International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Estimation of Adsorption Gas in Shale Gas Reservoir by Using Machine Learning Methods

Dennis Sabato Chinamo[1], Xiao-Qiang Bian[2], Zongyang Liu[3], Jing Cheng[4], Lan Huang[5]

[1,2,3]Petroleum Engineering School, Southwest Petroleum University, Chengdu 610500, China
[4]School of New Energy and Materials, Southwest Petroleum University, Nanchong 637001, China
[5]Sichuan Province Jinhe Geological Exploration Engineering CO., LTD. Chengdu 610051, China

Abstract:
Evaluation and development of shale gas deposits need exact calculation of adsorbed gas concentration. However, gas adsorption and desorption experiments are time-consuming and costly. Models based on physics and empirical correlations cannot anticipate these experiments. Based on geological factors, this study intends to construct a cost-effective and accurate machine-learning model to estimate adsorbed shale gas. In this study, 601 data points from shale gas reserves were utilized. These include reservoir temperature (T,°C), total organic carbon (TOC, wt%), vitrinite reflectance (Ro,%), Langmuir pressure (PL), and volume. Based on Support Vector regression (SVR), Particle Swarm Optimization (PSO), Grey Wolf Optimization (GWO), and Sparrow Search Algorithms (SSA) were created, trained, and evaluated. The accuracy of the different models is calculated by correlation coefficient (R^2), root mean square error (RMSE), average absolute relative deviation (AARD), and the duration of time each specified model (PSO-SVR, GWO-SVR, and SSA-SVR) takes to evaluate the effectiveness of each prediction model.
The results showed that three optimization models (PSO-SVR, GWO-SVR, and SSA-SVR) can make good predictions. However, the PSO-SVR model is the most accurate at predicting Langmuir pressure and volume and takes the least amount of time, with RMSE and R^2 values of 0.09990 and 0.9605, respectively. The GWO-SVR has an estimated RMSE of 0.1092 and R^2 of 0.9529, whereas the SSA-SVR has 0.1264 and 0.9368 for Langmuir volume. PSO-SVR had the lowest time requirement for Langmuir pressure data inputs, with RMSE and R^2 of 0.5017 and 0.9306, respectively. The findings indicate that all models can accurately forecast adsorbed gas, with the PSO-SVR model performing somewhat better than the GWO- and SSA-SVR models

Keywords; Adsorption gas, Shale Gas, Machine Learning, Grey Wolf optimization (GWO), Particle swarm optimization (PSO), Sparrow search algorithm (SSA)

1. Introduction

Shale reserves have emerged as a prominent area of interest in global oil and gas exploration. Shale oil and gas production can mitigate energy tensions, promote equilibrium in the supply structure, and facilitate the expansion of oil and gas exploration. There has been a widespread increase in new theories and technological progress in the global field of shale gas exploration. The improvements have significantly influenced the global exploration and development of shale gas deposits [1, 2, 3, 4, 5]. Shale gas, referred

to as natural gas contained inside shale formations, is a usually environmentally sustainable energy supply, consisting predominantly of methane, this fuel exhibits a higher level of combustion efficiency in comparison to other hydrocarbon fuels [6, 7, 8, 9, 10, 11, 12]. The components consist of solution gas, adsorbed gas, and free gas. Adsorbed gas refers to a significant quantity of gas that has undergone adsorption onto the surface of organic matter and clays inside the shale formation. On the other hand, solution gas denotes the gas that has undergone dissolution in the reservoir water or oil and free gas refers to the gas that is found within the pore space of shale rock. Temperature and pressure are the two variables that govern the volume of solution gas, as the pressure decreases below the bubble point, the gas that is dissolved in the liquid begins to escape and transforms into free gas [13]. Shale gas production has undergone significant expansion in recent years. The United States had a shale gas production of approximately 1.6% in 2000 [14]. Nevertheless, there was a notable increase rapidly in this statistic, reaching 47% by the year 2015 [15]. The shale gas output in China in 2013 was limited to a volume of 200 million cubic meters. However, the volume of shale gas increased to 1.3 billion cubic meters in 2014 [16] and subsequently increased rapidly to 4.47 billion cubic meters in 2015 [12]. According to the Environmental Impact Assessment (EIA) in 2013, the projected quantity of technically feasible shale gas resources worldwide is 207 trillion cubic meters (m^3). Shale gas is expected to play an increasingly important part in the global energy supply soon.

The development and evaluation processes for shale gas are fraught with numerous obstacles. Providing an accurate estimate of resources is one of the most pressing issues. As an example, the shale gas resources in China are estimated to be 31.6 trillion m^3 by the Chinese Ministry of Land and Resources and 25.1 trillion m^3 by the U.S. Energy Information Administration (EIA) [17, 18]. The ambiguity surrounding the calculation of resources will impact not only the selection of gas well sites at a small scale but also have an influence on the national and industrial policy-making process at a large scale. Therefore, it is imperative to devise a method to precisely assess shale gas reserves globally. Various techniques have been devised to estimate gas adsorption in shale formations, including experimental measurements, physics-based methods, and empirical correlations. Nevertheless, each of these solutions possesses undesirable drawbacks. Experimental measurements can be costly and time-consuming [19]. Moreover, the limited comprehension of the impact of several parameters on adsorption, such as reservoir temperature and shale mineralogy, renders physics-based or first-principal techniques highly unreliable in numerous instances. Researchers have attempted to find empirical relationships that accurately describe the adsorption properties of shale gas in various shale gas formations. This includes developing linear and nonlinear equations to represent the Langmuir pressure and volume in the Langmuir model [20] and the BET model [21]. The Langmuir model is widely acknowledged as a suitable method for fitting experimental data of shale gas adsorption, as supported by [22, 23]. The Langmuir model is a theoretical framework that describes the behavior of adsorption on a surface.

$$V = \frac{V_L}{1 + p_L / p} \tag{1}$$

where V is the adsorbed gas volume per ton of rock mass in m^3/t (cubic meter per ton of rock) at pressure p, and V_L is the Langmuir volume in m^3/t, which is the largest monolayer adsorption. P_L is the Langmuir pressure in MPa at which $V = V_L/2$. Quantitative relationships between P_L and V_L have been the subject of much research. As an illustration, in their study, [24] they developed the formula for Langmuir pressure $1/P_L = \exp(a/T + c)$ using temperature T, and they utilized TOC to get the predicted Langmuir volume $V_L = a \cdot TOC^b$, with a, b, and c serving as coefficients. [25] used TOC as a single predictor for both P_L and

V_L, resulting in $P_L = a \cdot TOC^b$ and $V_L = a \cdot TOC^b$. The small sample sizes and lack of other important geological characteristics in this research limit these models' predictive power and practicality. An equation for $P_L = \exp(a \cdot TOC^* + b \cdot \ln(T^*/R_O) + c)$ and an equation for $V_L = \exp(a \cdot TOC^* + b \cdot T^{*3} + c)$ was proposed [26] after non-dimensionalizing T, TOC, and Ro using statistical data analysis. which significantly reduced the average relative errors for Langmuir pressure and volume prediction as compared to previous models.

[27, 28, 29] They found further empirical correlations to forecast gas adsorption. However, these investigations were limited by the small number of data entries employed in each study, which was less than 70. For instance, [29] constructed an empirical model. The correlation was employed to estimate the Langmuir volume based on the TOC and total clay content. The anticipated Langmuir volume was next utilized to forecast the Langmuir pressure. Nevertheless, this model relied solely on 66 data points from China and Western Australia during its development and was restricted to making predictions specifically at room temperature [30]. Machine learning algorithms for estimating shale gas adsorption were sought by [31]. Models for the prediction of Langmuir parameters were established using artificial neural networks (ANN), multiple linear regression (MLR), and support vector machine (SVM). Their previous approach was There are 485 entries for the Langmuir volume submodel and 213 for the Langmuir pressure submodel. The data were acquired from shale reservoirs located in the United States, Canada, and China. With a training-to-testing ratio of 75:25, they discovered that RF produced the best accurate models for both volume and Langmuir pressure (R^2=0.84) and volume (R^2 0.81). Vitrite reflection (Ro), total organic carbon (TOC), and temperature (T) were the model input data utilized for the Langmuir pressure forecast. Whereas, temperature (T) and total organic carbon (TOC) needed for Langmuir volume prediction,

[32] constructed artificial intelligence models utilizing TSK-FIS, M-FIS, FNN, and SVM approaches. These models were trained using a dataset of more than 800 data points obtained from well logs and core data. The models were created and trained using data obtained from The Barnett Shale and were evaluated using Devonian shale data that had not been previously incorporated into the model. The FNN model had superior performance compared to the other models, achieving an average absolute percentage error (AAPE) of 12.02%. In contrast, M-FIS, SVM, and TSK-FIS exhibited higher AAPE values, with TSK-FIS having the highest AAPE of 15.62%.

[26] Created a statistical learning model using 301 data points from 19 shale reservoirs located in the United States, Canada, China, Germany, Sweden, Netherlands, and Brazil. The Langmuir pressure submodel utilized temperature (T), total organic carbon (TOC), and vitrinite reflection (Ro) as parameters. On the other hand, the Langmuir volume submodel only considered temperature (T) and total organic carbon (TOC). The Langmuir pressure and volume were predicted with average absolute errors of 27.67% and 23.76%, respectively.

To enhance the prediction of Langmuir pressure and Langmuir volume for gas adsorption computation in shale reservoirs using major geological factors such as T, TOC, and Ro, we created a suite of machine learning models. The first step is curating over 601 experimental data items derived from published experimental results. To further analyze the relationship between the Langmuir pressure and volume and geological characteristics, multilinear regression and machine learning models such as Grey Wolf Optimization (GWO), Particle swarm optimization (PSO), and Sparrow search algorithm (SSA) are next constructed. Lastly, the best models that can estimate gas adsorption values using reasonable and accurate Langmuir pressure and Langmuir volume are chosen by comparing and discussing the findings from these three regression approaches based on the correlation coefficient and model stability.

