is necessary to optimize the selection of these two parameters.

3.2.3. Extreme gradient boosting

XGB algorithm is an improved machine learning algorithm based on a gradient-boosting decision tree algorithm. Unlike the gradient boosting decision tree algorithm, which only uses the first derivative when optimizing the loss function, XGB performs the second-order Taylor expansion of the loss function. The algorithm adds regular terms to the objective function, so it can control the complexity of the model and reduce the variance of the model. After each iteration, the learning rate is assigned to leaf nodes, and the weight of each tree is reduced to provide a better learning space for the following (Torlay et al., 2017). In addition, XGB considers multithreading to improve the efficiency of the algorithm when the amount of data is large and memory is insufficient.

Similar to the RF algorithm, the number of decision trees and the maximum tree depth of decision trees are also two parameters that affect the fitting accuracy of the XGB algorithm. In addition, the XGB algorithm adds a learning rate parameter, which is used to control the degree of learning results of each decision tree in the training process.

3.2.4. Multilayer perceptron neural network

MLP is a feedforward artificial neural network. Feedforward neural network is an efficient nonlinear function fitting method that uses a gradient descent algorithm to minimize the loss function. MLP neurons have a hierarchical structure, generally consisting of an input layer, one or more hidden layers, and an output layer. The neurons in each layer are connected by weights, and the neurons in the output layer and the hidden layer also have thresholds to regulate the output (Panchal et al., 2011). The MLP adjusts the connection weights and related thresholds between neurons according to the training data during the learning process.

To make the MLP model better learn the nonlinear relationship between data, this paper sets three hidden layers for the MLP model. For the loss function optimization algorithm and activation function of the MLP model, this paper selects the stochastic gradient descent method and tanh function which converge faster. The optimizer's batch size is set to 200 to maximize the use of information from the training data set while maintaining reasonable computational efficiency. Using smaller batches speeds up training and reduces memory requirements, For the MLP model, the number of hidden layers and the number of neurons in each hidden layer are the main parameters that affect the performance of the algorithm. In addition, regularization parameters and learning rate are also important parameters that affect the prediction accuracy and training efficiency of the MLP model. Therefore, the number of neurons in the hidden layer, regularization parameters, and learning rate are taken as the parameters to be optimized for the MLP model in this paper.

3.2.5. Comparison of models

To reflect the differences among different models, this paper summarized the advantages and disadvantages of SVM, RF, XGB,

Table 3
Advantages and disadvantages of four machine learning models.

Model	Advantages	Disadvantages
SVM	Good generalization performance is also applicable to high-dimensional problems.	Sensitive to missing data, not suitable for large sample sets.
RF	Using random sampling, the trained model has a small variance and strong generalization ability.	In the noisy sample set, the model is easy to fall into overfitting.
XGB	The second order Taylor expansion and regularization terms are introduced to improve the accuracy and flexibility.	The process of pre-sorting is time-consuming.
MLP	Excellent self - adaptation, and self-learning function.	The training time is long and the optimization process is complicated.

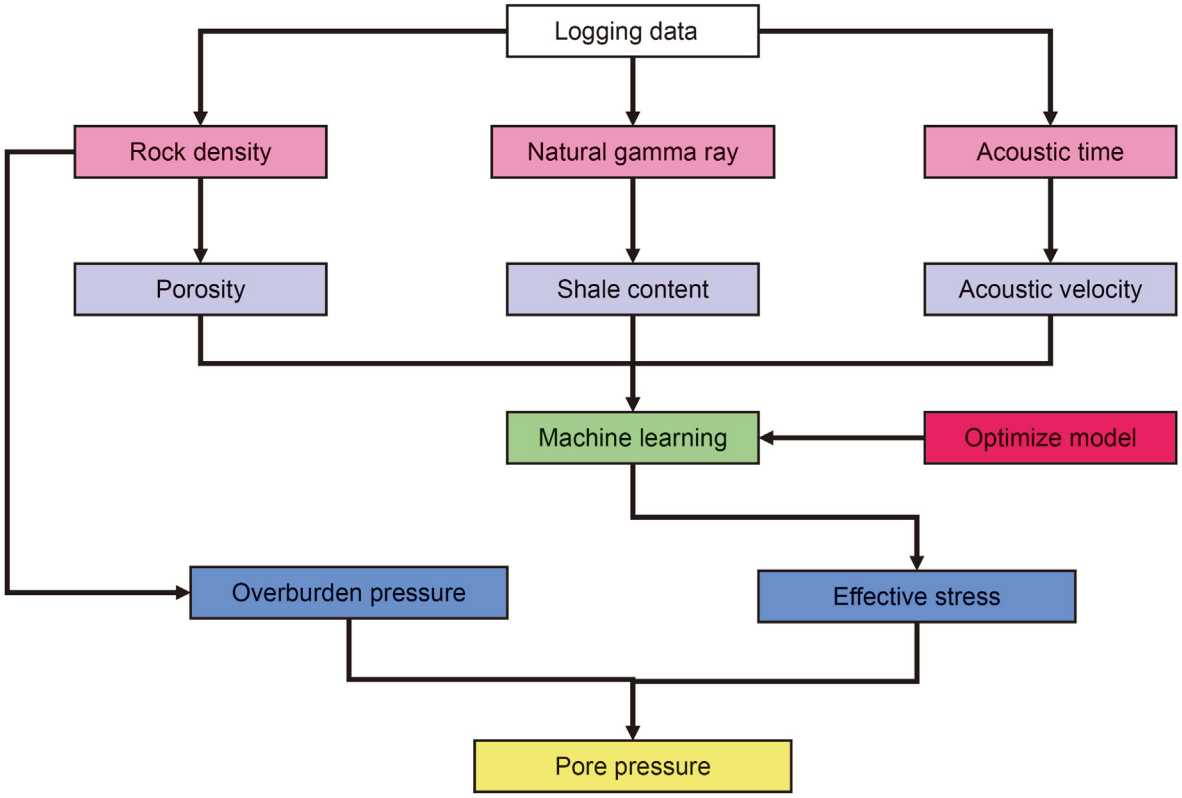

Fig. 8. Pore pressure prediction flow.

