


Article

Prediction of Key Development Indicators for Offshore Oilfields Based on Artificial Intelligence

Ke Li ^{1,2}, Kai Wang ^{1,2}, Chenyang Tang ^{1,2}, Yue Pan ^{1,2}, Yufei He ^{1,2}, Shaobin Cai ^{1,2}, Suidong Chen ³ and Yuhui Zhou ^{4,5,*} 

¹ Development Research Institute, China National Offshore Oil Corporation Research Institute, Beijing 100028, China; like2@cnooc.com.cn (K.L.); wangkai43@cnooc.com.cn (K.W.); tangchy4@cnooc.com.cn (C.T.); panyue@cnooc.com.cn (Y.P.)

² State Key Laboratory of Offshore Oil Exploitation, Beijing 100028, China

³ Hubei Key Laboratory of Oil and Gas Exploration and Development Theory and Technology, China University of Geosciences, Wuhan 430100, China

⁴ School of Petroleum Engineering, Yangtze University, Wuhan 430100, China

⁵ Western Research Institute, Yangtze University, Karamay 834000, China

* Correspondence: zhyhtree@yangtzeu.edu.cn

Abstract: As terrestrial oilfields continue to be explored, the difficulty of exploring new oilfields is constantly increasing. The ocean, which contains abundant oil and gas resources, has become a new field for oil and gas resource development. It is estimated that the total amount of oil resources contained in ocean areas accounts for 33% of the global total, while the corresponding natural gas resources account for 32% of the world's resources. Current prediction methods, tailored to land oilfields, struggle with offshore differences, hindering accurate forecasts. With oilfield advancements, a vast amount of rapidly generated, complex, and valuable data has piled up. This paper uses AI and GRN-VSN NN to predict offshore oilfield indicators, focusing on model-based formula fitting. It selects highly correlated input indicators for AI-driven prediction of key development metrics. Afterwards, the Shapley additive explanations (SHAP) method was introduced to explain the artificial intelligence model and achieve a reasonable explanation of the measurement's results. In terms of crude-oil extraction degree, the performance levels of the Long Short-Term Memory (LSTM) neural network, BP neural network, and ResNet-50 neural network are compared. LSTM excels in crude-oil extraction prediction due to its monotonicity, enabling continuous time-series forecasting. Artificial intelligence algorithms have good prediction effects on key development indicators of offshore oilfields, and the prediction accuracy exceeds 92%. The SHAP algorithm offers a rationale for AI model parameters, quantifying input indicators' contributions to outputs.

Keywords: offshore oilfield; development indicator prediction; artificial intelligence; SHAP



Citation: Li, K.; Wang, K.; Tang, C.; Pan, Y.; He, Y.; Cai, S.; Chen, S.; Zhou, Y. Prediction of Key Development Indicators for Offshore Oilfields Based on Artificial Intelligence. *Energies* **2024**, *17*, 4594. <https://doi.org/10.3390/en17184594>

Academic Editors: Muhammad Ishaq Bhatti and Mofazzal Hossain

Received: 17 July 2024

Revised: 28 August 2024

Accepted: 4 September 2024

Published: 13 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Oilfield development indicators are data produced along with the production of oilfields to record oilfield production conditions. After years of research by scholars, some indicator data that are closely related to oilfield development effects have been identified, and through the changing rules of these indicator data and possible future change trends, they provide guidance for oilfield development [1–3]. These data are usually divided into static indicators and dynamic indicators based on whether they change with production time. Such static indicators mainly include geological attributes such as structural type, sedimentary type, lithology, interlayer type, porosity, permeability, and saturation. They will change slightly over time, but this change is minimal, and they can be regarded as unchanging static indicators. On the contrary, indicators such as oil production speed, liquid production speed, water consumption rate, and production gas–oil ratio continue to change throughout the entire oilfield production process. Such indicators need to

be collected at fixed time intervals in order to record oilfield production changes, such as in oilfield production daily reports, oilfield production monthly reports, and oilfield production annual reports. In the actual production process, in addition to static indicators and dynamic indicators, special circumstances such as engineering conditions and manual adjustments need to be taken into consideration. In order to eliminate the influence of these factors, the study adopts part of the “Oilfield Development Level Classification” SY-T-6219:2023 [4]. Based on the indicators in this standard, the offshore oilfield development management indicators of the China National Offshore Oil Corporation are introduced, and quantitative characterization of these factors is carried out. The technical indicators include energy maintenance level, water-flooding reserve control degree, water-flooding reserve utilization degree, etc. (Table 1).

