

Article

Optimization of Abnormal Hydraulic Fracturing Conditions of Unconventional Natural Gas Reservoirs Based on a Surrogate Model

Su Yang ^{1,*}, Jinxuan Han ², Lin Liu ¹, Xingwen Wang ¹, Lang Yin ¹ and Jianfa Ci ¹

¹ Petroleum Engineering Technology Research Institute, Sinopec Southwest Oil and Gas Company, Deyang 618000, China

² School of Vehicle and Energy, Yanshan University, Qinhuangdao 066000, China; hanjx@ysu.edu.cn

* Correspondence: yangsusinopec1987@hotmail.com

Abstract: Abnormal conditions greatly reduce the efficiency of hydraulic fracturing of unconventional gas reservoirs. Optimizing the fracturing scheme is crucial to minimize the likelihood of abnormal operational conditions, such as pressure channeling, casing deformation, and proppant plugging. This paper proposes a novel machine learning-based method for optimizing abnormal conditions during hydraulic fracturing of unconventional natural gas reservoirs. Firstly, the main controlling factors of abnormal conditions are selected through a hybrid controlling analysis, upon which a surrogate model is established for predicting the occurrence probability of abnormal conditions, rather than whether abnormal conditions happen or not. Subsequently, a machine learning-based optimization algorithm is developed to minimize the occurrence probability of abnormal conditions, acknowledging their inevitability during the fracturing process. The optimal results demonstrate the proposed method outperforms traditional methods, on average. The proposed methodology is more in line with the needs of practical operation in an environment full of uncertainty.

Keywords: unconventional gas; abnormal conditions; machine learning; probability optimization; differential evolution



Citation: Yang, S.; Han, J.; Liu, L.; Wang, X.; Yin, L.; Ci, J. Optimization of Abnormal Hydraulic Fracturing Conditions of Unconventional Natural Gas Reservoirs Based on a Surrogate Model. *Processes* **2024**, *12*, 918. <https://doi.org/10.3390/pr12050918>

Academic Editor: Chuanliang Yan

Received: 26 February 2024

Revised: 19 April 2024

Accepted: 24 April 2024

Published: 30 April 2024



Copyright: © 2024 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

The prediction of abnormal working conditions plays an important role in the improvement of construction success and work efficiency during unconventional natural gas hydraulic fracturing [1]. Most unconventional natural gas is characterized by low permeability and low porosity, which makes it difficult to achieve better extraction results with conventional extraction methods [2]. Therefore, large-scale hydraulic fracturing is needed to expand the seepage channel to obtain industrial gas flow. Indeed, during large-scale hydraulic fracturing, it is common to encounter abnormal operating conditions, such as pressure channeling, casing deformation, and proppant plugging [3–5]. These abnormal conditions will lead to a series of serious consequences, such as production loss, environmental pollution, and waste of resources, and increase the difficulty and risk of the unconventional natural gas mining process [6]. Prior to conducting large-scale hydraulic fracturing operations, it is crucial to optimize the operational scheme and implement effective preventive measures [7]. This should be performed based on accurate predictions of these abnormal conditions. Currently, research on methods to prevent abnormal working conditions in hydraulic fracturing is centered around several aspects [8]. These include the establishment of complex working condition models, the utilization of multivariate sensor fusion technology, the application of intelligent algorithms, and the development of advanced control methods [9,10].

These approaches aim to enhance the ability to detect and mitigate abnormal conditions, thereby improving the overall operational efficiency and safety of hydraulic fracturing processes [11]. Certainly, it is important to acknowledge that these methods do have certain limitations and challenges. Some of the shortcomings include difficulties in establishing

accurate models and adjusting parameters, uncertainties associated with sensor data, the high complexity of algorithms, and increased computational overhead [12,13]. These limitations highlight the need for further research and development to address these issues and find more efficient and reliable solutions for preventing abnormal working conditions in hydraulic fracturing operations [14].

With the development of artificial intelligence, people have begun to turn to the use of machine learning methods to predict and prevent abnormal working conditions. A neural network algorithm is employed to monitor the oil pressure and casing pressure in the process of hydraulic fracturing in real-time to carry out early warning of accidents [15]. By using a support vector machine algorithm, the fracture parameters and reservoir properties (permeability and oil saturation) in the fracturing process are studied and analyzed to carry out prediction of the post-fracturing effect [16,17]. By using a restricted Boltzmann machine algorithm, post-fracturing production prediction of unconventional natural gas reservoirs is carried out to achieve prediction of the post-fracturing effect [18]. By using a decision tree algorithm, the open flow capacity of the gas well after fracturing is analyzed to evaluate the fracturing effect [19]. A decision tree algorithm is employed to analyze the open flow capacity of the gas well after fracturing to predict the effect after fracturing [20]. A data-driven method is employed to analyze the production of unconventional reservoirs such as shale and tight sandstone to predict the post-fracturing effect [21]. A random forest algorithm is employed to predict the production of fractured oil wells and the optimal fracturing parameters to predict the post-fracturing effect [22].

In this paper, a machine learning-based surrogate model is constructed to predict the occurrence probability of abnormal conditions. Additionally, a probabilistic optimization model for abnormal conditions is established based on the predictions of the machine learning model. In this optimization model, the prediction results of the machine learning model are not classified using a softmax operation. Instead, they directly output the expected probabilities of various abnormal conditions. According to the content of the article, optimization of fracturing construction parameters is achieved by optimizing the probability of abnormal conditions occurring. By leveraging a probabilistic optimization model based on machine learning predictions, construction parameters can be adjusted and optimized to minimize the likelihood of abnormal conditions during the fracturing process.

The rest of this paper is organized as follows: Section 2 describes the machine learning prediction model, including data collection and processing, principal control factor analysis, model construction and optimization, and comparative analysis of model results. Section 3 discusses the optimization model construction and the comparative analysis of a series of experimental results obtained using the optimization model algorithm. Finally, Section 4 is the conclusion of this study.

2. Prediction of the Odds of Abnormal Conditions Based on Machine Learning

2.1. Data Acquisition and Processing

In practice, due to improper behavior during the recording or operation process, the collected data was of low quality and needed to be preprocessed first. This preprocessing included data cleaning, segmentation, normalization, etc. Basic information and processing of the datasets used in this study are introduced in this section. According to the data analysis process shown in Figure 1, the analysis of the parameters required for abnormal conditions was performed.

Outlier screening primarily utilizes the 3σ method to identify abnormal data and determine abnormal conditions in order to facilitate screening. The sources of anomalies mainly consist of data anomalies (valuable) and entry anomalies (worthless, deleted, or modified) caused by accidents. The 3σ method is a commonly employed criterion for outlier screening. An outlier refers to a measured value that deviates from the mean value in a set of measured values by more than twice the standard deviation. During data processing, it is necessary to eliminate high anomalies, and whether to eliminate the abnormal values depends on the specific circumstances.