2. Methodology

Regression models are constructed to ascertain the quantitative correlations between Langmuir and geological factors. The model-building process comprises two distinct components: Variable selection. The Langmuir equation utilizes the variables P_L and V_L to represent the adsorption capacity of the shale formation. V_L refers to the greatest amount of gas that can be adsorbed, while P_L represents the pressure at which the adsorbed gas volume is half of V_L [20]. The Langmuir parameters may be determined in this study by utilizing geological characteristics, as there exists a robust correlation between the two. The selection of geological variables is based on the availability of data and their impact on P_L and V_L. Recent research has indicated that the primary elements that have a significant impact on P_L are the temperature of the reservoir (T), the amount of total organic carbon present (TOC), and the level of vitrinite reflectance (R_O) which is an indicator of the thermal maturity of the organic matter. Furthermore, the variables T, and TOC, exert influences on V_L, [13].

2.1 Support Vector Regression (SVR)

The Support Vector Regression (SVR) approach is utilized due to its capacity to effectively handle regression models for both linear and nonlinear input data, employing mathematical formulations known as kernels. The poly kernels employed in the Support Vector Regression (SVR) technique reorganize the input data according to its high-dimensional space while maintaining computational efficiency. According to [33], Support Vector Regression (SVR) employs the Vapnik-Chervonenkis (VC) theory to adopt a robust-insensitive approach to the loss function. This approach involves the creation of a symmetrical flexible tube, known as the ε-tube, with a minimal radius around the estimated function. The ε-tube penalizes both high and low misestimates, with data points located outside the tube being considered outliers. The value determines the width of the tube, where a smaller value corresponds to a lower error tolerance, while a larger value corresponds to a larger error tolerance. This value has an impact on the sparsity of the solution. The primary goal of Support Vector Regression (SVR) is to identify the optimal regression hyperplanes, which serve as decision boundaries for predicting the continuous output. The kernel is utilized to locate these hyperplanes within the high-dimensional input space. The ε value provides the distance between the boundary lines and the hyperplane, as previously mentioned. The convex optimization is performed using the iterative approach of sequential minimum optimization (SMO). The primary formulation of kernel SVR is specified as the objective function, [34]
denoted as

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i + \xi_i^* \tag{2}$$

Such that

$$w = \sum_{i=1}^{l}(a_i^* - a_i)\varphi(x_i) \tag{3}$$

Where l represents the number of training patterns, C is a tradeoff or regularization parameter that balances model complexity and training error, ξ_i and ξ_i^* are slack variables used to handle outliers above the ε range, and w denotes the weight vector, α_i, and α_i^* represents Lagrange multipliers are mathematical terms used in optimization problems. The function $\varphi(x_i)$ is a mapping function that transforms the input feature space x_i into a different space called the kernel space [33, 35] to solve the objective function and convert the optimization problem into its Wolfe dual form of the optimization problem.

$$\max_{\alpha,\alpha^*} \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) - \varepsilon \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(x_i, x_j) \quad (4)$$

Such that
$$\sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0$$
$$c \geq \alpha_i, \alpha_i^* \geq 0, i = 1.....,l \quad (5)$$

The target for the input vector x_i is represented by y_i and the kernel function $k(x_i, x_j)$ calculates the inner product $\langle \varphi(x_i), \varphi(x_j) \rangle$. Equations (5) and (6) can be solved using the SMO algorithm, as described by [35]. The regression hyperplane is represented by the function approximation.

$$f(x) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)k(x_i, x) \quad (6)$$

Where $k(x_i, x) = \varphi(x_i) \cdot \varphi(x)$

The accuracy of the model is then calculated by the normal approaches such as correlation coefficient (R^2) where the closer R^2 is to 1, the better the performance of the model is, root mean square error (RMSE) The closer the value of RMSE is to zero, the better the performance of the model, and average absolute relative deviation (AARD) The closer the value of AARD to zero, the better the performance of the model.

2.2. Particle swarm optimization (PSO)

Kennedy and Eberhart introduced PSO in 1995; it finds the best places to eat by analyzing fish movement patterns and bird aggregations. By updating a particle swarm iteratively and randomly initializing possible solutions, the technique addresses the global optimization problem. Particles relocate inside the cluster according to their previous best position ("pbest") and the global best position ("gbest") to discover the optimal solution [36]. PSO's low computational cost, simple implementation, and low CPU performance requirements make it a popular optimization strategy. The rules for computing the prior best position ("pbest") and global best position ("gbest") of each particle are:

$$pbest(i,t) = \arg\min[f(P_i(\mathrm{k})], \mathrm{i} \in \{1,2...,\mathrm{N}_p\} \quad (7)$$

$$gbest(t) = \arg\min[f(g_i(\mathrm{k}))] \quad (8)$$

The equation is as follows: i is the particle index, N_p is the total number of particles, t is the current iteration, f is the fitness function and p is the position. The following equation calculates the velocity of the particle (V) and position (p)

$$V_i(t+1) = \omega V_i(t) + c_1 t_1(pbest(i,t) - P_i(t)) + c_2 t_2(gbest(i) - P_i(t)) \quad (9)$$

$$P_i(t+1) = P_i(t) + V_i(t+1) \quad (10)$$

The formula $P_i(t)$ represents the particle current position, ω is the inertial weight that balances exploration and local development, and r_1 and r_2 are uniformly distributed random variables ranging from 0 to 1. The "acceleration coefficient" refers to the learning factors c_1 and c_2, typically equal to 2. In speed V, two approaches are commonly employed to limit speed: Velocity clamping and constriction coefficients.

Formula (9) includes “inertia”, “cognition” and “cooperation”, which enable particles to move in space with goals and find the optimal solution.

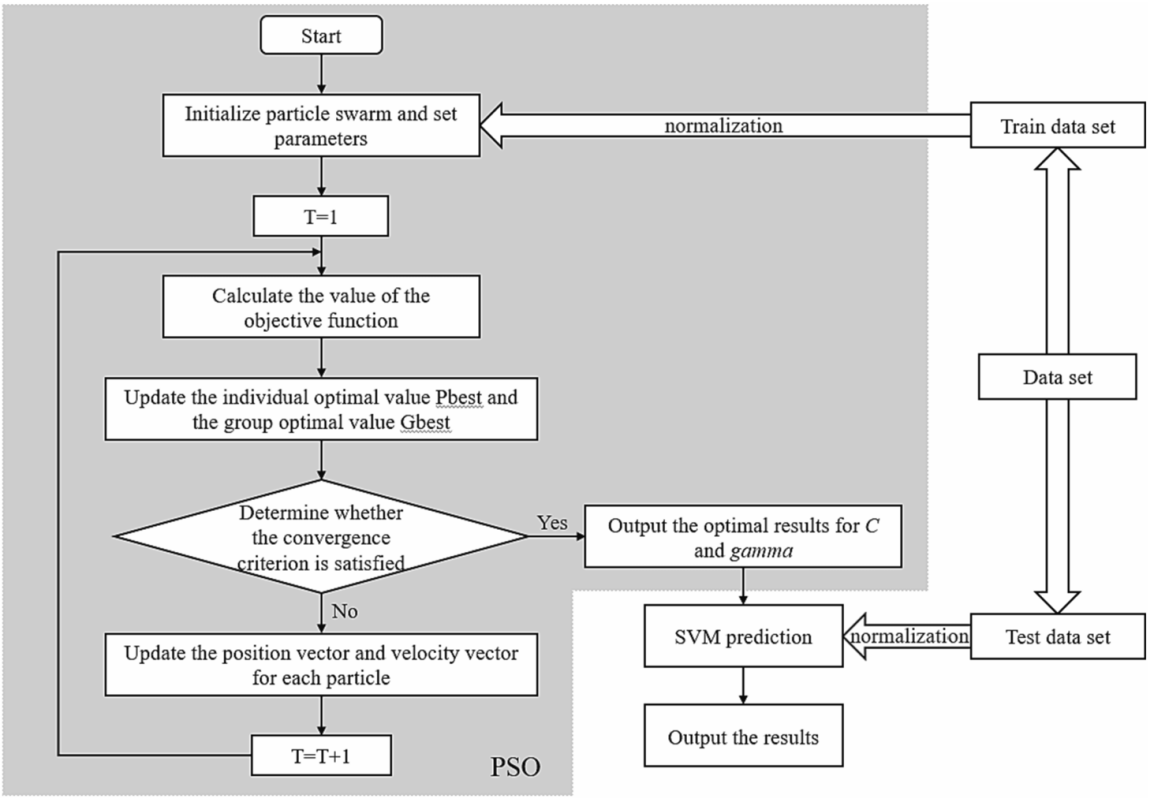

Fig.1 Flowchart of the PSO-SVR model

2.3. Grey Wolf optimization (GWO)

In 2014, Mirjalili and Lewis presented the GWO algorithm, which is based on grey wolves' social leadership and hunting ability. Grey wolves are considered apex predators. Fig 2 illustrates that these creatures are sociable and follow a rigid order of dominance. In this group, there are four types of members: α is the head of this group and belongs to the first social level. Decisions on when to wake up, where to sleep, and whether to hunt are primarily made by it. α is aided in decision-making and other group activities by β, who is situated on the second social level. The α Wolf candidate of the pack is after that β when α wolves reach maturity or pass away. The role of β is to advise α Wolf, administer discipline to the pack, obey α Wolf's commands, and issue commands to other wolves at the third and fourth social levels. In this set, β gives α informational input and also reinforces α's directive. Scouts, whistling wolves, long wolves, hunting wolves, and guarding wolves are all part of the third social level, which is represented by δ. In the first and second social levels, they were obligated to obey the α and β wolves, respectively. In the fourth level, they were given the ability to command the omega wolves.