Table 1. Classification of oilfield development indicators.

Dynamic Indicators	Static Indicator	Management Indicators
Gas–oil ratio	Oilfield classification	Oil well measures are efficient
Water consumption rate	Sedimentary phase type	Water injection well-injection rate
Production measures to increase oil volume	Reserve abundance	Injection qualification rate of sub-injection well section
Moisture content	Effective thickness	Comprehensive hourly rate of oil and water wells
Adjustments in the number of wells	Porosity	Energy level
Annual oil production	Penetration	Dynamic detection plan completion rate
Oil production rate	Saturation	Waterflooding reserves control degree
	Ground crude-oil density	Water-drive reserve utilization
	Viscosity	
	Reservoir type	
	Drive type	
	Medium-depth reservoir	

Traditional methods are suitable for scenarios with relatively small amounts of data and relatively simple geological conditions. For example, in the exploration stage or the early stage of new oilfield development, traditional methods can quickly provide preliminary prediction results. However, traditional methods have limited ability to deal with complex geological conditions and large amounts of data, and the prediction accuracy of traditional methods may gradually decline as field development deepens and geological conditions change. Artificial intelligence methods are suitable for scenarios with sufficient data volume or complex geological conditions, or those requiring high-precision prediction. In the middle and late stages of oilfield development or scenarios requiring refined management, artificial intelligence methods can provide prediction results which are more accurate.

In response to the above problems, this paper takes the prediction method of key development indicators of oilfields as the research object and carries out research on the prediction of key development indicators of oilfields based on artificial intelligence. By constructing neural network models such as BP and ResNet-50, this paper predicts the degree of oilfield recovery. At the same time, this paper also interprets the model based on the artificial intelligence model interpretation tool SHAP and fits the empirical formulas of each key development indicator for the target oilfield.

2. Key Development Indicator Prediction Method

2.1. Key Development Indicator Prediction Based on Traditional Methods

The prediction of oilfield development indicators is based on the historical data of the oilfield. By studying the historical development indicators of the oilfield, analyzing and

clarifying their changing rules, and combining the existing data to predict the changing trends of future development indicators, timely production adjustments are made based on the prediction results in order to obtain better production and development effects. The traditional prediction methods of oilfield development indicators can be divided into four categories based on their theoretical bases:

2.1.1. Classical Formula Prediction

Based on the classic formula prediction method, a variety of prediction curves have been derived, including decline curves, water-drive curves, injection–production relationship curves, oil and gas ratio curves, etc. Among them, water-drive characteristic curves are classified into four types: A, B, C, and D. The applicability of each water-flooding curve type is related to the viscosity of the crude oil. Specifically, the Type D water-flooding curve is more suitable when the crude-oil viscosity is lower, the Type A (or possibly another type, depending on the specific classification) water-flooding curve is more suitable when the crude-oil viscosity is medium, and the Type B water-flooding curve is more suitable when the crude-oil viscosity is higher. It is more suitable for type A and type C water-flooding characteristic curves [5]. Different curves can only predict specific oilfield development indicators. For example, the oil–gas ratio curve method is used to predict the production oil–gas ratio, while the water-drive curve is employed to predict indicators such as water cut and liquid production. However, the application of these characteristic curves needs to be under specific conditions. Only in this way can better accuracy be achieved. For instance, the water-drive characteristic curve is suitable for the medium-to-high water cut development stage of the oilfield, and the prediction accuracy will be affected in other stages [6]. Zhu Mingxia et al. (2022) [7] used the oil–water two-phase seepage theory to derive a new type of water-drive curve, and greatly improved the prediction accuracy for water content and geological reserves, reducing the average relative error levels to 2.1% and 5.2%. Zhu Lang (2022) [8] constructed a set of water-flooding characteristic curves for activated water flooding of heavy oil in offshore oilfields and used this characteristic curve to predict the increase in oil values. Deng Jingfu (2023) [9] used multiple nonlinear regression to fit and predict the production, water cut, and decline of the Bohai S oilfield. Zhang Jianda (2024) [10] combined the phase permeability curve and the water-flooding curve to predict oil production and water content. Ma Chao et al. (2022) [11] used water-drive characteristic curves to predict the recovery factor of an offshore oilfield.