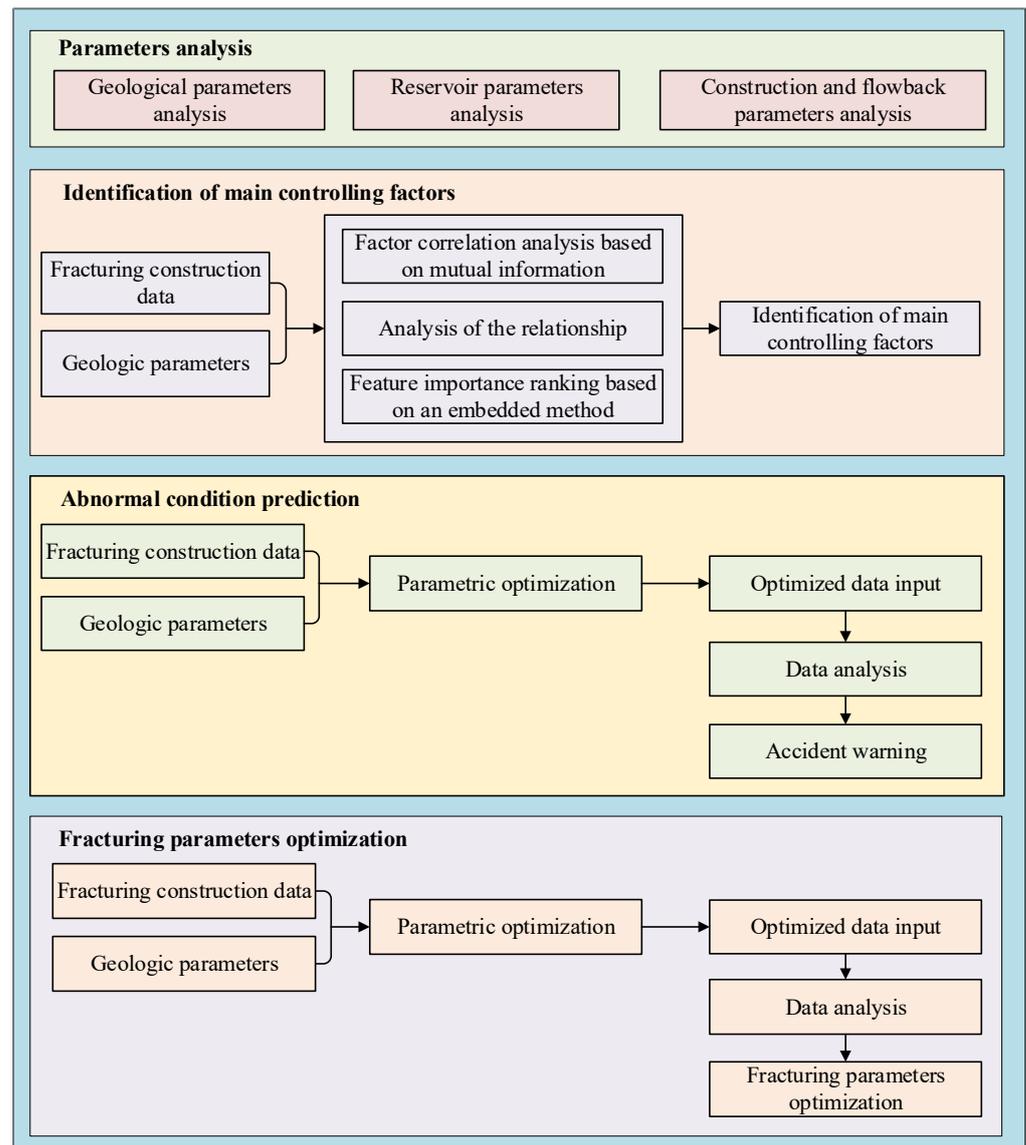


Figure 1. Flow chart of the proposed approach.

The K-Nearest Neighbor (KNN) classification algorithm is a relatively mature classification method in theory, and it is also a machine learning algorithm. The algorithm utilizes the correlation and similarity of data on different factors to fill in and correct missing values or outliers in data. The concept behind this method is that if a sample has the k most similar samples in the feature space (i.e., the nearest neighbors in the feature space), and a majority of these samples belong to a certain category, then it can be inferred that the sample also belongs to that category. Consequently, the KNN classification algorithm can be employed to automatically fill in default data. The flow chart illustrating the process of default value filling is depicted in Figure 2 [23,24].

The target oilfield reflects typical hydraulic fracturing of conventional shale gas in China. We performed the experiment on a real-world oilfield dataset sampled from the shale gas reservoir in Sichuan Province, China. Notably, the total dimension and records were 18 and 538, respectively. There were some outliers in the dataset, and we adopted related algorithms to identify and remove the outliers.

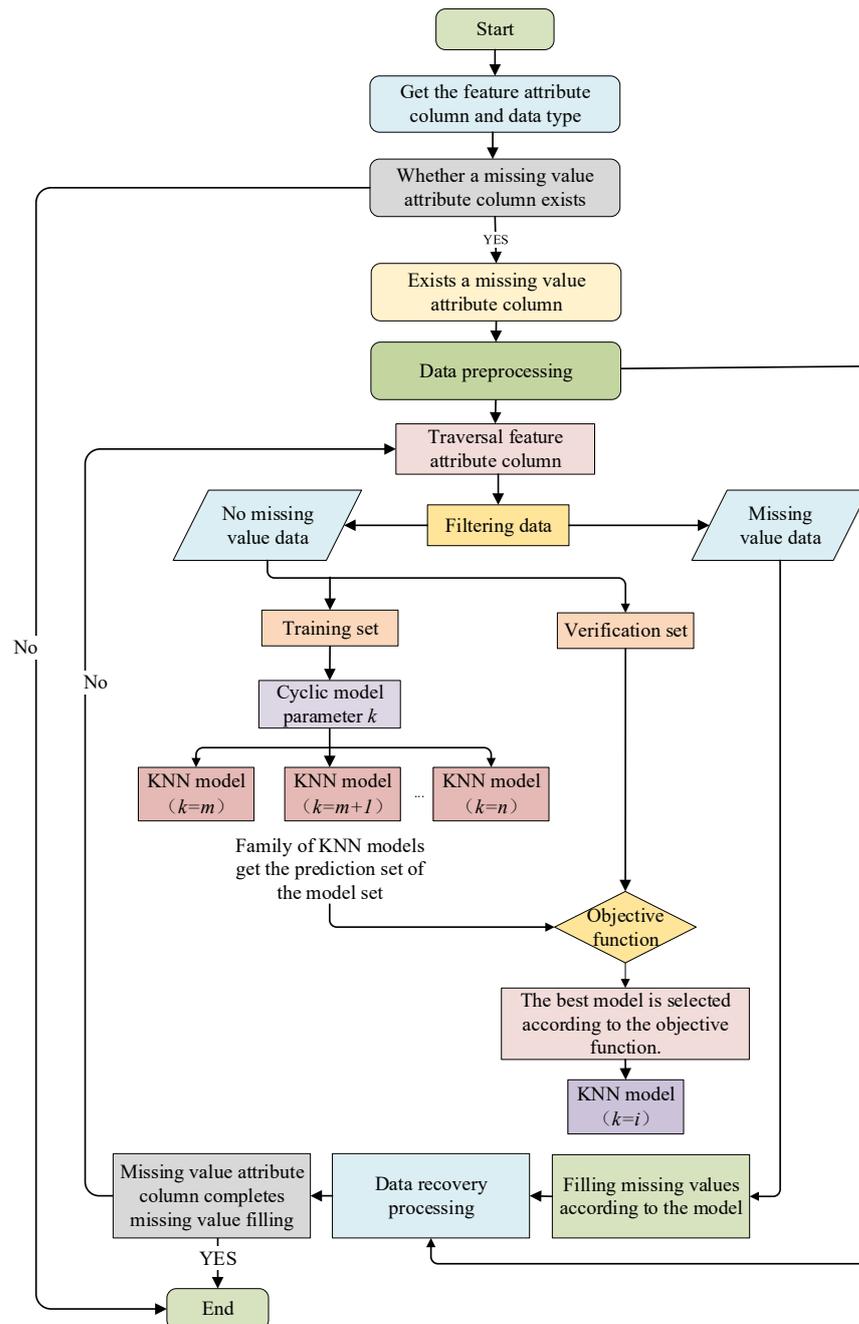


Figure 2. Data preprocessing.

2.2. Hybrid Analysis of the Main Controlling Factors of Abnormal Conditions

To determine the influence intensity of each controlling variable impacting on a target variable, five machine learning methods were utilized for further analysis, including correlation analysis, grey correlation analysis, main control factor identification based on mutual information, feature importance ranking based on embedded method, and Apriori correlation analysis, in which the occurrence of abnormal conditions was the target variable and the factors affecting the abnormal conditions were the input variables, whose dimension was 17, corresponding to Table 1. Based on the influence intensity values obtained, quantitative sorting was conducted from largest to smallest. The factors at the forefront of the sorting were selected as the main control factors that impact the occurrence of abnormal conditions [25,26].

Table 1. Parameters and the corresponding symbols.

Parameter Name	Symbols
Actual total number of temporary plugging sections	A ₁
Average cluster spacing (m)	A ₂
Average pump stop pressure gradient (%)	A ₃
Average segment length (m)	A ₄
Comprehensive sand ratio (%)	A ₅
Construction displacement (m ³)	A ₆
Flow back ratio (%)	A ₇
Normal temporary blocking pitch proportion range (%)	A ₈
Overall proportion of temporary plugging section (%)	A ₉
Normal temporary plugging ratio of temporary plugging section (%)	A ₁₀
Seam temporary plugging in place average pressure rise	A ₁₁
Seam temporary plugging in place pressure range	A ₁₂
The duration of the well is long (h)	A ₁₃
TOC (%)	A ₁₄
Total cumulative gas production (m ³)	A ₁₅
Total fracturing fluid (m ³)	A ₁₆
Young’s modulus (GPa)	A ₁₇

2.2.1. The Sorted Significance of Factors for Abnormal Conditions

- Rank Correlation Analysis

Rank correlation analysis, including correlation analysis, grey correlation analysis, and Apriori correlation analysis, is a type of non-parametric statistical method used to measure the correlation between two variables. Its main significance lies in the fact that the degree of correlation can be evaluated by the rank correlation coefficient, which is not affected by the data distribution [27].

