Prediction Technology of a Reservoir Development Model While Drilling Based on Machine Learning and Its Application

Xin Wang [1], Min Mao [1], Yi Yang [1], Shengbin Yuan [1], Mingyu Guo [2], Hongru Li [3], Leli Cheng [4,*], Heng Wang [4] and Xiaobin Ye [4]

1 China-France Bohai Geoservices Co., Ltd., Tianjin 300452, China; wangxin@cfbgc.com (X.W.); maomin@cfbgc.com (M.M.); yangyi@cfbgc.com (Y.Y.); yuansb@cfbgc.com (S.Y.)
2 Tianjin Branch of CNOOC Ltd., Tianjin 300459, China; guomy@cnooc.com.cn
3 CNOOC Energy Development Co., Ltd., Tianjin 300459, China; lihr2@cnooc.com.cn
4 Institute of Logging Technology and Engineering, Yangtze University, Jingzhou 434023, China; 2021710402@yangtzeu.edu.cn (H.W.); 2022711009@yangtzeu.edu.cn (X.Y.)
* Correspondence: chengleli@yangtzeu.edu.cn

Abstract: In order to further understand the complex spatial distribution caused by the extremely strong heterogeneity of buried hill reservoirs, this paper proposes a new method for predicting the development pattern of buried hill reservoirs based on the traditional pre-drilling prediction and post-drilling evaluation methods that mainly rely on seismic, logging, and core data, which are difficult to meet the timeliness and accuracy of drilling operations. Firstly, the box method and normalization formula are used to process and normalize the abnormal data of element logging and engineering logging, and then the stepwise regression analysis method is used to optimize the sensitive parameters of element logging and engineering logging. The Light Gradient Boosting Machine (LightGBM) algorithm, deep neural network (DNN), and support vector machine (SVM) are used to establish a new method for predicting the development pattern of buried hill reservoirs. Lastly, a comprehensive evaluation index F1 score for the model is established to evaluate the prediction model for the development pattern of buried hill reservoirs. The F1 score value obtained from this model's comprehensive evaluation index indicates that the LightGBM model achieves the highest accuracy, with 96.7% accuracy in identifying weathered zones and 95.8% accuracy in identifying interior zones. The practical application demonstrates that this method can rapidly and accurately predict the development mode of buried hill reservoirs while providing a new approach for efficient on-site exploration and decision-making in oil and gas field developments. Consequently, it effectively promotes exploration activities as well as enhances the overall process of oil and gas reservoir exploration.

Keywords: development mode; buried hill reservoirs; element logging; engineering logging; stepwise regression analysis; LightGBM algorithm

1. Introduction

Currently, the exploration and development of buried mountain oil and gas fields have emerged as a crucial area for augmenting offshore oil and gas storage and production [1,2]. Diverging from conventional reservoirs, burial hill reservoirs exhibit high heterogeneity with complex and diverse spatial characteristics, posing significant challenges in their evaluation and development. Extensive research has been conducted on the assessment of buried hill reservoirs, highlighting that lithology serves as the fundamental factor influencing reservoir development. Subsequently, geological processes such as tectonic movements and weathering leaching contribute to the formation of fractures or karst caves during later stages, resulting in pronounced heterogeneity within buried hill reservoirs, which hampers accurate prediction of their spatial distribution. However, these geological processes lead to distinct variations in reservoir reconstruction [3,4], consequently leading

to evident zonation patterns in terms of spatial types and properties within buried hill reservoirs. Therefore, investigating zonation developmental characteristics is highly significant for predicting superior reservoir zones and formulating effective exploration and development strategies.

Previous studies on the characteristic model of reservoir zonation development in buried hills primarily relied on seismic data, core data, well logging data, and well logging data to establish a zonation method based on the understanding of geological origin derived from the geological processes experienced by buried hills [5]. Currently, seismic data are predominantly utilized to predict the development of buried hill reservoirs prior to drilling. Zhang Zhijun et al. comprehensively employed well seismic data to investigate the seismic response characteristics of faults and fractures at various scales, thereby providing a theoretical foundation for fracture prediction in deep buried hill reservoirs [6]. Song Aixue et al. integrated forward modeling techniques to establish dominant seismic facies within different buried hill facies zones and combined them with lithology data to establish identification markers for zoning in buried hills [7]. The utilization of seismic data for zonation studies in buried hills mainly focuses on pre-drilling predictions; however, due to limitations in existing seismic data quality and complexities associated with fracture development within these formations, predicting favorable reservoirs often yields multiple solutions. Furthermore, numerous scholars have conducted a comprehensive evaluation of the buried hill development model using geological data such as core samples, thin sections, and well logging data [8,9]. Wang Deying et al. extensively employed various experimental methods, including core analysis, X-ray diffraction, scanning electron microscopy, conventional physical properties testing, zircon dating, mineral dissolution simulation, statistical analysis methods to investigate the geological characteristics, genetic mechanisms, and a development model of gneiss weathering crust reservoirs in the Bohai Sea [2,10,11]. Shenwei et al. successfully achieved qualitative and quantitative characterization of each phase zone by integrating core samples, rock flakes, conventional logging data, imaging logging data, and array acoustic logging data [12–14]. Although these approaches are comprehensive in nature, they rely on post-drilling logging data and coring experiment results, which may not fulfill the requirement for effective operational decision-making [15]. Therefore, an efficient drilling-based method is urgently needed to determine the zonal development mode of buried hill reservoirs.

Due to the exorbitant costs associated with offshore oil and gas drilling, limited real-time data availability, and inadequate evaluation methods for mining parameters, accurately predicting the zonal development mode of buried hill reservoirs through drilling is of utmost importance [16–19]. This paper proposes an intelligent learning algorithm-based evaluation method for assessing the development model of buried hill reservoirs while drilling, leveraging drilling engineering logging and element logging technology. The proposed approach effectively guides reservoir evaluation and facilitates the formulation of development plans, yielding promising results in our research area.

