

Article

A Powerful Prediction Framework of Fracture Parameters for Hydraulic Fracturing Incorporating eXtreme Gradient Boosting and Bayesian Optimization

Zhe Liu ^{1,2}, Qun Lei ^{1,2}, Dingwei Weng ^{1,2}, Lifeng Yang ^{1,2,*}, Xin Wang ^{1,2}, Zhen Wang ^{1,2}, Meng Fan ^{1,2} and Jiulong Wang ³

- ¹ CNPC Key Laboratory of Oil and Gas Reservoir Stimulation, Langfang 065007, China; liuzhe69@petrochina.com.cn (Z.L.); leiqun@petrochina.com.cn (Q.L.); wendw69@petrochina.com.cn (D.W.); wangx69@petrochina.com.cn (X.W.); wangzhen69@petrochina.com.cn (Z.W.); fanmeng2020@petrochina.com.cn (M.F.)
- ² Research Institute of Petroleum Exploration and Development, PetroChina, Beijing 100083, China
- ³ Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China; jlwang@cnic.cn
- * Correspondence: yanglf69@petrochina.com.cn

Abstract: In the last decade, low-quality unconventional oil and gas resources have become the primary source for domestic oil and gas storage and production, and hydraulic fracturing has become a crucial method for modifying unconventional reservoirs. This paper puts forward a framework for predicting hydraulic fracture parameters. It combines eXtreme Gradient Boosting and Bayesian optimization to explore data-driven machine learning techniques in fracture simulation models. Analyzing fracture propagation through mathematical models can be both time-consuming and costly under conventional conditions. In this study, we predicted the physical parameters and three-dimensional morphology of fractures across multiple time series. The physical parameters encompass fracture width, pressure, proppant concentration, and inflow capacity. Our results demonstrate that the fusion model applied can significantly improve fracture morphology prediction accuracy, exceeding 0.95, while simultaneously reducing computation time. This method enhances standard numerical calculation techniques used for predicting hydraulic fracturing while encouraging research on the extraction of unconventional oil and gas resources.

Keywords: hydraulic fracture; fracture parameters; machine learning; eXtreme Gradient Boosting model; unconventional reservoir



Citation: Liu, Z.; Lei, Q.; Weng, D.; Yang, L.; Wang, X.; Wang, Z.; Fan, M.; Wang, J. A Powerful Prediction Framework of Fracture Parameters for Hydraulic Fracturing Incorporating eXtreme Gradient Boosting and Bayesian Optimization. *Energies* **2023**, *16*, 7890. <https://doi.org/10.3390/en16237890>

Academic Editor: Jalel Azaiez

Received: 20 October 2023

Revised: 25 November 2023

Accepted: 26 November 2023

Published: 3 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hydraulic fracturing is the process of using a high-pressure pump on the surface to inject fracturing fluid with high viscosity into the formation through the well bore, causing the formation rock to fracture and expand to form artificial fractures. This technology has been used extensively over the last few decades and has become one of the key technologies in the global oil and gas industry [1]. The study of the geometry, length, width and depth characteristics of fractures is essential in the hydraulic fracturing process, because this research is important for assessing the feasibility of hydraulic fracturing techniques, optimizing the fracturing parameters, and predicting the response of the subsurface production.

Hydraulic fracturing research, which began in the last century and continues to evolve today, can be divided into the following categories.

1. **In situ observation methods:** This method focuses on observing and recording at the fracturing site to study the phenomena that occur during hydraulic fracturing. Representative field observation methods include seismic exploration, surface deformation monitoring, and groundwater level monitoring. American petroleum engineer John F.

Schlumberger invented electronic logging, which can characterize rock fractures by measuring the speed and path of sound waves propagating through rock, in the 1930s; Hiroaki Niitsuma [2] used the seismic exploration method to determine the fracture and stress of underground rocks; and Zou [3] obtained three-dimensional images and internal fracture information of underground rocks using computed tomography.

2. Laboratory experimental methods: This method mainly uses laboratory experiments to study the physical and mechanical properties during hydraulic fracturing. J. Groenenboom [4] used the velocity and path of ultrasonic wave propagation in rock to determine the fracture characteristics of rock; Liu [5] used a triaxial hydraulic fracturing experimental setup to study the effects of prefabricated fracture parameters, fracturing fluid viscosity and rock physical properties on hydraulic fracturing.
3. Mathematical modeling method: This method mainly describes the physical and mechanical phenomena in the hydraulic fracturing process by establishing mathematical models. Axel KL Ng [6] proposed a numerical procedure based on the finite element method to predict hydraulic fractures in the core wall of earth and rock dams; Torres Sergio Andres Galindo [7] proposed a discrete element simulation method of hydraulic fracture process in oil wells considering the elastic properties of rocks and the Mohr–Coulomb fracture criterion; and Jeroen Groenenboom [8] proposed a finite difference model to characterize fractures using seismic waves generated by an active seismic source.

However, all of the above methods have limitations. In situ observation methods generally require more equipment and are more geologically demanding. With laboratory methods, it is difficult to control for the effects of environmental factors and they are usually used only to study short-term results, making it difficult to predict long-term results. Mathematical modeling methods require longer and more costly data collection, model construction and solution processes. They also require a series of assumptions about the actual problem that may not match the actual situation, leading to inaccuracy and reduced reliability of the model. These are all issues that need to be addressed in current hydraulic fracturing projects [9–12].

In recent years, with the continuous development of new technologies, machine learning techniques have made significant breakthroughs in the petroleum field [13–16]. Yang [17] established a coal-bed methane fractured well capacity prediction model based on a combined approach of recurrent neural networks and multilayer perceptron. Shen [18] established a fractured well capacity prediction model based on convolutional neural networks and U-Net architecture for real-time diagnosis of fracturing conditions. Saeed Salehinia [19] presented a method based on nonlinear system identification modeling for predicting the PVT properties of crude oil systems. Pejman Tahmasebi [20] presented a fast independent artificial neural network architecture for permeability prediction. Pejman Tahmasebi [21] developed a neural network workflow to provide a systematic approach to solving various petroleum engineering problems. F. Torabi [22] developed an artificial neural network model for predicting and calculating oil viscosity by applying a large amount of experimental data on dead, saturated and undersaturated oil viscosities of different samples from Iranian reservoirs. Aref Hashemi Fath [23] utilized multilayer perceptron and a radial basis function to predict oil viscosity and radial basis function neural networks to predict the dissolved gas-oil ratio of crude oil systems. Mohammadzaheri Morteza [24] proposed a new shallow artificial neural network to model the multiphase flow of submerged oil electric pumps.

We have improved the model to address the drawbacks of the traditional hydraulic fracturing process. We use a data-driven approach to speed up the computational process of traditional mathematical modeling and input environmental factors under different operating conditions to improve the generalizability of the model prediction.

In this study, we identified three model evaluation parameters and introduced four machine learning model algorithms to model the numerical simulation process based on the sampled dataset. By fitting the dataset and comparing the prediction accuracy and speed,

we selected the eXtreme gradient boosting model as the model for numerical simulation of artificial fractures. Compared with traditional numerical simulation methods, the eXtreme gradient boosting model does not need to model the physical process or solve complex nonlinear differential equations—it can be directly trained and predicted by using existing data. Therefore, this model avoids complex numerical calculations and achieves improved prediction accuracy and adaptive optimization of the model by continuously and iteratively modifying the hyperparameters using a Bayesian optimization algorithm.

The overall structure of the study takes the form of five chapters, including this Section 1. Section 2 introduces the main methods and models used in the research of this study. Section 3 presents the application of the model to practical engineering applications. Section 4 details the results and discusses them. Section 5 presents the conclusions of this study.

2. Research Methodology and Model Construction

2.1. Numerical Simulation

Because the large error and the small amount of statistical data in real production cannot match the data quality requirements of machine learning, in this study we used numerical simulation to generate datasets. Specifically, we used the software to simulate and solve the fluid–solid coupled fracture propagation and proppant transport problem using three-dimensional displacement discontinuity [25] and finite volume methods [26]. This method allows us to obtain parameters such as fracture height, fracture length, conductivity, etc. under different conditions, providing a reliable database for subsequent studies. We considered a variety of different physical models and boundary conditions to accurately simulate actual fracture morphological expansion during the numerical simulations [27–31]. At the same time, we performed several iterations of the experiments to ensure the reliability and stability of the data. Finally, we obtained a statistically significant dataset that can be used to support our study.

2.2. eXtreme Gradient Boosting Model

The eXtreme gradient boosting (XGBoost) algorithm implements parallel computing, efficient processing of sparse data, and optimization of memory usage to achieve a balance between model predictive effectiveness and computational speed.

After the model is built, it is necessary to evaluate how well the model predicts the unknown data, so we specify the loss function to be used as an indicator of the predictive ability of the model. Space complexity and time complexity are two important metrics for evaluating the computational efficiency of different models and algorithms. XGBoost achieves this performance through the introduction of a model complexity metric for the objective function, which is a traditional loss function plus model complexity:

$$\text{Obj} = \sum_{i=1}^m l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

where i represents the i th sample in the dataset, m denotes the number of samples in the dataset into which the k th tree is imported, and the total number of trees generated is denoted by K , y_i and $\hat{y}_i$, which are the outputs of the model and the actual outputs, $l()$ is the loss function, f_k is the specific structure of the k th tree, and $\Omega()$ is its regularization factor. The first part is similar to the loss function of the traditional model and is used to compare the difference between actual and predicted values. The second part of the objective function measures the complexity of the tree model through the structure of the tree, which contains a variety of forms. The root mean square error of the traditional loss function is commonly used to measure the prediction variance of the XGBoost model. Since the second part is essentially independent of the feature matrix, it is minimized by finding the minimum value to obtain the optimum at each iteration. Minimizing the objective function means minimizing both the loss function and the model complexity,

thus achieving a reduction in the model prediction error rate while reducing the model complexity.

2.3. Hyperparameter Optimization

Hyperparameter optimization is required when using machine learning in engineering practice. Hyperparameters are parameters set in machine learning algorithms that are mainly used to control the performance of the model in terms of complexity, training speed, and generalization ability. Different combinations of hyperparameters are one of the key factors affecting the accuracy of XGBoost models. In the model, the commonly used hyperparameters include learning rate, tree depth, number of trees, column sampling rate, and so on [32–34].

If the learning rate is too high, the weights of the base learners will shrink too fast, resulting in the model not being able to reach the optimal state; if the learning rate is too low, the training time of the model will be longer because the base learners will need more numbers to reach the same accuracy. In addition, the depth of the tree and the number of trees are important hyperparameters that affect the accuracy of the model. Deeper trees fit the data better, but tend to overfit; shallower trees tend to underfit. In addition, the column sampling rate is the proportion of feature samples used for segmentation on each node during the construction of each tree. Too small a column sampling rate will result in an underfitting of the model, while too large a rate will result in an overfitting of the model.

Therefore, the accuracy of the XGBoost model can be effectively improved by choosing to adjust the hyperparameter combinations. Common methods for adjusting hyperparameters include grid search, stochastic search and Bayesian optimization. In order to find the best combination of hyperparameters quickly and automatically, this study uses Bayesian optimization to automatically adjust the parameters and realize adaptive model optimization.