The calculation formula for the rank correlation coefficient, specifically Spearman’s rank correlation coefficient, grey correlation analysis, and Apriori correlation analysis, is as follows:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \tag{1}$$

where ρ represents Spearman’s rank correlation coefficient, d denotes the rank difference corresponding to the two variables, and n is the number of samples.

Based on the rank correlation analysis, the sorted significance of factors impacting on pressure channeling, casing deformation, and proppant plugging were analyzed. The results, named as A_i^c , are shown in Figure 3 and Table 2.

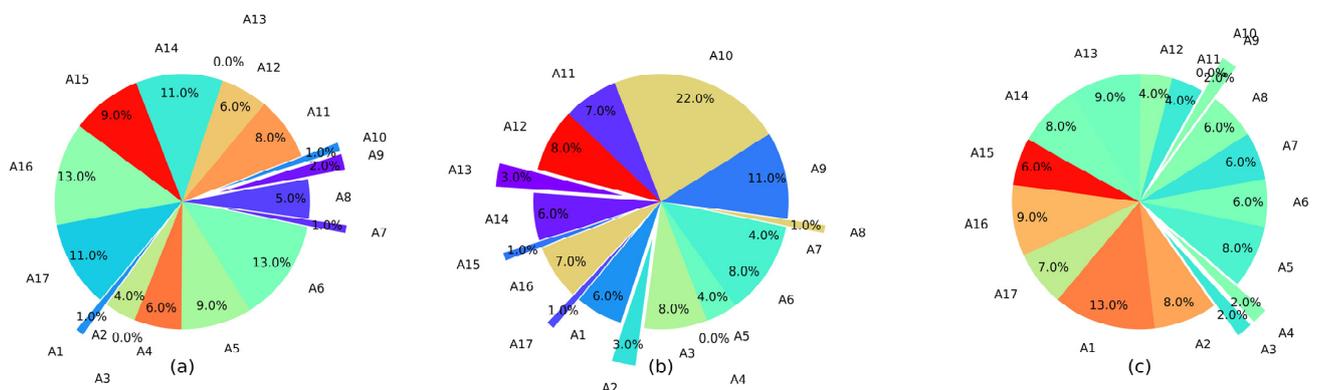


Figure 3. The correlation value between factors and the three abnormal conditions based on the rank correlation analysis. In (a–c), A_i ($i = 1, 2, \dots$) represents different main control factors. For example, A_1 represents the actual total number of temporary plugging sections, and A_2 denotes the average cluster spacing. For more details, please refer to Table 1.

Table 2. The sorted significance of factors for each abnormal condition based on the rank correlation analysis.

Abnormal Conditions	Sorted Significance of Factors
pressure channeling	$A_{16} > A_6 > A_{14} > A_{17} > A_5 > A_{15} > A_{11} > A_{12} > A_4 > A_8 > A_2 > A_9 > A_7 > A_1 > A_{10} > A_{13} > A_3$
casing deformation	$A_{10} > A_9 > A_6 > A_3 > A_{12} > A_{11} > A_{16} > A_1 > A_{14} > A_7 > A_5 > A_2 > A_{13} > A_{17} > A_{15} > A_8 > A_4$
proppant plugging	$A_1 > A_{16} > A_{13} > A_{14} > A_2 > A_5 > A_{17} > A_8 > A_{15} > A_7 > A_6 > A_{11} > A_{12} > A_3 > A_4 > A_9 > A_{10}$

- Mutual Information

Mutual information is an index employed to measure the correlation or dependence between two random variables. Its calculation formula is as follows:

$$I(X; Y) = \sum P(X, Y) \log \sum \left[\frac{P(X, Y)}{P(X)P(Y)} \right] \tag{2}$$

where X and Y are two random variables, $P(X, Y)$ is the probability of X and Y occurring at the same time, and $P(X)$ and $P(Y)$ are the probability of X and Y occurring alone, respectively [28].

Based on mutual information, the sorted significance of factors impacting on pressure channeling, casing deformation, and proppant plugging were analyzed. The results, named as A_i^m , are shown in Figure 4 and Table 3.

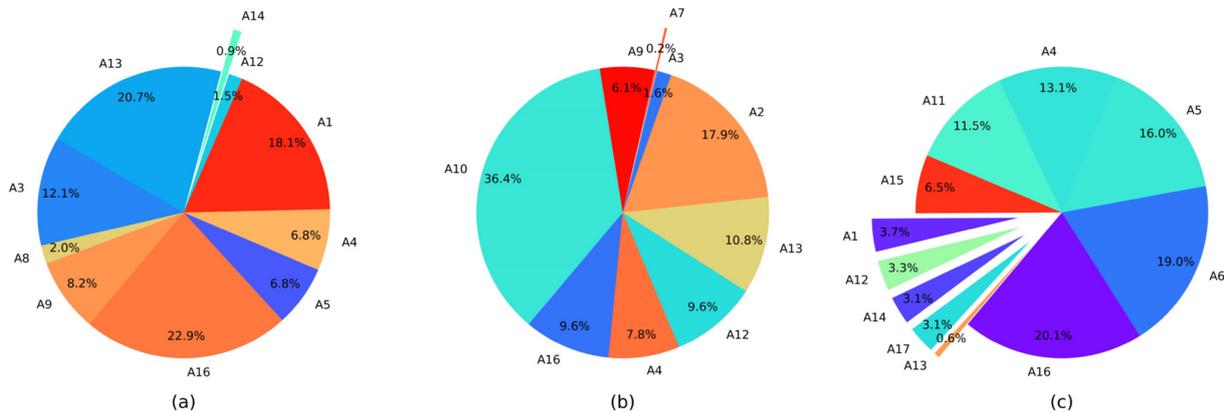


Figure 4. The correlation value between factors and the three abnormal conditions based on mutual information (a–c).

Table 3. The sorted significance of factors for each abnormal condition based on mutual information.

Abnormal Conditions	Sorted Significance of Factors
pressure channeling	$A_{16} > A_{13} > A_1 > A_3 > A_9 > A_4 > A_5 > A_8 > A_{12} > A_{14} > A_{10} > A_7 > A_2 > A_{17} > A_{15} > A_{11}$
casing deformation	$A_{10} > A_2 > A_{13} > A_{12} > A_{16} > A_4 > A_9 > A_3 > A_7 > A_8 > A_{17} > A_{14} > A_1 > A_{15} > A_{11} > A_5 > A_6$
proppant plugging	$A_{16} > A_6 > A_5 > A_4 > A_{11} > A_{15} > A_1 > A_{12} > A_{14} > A_{17} > A_{13} > A_{10} > A_9 > A_8 > A_7 > A_3 > A_2$

- Embedded Method

The embedded method is a machine learning method for feature selection. Its main significance is to improve model performance and generalization ability by automatically selecting the most relevant and important features during model training [29].

Taking L1 regularization as an example, the calculation formula is as follows:

$$\min \text{Loss}(\theta) + \lambda^* \|\theta\|_1 \tag{3}$$

where $\text{Loss}(\theta)$ represents the loss function of the model, θ denotes the parameter of the model, $\|\theta\|_1$ represents the L1 norm (the sum of absolute values) of θ , and λ denotes the regularization parameter, which is used to control the degree of feature selection.

Based on the embedded method, the sorted significance of factors impacting on pressure channeling, casing deformation, and proppant plugging were analyzed. The results, named as A_i^e , are shown in Figure 5 and Table 4.