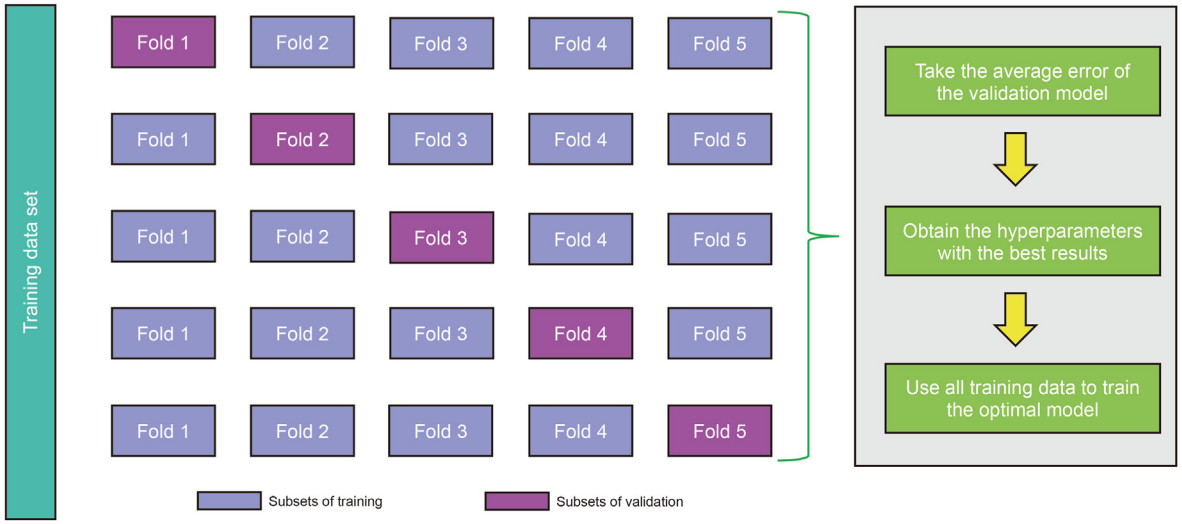

Fig. 9. The workflow of 5-fold cross-validation.

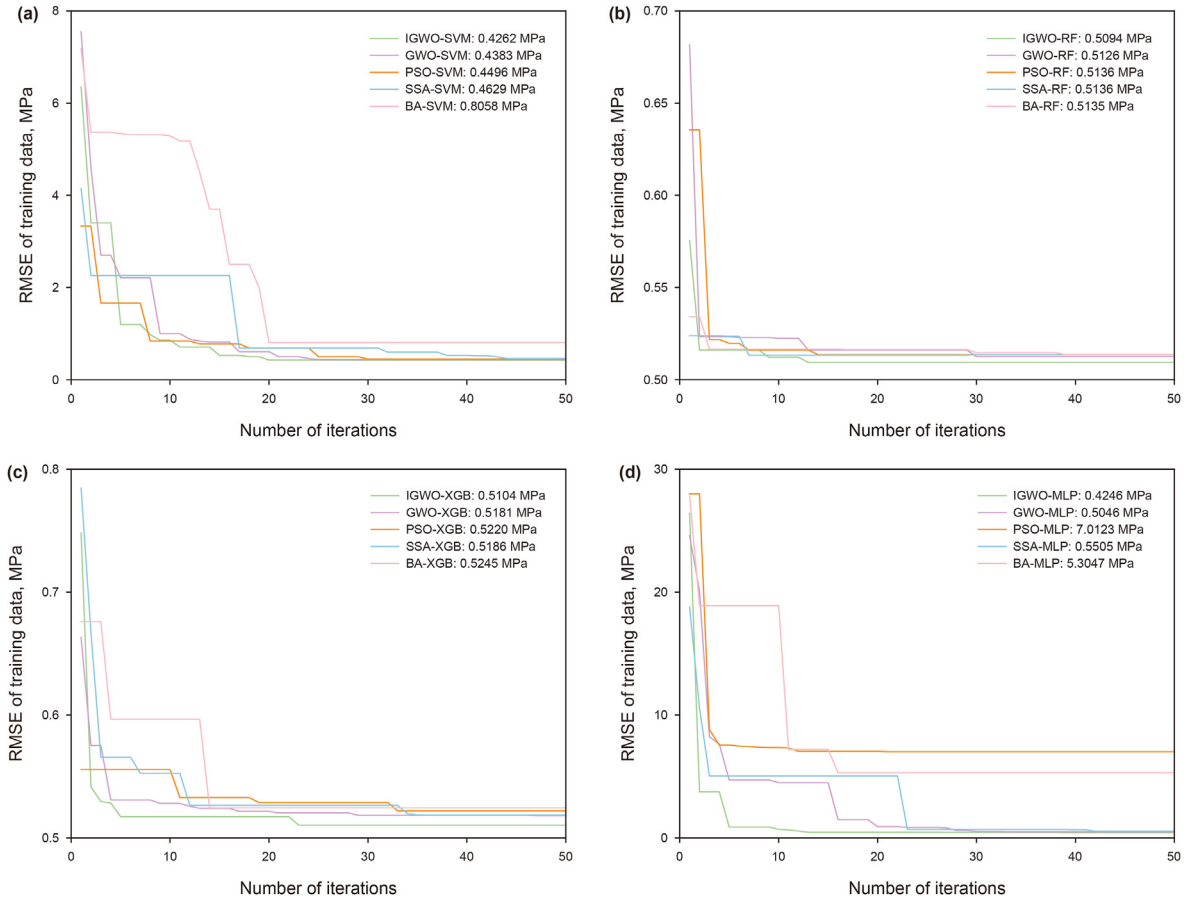

Fig. 10. RMSE decline curves for different hybrid machine learning optimization models: (**a**) SVM; (**b**) RF; (**c**) XGB; (**d**) MLP.

Table 4
Comparison of optimization results of the four methods.

Model	Hyperparameter	Search space	Result of optimization	Best RMSE
SVM	C	[0,1000]	53	0.4262
	gamma	[0,1000]	0.1887	
RF	n_estimators	[1,100]	30	0.5094
	max_depth	[1,100]	19	
XGB	n_estimators	[1,100]	98	0.5104
	max_depth	[1,100]	40	
	learning_rate	[0,1]	0.0769	
MLP	hidden_layer_sizes	([1,100], [1,100], [1,100])	(18, 59, 44)	0.4246
	alpha	[0,1]	0	
	learning_rate_init	[0,1]	0.0024	

and MLP models. Table 3 lists the advantages and disadvantages of each of the four models in use.

3.3. Forecasting process

The effective stress theorem and acoustic velocity model are the basis of pore pressure prediction. Eq. (17) describes the effective stress theorem. Eq. (18) describes the acoustic velocity model. The acoustic velocity model shows that the velocity decreases with the increase in porosity and shale content and increases with the increase in effective stress. It can be seen that one way to predict effective stress is to use acoustic velocity, shale content, and porosity as input characteristics to build a nonlinear model between effective stress and them. Therefore, using the fitting ability of the machine learning model, the nonlinear model of acoustic velocity, shale content, porosity, and effective stress can be established. Then the pore pressure is obtained according to the effective stress theorem.

$$P_E = P_O - P_P \tag{17}$$

$$V_P = 5.77 - 6.94\varphi - 1.73\sqrt{V_{sh}} + 0.446\left(P_E - e^{-16.7P_E}\right) \tag{18}$$

where P_E is effective stress, MPa; P_O is overburden pressure, MPa; P_P is pore pressure, MPa; V_P is acoustic velocities, km/s. The overall flow of the machine learning model optimized based on the intelligent optimization algorithm to predict the pore pressure is shown in Fig. 8.