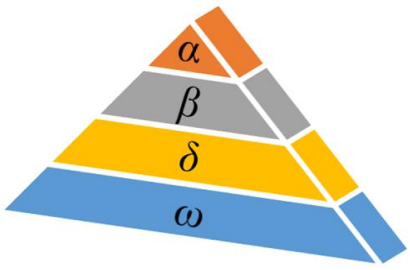

Fig.2 Grey Wolf Hierarchy

The fourth social level, which is the lowest level in the group, is occupied by ω As a scapegoat, it must carry out the bidding of every higher-ranking wolf. There is a small role for ω in the group, but issues with group conflict are inevitable in the absence of ω.

The social structure of a grey wolf pack is just as important as their hunting techniques, which include following, encircling, and eventually attacking prey. As previously said, this is the procedure that the GWO optimization algorithm will adhere to. Order in the social hierarchy is paramount, followed by the pursuit, encirclement, and eventual attack of prey. We rank the solutions as follows: α Wolf for the fittest, β Wolf for the second-best, δ Wolf for the third-best, and ω Wolf for the remaining candidates. The optimization process of the GWO method is led by α wolves, β wolves, and δ wolves, while the remaining ü wolves adhere to these three kinds of wolves. To model the process of pursuing, encircling, and ultimately attacking a target, the following equation is utilized:

$$D = C^{\circ} X_p(t) - X(t) \tag{11}$$

$$X(t+1) = X_p(t) - A^{\circ} D \tag{12}$$

Where $X_P(t)$ is the prey position vector and X is the grey wolf position vector, t is the number of current iterations, A and C are vectors representing the synergy coefficients. This equation determines the vectors A and C.

$$A = 2a^{\circ} r_1 - a \tag{13}$$

$$C = 2r_2 \tag{14}$$

While r_1 and r_2 are arbitrary vectors in the interval [0,1], elements decrease linearly from 2 to 0 as the iteration progresses. To discover the best answer to the optimization problem, the GWO algorithm used the previously mentioned encircling prey mechanism in conjunction with the simulated social hierarchy leadership. As a result, other search agents are compelled to revise their positions concerning the top three solutions that have been saved by the algorithm. The optimization process involves. To simulate search results and identify interesting parts of the search space, the following formula is continually executed on each search agent [37]

$$D_{\alpha} = C_1^{\circ} X_{\alpha} - X \tag{15}$$

$$D_{\beta} = C_2^{\circ} X_{\beta} - X \tag{16}$$

$$D_{\delta} = C_3^{\circ} X_{\delta} - X \tag{17}$$

$$X_1 = X_{\alpha} - A_1^{\circ} D_{\alpha} \tag{18}$$

$$X_2 = X_{\beta} - A_2^{\circ} D_{\beta} \tag{19}$$

$$X_3 = X_\delta - A_3^\circ D_\delta \tag{20}$$

$$X = (t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{21}$$

Using this equation, we can revise all of the possible solutions. In summary, the GWO method determines α Wolf, β Wolf, and δ Wolf through the calculation of the corresponding fitness. Then, using the best three solutions, it calculates the potential position of the prey. Consequently, α Wolf, β Wolf, and δ Wolf determine the potential whereabouts of prey, and the other wolves adhere to their lead. At last, the GWO algorithm is stopped when the termination condition is satisfied [38].

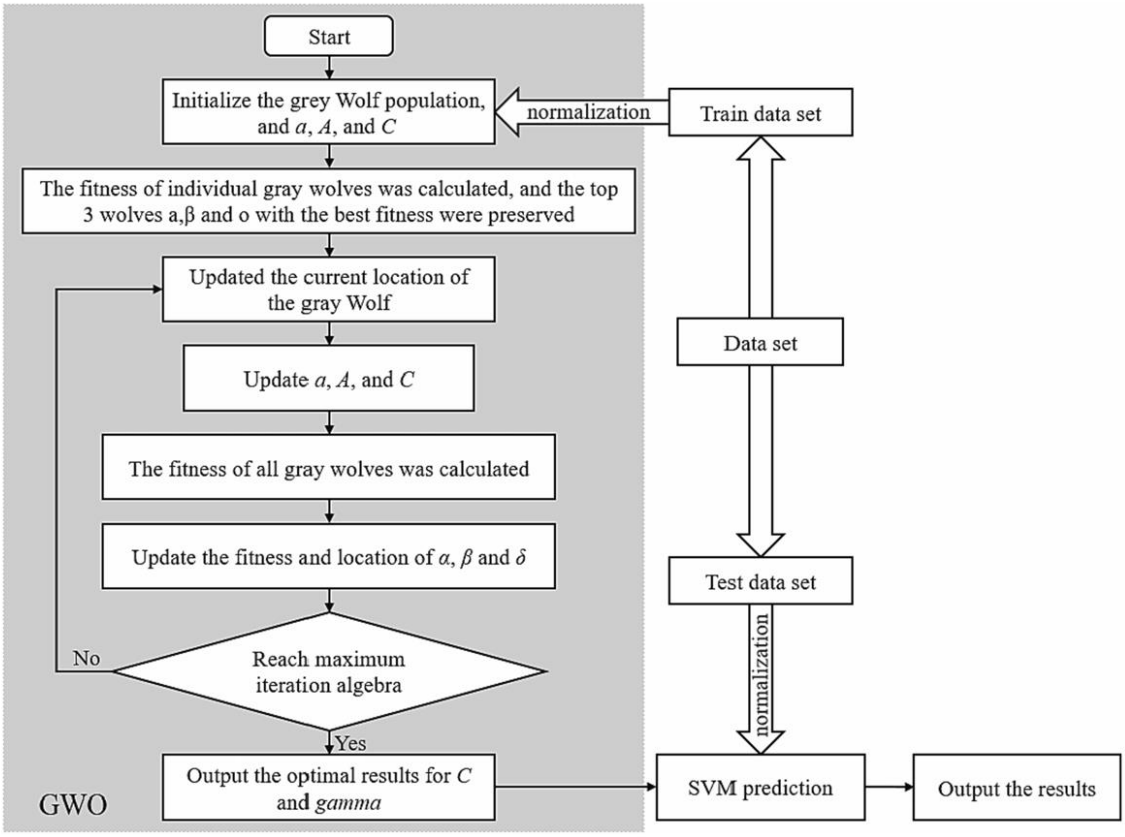

Fig.3 GWO algorithm mechanism flow chart.

2.4. Sparrow search algorithm (SSA)

The Sparrow Search Algorithm, or SSA, was suggested by Seyedali and Mirjalili. The following are the key points of SSA. First It all starts with initializing the individual positions of the population. second, the ideal person is chosen based on the optimum fitness value that has been computed. Finally, the best person is chosen to lead, and their position is changed to that of the food source. To put it another way, this is the same as keeping track of where each food supply is at its ideal throughout the iteration, which makes it harder for followers to reach that spot and improves algorithm convergence [39] The sparrow's position is determined by the following matrix

$$X = \begin{Bmatrix} X_{1,1}X_{1,2}.....X_{1d} \\ X_{2,1}X_{2,2}.....X_{2d} \\ X_{n,1}X_{n,2}.....X_{n,d} \end{Bmatrix} \tag{22}$$

The equation (22) represents the relationship between the number of sparrows (n) and the dimension (d) of the variable being optimized. The fitness values of all sparrows are denoted by the following vectors:

$$f_x = \begin{bmatrix} (\{X_{1,1}X_{1,2}.....X_{1,d}\}) \\ (\{X_{2,1}X_{2,2}.....X_{2,d}\}) \\ (\{X_{n,1}X_{n,2}.....X_{n,d}\}) \end{bmatrix} \quad (23)$$

In Equation (23), the variable n denotes the quantity of sparrows, while the value of each row in FX signifies the fitness value of the birds. Within the SSA, producers with superior health values are given precedence in accessing food throughout the search procedure. Furthermore, producers, who bear the responsibility of locating sustenance and guiding the mobility of the whole population, can get food from a diverse array of locations. The location of the producer during each iteration is decided by the following formula, as specified by the first two rules:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \times \exp\left(\frac{-i}{\alpha \times iter_{\max}}\right), R_2 < ST \\ X_{i,j}^{t} + QL, R_2 \geq ST \end{cases} \quad (24)$$

Equation (24) uses the current iteration count, denoted as t, and a random variable j = 1, 2, 3, 4... d. $X_{i,j}^{t}$ represents the JTH dimension in the iteration t. R^2, which stands for the warning value, is a random integer in the interval [0,1], and α is a random number between 0 and 1. $iter_{\max}$ is the constant with the highest number of repetitions. Additionally, the value of the safety valve, denoted as ST, is a random integer between 0.5 and 1. Q is a normally distributed integer, and L is a 1 × d matrix with 1 as the value of each member. The producer goes into prolonged search mode when $R^2 < ST$, which indicates that there are no predators around. If R^2 is more than or equal to ST, it means that sparrows have seen a predator and must swiftly fly to a safer place, and some of the scavengers keep their eyes on the producers. Once they see that the producers have better food, they start competing with the producers.

These scavengers perform the following formula:

$$X_{i,j}^{t+1} = \begin{cases} Q \times \exp\left(\frac{X_{worst}^{t} - X_{i,j}^{t}}{i^2}\right), i < \frac{n}{2} \\ X_{P}^{t+1} + \left[X_{i,j}^{t} - X_{i,j}^{i+1}\right].A^{+}.L \end{cases} \quad (25)$$

Equation (25) gives the producer's optimal location as X_P and the poorest possible position in the current area as X_{worst}. A means 1 × d matrix, where each entry is -1 and $A^{+}=A^{T}(AA^{T})^{-1}$, When i >n/2, the scavengers with poor fitness value of i are the most likely to starve.

