processes

MDPI

Article

Drilling Overflow Diagnosis Based on the Fusion of Physical and Intelligent Algorithms

Yadong Shi [1], Hongda Hao [1,*], Rentong Liu [2], Song Deng [1], Chaowei Li [1], Qiu Li [1] and Chengguo Liu [1]

[1] School of Petroleum and Natural Gas Engineering, Changzhou University, Changzhou 213164, China; shiyd_yd@163.com (Y.S.); dengsong@cczu.edu.cn (S.D.); chw25@163.com (C.L.); liq@cczu.edu.cn (Q.L.); cgliu@cczu.edu.cn (C.L.)
[2] CNPC Engineering Technology R&D Co., Ltd., Beijing 102206, China; liurtdr@cnpc.com.cn
* Correspondence: hongdahao@cczu.edu.cn

Abstract: The diagnosis of overflow risk has always been an important area of research in drilling operations in the field of oil and gas engineering. In the face of the limitations and lag of traditional overflow diagnosis methods, the practical application effect of existing models and methods is not obvious, and there is no integration of the physical model and the intelligent algorithm model for overflow diagnosis, this paper proposes a method of adaptive weight fusion of physical model and intelligent algorithm model diagnosis results. Based on the fusion of the physical model and the intelligent algorithm model, an overflow diagnosis model of managed pressure drilling is established. The research results show that the fusion model in this paper can combine the accuracy weight of the physical model and the intelligent algorithm model for the intelligent diagnosis of overflow risk, which improves the mechanization and interpretability of the model and diagnosis results while ensuring accuracy and efficiency. And the intelligent algorithm model used in the fusion model is superior to other algorithm models. The overflow diagnosis accuracy of the fusion model on the test set samples reaches more than 98%, and the accuracy of the validation set is 94.25%. The content of this study provides guidance for drilling overflow diagnosis and model fusion methods.

Keywords: overflow diagnosis; physical model; intelligent algorithm model; adaptive weight fusion; managed pressure drilling; model fusion

check for updates

Academic Editor: Maoyin Chen

Received: 12 January 2025
Revised: 29 January 2025
Accepted: 11 February 2025
Published: 18 February 2025

Citation: Shi, Y.; Hao, H.; Liu, R.; Deng, S.; Li, C.; Li, Q.; Liu, C. Drilling Overflow Diagnosis Based on the Fusion of Physical and Intelligent Algorithms. *Processes* **2025**, *13*, 577. https://doi.org/10.3390/pr13020577

Copyright: © 2025 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

Drilling and completion engineering is a key link in the process of oil and gas exploration and development, and its cost accounts for about 50% of the total cost [1]. Among them, drilling engineering is an extremely complex underground drilling operation, full of concealment and uncertainty. For oil and gas resources in complex geological environments, such as offshore and ultra-deep reservoirs [2], as the drilling depth increases, the challenges of high temperature and pressure, narrow the safe density window (the pressure range between the three pressure curves of formation fracture pressure, formation pore pressure, and formation collapse pressure that does not cause formation fracture, overflow, or collapse), and other complex conditions [3,4] lead to multiple difficulties in drilling efficiency, risk, and cost [1]. As a result, well control risks are prominent and complex downhole situations occur frequently. Overflow is one of the most common complicated situations in the drilling process. In the process of drilling, when the density of the drilling fluid is less than the density of the formation fluid and the pressure of the formation fluid cannot be balanced, the bottom hole pressure will be less than the formation pressure,

and the oil, gas and water in the formation will be pressed into the wellbore, resulting in overflow. For gas wellbore, gas intrusion and oil well kick will occur. If it is not handled in time, it will further deteriorate into a serious blowout within 10–30 min [5] and then cause great losses. Therefore, it is necessary to timely and accurately diagnose the overflow risk during drilling.

However, traditional empirical- and physics-based methods are limited and struggle to cope with the increasingly complex drilling process [1,6]. Traditional overflow detection methods require the monitoring of surface parameters (e.g., mud flow rate, mud volume) [7] or installing more sensors on the drilling platform [8]. These methods are based on obvious changes in physical parameters and surface phenomena to judge the overflow, which generally has time lags and depends on the experience, responsibility and consciousness of the monitor. Prior drilling parameters are not effectively utilized, which has great limitations and cannot provide an early diagnosis of overflow [9,10].

In order to avoid or reduce the drilling complexity caused by pressure, Managed Pressure Drilling (MPD) technology is mostly adopted [11]. MPD technology controls the wellbore pressure and distribution by adjusting the drilling parameters [12], which basically ensures that the bottomhole pressure is within a narrow density window and achieves the purpose of safe drilling. However, managed pressure drilling cannot completely maintain the bottomhole pressure within a narrow density window [13], but fortunately, it provides relatively rich and fine-managed pressure drilling data. Therefore, a variety of drilling parameters can be effectively used to diagnose the overflow through an intelligent algorithm model.

Compared with other methods, artificial intelligence and big data technology have strong nonlinear fitting and information mining capabilities, and have significant advantages in solving complex problems [1]. Farahani et al. presented a robust model to apply LSSVM for predicting the thixotropic fluid rheology [14]. Norouzi et al. applied a novel Hybrid Particle Swarm Optimization-Simulated Annealing method (HPSOSA) to develop a new correlation to predict CO_2-Oil MMP in a wide range of thermodynamic and reservoir conditions [15]. Considering the large amount of real-time data generated in drilling operations every day, drilling engineering also benefits greatly from the application of ML [16]. Through testing, Muojeke S et al. found that the combination of supervised data-driven modeling and downhole parameter monitoring can effectively perform early kick detection [17]. Liang H et al. and Si Menghan et al. established a drilling overflow diagnosis model by using an optimized BP neural network [18,19]. Shi X proposed a real-time spillage judgment method based on random forest and used random forest to analyze the importance of features in the primary election [20]. Wang Z et al. selected hook load, bit pressure, torque, flow rate, penetration rate and riser pressure as the feature parameters based on the analysis of the gas surge mechanism, and tested a variety of them, and found that the ensemble learning RUSBoost algorithm model had the best effect on early monitoring of gas surge [21]. Zhang Z et al. and Zhang F et al. used the method based on feature clustering of time series data to diagnosis overflow of the drilling process [3,22]. Muojeke et al. used ANN to establish a binary classifier for the occurrence of overflow according to the downhole monitoring parameters when the overflow occurred [17]. Yin Q et al. used drilling data from surface monitoring, riser monitoring and downhole monitoring to find that the LSTM-RNN-based model has obvious advantages in overflow classification and early monitoring [23,24]. Fu J et al. and Wang Y et al., respectively, realized continuous prediction and real-time warning of overflow through LSTM-CNN [10,25]. By analyzing the relationship between different drilling activities and drilling data, Sule I et al. discussed the safety and reliability assessment of managed pressure drilling operations and finally used the Bayesian network model to analyze the well surge [11]. GAO X et al. proposed

an early overflow identification model based on slow feature analysis for the process of managed pressure drilling [26].

In intelligent drilling and completion, the fusion of data-driven and physics-based models and the interpretability of intelligent algorithms is at the forefront of research [1]. Therefore, model fusion has become the hot and difficult point of current research. Li F et al. summarized the data–physical fusion method into four modes: parallel mode, serial mode, guidance mode and feedback mode [27]. Zhang F et al. fused the transient solid transport model, the drillstring model, and the data-driven model to assess the drilling stuck risk [28]. Liu W et al. proposed a series fusion of a physical model and a data-driven model for the identification and early warning of managed pressure drilling overflow [29]. Bi J et al. combined oil and gas percolation theory with a pure data-driven percolation surrogate model, and the research results show that the fusion model has better prediction accuracy and transferability [30]. Zhu Z et al. constructed a hybrid model of ANN and particle swarm optimization to predict bottomhole pressure through the constraints of the multiphase flow mechanism, which improved the accuracy and stability of the model [31]. Ji Y calculated the bottomhole pressure through a simplified physical model and used the drilling parameters as the input of the BP neural network model optimized by a genetic algorithm to realize the accurate identification of overflow and loss accidents in the drilling process [32]. Mohammad G et al., in order to reduce the insufficient learning ability, weak generalization ability, limited ability to deal with noise and outliers, and poor interpretability of a single model, used Bayesian maximum entropy to combine multiple ML algorithms to predict water quality variables [33]. Li Y et al. proposed a drillstring fault warning framework based on multi-model fusion [34]. Erge O et al. combined the HMM (hidden Markov model) algorithm with a physics-based model to predict pressure more accurately and robustly [35]. According to the model fusion and fusion methods in the existing literature, it can be summarized as the following three forms: the fusion of mechanism or physical model and intelligent algorithm model; the fusion of multiple intelligent algorithm models; and an intelligent algorithm model based on physical constraints.

In summary, there are traditional physical model methods and data-driven artificial intelligence methods for the research of drilling overflow diagnosis. However, there has been no practical case where the two are integrated and used together. Therefore, based on the data of managed pressure drilling, this paper proposes a fusion method of the physical diagnosis model and the intelligent algorithm diagnosis model. Based on adaptive weight fusion, the overflow diagnosis model based on the comparison between the mechanism of ECD (Equivalent Circulating Density) and the equivalent density of formation pore pressure is fused with the overflow diagnosis model based on the Bayesian optimization algorithm optimized GBDT (Gradient Boosting Decision Trees) algorithm. To realize the diagnosis of overflow risk in managed pressure drilling.

2. Methods

In this paper, based on preprocessed MPD data, a physical diagnosis model is proposed to compare the pore pressure equivalent density with ECD using cubic polynomial regression. Additionally, an intelligent algorithm diagnosis model is introduced, where the GBDT algorithm is optimized through a Bayesian optimization algorithm, which is used to diagnose drilling overflow. The consistency of the diagnosis results from both models is then evaluated. When the results are consistent, they are directly output. In case of inconsistent results, a fusion diagnosis model based on an adaptive weight integration fusion method is applied for further diagnosis. The process of establishing the fusion

diagnosis model is shown in Figure 1 below. The specific details of the methods will be described in this chapter.