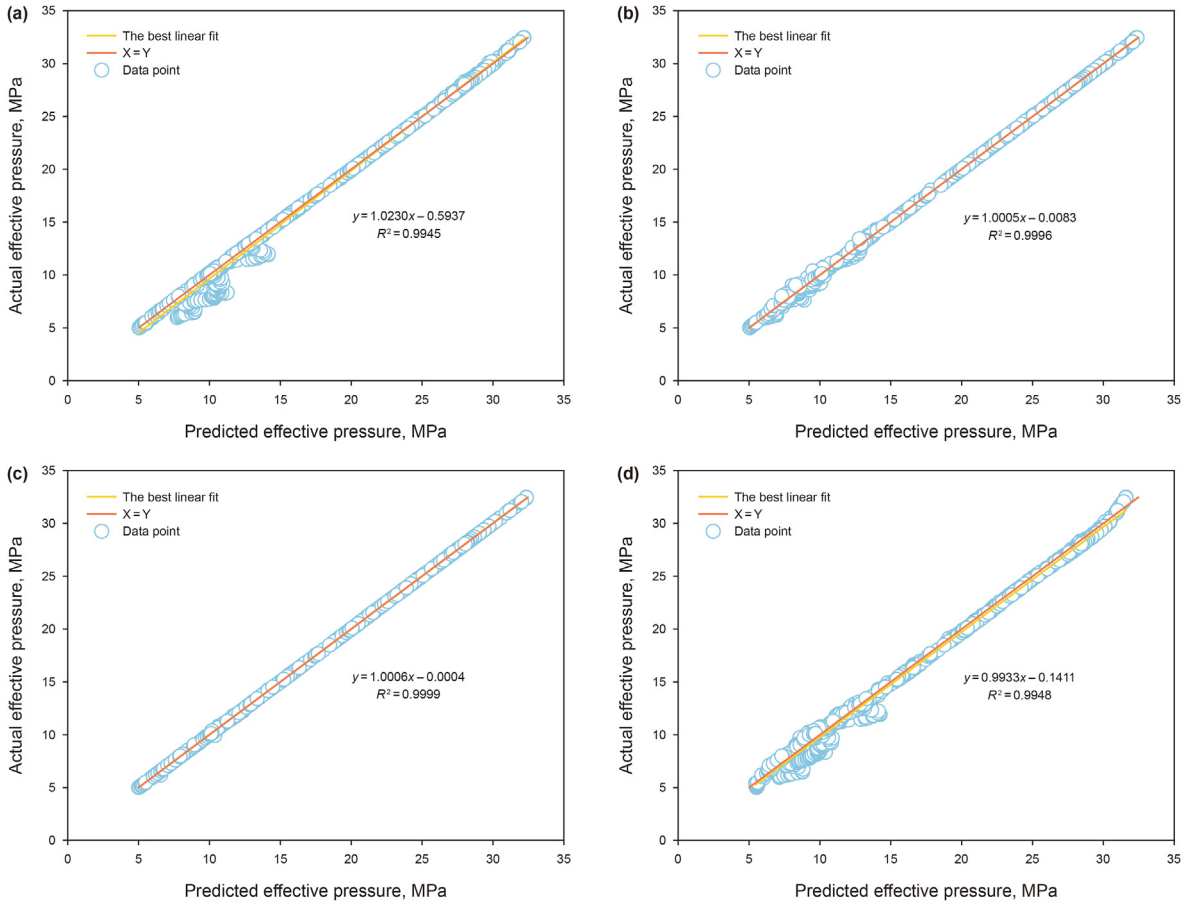

Fig. 11. The linear fitting results of the four machine learning models for the predicted and actual effective stress values of Well 1: (**a**) SVM; (**b**) RF; (**c**) XGB; (**d**) MLP.

3.4. Model evaluation index

The root means the square error is used to measure the deviation between the observed value and the predicted value. It is a statistical index often used to measure the prediction accuracy of machine learning models. The smaller the value of RMSE, the higher the prediction accuracy of the model. The coefficient of determination is commonly used to evaluate the degree of agreement between the predicted and actual values of regression models, where the larger the R^2, the better the fitting effect. Therefore, RMSE and R^2 were selected in this paper as the evaluation index. These metrics can be expressed mathematically as shown in Eqs. (19) and (20):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_{ip}\right)^2} \tag{19}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_{ip}\right)^2}{\sum_{i=1}^{n}\left(y_i - y_m\right)^2} \tag{20}$$

where y_i is the true value, y_m is the average value of the true value, and y_{ip} is the predicted value.

3.5. Model validation method

The traditional method of data set partitioning is to divide the data set into a training set and a test set. Most of the data is used as the training set, and a few data is used as the test set. The performance of the model on the test set is improved by adjusting the hyperparameters. This approach identifies the models that perform best on the test set. But its performance on the new data is still unknown. Unlike traditional data set partitioning approaches, *K*-fold cross-validation is a model evaluation method that avoids the limitations and particularity of fixed data sets. In this method, the training set data is equally divided into *k* groups, and each subset data is verified once, and the rest *k*-1 subset data is used as the training set. It is important to note that these data were randomly assigned to ensure that the model fully learned the changes throughout the well. Then we have *k* models. The average value of the evaluation results of the validation sets of the *k* models was taken as the final performance index of the model. The hyperparameter combination with the best performance under *K*-fold cross-validation will be used as the best combination of the new model, and the model will be retrained with the training set data. At this time, the test set data can be used as the new sample data that has not been learned to test the generalization ability of the model. The common choices for *k* values are usually 3, 5 or 10. To avoid too long a training time for the model, this paper chooses the method of 5-fold cross-validation. The workflow of 5-fold cross-validation is shown in Fig. 9.