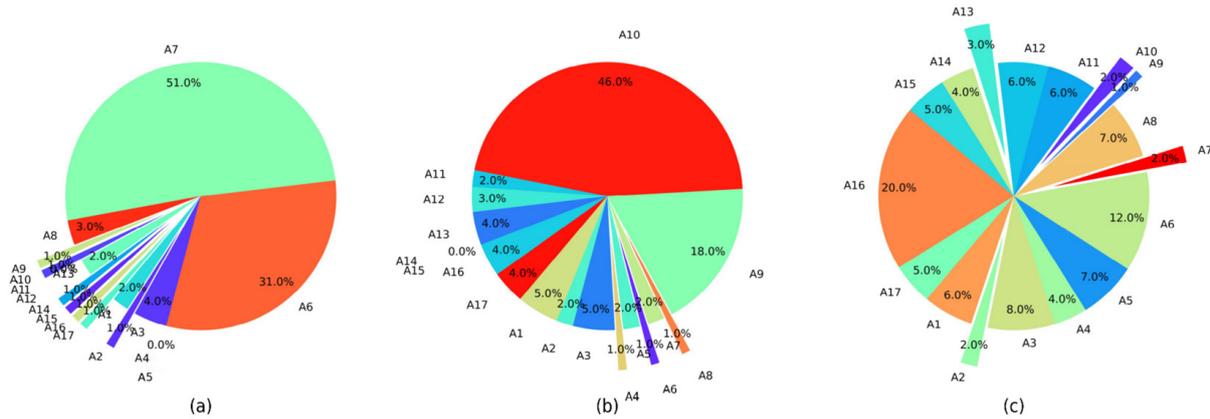


Figure 5. The correlation value between factors and the three abnormal conditions based on the embedded method (a–c).

Table 4. The sorted significance of factors for each abnormal condition based on the embedded method.

Abnormal Conditions	Sorted Significance of Factors
pressure channeling	$A_7 > A_6 > A_3 > A_8 > A_{13} > A_1 > A_2 > A_{16} > A_9 > A_{15} > A_{10} > A_{14} > A_4 > A_{17} > A_5 > A_{12} > A_{11}$
casing deformation	$A_{10} > A_9 > A_3 > A_1 > A_{13} > A_{16} > A_{12} > A_{11} > A_2 > A_5 > A_7 > A_6 > A_4 > A_8 > A_{15} > A_{17} > A_{14}$
proppant plugging	$A_{16} > A_6 > A_3 > A_5 > A_8 > A_{11} > A_1 > A_{12} > A_{15} > A_{17} > A_4 > A_{14} > A_{13} > A_7 > A_{10} > A_2 > A_9$

2.2.2. The Main Controlling Factors of Abnormal Conditions

Through the analysis of the sorted significance of factors impacting on abnormal conditions in Tables 2–4 of Section 2.1, we found that different methods in evaluating the significance of factors have distinct results. To alleviate the effects caused by the difference existing in machine learning, we proposed a comprehensive method for determining the main controlling factors from the perspectives of subjective and objective. For the perspective of subjective, the experts’ experience method was applied to calculate the sorted significance of factors, named as (j, A_i^s) , where $i, j = 1, 2, \dots, 17$, and i, j denote the indexes of factors and sorting significance of factors, respectively. Then, the combined weighting method can be formulized as follows:

$$A_i = \alpha_i[A_i^s] + \beta_i[A_i^o], \quad i = 1, 2, \dots, 17 \tag{4}$$

where A_i is the sorted significance of factors; α_i and β_i denote the subjective and objective weights, respectively; $[\cdot]$ denotes a function aiming at returning the indexes of sorting significance of factors; and $A_i^o = [A_i^c, A_i^m, A_i^e]$ denotes the objective sorted significance of factors.

By setting $\alpha_i = 0.4$, $\beta_i = [0.1, 0.2, 0.3]$, we obtained the comprehensive sorting results of significance regraded to each factor impacting on abnormal conditions, as shown in Table 5.

Table 5. The sorted significance of factors for each abnormal condition based on subjective and objective analysis.

Abnormal Conditions	Sorted Significance of Factors
pressure channeling	$A_6 > A_{16} > A_3 > A_{13} > A_7 > A_1 > A_8 > A_2 > A_{14} > A_9 > A_4 > A_{17} > A_{10} > A_{17} > A_5 > A_{12} > A_{11}$
casing deformation	$A_{16} > A_3 > A_2 > A_{12} > A_4 > A_6 > A_7 > A_8 > A_{10} > A_9 > A_{14} > A_{13} > A_1 > A_{15} > A_{11} > A_5 > A_{17}$
proppant plugging	$A_{16} > A_6 > A_4 > A_{14} > A_3 > A_1 > A_{12} > A_5 > A_{15} > A_{17} > A_7 > A_{10} > A_{11} > A_9 > A_8 > A_{13} > A_2$

According to Table 5, we can conclude that the main controlling factors that have a greater impact on abnormal conditions are the average cluster spacing A_2 , the average pump stop pressure gradient A_3 , the average segment length A_4 , the construction displacement A_6 , the flow back ratio A_7 , the TOC A_{14} , and the total fracturing fluid A_{16} .

2.3. Machine Learning-Based Surrogate Models

2.3.1. Model Construction

Firstly, taking the abnormal conditions as the output variables and the main controlling factors affecting abnormal conditions as the input variables, a machine learning model was established aimed at predicting the occurrence probability of abnormal conditions. Secondly, the field data were divided into training data with the size of training data set as 0.8, and test data with the size of test data set as 0.2, and then the model was trained and tested based on corresponding data, and consequently obtained multiple training confusing matrices and test confusing matrices for the three typical abnormal conditions. Thirdly, the results in confusing matrices were employed to judge whether the selected main control factors were reasonable. If reasonable, the current main controlling factors were retained for the next step. If it was unreasonable, we analyzed whether the unreasonable reason was improper selection of the main controlling factors or improper selection of models. Finally, according the unreasonable reason, the corresponding experiment was adjusted. If the main controlling factors were improperly selected, the quantitative sorting steps $Y = F(X_1)$ of the main controlling factors in the analysis of abnormal conditions were returned. If the model was improperly selected, the experiment went back to the model selection step in the establishment of the machine learning model [30,31].

The above clearly illustrates how reasonable and unreasonable subsequent processes should be carried out when the results of the confusion matrix are used to determine whether the selected control factors are reasonable, with emphasis on the treatment process when the results are unreasonable. The specific process is clearly shown in Figure 6 below.

Taking the occurrence of abnormal conditions $Y = f(X_1, X_2, \dots, X_n)$ as the output variable and the main control factors affecting abnormal conditions as the input variables, a machine learning model aimed at predicting abnormal conditions was established based on the negative log-likelihood loss function, which is shown as follows:

$$L(y, Y) = -(y * \log(Y) + (1 - y) * \log(1 - Y)) \quad (5)$$

where X represents the occurrence index of pressure channeling, casing deformation, and proppant plugging; X_i represents the influencing factors; f represents the regression model; and y, Y represent the real and predicted values of whether the abnormal working condition occurs, respectively.

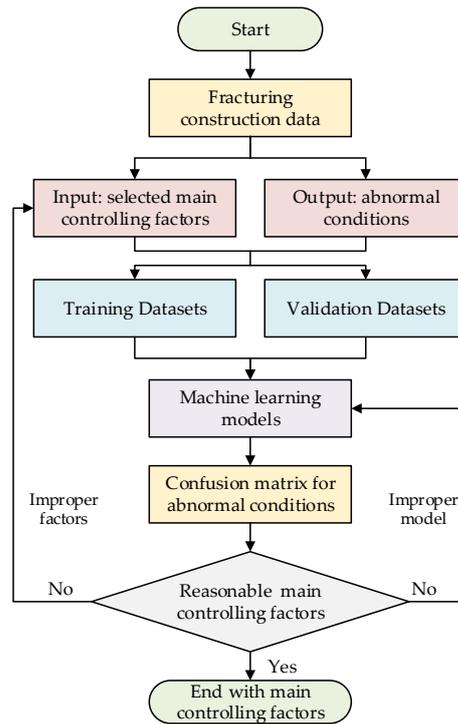


Figure 6. A flow chart of subsequent results using the results of the confusion matrix to determine whether the selected control factors are reasonable.

According to the main controlling factors of abnormal conditions obtained from Table 5, this section uses a variety of typical machine learning algorithms, including SVM, decision tree, random forest, extra trees, extra tress forest, gradient boosting, AdaBoost, bagging, gradient boosting trees, XGBoost, LightGBM, CatBoost, and CART, to model and predict the abnormal conditions in the study block, and uses 20% of the model data to validate. At the same time, the input and output parameters of the model, taking the abnormal conditions as the output parameters and the main controlling factors affecting abnormal conditions as the input parameters, are listed in Table 6.