Bayesian optimization [35,36] is a global algorithm for optimizing black-box functions to find the minimum or maximum value without knowing the exact form. Bayesian optimization makes use of Bayesian inference to find the optimum, approximating the objective function by building probabilistic models on experimental data. The probabilistic model of Bayesian optimization can be expressed as:

$$P(z|D) = \frac{P(D|z)P(z)}{P(D)} \quad (2)$$

where z is the unknown objective function, D is the set of known observations $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ while $y_n = z(x_n) + \varepsilon_n$, and ε is the observation error. $P(D|z)$ is the likelihood distribution of y , and $P(z)$ is the prior probability distribution of z , which represents the assumptions about the state of the unknown function. $P(z|D)$ represents the posterior probability distribution of z , which describes the confidence level of the unknown objective function after correcting the prior probabilities according to the set of observations.

The main process of using Bayesian optimization techniques to implement the XGBoost model for hyperparameter optimization is as follows.

Prepare the dataset and divide it into training and test sets. Specify the range and type of hyperparameter values and determine the optimization objective. Run the Bayesian optimization algorithm to search for the optimal combination of hyperparameters and keep updating the search results until a predetermined number of iterations or time is reached. Finally, the performance of the optimal hyperparameter combinations is tested on the test set and the prediction accuracy of the model is evaluated.

3. Practical Use of Model

3.1. Background of the Project

The reservoir target area is located in the transition area between the south Sichuan low steep tectonic belt and the southwest Sichuan low fold tectonic belt, and the structural

characteristics are relatively gentle. The longitudinal comprehensive interpretation and evaluation of the reservoirs are four sub-formations of the Longmaxi Formation and Wufeng Formation, with an overall thickness of 69.3 m. The curves of organic carbon, gas content and brittleness index gradually increase from top to bottom, and the reservoirs gradually become better from top to bottom. The continuous thickness of Class I reservoirs is 10 m. The reservoir fracture pressure ranges from 104.4 to 139.5 MPa, with an average of 119.6 MPa. The minimum ground stress is 90.0–106.7 MPa and the average is 97.3 MPa. The maximum ground stress is 104.6–123.0 MPa, with an average of 111.8 MPa. Vertical ground stress is 102.7–104.5 MPa, with an average of 103.6 MPa. Average maximum and minimum principal stress difference is 14.5 MPa.

3.2. Methods of Construction and Use of the Dataset

In this study, a numerical simulation program was used to generate fracture extensions. The numerical simulation program is based on the three-dimensional displacement discontinuity method to solve the fluid–solid coupling problem during fracture growth, which takes into account the effects of multiple factors, such as interlayer stress inhomogeneity, matrix filtration, interstitial stress interference, multiple types of fluid systems, and transport of proppant with multiple particle sizes. It is used to describe the fracture deformation using the three-dimensional displacement discontinuity method (3D-DDM), which defines the amount of displacement discontinuity as:

$$\begin{cases} D_1(x_1, x_2, 0) = u_1(x_1, x_2, 0^-) - u_1(x_1, x_2, 0^+) \\ D_2(x_1, x_2, 0) = u_2(x_1, x_2, 0^-) - u_2(x_1, x_2, 0^+) \\ D_3(x_1, x_2, 0) = u_3(x_1, x_2, 0^-) - u_3(x_1, x_2, 0^+) \end{cases} \quad (3)$$

where u is the displacement of the object in different states, D_1 is the shear displacement discontinuity, D_2 is the normal displacement discontinuity, and the negative D_3 indicates fracture opening. With this formulation, the fracture is discretized into a series of structured rectangular cells, and the main variables such as pressure are located at the central nodes of the cells. For a single fracture cell, which is subjected to constant shear and normal displacement discontinuities, the analytical solution of the stress field is:

$$\begin{cases} \sigma_{sl}^i = \sum_{j=1}^N K_{sl,sl}^{ij} D_{sl}^j + \sum_{j=1}^N K_{sl,sh}^{ij} D_{sh}^j + \sum_{j=1}^N K_{sl,nn}^{ij} D_{nn}^j \\ \sigma_{sh}^i = \sum_{j=1}^N K_{sh,sl}^{ij} D_{sl}^j + \sum_{j=1}^N K_{sh,sh}^{ij} D_{sh}^j + \sum_{k=1}^N K_{sh,nn}^{ij} D_{nn}^j \\ \sigma_{nn}^i = \sum_{j=1}^N K_{nn,sl}^{ij} D_{sl}^j + \sum_{j=1}^N K_{nn,sh}^{ij} D_{sh}^j + \sum_{j=1}^N K_{nn,nn}^{ij} D_{nn}^j \end{cases} \quad (4)$$

where K is the boundary influence coefficient and N is the total number of cells. The equations controlling the flow of liquids and sand-carrying fluids within the fracture are:

$$\frac{\partial w}{\partial t} = \nabla \cdot q + q_i - q_L \quad (5)$$

$$q = -K(w) \tilde{Q}(c, w) [\nabla p - \rho g] \quad (6)$$

$$q_i = \delta(x - x_0, y - y_0) Q(x, y) \quad (7)$$

$$q_L = \frac{2C_L}{\sqrt{t - \tau(x, y)}} \quad (8)$$

$$K(w) = \frac{w^3}{12\mu^f} \quad (9)$$

$$\rho = (1 - \phi) \rho^f + \phi \rho^p = (1 - c\phi_{\max}) \rho^f + c\phi_{\max} \rho^p \quad (10)$$

$$\tilde{Q}(c, w) = (1 - c)^\beta + \frac{a^2}{w^2} c \tilde{D} \quad (11)$$

where w is the flow displacement of the liquid and the sand-carrying liquid in the fracture, q is the flux of sand-carrying liquid, q_i is the injected volume per unit time per unit area in m/s, μ^f is the viscosity of the fracturing fluid, ρ is the density of the sand-carrying liquid, $\tilde{Q}(c, w)$ is used to describe the flow of sand-carrying liquids, $\beta = 1.5$, $\tilde{D} = 8(1 - \varphi_{\max})^\alpha / 3\varphi_{\max}$, $c = \varphi / \varphi_{\max}$, φ is the proppant concentration, and $\varphi_{\max} = 0.585$, $\alpha = 4.1$, a is the radius of the proppant particles.

The equation controlling the transport of proppant in the fractures is given by:

$$\frac{\partial cw}{\partial t} = \nabla \cdot q^p + q_i^p \quad (12)$$

where q^p is the volume of proppant injected per unit time per unit area in m/s.

Input data were modified in the program to simulate injection under different conditions, and the different associated fracture extension data were obtained by running the program. The fracture extension data include fracture area, fracture height, fracture length, maximum fracture width, displacement of each fracture, friction of each fracture, and fracture volume. The program outputs the results to different PFC files for storage.

Specifically, this can be accomplished by varying the displacement, time, proppant volume concentration, and fracturing fluid type. The specific parameters can be varied in the following ranges: pumping time from 1 to 30 min, discharge volume from 2 to 12, proppant concentration from 0 to 0.1 and 0 in the first row, and fluid volume/total fluid volume greater than 30% at 0 concentration.

In conclusion, this study mainly simulates different fracturing operations by controlling the number of perforation clusters, perforation cluster coordinates, pump displacement, pumping time, proppant volume concentration, and type of fracturing fluids, simulates different geomechanical properties under different blocks by controlling the Young's modulus, Poisson's ratio, and the maximum and minimum principal stresses of the formation, and simulates the fracturing of horizontal wells and vertical wells by varying the parameters of the well trajectory coordinates. A series of datasets were generated by randomly combining the values of different parameters according to normal distribution and sampling them with a confidence interval of 0.9 as input parameters. Through this method, 2000 sets of data were generated to form datasets, and the datasets were categorized into training and test sets for further use.

3.3. Research Technology Route

The main flow of this study is shown in Figure 1. The overall technical route shown in the figure includes three parts: data input, machine learning and data output. Firstly, the fracturing data is prepared and input into the numerical simulation program to generate a dataset. Next, the dataset is trained using the selected machine learning model, and the model is continuously modified or hyper-parameterized until the prediction accuracy meets expectations. Finally, the trained model is output, and the model can optimize and accelerate the original numerical simulation software.

The input data processing part includes data cleaning, data conversion, data integration and data dimensionality reduction, the machine learning part mainly includes model evaluation, Bayesian optimization and model adaptive optimization, and the data output part includes model generation and packaged output.

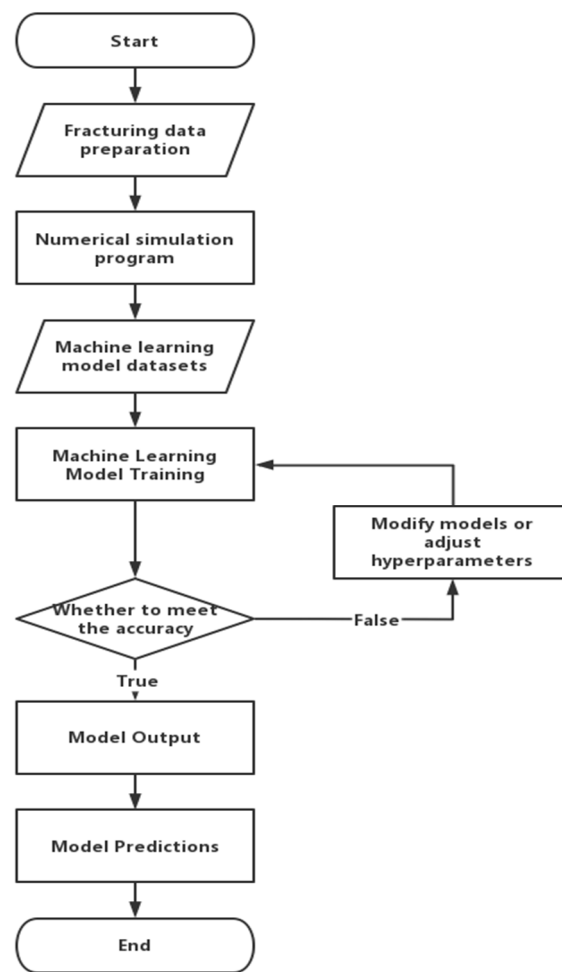


Figure 1. The overall research technology flow of the project.

3.4. Model Evaluation Metrics

This study used the coefficient of determination R^2 to assess the accuracy of the model. The coefficient of determination R^2 is used in statistics to determine the explanatory power of a statistical model and to measure the proportion of the variation in the dependent variable that can be explained by the independent variable [37,38].

For simple linear regression, the coefficient of determination is the square of the sample correlation coefficient. When other regression independent variables are added, the coefficient of determination changes to the square of the multiple correlation coefficient.

Assume that a dataset includes a total of n observations $y_1, y_2, \dots, y_n$, with corresponding model predictions $f_1, f_2, \dots, f_n$. Defining the residuals, the average observation is:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (13)$$

The total sum of squares can then be obtained as:

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (14)$$

The regression sum of squares:

$$SS_{reg} = \sum_i (f_i - \bar{y})^2 \quad (15)$$

The residual sum of squares:

$$SS_{res} = \sum_i (y_i - f_i)^2 = \sum_i e_i^2 \quad (16)$$

From this, the coefficient of determination can be defined as:

$$R^2 \equiv 1 - \frac{SS_{res}}{SS_{tot}} \quad (17)$$

The closer this statistic is to 1, the better the fit of the model.

In this study, we explicitly state that when the R^2 is greater than 0.9, we consider the model to be as accurate as required.