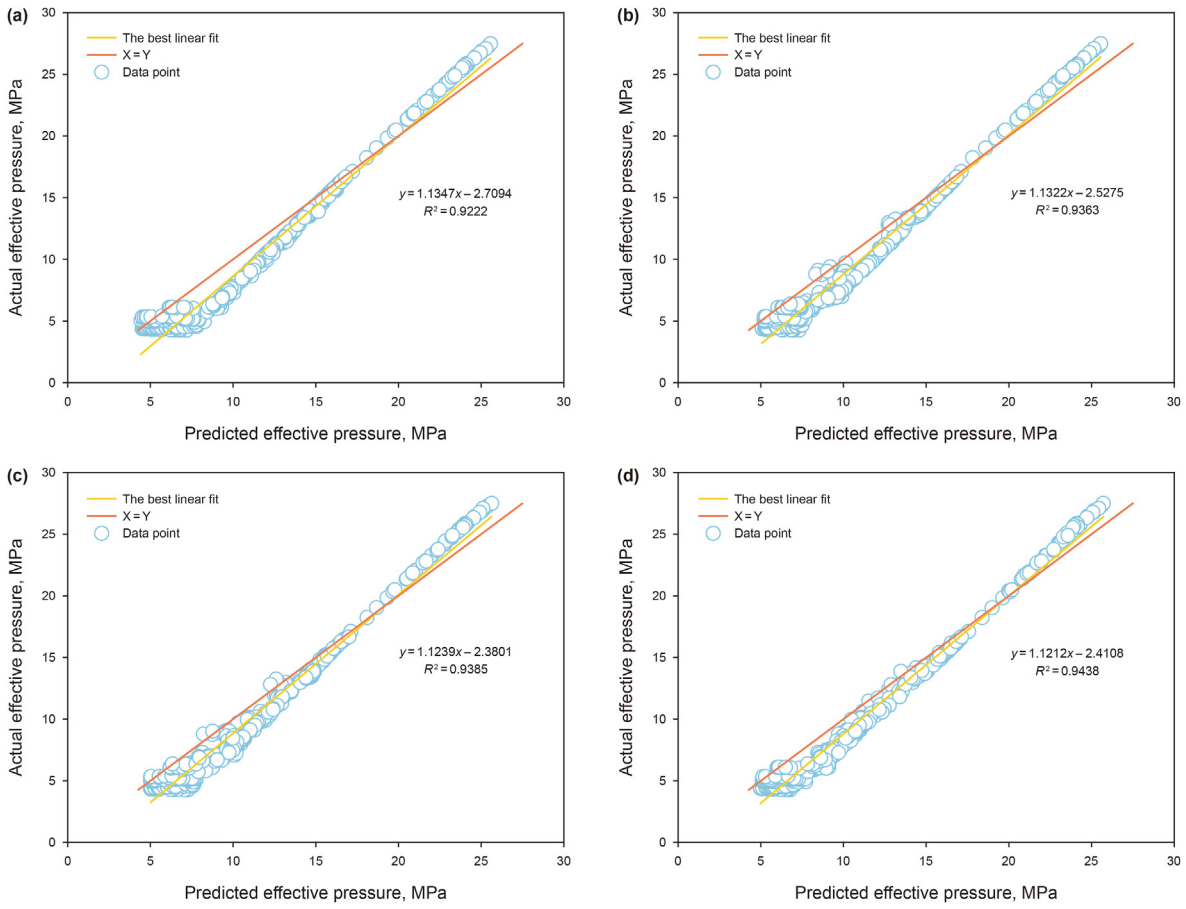

Fig. 12. The linear fitting results of the four machine learning models for the predicted and actual effective stress values of Well 2: (**a**) SVM; (**b**) RF; (**c**) XGB; (**d**) MLP.

4. Results and discussion

4.1. Data preprocessing

Machine learning is a data-driven approach that requires more data sets to train models. Indeed, the data quality in machine learning prediction is just as important as the quality of the regression or the classification models. To improve the generalization performance of prediction results, it is necessary to process the training data. Therefore, before training the machine learning model, the training feature values need to be converted into the same range.

Data normalization can be done by placing the value of the smaller feature (porosity and shale content) in the same order of magnitude as the larger feature value (acoustic velocity), to eliminate the influence that the larger value may have on the smaller value in the training model. Data normalization can not only improve the training accuracy of the model but also accelerate the convergence rate of the model. The max-min method is the most widely used data normalization method in machine learning. This method scales all data between 0 and 1, eliminating dimensional effects. This method can be expressed as described in Eq. (21):

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{21}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of this parameter in all samples, respectively, and x and x^* are the values of the current sample points and the normalized values of this parameter.

4.2. Developing prediction models

This paper implements machine learning models based on intelligent algorithms using Python programming. In this paper, four kinds of machine learning models and five kinds of intelligent optimization algorithms are combined respectively to form 20 kinds of hybrid machine learning optimization algorithms. To reflect the performance gap between the selected algorithms, the same number of population and algorithm iteration times were set to train the machine learning model. Considering that the machine learning model in this paper needs to optimize 2 to 5 hyperparameters, the optimal solution can be found with a small number of populations and iterations. Therefore, in this paper, the population of the intelligent optimization algorithm is 5, and the maximum number of iterations is 50.

For the training process of the model, RMSE is used as the performance index of cross-validation in this paper. The RMSE descent process of the machine learning model optimized by the intelligent optimization algorithm is shown in Fig. 10. It can be seen from Fig. 10 that the four machine learning models optimized by the IGWO algorithm all obtain the best RMSE results. In addition, the optimization performance of the BA and PSO algorithm on the MLP model is poor. This shows that the BA and PSO algorithms can not find the potential location of the global optimal solution well on the high-dimensional features. In the early iteration stage of the algorithm, the optimal target location is not determined, so it falls into the local optimal solution. In terms of convergence speed, the IGWO algorithm is also significantly faster than other intelligent optimization algorithms. Therefore, the IGWO algorithm also has more opportunities to find the position of the optimal solution in the late