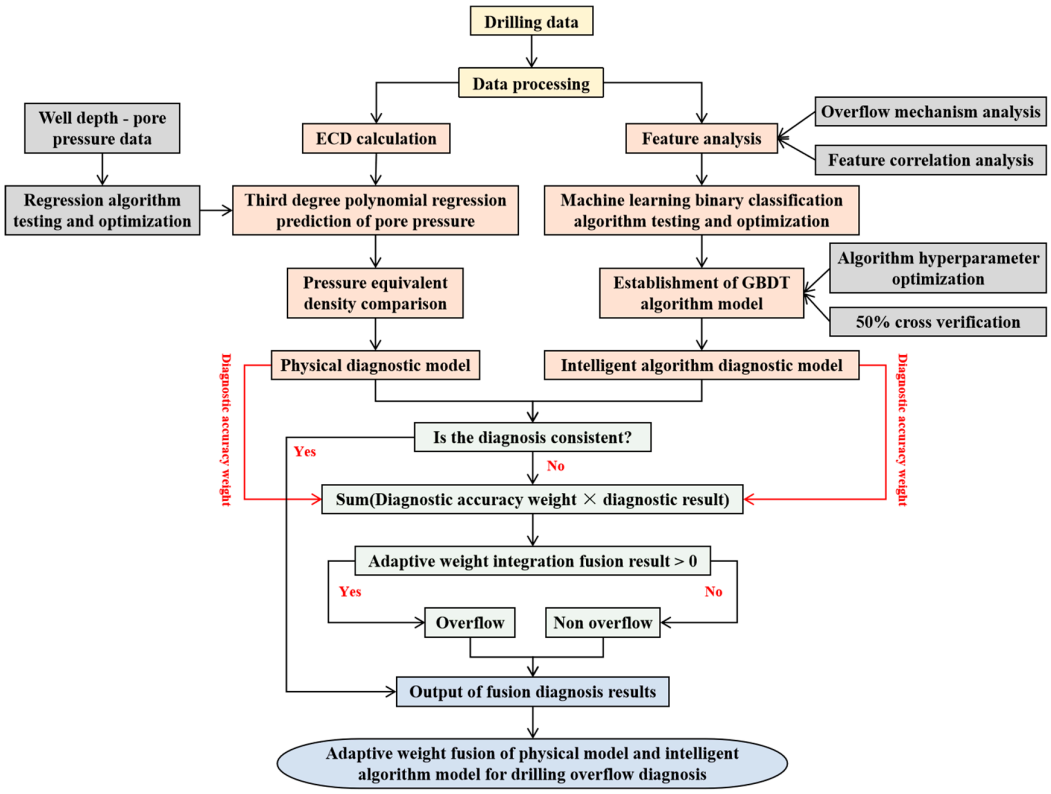


Figure 1. The overall research method and process for establishing the drilling overflow fusion diagnosis model.

All the code used in this research was implemented in Python (version 3.9.13) through Jupyter Notebook (version 6.4.12). The machine learning models, Bayesian optimization algorithms, and various evaluation methods were all implemented using classes from different sub-modules of the scikit-learn (sklearn) library. Specifically, for example, the GradientBoostingClassifier from sklearn.ensemble was used to build the GBDT model, and the BayesSearchCV from the scikit-optimize (skopt) library was used for hyperparameter optimization of the GBDT model, among other methods, which are not described in detail here. All the libraries and classes used in the methods and algorithms are open source.

2.1. Feature Analysis

2.1.1. Feature Mechanism Analysis

By referring to the description and application of features in the relevant literature, combined with the drilling overflow mechanism and the phenomena or changes of the feature parameters before and after the overflow accident, we sorted out the phenomena or changes of the related feature of the overflow mechanism, as shown in Table 1.

Table 1. Phenomena or changes of related features of overflow mechanism.

Feature Name	Feature Phenomena or Changes
Outlet drilling fluid flow rate	Increase
Total pool volume	Rise
Returned drilling fluid	With oil droplets/bubbles
Rate of penetration	Speed up or empty suddenly
Hook load(The axial force borne by the hook)	Increase

Table 1. *Cont.*

Feature Name	Feature Phenomena or Changes
Export drilling fluid density	Decrease
Hydrostatic column pressure	Decrease
Annular pressure loss	Decrease

2.1.2. Feature Correlation Analysis

The feature dimension of data will affect the prediction efficiency and accuracy of the model, as well as the generalization ability of the model prediction. Therefore, feature analysis and key feature selection are indispensable steps. Feature selection, selecting subsets from the original feature dataset, is an inclusive relationship and does not change the original feature space. From the perspective of visualizing feature analysis results, this paper avoids the uninterpretability and black-box effect caused by directly selecting algorithms and their analysis results, such as Principal Component Analysis (PCA) and Kernel Principal Component Analysis (KPCA).

Therefore, we decided to use the correlation analysis method for the analysis of algorithm features. When choosing the correlation analysis algorithm, we analyze and compare some existing algorithms to select the best one with good correlation analysis effect and fast running speed.

The computational complexity for each correlation analysis method was determined based on the following general principles:

Low computational complexity: Methods that require basic operations (e.g., simple arithmetic or sorting) and can be computed efficiently with relatively little computational resources, even for large datasets. Examples include Pearson, Spearman, and Kendall correlation methods.

Moderate computational complexity: Methods that involve slightly more complex operations, such as dealing with categorical data or computing contingency tables. The Chi-square test falls into this category.

High computational complexity: Methods that require advanced computations, such as entropy-based methods or techniques that calculate mutual dependence between variables. For instance, mutual information involves calculating probabilities and can be computationally intensive, especially for large datasets.

The correlation analysis methods tested are shown in Table 2.

Table 2. Correlation analysis methods.

Correlation Analysis Method	Scope of Applicability	Computational Complexity
Pearson	Continuous, linear data, normally distributed	low
Spearman	Linear and nonlinear data	low
Kendall	Discontinuous, non-normally distributed data, fixed sequence	low
Distance correlation	Linear and nonlinear data	higher
Chi-square test	Classification data	moderate
Mutual information	Linear and nonlinear data	high

Through comparative analysis, it can be found that Pearson is used to measure the linear correlation between two variables, and it has requirements for data distribution; Chi-square test is suitable for categorical data [36]; Kendall is usually used to sort data with specific attributes; therefore, it is not suitable. Spearman, distance correlation and mutual information are all suitable for linear and nonlinear correlation between two

variables. Spearman has smaller computational complexity and better robustness; therefore, Spearman correlation analysis method is more suitable.

Spearman correlation analysis method was used for data-driven feature analysis and optimization. The correlation analysis formula is as follows: Formula (1).

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

where ρ is Spearman correlation coefficient; d_i denotes the difference of rank values of the ith data pair. n is the total number of observed samples.

2.2. Physical Diagnosis Model

2.2.1. Regression of Pore Pressure Equivalent Density

Normally, as the well depth increases, the formation pressure and temperature also increase, resulting in an increase in the pore pressure equivalent density. Although the pore pressure equivalent density is affected by many factors, including formation pressure, temperature, rock type, rock pore structure, etc. However, in this study, starting from the perspective that the pore pressure equivalent density of adjacent wells in the same block is approximately consistent at the same well depth, we learn the relationship between the well depth and the pore pressure equivalent density of the block through the machine learning regression algorithm and then establish the regression model for the prediction of the pore pressure equivalent density.

In this study, cubic polynomial regression model is selected, which is more suitable for discrete but with a certain trend of small sample data than linear regression and other machine learning regression methods.

2.2.2. Physical Diagnosis Model of Drilling Overflow

After inputting the MPD data for the physical diagnosis model of drilling overflow, the equivalent density of pore pressure obtained from the regression of each data will be compared with the ECD in the data. When the equivalent density of formation pore pressure is greater than ECD, overflow may occur; otherwise, overflow will not occur. After the comparison results are obtained, the overflow diagnosis results of the physical model can be obtained. The model method flow is shown in Figure 2. To this end, we write the physical model of overflow diagnosis using Python to ensure it is data driven. After importing the data, we can obtain the overflow diagnosis situation of each piece of data.

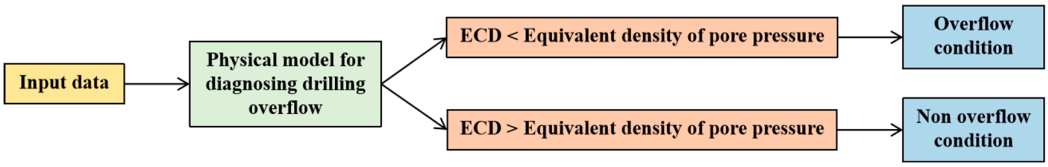


Figure 2. Diagnosis process of physical diagnosis model for drilling overflow.

2.3. Intelligent Algorithm Diagnosis Model

2.3.1. GBDT Algorithm

(1) Principle of GBDT algorithm

GBDT (Gradient Boosting Decision Trees) is an ensemble learning algorithm. Based on the idea of gradient boosting, it uses an additive model (linear combination of basis functions) and a forward step-by-step algorithm, and the decision tree (Classification and Regression Tree, CART Tree) is used as the basis function, and the accuracy of the overall model is gradually improved by iteratively building the decision tree model.

In the GBDT algorithm model, each decision tree model, that is, each weak classifier, tries to capture the information of the samples that the previous model predicted incorrectly through training, so as to gradually reduce the residual error. Finally, the prediction results of multiple decision tree models were weighted and fused to obtain the final model prediction result. The implementation principle of gbdt algorithm is shown in Figure 3. The principle of GBDT algorithm is shown in Formula (2).

$$F_m(x) = \sum_{m=1}^{M} T(x;\theta_m) \tag{2}$$

where $F_m(x)$ is the predicted value of the GBDT model with m decision trees for input sample x after m iterations. x is the feature matrix formed by the input data, and the shape of this feature matrix is the number of feature parameters multiplied by the number of sample data. $T(x;\theta_m)$ is a decision tree; x is the input sample. θ_m are the parameters of the decision tree. M is the number of decision trees.

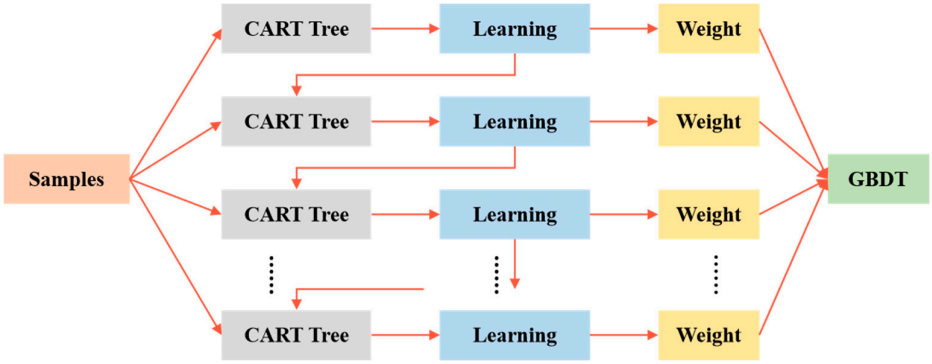


Figure 3. Implementation principle of gbdt algorithm.

(2) Hyperparameters of GBDT algorithm

The names and meanings of the commonly used hyperparameters involved in the GBDT algorithm, and their influence on the GBDT algorithm model are shown in Table 3 below.

Table 3. Hyperparameters of GBDT algorithm.

Name	Meaning	Effect
n_estimators	Number of iterations	The higher the value, the higher the complexity of the model
learning_rate	Learning rate	Smaller values require more iterations
max_depth	Tree depth	Larger values are prone to overfitting
subsample	Number of samples used to train each subtree	A larger value reduces the risk of overfitting
min_samples_split	Minimum number of samples inside a node	Node splitting threshold
min_samples_leaf	Minimum number of samples inside a leaf node	Leaf node splitting threshold

2.3.2. Model Optimization

In machine learning, the performance of a model is often affected by hyperparameters. Selecting appropriate hyperparameters can significantly improve the generalization ability and prediction accuracy of the model so that the model can better adapt to the actual application scenario. However, manual hyperparameter tuning is usually a time-consuming and tedious task, as hyperparameter selection often requires repeated evaluation and tuning of model performance.

In order to solve this problem, this study used the Bayesian optimization algorithm to optimize the hyperparameters of the GBDT model, and the five-fold cross-validation method was used to assist the performance evaluation when selecting parameters.