2. Geological Setting

The Bozhong A gas field is situated in the southwest of the Bozhong Depression, encompassed by both the Bozhong and Huanghekou Depressions. It represents a near north–south structural ridge that is surrounded by the Bozhong Southwest Depression, Bozhong main depression, and Huanghekou Depression [20]. The primary reservoirs containing gas are Archaean metamorphic buried hills. This gas field constitutes a fracture-dominated massive condensate gas reservoir with a burial depth ranging from 3870 m to 4700 m. The homogeneity of the buried hill reservoir in Archean is primarily controlled by geological factors such as paleogeomorphology, tectonic movements, and weathering leaching. Due to variations in weathering and structural homogenization, the buried hill reservoir exhibits distinct characteristics of a weathering zone and an inner zone from top to bottom (Figure 1). Among them, the weathered zone is influenced by both faulting and weathering processes, resulting in the development of reservoir spaces such as structural

fractures, weathered fractures, and corrosion holes. The continuity of the reservoir plane is excellent, with a thick high part gradually thinning towards the low part, exhibiting layered characteristics. Overall, the vertical thickness of gas drilled within the weathering zone ranges from 125.1 m to 227.0 m, averaging 170 m. The content of dark minerals in this zone is relatively low, with an average of 5.7%. Interpreted porosity averages at 4.4%, permeability at 3.3 mD on average, and a net-to-gross ratio ranging from 45% to 69%. With increasing depth inside the buried hill structure, geological stress gradually reduces its influence while faults primarily control reservoir development; fractures are distributed along these faults in a zonal pattern [21]. The gas layer thickness within this inner zone measures approximately 171.6 m $\perp$ 165.0 m with higher dark mineral content compared to that found in the weathered zone. The logging interpretation indicates an average porosity and permeability of 2.9% and 2.0 mD, respectively, while the net hair ratio stands at around 28.0%, lower than that observed in the weathered zone. The buried hill reservoir as a whole is characterized by the longitudinal variation in physical property deterioration from the weathering zone to the inner zone.
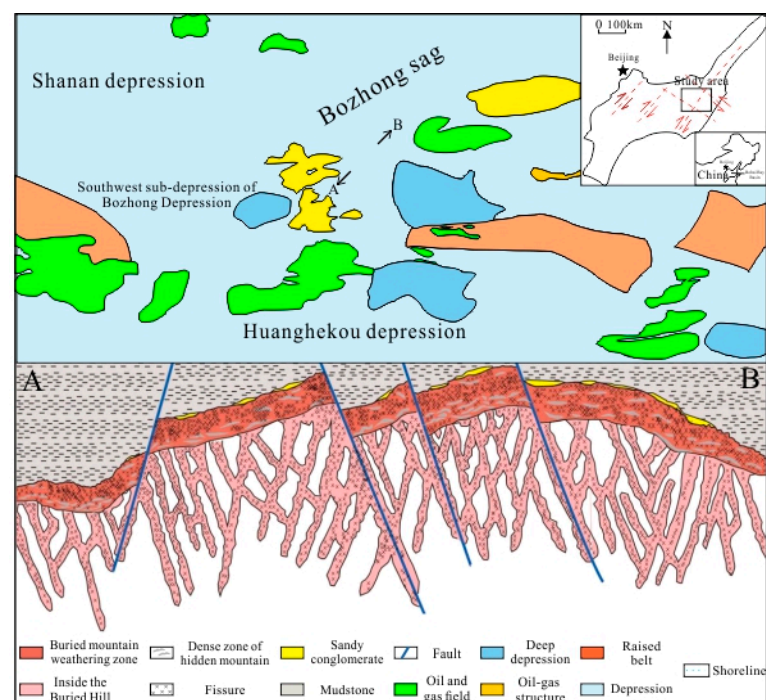

Figure 1. Reservoir development model of a buried hill in the Bozhong A gas field, (A,B) is the location of the lower profile.

Based on the aforementioned characteristics of reservoir development patterns and fracture distribution, the development strategy for the Archaean buried hill in this gas field has been determined. A set of development strata is adopted primarily to exploit the reserves within the weathering zone of the buried hill. Therefore, it is crucial to employ geological model evaluation technology for vertical zonation of buried hills in order to optimize the oil and gas field development plan.

3. Material and Method

Element logging and engineering parameter logging are high-performance real-time data generated during drilling. The rock type is an internal factor that controls the formation of fractures in buried hills, primarily influencing the difficulty of fracture formation and the storage and permeability spaces within buried hills under external forces [22]. Continuous analysis of rock cuttings using element logging technology enables timely and effective monitoring of changes in major elements and trace elements, reflecting the evolving characteristics of rock types [23]. Real-time monitoring of pore and fracture

development in buried hill reservoirs can reflect changes in drilling parameters, indicating the presence of drillable reservoir space within weathering zones. Therefore, drilling engineering parameters often exhibit a strong correlation with effective reservoir development. Consequently, we propose a machine learning-based prediction method for reservoir development patterns utilizing these two LWD techniques [24].

3.1. Data Collection and Processing

The parameters of elemental logging and engineering logging were calculated for 18 wells in the research area. Firstly, the sample underwent pretreatment. The selected elements logging measured the main elements, including Na, Mg, Al, Si, P, S, Cl, K, Ca, Ba, Ti, Mn, and Fe. Varying influence weights led to the exclusion of trace elements such as V, Ni, Sr, and Zr. Simultaneously, the combination of the drilling parameters, comprising torque, fracture pressure gradient, weight on bit, penetration rate, and drilling time, was studied. Additionally, the sampling intervals differed between the two logging technologies, with a 5 m interval for element logging and a 1 m interval for engineering logging. The average engineering parameter corresponding to a depth up to 5 m from the element parameter was considered as the depth engineering parameter value. Finally, this value was further processed.

The box plot method is used to characterize the abnormal situation of the data, to detect and identify the abnormal value of the data, and to remove it from the sample. As shown in Figure 2, the value Q3, corresponding to 75% of the subpoints of the data, is the upper quartile, the value Q2, corresponding to 50% of the subpoints of the data, is the median, and the value Q1, corresponding to 25% of the subpoints of the data, is the lower quartile. The calculation formula for the upper limit is Q3 + 1.5 (Q3 − Q1), and the calculation formula for the lower limit is Q1 − 1.5 (Q3 − Q1). When using a box chart to identify outliers, when the datum value is greater than the upper limit of the box chart or less than the lower limit of the box chart, it is judged to be an outlier [25].
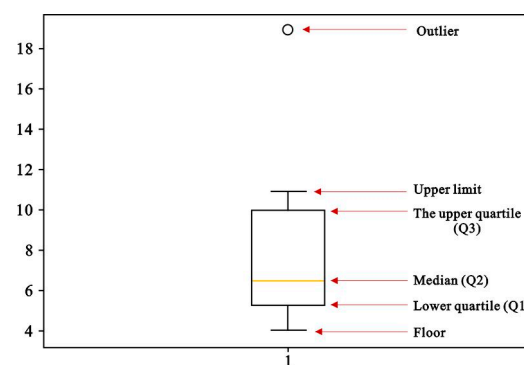

Figure 2. "Box Diagram" abnormal data processing.