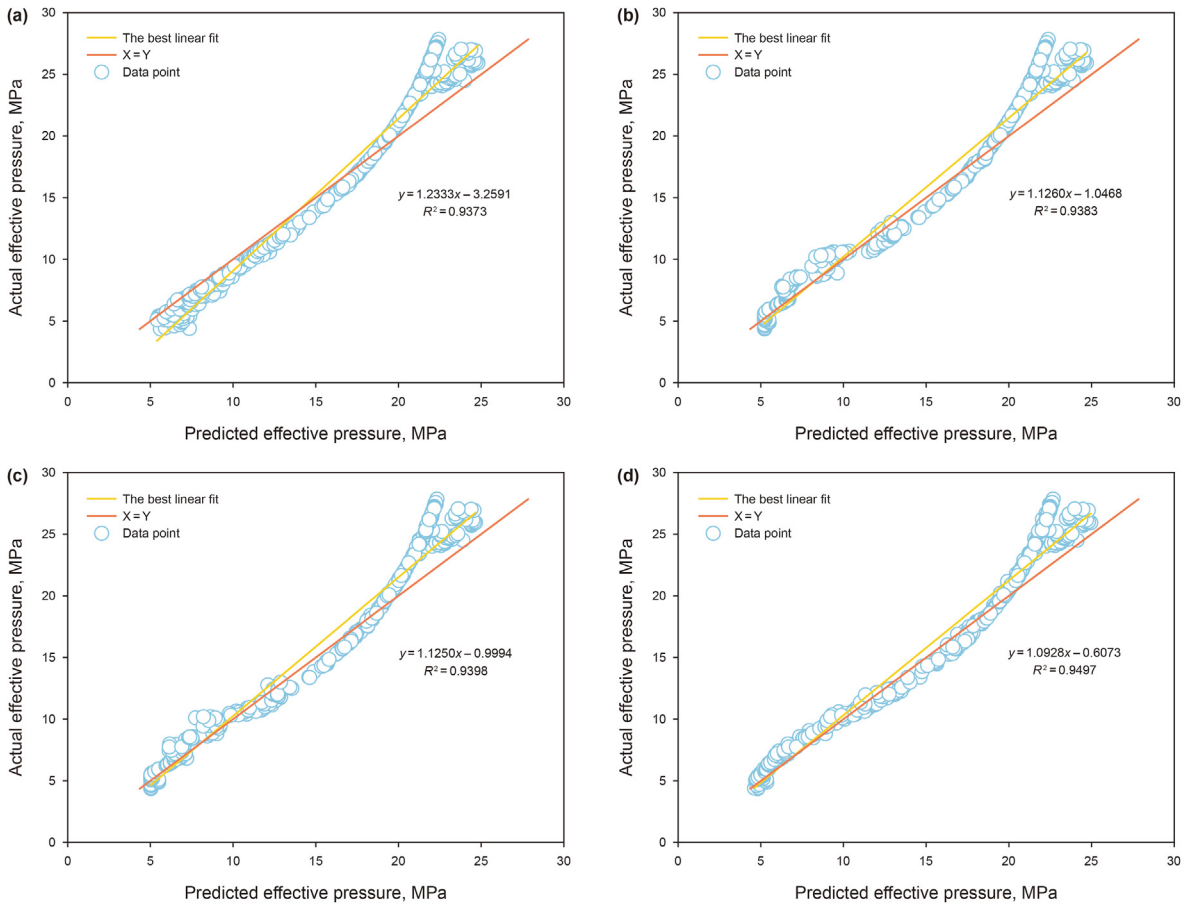

Fig. 13. The linear fitting results of the four machine learning models for the predicted and actual effective stress values of Well 3: (**a**) SVM; (**b**) RF; (**c**) XGB; (**d**) MLP.

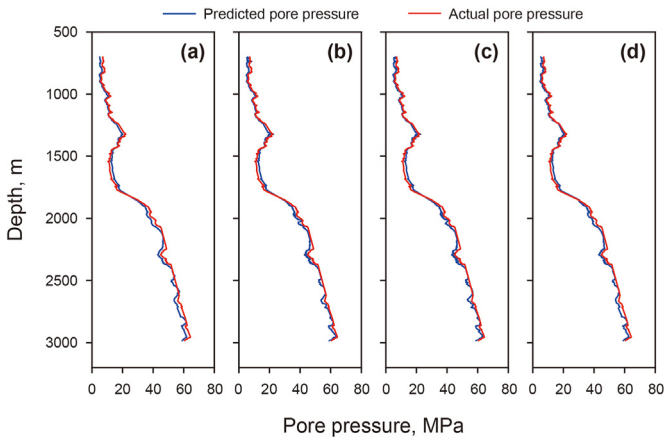

Fig. 14. The results of comparison of the four machine learning models for the predicted and actual pore pressure values of Well 2: (**a**) SVM; (**b**) RF; (**c**) XGB; (**d**) MLP.

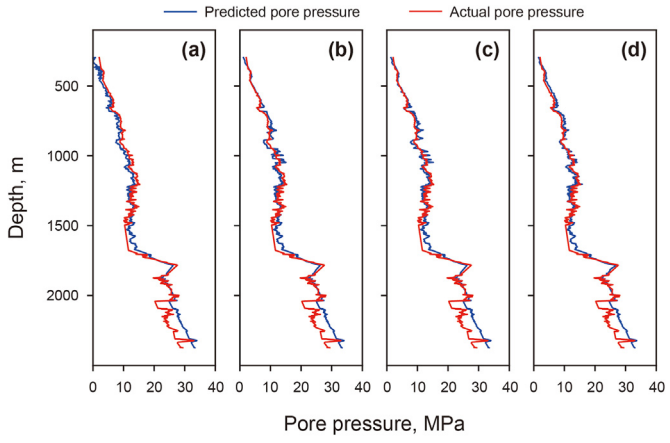

Fig. 15. The results of comparison of the four machine learning models for the predicted and actual pore pressure values of Well 3: (**a**) SVM; (**b**) RF; (**c**) XGB; (**d**) MLP.

iteration period and improve the solving accuracy of the algorithm. In addition, the RMSE result of the MLP model was 0.4246 MPa, which was lower than 0.4262 MPa for the SVM model, 0.5094 MPa for the RF model, and 0.5104 MPa for the XGB model. This shows that the MLP model has the best predictive generalization ability under 5-fold cross-validation. Table 4 shows the hyperparameter results of four machine-learning models optimized by the IGWO algorithm.

Since the cross-validation is only used to evaluate the generalization ability of different combinations of hyperparameters in the training data set, it is also necessary to apply the hyperparameters in Table 4 with the optimal performance under the 5-fold cross-validation to the training of all Well 1 data. Fig. 11 shows the linear fitting results of the four machine learning models for the predicted and actual effective stress values of Well 1. The four models in Fig. 11 show high nonlinear fitting accuracy in training data. The RF and XGB models were close to a perfect fit, showing almost identical predicted and actual values in Well 1.