In hyperparameter optimization, the performance of each hyperparameter combination is measured by the average test accuracy obtained by five-fold cross-validation. Five-fold cross-validation is a robust method for evaluating model performance and selecting hyperparameters. It involves splitting the available data into five equal-sized subsets, or "folds". The model is trained on four of these subsets (the training set) and evaluated on the remaining subset (the validation set). This process is repeated five times, with each subset used as the validation set once, while the remaining four subsets are used for training. The performance of the model is then averaged over the five iterations to obtain a more reliable estimate of its generalization ability. The implementation principle of five-fold cross-validation method is shown in Figure 4 below. By setting the parameter cv = 5, the model can more accurately evaluate the performance of the model under different hyperparameter settings and help select the best combination of hyperparameters, avoiding excessive dependence on a specific training and test set.

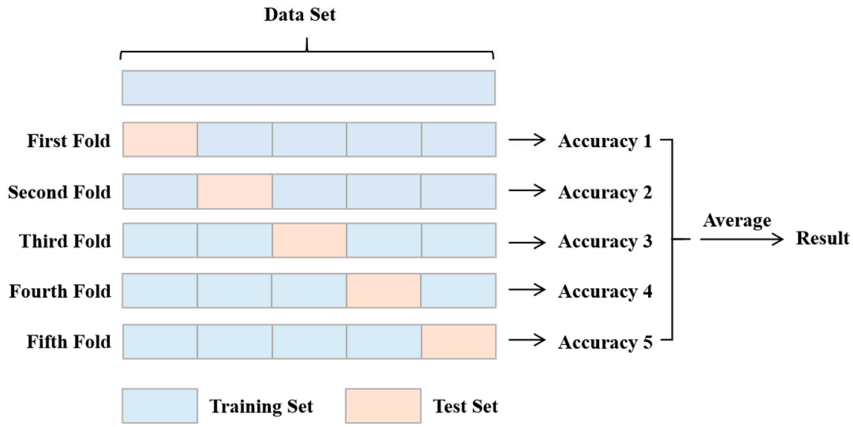


Figure 4. Implementation principle of five-fold cross-validation method.

2.4. Model Fusion Method

The model fusion method used in this study is based on the similarities and differences between the diagnosis results of the physical model and the intelligent algorithm model to analyze the diagnosis results and fuse the adaptive weights.

Firstly, the diagnosis results of the physical model and the intelligent algorithm model were compared, and the final fusion result was directly obtained when the diagnosis results of the two models were consistent. When the diagnostic results of the two models are inconsistent, from the perspective of the diagnostic accuracy of the two models for the overflow conditions in the test phase, the diagnostic results of the physical model and the diagnostic results of the intelligent algorithm model are adaptively fused, and the overflow diagnosis model based on the fusion of the physical model and the intelligent algorithm model is obtained. In this case, 0 is used as the limit for diagnosing overflow. If it is greater than 0, it indicates overflow; if it is less than 0, it indicates non-overflow. When the physical model diagnosis result is the overflow condition (1), and the intelligent algorithm model diagnosis result is the non-overflow condition (0), if the overflow diagnosis accuracy of the physical model $\times 1$ + the non-overflow diagnosis accuracy of the intelligent algorithm model $\times(-1) > 0$, the fusion diagnosis result is the overflow condition, otherwise it is the non-overflow condition. When the diagnosis result of the physical model is the non-overflow condition (0), and the diagnosis result of the intelligent algorithm model is the overflow condition (1), if the accuracy of the physical model non-overflow diagnosis

×(−1)+ the accuracy of the intelligent algorithm model overflow diagnosis ×1 > 0, the fusion result is the overflow condition, otherwise it is the non-overflow condition. The model adaptive weight fusion diagnosis method used in this study is shown in Table 4 below. The set weight is not limited to the accuracy of the model in the test phase for the overflow diagnosis, and the percentage of the two models in the fusion result can also be set artificially.

Table 4. Adaptive weight fusion drilling overflow diagnosis method.

Prediction Results		Intelligent Algorithm Model	
		0 (Non-Overflow)	1 (Overflow)
Physical model	0 (non-overflow)	0 (non-overflow)	$P_{WL0} \times (-1) + P_{ML1} \times 1 > 0$
	1 (overflow)	$P_{WL1} \times 1 + P_{ML0} \times (-1) > 0$	1 (overflow)

where P_{WL0} is the accuracy of the physical model to diagnose the non-overflow condition, and P_{WL1} is the accuracy of the physical model to diagnose the overflow condition. P_{ML0} is the accuracy of the intelligent algorithm model in diagnosing non-overflow conditions, and P_{ML1} is the accuracy of the intelligent algorithm model in diagnosing overflow conditions.

3. Case Studies

3.1. Data Selection and Preprocessing

The research data used in this study are from the middle ancient block of the Tarim oil field, which has a harsh geological environment and has the features of deep reservoir burial (5000~7000 m), high temperature (130~170 °C), high pressure (170 mpa), and low porosity and low permeability. The formation fluid medium is complex and the reservoir is highly heterogeneous. The development of fractures and holes in the target zone is easy to gush and leak, which brings great challenges to the drilling and completion process. We selected managed pressure drilling data collected from some sections of several wells with frequent overflows, in which the number of overflows in a single well was up to 11. After consulting the drilling history and complex situation in the data and coding the target features, the preliminary Excel data sheet was formed.

In order to prepare for subsequent feature analysis and model training and prediction, this study used data preprocessing methods such as target feature coding, data cleaning and under-sampling processing before model establishment.

Target feature coding processing is to add an additional column of “overflow” to the data. According to the complex working condition records in the well history data, the corresponding overflow situation of each piece of data is determined, and the 0–1 coding is adopted. The target value is 0 or 1, and “0” means no overflow and “1” means overflow.

Data cleaning processing is mainly to clean the features and abnormal data. According to the data description of well history data and the actual data situation, we initially delete the feature columns that are irrelevant to the overflow and then delete the feature columns that have not been collected and the feature columns with poor accuracy. Then, the missing or all 0 data are removed using data cleaning, and the triple σ criterion is used to detect and deal with abnormal data.

Considering the balance of the target categories, we perform data undersampling processing, remove the data samples with a relatively long period of overflow, and ensure that the ratio of overflow and non-overflow data is not less than 1:4.

Finally, through data record explanation and data preprocessing, 7388 pieces of data were obtained, in which the ratio of overflow and non-overflow data was 4491:2897, and each piece of data had 31 features. In addition, an additional 400 pieces of data were selected as the final fusion test data, and the ratio of overflow and non-overflow data was 200:200.

3.2. Main Features Affecting Overflow

In order to ensure the correlation between features and target overflow, we used all 7388 pieces of data after data preprocessing for Spearman correlation analysis and obtained the correlation diagram between features and overflow, as shown in Figure 5 below.

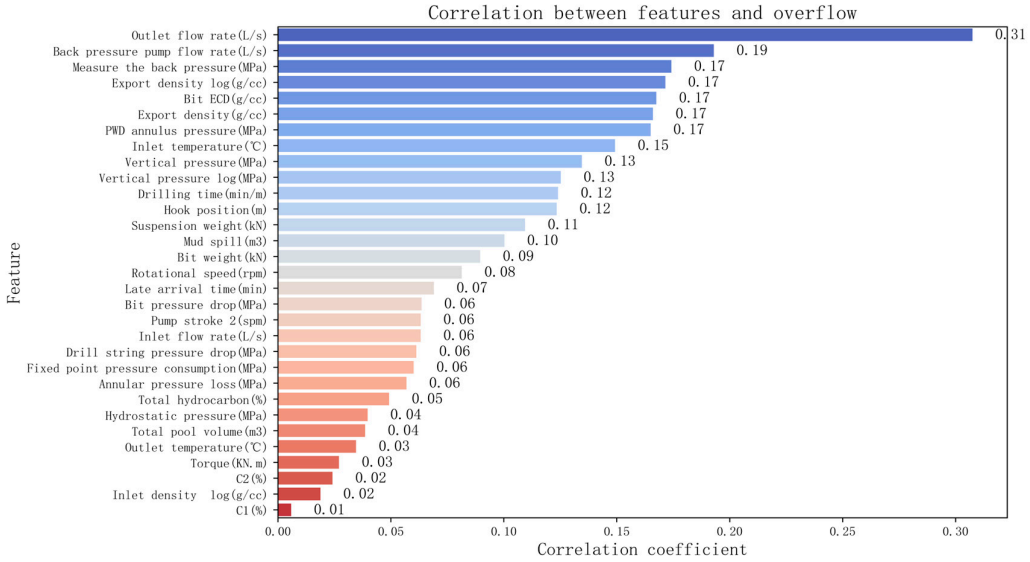


Figure 5. Correlation between features and overflow.

The correlation between 31 features and overflow can be clearly seen in Figure 5. In the Figure 5, Cool colors (blue) represent negative correlations, where the feature is inversely related to the target variable. Warm colors (red) represent positive correlations, where the feature is positively related to the target variable. The intensity of the color indicates the strength of the correlation: Darker shades represent stronger correlations (whether positive or negative). Lighter shades represent weaker correlations. In order to improve the generalization ability of the model and avoid the influence of feature redundancy on the complexity and prediction accuracy of the model, the correlation coefficient of 0.1 is taken as the limit, and the features with low correlation and duplication are discarded. Finally, through the feature analysis, we obtain the following eleven main features affecting the overflow: mud spill (m^3), suspension weight (kN), drilling time (min/m), vertical pressure (MPa), inlet temperature (°C), PWD annular pressure (MPa), export density (g/cc), bit ECD (g/cc), measured back pressure (MPa), back pressure pump flow (L/s), and outlet flow (L/s).

3.3. Physical Diagnosis Model

3.3.1. Physical Model Establishment

First, we consulted the well history data of each well and found that each well only had a very small amount (two–three pieces) of pore pressure equivalent density obtained from field tests, which could not meet the needs of physical model testing. To do this, we investigated and reviewed the well history data of six additional wells in the block, collecting a total of more than 20 pieces of well depth and pore pressure equivalent density data. Since managed pressure drilling is often used in deep well drilling, in order to ensure that the regression model can better learn the relationship between well depth and pore pressure equivalent density, and to ensure that the regression model can adapt to the data segment used by the fusion model, we retained the data in the vicinity of 4000–6500 m after screening. Through the cubic polynomial regression algorithm, the equivalent density function formula of well depth and pore pressure, as shown in Equation (3), is obtained,

and the fitting of the data is shown in Figure 6 below. Therefore, we established a cubic polynomial regression model for predicting the equivalent density of pore pressure.

$$y = 0.4941 + 2.9747 \times 10^{-4}x - 4.2514 \times 10^{-8}x^2 + 2.1443 \times 10^{-12}x^3 \quad (3)$$

where x is the well depth, m, and y is the pore pressure equivalent density, g/cm^3.