4. Results and Discussion

4.1. Model Preference

In order to select the model with the highest prediction accuracy for further learning and prediction, four different models are selected for pre-validation in this study. Pre-validation [39,40] is the process of comparing multiple models before selecting one. By comparing the accuracy and generalization ability of the predictions for the same dataset, the best model is selected and the risk of overfitting is avoided. To ensure that the error between learning and prediction is reduced, when using the `train_test_split` [41] function of the Scikit-learn library in Python, the same hardware equipment is used, the dataset is processed accordingly, and the same random numbers are chosen to ensure that the same training and test sets are used for different models.

A comparison of the prediction accuracy of the different models is shown in Figure 2.

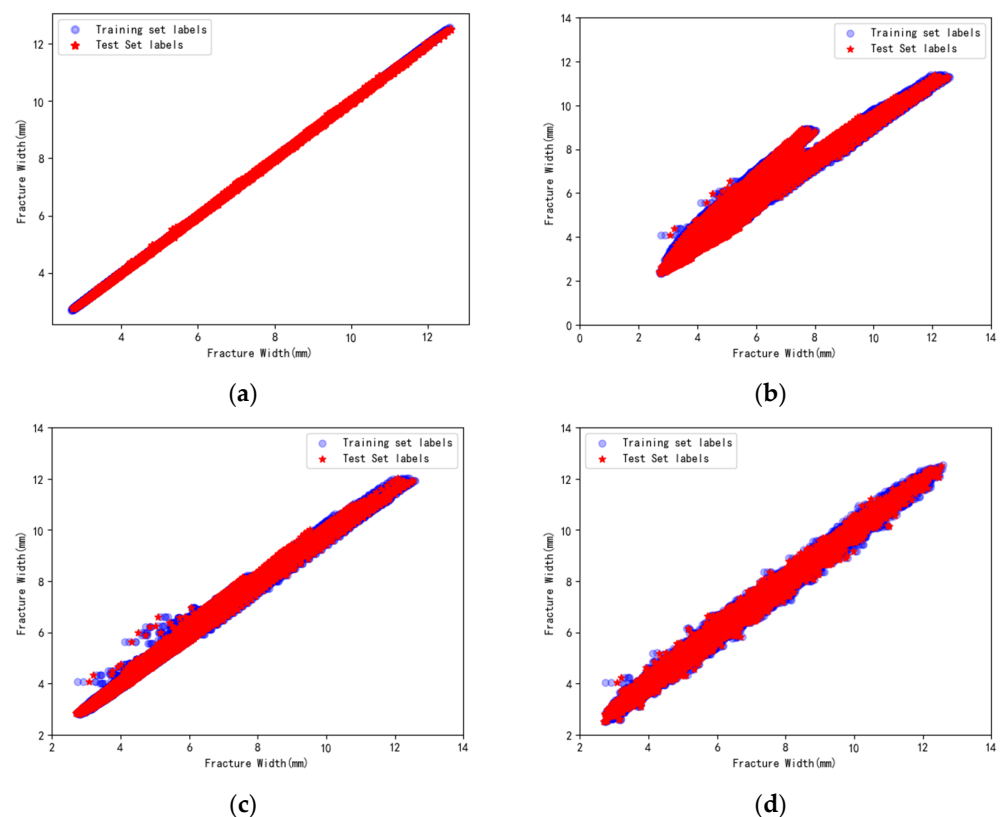


Figure 2. Comparison of prediction accuracy of different models ((a) eXtreme gradient boosting model; (b) Bayesian ridge regression model; (c) support vector machine regression model; (d) multi-layer perceptron regression model).

Figure 2 and Table 1 show the prediction performance of the models built by eXtreme gradient boosting, Bayesian ridge regression, support vector machine regression, and multilayer perceptron regression models. The most accurate XGBoost model has a coefficient of determination of 0.9985, which is more accurate than the other models. Therefore, the XGBoost model was chosen for subsequent training and prediction in this study.

Table 1. Comparison of prediction accuracy—R2 scores of different models.

Model	R2 Score
eXtreme gradient boosting model	0.9985
Bayesian ridge regression model	0.8645
Support vector machine regression model	0.9395
Multilayer perceptron regression model	0.9736

4.2. Bayesian Optimization

The hyperparameter results obtained by Bayesian optimization during the model training fitting phase are shown in Table 2.

Table 2. Hyperparameter results of XGBoost model based on Bayesian optimization in model prediction stage in model training stage.

Hyperparameters	Range	Optimal Parameter Results
Number of decision trees	1–300	78
Maximum depth of the tree	1–25	3
Minimum sample size for splitting	2–10	2
Bootstrap sampling	True/False	True
Number of Bayesian iterations	/	40
k-fold cross-validation	/	5

The test dataset was validated using the XGBoost model before and after Bayesian optimization. The data accuracy comparison results before and after optimization are obtained as shown in Table 3.

Table 3. Comparison of model prediction accuracy before and after optimization.

Model	R2 Score
Pre-optimization	0.9533
Post-optimization	0.9907

The comparison shows that the predicted R2 of the model after Bayesian optimization increases from 0.9533 to 0.9907, which proves the effectiveness of Bayesian optimization.

4.3. Analysis of the Results of the Time-Series Fit to the Two-Dimensional Data

In fracture analysis, the study of fracture changes from initial to steady state can provide critical information. This information reflects the mechanical behavior of the rock and the evolution of the fracture extension, orientation and morphology, as well as the various changes in the fracture process. Based on this information, we can assess the nature of stress distribution in the rock, the size and distribution of the fracture, and the stability of the rock.

In order to study the characteristics of different fractures over time, we obtained different parameters for each fracture from the database and normalized them. The parameters included area, length, maximum fracture width, displacement per fracture, friction per fracture and fracture volume. To further analyze the evolution of the fracture, we sampled four different groups of fractures at equal intervals. Each fracture was sampled at a time step, from which 30 samples were taken to plot the predicted results. This time-step sampling method avoids problems such as overfitting and underfitting.

4.3.1. Analysis of Fracture Length Results for Two-Dimensional Data

This study selects the fracture length parameter for sampling prediction. The variation in fracture length directly reflects the fracture expansion, and the measurement of fracture length is simple and easy for controlling the accuracy. Therefore, the extension pattern of fracture length during hydraulic fracturing is an important parameter to study in hydraulic fracturing [42,43].

This study extracts data from the fracture length file for different fractures according to the time step and divides the training set and test set for learning and prediction. The results obtained are plotted in the form of a line graph in the same chart, where the horizontal coordinate is the time and the vertical coordinate is the fracture length. The results are shown in Figure 3. At the same time, we calculated each fracture length prediction accuracy error, and the results are shown in Table 4.

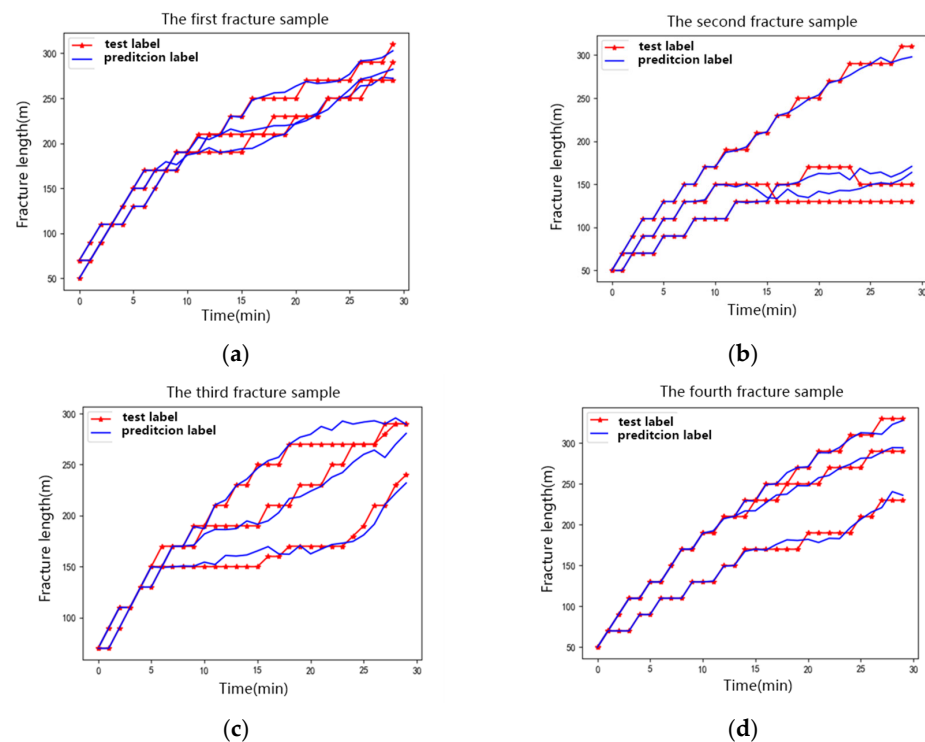


Figure 3. Different fracture length sampling predictions ((a) the first fracture; (b) the second fracture; (c) the third fracture; (d) the fourth fracture).

Table 4. Fracture length prediction accuracy error.

	R2 Score	MAE	MSE
The first fracture	0.9667	0.0125	0.000376
The second fracture	0.9726	0.0131	0.000414
The third fracture	0.9614	0.0138	0.000463
The fourth fracture	0.9692	0.0129	0.000406

From the results, it can be seen that the expansion process of fracture length can be divided into three main expansion stages: in the initial hydraulic fracturing pre-proppant pumping time step, the expansion rate of fracture length grows linearly with time. With the increasing pumping time, the pumping program enters the sand addition stage, the expansion rate of fracture length gradually decreases, the slope of the line graph of fracture length change gradually decreases, and the shape of the curve starts to change. In some samples, some fractures will eventually reach a stable state so as not to continue to increase the length, realizing the process of fracture making to proppant filling. In the initial pre-fluid stage of fracture length extension, because the length extension of the fracture is

affected by fewer factors, its influence features are fewer, so the model learns and predicts it more accurately and can fit the test set label better. As the fracturing process continues, because the fracture length extension is affected by more and more factors and its related feature values are increasing, the complexity of the model is increased, which led to a gradual decrease in the fitting accuracy.

However, by comparing the overall fitting degree (Table 4), it can be calculated that the average R^2 is 0.967475, the average MAE is 0.013075, and the average MSE is 0.00041475, where the average R^2 exceeds the criterion of 0.9 specified in the article, and the average MAE and MSE results are less than two percent and five ten thousandths of a percent, respectively, which proves that the model has a high prediction accuracy. The model is proved to have high prediction accuracy and meets the expected goal specified in the article. The model can still predict the expansion of fracture length in the hydraulic fracturing process relatively well and thus can be used in engineering practice.

4.3.2. Analysis of Fracture Area Results for Two-Dimensional Data

In this study, fracture area was chosen as the target for sampling and prediction because it is one of the most important indicators for assessing the size and distribution of rock fractures and rock stability. By predicting the fracture area, the changing pattern of fracture in the time series can be understood and can be used to assess the stress distribution, fracture extension rate and direction, and other properties of the formation during hydraulic fracturing [44,45].

We plotted the test dataset and the predicted dataset on the same graph for comparison. By comparing the fit of the two folded curves, a simple preliminary judgment of the accuracy of the model predictions can be made. In general, the higher the overlap of the two curves, the better the fit of the model predictions, thus indicating the higher accuracy of the model predictions. The horizontal coordinate represents the time and the vertical coordinate represents the fracture area. The plotting results are shown in Figure 4. What is more, we calculated each fracture area prediction accuracy error and the results are shown in Table 5.