In addition, in order to eliminate the impact on subsequent machine learning, the input data are standardized, and the processing formula is as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x' is the standardized value of the feature parameter, x is the characteristic parameter value, and x_{max} is the maximum value of the feature parameter. x_{min} is the minimum value of the feature parameter.

The standard post-values of the feature parameters after data deletion are taken as the input parameters of the model.

3.2. The Parameters Were Selected by Stepwise Regression Analysis

The contribution of each feature to the dependent variable will be determined through stepwise regression, which determines whether a feature should be included in the regression model [26]. If a newly introduced variable fails to maintain its significance based on F-test, it will be eliminated. This ensures that important variables are always retained in the model, reducing input dimensions and mitigating the risk of overfitting [27]. The implementation steps are as follows:

When A significance level α is given, the critical value $F\alpha(m, n - m - 1)$ of the rejection domain is determined according to the degree of freedom m (that is, the number of features) and $n - m - 1$. For each feature parameter $x_i(1 \leq i \leq m)$, a unitary linear regression model is fitted, and the statistic Q is regarded as an empty set, that is, $SSE(Q) = SST$, then $SST = (x_i|Q) = SSR(x_i)$, $MSE(Q, x_i) = MSE(x_i)$, and each F_i are calculated:

$$F_i = \frac{SSR(x_i)}{MSE(x_i)}, i = 1, 2, \ldots m \tag{1}$$

$$F_{i1} = \max_{1 \leq i \leq m} F_i \tag{2}$$

In the aforementioned formula, SSE represents the sum of squared errors between the predicted values of the model and the original values. SSR denotes the sum of squared differences between the predicted values of the model and the mean value of the original data. SST signifies the sum of squared differences between the original data and its mean value. MSE stands for mean square error, while $F\alpha$ corresponds to the critical value in the rejection domain.

If $Fim > F\alpha$, it indicates that regression model with characteristic parameter x_i is selected as the current model; otherwise, no independent variables are introduced into this model. By incorporating the remaining $m - 1$ characteristic parameters into the current model, we obtain $m - 1$ binary regression models.

$$F_i = \frac{SSR(x_i|x_{i1})}{MSE(x_{i1}, x_i)}, i \neq 1 \tag{3}$$

$$F_{i2} = \max_{i \neq i1} F_i \tag{4}$$

If $F_{i2} > F\alpha$, the selection ends, and the selected model parameters are optimal. If $F_{i2} > F\alpha$, then the feature parameter x_{i2} is selected into the model, and x_{i2} is introduced according to the above calculation, which still has a significant impact on y:

$$F_i = \frac{SSR(x_{i1}|x_{i2})}{MSE(x_{i1}, x_{i2})} \tag{5}$$

If $F_{i2} \leq F\alpha$, then x_{i1} needs to be eliminated, and the model containing x_{i2} is the optimal model.

Based on the selected model from the previous step, the remaining $m - 2$ feature parameters are incorporated into the current model, and their F-values are calculated for fitting and merging, determining whether to introduce these parameters. Finally, these steps are iterated until all feature parameters have been either selected or eliminated, resulting in the preferred set of remaining parameters.

3.2.1. The Parameters of Logging Elements Were Optimized by Stepwise Regression Analysis

The dependent variable in this study was the type of reservoir development model, while the independent variables were considered as the main elements. Table 1 and Figure 3 present the analysis of variance and the model, respectively. In the stepwise regression analysis, each variable was introduced individually based on its importance [28]. As shown in Table 2, Al was the first variable to enter. At this stage, SSR (the sum of squares of

differences between predicted and actual types) equaled 23.669 with a degree of freedom equal to 1; SSE (the sum of squares error between predicted and actual types) equaled 163.277 with a degree of freedom equal to 756. The mean square error can be derived from this $MSE = \frac{SSR}{P} = 23.669$, mean residual sum of squares $MSR = \frac{SSE}{n-p-1} = 0.216$, hence, $F = \frac{MSR}{MSE} = 109.579$, the F-value is the result of the F-test.

Table 1. Analysis of variance of logging element parameters.

	Model	Quadratic Sum	DOF	Mean Square	F-Value	Significance p
a	Regression	23.669	1	23.669	109.593	<0.001
	Residual error	163.277	756	0.216	—	—
	Summary	186.946	757	—	—	—
b	Regression	34.840	2	17.420	86.468	<0.001
	Residual error	152.106	755	0.201	—	—
	Summary	186.946	757	—	—	—
c	Regression	48.040	3	16.013	86.922	<0.001
	Residual error	138.906	754	0.184	—	—
	Summary	186.946	757	—	—	—
d	Regression	51.387	4	12.847	71.362	<0.001
	Residual error	135.559	753	0.180	—	—
	Summary	186.946	757	—	—	—
e	Regression	53.409	5	10.682	60.153	<0.001
	Residual error	133.537	752	0.178	—	—
	Summary	186.946	757	—	—	—
f	Regression	55.306	6	9.218	52.586	<0.001
	Residual error	131.640	751	0.175	—	—
	Summary	186.946	757	—	—	—
g	Regression	60.047	7	8.578	50.699	<0.001
	Residual error	126.899	750	0.169	—	—
	Summary	186.946	757	—	—	—
h	Regression	61.743	8	7.718	46.171	<0.001
	Residual error	125.203	749	0.167	—	—
	Summary	186.946	757	—	—	—