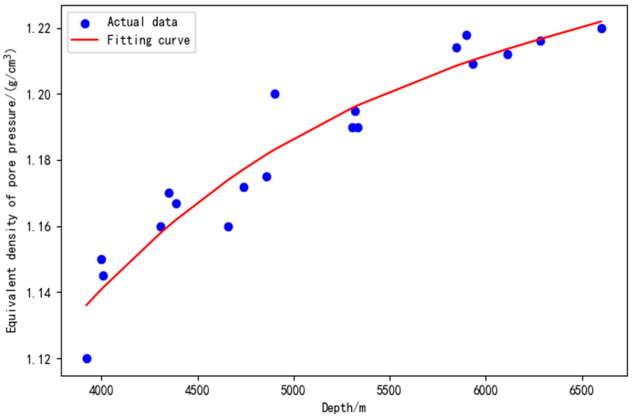


Figure 6. Fitting of borehole depth–pore pressure equivalent density by 3rd-degree polynomial regression.

By invoking the cubic polynomial regression model of pore pressure equivalent density prediction, we predicted the pore pressure equivalent density of 7388 pieces of data trained and tested using the fusion model and obtained the test samples and pore pressure equivalent density prediction results, as shown in Figure 7 below, which were used for the next step of testing and evaluation of the physical diagnosis model.

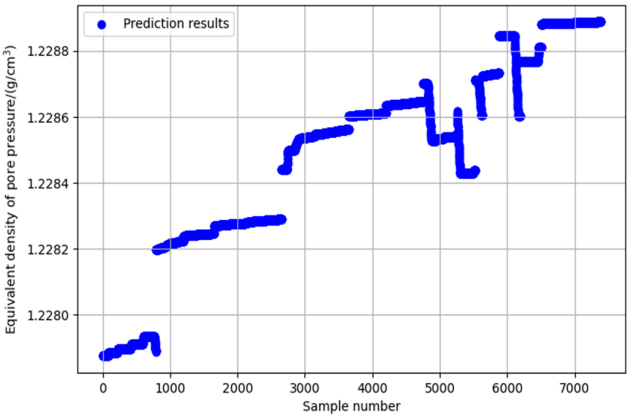


Figure 7. Prediction results of pore pressure equivalent density in test set.

In the 7388 pieces of data from the fusion model training and testing, the pore pressure equivalent density predicted by regression is compared with the bit ECD (g/cc) calculated by field experts in the data through the physical diagnosis model of overflow as shown in the formula, and the diagnosis result of the physical model of drilling overflow corresponding to each sample can be obtained.

$$0.4941 + 2.9747 \times 10^{-4}x - 4.2514 \times 10^{-8}x^2 + 2.1443 \times 10^{-12}x^3 > ECD \quad (4)$$

Through comparison with the actual overflow situation corresponding to the data, it is concluded that the diagnostic accuracy of the physical diagnosis model of drilling

overflow is 62.11%, and the comparison between the overflow diagnosis results of the physical diagnosis model and the actual overflow results is shown in Figure 8.

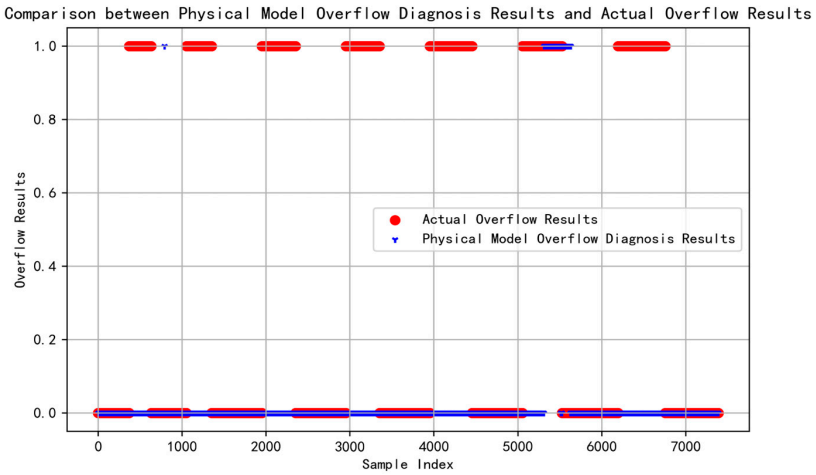


Figure 8. Comparison between the overflow diagnosis results of the physical model of the test set and the actual overflow results.

Subsequently, the test effect of the physical model was evaluated, and the evaluation report of the model was obtained, as shown in Table 5. It can be found that precision, recall, and f1-score are the same as accuracy, and the diagnostic effect of 0 (non-overflow) is average, while that of 1 (overflow) is poor. In addition, the confusion matrix of the model is shown in Figure 9 and the ROC curve is shown in Figure 10. It can be seen that there are only 2697 non-overflow samples and 102 overflow samples with diagnostic errors in the model for overflow diagnosis, and the diagnostic accuracy of the model for non-overflow conditions is 66.23% and 33.77%, respectively. As the weight of the subsequent fusion model, the physical model is used to diagnose the overflow and non-overflow conditions. The ROC curve is only slightly higher than the default random classifier, and the AUC of 0.523 is only slightly greater than 0.5, indicating that the physical model is only moderately effective for overflow diagnosis.

Table 5. Physical model evaluation report.

	Precision	Recall	f1-Score	Support
0	0.6194	0.9773	0.7582	4491
1	0.6623	0.0690	0.1250	2897
accuracy			0.6211	7388
macro avg	0.6408	0.5232	0.4416	7388
weighted avg	0.6362	0.6211	0.5099	7388

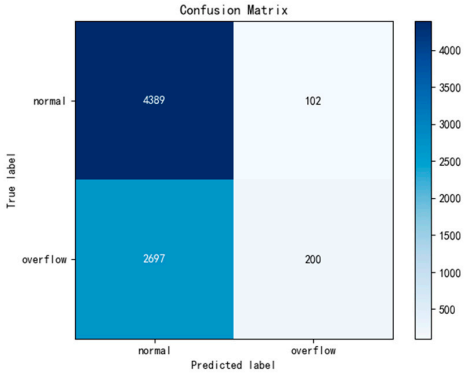


Figure 9. Confusion matrix of physical model.

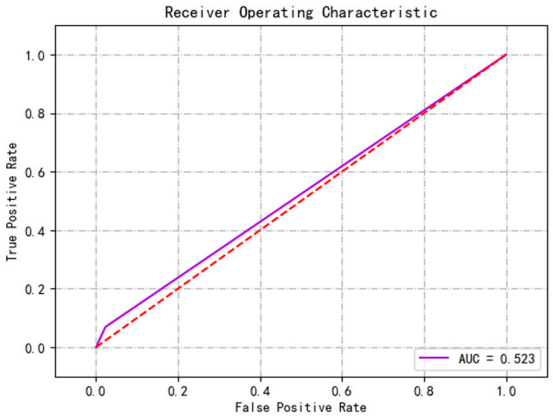


Figure 10. ROC curve of physical model.

3.3.2. Physical Model Diagnosis Results

Through the established intelligent diagnosis model of drilling overflow, 400 pieces of data in the verification set were also diagnosed, and the prediction results of pore pressure equivalent density, as shown in Figure 11, were obtained.

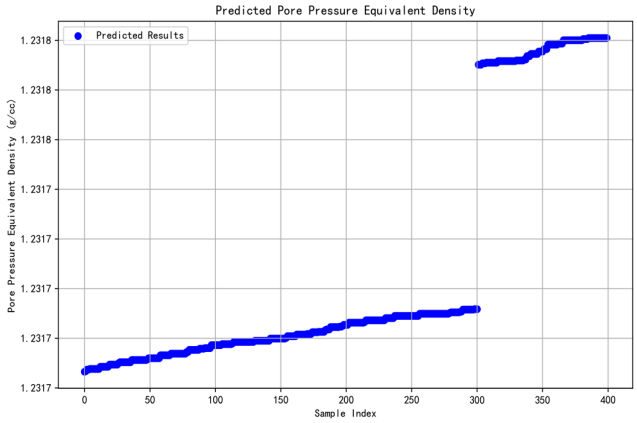


Figure 11. Prediction of equivalent density of pore pressure in verification set.

Then, the pore pressure equivalent density obtained from the last 400 validation sets was compared with the ECD in the data to obtain the physical model overflow diagnosis results for each sample. Then, the overflow diagnosis of the physical model and the comparison of the results can be obtained, and the sample overflow diagnosis results, as shown in Figure 12 below, can be compared with the actual results.

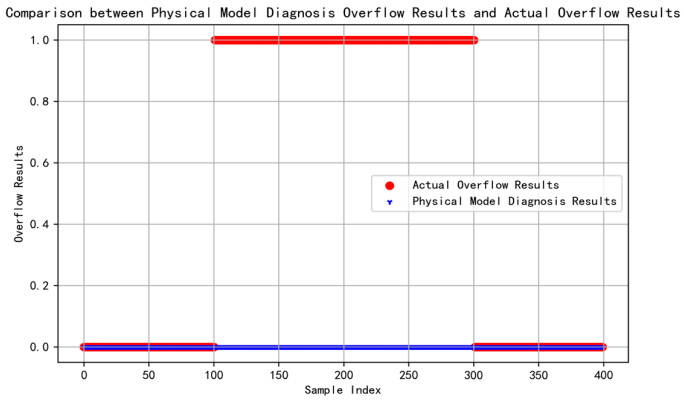


Figure 12. Comparison between the overflow diagnosis results of the verification set physical model and the actual results.

As can be seen from Figure 12, in the verification set, the diagnosis results of the physical model are all non-overflow conditions, and the diagnosis accuracy rate of the model is 50% after further evaluation, with poor effect.

3.4. Intelligent Algorithm Diagnosis Model

3.4.1. Bayesian Optimization GBDT Model Establishment

In this study, the GBDT model was established by using sklearn.ensemble library in Python programming language. The feature matrix and target of the model were selected by feature names. In total, 70% of the processed data were taken as training data, 30% as test data, and random seed random_state = 42 was set.

When setting the initial search space of Bayesian optimization, we combine the characteristics and prior knowledge of the GBDT algorithm, select according to the default value of each hyperparameter and the actual situation of the problem, and give a reasonable prior distribution for each hyperparameter magnitude, that is, set the sampling method corresponding to each hyperparameter, and determine the hyperparameter of the prior distribution. This prior knowledge can help the Bayesian optimization algorithm to better guide the search for optimal hyperparameters.

We define search_space as a dictionary of Bayes algorithm hyperparameter search space. Each key in the dictionary represents a hyperparameter name, and the corresponding value includes the range of possible values of the hyperparameter and the sampling method. The dictionary content is listed in Table 6.

Table 6. Bayesian optimization GBDT hyperparameter dictionary.

Name	Search Space	Sampling Mode	Default Value	Optimized Value
n_estimators	(50, 500)	Arithmetic sequence	100	410
learning_rate	(0.001, 0.5)	Log-uniform	0.1	0.095
max_depth	(3, 10)	Arithmetic sequence	3	10
subsample	(0.5, 1.0)	Uniform	1	0.69
min_samples_split	(2, 20)	Arithmetic sequence	2	4
min_samples_leaf	(1, 20)	Arithmetic sequence	1	5

Next, using the defined search space, we create a BayesSearchCV object opt. Specify the following parameters, when creating the object model: to optimize the model GradientBoostingClassifier (); search_space: hyperparameter search space, that is, the search space of the six hyperparameters defined above; n_iter: number of iterations, that is, how many hyperparameter combinations the algorithm tries, set to 30; cv: cross verification fold, set to five; scoring: evaluation index using accuracy; random_state: random seed, set to 42.