2.1.2. Prediction Using the Hydrodynamic Formula Method

The hydrodynamic formula method is based on fluid-mechanics formulas. This part of the prediction method encompasses the seepage mechanics prediction model, the equivalent seepage resistance model, the piston flow method, and the non-piston flow method. This prediction method is primarily used in the early stage of oilfield development. Although it has a solid theoretical basis, when it is combined with practice, the prediction effects will greatly differ due to variations in actual conditions. Zong Huifeng (2007) [12] used the hydrodynamic formula to optimize the development-based effects of water-driven oilfields and effectively improve the recovery rate. Guo Wenmin (2016) [13] combined the hydrodynamic control improvement measures and other oilfield characteristics to construct a hydrodynamic method for injection and production regulation in the ultra-high-water-content period to improve the water-drive control intensity. Gao Min (2021) [14] optimized the development method for the high-water content period by using fault-block reservoirs based on hydrodynamics (Table 2).

Table 2. Classic formula.

Formula Name	Functional Relationship
Type A water-drive curve	$\log Wp = a + b \times Np$
Type B water-drive curve	$\log Lp = a + b \times Np$
Type C water-drive curve	$Lp/Np = a + b \times Lp$
Type D water-drive curve	$Lp/Np = a + b \times Np$
Hyperbolic decline curve	$q(t) = q' \left(1 + nDt\right)^{\frac{1}{n}}$
Injection–production relationship curve	$\lg(W_I - F) = C + DN_p$

2.1.3. The Material Balance Equation Method of Prediction

The material balance equation method considers the oilfield development process as a container, in which oil, gas, and water are the substances. This method postulates that these three substances always adhere to the material balance equation throughout the entire development process [15]. In the prediction process, the material balance equation is divided into the material balance equation for closed elastic drive reservoirs and the material balance equation for unclosed elastic drive reservoirs. The principle of using the material balance equation for prediction is simple, but its effectiveness in refined predictions is not very ideal. Wang Di et al. (2021) [16] constructed the material balance equation of a buried-hill condensate gas reservoir using the material balance principle and estimated the dynamic reserves of the corresponding work area. Gu Hao et al. (2022) [17] modified the material balance equation of the ultra-deep reservoir, estimated the dynamic geological reserves of the work area, and predicted the change of dynamic geological reserves after the reservoir pressure drop increased (Table 3).

Table 3. Material balance equation formula.

Formula Name	Functional Relationship	Describe
Material balance of closed elastic flooding reservoir	$N_p B_o = \Delta V_W + \Delta V_p + \Delta O$	Elastic cumulative oil production = expansion volume of crude oil + expansion volume of bound water + shrinkage volume of rock pores
	$N_p B_o + W_p B_W = CB_{oi} N \Delta p + W_e$	Cumulative oil production of the reservoir + cumulative water production = total elastic expansion of the reservoir + edge water intrusion

2.1.4. Reservoir Numerical Simulation Prediction

Reservoir numerical simulation employs computers to solve mathematical models of oilfields, simulating the flows of oil and water within underground reservoirs. Through model selection, sensitivity testing, data input, and history matching, oilfield development indicators such as water cut and production can be dynamically predicted [18]. This method can simulate the oil and water flow in various heterogeneous reservoirs and is suitable for development planning and adjustment.

2.2. Development Indicator Prediction Method Combined with Artificial Intelligence

With the continuous development of artificial intelligence, oilfield development indicator prediction methods combined with artificial intelligence have been continuously proposed in recent years. Han Rong et al. (2000) [19] proposed a method using a BP neural network to quickly predict the production of a single well in an oilfield. The predicted results from the prediction model for the liquid production, oil production and gas production of the oil well show that the BP neural network prediction model can improve the prediction accuracy of the liquid production, oil production and gas production of the