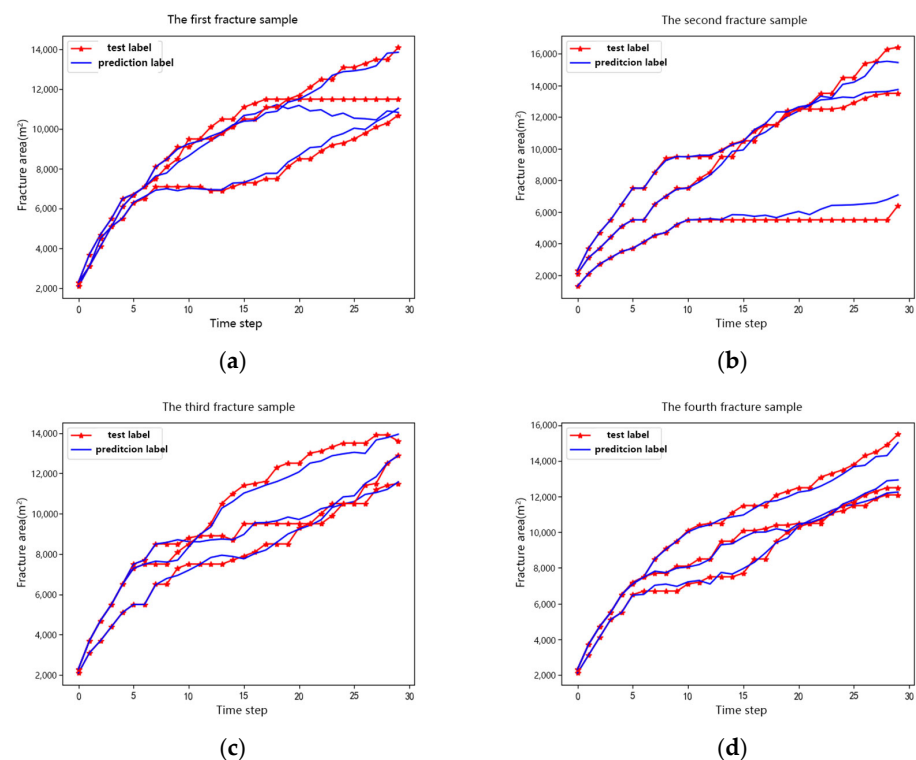


Figure 4. Different fracture area sampling predictions ((a) the first fracture; (b) the second fracture; (c) the third fracture; (d) the fourth fracture).

Table 5. Fracture area prediction accuracy error.

	R2 Score	MAE	MSE
The first fracture	0.9816	0.0091	0.000225
The second fracture	0.9792	0.0116	0.000377
The third fracture	0.9793	0.0111	0.000311
The fourth fracture	0.9820	0.0093	0.000209

As can be seen from Figure 4, the fracture area increased throughout the process as the fracturing fluid was pumped in. In the first 5 min time step, the fracture expansion rate is faster. As the fracturing fluid continues to be pumped in, the fracture area continues to increase, but the rate of fracture area increase gradually decreases due to the influence of formation stress. In some samples, the fracture area eventually stabilized as the time step increased, reaching equilibrium under the combined influence of various factors. The model can predict the change rule of the area of each fracture under different samples with high accuracy under the initial time step and the curves can basically overlap, but the prediction accuracy decreases with the increase in the time step. It can be concluded from this study that under the initial injection fracturing state, the increase in fracture area is mainly related to the pressure of the injected fluid and the nature of the formation, which is basically linear. With the continued injection of fracturing fluid, the change in the pressure state in the fracture leads to the change in the expansion rate of the fracture, and at the same time, the expansion rate is also to a certain extent affected by the expansion of the surrounding fractures, which results in the change in the expansion rate of the fracture change.

The overall fitting accuracy can be obtained through calculation, and from Table 5, the average R2 is 0.967475, the average MAE is 0.013075, and the average MSE is 0.00041475, in which the average R2 exceeds the standard of 0.9 stipulated in the article, and the average MAE and MSE results are less than two percent and five ten thousandths of a percent respectively. This proves that the model has a high prediction accuracy and achieves the expected goal of the article. The above rule of change is in line with common sense in engineering practice and has high credibility.

4.4. Feature Importance Analysis

Feature importance refers to the extent to which different features contribute to the performance of a model in a machine learning model. In practice, analyzing feature importance helps to gain insight into how the model works and helps to select features to improve the predictive performance of the model.

There are a variety of methods for analyzing feature importance, of which the common ones include [46–50] the following.

1. Tree model-based feature importance analysis methods: For example, in tree model-based algorithms such as random forest and GBDT, feature importance metrics (information gain, Gini index, etc.) can be used to assess the degree of impact of each feature on the performance of the model.
2. Linear model-based feature importance analysis methods: In linear regression, logistic regression and other linear model-based algorithms, feature coefficients or normalization coefficients can be used to assess the degree of contribution of each feature to the model.
3. Feature importance analysis method based on model adjustment: The impact of each feature on the performance of the model can be assessed by gradually removing certain features and then comparing the change in model performance. After obtaining the feature importance assessment results, feature selection or adjustment can be carried out based on the assessment results, and the improvement in model performance and efficiency can be achieved by retaining features with high importance, eliminating redundant features, or merging features.

Since the XGBoost model used in this study is based on the fundamentals of tree model construction, we used an integrated learning computational approach based on decision trees to calculate feature importance. In this study, we extracted the data of the 30 min time step and showed the method of feature importance by plotting a histogram. The plotted results are shown in Figure 5.

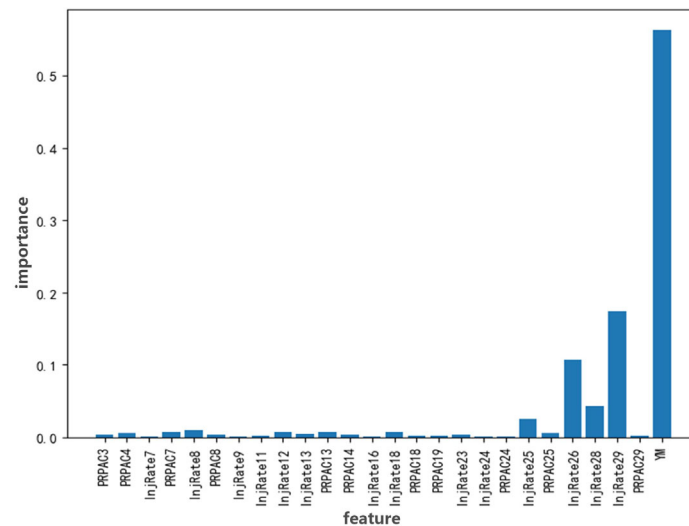


Figure 5. Comparison of the importance of different features at the 30 min time step.

From the results, it can be seen that the most important feature in hydraulic fracturing is the Young's modulus of the formation properties. The Young's modulus can be used to analyze the deformation extension, stability and bearing capacity of fracturing.

After considering the formation properties, the second important feature affecting the fracturing pattern in 30 min is the injection rate of hydraulic fracturing fluid from 25 to 29 min, and the closer the injection rate is to the 30 min time step, the more obvious the influence on the final fracturing pattern is. It can be found that the effect of fracking fluid on fracture extension during hydraulic fracturing is not effective in time, but has a certain lag and cumulative nature. With the increase in time, the effect of the fracking fluid injected in the previous time step will be gradually weakened, while the effect of the current time step will not be effective immediately, but the effect on the fracture extension will last for several time steps. This is an important reason that real-time fracture detection and control of fracturing fluid injection are needed in engineering practice.

4.5. Analysis of the Results of the 3D Data Time-Series Fitting

In order to study the changes in fracture morphology during hydraulic fracturing, this study visualizes the data predicted by the model and generates directly observable 3D images to facilitate the study of the whole process.

The main principles of 3D visualization are as follows. In this study, fracture data are generated by numerical simulation software and specific fracture grids are generated at different time steps, with an ever-increasing number of grids to simulate the continuous expansion of the fracture. In this study, a coordinate cell containing four vertices and a center point can be generated each time, it is shown in the Figure 6, take the prediction result of fracture width as an example, the software defines a quadrilateral grid based on the coordinates of these five nodes in the x, y, and z directions, for a total of 15 pieces of data. In addition, annotated data within the grid will be generated, including data on fracture widths, inflow capacity, and fracture pressure. When plotting the 3D visualization, a complete fracture visualization is generated by plotting the five nodes of each cell and the labeled data within the grid.

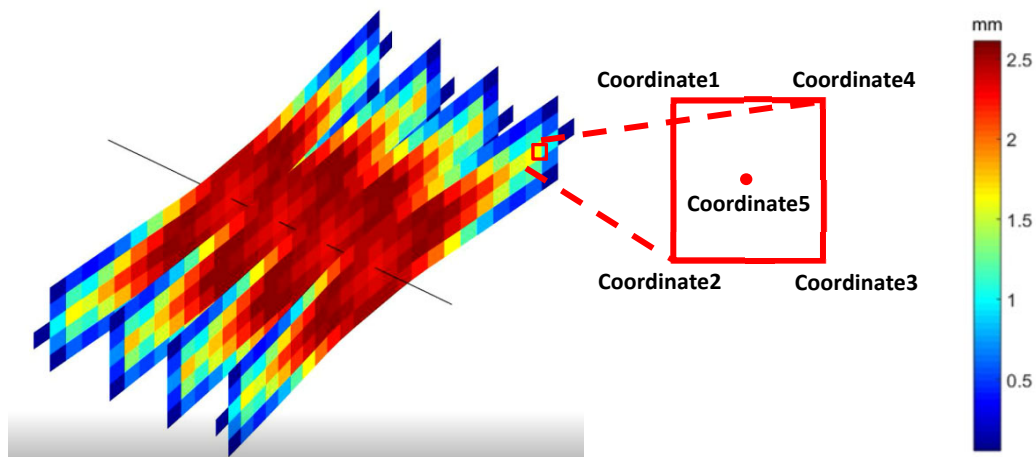


Figure 6. 3D visualization diagram of fracture morphology.

Prediction results often need to be recombined and reordered to achieve a more intuitive and readable output. In this process, MATLAB is a very powerful tool that can help visualize the prediction results. In this study, the 3D plotting function in MATLAB can be used to visualize the fracture morphology under different conditions so as to observe more intuitively the change in fracture morphology under different conditions. At the same time, various functions in MATLAB can be used to further analyze and process the results so as to get more comprehensive and accurate information. This not only improves the readability and visualization of the prediction results but also provides a better understanding of the prediction ability of the model, which can provide reference and support for subsequent model improvement and application.

In this study, the fracture width data at the 10 min, 20 min and 30 min time points were selected as the basis for the visualization operation, and the output results are shown in Figures 7–9.