The fit() method is implemented to optimize the hyperparameter combination and applied to the trained model for accuracy test to find a better hyperparameter combination. Finally, we obtained the optimized hyperparameter values as shown in Table 6 above, and based on this, established the intelligent diagnosis model of drilling overflow optimized GBDT model by Bayesian optimization algorithm.

In order to observe the combination of hyperparameters and the corresponding accuracy rate, and to judge whether there is a local optimal situation in the iterative optimization process, we draw the double *Y*-axis diagram of the values of each hyperparameter in 30 iterations and the average test accuracy rate of the 50-fold cross-verification, as shown in Figure 13a–d. The values of max_depth, min_samples_split, and min_samples_leaf, as well as learning_rate and subsample, are in the same graph.

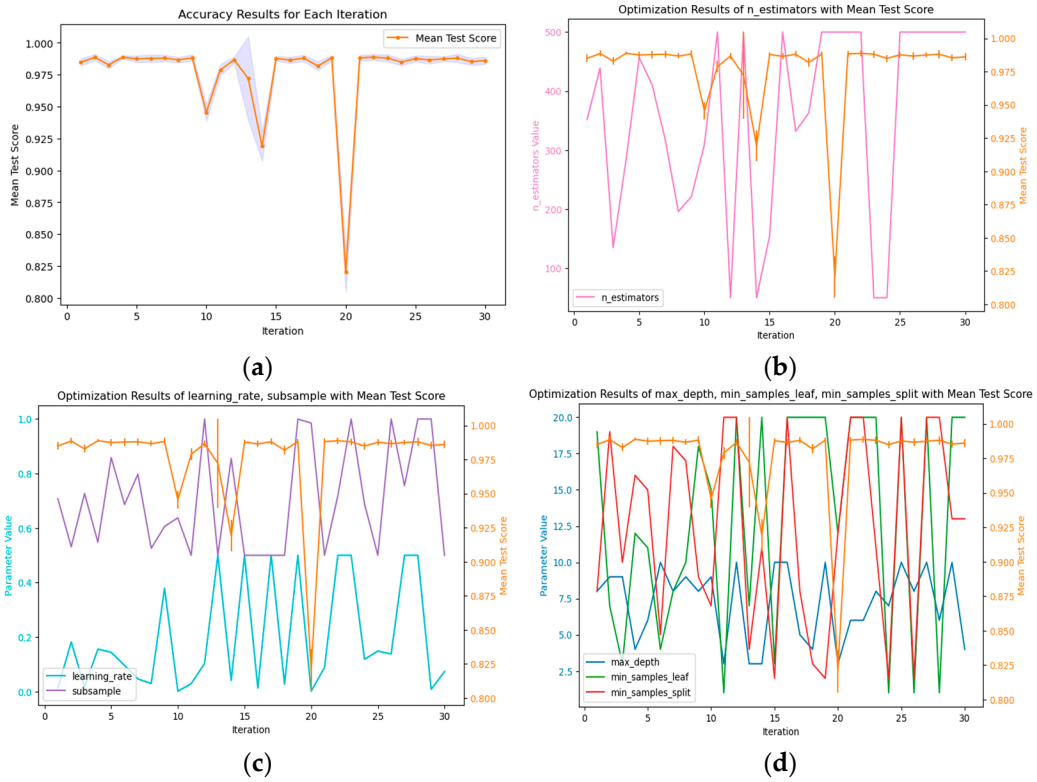


Figure 13. (**a**) Diagnostic accuracy of 30 iterations. (**b**) Values of n_estimators for 30 iterations. (**c**) Values of learning_rate and subsample were iterated for 30 times. (**d**) Values of max_depth, min_samples_split and min_samples_leaf.

It can be found that the selection of hyperparameters comes from the fourth iteration optimization, at which time the model achieves the highest average test accuracy of 98.88%. After keeping six significant digits, we found that the average test accuracy of the 4th and 22nd iteration optimization was 0.988784, and the model chose the highest result in the first iteration. In addition, from the general trend, the average test accuracy rate after hyperparameter optimization does not fluctuate significantly. Among the 30 iterations, except for the 20th iteration, the average test accuracy rate of the hyperparameter combination for 27 iterations is above 97%, and the selection of the best hyperparameter does not appear to be locally optimal in the optimization process. It can be concluded that the selection of GBDT model hyperparameters by the Bayesian optimization algorithm is relatively robust and robust. Further, the overflow diagnosis results of the Bayesian optimized GBDT model test set are shown in Figure 14.

Then, the model was evaluated and the evaluation report of the model was obtained, as shown in Table 7. It can be found that precision, recall and f1-score have high values as well as accuracy, with good effects in the classification of 0 (non-overflow) and 1 (overflow). The diagnostic accuracy of the model is 98.95% for non-overflow working conditions and 98.75% for overflow working conditions, which is used as the weight of the subsequent fusion model for the intelligent algorithm model to diagnose overflow working conditions and non-overflow working conditions.

Combined with the confusion matrix of the model as shown in Figure 15 and the ROC curve as shown in Figure 16, the GBDT algorithm improved the accuracy rate of overflow diagnosis by 1.05% compared with that of before the optimization. In the test set, the error diagnosis of overflow condition was reduced from 22 samples to 11 samples. For the error diagnosis of non-overflow condition, the number of samples was reduced from 26 to 14. Overall, the number of samples with 23 false diagnoses was reduced by nearly half. The

ROC curve approximates the isosceles right triangle, and the AUC also achieves a good effect of 0.9888. It can be seen that the model hyperparameter optimization is necessary. The good results of the model after hyperparameter optimization are proved again.

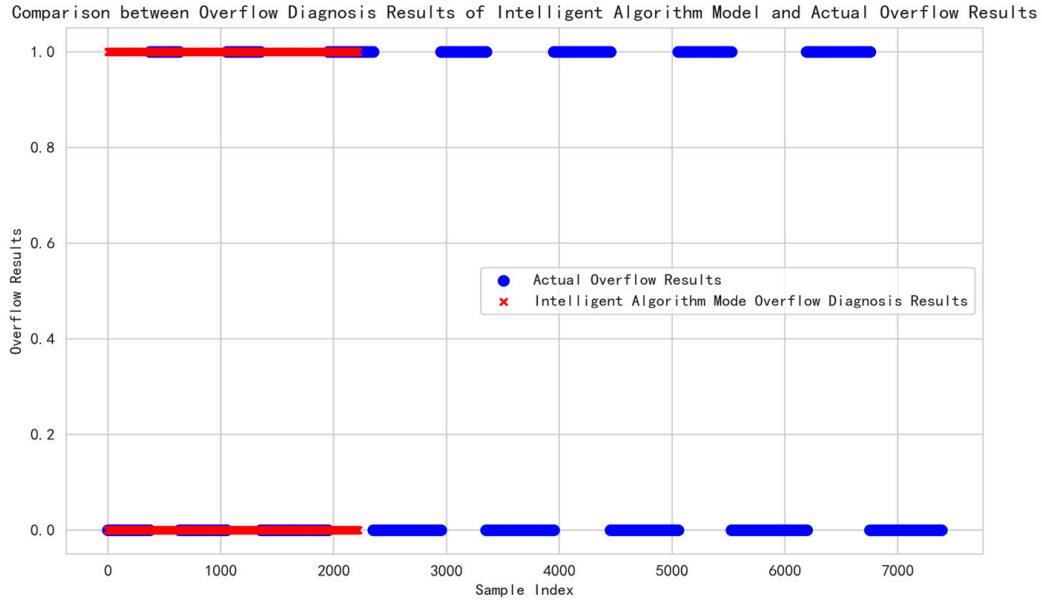


Figure 14. Comparison between the overflow diagnosis results of the intelligent algorithm model of the verification set and the actual results.

Table 7. Evaluation report of GBDT model test set for Bayesian optimization.

	Precision	Recall	f1-Score	Support
0	0.9917	0.9917	0.9917	1333
1	0.9876	0.9876	0.9876	884
accuracy			0.9901	2217
macro avg	0.9897	0.9897	0.9897	2217
weighted avg	0.9901	0.9901	0.9901	2217

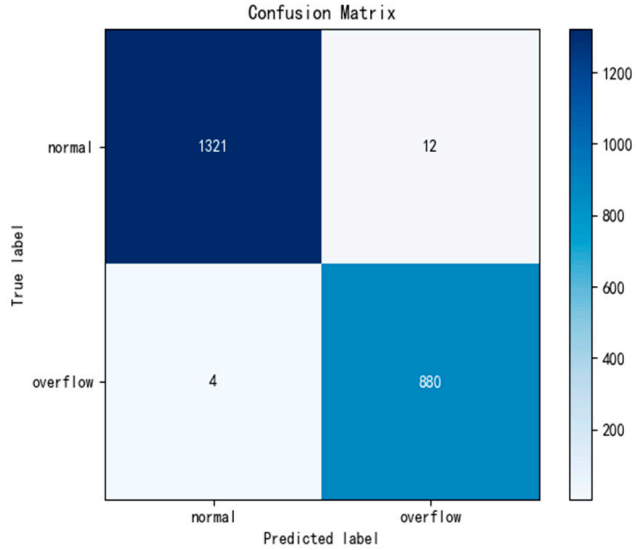


Figure 15. Confusion matrix of GBDT model test set based on Bayesian optimization.

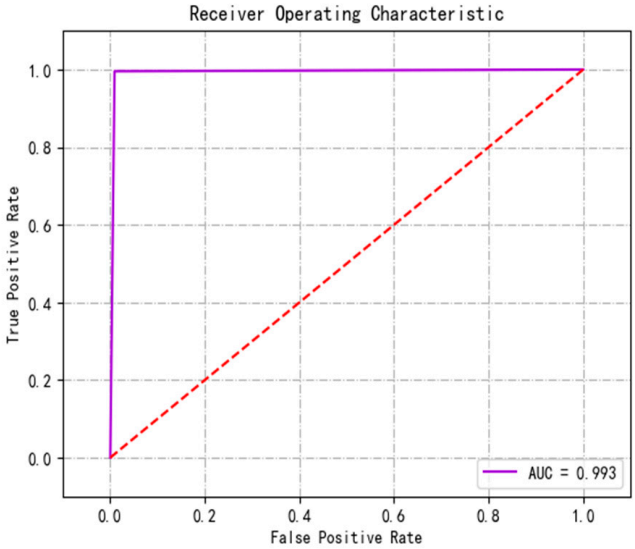


Figure 16. ROC curve of GBDT model test set for Bayesian optimization.

3.4.2. Intelligent Algorithm Model Diagnosis Results

The drilling overflow intelligent diagnosis model of the GBDT model was optimized using the established Bayesian optimization algorithm, the overflow situation of 400 pieces of data in the validation set was diagnosed, and the diagnostic results were compared with the actual results. Finally, the sample overflow diagnosis results were compared with the actual results, as shown in Figure 17.