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta\left(X_{i,j}^{t} - X_{best}^{t}\right), f_i > fg \\ X_{i,j}^{t} + K \times \left(\frac{X_{i,j}^{t} - X_{worst}^{t}}{(f_1 - f_w) + \varepsilon}\right), f_i = f_g \end{cases} \quad (26)$$

Equation (26) indicates that the current optimum location, X_{best}, is determined by the step size control parameter, β, whose value is a normally distributed random integer with an average of 0. The value of K may be any integer between -1 and 1. f_g and f_w stand for the optimal and optimally worst fitness values, respectively [40]

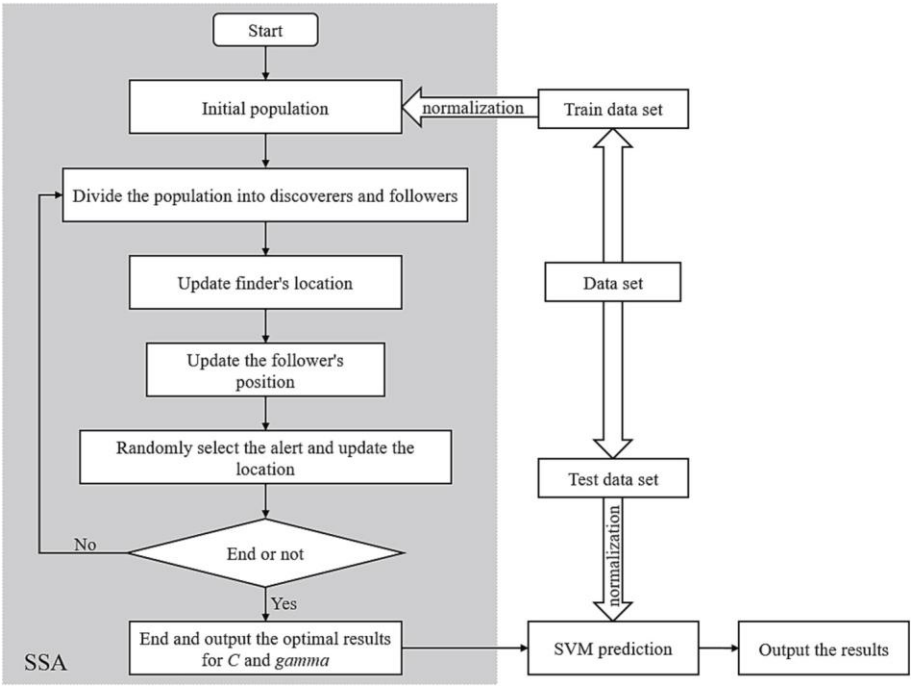

Fig4. SSA algorithm mechanism flow chart

3. Data description and preprocessing

Since increasing the amount of data forms the basis of statistical learning, doing so is essential for creating a reliable estimating model. Most current models require more than 10 data points to build a reliable model. This study compiles raw data points from many experiments and studies to circumvent the data size limitation [41, 45, 42, 43, 44, 49, 13, 47, 50, 46] .Table 1 shows the Langmuir volume data and Langmuir pressure data used in this research. The variety of data sources makes the model suitable for general use, before using the raw data to develop models, it needs to be cleaned, two simple processes comprise data preparation in this study.

Table 1. Data source information

	Reference	Number of data points
V_L data	[41]	336
	[13]	6
	[42]	12
	[43]	12
	[44]	55
	[45]	40
	[46]	8
	[47]	41

	[48]	16
P_L data	[45]	40
	[43]	12
	[49]	4
	[50]	19

The first process is to ensure that each data point contains both dependent and independent variables The analysis involved examining the data entries on Langmuir pressure with temperature T, TOC, and Ro, as well as the Langmuir volume with T and TOC. As more variables are merged for analysis, the data set size continues to decrease because not all variables are available in the data entries. The second process was geological characteristics were reduced to more practical ranges by considering the actual reservoir conditions of the shale gas fields currently being developed commercially. In this research, the measurement temperature is limited to 122.3 °C, the total organic carbon (TOC) ranges from 0% to 11.7%, the Langmuir pressure is limited to 0 to 39.43 MPa, and the maximum recorded Ro value is 4.2%. Reservoirs with higher Ro values are indicative of being over-matured. Upon applying the specified limits to the data, a total of 75 entries for Langmuir pressure data and 526 entries for Langmuir volume data remain for regression analysis. The pre-processed data is randomly divided into training data and testing data according to the ratio of 8:2.

4. Results and discussion

To find out how well an intelligent model works, developers often use statistical assessment techniques. R^2, RMSE, and AARD% are three statistical measures used in this research to verify and confirm the developed integrity of the model. Table 2 shows the parameters, where y_i represents the predicted value, $\hat{y}$ represents the anticipated value, $\bar{\hat{y}}$ represents the mean of the expected value, and n is the number of data pieces. Along with this, the runtime is another factor in evaluating its performance.

Table 2. Performance metrics

Name	Abbreviation	Expression
Correlation coefficient	R^2	$1-\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 / \sum_{i=1}^{n}(y_i-\bar{\hat{y}})^2$
Root mean square error	$RMSE$	$\sqrt{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 / n}$
Average absolute relative deviation	$AARD\%$	$\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{y_i-\hat{y}_i}{\hat{y}_i}\right\|\times 100$

Assuming RMSE and AARD values are close to 0, and R^2 is close to 1, the suggested model becomes very accurate and exact. All data, training data, and test data for all three systems and the full dataset are represented by the values of the statistical parameters in Fig 10 and Fig 16. A lower AARD and RMSE value, together with a higher R^2 value, indicates that the proposed model is accurate for each system, as shown in Fig 10 and Fig 16. There are separate entries for the analysis of Langmuir pressure and Langmuir volume. T, TOC, and Ro are the geological parameters that are examined for Langmuir pressure, while T

and TOC are the ones that are examined for Langmuir volume. The same methods have been used to train and test PSO-SVR, GWO-SVR, and SSA-SVR models for both of them.

Langmuir volume V_L

Analysis of 526 data items, including T, TOC, and VL, demonstrates that PSO-SVR models outperform other models. Additionally, GWO-SVR and SSA-SVR exhibit superior predictive capabilities, with the order of performance being PSO-SVR > GWO-SVR > SSA-SVR. The correlation coefficient R^2 of PSO-SVR models on the total data set is 0.9605, which indicates a very robust link established for these 526 data items. The Average Absolute Relative Deviation (AARD%) obtained from the PSO-SVR models is 3.28 and the RMSE is 0.0999 indicating significantly superior performance compared to other models in predicting Langmuir volume.

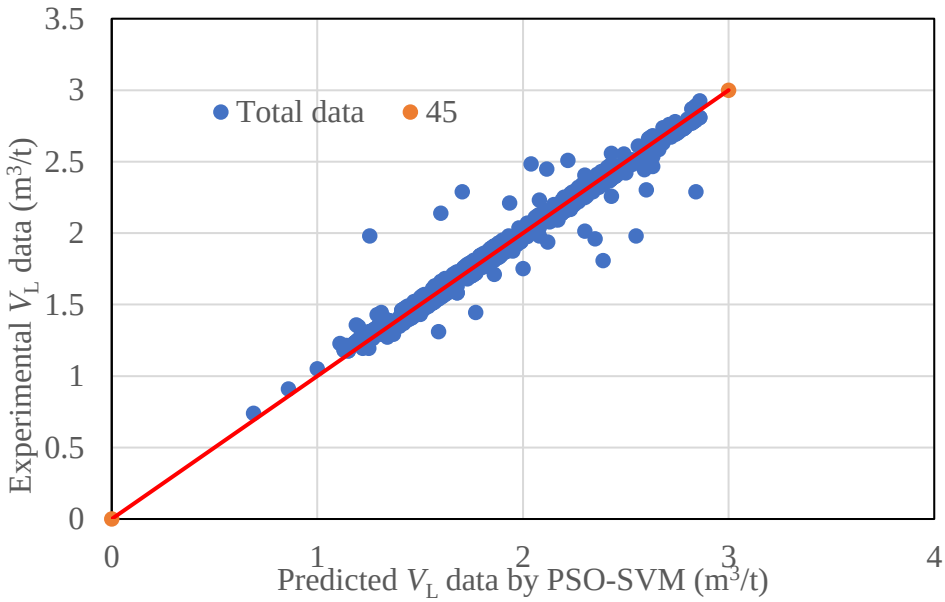

Fig. 5 Experimental and predicted Langmuir volume values from PSO-SVR.

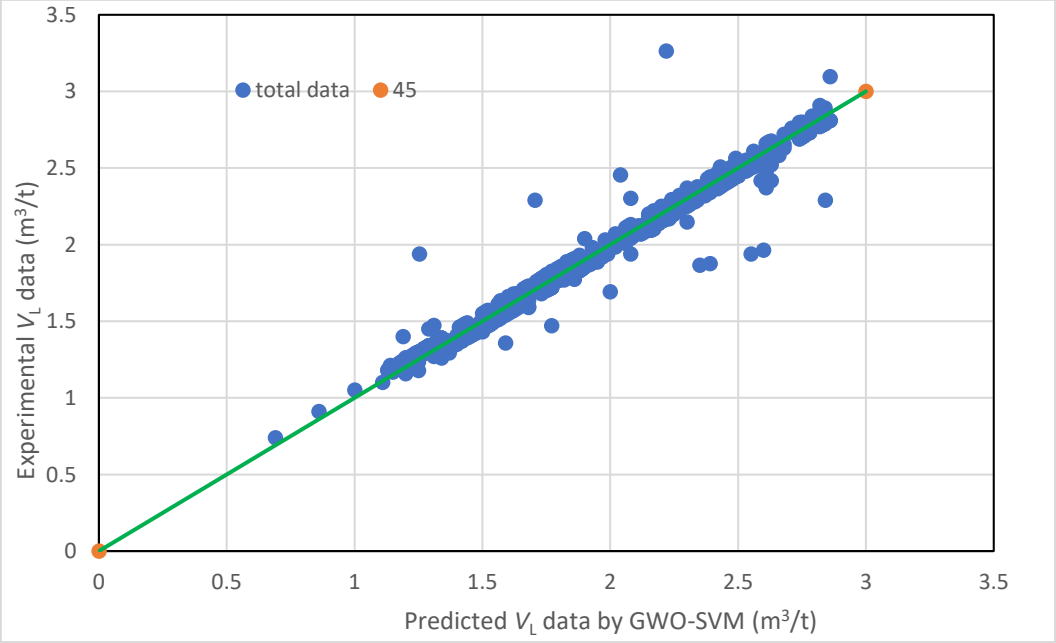

Fig.6 Experimental and predicted Langmuir volume values from GWO-SVR.