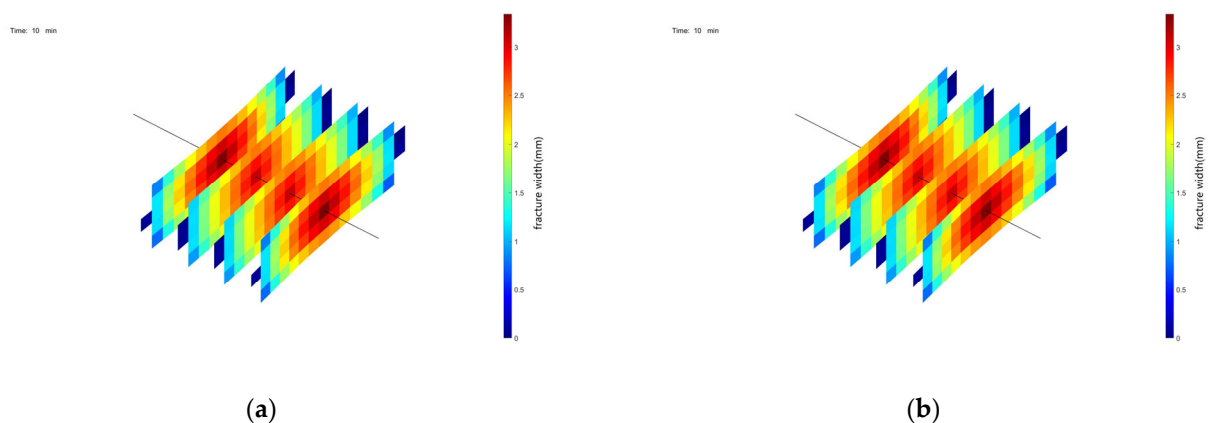


Figure 7. 3D visualization of fracture width at 10 min time step ((a) numerical simulation method; (b) model fitting methods).

In Figures 7–9, the left side is the 3D result plotted from the data generated by the numerical simulation method, while the right side is by the machine learning model. The average prediction accuracy of all grids at different time steps is obtained by calculating the results, as shown in Table 6.

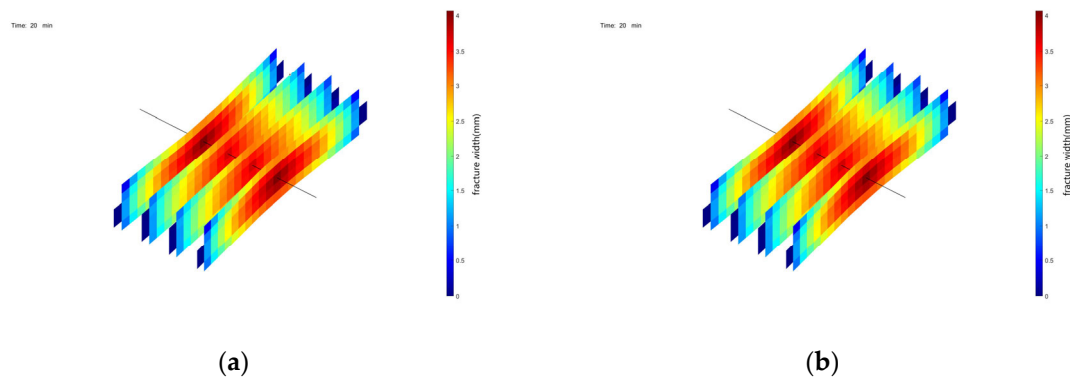


Figure 8. 3D visualization of fracture width at 20 min time step ((a) numerical simulation method; (b) model fitting methods).

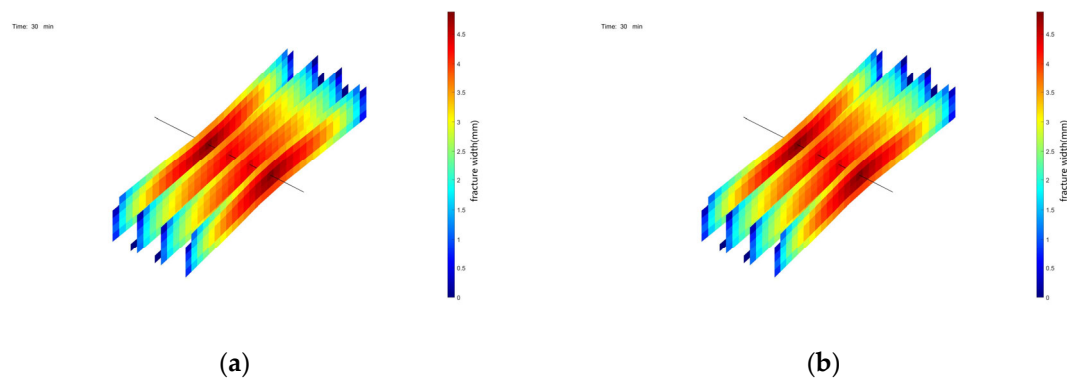


Figure 9. 3D visualization of fracture width at 30 min time step ((a) numerical simulation method; (b) model fitting methods).

Table 6. Average prediction accuracy at different time steps.

Time Step	R2 Score
10 min	0.9976
20 min	0.9803
30 min	0.9782

It is found that the model can effectively predict the morphology of 3D fractures at different time steps. By comparing the two different outputs, this study can tentatively conclude that the fracture generation principle of the numerical simulation software can be obtained, as constantly increasing the corresponding value of 0 to both sides of the “empty grid” is used to simulate the increase of cracks in the length direction.

The three-dimensional morphology of the fractures can be initially schematically obtained. As shown in Figures 10 and 11, the model predicted proppant concentrations and proppant placement concentrations at different time steps. It can be assumed that the predicted values are sufficiently similar in accuracy to the original data for application in engineering practice.

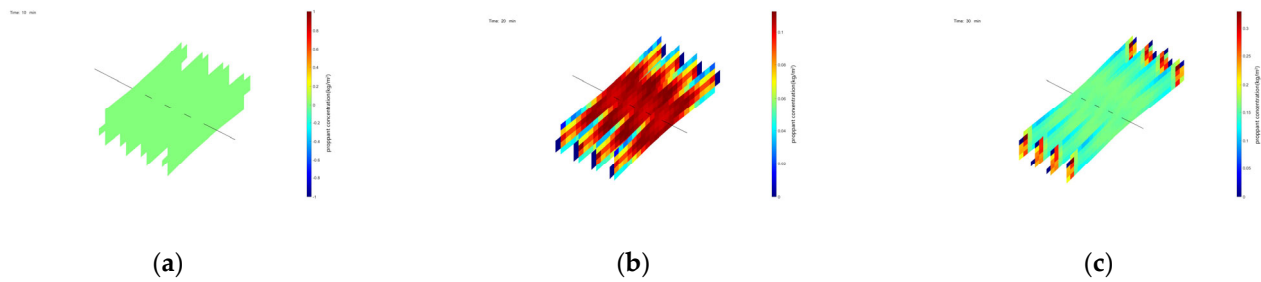


Figure 10. Model fitting to predict proppant concentration ((a) 10 min; (b) 20 min; (c) 30 min).

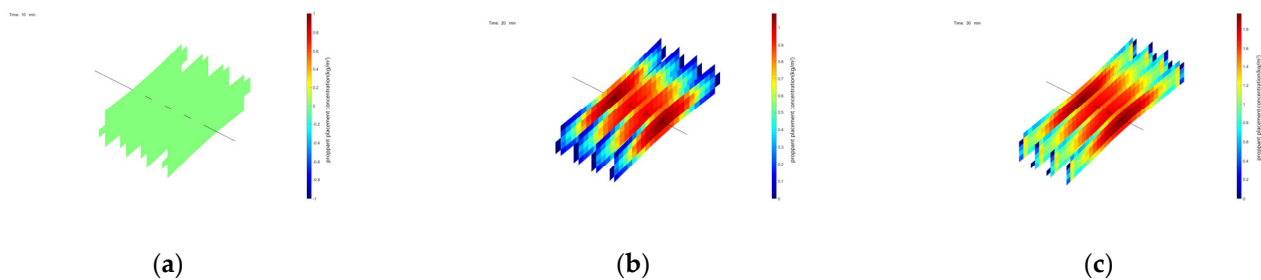


Figure 11. Model fitting to predict proppant placement concentration ((a) 10 min; (b) 20 min; (c) 30 min).

4.6. Actual Engineering Applications

Vertical well fracturing is the most widely used production enhancement technology in oil and gas exploration and development. Vertical well fracturing has the advantages of high production capacity, low extraction cost, simple operation and low risk. In addition, vertical well fracturing technology can be used to increase the production of exploited oil and gas wells and effectively extend the life and production capacity of oil and gas wells. Therefore, this section focuses on the proxy modeling study and prediction of single-layer fracturing in vertical wells. In this section, the fracture width, the fluid pressure in the fracture, and the volumetric and spreading concentrations of proppant in the fracture are selected as the main features for prediction.

Different parameter optimizations are needed for different forecasting requirements, in order to test the applicability of the fracturing simulation machine learning model. This chapter conducts a vertical well single-layer fracturing simulation prediction in which Bayesian optimization of hyper-parameter adaptive tuning optimization is implemented and the results are shown in Table 7.

Table 7. Hyperparameter results of XGBoost model based on Bayesian optimization in model prediction stage.

Hyperparameters	Range	Optimal Parameter Results
Number of decision trees	1–300	125
Maximum depth of the tree	1–25	6
Minimum sample size for splitting	2–10	3
Bootstrap sampling	True/False	True
Number of Bayesian iterations	/	40
k-fold cross-validation	/	5

4.6.1. Prediction of Actual Fracture Width in the Vertical Well Projects

The vertical well is designed with a total of two perforations, where the deeper perforation corresponds to a lower formation pressure and the shallower perforation corresponds to a higher formation pressure. The predicted fracture widths for the different

production process time steps in the vertical well were output in 3D visualization, and the results are shown in Figure 12.