Predictive variable: a—Al; b—Al, S; c—Al, S, Si; d—Al, S, Si, Mn; e—Al, S, Si, Mn, Na; f—Al, S, Si, Mn, Na, Ca; g—Al, S, Si, Mn, Na, Ca, K; h—Al, S, Si, Mn, Na, Ca, K, Fe.
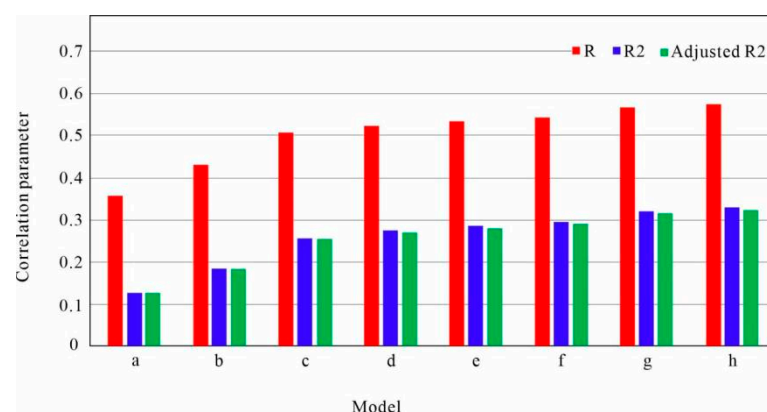

**Figure 3.** Model effect. "a–h" represents the different models in Table 1.

Table 2. Analysis of variance of engineering parameters.

	Model	Quadratic Sum	DOF	Mean Square	F-Value	Significance p
a	Regression	48.585	1	48.585	265.464	<0.001
	Residual error	138.361	756	0.183	—	—
	Summary	186.946	757	—	—	—
b	Regression	54.023	2	27.011	153.425	<0.001
	Residual error	132.923	755	0.176	—	—
	Summary	186.946	757	—	—	—
c	Regression	55.756	3	18.585	106.817	<0.001
	Residual error	131.190	754	0.174	—	—
	Summary	186.946	757	—	—	—
d	Regression	57.521	4	14.380	83.665	<0.001
	Residual error	129.425	753	0.172	—	—
	Summary	186.946	757	—	—	—
e	Regression	58.251	5	11.650	68.076	<0.001
	Residual error	128.695	752	0.171	—	—
	Summary	186.946	757	—	—	—

Predictive variable: a—Fracture pressure gradient; Drilling time; b—Fracture pressure gradient, Drilling time, Drilling rate; c—Fracture pressure gradient, Drilling time, Drilling rate, Torque; d—Fracture pressure gradient, Drilling time, Drilling rate, Torque, Bit pressure. Dependent variable: e—Reservoir development model type.

In Figure 3, R represents the correlation, which exhibits a gradual increase with the introduction of the preferred parameter. The decision coefficient R^2 signifies the percentage that the fitted model can explain the variation in the dependent variable. The progressive increment of R^2 indicates improved effectiveness of the regression model, while adjusted R^2 is employed to evaluate how well the model fits due to an increase in independent variables. According to Figure 3, upon introducing the Al element, there is a complex decision coefficient R^2 value of 0.127 for this model and a significance level $p = 0.000b < 0.05$, confirming significant regression coefficients. Following regression principles, Al should be included as an optimal variable by introducing one element at a time and repeating these steps accordingly. Eventually, the parameters selected include Al, S, Si, Mn, Na, Ca, K, and Fe, whereas other elements are eliminated from consideration.

In Figure 4, examining standardized residuals and measured cumulative probability reveals that final feature distribution generally adheres to normality assumptions with close proximity between the curve and diagonal line, indicating a good fitting effect of our model.
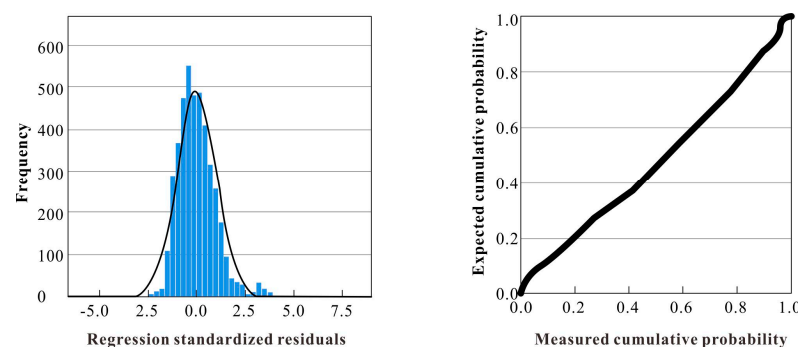
Figure 4. Normalized residuals and cumulative probabilities of parameter preference regression.

3.2.2. Using Stepwise Regression Analysis to Optimize Drilling Engineering Parameters

The progressive regression analysis method was employed to optimize the drilling engineering parameters, with the geological model type considered as the dependent variable and fracture pressure gradient, drilling time, drilling rate, torque, weight on bit, and *Dc* index taken as independent variables for gradual analysis. The summary of the stepwise regression analysis model and variance analysis is presented in Table 2 and Figure 5, respectively. The optimized parameters include fracture pressure gradient, drilling time, drilling rate, torque, and weight on bit.
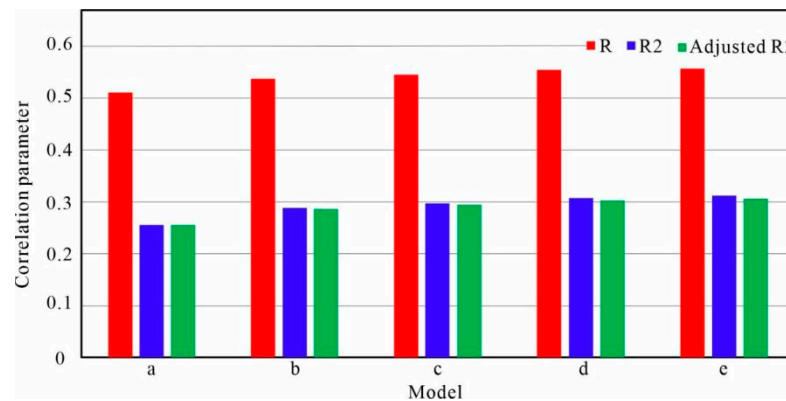

Figure 5. Model effect. "a–e" represents the different models in Table 2.