In SVM and MLP models, the predicted results of some low

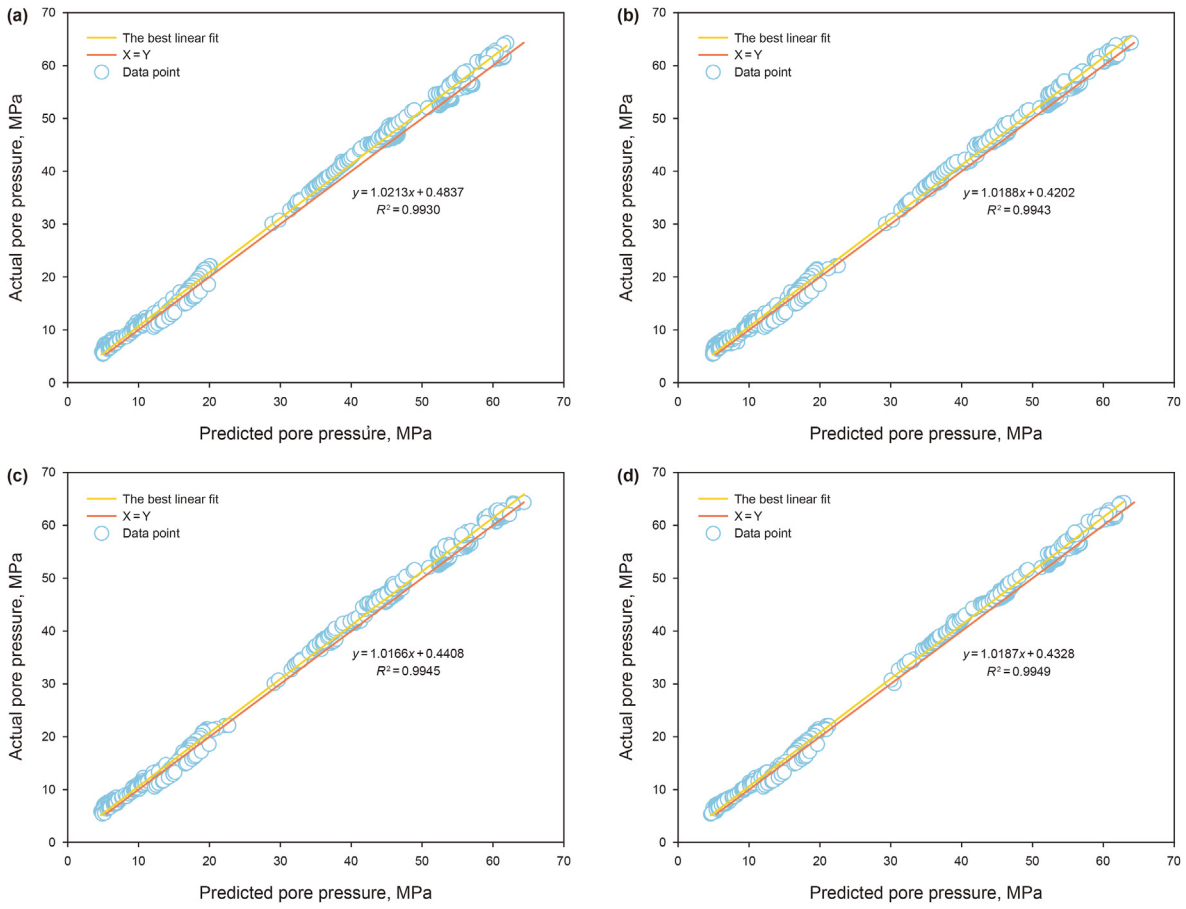

Fig. 16. The linear fitting results of the four machine learning models for the predicted and actual pore pressure values of Well 2: (**a**) SVM; (**b**) RF; (**c**) XGB; (**d**) MLP.

effective stress values are higher than the actual values. This is because, at the lower effective stress, there are some low effective stress values caused by abnormally high pressure. Since the acoustic velocity is proportional to the effective stress, however, higher acoustic velocity will produce lower effective stress at these abnormally high pressures. Therefore, for these points, porosity and shale content should be used to correct the predicted results of effective stress. In the process of model fitting, the SVM and MLP models ignore some marginal data, resulting in the phenomenon of underfitting, which leads to the reduction of the generalization ability of the model. As a result, some predicted values were higher than actual values in well 1. However, RF and XGB models consider more outliers to fit the model, so the predicted and actual values are very close to each other. However, this does not mean that RF and XGB models will be better at predicting new samples because overfitting can occur.

4.3. Model prediction results

Although this paper has proved that the machine learning model fitted under the hyperparameter has certain generalization ability through the 5-fold cross-validation. However, such high training accuracy may still lead to overfitting problems, and the performance of low prediction accuracy may appear in the new data set. Therefore, the performance of these four models will be tested with data from Well 2 and Well 3. Figs. 12 and 13 show the linear fitting results of the four machine learning models for the predicted and actual effective stress values of Well 2 and Well 3. The R^2 of the MLP model was the highest in the data of Well 2, reaching 0.9438. This result was better than 0.9385 for the XGB model, 0.9363 for the RF model, and 0.9222 for the SVM model. The R^2 of the MLP model was the highest in the data of Well 3, reaching 0.9497. This result was better than 0.9398 for the XGB model, 0.9383 for the RF model, and 0.9373 for the SVM model. Figs. 12 and 13 show that the MLP model has the best predictive performance of effective stress on the data from Well 2 and Well 3, which also verify that the generalization ability of the MLP model is superior to the SVM, RF, and XGB models under the 5-fold cross-validation.

It can be seen that although the RF and XGB models had higher fitting accuracy in Well 1, they did not show the same higher prediction accuracy in Well 2 and Well 3. While the SVM model performed well on the 5-fold cross-validation, it did not perform well on the prediction results of Well 1, Well 2, and Well 3. Compared with the SVM model, the MLP model has a better fitting effect on the training set, but it is not as close to complete fitting as the RF and XGB models. Therefore, the MLP model has better generalization ability. It can be seen from the prediction results of Well 2 that the predicted value is higher than the actual value at the lower effective stress as in Well 1. This is because both Wells are in the same area, and although the pore pressure changes in different ways, the sound velocity is similar to the effective stress changes in the area. From the prediction results of Well 3, the predicted results are closer to the actual value at the lower effective stress, but the predicted value is lower than the actual value at the higher point. According to the actual value of pore pressure in Well 3, abnormal low pressure occurred in the deep formation. As a result, the actual

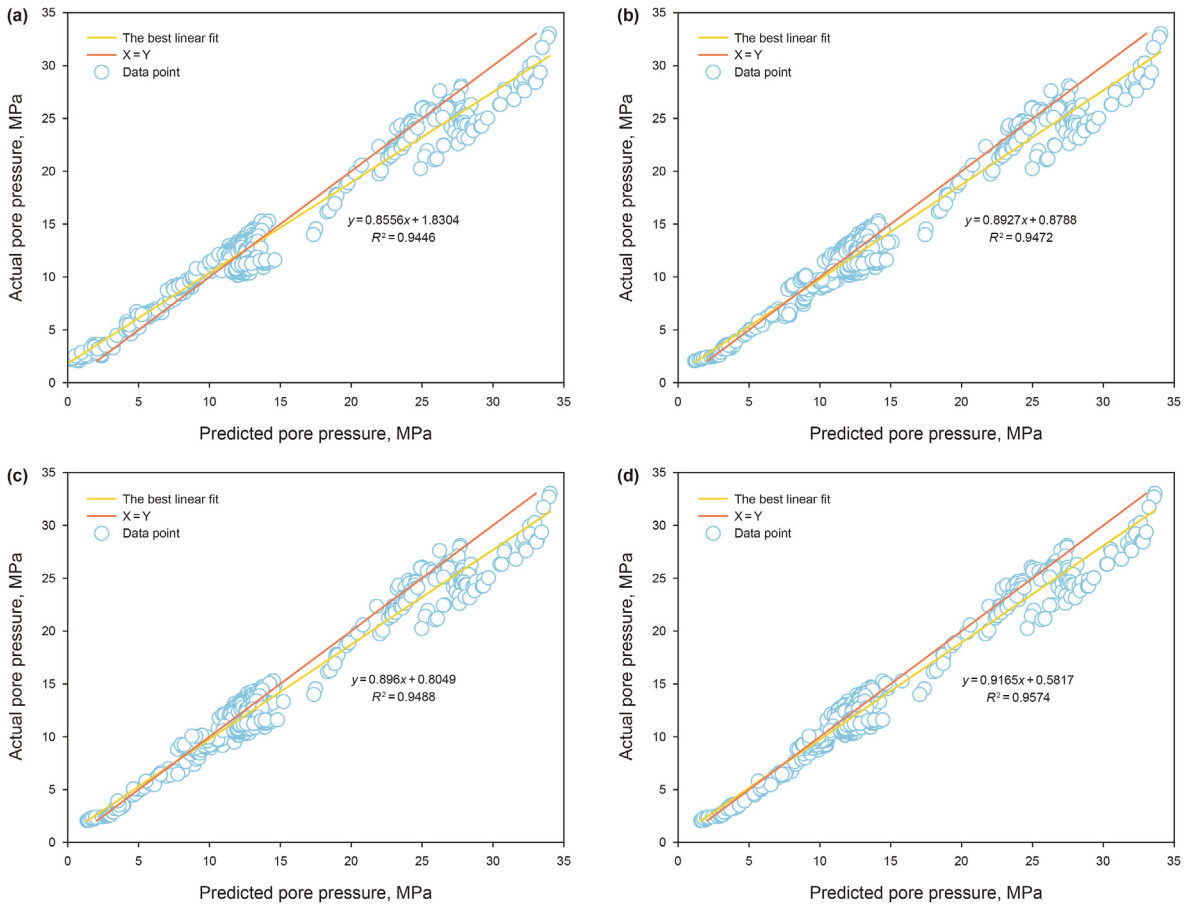

Fig. 17. The linear fitting results of the four machine learning models for the predicted and actual pore pressure values of Well 3: (**a**) SVM; (**b**) RF; (**c**) XGB; (**d**) MLP.