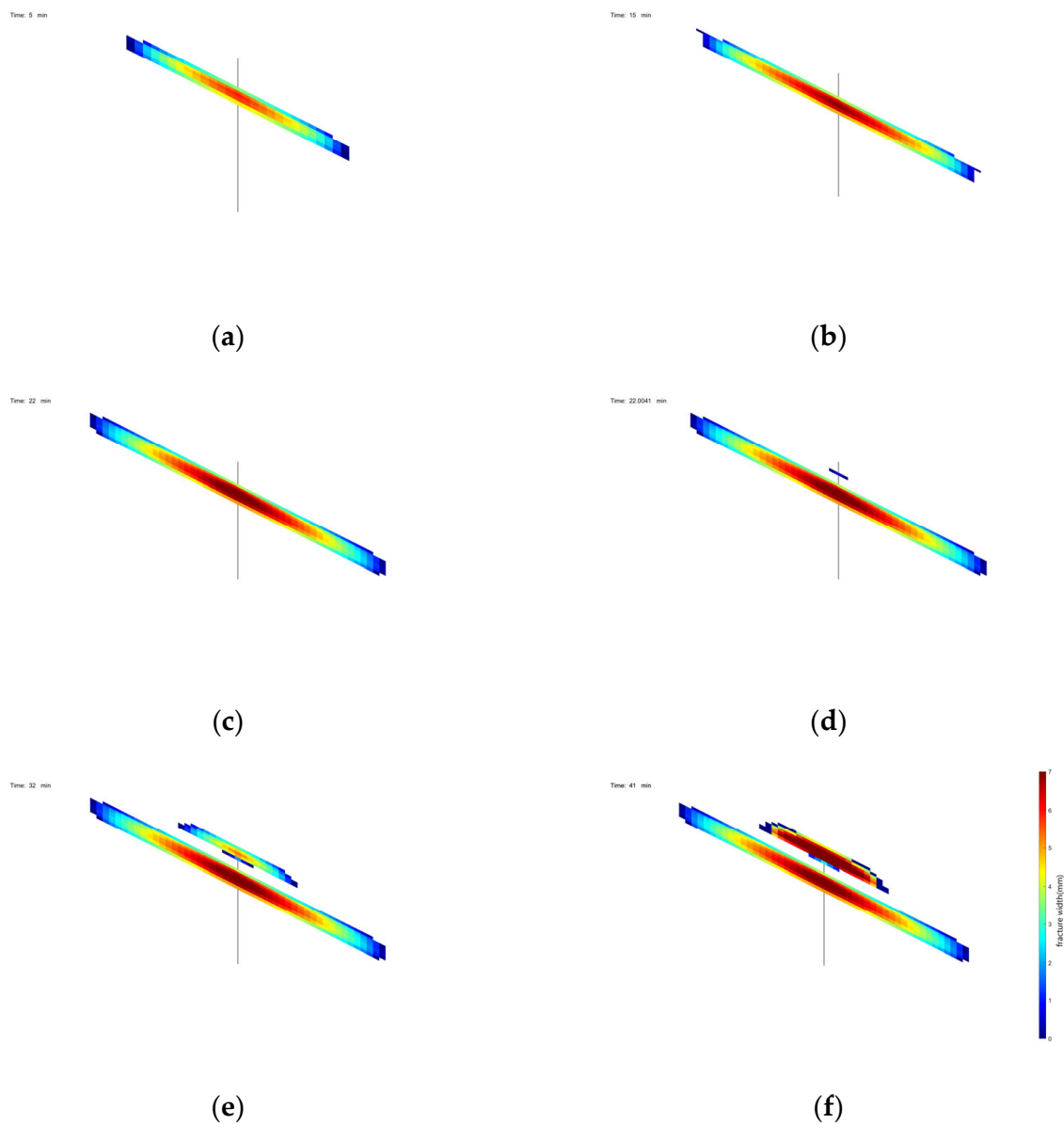


Figure 12. 3D visualization of fracture width prediction results at different time steps ((a) 5 min time step; (b) 15 min time step; (c) 22 min time step; (d) 23 min time step; (e) 32 min time step; (f) 41 min time step).

From the results in Figure 13, it can be found that the fracturing process is not exactly the same due to the different stratigraphic stresses at different depth conditions in the formation. Fractures are more likely to appear at depths with lower formation pressures. Fractures appear earlier at deeper locations in vertical wells than at shallower locations, and the upper fracture extends only after the time step of 22 min, but due to the difference in formation pressures, the extent of the upper fracture is still much smaller than that of the lower fracture.

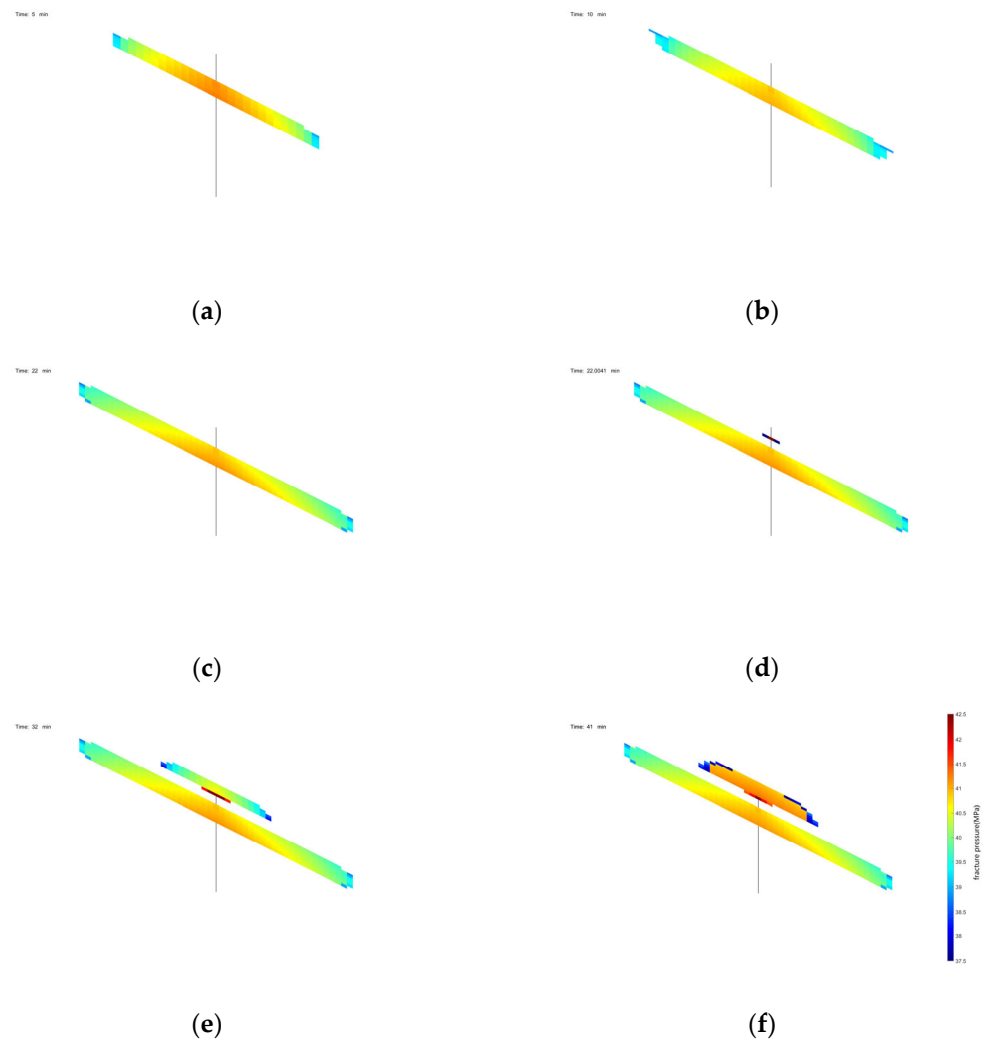


Figure 13. 3D visualization of fracture pressure prediction results at different time steps ((a) 5 min time step; (b) 15 min time step; (c) 22 min time step; (d) 23 min time step; (e) 32 min time step; (f) 41 min time step).

4.6.2. Prediction of Actual Fracture Pressures in the Vertical Well Projects

The predicted fracture pressures for different time steps of the production process in a vertical well are output in a 3D visualization, and the results are shown in Figure 13.

Similarly to the way the internal width of the fracture expands, the process of pressure change during fracture expansion varies at different locations. Under the effect of gravity, the fracturing fluid is injected and accumulates in the lower part of the fracture, resulting in a higher pressure in the lower part than the pressure at different heights of the grid at the same location.

4.6.3. Prediction of Actual Proppant Concentration in the Vertical Well Projects

Throughout the injection process, it is divided into two different injection liquids, in which the concentration of the injected liquid is 0 for 0 to 10 min. From 10 min onwards, it is replaced by another injection liquid of a fixed concentration, which is no longer 0 at that time. The predicted proppant concentration for different time steps of the production process in a vertical well is output in a 3D visualization, and the results are shown in Figure 14.

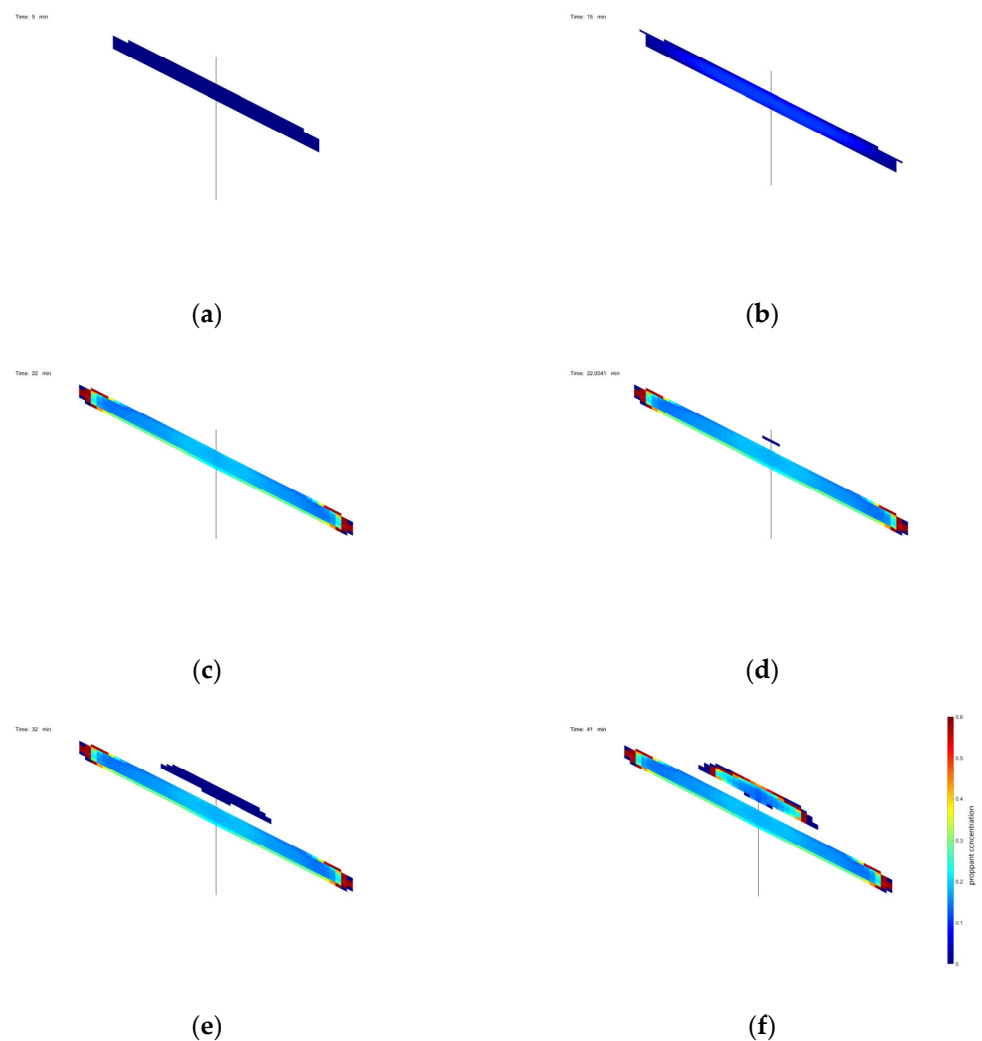


Figure 14. 3D visualization of proppant concentration prediction results at different time steps ((a) 5 min time step; (b) 15 min time step; (c) 22 min time step; (d) 23 min time step; (e) 32 min time step; (f) 41 min time step).

According to the results, it can be seen that after 10 min, the proppant concentration in the whole fracture is increasing with the continuous injection of fracturing fluid, because at this time, the proppant concentration in the injected fluid is no longer 0. In the fracture close to the injection hole, the proppant fixation rate is higher due to the accumulation of proppant. In the middle of the fracture, the volumetric concentration of proppant was lower than that at the edges, which was due to the wider width of the fracture, where the proppant had not yet accumulated, resulting in a larger volume of fracturing fluid inside, whereas at the outer edges of the fracture, the concentration of proppant was lower, which was due to the fact that the fracturing fluid had just been injected into the fracture. This showed the concentric step phenomenon of decreasing and then increasing from the inside to the outside.