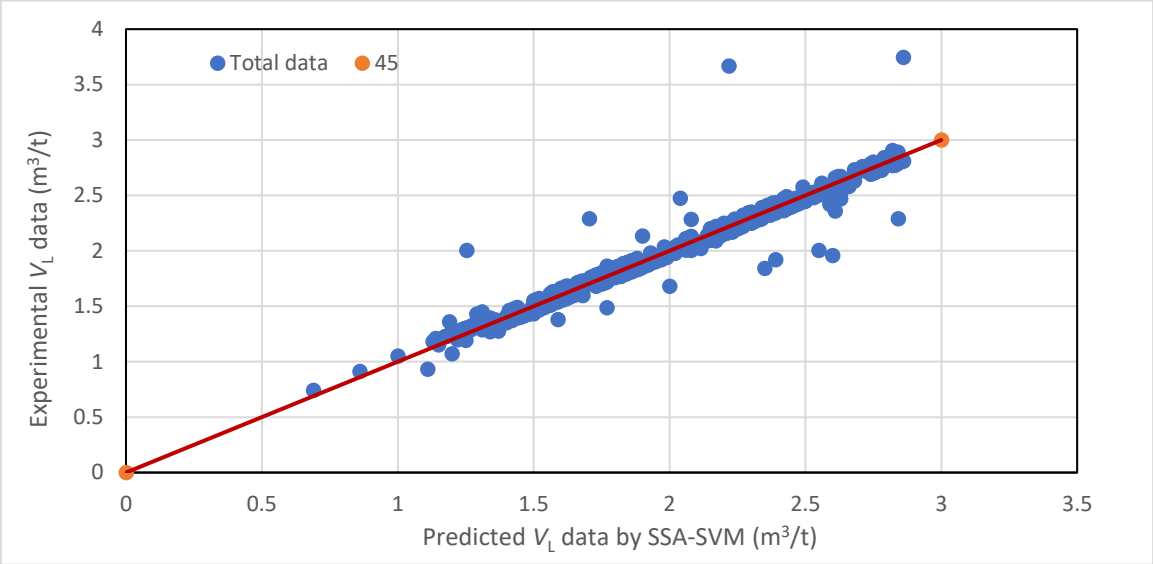

Fig.7 Experimental and predicted Langmuir volume values from SSA-SVR.

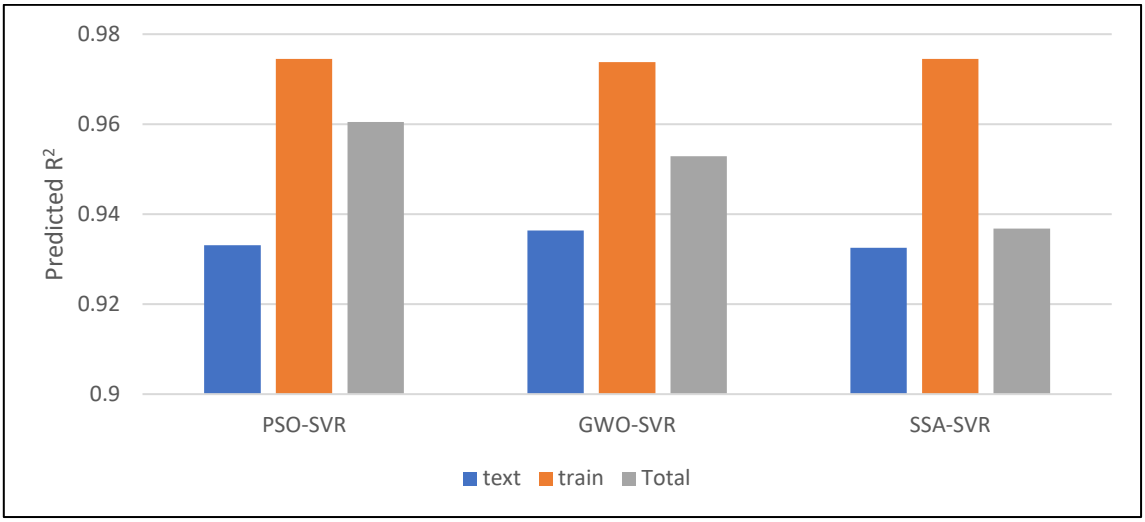

Fig. 8. Correlation coefficient R^2 for PSO-SVR, GWO-SVR, and SSA-SVR models for V_L

Fig. 9. Average Absolute Relative Deviation AARD% for PSO-SVR, GWO-SVR, and SSA-SVR models for V_L

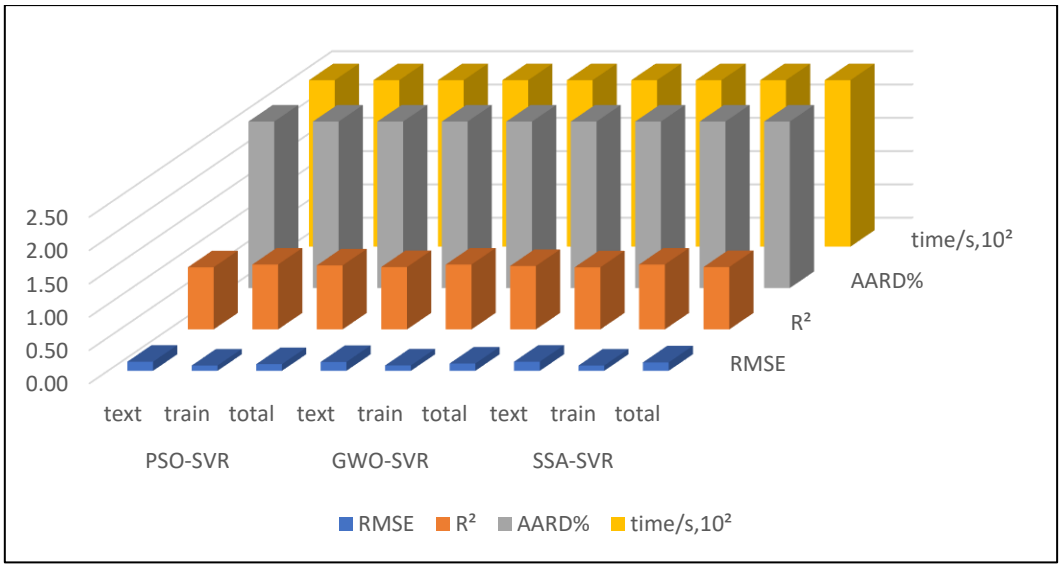

Fig .10. Statistical analysis and comparison of different ML models for V_L

Langmuir Pressure P_L

Analysis of 75 data items, including T, TOC, Ro, and P_L, demonstrates that PSO-SVR models outperform other models. Additionally, GWO-SVR and SSA-SVR exhibit superior predictive capabilities with the order of performance being PSO-SVR > GWO-SVR > SSA-SVR. The correlation coefficient R^2 of the PSO-SVR model on the whole data set is 0.9309, which indicates a very robust link established for these 75 data items. The Average Absolute Relative Deviation (AARD%) obtained from the PSO-SVR models is 7.68, and RMSE is 0.5017 with a minimum time of 5.85 seconds indicating significantly superior performance compared to other models in predicting Langmuir volume.

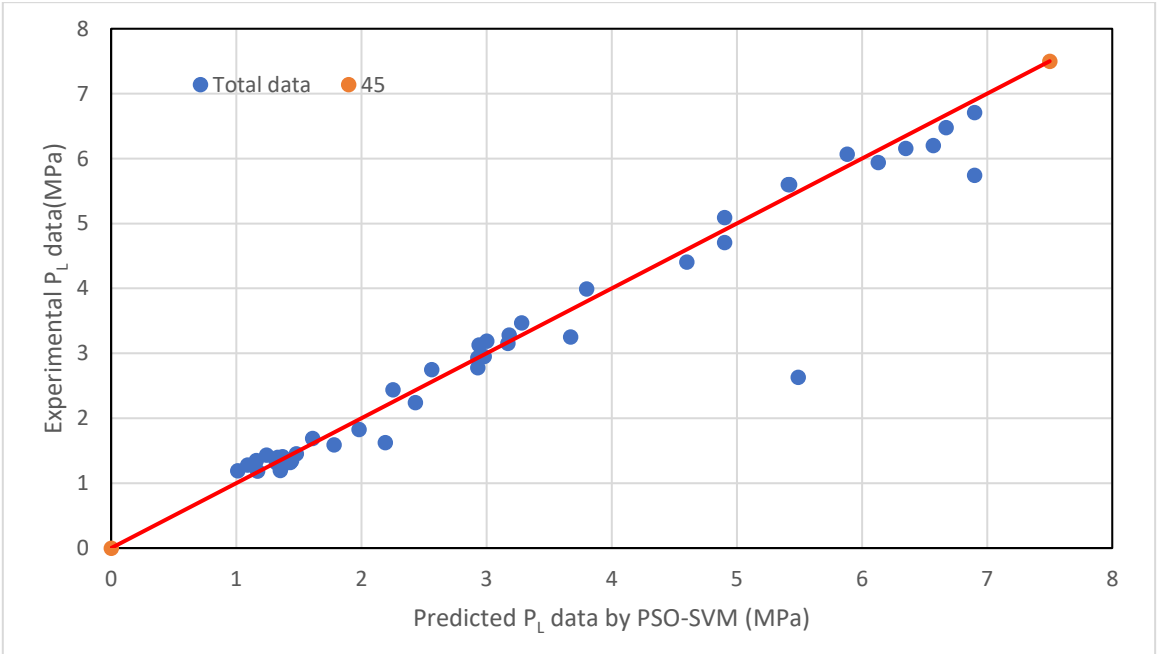

fig. 11. Experimental and predicted Langmuir pressure values from PSO-SVR.