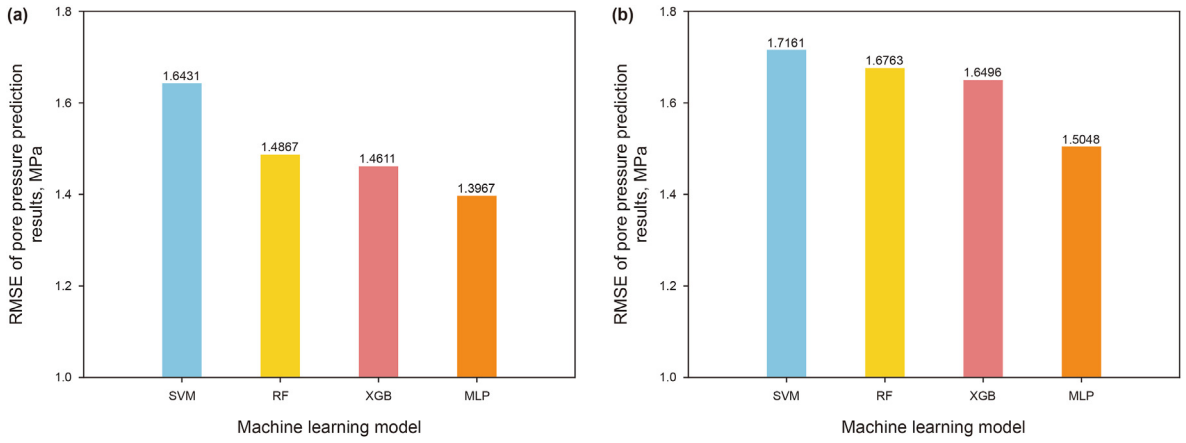

Fig. 18. The RMSE comparison results for the four machine learning models: (**a**) Well 2; (**b**) Well 3.

value of effective stress is larger than the normal condition, resulting in the phenomenon that the predicted value is lower than the actual value.

In this paper, the overburden pressure is subtracted from the effective stress value predicted by the machine learning model to obtain the predicted pore pressure. Figs. 14 and 15 show the results of the comparison of the four machine learning models for the predicted and actual pore pressure values of Well 2 and Well 3.

Figs. 16 and 17 show the linear fitting results of the four machine learning models for the predicted and actual pore pressure values of Well 2 and Well 3. Compared with the prediction of effective stress, the R^2 value of pore pressure prediction is further improved. The R^2 values of the MLP model were the highest in the data of Well 2 and Well 3, reaching 0.9949 and 0.9574, which were better than 0.9945 and 0.9488 for the XGB model, 0.9943 and 0.9472 for the RF model, and 0.9930 and 0.9446 for the SVM model. It can be seen that the IGWO-MLP model has the best R^2 performance of pore pressure.

In addition to comparing R^2 results, the RMSE values of the four models were compared in Well 2 and Well 3. Fig. 18 shows the

RMSE comparison results for the four machine learning models. The RMSE value of the MLP model is 1.3967 and 1.5048 MPa, which is lower than 1.4611 and 1.6496 MPa for the XGB model, 1.4867 and 1.6763 MPa for the RF model and 1.6431 and 1.7161 MPa for the SVM model. In the RMSE results, the MLP model also achieved the best performance. It can be seen that the IGWO-MLP hybrid model in this paper has the best performance in predicting effective stress not only in Well 1 trained with the 5-fold cross-validation method but also in the test data in Well 2 and Well 3. After the pore pressure is obtained by using the effective stress theorem, the machine learning model shows good generalization ability in the prediction of pore pressure. Considering the performance of all the hybrid models, the IGWO-MLP model is found to be the most suitable hybrid model for the establishment of pore pressure prediction.

5. Conclusion

Based on intelligent optimization and machine learning algorithm, this paper realizes an intelligent prediction method for pore pressure. The study's findings can be mainly summarized as follows:

(1) The three enhanced strategies proposed in this paper can increase the ability of the IGWO algorithm to determine the global optimal solution position, the local solution's accuracy, and the algorithm's convergence speed. On seven benchmark function tests, compared to the GWO, PSO, SSA, and BA, the IGWO algorithm performed better.
(2) In this study, an intelligent prediction model of effective stress was established based on acoustic velocity, shale content, and porosity. Combined with the effective stress theorem, the predicted value of pore pressure is obtained. Among multiple hybrid machine learning optimization algorithms, IGWO-MLP shows the best performance on the 5-fold cross-validation. The RMSE result of the MLP model was 0.4246 MPa, which was lower than 0.4262 MPa for the SVM model, 0.5094 MPa for the RF model, and 0.5104 MPa for the XGB model.
(3) IGWO-MLP shows the best prediction accuracy in the prediction results of pore pressure in Well 2 and Well 3. The R^2 of the MLP model was the highest in the data of Well 2 and Well 3, reaching 0.9949 and 0.9574. This result was better than 0.9945 and 0.9488 for the XGB model, 0.9943 and 0.9472 for the RF model, and 0.9930 and 0.9446 for the SVM model. In addition, the RMSE value of the MLP model is 1.3967 and 1.5048 MPa, which is lower than 1.4611 and 1.6496 MPa for the XGB model, 1.4867 and 1.6763 MPa for the RF model, 1.6431 and 1.7161 MPa for the SVM model.

The input features of the model in this paper are determined using the acoustic velocity model. When the acoustic velocity is abnormally high in shallow strata, the prediction accuracy of effective stress will be decreased. In the future, the machine learning model can be further optimized to improve the prediction accuracy when a more suitable multivariable parametric model is identified for predicting pore pressure. In addition, cross-well prediction usually finds performance deterioration due to the non-iid issue. The generalizability of the model could be improved by domain adaptation if the training and testing wells are very different in data distribution in the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Atashbari, V., Tingay, M., Zareian, M.H., 2012. Compressibility method for pore pressure prediction. In: Abu Dhabi International Petroleum Conference and Exhibition. Society of Petroleum Engineers. https://doi.org/10.2118/156337-MS.

Azadpour, M., Manaman, N.S., Kadkhodaie-Ilkhchi, A., et al., 2015. Pore pressure prediction and modeling using well-logging data in one of the gas fields in south of Iran. J. Petrol. Sci. Eng. 128, 15–23. https://doi.org/10.1016/j.petrol.2015.02.022.