Table 6. Controlling factors and output variables.

Input Parameter	Output Variable
$A_2, A_3, A_4, A_6, A_7, A_{14}, A_{16}$	Whether pressure channeling occurs Whether casing deformation occurs Whether proppant plugging occurs

2.3.2. Model Validation

By using various machine learning algorithms (support vector machine and decision tree) and ensemble learning algorithms (random forest, extra trees, etc.), prediction and modeling were carried out based on three abnormal conditions in the study area. Thus, the training and verification results of each algorithm for casing deformation, pressure channeling, and proppant plugging prediction were obtained from corresponding experiments. The results are shown in Table 7 [32,33].

By comparing the performance of various machine learning algorithms (support vector machine, decision tree, random forest, extra trees, etc.) in abnormal condition prediction, it was found that the extra trees and AdaBoost methods had optimal performance in predicting casing deformation. Random forest, AdaBoost, and bagging methods had optimal performance in predicting pressure channeling, while the random forest and extra trees methods had optimal performance in predicting proppant plugging. Overall, ensemble learning, including extra trees, random forest, AdaBoost and bagging, had higher accuracy

in the training and testing performances of predicting casing deformation, pressure channeling, and proppant plugging. This is because, in the case of a small sample size, ensemble learning can realize the integration of weak learners, and then can comprehensively utilize the diversity of sample attributes to improve the accuracy of fitting and prediction.

Table 7. Fitting results of machine learning model for abnormal working condition prediction.

Algorithm	Casing Deformation		Pressure Channeling		Proppant Plugging	
	AOTS	AOVS	AOTS	AOVS	AOTS	AOVS
SVM	0.71	0.83	0.92	0.88	0.95	0.94
Decision tree	1.00	0.61	1.00	0.88	1.00	0.88
Random forest	1.00	0.72	1.00	0.94	1.00	0.94
Extra trees	1.00	0.77	1.00	0.88	1.00	0.94
Extra trees forest	1.00	0.72	1.00	0.83	1.00	0.88
Gradient Boosting	1.00	0.66	1.00	0.88	1.00	0.83
Ada Boost	1.00	0.83	1.00	0.94	1.00	0.88
Bagging	0.95	0.66	0.98	0.94	1.00	0.88
Gradient Boosting Trees	1.00	0.67	1.00	0.87	1.00	0.85
XG Boost	1.00	0.73	1.00	0.84	1.00	0.83
Light GBM	1.00	0.76	1.00	0.78	1.00	0.85
Cat Boost	1.00	0.74	1.00	0.92	1.00	0.83
CART	0.98	0.76	0.98	0.93	1.00	0.82

Note: the accuracy of the training set and the verification set for each algorithm are denoted by AOTS and AOVS, respectively.

On average, the AdaBoost algorithm had higher accuracy in the training and testing performances of predicting casing deformation, pressure channeling, and proppant plugging situations. Next, the confusion matrices of the ensemble learning algorithm on casing deformation, pressure channeling, and proppant plugging in the training set and test set are presented in Figure 7.

		predicted value	
		0	1
true value	0	71 (TN)	12 (FP)
	1	6 (FN)	18 (TP)

Confusion matrix based on casing deformation (a)

		predicted value	
		0	1
true value	0	95 (TN)	0 (FP)
	1	6 (FN)	6 (TP)

Confusion matrix based on pressure channeling (b)

		predicted value	
		0	1
true value	0	94 (TN)	6 (FP)
	1	7 (FN)	0 (TP)

Confusion matrix based on proppant plugging (c)

Figure 7. Confusion matrix based on different conditions.

As shown in Figure 7a–c represent the test confusion matrices based on pressure channeling, casing deformation, and proppant plugging obtained by the AdaBoost algorithm, respectively. The test confusion matrix is further explained below. The test confusion matrix in (a) shows that there are 83 observations labeled as 0 where 71 prediction results (TN) are correct and 12 prediction results (FP) are incorrect, while there are 24 observations labeled as 1 where 6 prediction results (FN) are incorrect and 18 prediction results (TP) are correct. Similarly, Figure 7b,c test confusion matrices are also explained in this way. In particular, 0 means that the corresponding abnormal conditions do not occur, and 1 means that the corresponding abnormal conditions occur. Therefore, the accuracy of the verification sets of casing deformation, pressure channeling, and proppant plugging obtained by the AdaBoost algorithm were 83%, 94%, and 88%, respectively.

$$\text{accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{FP} + \text{FN} + \text{TP}) \quad (6)$$

Overall, it can be found that the AdaBoost algorithm had higher accuracy in the training and testing stages of predicting pressure channeling, casing deformation, and proppant plugging situations.

3. Optimization of the Fracturing Scheme Based on a Surrogate Model

In the process of unconventional natural gas hydraulic fracturing, it is impossible to completely avoid the occurrence of abnormal conditions due to the uncertainty of the construction site. The only thing we can do is to minimize the probability of abnormal conditions by optimizing the prediction of abnormal conditions, thereby improving the efficiency and safety of fracturing operations.

3.1. Model Construction

To address continuous optimization in classification tasks, the aforementioned machine learning methods were employed to calculate the occurring probability, not the occurring classification. Here, the output of the machine learning model was a three-dimensional vector $Y = (y_1, y_2, y_3)$ where y_1 , y_2 , and y_3 are real values in $[0, 1]$, standing for the occurring classification of abnormal conditions, pressure channeling $[1, 0, 0]$, casing deformation $[0, 1, 0]$, and proppant plugging $[0, 0, 1]$, respectively.

Then, a single-objective optimization model was established to reduce the occurrence of abnormal conditions. The decision variables were the controlling factors shown in Table 6, including average cluster spacing, average pump stop pressure gradient, average segment length, construction displacement, flow back ratio, TOC, and total fracturing fluid, represented by X_1, \dots, X_6 . The constraints were engineering parameters and their drainage region, including high-quality reservoir thickness, high-quality reservoir drilling rate, maximum principal stress, minimum principal stress, TOC, gas content, porosity, total hydrocarbons, etc. Moreover, the parameter θ denotes the model parameter.

Then, we have the following objective function:

$$\begin{aligned} \min \sum_{i=1}^n f_i(X, \theta^i) p_i(X, \theta^i) \\ \text{s.t. } \underline{X}_i \leq X_i \leq \bar{X}_i, 1 \leq i \leq 6 \end{aligned} \quad (7)$$

where $f_i(X, \theta^i)$ represents whether the abnormal condition occurs or not; $p_i(X, \theta^i)$ represents the occurrence probability of abnormal conditions [34,35]; and \underline{X}_i and \bar{X}_i represent the lower and upper bound of the i -th construction parameter, respectively.

A genetic algorithm, also known as a genetic optimization algorithm (GO) is a heuristic search algorithm, inspired by the theory of biological evolution, and used to solve optimization problems. A genetic algorithm simulates the process of biological evolution, and searches the solution space of the problem by simulating natural selection, crossover, and mutation mechanisms to find the optimal solution or approximate optimal solution. The flow chart illustrating the process of a GO is depicted in Figure 8.

The core idea of particle swarm optimization (PSO) is derived from the behavior law of birds foraging (Table 8). By using the information sharing of individuals in a group, the whole group can complete the retrieval of each region in the exploration of space, and finally find the location of food, that is, the existence of the optimal solution. Then, we have the following interpretation function:

$$v^{k+1} = \omega * v^k + c_1 * r_1 * (P_{pbest}^k - x^k) + c_2 * r_2 * (P_{gbest}^k - x^k) \quad (8)$$

A differential evolution algorithm (DE) is a kind of efficient global optimization algorithm. It is also a heuristic search algorithm based on groups, where each individual in the group corresponds to a solution vector. Now, the DE is widely used to solve complex optimization problems, and has achieved very good results. It simulates the process of biological evolution and generates a set of solutions through continuous evolution in order

to find the optimal solution to the problem. The flow chart illustrating the process of a DE is depicted in Figure 9.