4.6.4. Prediction of Actual Hydraulic Conductivity in the Vertical Well Projects

The inflow capacity reflects the ability of the fracturing fluid to be transported in the fracture during the hydraulic fracturing process. In general, the higher the inflow capacity, the faster the fracturing fluid expands in the fracture, given the same formation properties. Figure 15 shows the 3D visualization of the predicted results of the actual inflow capacity for the vertical well project at different time steps.

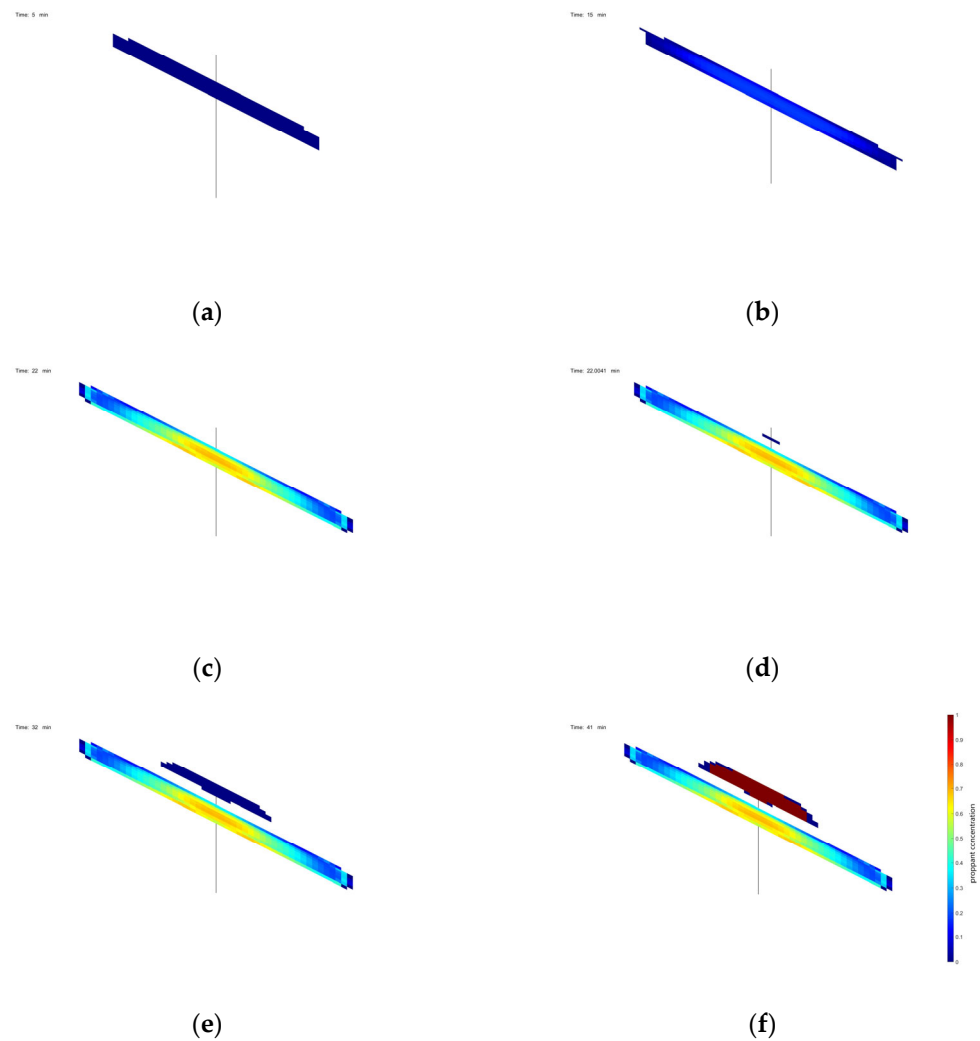


Figure 15. 3D visualization of inflow capacity prediction results at different time steps ((a) 5 min time step; (b) 15 min time step; (c) 22 min time step; (d) 23 min time step; (e) 32 min time step; (f) 41 min time step).

The results show that as the fracturing fluid is continuously injected, the inflow capacity increases with each increasing time step, and as the fracturing perforation is located in the center, the inflow capacity in the entire fracture shows a gradient that gradually decreases along the central perforation towards the marginal location.

4.7. Computational Time Analysis

In engineering practice, simulation using numerical simulation algorithms consumes a lot of computational power and time and the results have a certain lag, which greatly delays the project progress. Therefore, this study proposes utilizing machine learning to learn and predict the data in order to improve the computational speed and facilitate the engineering progress. In order to verify the speed improvement in machine learning compared to numerical simulation methods, different methods are used to compute the same data simultaneously and calculate the time used for comparison and plotting.

In this study, we used the same hardware to avoid errors. Fifty samples were taken from the input data using random sampling. Calculations were performed using the established machine learning model and the original numerical simulation software. The numerical simulation software uses the subprocess module to generate processes for numerical simulations and obtain their return states. The runtimes of the numerical simulations and model predictions were exported to a .csv file and saved separately. At the

same time, the times used by both methods are plotted on the same graph. The horizontal coordinate of the graph is the number of 50 randomly sampled samples, and the vertical coordinate is the time for each method. In order to facilitate the distinction, the numerical simulation method is plotted as a bar graph in blue with the vertical coordinates on the left side, ranging from 0 to 100 s, and the model prediction method is plotted as a line graph in red with the coordinates on the right side, ranging from 0 to 5 s, and the results are shown in Figure 16.

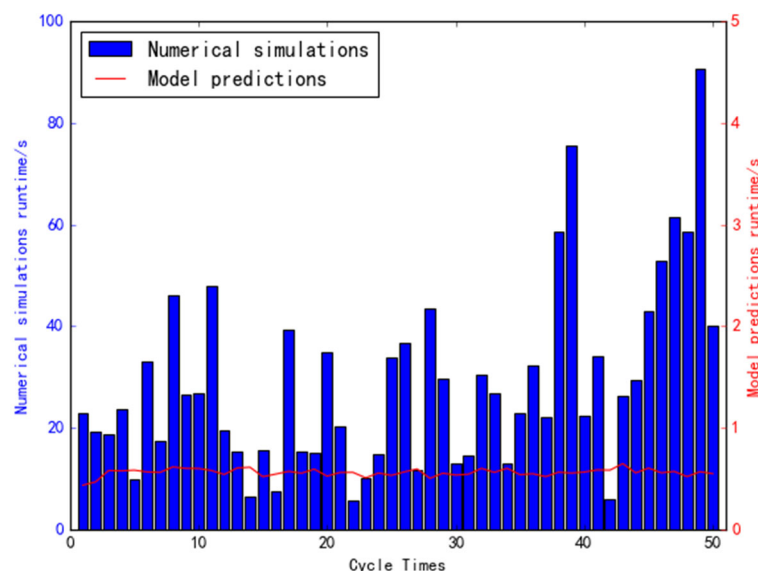


Figure 16. Numerical simulation versus model prediction runtime.

The results show that when numerical simulation is used to calculate the data, the time distribution varies greatly from 5 s to 100 s due to the limitations of different factors such as computer memory and the difficulty of solving the linear equations, which takes up a large amount of memory and arithmetical power. In contrast, when using machine learning models for data prediction, the time required is only between 0.1 and 0.8 s, and the time distribution is stable within the same time period, requiring much less memory and arithmetical power than the resources required by numerical simulation methods. This can be considered as fulfilling the requirement of accelerating the algorithmic process of numerical simulation proposed in this study.

5. Conclusions

The aim of this study was to optimize the acceleration of numerical simulation procedures using machine learning methods. The fracture extension process was simulated by a numerical simulation program for extension generation, generating fracture extension data for different input conditions and generating the corresponding datasets. Four different machine learning models were selected for prediction accuracy comparison, and the obtained datasets were imported into the models for prediction comparison. The limiting gradient model was selected as the subsequent machine prediction model, which was Bayesian-optimized for prediction and comparison of the numerical simulation data from the co-engineering practice.

As a result of the study, the following conclusions were reached.

1. For the analysis of the spatiotemporal fitting results of 2D data, it can be found that the model is able to predict the change rule of the length and the area of each fracture under different samples with relatively high accuracy under the initial time step after completing the training through the training set. However, with increasing time steps, the prediction accuracy decreases, and the change in geological conditions during the

hydraulic fracturing injection process leads to the above change rule of the model prediction accuracy.

2. The data obtained from the model predictions are visualized to generate three-dimensional images that can be directly observed. Comparing the prediction results of numerical simulation and machine learning models, it can be found that the numerical simulation software generates cracks by continuously adding “empty grids” with a value of 0 on both sides to simulate the increase in cracks in the length direction. Under three-dimensional conditions, the morphological changes during crack expansion can be visually depicted.
3. The hyperparameters of the XGBoost-based fracturing simulation model were adaptively optimized using a Bayesian optimization algorithm. The predicted fracture width, fracturing pressure, proppant concentration and inflow capacity of a single fracture in an actual vertical well have R2 values above 0.94, which makes the optimization of the model somewhat universal.
4. The study compares the time required to calculate the fracture morphology by different methods. Compared with the traditional numerical simulation methods, the average computation time of the prediction methods using machine learning models is improved by more than 95%, which greatly optimizes the computation process and reduces the required computational resources.

Author Contributions: Conceptualization, Z.L. and Q.L.; methodology, Z.L.; software, J.W.; validation, L.Y. and J.W.; investigation, X.W.; resources, Z.W.; data curation, J.W.; writing—original draft preparation, Z.L.; writing—review and editing, Q.L. and D.W.; visualization, M.F.; supervision, Q.L.; project administration, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the CNPC Science and Technology Project of Software Development on Volume Fracturing Optimization and Design (2020B-4118).

Data Availability Statement: The data are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mahmoud, A.; Gowida, A.; Aljawad, M.S.; Al-Ramadan, M.; Ibrahim, A.F. Advancement of Hydraulic Fracture Diagnostics in Unconventional Formations. *Geofluids* **2021**, *2021*, 4223858. [\[CrossRef\]](#)
2. Niitsuma, H.; Fehler, M.; Jones, R.; Wilson, S.; Albright, J.; Green, A.; Baria, R.; Hayashi, K.; Kaieda, H.; Tezuka, K.; et al. Current status of seismic and borehole measurements for HDR/HWR development. *Geothermics* **1999**, *28*, 475–490. [\[CrossRef\]](#)
3. Zou, Y.; Zhang, S.; Zhou, T.; Zhou, X.; Guo, T. Experimental Investigation into Hydraulic Fracture Network Propagation in Gas Shales Using CT Scanning Technology. *Rock Mech. Rock Eng.* **2016**, *49*, 33–45.
4. Groenenboom, J.; Romijn, R. Acoustic monitoring of hydraulic fracture growth. *First Break* **1997**, *15*, 295–303.
5. Liu, Z.; Wang, S.; Ye, H.; Yang, L.; Feng, F.; Lian, H.; Yang, D. Experimental study on the effects of pre-cracks, fracturing fluid, and rock mechanical characteristics on directional hydraulic fracturing with axial pre-cracks. *Geomech. Geophys. Geo-Energy Geo-Resour.* **2021**, *7*, 29. [\[CrossRef\]](#)
6. Axel, K.L.N.; John, C.S. A case study of hydraulic fracturing using finite element methods. *Can. Geotech. J.* **1999**, *36*, 861–875.
7. Galindo, T.S.A.; Muñoz, C.J.D. Simulation of the hydraulic fracture process in two dimensions using a discrete element method. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2007**, *75* Pt 2, 066109. [\[CrossRef\]](#)
8. Groenenboom, J.; Falk, J. Scattering by hydraulic fractures: Finite-difference modeling and laboratory data. *Geophysics* **2012**, *65*, 612–622. [\[CrossRef\]](#)
9. Ma, H.; Zhao, W.; Zhao, Y.; He, Y. A Data-Driven Oil Production Prediction Method Based on the Gradient Boosting Decision Tree Regression. *Comput. Model. Eng. Sci.* **2022**, *134*, 1773–1790. [\[CrossRef\]](#)
10. Song, H.; Zhu, J.; Wei, C.; Wang, J.; Du, S.; Xie, C. Data-driven physics-informed interpolation evolution combining historical-predicted knowledge for remaining oil distribution prediction. *J. Pet. Sci. Eng.* **2022**, *217*, 110795. [\[CrossRef\]](#)
11. Chen, G.; Luo, X.; Jiao, J.J.; Xue, X. Data-driven evolutionary algorithm for oil reservoir well-placement and control optimization. *Fuel* **2022**, *326*, 125125. [\[CrossRef\]](#)
12. Qin, S.; Liu, J.; Yang, X.; Li, Y.; Zhang, L.; Liu, Z. Predicting Heavy Oil Production by Hybrid Data-Driven Intelligent Models. *Math. Probl. Eng.* **2021**, *2021*, 5558623. [\[CrossRef\]](#)