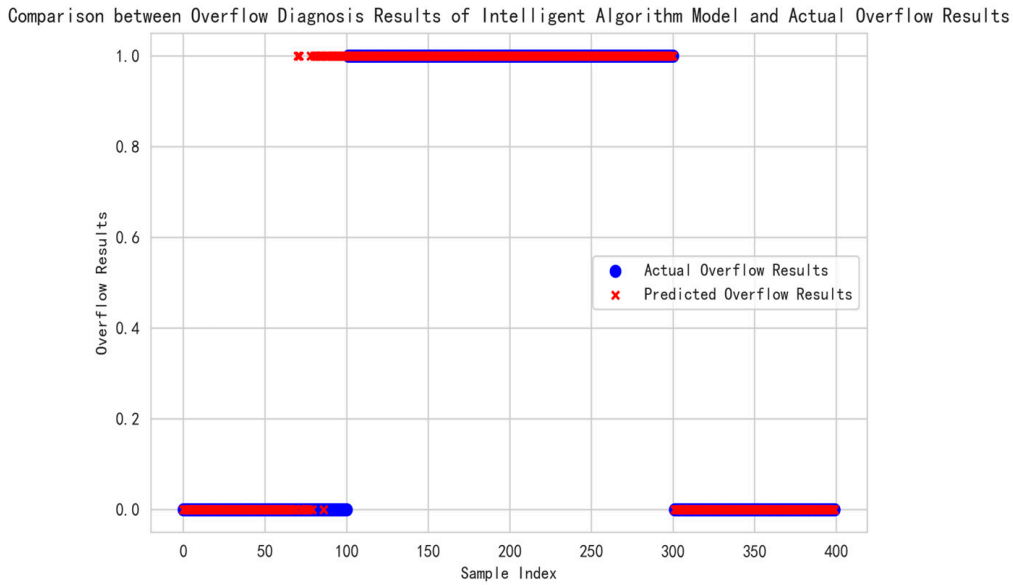


Figure 17. Comparison between the overflow diagnosis results of the intelligent algorithm model of the verification set and the actual results.

As can be seen from Figure 17, on the verification set, the intelligent algorithm model diagnosis results are relatively good. The diagnostic results of the model were further evaluated, and the diagnostic accuracy of the model was 94.25%. Using the confusion matrix in Figure 18, it can be found that 23 overflow samples were wrongly diagnosed, and the wrong diagnosis samples were relatively concentrated. As can be seen from Figure 19, the ROC curve is also good, the error rate of horizontal coordinate is small, and the AUC value is 0.9425.

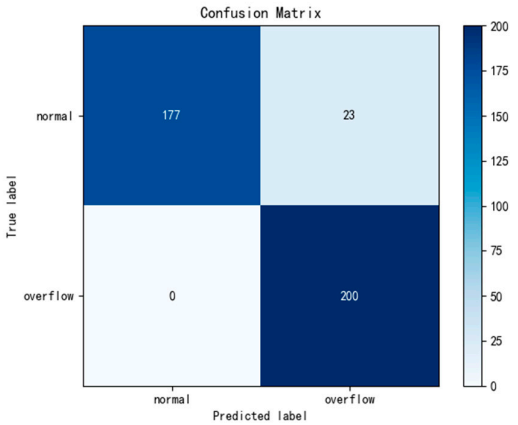


Figure 18. Confusion matrix of GBDT model validation set for Bayesian optimization.

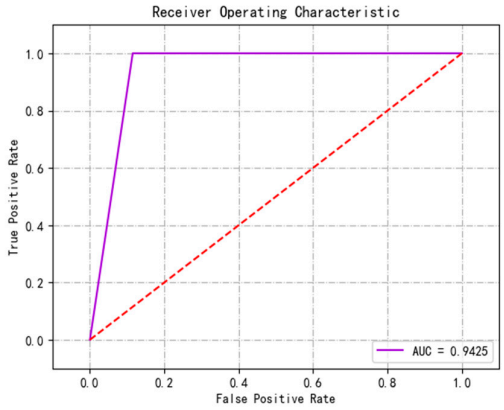


Figure 19. ROC curve of GBDT model test set with Bayesian optimization.

3.5. Fusion Model Overflow Diagnosis Results

Based on the diagnosis results of the physical model in Section 3.3.2 and the intelligent algorithm model in Section 3.4.2 above, we further analyze and integrate the diagnosis results of the two models, and establish a drilling overflow diagnosis model that integrates the physical model and the intelligent algorithm model.

In the actual drilling process, the field drilling personnel prefer to trust the diagnosis results of the physical model because it is more direct and has relatively strong mechanical rationality. In addition, combined with the physical model test in Section 3.3, it has a certain effect on the diagnosis of non-overflow conditions. Therefore, when the diagnostic results of the intelligent algorithm model and the physical model are inconsistent, an additional weight of 10% is assigned to the diagnostic accuracy of the physical model on the basis of the comprehensive weight optimization fusion mentioned in Section 2.4.

First, the accuracy of the physical model and the intelligent algorithm model on the test set is loaded by code adaptively, as follows: The accuracy of the physical model in the diagnosis of non-overflow conditions and overflow conditions is wl_accuracy_0 = 0.6625 and wl_accuracy_1 = 0.3377, respectively. The accuracy of the intelligent algorithm model in the diagnosis of non-overflow conditions and overflow conditions is ml_accuracy_0 = 0.9917 and ml_accuracy_1 = 0.9876, respectively. Then, we can obtain the model fusion diagnosis method, as shown in Table 8 below. When the diagnosis results of the two models are consistent, the results can be obtained directly. If they are inconsistent, the two formulas in Table 8 are used for diagnosis. If they are greater than 0, the diagnosis result is overflow. On the contrary, the diagnosis result is non-overflow. In summary, the overflow diagnosis result of the fusion model can be obtained.

Table 8. Adaptive weight fusion drilling overflow diagnosis model.

Prediction Results		Intelligent Algorithm Model	
		0	**1**
Physical model	0	0	$(0.6625 + 0.1) \times (-1) + (0.9917–0.1) \times 1 > 0$
	1	$0.3377 \times 1 + 0.9876 \times (-1) > 0$	1

The overflow diagnosis results of the fusion model can be compared with the actual results by testing 400 pieces of data from the validation set.

As can be seen from Figure 20, in the verification set, the fusion model diagnosis results are the same as the intelligent algorithm model, and relatively good diagnosis results are obtained.

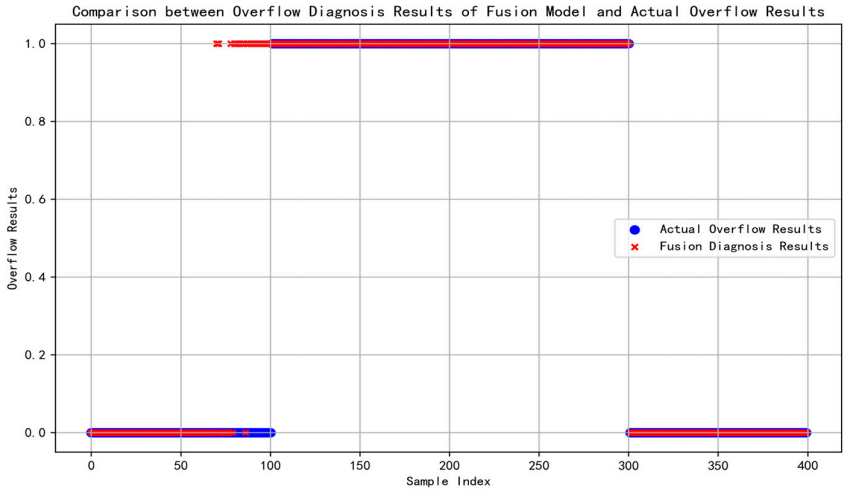


Figure 20. Comparison between the overflow diagnosis results of the validation set fusion model and the actual results.

4. Discussion

4.1. Influence and Optimization of Algorithm Model

4.1.1. Optimization of Regression Fitting Algorithm Model

Before selecting the cubic polynomial regression algorithm model, we used the data of well depth and pore pressure equivalent density to test the commonly used regression algorithm model to select the more suitable model. To this end, we screened linear regression, support vector machine regression, decision tree regression, random forest regression and polynomial regression methods, respectively, and obtained the regression fitting of well depth and pore pressure equivalent density, as shown in Figure 21 below.

As can be seen from Figure 21, support vector machine regression, decision tree regression and random forest regression are obviously not applicable, and linear regression cannot learn the relationship between well depth and pore pressure equivalent density robustness. Quadratic polynomial regression, which tends to flatline around 6000 m and approaches the apex of the parabola at 6500 m, also does not fit. The fitting curve of quartic polynomial regression is too rigid, and the effect in subsequent data pairs is not good, and there is a downward trend. Therefore, cubic polynomial regression is chosen as the method to predict the equivalent density of pore pressure by well depth.

In practice, formation pressure prediction may need to be combined with more logging data of adjacent wells in the same block for regression prediction, so as to obtain more accurate formation pressure prediction results in drilling or belt drilling, and thus improve the accuracy of the physical model's overflow diagnosis.

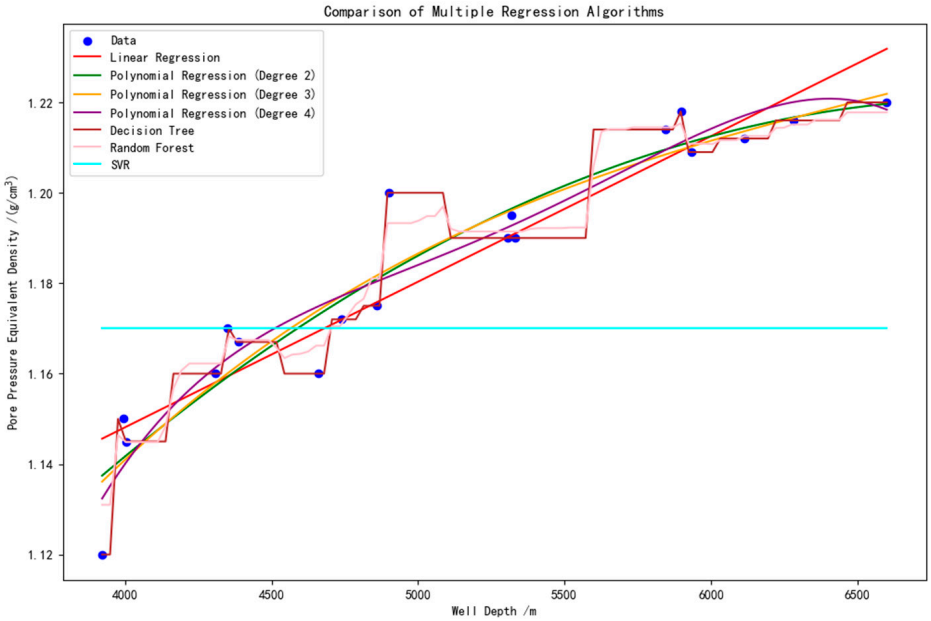


Figure 21. Data fitting of seven regression algorithms.

It is worth noting that the diagnostic results of the physical model are entirely based on numerical comparisons. Although we predicted the pore pressure equivalent density, it is still not an exact actual result. Moreover, in the case of a narrow safe density window, the ECD is very close to the pore pressure equivalent density. This leads to a relatively low accuracy in the physical model's diagnostics. Therefore, this aspect can be further explored in the future.