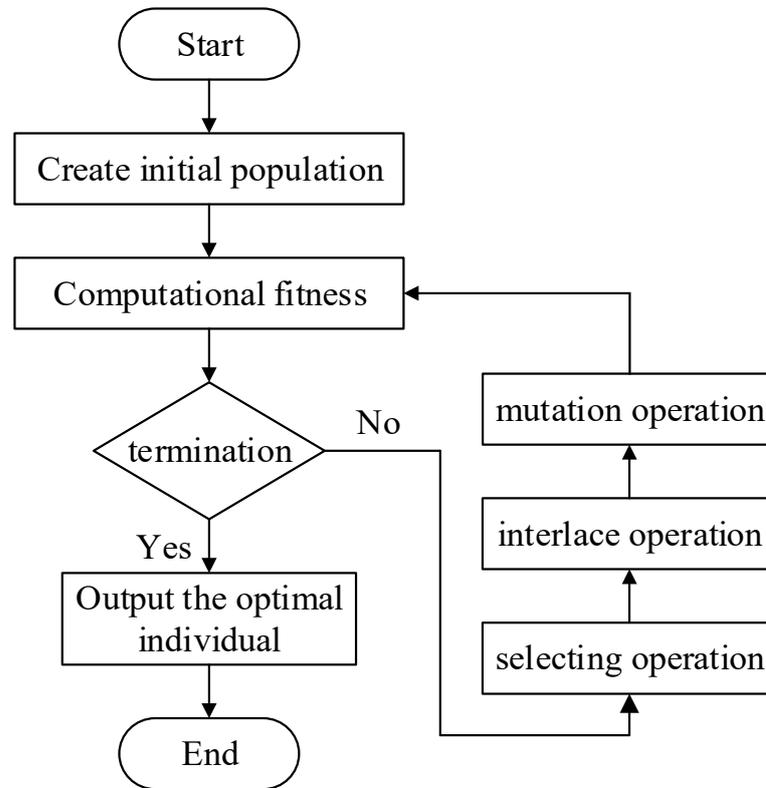


Figure 8. Flow chart of a genetic optimization algorithm.

Table 8. Description of parameters in the PSO algorithm expression.

Parameter	Meaning
k	Iterations
ω	Inertia weight
c_1	Individual learning factor
c_2	Group learning factor
r_1, r_2	Random numbers in [0, 1] increase the randomness of the search
x^k	In the k iteration, the position of the particle
v^k	In the k iteration, the displacement of the particle
p_{pbest}^k	The optimal location to which individual particles arrive
p_{gbest}^k	The optimal location for the particle population to reach

3.2. Model Solution and Result Comparison

As shown in Table 9, a series of machine learning methods are applied to map the classification into probability, aiming at optimizing the hydraulic fracturing scheme. The mapping probability denotes the occurrence probability of abnormal conditions, whose results should approximate 1, e.g., $[1, 0, 0] \rightarrow 0.86$ means the occurring probability of pressure channeling is 0.86 under the extra trees method for well 1.

Due to the machine learning model in Section 2 not having explicit expressions, the generic gradient-based optimization algorithm was no longer applicable. Therefore, it was necessary to employ heuristic algorithms based on evolutionary strategies, including a genetic optimization algorithm (GO), particle swarm optimization (PSO) algorithm, and differential evolution algorithm (DE) to achieve scheme optimization [36–40].

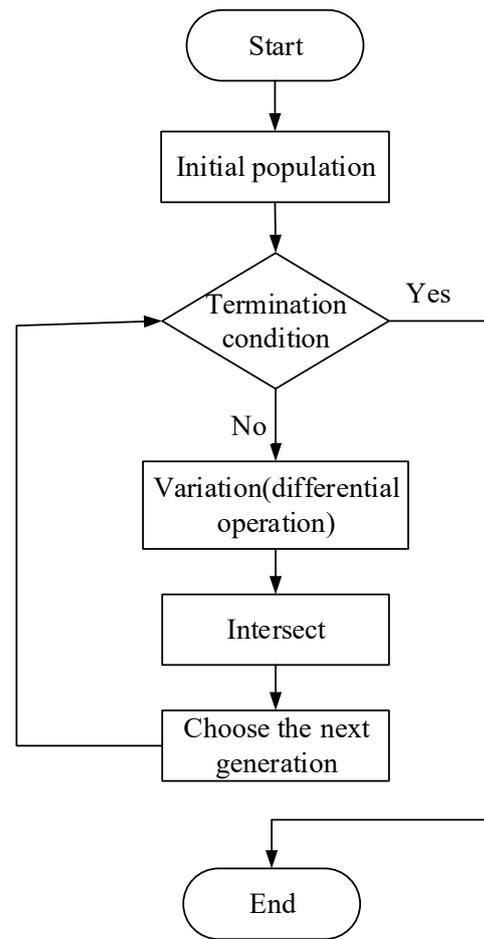


Figure 9. Flow chart of a differential evolution algorithm.

Table 9. The predictive probability of abnormal conditions under machine learning.

Well	Extra Trees			Decision Tree			Random Forest		
	Well 1	Well 2	Well 3	Well 1	Well 2	Well 3	Well 1	Well 2	Well 3
Abnormal conditions	[1, 0, 0]	[0, 1, 0]	[0, 0, 1]	[1, 0, 0]	[0, 1, 0]	[0, 0, 1]	[1, 0, 0]	[0, 1, 0]	[0, 0, 1]
Occurring Probability	0.86	0.99	0.74	0.98	0.99	0.96	0.86	0.98	0.72
Well	Logistic Regression			Ada Boost			Bagging		
	Well 1	Well 2	Well 3	Well 1	Well 2	Well 3	Well 1	Well 2	Well 3
Abnormal conditions	[1, 0, 0]	[0, 1, 0]	[0, 0, 1]	[1, 0, 0]	[0, 1, 0]	[0, 0, 1]	[1, 0, 0]	[0, 1, 0]	[0, 0, 1]
Occurring Probability	0.58	0.58	0.58	0.68	0.82	0.89	0.98	0.98	0.83

Note: [1, 0, 0], [0, 1, 0], and [0, 0, 1] denote the occurrence of pressure channeling, casing deformation, and proppant plugging, respectively.

By using three optimization algorithms to optimize the three abnormal conditions, we found that the optimization effects of these three optimization algorithms were different. From this, we present data on the changes in the values of the main control factors before and after the optimization of the three algorithms, as well as the corresponding changes in the probability of occurrence of abnormal working conditions (Figure 10).

According to Table 10, comparing the occurring probability of abnormal conditions before and after optimization, we can conclude the following observations for better understanding the optimization results: First, for wells 1–3, any optimization algorithm (GO, PSO, and DE) can better reduce the probability of occurrence of any abnormal conditions (pressure channeling, casing deformation, and proppant plugging) by the corresponding optimized construction parameters, e.g., the occurring probability of pressure channeling

in well 1 reduces from 0.86 to 0.27 under the GO, where $A_2 = 9.01, A_4 = 76.19, A_5 = 3.95, A_6 = 12.05, A_7 = 35.19, A_{14} = 5.07,$ and $A_{16} = 43,869$ are optimized as $A_2 = 8.64, A_4 = 66.07, A_5 = 7.15, A_6 = 10.90, A_7 = 46.34, A_{14} = 6.65,$ and $A_{16} = 27,148$. Second, to avoid pressure channeling in well 1, the GO and DE were more efficient than the PSO algorithm, i.e., the original probability 0.86 was optimized as 0.27 by the GO and DE, while the optimized result of PSO was 0.30; to avoid casing deformation in well 2, the DE was more efficient than the GO and PSO, i.e., the original probability 0.98 was optimized as 0.34 by the DE, while the optimized results of the GO and PSO were 0.35 and 0.36, respectively; to avoid proppant plugging in well 3, the PSO algorithm was more efficient than the GO and DE, i.e., the original probability 0.72 was optimized as 0.28 by PSO, while the optimized result of the GO and DE was 0.29. Third, the average optimization reduction probabilities of the three algorithms, the GO, PSO, and DE, for abnormal conditions were $0.86 - [0.27, 0.30, 0.27] = [0.59, 0.56, 0.59], 0.98 - [0.35, 0.36, 0.34] = [0.63, 0.62, 0.64],$ and $0.72 - [0.29, 0.28, 0.29] = [0.43, 0.44, 0.43],$ respectively, which means the PSO algorithm has the most significant optimization effect and the DE has the lowest optimization result in occurring probability.