Bektas, E., Miska, S.Z., Ozbayoglu, E.M., et al., 2015. Application of Kalman filter to predictions of pore pressure while drilling. In: SPE Annual Technical Conference and Exhibition. OnePetro. https://doi.org/10.2118/174908-MS.

Bowers, G.L., 1995. Pore pressure estimation from velocity data: accounting for overpressure mechanisms besides undercompaction. SPE Drill. Complet. 10 (2), 89–95. https://doi.org/10.2118/27488-PA.

Czerniak, M., 2017. RhoVe method: a new empirical pore pressure transform. Mar. Petrol. Geol. 86, 343–366. https://doi.org/10.1016/j.marpetgeo.2017.04.019.

Dhargupta, S., Ghosh, M., Mirjalili, S., et al., 2020. Selective opposition based grey wolf optimization. Expert Syst. Appl. 151, 113389. https://doi.org/10.1016/j.eswa.2020.113389.

Eaton, B.A., 1972. The effect of overburden stress on geopressure prediction from well logs. J. Petrol. Technol. 24 (8), 929–934. https://doi.org/10.2118/3719-PA.

Eberhart-Phillips, D., Han, D.H., Zoback, M.D., 1989. Empirical relationships among seismic velocity, effective pressure, porosity, and clay content in sandstone. Geophysics 54 (1), 82–89. https://doi.org/10.1190/1.1442580.

Emary, E., Zawbaa, H.M., Hassanien, A.E., 2016. Binary grey wolf optimization approaches for feature selection. Neurocomputing 172, 371–381. https://doi.org/10.1016/j.neucom.2015.06.083.

Fawagreh, K., Gaber, M.M., Elyan, E., 2014. Random forests: from early developments to recent advancements. Syst. Sci. Control Eng.: An Open Access J. 2 (1), 602–609. https://doi.org/10.1080/21642583.2014.956265.

Fayed, H.A., Atiya, A.F., 2019. Speed up grid-search for parameter selection of support vector machines. Appl. Soft Comput. 80, 202–210. https://doi.org/10.1016/j.asoc.2019.03.037.

Fu, J., Wen, X.H., 2018. A regularized production-optimization method for improved reservoir management. SPE J. 23 (2), 467–481. https://doi.org/10.2118/189457-PA.

Guria, C., Goli, K.K., Pathak, A.K., 2014. Multi-objective optimization of oil well drilling using elitist non-dominated sorting genetic algorithm. Petrol. Sci. 11 (1), 97–110. https://doi.org/10.1007/s12182-014-0321-x.

Hottman, C.E., Johnson, R.K., 1965. Estimation of formation pressures from log-derived shale properties. J. Petrol. Technol. 17 (6), 717–722. https://doi.org/10.2118/1110-PA.

Hu, L., Deng, J., Zhu, H., et al., 2013. A new pore pressure prediction method-back propagation artificial neural network. Electron. J. Geotech. Eng. 18, 4093–4107. https://www.researchgate.net/publication/287438742.

Huang, H., Li, J., Yang, H., et al., 2022. Research on prediction methods of formation pore pressure based on machine learning. Energy Sci. Eng. 10, 1886–1901. https://doi.org/10.1002/ese3.1112.

Jorden, J.R., Shirley, O.J., 1966. Application of drilling performance data to overpressure detection. J. Petrol. Technol. 18 (11), 1387–1394. https://doi.org/10.2118/1407-PA.

Kuang, L., Liu, H., Ren, Y., et al., 2021. Application and development trend of artificial intelligence in petroleum exploration and development. Petrol. Explor. Dev. 48 (1), 1–14. https://doi.org/10.1016/S1876-3804(21)60001-0.

Li, D.L., Ma, J.F., Li, L., et al., 2022. Prediction method of formation pressure in abnormal pressure reservoir in Bozhong sag. Prog. Geophys. 37 (3), 1266–1273. https://doi.org/10.6038/pg2022FF0043 (in Chinese).

Mirjalili, S., Mirjalili, S.M., Lewis, A., 2014. Grey wolf optimizer. Adv. Eng. Software 69, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007.

Moazzeni, A., Haffar, M.A., 2015. Artificial intelligence for lithology identification through real-time drilling data. J. Earth Sci. Climatic Change 6 (3), 1–4. https://doi.org/10.4172/2157-7617.1000265.

Muther, T., Dahaghi, A.K., Syed, F.I., et al., 2022. Physical laws meet machine intelligence: current developments and future directions. Artif. Intell. Rev. 1–67. https://doi.org/10.1007/s10462-022-10329-8.

Panchal, G., Ganatra, A., Kosta, Y.P., et al., 2011. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. Int. J. Comput. Theor. Eng. 3 (2), 332–337. https://doi.org/10.7763/IJCTE.2011.V3.328.

Rashidi, M., Asadi, A., 2018. An artificial intelligence approach in estimation of formation pore pressure by critical drilling data. In: 52nd US Rock Mechanics/Geomechanics Symposium. OnePetro.

Rebentrost, P., Mohseni, M., Lloyd, S., 2014. Quantum support vector machine for big data classification. Phys. Rev. Lett. 113 (13), 130503.

Sayers, C.M., Smit, T.J.H., Van-Eden, C., et al., 2003. Use of reflection tomography to predict pore pressure in overpressured reservoir sands. In: SEG Technical Program Expanded Abstracts 2003. Society of Exploration Geophysicists, pp. 1362–1365. https://doi.org/10.1190/1.1817541.

Syed, F.I., Alshamsi, A., Dahaghi, A.K., et al., 2022a. Application of ML & AI to model petrophysical and geomechanical properties of shale reservoirs–A systematic literature review. Petroleum 8 (2), 158–166.

Syed, F.I., Alshamsi, M., Dahaghi, A.K., et al., 2022b. Artificial lift system optimization using machine learning applications. Petroleum 8 (2), 219–226.

Terzaghi, K., Peck, R.B., Mesri, G., 1996. Soil Mechanics in Engineering Practice. John

Wiley & Sons.
Torlay, L., Perrone-Bertolotti, M., Thomas, E., et al., 2017. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. Brain inform. 4 (3), 159–169. https://doi.org/10.1007/s40708-017-0065-7.
Wang, W., Men, C., Lu, W., 2008. Online prediction model based on support vector machine. Neurocomputing 71 (4–6), 550–558. https://doi.org/10.1016/j.neucom.2007.07.020.
Xu, Z., Yan, X., Yang, X., 2010. Casing life prediction using Borda and support vector machine methods. Petrol. Sci. 7 (3), 416–421. https://doi.org/10.1007/s12182-010-0087-8.
Yu, H., Chen, G., Gu, H., 2020. A machine learning methodology for multivariate pore-pressure prediction. Comput. Geosci. 143, 104548. https://doi.org/10.1016/j.cageo.2020.104548.
Zhang, G., Davoodi, S., Shamshirband, S., et al., 2022. A robust approach to pore pressure prediction applying petrophysical log data aided by machine learning techniques. Energy Rep. 8, 2233–2247. https://doi.org/10.1016/j.egyr.2022.01.012.