oil well; Ren Baosheng (2008) [20] proposed an insensitive support vector machine and introduced it into the prediction model of oilfield development dynamic indicators, which effectively solved the overfitting problem caused by using too limited an amount of sample data in the traditional method and effectively improved the generalization ability of the model; Ma Linmao et al. (2015) [21] used a genetic algorithm to optimize the BP neural network and applied it to the production prediction of the high water-content period in the BED test area of Daqing Oilfield; Zhao Ling et al. (2018) [22] proposed a process support vector regression machine algorithm (PSVR) for process parameter optimization using the turbine algorithm, predicted the liquid production and water content, and obtained good prediction results; Zhang Yuhang (2016) [23] proposed an improved particle swarm discrete process neural network model through comparative research, predicted the oil production and liquid production of the oilfield, and obtained better prediction results; Chen Chenglong (2022) [24] used the BP neural network, improved by a genetic algorithm, to predict the water content, cumulative oil production, and recovery rates of production wells in the eastern transition zone of North Zone 1 of the Sazhong Development Zone of Daqing Oilfield, and potential wells were identified based on this prediction; Zhong Yihua et al. (2020) [25] used deep learning convolutional neural networks and recurrent neural networks to mine the reservoir-characteristic patterns and development-dynamic change laws of the oilfield development system, and, using the ELMO-based residual multi-head selection joint extraction model of deep learning entities and relationships, proposed a method, knowledge base, and model library for mining the best prediction model based on reservoir type and development stage; Li Tiening (2016) [26] used the Elman network, optimized by an improved genetic algorithm, to predict the water content of a single well, and used the double hidden layer process neural network combined with particle swarm algorithm to predict the oilfield production; Ha and Nguyen et al. (2002) [27] used an MNN neural network to predict monthly production; in Hu et al. (2019) [28], the GRU neural network, improved by principal component analysis, was used to predict oil production, and compared with the BP neural network, as improved by principal component analysis, with the results showing that the PCA-GRU neural network achieved higher accuracy. Dang Chen (2023) [29] used LSTM, the LSTM algorithm as improved by genetic algorithm, and a particle swarm optimization algorithm to develop a warning model for four oilfield development indicators. Zhu Bilei (2024) [30] constructed a CNN-Bi LSTM oil production and water content prediction model. Qu Qing (2024) [31] used the C bi GRU-Attention model to predict oil production and water production. Sun Dongming (2021) [32] predicted oilfield development indicators based on a radial basis process neural network. Fan Sen (2023) [33] predicted the injection volume of stratified water injection based on CNN-LSTM (Table 4).

Table 4. Development indicator prediction methods combined with artificial intelligence.

Researchers	Predictive Indicators	Method
Han Rong et al. (2000) [19]	Oil well liquid production, oil production, gas production	BP neural network
Ma Linmao et al. (2015) [21]	Yield prediction during high water content period	GA-BP
Zhang Yuhang (2016) [23]	Oil production, liquid production	Improved discrete process neural network model using particle swarm
Li Tiening (2016) [26]	Moisture content, oilfield production	Elman network optimized by improved genetic algorithm, Double hidden layer process neuron network combined with particle swarm algorithm

Table 4. Cont.

Researchers	Predictive Indicators	Method
Zhao Ling et al. (2018) [22]	Liquid production, moisture content	Turbine algorithm optimization, Process support vector regression machine algorithm (PSVR)
Chen Chenglong (2022) [24]	Water content, cumulative oil production, recovery factor	GA-BP
Ha et al. (2002) [27]	Monthly production	MNN neural network
Hu et al. (2019) [28]	Oil production	GRU neural network improved by principal component analysis

3. Feature Correlation Analysis and Key Development Indicator Prediction Algorithm

3.1. Feature Correlation Analysis

In the actual prediction processes for key development indicators, a greater number of development indicators participating in the prediction does not mean higher accuracy. The participation of low-correlation or non-correlation development indicators will reduce the prediction accuracy values for key development indicators. Therefore, before the key development indicator prediction work is carried out, it is also necessary to conduct a feature correlation analysis between the selected indicator system and the key development indicators. Currently, commonly used correlation analysis methods include conventional calculation methods such as the correlation coefficient method and grey correlation analysis method. There are also feature correlation analysis algorithms based on deep learning.

3.1.1. Correlation Coefficient Method

The correlation coefficient method usually evaluates the correlation between features by constructing a linear functional relationship between two features. However, in the actual application process, it will be found that the correlation between features is not always a simple linear relationship, but also includes many other types of correlations, such as exponential correlation and polynomial correlation. Therefore, there are many ways to calculate the correlation coefficient. Commonly used correlation coefficients include the Pearson correlation coefficient, Spearman rank correlation coefficient, and Kendall rank correlation coefficient, etc.