4.1.2. Optimization of Classification Diagnosis Algorithm Model

Before selecting the GBDT algorithm model, we made use of the original training and test data, followed the same data preprocessing, data partitioning and selection features as in the case study, and conducted the common machine learning algorithm model for binary classification problems. For example, naive Bayes, KNN, logistic regression, decision tree, neural network, SVM, AdaBoost, GBDT, etc., were tested for prediction accuracy, and the results shown in Figure 22 below were obtained.

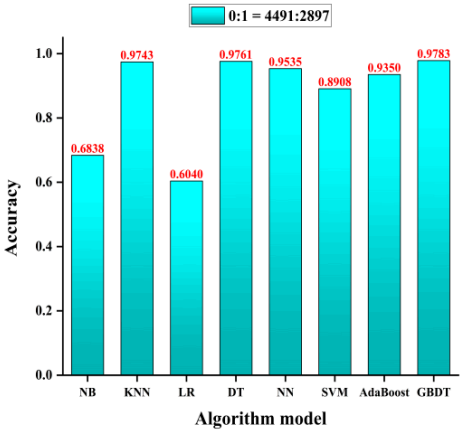


Figure 22. Diagnosis accuracy of overflow conditions of eight algorithm models.

From the comparison of the accuracy of various algorithm models in Figure 22, it can be found that the accuracy of KNN, DT, NN and GBDT algorithm models is higher, all exceeding 95%. As can be seen from Figure 23, the AUC value of the KNN, DT, and

GBDT algorithm models also exceeds 95%, followed by the AUC value of the NN algorithm model. The effect of the GBDT algorithm model based on the decision tree is slightly better than that of the decision tree algorithm model, and it is also better than AdaBoost, which is also an integrated learning algorithm.

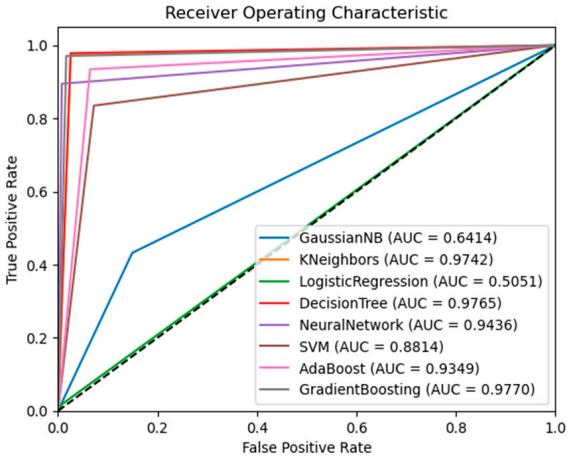


Figure 23. Diagnosis ROC curve of overflow conditions of eight algorithm models.

In view of the high prediction accuracy of the four algorithm models, the accuracy of the algorithm model can not be further measured. Therefore, we further evaluated the four models and obtained the confusion matrix of each algorithm model, as shown in Figure 24a–d below, as well as the precision, recall and f1-score of diagnosing normal and overflow conditions, as shown in Figure 25a,b below.

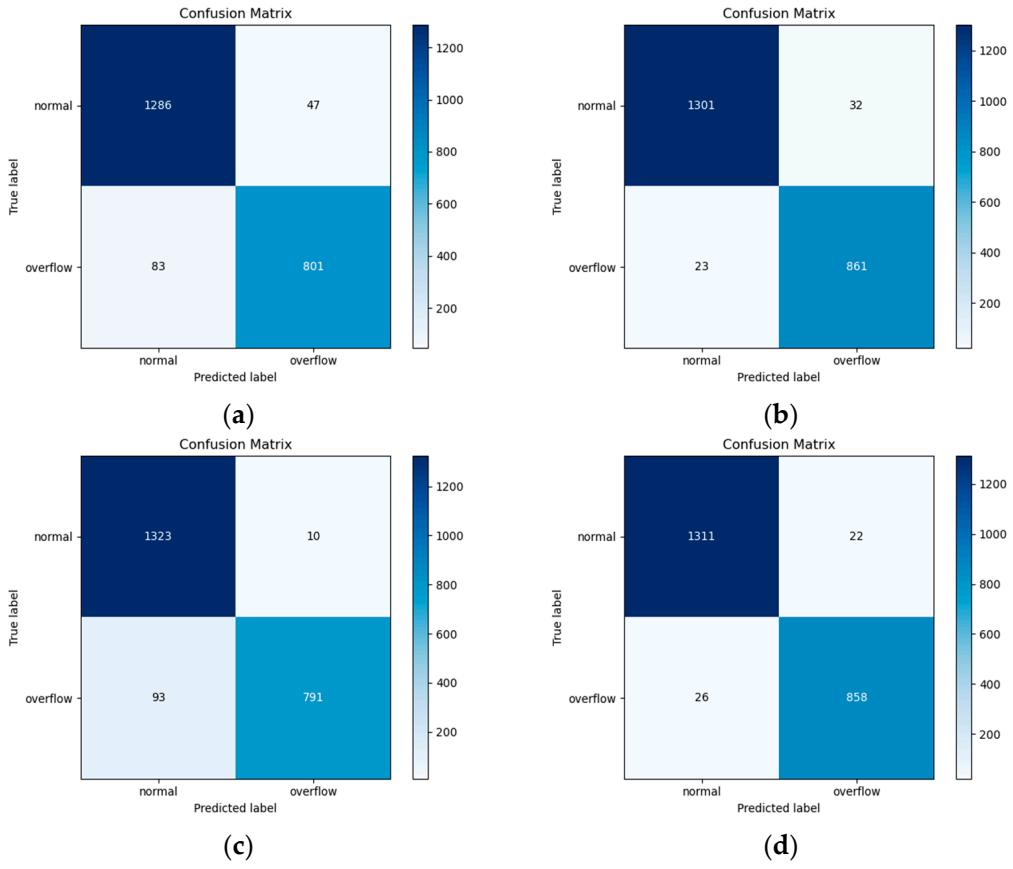
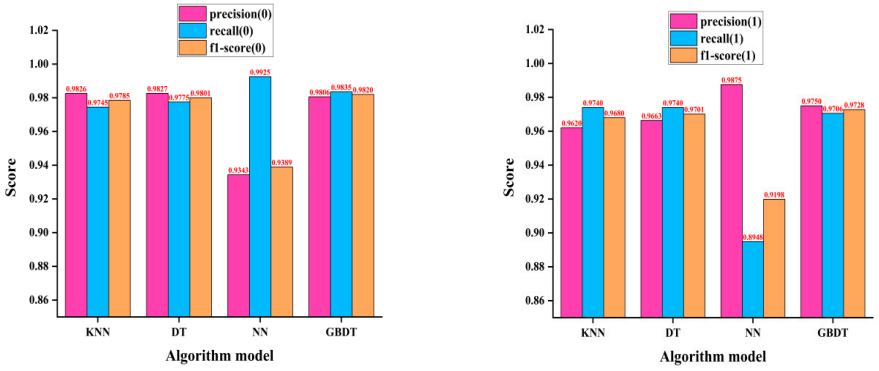


Figure 24. (**a**) KNN. (**b**) DT. (**c**) NN. (**d**) GBDT.



(**a**) (**b**)

Figure 25. (**a**) Diagnosis scores of the four algorithm models for non-overflow conditions. (**b**) Overflow condition diagnosis scores of four algorithm models.

Through a comparison of the confusion matrix, it can be seen that the number of error diagnosis samples of the KNN algorithm model is 130, the DT algorithm model is 55, the NN algorithm model is 103, and the GBDT algorithm model is 48. It can be seen that the overall effect corresponds to the accuracy rate. In the diagnosis of overflow conditions, the GBDT algorithm model has the best effect and the smallest number of wrong diagnosis samples.

According to Figure 25a,b, compared with KNN, DT and NN, the GBDT algorithm model has better and more stable results in the precision, recall and f1-score evaluation criteria.

Therefore, through the analysis of a variety of evaluation criteria, we finally selected the GBDT algorithm to build an intelligent algorithm model for the diagnosis of overflow conditions.

4.2. Analysis of Model Fusion Method

In this study, the physical model of overflow diagnosis is based on the qualitative diagnosis results of formation pressure and ECD comparison. The overflow diagnosis of the intelligent algorithm model is also the qualitative diagnosis result of the model binary classification output. Aiming at the fusion of the physical model and intelligent algorithm model, when the diagnosis results of the physical model and the intelligent algorithm model are inconsistent, we adopt the comprehensive weight optimization fusion method. The integration of the physical model and the intelligent algorithm model can effectively improve the overall machine rationality and accuracy of the model on the basis of the two separate models. Compared with the traditional physical model overflow diagnosis, the fusion model in this paper can use the collected data to directly diagnose in2332 close to real time, avoiding lag. Compared with the overflow diagnosis of the existing intelligent algorithm model, the fusion model adopts the field calculation data ECD and the model to predict the pore pressure equivalent density of the data on the basis of the intelligent algorithm model. In the diagnosis of overflow, it is of great help to engineers to regulate parameters and analyze and solve the overflow risk problem. In addition, through the test, the diagnostic accuracy of the fusion model for overflow risk is more than 98%, and the application accuracy is also up to 93.5%, which is relatively good compared with other existing method models.

Although in the modeling of intelligent algorithm model, binary classification modeling can be changed to quantitative probabilistic modeling, and the judgment basis of binary classification also comes from probabilistic prediction results. However, the probabilistic prediction model based on binary target data inevitably has some problems such as poor interpretability and unconvincing. Secondly, compared with the qualitative diagnosis

results of the physical model, the fusion of qualitative and quantitative methods has certain irrationality in the fusion of different methods and results of different properties. And in the actual situation, for the overflow risk, the initial attention is more qualitative whether overflow. Therefore, the research of the model fusion method on the drilling overflow problem needs further discussion.

4.3. Research Significance and Application Scenarios

In this study, the physical model, intelligent algorithm model and fusion model, all of which are applicable for diagnosing overflow risk in oil- and gas-managed pressure drilling operations, aim to improve the efficiency and safety of managed pressure drilling and reduce NPT (non-productive time) and economic losses. The designed code program can be deployed on the field computer or can be packaged and integrated into the existing intelligent drilling software for real-time monitoring and diagnosis. In particular, when the MPD equipment can not effectively ensure the safety of drilling, it can play a role.

Due to the MPD data used in this study, the research methods are primarily applicable to the field of MPD operation. If MPD is not used in the field, that is, if fine MPD data are not available, mud logging data can be used, but regression prediction of ECD and re-characterization are required. Therefore, the conventional mud logging operation can be further discussed in the future.

5. Conclusions

Starting with drilling overflow diagnosis and model fusion methods, this study adopts pressure control data collected from field-managed pressure drilling, and on the basis of data feature mechanism analysis and feature Spearman correlation analysis feature selection, after data processing, the drilling overflow physical diagnosis model based on cubic polynomial regression and the drilling overflow intelligent algorithm diagnosis model based on the GBDT algorithm optimized using the Bayesian optimization algorithm are established, respectively. The drilling overflow diagnosis model based on a physical model and an intelligent algorithm is established using the comprehensive weight optimization fusion method, and the drilling overflow risk diagnosis is realized. Based on this study, the following conclusions could be obtained:

(1) The cubic polynomial algorithm model has a good application in the prediction of formation pore pressure equivalent density and provides a method for the physical model of overflow diagnosis by comparing ECD with the formation pore pressure equivalent density. By optimizing the hyperparameters of the algorithm model, the prediction accuracy of the algorithm model can be improved. The GBDT algorithm model optimized using the Bayesian optimization algorithm has a good effect on overflow risk diagnosis, achieving 98.88% accuracy in the test set and 93.5% accuracy in the verification set.