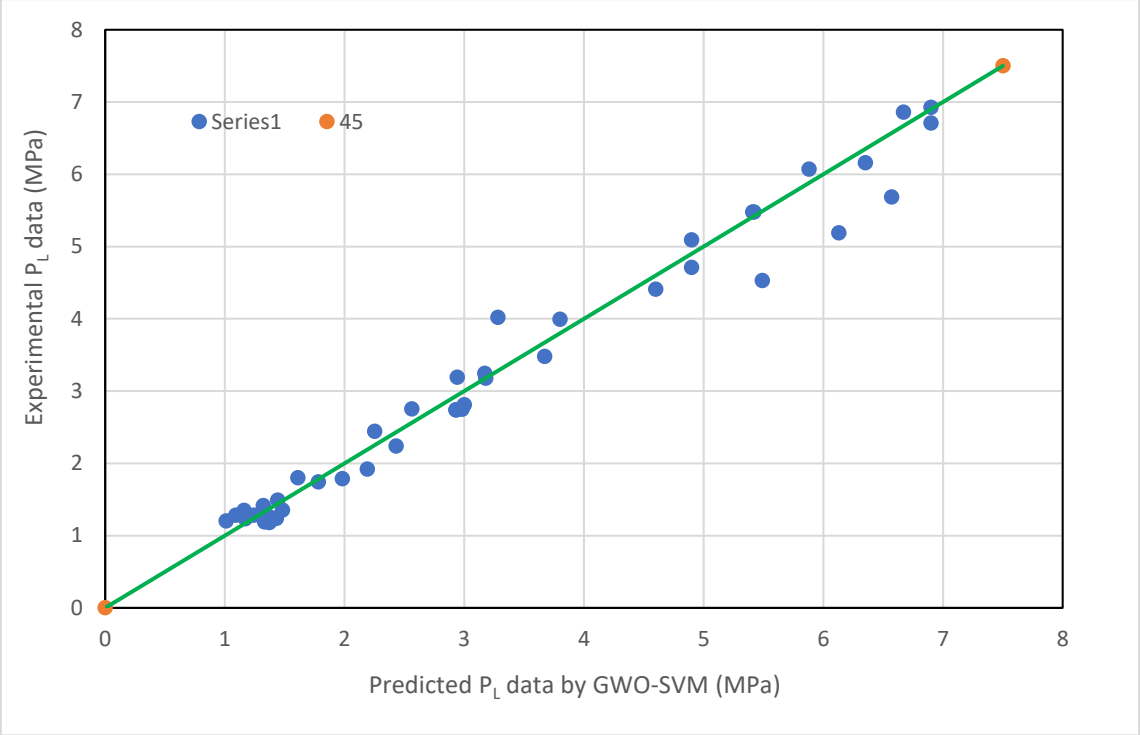

Fig.12.Experimental and predicted Langmuir pressure values from GWO-SVR.

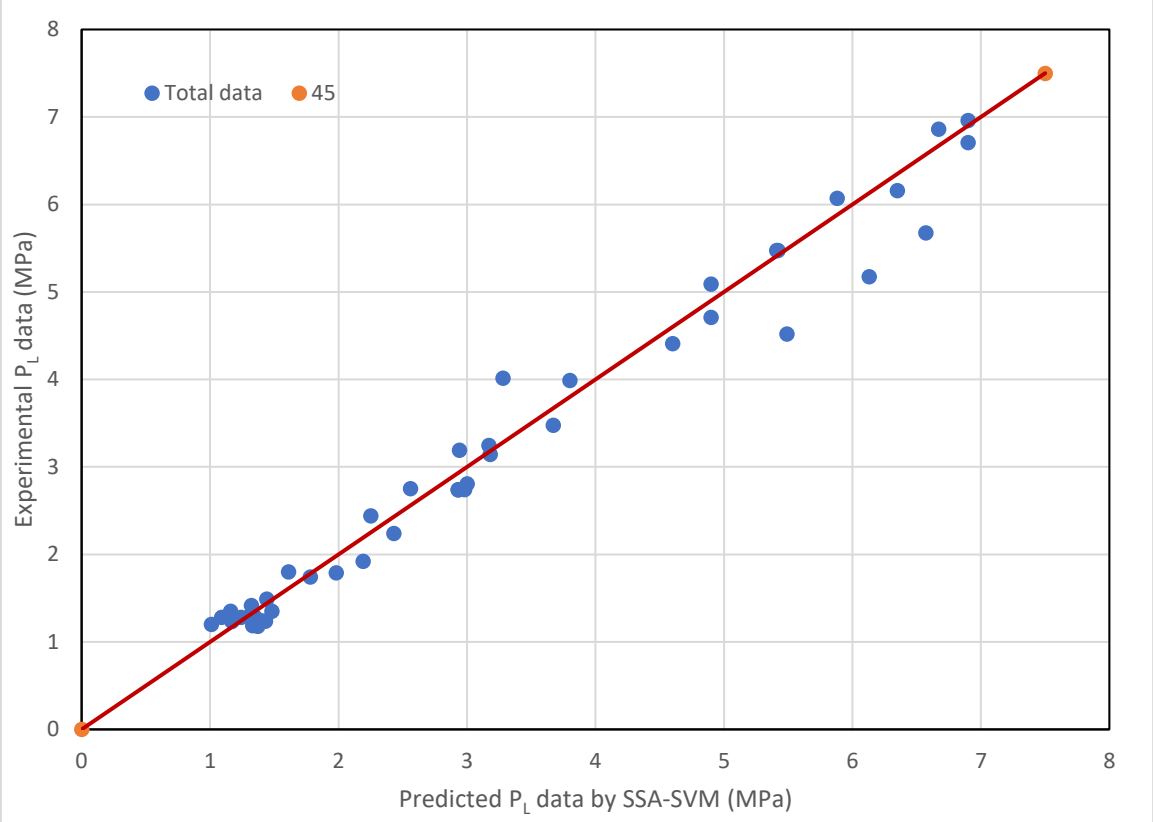

Fig 13. Experimental and predicted Langmuir pressure values from SSA-SVR.

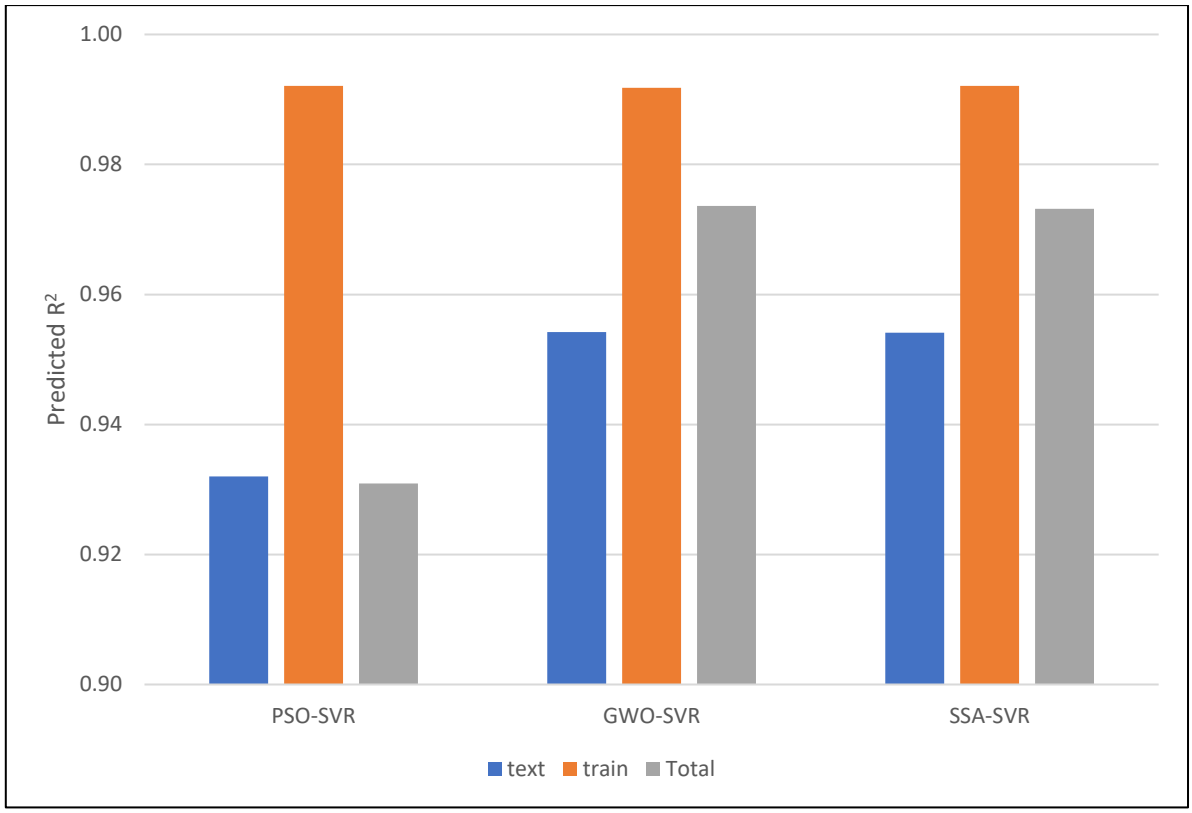

Fig 14. Correlation coefficient R^2 for PSO-SVR, GWO-SVR, and SSA-SVR models for P_L