3.3. Model Establishment and Evaluation

Compared to traditional neural networks, the deep neural network (DNN) possesses certain advantages. The increased depth resulting from multiple hidden layers significantly enhances its performance [29]. However, this improvement comes at the cost of a larger number of parameters, such as initial network topology values, weights, and thresholds. Consequently, the model becomes more complex and prone to overfitting. Support vector machine (SVM) is a supervised learning model commonly employed for pattern recognition, classification, and regression analysis [30]. Nevertheless, SVM exhibits limitations in handling large sample sizes and utilizing kernel functions effectively for linear non-fractional data selection criteria, which are absent [31]. It also demonstrates heightened sensitivity towards missing or noisy data while failing to provide robust support for feature diversity [32]. In this study, we employ the LightGBM algorithm based on gradient-boosting trees to construct our model. This approach not only supports multi-class features but also reduces time and space overhead during unilateral gradient sampling compared to traditional machine learning algorithms when traversing all feature values. Additionally, it offers benefits such as low memory consumption, reduced computational costs, fast training speed with high accuracy rates while efficiently processing massive datasets [33–36].

The GBDT model is an integrated tree-based approach that effectively addresses classification and regression problems [37]. By iteratively fitting the residuals of the previous model with new trees, a complete model comprising K trees is trained, and the final predicted value is obtained by summing up the corresponding results from each tree. In line with formula 6, the fundamental concept behind the LightGBM algorithm lies in combining M weak regression trees into robust ones [38].

$$F(x) = \sum_{k=1}^{K} f_k(x) \tag{6}$$

The innovation of LightGBM lies in the incorporation of novel technologies, namely Exclusive Feature Bunching (EFB) and Gradient-based One-Side Sampling (GOSS), into the histogram-based GBDT algorithm. EFB enables the fusion and binding of certain fea-

tures to effectively reduce feature dimensionality without compromising accuracy. Meanwhile, the GOSS algorithm ensures accurate estimation of information gain while reducing sample size.

In this study, a total of 1608 data combinations from 18 wells were utilized as training samples. The aforementioned preferred logging elements and drilling engineering parameters were employed as input variables for the model, while the labeled–coded types of buried hill reservoir development patterns served as the output variables. Consequently, a machine learning-based prediction model for buried hill reservoir development patterns was established. To enhance the accuracy assessment of the model, it is essential to establish a comprehensive set of evaluation indices [39–41]. Traditionally, prediction accuracy has been used to measure model performance; however, in cases where sample sizes are imbalanced, accuracy alone fails to reflect the true predictive capability of the model. Therefore, this paper proposes utilizing the F1 score index for comprehensive evaluation purposes. The formula for calculating the F1 score is presented below.

$$precision = TP/(TP + FP) \tag{7}$$

$$recall = TP/(TP + FN) \tag{8}$$

$$F1_{score} = 2 \times precision \times recall/(precision + recall) \tag{9}$$

In the formula, *precision* is the accuracy rate, *recall* is the recall rate, and the classification results of pattern recognition can generally be classified into four categories, *TP* is the true example, *FP* is the false positive example, *FN* is the false counter-example, and *TN* is the true counter-example.

The F1 score demonstrates a robust amalgamation of the detection rate and recall rate, rendering it a more comprehensive evaluation metric.

3.4. Model Verification

Based on previous knowledge of the buried hill reservoir development pattern by conventional logging curves, imaging logging, wall core, core, and slice, the accuracy and reliability of the prediction of the buried hill reservoir development pattern based on the learning model of the base tool in this paper are verified. The difference in reservoir performance between the weathered zone and inner zone is also evident in the physical properties, drilling time, resistivity, and other electrical measurement curves. Due to the relatively developed fractures in the weathered zone, the average linear fracture density of imaging logging is 3–6 fractures/m, showing good reservoir performance; the average porosity of logging interpretation is 2.4–6.5%, and the net gross ratio is 0.33–0.62. However, the reservoir in the inner zone is poor as a whole, with an average linear fracture density of 0.8–1.2 fractures/m. The average porosity of logging interpretation is 1.7–3.9%, and the net-to-gross ratio is less than 0.35. It can be seen from the electrical measurement curve that cracks develop in the weathering zone, the drilling time and resistivity are relatively low, the drilling time is 8–29 min/m, and the resistivity is 170–1100 $\Omega \cdot$m. The inner zone dense layer is relatively developed, and the drilling time is generally higher, ranging from 12 to 52 min/m, and the resistivity is significantly higher than that of the weathering zone, ranging from 700 to 22,000 $\Omega \cdot$m. The core shows that the whole fracture of the weathering zone is relatively developed, and the analysis of the cast thin slice and scanning electron microscope shows that the micro-reservoir space of the weathering zone is mainly micro-fracture, followed by the dissolution porosity, which is distributed in a beaded pattern along the micro-fracture. The main reservoir space type of the inner zone is that weathering fractures and dissolution pores formed by weathering leaching are less developed, and the matrix is relatively dense [42].

4. Result

During the drilling process of metamorphic rock buried hill in well N1 of the Bozhong A gas field, the optimal element logging parameters Al, S, Si, Mn, Na, Ca, K, Fe, and