(2) An overflow risk diagnosis method based on the fusion of a physical model and an intelligent algorithm model is proposed. Overflow diagnosis is carried out by the physical model and the intelligent algorithm model, respectively, and then the fusion model constructed by the adaptive weight fusion method is used to obtain the model fusion diagnosis results. The proposed new method of fusion model construction can improve the overall mechanical rationality and accuracy of the model, and the application of model fusion in drilling overflow was made and given a reference.

(3) The new method used in this study to diagnose drilling overflow risk is of great significance for promoting the safety of the drilling system and the automation of the drilling process.

Author Contributions: Conceptualization, H.H. and C.L. (Chaowei Li); Methodology, Y.S., S.D. and C.L. (Chengguo Liu); Software, Y.S.; Validation, S.D.; Formal analysis, H.H. and C.L. (Chaowei Li); Investigation, Q.L. and C.L. (Chengguo Liu); Resources, R.L.; Data curation, H.H. and R.L.; Writing—original draft, Y.S.; Writing—review & editing, Y.S.; Visualization, Q.L.; Funding acquisition, S.D. and C.L. (Chaowei Li). All authors have read and agreed to the published version of the manuscript.

Funding: The authors are grateful for the support of the National Key R&D Program of China (2022YFC2806403) and the CNPC-CZU Innovation Alliance (2021DQ06).

Data Availability Statement: The datasets presented in this article are not readily available because project restrictions, the data is confidential and cannot be made public.

Conflicts of Interest: Author Rentong Liu was employed by the company CNPC Engineering Technology R&D Co., Ltd. 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 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

1. Li, G.; Song, X.; Tian, S.; Zhu, Z. Intelligent drilling and completion: A review. *Engineering* **2022**, *18*, 33–48. [CrossRef]
2. Abughaban, M.; Alshaarawi, A.; Meng, C.; Ji, G.; Guo, W. Optimization of drilling performance based on an intelligent drilling advisory system. In Proceedings of the International Petroleum Technology Conference, Beijing, China, 26–28 March 2019.
3. Zhang, Z.; Lai, X.; Wu, M.; Chen, L.; Lu, C.; Du, S. Fault diagnosis based on feature clustering of time series data for loss and kick of drilling process. *J. Process Control* **2021**, *102*, 24–33. [CrossRef]
4. Zhang, Z.; Zhao, Y.; Wang, J.; Wang, G.; Zhong, L.; Lu, J.; Yuan, J. A real time monitoring method for well-kick and lost circulation based on distributed fiber optic temperature measurement. *Geoenergy Sci. Eng.* **2023**, *229*, 212087. [CrossRef]
5. Duan, S.; Song, X.; Cui, Y.; Xu, Z.; Liu, W.; Fu, J.; Zhu, Z.; Li, D. Intelligent kick warning based on drilling activity classification. *Geoenergy Sci. Eng.* **2023**, *222*, 211408. [CrossRef]
6. Akinsete, O.; Adesiji, B. Bottom-hole pressure estimation from wellhead data using artificial neural network. In Proceedings of the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 5–7 August 2019.
7. Nhat, D.; Venkatesan, R.; Khan, F. Data-driven Bayesian network model for early kick detection in industrial drilling process. *Process Saf. Environ. Prot.* **2020**, *138*, 130–138. [CrossRef]
8. Alouhali, R.; Aljubran, M.; Gharbi, S.; Al-yami, A.; Aramco, S. Drilling through data: Automated kick detection using data mining. In Proceedings of the SPE International Heavy Oil Conference and Exhibition, Kuwait City, Kuwait, 10–12 December 2018.
9. Li, W.; Yang, X.; Chen, X.; Chen, H.; Li, Y.; Li, X. Research and application of downhole blowout prevention system while drilling: A review. *J. Pet. Sci. Eng.* **2020**, *188*, 106882.
10. Fu, J.; Liu, W.; Han, X.; Li, F. CNN-LSTM Fusion Network Based Deep Learning Method for Early Prediction of Overflow. *China Pet. Mach.* **2021**, *49*, 16–22.
11. Sule, I.; Khan, F.; Butt, S.; Yang, M. Kick control reliability analysis of managed pressure drilling operation. *J. Loss Prev. Process Ind.* **2018**, *52*, 7–20. [CrossRef]
12. Chen, X.; He, M.; Xu, M.; Wang, S.; Dai, B. Early gas kick detection-inversion-control integrated system: The significance of applications of managed pressure drilling: A review. *Geoenergy Sci. Eng.* **2023**, *229*, 212134. [CrossRef]
13. Xu, Z. *Research and Design of Overflow Monitoring and Diagnosis System of Controlled Pressure Drilling Based on Cloud Computing*; Southwest Petroleum University: Chengdu, China, 2019.
14. Farahani, M.; Reza, S.; Saeid, J. A robust modeling approach for predicting the rheological behavior of thixotropic fluids. In Proceedings of the 80th EAGE Conference and Exhibition 2018, Copenhagen, Denmark, 11–14 June 2018; Volume 1, pp. 1–5.
15. Norouzi, S.; Nazari, M.; VasheghaniFarahani, M. A novel hybrid particle swarm optimization-simulated annealing approach for CO_2-oil minimum miscibility pressure (MMP) prediction. In Proceedings of the 81th EAGE Conference and Exhibition 2019, London, UK, 3–6 June 2019; Volume 1, pp. 1–5.
16. Mirza, M.; Ghoroori, M.; Chen, Z. Intelligent Petroleum Engineering. *Engineering* **2022**, *18*, 27–32. [CrossRef]
17. Muojeke, S.; Venkatesan, R.; Khan, F. Supervised data-driven approach to early kick detection during drilling operation. *J. Pet. Sci. Eng.* **2020**, *192*, 107324. [CrossRef]
18. Liang, H.; Zou, J.; Liang, W. An early intelligent diagnosis model for drilling overflow based on GA–BP algorithm. *Clust. Comput.* **2019**, *22*, 10649–10668. [CrossRef]
19. Si, M. *Research on Real-Time Early-Warning Method of Drilling Overflow*; Southwest Petroleum University: Chengdu, China, 2016.

20. Shi, X.; Zhou, Y.; Zhao, L.; Jiang, H. *Study on How to Determine Kicks and Losses in Real Time Spill on Basis of Random Forest Method*; Drilling & Production Technology: Chengdu, China, 2020; pp. 9–12.
21. Wang, Z.; Chen, G.; Zhang, R.; Zhou, W.; Hu, Y. Early monitoring of gas kick in deepwater drilling based on ensemble learning method: A case study at South China Sea. *Process Saf. Environ. Prot.* **2023**, *169*, 504–514. [CrossRef]
22. Wang, Q.; Zhang, F.; Li, Z.; Wang, Y.; Fang, H. Real-time intelligent drilling monitoring technique based on the coupling of drilling model and artificial intelligence. *Oil Drill. Prod. Technol.* **2020**, *42*, 6–15.
23. Yin, Q.; Yang, J.; Tyagi, M.; Zhou, X.; Hou, X.; Cao, B. Field Data Analysis and Risk Assessment of Gas Kick during Industrial Deepwater Drilling Process Based on Supervised Learning Algorithm. *Process Saf. Environ. Prot.* **2020**, *146*, 10–24. [CrossRef]
24. Yin, Q.; Yang, J.; Li, Z.; Huang, Y.; Luo, M.; Wang, B.; Tyagi, M.; Xu, G.; Zhao, X. A Field Case Study of Managed Pressure Drilling in Offshore Ultra High-Pressure High-Temperature Exploration Well in the South China Sea. *SPE Drill. Complet.* **2020**, *35*, 503–524. [CrossRef]
25. Wang, Y.; Hao, J.; Zhang, F.; Wei, Q.; Peng, Z.; Duan, M. Adaptive LSTM early warning method for kick detection in drilling. *Control Theory Appl.* **2022**, *39*, 441–448.
26. Gao, X.; Li, H.; Wang, Y.; Chen, T.; Zhou, X.; Zhong, L. Fault Detection in Managed Pressure Drilling Using Slow Feature Analysis. *IEEE Access* **2018**, *6*, 34262–34271. [CrossRef]
27. Li, F.; Wang, Q.; Hu, J.; Tang, Y. Combined Data-driven and Knowledge-driven Methodology Research Advances and Its Applied Prospect in Power Systems. *Proc. CSEE* **2021**, *41*, 4377–4390.
28. Zhang, F.; Islam, A.; Zeng, H.; Chen, Z.; Zeng, Y.; Wang, X.; Li, S. Real Time Stuck Pipe Prediction by Using a Combination of Physics-Based Model and Data Analytics Approach. In Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, United Arab Emirates, 11–14 November 2019.
29. Liu, W.; Fu, J.; Deng, S.; Huang, P.; Zou, Y.; Shi, Y.; Cai, C. Overflow Identification and Early Warning of Managed Pressure Drilling Based on Series Fusion Data-Driven Model. *Processes* **2024**, *12*, 1436. [CrossRef]
30. Bi, J.; Li, J.; Wu, K.; Chen, Z.; Gao, Y.; Feng, D.; Zhang, S.; Li, X. A data-driven flow surrogate model based on a data-driven and physics-driven method. *Pet. Geol. Recovery Effic.* **2023**, *30*, 104–114.
31. Zhu, Z.; Liu, Z.; Song, X.; Zhu, S.; Zhou, M.; Li, G.; Duan, S.; Ma, B.; Ye, S.; Zhang, R. A physics-constrained data-driven workflow for predicting bottom hole pressure using a hybrid model of artificial neural network and particle swarm optimization. *Geoenergy Sci. Eng.* **2023**, *224*, 211625. [CrossRef]
32. Ji, Y. *Research on Intelligent Diagnosis Method of Downhole Overflow Leakage Risk Based on Drilling Big Data and Physical Model*; China University of Petroleum: Beijing, China, 2018.
33. Zamani, M.; Nikoo, M.; Niknazar, F.; Al-Rawas, G.; Al-Wardy, M.; Gandomi, A. A multi-model data fusion methodology for reservoir water quality based on machine learning algorithms and bayesian maximum entropy. *J. Clean. Prod.* **2023**, *416*, 137885. [CrossRef]
34. Li, Y.; Cao, W.; Gopaluni, R.; Hu, W.; Wu, M. Early warning of drillstring faulty conditions based on multi-model fusion in geological drilling processes. *J. Process Control* **2023**, *126*, 26–35. [CrossRef]
35. Erge, O.; Oort, E. Hybrid physics-based and data-driven modeling for improved standpipe pressure prediction. In Proceedings of the SPE/IADC Drilling Conference and Exhibition, Online, 8–12 March 2021.
36. Huang, K.; Xiong, B.; Cui, G.; Li, Z.; Bao, B. Dynamic Identification Method of Demagnetization Fault of Double V-Shaped PMSM Based on Pearson Correlation Analysis. *Trans. China Electrotech. Soc.* **2024**, *39*, 7111–7125.

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.