13. Du, S.; Wang, M.; Yang, J.; Zhao, Y.; Wang, J.; Yue, M.; Xie, C.; Song, H. An enhanced prediction framework for coalbed methane production incorporating deep learning and transfer learning. *Energy* **2023**, *282*, 128877. [\[CrossRef\]](#)
14. Du, S.; Wang, J.; Wang, M.; Yang, J.; Zhang, C.; Zhao, Y.; Song, H. A systematic data-driven approach for production forecasting of coalbed methane incorporating deep learning and ensemble learning adapted to complex production patterns. *Energy* **2023**, *263*, 126121. [\[CrossRef\]](#)
15. Song, H.; Du, S.; Yang, J.; Zhao, Y.; Yu, M. Evaluation of hydraulic fracturing effect on coalbed methane reservoir based on deep learning method considering physical constraints. *J. Pet. Sci. Eng.* **2022**, *212*, 110360. [\[CrossRef\]](#)
16. Du, S.Y.; Zhao, X.G.; Xie, C.Y.; Zhu, J.W.; Wang, J.L.; Yang, J.S.; Song, H.Q. Data-driven production optimization using particle swarm algorithm based on the ensemble-learning proxy model. *Pet. Sci.* **2023**, *20*, 2951–2966. [\[CrossRef\]](#)
17. Yang, R.; Qin, X.; Liu, W.; Huang, Z.; Shi, Y.; Pang, Z.; Zhang, Y.; Li, J.; Wang, T. A Physics-Constrained Data-Driven Workflow for Predicting Coalbed Methane Well Production Using Artificial Neural Network. *SPE J.* **2022**, *27*, 1531–1552. [\[CrossRef\]](#)
18. Shen, Y.; Cao, D.; Ruddy, K.; de Moraes, L.F.T. Near Real-Time Hydraulic Fracturing Event Recognition Using Deep Learning Methods. *SPE Drill. Complet.* **2020**, *35*, 478–489. [\[CrossRef\]](#)
19. Shi, J.; Shen, B. Approximation schemes of stresses on elements for the three-dimensional displacement discontinuity method. *Eng. Anal. Bound. Elem.* **2014**, *48*, 63–72. [\[CrossRef\]](#)
20. Fath, H.A.; Madanifar, F.; Abbasi, M. Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. *Petroleum* **2020**, *6*, 80–91. [\[CrossRef\]](#)
21. Mohammadzaheri, M.; Tafreshi, R.; Khan, Z.; Ghodsi, M.; Franchek, M.; Grigoriadis, K. Modelling of petroleum multiphase flow in electrical submersible pumps with shallow artificial neural networks. *Ships Offshore Struct.* **2020**, *15*, 174–183. [\[CrossRef\]](#)
22. Salehinia, S.; Salehinia, Y.; Alimadadi, F.; Sadati, S.H. Forecasting density, oil formation volume factor and bubble point pressure of crude oil systems based on nonlinear system identification approach. *J. Pet. Sci. Eng.* **2016**, *147*, 47–55. [\[CrossRef\]](#)
23. Tahmasebi, P.; Hezarkhani, A. A fast and independent architecture of artificial neural network for permeability prediction. *J. Pet. Sci. Eng.* **2012**, *86–87*, 118–126. [\[CrossRef\]](#)
24. Al-Bulushi, N.I.; King, P.R.; Blunt, M.J.; Kraaijveld, M. Artificial neural networks workflow and its application in the petroleum industry. *Neural Comput. Appl.* **2012**, *21*, 409–421. [\[CrossRef\]](#)
25. Torabi, F.; Abedini, A.; Abedini, R. The Development of an Artificial Neural Network Model for Prediction of Crude Oil Viscosities. *Pet. Sci. Technol.* **2011**, *29*, 804–816. [\[CrossRef\]](#)
26. Chen, X.; Tang, C.; Du, Z.; Tang, L.; Wei, J.; Ma, X. Numerical simulation on multi-stage fractured horizontal wells in shale gas reservoirs based on the finite volume method. *Nat. Gas Ind. B* **2018**, *6*, 347–356. [\[CrossRef\]](#)
27. Zou, J.; Jiao, Y.Y.; Tan, F.; Lv, J.; Zhang, Q. Complex hydraulic-fracture-network propagation in a naturally fractured reservoir. *Comput. Geotech.* **2021**, *135*, 104165. [\[CrossRef\]](#)
28. Ma, X.; Sun, Y.; Guo, W.; Jia, R.; Li, B. Numerical simulation of horizontal well hydraulic fracturing technology for gas production from hydrate reservoir. *Appl. Ocean. Res.* **2021**, *112*, 102674. [\[CrossRef\]](#)
29. Peng, Z.; Desheng, Z. Study on fracture propagation and interaction mechanism during hydraulic fracturing. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *621*, 012134.
30. Barba, D.F.; Picano, F. A Novel Approach for Direct Numerical Simulation of Hydraulic Fracture Problems. *Flow Turbul. Combust.* **2020**, *105*, 335–357. [\[CrossRef\]](#)
31. Davarpanah, A.; Mirshekari, B. Numerical modelling of hydraulic fracturing procedure in hydrocarbon reservoirs. *Model. Earth Syst. Environ.* **2019**, *5*, 1297–1302. [\[CrossRef\]](#)
32. Shi, Y.; Qi, H.; Qi, X.; Mu, X. An efficient hyper-parameter optimization method for supervised learning. *Appl. Soft Comput. J.* **2022**, *126*, 109226. [\[CrossRef\]](#)
33. Zhang, L.; Zhang, J.; Wang, Y. A Novel Classification Model based on Ensemble Feature Selection and Hyper-parameter Optimization. *J. Artif. Intell. Pract.* **2022**, *5*, 11–17.
34. Chukwuemeka, O.N.; MacGregor AO, J.; Chukwualuka, L.N.; Ferguson, O. Automated Machine Learning Models and State-Of-The-Art Effort in Mitigating Combined Algorithm Selection and Hyperparameter Optimization Problems: A Review. *Mach. Learn. Res.* **2022**, *7*, 1–7.
35. Nametala, C.A.; Faria, W.R.; Pereira Júnior, B.R. On the performance of the Bayesian optimization algorithm with combined scenarios of search algorithms and scoring metrics. *Genet. Program. Evolvable Mach.* **2022**, *23*, 193–223. [\[CrossRef\]](#)
36. Vahid, R.; Somayeh, M.; Erik, C. Using Bayesian optimization algorithm for model-based integration testing. *Soft Comput.* **2021**, *26*, 3503–3525.
37. Dobrev, P.; Dyk, V.E.E.; Vorster, J.F. New approach to evaluating predictive models of photovoltaic systems. *Sol. Energy* **2020**, *204*, 134–143. [\[CrossRef\]](#)
38. Martins, C.J.; Sousa, A.L. *Bioelectronic Vision: Retina Models, Evaluation Metrics and System Design*; World Scientific Publishing Company: Singapore, 2009.
39. Wood, N.T.; Peoples, C.M. Pre-validation and performance prediction using pressure monitoring to evaluate HPLC method development changes. *Chromatographia* **2012**, *75*, 441–448. [\[CrossRef\]](#)
40. Tibshirani, J.R.; Efron, B. Pre-validation and inference in microarrays. *Stat. Appl. Genet. Mol. Biol.* **2011**, *1*. [\[CrossRef\]](#)
41. Anita, R.; Dávid, B.; Károly, H. Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification. *Molecules* **2021**, *26*, 1111.

42. Liu, D.; Lecampion, B.; Garagash, I.D. Propagation of a fluid-driven fracture with fracture length dependent apparent toughness. *Eng. Fract. Mech.* **2019**, *220*, 106616. [[CrossRef](#)]
43. Wang, Z.; Li, W.; Qiao, L.; Liu, J.; Yang, J. Hydraulic properties of fractured rock mass with correlated fracture length and aperture in both radial and unidirectional flow configurations. *Comput. Geotech.* **2018**, *104*, 167–184. [[CrossRef](#)]
44. Yang, L.; Wu, S.; Gao, K.; Shen, L. Simultaneous propagation of hydraulic fractures from multiple perforation clusters in layered tight reservoirs: Non-planar three-dimensional modelling. *Energy* **2022**, *254*, 124483. [[CrossRef](#)]
45. Luo, T.; Liu, S.; Liu, S. A productivity model for vertical wells with horizontal multi-fractures. *Int. J. Oil Gas Coal Technol.* **2022**, *31*, 225–241. [[CrossRef](#)]
46. Luo, T.; Liu, S.; Liu, S. Using XGBoost Regression to Analyze the Importance of Input Features Applied to an Artificial Intelligence Model for the Biomass Gasification System. *Inventions* **2022**, *31*, 225–241.
47. Coffey, B.; Kubacki, M.; Wen, Y.; Zhang, T.; Barajas, C.A.; Gobbert, M.K. Machine Learning with Feature Importance Analysis for Tornado Prediction from Environmental Sounding Data. *PAMM* **2021**, *20*, e202000112. [[CrossRef](#)]
48. Wu, M.H.; Cai, S.; Zeng, C.Y.; Wang, Z.F.; Zhao, N.; Zhu, L.; Wang, J. Big data analytics: An improved method for large-scale fabrics detection based on feature importance analysis from cascaded representation. *Int. J. Grid Util. Comput.* **2021**, *12*, 81–93. [[CrossRef](#)]
49. Qiu, C.; Schmitt, M.; Mou, L.; Ghamisi, P.; Zhu, X.X. Feature Importance Analysis for Local Climate Zone Classification Using a Residual Convolutional Neural Network with Multi-Source Datasets. *Remote Sens.* **2018**, *10*, 1572. [[CrossRef](#)]
50. Ma, D.; Xiao, J.; Li, Y.; Diao, Y.; Guo, Y.; Li, M. Feature importance analysis in guide strand identification of microRNAs. *Comput. Biol. Chem.* **2011**, *35*, 131–136. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.