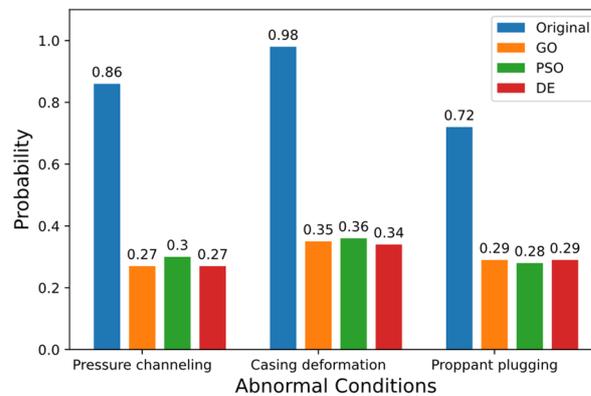


Figure 10. Comparison of the probability before and after optimization.

Table 10. Comparison of the results before and after optimization.

Pressure Channeling before and after Optimization in Well 1.								
Pressure Channeling	A_2	A_4	A_5	A_6	A_7	A_{14}	A_{16}	Probability
Original input/output	9.01	76.19	3.95	12.05	35.91	5.07	43,869	0.86
Optimized by GO	8.64	66.07	7.15	10.90	46.34	6.85	27,148	0.27
Optimized by PSO	23.14	57.68	7.55	13.55	47.62	6.76	20,000	0.30
Optimized by DE	8.64	66.07	7.13	10.90	46.73	6.69	27,148	0.27
Casing deformation before and after optimization in well 2.								
Casing deformation	A_2	A_4	A_5	A_6	A_7	A_{14}	A_{16}	Probability
Original input/output	9.51	72.1	3.78	12.00	32.03	3.83	45,838	0.98
Optimized by GO	10.18	76.79	4.09	11.36	29.12	5.31	46,849	0.35
Optimized by PSO	10.09	93.09	4.39	9.78	19.23	5.42	33,382	0.36
Optimized by DE	10.18	76.79	3.87	11.36	28.85	5.57	46,849	0.34
Proppant plugging before and after optimization in well 3.								
Proppant plugging	A_2	A_4	A_5	A_6	A_7	A_{14}	A_{16}	Probability
Original input/output	8.91	74.35	7.18	13.50	23.11	3.46	38,554	0.72
Optimized by GOA	10.68	93.20	6.38	12.69	22.31	3.24	32,400	0.29
Optimized by PSOA	9.90	100.0	7.56	14.06	21.95	3.87	33,125	0.28
Optimized by DEA	10.68	93.20	6.29	12.69	22.62	2.89	32,400	0.29

Note: The GO, PSO algorithm, and DE are denoted by bold **GO**, **PSO**, and **DE**, respectively. $A_2, A_4, A_5, A_6, A_7, A_{14},$ and A_{16} denote the average cluster spacing, average segment length, comprehensive sand ratio, construction displacement, flow back ratio, TOC, and total fracturing fluid parameter, respectively.

According to the optimized results of features A_2 , A_4 , A_5 , A_6 , A_7 , A_{14} , and A_{16} in Table 10, reducing the A_6 (construction displacement) helps to reduce the probability of pressure channeling, casing deformation, and proppant plugging, and increasing the A_4 (average segment length) helps to reduce the probability of casing deformation and proppant plugging.

4. Conclusions

This paper firstly focuses on establishing a machine learning model to analyze the primary controlling factors that affect abnormal conditions during unconventional natural gas hydraulic fracturing. Then, to avoid the occurrence of abnormal conditions, a series of machine learning methods were applied to predict abnormal conditions in hydraulic fracturing. Subsequently, under the pretrained model of prediction, three typical optimization algorithms, including the GO, PSO algorithm, and DE, were conducted to optimize the fracturing parameters for reducing the occurrence probability of abnormal conditions.

Based on the analysis of the primary control factors and the optimization model prediction, the specific conclusions of this study are as follows:

(1) The main factors affecting abnormal conditions.

For pressure channeling, the main controlling factors are the total amount of fracturing fluid, liquid strength, comprehensive sand ratio, maximum sand ratio, and average cluster spacing.

For casing deformation, the main controlling factors are liquid strength, Young's modulus, and the mechanical brittleness index.

For proppant plugging, the main controlling factors are 100–200 mesh sand dosage, construction displacement, total amount of fracturing fluid, and sand strength.

(2) Prediction of abnormal conditions.

By comparing diverse machine learning algorithms, it can be concluded that the random forest algorithm, additional tree algorithm, and AdaBoost algorithm have higher accuracy when training and testing whether abnormal conditions occur, which provides a fundamental for choosing the optimal machine learning method for optimization in avoiding the occurrence of abnormal conditions.

(3) A single-objective optimization model to reduce the occurrence probability of abnormal conditions was established.

(4) By comparing three different optimization algorithms, it can be concluded that different evolution algorithms have different optimization effects.

(5) In particular, proper parameters in hydraulic fracturing can help to reduce the probability of pressure channeling, casing deformation, and proppant plugging, and increasing the average segment length helps to reduce the probability of casing deformation and proppant plugging.

In addition to the factors collected in this paper, abnormal conditions may also be related to factors such as in-situ stress distribution, well completion, natural fractures, and adjacent well construction. These data are typical multi-source heterogeneous data. Consequently, future research will aim to develop a predictive method for abnormal conditions that incorporates spatiotemporal information. Additionally, the goal is to establish a more accurate and reasonable optimization model for abnormal conditions.

Author Contributions: Conceptualization, S.Y. and L.L.; methodology, S.Y. and J.H.; software, S.Y. and L.Y.; validation, S.Y., J.H. and L.L.; formal analysis, S.Y.; investigation, S.Y.; resources, J.C. and X.W.; data curation, L.L.; writing—original draft preparation, S.Y. and J.H.; writing—review and editing, S.Y.; visualization, L.Y.; supervision, L.L. and X.W.; project administration, S.Y. and X.W.; funding acquisition, S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC was funded by the Science and Technology Project of Sinopec Southwest Oil and Gas Company, grant number KJ-702-2220.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: All the editors and anonymous reviewers are gratefully acknowledged.