engineering parameters such as rupture pressure gradient, drilling time, drilling rate, torque, and bit weight were predicted using the LightGBM algorithm model. It is predicted that the boundary depth between the weathering zone and the inner zone of the well is 4210.0 m. After drilling operations concluded successfully, the analysis data from the logging curve analysis, image logging interpretation, wireline sidewall core sampling, and thin section examination indicate that above 4205.0 m formation resistivity, values are below 200 Ω, making it evident that rock density falls below 2.73 g/cm^3 while actual porosity ranges between 1.2% and 12.8%, which can be attributed to weathering processes and tectonic activities. The reservoir space is predominantly characterized by pores and fractures. The overall resistivity of the inner fracture zone generally ranges between 200 Ω and 2000 Ω, while the rock density exhibits an overall increase, typically exceeding 2.71/cm^3. The actual porosity falls within the range of 1.3% to 3.1%, primarily influenced by structural dynamics, with cracks dominating the reservoir space. These findings indicate that the boundary depth between the genuine weathering crust and the inner zone lies at a depth of 4205.0 m (Figure 6). Moreover, these studies demonstrate that drilling predictions align with the actual development model of buried hills.
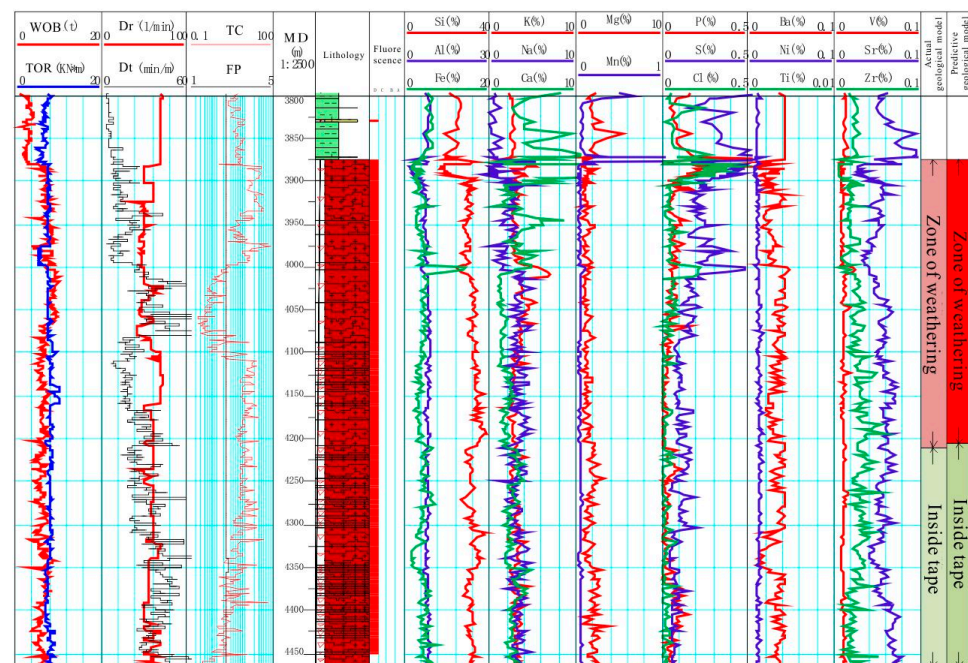

Figure 6. Reservoir development model prediction of a buried hill well N1 in the Bozhong A gas field.

During the drilling process itself, based on predictions from our reservoir development model regarding boundaries between weathered zones and internal zones, we decided to conduct production testing in the upper weathered zone initially before utilizing a 9.53 mmPC oil nozzle for further production testing purposes. This approach resulted in an average daily oil production rate of 176.94 m^3/d alongside an average daily gas production rate of 178,586 m^3/d—indicative of high productivity levels observed in this wellbore scenario. The accuracy and reliability of our prediction model are further substantiated through these outcomes. Compared with logging, imaging logging, core, and cast thin sections, the development model of the buried hill in the Bozhong Depression can be judged at least 15 days earlier, and the cost can save at least 30%.

5. Discussion

By utilizing 150 data combinations from two wells as test samples, we validated the efficacy of three machine learning models, namely LightGBM, SVM, and DNN. As depicted in Figure 7, the F1 score values demonstrate that LightGBM exhibits the highest accuracy

in identifying weathering zones at an impressive rate of 96.7%, along with a commendable accuracy of 95.8% for inside zone identification. Following closely is the DNN with accuracies of 88.6% and 90.2% for the weathering zone and inside zone identification, respectively. In contrast, SVM achieves lower accuracies of 84.6% and 87.6% for the weathering zone and inside zone identification, respectively, when compared to the other two models.
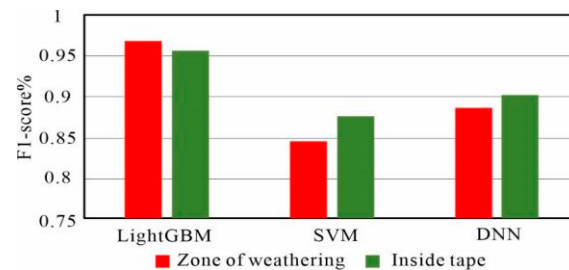

Figure 7. Comparison of accuracy of prediction results of the three models.

In summary, it can be observed that different models exhibit certain variations in predicting reservoir development patterns. Notably, the LightGBM algorithm demonstrates superior performance, highlighting the superiority of gradient-boosting tree models over classical machine learning methods like SVM. This finding underscores the advanced nature and predictive capabilities of the LightGBM algorithm in forecasting buried hill reservoir development patterns.

6. Conclusions

1. A machine learning-based approach is proposed for predicting the development mode of buried hill reservoirs during drilling. Firstly, a multi-parameter fusion technique is employed to integrate element logging and engineering logging data obtained while drilling, followed by optimization of sensitive parameters using stepwise regression analysis. Subsequently, a prediction model for the development mode of buried hill reservoirs is established using the LightGBM algorithm, providing a novel method for rapid prediction in this context. The accuracy of the proposed model surpasses previous approaches.

2. Through the validation of three machine learning models, namely LightGBM, SVM, and DNN, it is demonstrated that LightGBM exhibits the highest accuracy in identifying the weathering zone with a remarkable precision of 96.7% while achieving an accuracy rate of 95.8% for identifying the inside zone. Following closely is the DNN, which attains accuracies of 88.6% and 90.2% for the weathering zone and inside zone identification, respectively. The SVM model demonstrates an identification accuracy of 84.6% for the weathering zone and 87.6% for inner zone recognition correspondingly. Consequently, it can be concluded that the LightGBM algorithm model holds great potential in predicting reservoir development patterns within this oilfield. The ideas and methods in this paper can be further applied to the development and production of other oil fields so as to improve the efficiency of exploration and development.

3. In contrast to the transmission evaluation method, this prediction approach employs MWD data for assessment and offers an intelligent technical solution for prediction in scenarios with limited data. It exhibits characteristics of enhanced prediction speed and heightened accuracy. This methodology can serve as a robust foundation for efficient field exploration and development decision-making, thereby effectively advancing the progress of oil and gas reservoir exploration and development in this region.

Data Availability Statement: The datasets presented in this article are not readily available because the well data is confidential to commercial companies.