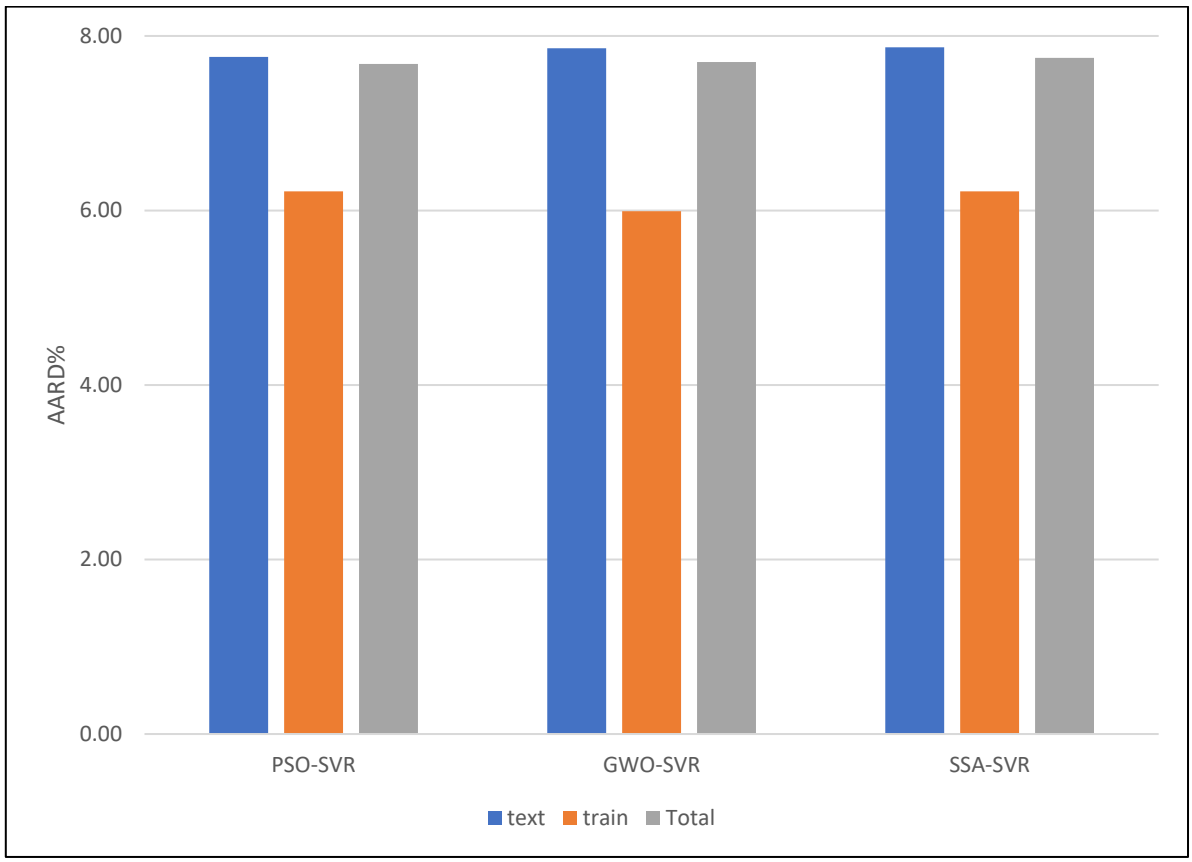

Fig.15. Average Absolute Relative Deviation AARD% for PSO-SVR, GWO-SVR, and SSA-SVR models for P_L

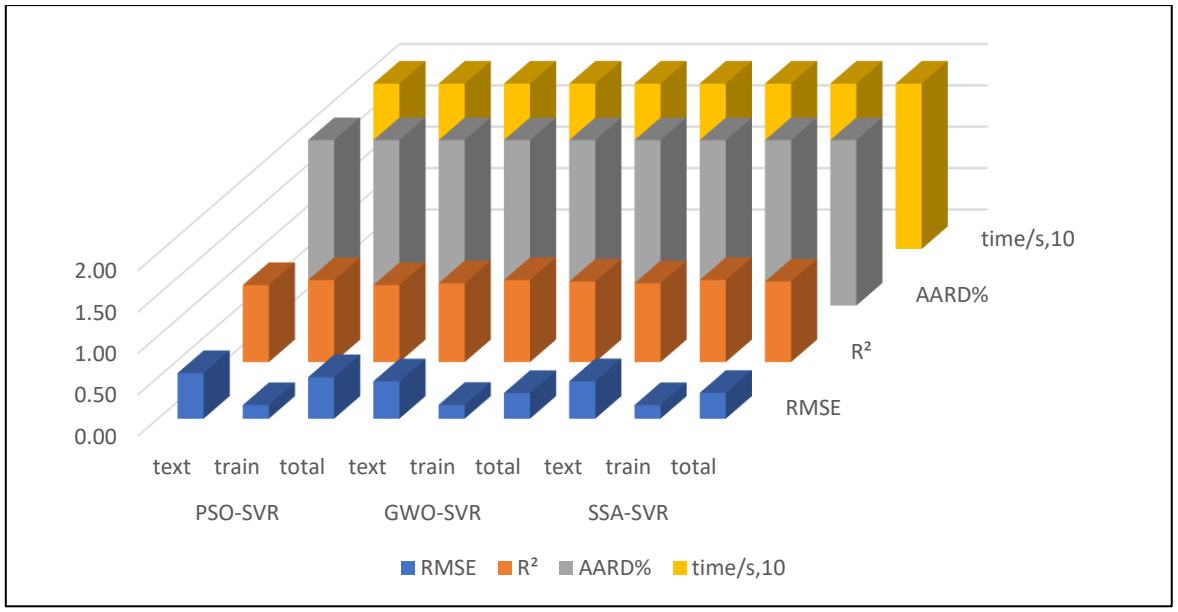

Fig.16. Statistical analysis and comparison of different ML models for P_L

Conclusion

The artificial intelligence models used in this research achieved remarkable accuracy in predicting Langmuir pressure and Langmuir volume compared to an existing model. Based on the findings of this investigation, we can conclude as follows:

1. Artificial intelligence models developed with PSO-SVR, GWO-SVR, and SSA-SVR, can predict the shale gas adsorption parameters (Langmuir pressure and Langmuir volume) with higher accuracy than SVR
2. After comparing and contrasting the three ML model prediction outcomes and prediction times, it was determined that PSO-SVR had the best average prediction results.
3. It is possible to construct models for specific basins to obtain more precise estimates. Unfortunately, data availability is a limiting factor here because good estimations from ML models need a massive amount of data.

References

1. Jarvie, D. M.; HIl.; R. J.; Ruble.; T.E.; Pollastro, R.M.;, "Unconventional shale-gas systems: The Mississippian Barnett Shale of north-central Texas as one model for thermogenic shale-gas assessment," *AAPG BULL,* vol. 91, pp. 475-499, 2007.
2. Curtis, M.E.; Cardott, G.J.; Sondergeld, C.H.; Rai, C.S., "Development of organic porosity in the woodford shale with increasing thermal maturity.," *Int. J. Coal. Geol.,* vol. 103, pp. 26-31, 2012.
3. Wang, D.; Shao, L.; Li, M.; Lv, D.; Liu, H., "Hydrocarbon generation characteristics, reserving performance and perservation conditions of continental coal measure shale gas: a case study of mid jurassic shale gas in the Y an'an formation, Ordos Basin," *J. Petrol.Sci. Eng,* vol. 145, pp. 609-628, 2016.
4. Wang, S.f.; Dong, D. Z.; Wang, Y. M.; Li, X. J.; Guan, Q. Z. , "Sedimentary geochemical proxies for paleoenvironment interpretation of organic-rich shale: A case study of the Lower Silurian Longmaxi Formation, Southern Sichuan Basin, China," *J. Nat. Gas Sci .Eng,* vol. 28, pp. 691-699, 2016b.