Conflicts of Interest: Authors Su Yang, Lin Liu, Xingwen Wang, Lang Yin and Jianfa Ci were employed by the company Sinopec Southwest Oil and Gas Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The Sinopec Southwest Oil and Gas Company had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Zhang, H.; Zhang, G.; Wang, G.; Wang, L.; Liu, Y.; Ren, Y.; Zheng, S. Classification and prediction method for ROP based on genetic algorithm optimization random forest model. *Sci. Technol. Eng.* **2022**, *22*, 15572–15578.
- Bao, J.; Yang, C.; Xu, J.; Liu, H.; Wang, G.; Zhang, G.; Zhou, D. A fully coupled and full 3D finite element model for hydraulic fracturing and its verification with physical experiments. *J. Tsinghua Univ.* **2021**, *61*, 833–841.
- Yang, P.; Zhang, S.; Zou, Y.; Lim, J.; Ma, X.; Tian, G.; Wang, J. Fracture propagation, proppant transport and parameter optimization of multi-well pad fracturing treatment. *Pet. Explor. Dev.* **2023**, *50*, 1225–1235. [[CrossRef](#)]
- Gong, D.; Chen, J.; Qu, Z.; Guo, T. Effect of radial well guidance on hydraulic fracturing crack propagation mechanism. *J. Southwest Pet. Univ.* **2018**, *40*, 122.
- Wang, J.; Qu, Z.; Guo, T.; Chen, M.; Lü, M. Numerical simulation of post-fracturing flowback considering fracturing fluid imbibition. *J. Shenzhen Univ. Sci. Eng.* **2023**, *40*, 56.
- Fei, W.; Tong, Z.; Jiabin, X.; Jingjing, G.; Jieli, S.; Shicheng, Z.; Yushi, Z. Fracturing pump-stopping pressure drop model considering proppant migration. *Acta Pet. Sin.* **2023**, *44*, 647.
- Liu, X.; Guo, T.; Qu, Z. Feasibility evaluation model and application of hydraulic fracturing in hydrate reservoir. *J. Cent. South Univ.* **2022**, *53*, 1058–1068.
- Zhang, X.; Wang, X.Z.; Guo, T.K. Experiment on evaluation of temporary plugging agent for in-fracture steering fracturing in Shunbei oilfield. *Lithol. Reserv.* **2020**, *32*, 170–176.
- Guo, T.; Lyu, M.; Chen, M.; Xu, Y.; Weng, D.; Qu, Z.; Liu, X. Proppant transport law in multi-branched fractures induced by volume fracturing. *Pet. Explor. Dev.* **2023**, *50*, 955–970. [[CrossRef](#)]
- Zheng, D.; Miska, S.; Ziaja, M.; Zhang, J. Study of anisotropic strength properties of shale. *AGH Drill. Oil Gas* **2019**, *36*, 93–112. [[CrossRef](#)]
- Lyu, M.; Qu, Z.; Guo, T.; Wang, X.; Li, J.; Wang, Z. Formation process and influence factors of proppant agglomerates in channel fracturing. *J. Xi'an Shiyou Univ.* **2021**, *36*, 63–69.
- Zhao, Z.; Zhang, Y.; Wang, J.; Qu, Z.; Guo, T.; Lu, M. Experiment on proppant placement for channel fracturing. *J. China Univ. Pet.* **2019**, *43*, 104–111.
- Zheng, D.; Turhan, C.; Wang, N.; Ashok, P.; van Oort, E. Prioritizing Wells for Repurposing or Permanent Abandonment Based on Generalized Well Integrity Risk Analysis. In Proceedings of the IADC/SPE International Drilling Conference and Exhibition, Galveston, TX, USA, 5–7 March 2024.
- Fan, Y.; Zhou, F.; Jiang, T.; Zhang, S.; Bai, S.; Zhao, X.; Yang, Q.; Yu, W. Research and Application of Environmentally Friendly Variable-viscosity Fracturing Fluid for Shale Gas. *Spec. Oil Gas Reserv.* **2023**, *30*, 147–152.
- Zou, Y.; Shi, S.; Zhang, S.; Li, J.; Wang, F.; Wang, J.; Zhang, X. Hydraulic fracture geometry and proppant distribution in thin interbedded shale oil reservoirs. *Pet. Explor. Dev.* **2022**, *49*, 1185–1194. [[CrossRef](#)]
- Zhang, L.; Qu, Z.; Lyu, M.; Guo, T.; Liu, X.; Wang, X. Support effect of different particle proppant combinations on complex fractures. *Fault Block Oil Gas Field* **2021**, *28*, 278–283.
- Jia, B.; Xian, C.; Tsau, J.S.; Zuo, X.; Jia, W. Status and outlook of oil field chemistry-assisted analysis during the energy transition period. *Energy Fuels* **2022**, *36*, 12917–12945. [[CrossRef](#)]
- Chen, Y.; Qu, Z.; Ding, Y.; Guo, T.; Bai, Y.; Wang, J. Unified backflow model after multilayer hydraulic fracturing. *Fault-Block Oil Gas Field* **2020**, *27*, 484–488.
- Qi, L.; Fengtao, Q.; Jingbin, H. Prediction model of mechanical ROP during drilling based on BAS-BP. *J. Xi'an Pet. Univ.* **2021**, *36*, 89–95.
- Qu, Z.Q.; Zhang, W.; Guo, T.K.; Sun, J.; Gong, F.C.; Tian, Y.; Li, X.L. Numerical simulation of heat mining performance of hot dry rocks with fracture network based on a local thermal non-equilibrium model. *J. China Univ. Pet.* **2019**, *43*, 90–98.
- Xu, R.; Guo, T.; Qu, Z.; Chen, M.; Qin, J.; Mou, S.; Zhang, Y. Numerical simulation of fractured imbibition in a shale oil reservoir based on the discrete fracture model. *Chin. J. Eng.* **2022**, *44*, 451–463.
- Pan, Y.; Wang, Y.; Che, M.; Liao, R.; Zheng, H. Post-fracturing production prediction and fracturing parameter optimization of horizontal wells based on grey relation projection random forest algorithm. *J. Xi Shiyou Univ.* **2021**, *36*, 71–76.
- Huang, A.; Xu, R.; Chen, Y.; Guo, M. Research on multi-label user classification of social media based on ML-KNN algorithm. *Technol. Forecast. Soc. Change* **2023**, *188*, 122271. [[CrossRef](#)]

24. Huang, L.; Song, T.; Jiang, T. Linear regression combined KNN algorithm to identify latent defects for imbalance data of ICs. *Microelectron. J.* **2023**, *131*, 105641. [[CrossRef](#)]
25. Tong, S.; Mohanty, K.K. Proppant transport study in fractures with intersections. *Fuel* **2016**, *181*, 463–477. [[CrossRef](#)]
26. Zheng, D.; Ozbayoglu, E.; Miska, S.Z.; Liu, Y.; Li, Y. Cement sheath fatigue failure prediction by ANN-based model. In Proceedings of the Offshore Technology Conference, Houston, TX, USA, 2–5 May 2022.
27. Abbasi, M.A. A comparative Study of Flowback Rate and Pressure Transient Behaviour in Multifractured Horizontal Wells. Master's Thesis, University of Alberta, Edmonton, AL, Canada, 2013.
28. Wu, Z.; Dong, L.; Cui, C.; Cheng, X.; Wang, Z. A numerical model for fractured horizontal well and production characteristics: Comprehensive consideration of the fracturing fluid injection and flowback. *J. Pet. Sci. Eng.* **2020**, *187*, 106765. [[CrossRef](#)]
29. Yamina, A.; Hammerquist, C.; Ouenes, A. Anisotropic damage mechanics for asymmetric hydraulic fracture height propagation in a layered unconventional gas reservoir. *J. Nat. Gas Sci. Eng.* **2019**, *67*, 1–13.
30. Cui, C.; Wang, Z.; Wu, Z.; Cheng, X.; Ye, Y. Comprehensive proppant settling model in hydraulic fractures of unconventional gas reservoir considering multifactorial influence. *Arab. J. Geosci.* **2020**, *13*, 788. [[CrossRef](#)]
31. Wang, K.; Zhang, G.; Wang, Y.; Zhang, X.; Li, K.; Guo, W.; Du, F. A numerical investigation of hydraulic fracturing on coal seam permeability based on PFC-COMSOL coupling method. *Int. J. Coal Sci. Technol.* **2022**, *9*, 10. [[CrossRef](#)]
32. Abbasi, M.A.; Dehghanpour, H.; Hawkes, R.V. Flowback analysis for fracture characterization. In Proceedings of the SPE Canadian Unconventional Resources Conference, Calgary, AB, Canada, 30 October–1 November 2012.
33. Alotaibi, M.A.; Miskimins, J.L. Slickwater proppant transport in hydraulic fractures: New experimental findings and scalable correlation. *SPE Prod. Oper.* **2018**, *33*, 164–178. [[CrossRef](#)]
34. Al Tammar, M.J.; Agrawal, S.; Sharma, M.M. Effect of geological layer properties on hydraulic-fracture initiation and propagation: An experimental study. *SPE J.* **2019**, *24*, 757–794. [[CrossRef](#)]
35. Hu, X.; Wu, K.; Song, X.; Yu, W.; Zuo, L.; Li, G.; Shen, Z. Development of a new mathematical model to quantitatively evaluate equilibrium height of proppant bed in hydraulic fractures for slickwater treatment. *SPE J.* **2018**, *23*, 2158–2174. [[CrossRef](#)]
36. Veza, I.; Spraggon, M.; Fattah, I.R.; Idris, M. Response surface methodology (RSM) for optimizing engine performance and emissions fueled with biofuel: Review of RSM for sustainability energy transition. *Results Eng.* **2023**, *18*, 101213. [[CrossRef](#)]
37. Nakkeeran, G.; Krishnaraj, L. Prediction of cement mortar strength by replacement of hydrated lime using RSM and ANN. *Asian J. Civ. Eng.* **2023**, *24*, 1401–1410. [[CrossRef](#)]
38. Khatri, K.A.; Shah, K.B.; Logeshwaran, J.; Shrestha, A. Genetic algorithm based techno-economic optimization of an isolated hybrid energy system. *CRF* **2023**, *8*, 1447–1450.
39. Du, W.; Ma, J.; Yin, W. Orderly charging strategy of electric vehicle based on improved PSO algorithm. *Energy* **2023**, *271*, 127088. [[CrossRef](#)]
40. Song, Y.; Zhao, G.; Zhang, B.; Chen, H.; Deng, W.; Deng, W. An enhanced distributed differential evolution algorithm for portfolio optimization problems. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106004. [[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.