Conflicts of Interest: Xin Wang, Min Mao, Yi Yang, Shengbin Yuan were employed by the company China-France Bohai Geoservices Co., Ltd. Mingyu Guo were employed by the company Tianjin Branch of CNOOC Ltd. Hongru Li were employed by the company CNOOC Energy Development Co., Ltd. The 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 companies had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Hu, Z.; Xu, C.; Yang, B.; Huang, Z.; Su, W. Genetic mechanism of the granite buried-hill reservoir of the Penglai 9-1 oilfield in Bohai Sea. *Pet. Res.* **2017**, *2*, 355–365. [CrossRef]
2. Zhu, X.; Cai, H.; Wang, X.; Zhu, Q.; Meng, Z. Research and Application of Water Flooding Timing and Method for Blocky Bottom Water Fractured Buried Hill Reservoir. *J. Power Energy Eng.* **2019**, *7*, 1–10. [CrossRef]
3. Deying, W.; Qingbin, W.; Xiaojian LI, U.; Meng, Z.; YiWei, H. Characteristics and developing patterns of gneiss buried hill weathering crust reservoir in the sea area of the Bohai Bay basin. *Acta Petrol. Sin.* **2019**, *35*, 1181–1193. [CrossRef]
4. Xiaofeng, D.; Xiaojian, L.; Xintao, Z.; Yongjun, L.; Yanhong, X. Characteristics and controlling factors of Archen metamorphic reservoir in Bohai sea area. *China Offshore Oil Gas* **2021**, *33*, 15–27.
5. Tong, K.; Li, B.; Dai, W.; Zheng, H.; Zhang, Z.; Cheng, Q.; Wang, J.; Fang, N. Sparse well pattern and high-efficient development of metamorphic buried hills reservoirs in Bohai Sea area, China. *Pet. Explor. Dev.* **2017**, *44*, 625–635. [CrossRef]
6. Zhijun, Z.; Guangrui, X.; Yao, L. Seismic response characteristics and prediction techniques of inside fractures in buried-hill of metamorphic rock in Bozhong 19-6 oilfield. *Oil Geophys. Explor.* **2021**, *56*, 845–852+675.
7. Song, A.; Yang, J.; Yang, J. Zoning characteristics of buried hill reservoir and prediction of favorable reservoir in the western deep water area of South China Sea. *China Offshore Oil Gas* **2020**, *32*, 54–63.
8. Ge, J.; Wu, S.; Touré, D.; Cheng, L.; Miao, W.; Cao, H.; Pan, X.; Li, J.; Yao, M.; Feng, L. Analysis on biomass and productivity of epilithic algae and their relations to environmental factors in the Gufu River basin, Three Gorges Reservoir area, China. *Environ. Sci. Pollut. Res.* **2015**, *24*, 26881–26892. [CrossRef]
9. Gu, Y.; Zhang, D.; Bao, Z. A new data-driven predictor, PSO-XGBoost, used for permeability of tight sandstone reservoirs: A case study of member of chang 4+5, western Jiyuan Oilfield, Ordos Basin. *J. Pet. Sci. Eng.* **2021**, *199*, 108350. [CrossRef]
10. Hong, Y.; Ye, C.; Li, S.; Yang, Y.; Shu, X.; Zhang, J.; Zhang, Q. Advancing Analysis of Spatio-Temporal Variations of Soil Nutrients in the Water Level Fluctuation Zone of China's Three Gorges Reservoir Using Self-Organizing Map. *PLoS ONE* **2015**, *10*, e0121210.
11. Jingshan, C.; Zhong, L.; Zhenyu, W.; Xiucheng, T.; Ling, L.; Qing, M. Paleokarstification and reservoir distribution of Ordovician carbonate rocks in Tarim Basin. *Acta Sedimentol. Sin.* **2007**, *25*, 858–868.
12. Wei, S.; Jun, T.; Jinbo, W.; Baoyong, T.; Ze, H.; Lei, C. Logging characterization method and application of vertical zonation of buried hill reservoir. *China Pet. Explor.* **2023**, *28*, 144–153.
13. Verma, A.K.; Deb, D.; Dey, A.C.; Roy, S.; Singh, A.K.; Avadhani, V.L.N.; Tiwari, R.R. Development of one dimensional geomechanical model for a tight gas reservoir. *Sci. Rep.* **2021**, *11*, 21433. [CrossRef]
14. Kumar, I.; Tripathi, B.K.; Singh, A. Synthetic well log modeling with light gradient boosting machine for Assam-Arakan Basin, India. *J. Appl. Geophys.* **2022**, *203*, 104697. [CrossRef]
15. Salone, R.; De Paola, C.; Carbonari, R.; Rufino, F.; Avino, R.; Caliro, S.; Cuoco, E.; Santi, A.; Di Maio, R. High-resolution geoelectrical characterization and monitoring of natural fluids emission systems to understand possible gas leakages from geological carbon storage reservoirs. *Sci. Rep.* **2023**, *13*, 18585. [CrossRef]
16. Choi, W.; Choi, T.; Heo, S. A Comparative Study of Automated Machine Learning Platforms for Exercise Anthropometry-Based Typology Analysis. Performance Evaluation of AWS SageMaker, GCP VertexAI, and MS Azure. *Bioengineering* **2023**, *10*, 891. [CrossRef]
17. Dev, V.A.; Eden, M.R. Formation lithology classification using scalable gradient boosted decision trees. *Comput. Chem. Eng.* **2019**, *128*, 392–404. [CrossRef]
18. Gu, Y.; Zhang, D.; Lin, Y.; Ruan, J.; Bao, Z. Data-driven lithology prediction for tight sandstone reservoirs based on new ensemble learning of conventional logs. A demonstration of a Yanchang member, Ordos Basin. *J. Pet. Sci. Eng.* **2021**, *207*, 109292. [CrossRef]
19. Nguyen, T.B.N.; Bae, W.; Nguyen, L.A.; Dang, T.Q.C. A New Method for Building Porosity and Permeability Models of a Fractured Granite Basement Reservoir. *Pet. Sci. Technol.* **2014**, *32*, 1886–1897. [CrossRef]

20. Wu, J.; Zhou, D.; Lv, D.; Niu, C.; He, D.; Li, G.; Wang, X. Structural interpretation of inner buried hill under restricted data availability—A case study in KL-X area, Bohai Bay Basin, SEG Technical Program Expanded Abstracts. In Proceedings of the 2014 SEG Annual Meeting, Denver, CO, USA, 26–31 October 2014; pp. 1689–1693.