5. LV, D. M; Wang, D. D.; Li, Z. X.; Liu, H. Y.; Li, Y., "Depositional environmemt, sequence stratigraphy and sedimemtary minerationation mechanism in the coal bed and oil shale bearing succession: a case study from the paleogene Huangxian Basin of china," *J. Petrol. S CI. Eng.,* vol. 148, pp. 32-51, 2018.
6. Aguilera, R.F., "The role of natural gas in a low carbon Asia Pacific.," *Applied Energy,* vol. 113, pp. 1795-1800, 2014.
7. Stamford, L.; A. Azapagic., "Life cycle environmental impacts of UK shale gas," *Applied Energy,* vol. 134, pp. 506-518, 2014.
8. Arora, V.; Y. Cai., "U.S. natural gas exports and their global impacts," *Applied Energy,* vol. 120, pp. 95-103, 2014.
9. Howarth, R.; A. Ingraffea.; T, Engelder., "Natural gas: Should fracking stop? Nature," *Cornell University1, Pennsylvania State University,* vol. 477, pp. 271-275, 2011.
10. Kerr, R. A., "Natural gas from shale bursts onto the scene," *Science,* vol. 328(5986), pp. 1624-1626, 2010.
11. Yuan, J.; D. Luo.; L. Feng., "A review of the technical and economic evaluation techniques for shale gas development," *Applied Energy,* vol. 148, pp. 49-65, 2015.
12. Weijermars, R., "US shale gas production outlook based on well roll-out rate scenarios," *Applied Energy,* vol. 124, pp. 283-297, 2014.
13. Chen, Y.; Jiang, S.; Zhang, D.; Liu, C., "An adsorbed gas estimation model for shale gas reservoirs via statistical learning," *Applied Energy,* vol. 197, pp. 327-341, 2017.
14. Wang, Z.; A. Krupnick., "A Retrospective Review of Shale Gas Development in the United States: What Led to the Boom?," *Resources for the Future,* 2013.
15. Wade, "Natural Gas Annual 2015, Energy Information Administration (EIA)," 2016.
16. Zhang, H.; D, Cao., "Molecular simulation of displacement of shale gas by carbon dioxide at different geological depths," *Chemical Engineering Science,* vol. 156, pp. 121-127, 2016.
17. Kuuskraa, V.; S,H, Stevens.; K,D, Moodhe., "Technically recoverable shale oil and shale gas resources: an assessment of 137 shale formations in 41 countries outside the United States," *Department of Energy,* 2013.
18. Zhang, D., "China shale gas resource potential evaluation survey," *Beijing Geological Publishing House,* 2012.
19. Sander, R.; Pan, Z.; Connell, L.D.; Camilleri, M.; Grigore, M.; Yang, Y., "Controls on methane sorption capacity of Mesoproterozoic gas shales from the Beetaloo Sub-basin Australia and global shales," *International Journal of Coal Geology,* vol. 199, p. 65–90, 2018.
20. Langmuir, I., "The adsorption of gases on plane surfaces of glass, mica and platinum," *Journal of the American Chemical society,* vol. 40(9), p. 1361–1403, 1918.
21. Yu, W.; Sepehrnoori, K.; Patzek, T.W., "Modeling gas adsorption in Marcellus shale with Langmuir and bet isotherms," *SPE Journal,,* vol. 21(02), p. 589–600, 2016.
22. Gasparik, M., "High-pressure methane sorption isotherms of black shales from the Netherlands," *Energy & Fuels,* vol. 26(8), p. 4995–5004, 2012.
23. Dilmore, R., "Experimental characterization of Marcellus shale outcrop samples, and their interactions with carbon dioxide and methane," *Department of Energy, National Energy Technology Laboratory,* 2015.

24. Zhang, T.; Ellis, G.S.; Ruppel, S.C.; Milliken, K.; Yang, R., "Effect of organic-matter type and thermal maturity on methane adsorption in shale-gas systems," *Organic geochemistry,* vol. 47, p. 120–131, 2012.
25. Zhong, G.H.; Xie, B., Zhou, X.; Peng, X.; Tian, C., "A logging evaluation method for gas content of shale gas reservoirs in the Sichuan Basin," *Nat. Gas Ind,* vol. 36, p. 43–51, 2016.
26. Chen, Y.; Jiang, S.; Zhang, D.; Liu, C., "An adsorbed gas estimation model for shale gas reservoirs via statistical learning," *Applied energy,* vol. 197, p. 327–341, 2017.
27. Ji, W., Song.; Y., Jiang, Z.; Wang, X.; Bai, Y.; Xing, J., "Geological controls and estimation algorithms of lacustrine shale gas adsorption capacity: A case study of the Triassic strata in the southeastern Ordos Basin, China.," *International Journal of Coal Geology,* vol. 134–135, p. 61–73, 2014.
28. Tang, L., Song, Y., Jiang, Z., Pang, X., Li, Z., Li, Q., Li, W., Tang, X., & Pan, A, "Influencing Factors and Mathematical Prediction of Shale Adsorbed Gas Content in the Upper Triassic Yanchang Formation in the Ordos Basin, China," *Minerals,* vol. 9(5), p. 265, 2019.
29. Zou, J.; Rezaee, R., "A Prediction Model for Methane Adsorption Capacity in Shale Gas Reservoirs," *Energies,* vol. 12 (2), p. 280, 2019.
30. Aly, M.; Abdulraheem, A.; Moawad, T., " A Comparative Study Between Artificial Intelligence Models and a Statistical Learning Model to Predict Gas Adsorption in Shales," *SPE-212742-MS,* 2023.
31. Wang, L.; Liu, M.; Altazhanov, A.; Syzdykov, B.; Yan, J,; Meng, X.; Jin, K., "Data Driven Machine Learning Models for Shale Gas Adsorption Estimation," 2020.
32. Mahmoud, A. A.; Elkatatny, S.; Ali, A. Z.; Abouelresh, M.; Abdulraheem, A., "Evaluation of the Total Organic Carbon (TOC) Using Different Artificial Intelligence Techniques," *Sustainability,* vol. 11(20), p. 5643, 2019.
33. Awad, M.; Khanna, R., Chapter 4: Support Vector Regression, Apress ed., In Efficient Learning Machines, 2015, p. 67–80.
34. Irfan, A, S.; Azli, M N.; Abdulkareem, A, F.; Padmanabhan, E., "Multivariate Based Analysis of Methane Adsorption Correlated to Toc and Mineralogy Impact from Different Shale Fabrics," *unconventional resources technology coference,* 2021.
35. Yeh, C. Y.; Huang, C.W.; Lee, S.J.;, "A multiple-kernel support vector regression approach for stock market price forecasting," *Expert Systems with Applications,* vol. 38(3), p. 2177–2186, 2011.
36. F. Y. A.; B. Teslim Olayiwola A,; C. A. A., " Application of Artificial Intelligence-Based Predictive Methods in Ionic Liquid Studies A Review," *Fluid Phase Equilib,* 2021.
37. S.S.S.A. Mirjalili,; S.S.S.S. Sarem,; S.M.S.S Mirjalili,; L.DS.S. Coelho., "Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization," *Expart syst,* pp. 106-119, 2016.
38. X. B. X. S. Bian,; L.Z. L. S.Zhang,; Z.D.Z.S. Du,; J.C.J.S. Chen,; J.Z.J.S. Zhang, "Prediction of sulfur solubility in supercritical sour gases using grey wolf optimizer-based support vector machine," *J. MOL.Liq,* pp. 431-438, 2018.
39. Seyedali Mirjalili., Amir H. Gandomi., Seyedeh Zahra Mirjalili., Shahrzad Saremi.,Hossam Faris d, Seyed Mohammad Mirjalili, "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems," *Advances in Engineering Software,* pp. 163-191, 2017.
40. Jiankai Xue., Bo Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems Science & Control Engineering ,* vol. 8, no. 1, 2020.

41. Ahmad, M. W.; Reynolds, J.; Rezgui, Y., "Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees," *Journal of Cleaner Production,* vol. 203, p. 810–821, 2018.
42. Breiman, L., "Random forests," *Machine Learning,* vol. 45, pp. 5-32, 2001.
43. Tian, H., Tengfei Li., Tongwei Zhang., Xianming Xiao, "Characterization of methane adsorption on overmature Lower Silurian–Upper Ordovician shales in Sichuan Basin, southwest China: Experimental results and geological implications," *International Journal of Coal Geolog,* 2016.
44. Matus Gasparik., Pieter Bertier, Yves Gensterblum., Amin Ghanizadeh., Matus Gasparika, Pieter Bertier b, Yves Gensterblum a, Amin Ghanizadeha, Ralf Littke , "Geological controls on the methane storage capacity in organic-rich shales," *International Journal of Coal Geology,* vol. 123, pp. 34-51, 2014.
45. Wenming Ji.,Yan Song., Zhenxue Jiang., Lei Chen., Zhuo Li., Xiao Yang., Mianmo Meng, "Estimation of marine shale methane adsorption capacity based on experimental investigations of Lower Silurian Longmaxi formation in the Upper Yangtze Platform, south China," *Marine and Petroleum Geology,* vol. 68, pp. 94-106, 2015.
46. Jingqiang Tana., Philipp Wenigerb., Bernhard Kroossb., Alexej Merkelb., Brian Horsfield., Jinchuan Zhangc., Christopher J.. Borehamd., Ger van Graase., Bruce Alastair Toch.er, "Shale gas potential of the major marine shale formations in the Upper Yangtze Platform, South China, Part II: Methane sorption capacity," *Fuel,* vol. 129, pp. 204-218, 2014.
47. Peng Li., Zhenxue Jiang., Min Zheng., HeBi., Lei Chen, "Estimation of shale gas adsorption capacity of the Longmaxi Formation in the Upper Yangtze Platform, China," *Journal of Natural Gas Science and Engineering,* vol. 34, pp. 1034-1043, 2016.
48. Jie Zou., Reza Rezaee., Kouqi Liu, "Effect of Temperature on Methane Adsorption in Shale Gas Reservoirs.," *energy&fuel,* vol. 31, pp. 12081-12092, 2017.
49. "Experimental Characterization of Marcellus Shale Outcrop Samples, and their Interactions with Carbon Dioxide and Methane," *U.S department of enenrgy. National energy technology laboratory,* vol. 1, 2015.
50. Fangwen Chen., Shuangfang Lu., Xue Ding., Xipeng He., Huilin Xing, " Evaluation of adsorbed gas amount in a shale reservoir using the three compositions adsorbing methane (TCAM) method: A case from the Longmaxi shale in Southeast Chongqing, China," *Energy& Fuels,* 2017.
51. FangwenChen., ShuangfangLu., XueDing., Xipeng He, "Shale gas reservoir characterization: A typical case in the Southeast Chongqing of Sichuan Basin, China," *PLOS ONE,* 2018.
52. Wenrui Shi, Xingzhi Wang, Chaomo Zhang, Aiguo Feng, Zisang Huang, "Experimental study on gas content of adsorption and desorption in Fuling," *Journal of Petroleum Science and Engineering,* vol. 180, pp. 1069-1076, 2019.