21. Changgui, X.; Xiaofeng, D.; Xiaojian, L.; Wei, X.; Yiwei, H. Formation mechanism of high-quality deep buried-hill reservoir of Archaeanmetamorphic rocks and its significance in petroleum exploration in Bohai Sea area. *Oil Gas Geol.* **2020**, *41*, 235–247+294.

22. Yongan, X.; Yongbo, C.; Yuanyuan, Z. Recent breakthroughs in hydracarbon exploration in Bohai Sea. *China Offshore Oil Gas* **2015**, *27*, 1–9.

23. Habibnia, B.; Momeni, A. Reservoir Characterization in Balal Oil Field by Means of Inversion, Attribute, and Geostatistical Analysis Methods. *Pet. Sci. Technol.* **2012**, *30*, 1609–1618. [CrossRef]

24. Lovrić, M.; Pavlović, K.; Žuvela, P.; Spataru, A.; Lučić, B.; Kern, R.; Wong, M.W. Machine learning in prediction of intrinsic aqueous solubility of drug-like compounds. Generalization, complexity, or predictive ability? *J. Chemom.* **2021**, *35*, 7–8. [CrossRef]

25. Merembayev, T.; Kurmangaliyev, D.; Bekbauov, B.; Amanbek, Y. A Comparison of Machine Learning Algorithms in Predicting Lithofacies. Case Studies from Norway and Kazakhstan. *Energies* **2021**, *14*, 1896. [CrossRef]

26. Reinwarth, B.; Miller, J.K.; Glotzbach, C.; Rowntree, K.M.; Baade, J. Applying regularized logistic regression (RLR) for the discrimination of sediment facies in reservoirs based on composite fingerprints. *J. Soils Sediments* **2017**, *17*, 1777–1795. [CrossRef]

27. Zhang, S.; Gu, Y.; Gao, Y.; Wang, X.; Zhang, D.; Zhou, L.; Kong, H. Petrophysical Regression regarding Porosity, Permeability, and Water Saturation Driven by Logging-Based Ensemble and Transfer Learnings. A Case Study of Sandy-Mud Reservoirs. *Geofluids* **2022**, *2022*, 9443955. [CrossRef]

28. Plaksina, T.; Gildin, E. Applied method for production design optimization under geologic and economic uncertainties in shale gas reservoirs. *Int. J. Model. Simul.* **2017**, *38*, 67–82. [CrossRef]

29. Song, J.; Liu, G.; Jiang, J.; Zhang, P.; Liang, Y. Prediction of Protein–ATP Binding Residues Based on Ensemble of Deep Convolutional Neural Networks and LightGBM Algorithm. *Int. J. Mol. Sci.* **2021**, *22*, 939. [CrossRef]

30. Tang, M.; Zhao, Q.; Ding, S.X.; Wu, H.; Li, L.; Long, W.; Huang, B. An Improved LightGBM Algorithm for Online Fault Detection of Wind Turbine Gearboxes. *Energies* **2020**, *13*, 807. [CrossRef]

31. Xie, F.; Zhang, W.; Shen, M.; Zhang, J. Multi-condition dynamic model control strategy of the direct drive motor of electric vehicles based on PIO–LightGBM algorithm. *J. Power Electron.* **2022**, *23*, 499–509. [CrossRef]

32. Zhang, X.-x.; Deng, T.; Jia, G.-z. Nuclear spin-spin coupling constants prediction based on XGBoost and LightGBM algorithms. *Mol. Phys.* **2019**, *118*, e1696478. [CrossRef]

33. Cai, J.; Li, X.; Tan, Z.; Peng, S. An assembly-level neutronic calculation method based on LightGBM algorithm. *Ann. Nucl. Energy* **2021**, *150*, 107871. [CrossRef]

34. Fang, Q.; Shen, B.; Xue, J. A new elite opposite sparrow search algorithm-based optimized LightGBM approach for fault diagnosis. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *14*, 10473–11049. [CrossRef] [PubMed]

35. Gu, L.; He, Y.; Liu, H.; Wei, Z.; Guo, J. Metasurface meta-atoms design based on DNN and LightGBM algorithms. *Opt. Mater.* **2023**, *136*, 113471. [CrossRef]

36. Guo, X.; Gui, X.; Xiong, H.; Hu, X.; Li, Y.; Cui, H.; Qiu, Y.; Ma, C. Critical role of climate factors for groundwater potential mapping in arid regions: Insights from random forest, XGBoost, and LightGBM algorithms. *J. Hydrol.* **2023**, *621*, 129599. [CrossRef]

37. Hamed EA, R.; Salem MA, M.; Badr, N.L.; Tolba, M.F. An Efficient Combination of Convolutional Neural Network and LightGBM Algorithm for Lung Cancer Histopathology Classification. *Diagnostics* **2023**, *13*, 2469. [CrossRef] [PubMed]

38. Huang, B.; Wang, C. Research on Data Analysis of Efficient Innovation and Entrepreneurship Practice Teaching Based on LightGBM Classification Algorithm. *Int. J. Comput. Intell. Syst.* **2023**, *16*, 145. [CrossRef]

39. Liu, Y.; Zhu, R.; Zhai, S.; Li, N.; Li, C. Lithofacies identification of shale formation based on mineral content regression using LightGBM algorithm: A case study in the Luzhou block, South Sichuan Basin, China. *Energy Sci. Eng.* **2023**, *11*, 4256–4272. [CrossRef]

40. Meng, Y.; Yang, Q.; Chen, S.; Wang, Q.; Li, X. Multi-branch AC arc fault detection based on ICEEMDAN and LightGBM algorithm. *Electr. Power Syst. Res.* **2023**, *220*, 109286. [CrossRef]

41. Pan, H.; Li, Z.; Tian, C.; Wang, L.; Fu, Y.; Qin, X.; Liu, F. The LightGBM-based classification algorithm for Chinese characters speech imagery BCI system. *Cogn. Neurodynamics* **2022**, *17*, 373–384. [CrossRef]

42. Tingen, F.; Tao, N.; Hongjun, F.; Shuai, W.; Dakun, X.; Jianghua, L. Geological model and development strategy of Archaean buried hill reservoir in BZ19-6 condensate field. *China Offshore Oil Gas* **2021**, *33*, 85–92.