Pearson's correlation coefficient is calculated in such a way that, firstly, the mean values of features X , Y need to be calculated, and then the covariance $Cov(X, Y)$ between the features X , Y as well as the standard deviation σ_x and σ_y for each of X , Y are computed, and the correlation coefficients are obtained:

$$E(X) = \frac{\sum_{i=1}^n x_i}{n}, E(Y) = \frac{\sum_{i=1}^n y_i}{n} \quad (1)$$

$$Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - E(X))(Y_i - E(Y))}{n} \quad (2)$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (X_i - E(X))^2}{n}}, \sigma_y = \sqrt{\frac{\sum_{i=1}^n (Y_i - E(Y))^2}{n}} \quad (3)$$

$$Pearson = \frac{Cov(X, Y)}{\sigma_x \sigma_y} \quad (4)$$

The calculation of the Spearman's correlation coefficient needs to be evaluated first by ranking the data separately, i.e., ranking the size of each data element after arranging them

in an order from smallest to largest, and calculating the correlation coefficient through the grade difference of the corresponding positions of the two groups of data:

$$d_i = X_i - Y_i \quad (5)$$

$$Spearman = 1 - \frac{6 \times \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (6)$$

Kendall's rank correlation coefficient evaluates the correlation by measuring the consistency of the ranking or ordering of the variables, similar to Spearman's rank correlation coefficient. However, the key difference lies in how they assess the relationship between data pairs. Spearman's rank correlation coefficient considers the relative distance error between the ranks of corresponding data pairs, meaning it focuses on the differences in the positions of the data pairs in their respective rankings. In contrast, Kendall's rank correlation coefficient assesses the concordance of the changes in the rankings of the two sets of variables, specifically by counting the number of concordant and discordant pairs. A concordant pair refers to a situation in which the ranks of the corresponding elements in two sets of variables agree in their ordering (i.e., both increase or both decrease), whereas a discordant pair refers to a situation in which their ordering disagrees. Kendall's tau coefficient is calculated based on the difference between the number of concordant pairs and the number of discordant pairs, divided by the total number of possible pairs. This approach makes Kendall's coefficient less sensitive to outliers and more robust in certain scenarios compared to Spearman's coefficient. For two random variables X, Y take the corresponding data pairs (x_i, y_i) and (x_j, y_j) , where $i < j$. If $x_i < x_j$ and $y_i > y_j$ or $x_i > x_j$ and $y_i > y_j$, it is determined that there is a consistency between this group of variables; otherwise there is no consistency. As for the special case of $x_i = x_j$ and $y_i = y_j$, it is considered inconsistent and not contradictory. For Kendall's correlation coefficient the formula is as follows:

$$Kendall = \frac{n_c - n_d}{n(n-1)/2} \quad (7)$$

where n_c is the number of data pairs with consistency, n_d is the number of data pairs without consistency, and n is the total number of data pairs.

3.1.2. Grey Correlation Analysis

Grey relational analysis (GRA) is a method in grey system theory [34] which can be used to determine the correlations between various indicators. Unlike the correlation coefficient method, which analyzes data pairwise, the grey relational analysis method treats all indicators as a whole, and the analysis of correlation is based on the whole system [35–38]. Grey relational analysis requires, first, clarifying the parent sequence and the subsequence. In this study, the parent sequence is the key development indicator, and the subsequences are the selected development indicators within the system. After clarifying the parent sequence and the subsequences, each sequence needs to be de-dimensionalized using normalization and averaging methods, and then the correlation coefficients are calculated.

Firstly, it is necessary to perform the absolute difference calculation for the subsequence and the parent sequence after preprocessing, respectively, i.e., $|x_i - x_0|, i \in [1, n]$, x_0 is the value of the parent sequence. Interpolation of all data needs to be filtered to determine the global minimum and global maximum:

$$a = \min_i \min_j |x_0(k) - x_i(k)| \quad (8)$$

$$b = \max_i \max_j |x_0(k) - x_i(k)| \quad (9)$$

The grey correlation coefficient is calculated:

$$\zeta_i(k) = \frac{a + \rho \cdot b}{|x_0(k) - x_i(k)| + \rho \cdot b} \quad (10)$$

where ρ is the discrimination coefficient, taking a value between $[0, 1]$, and usually taking the value of 0.5. The mean value of $\zeta(k)$ obtained after the calculation is calculated subsequently, and the calculated value is the grey correlation between the current sequence and the parent sequence. The closer the grey correlation is to 1, the stronger the correlation between the two variables.

3.1.3. Artificial Intelligence Correlation Analysis

The correlation between features can also be analyzed using an artificial neural network model. Typically, a specific neural network is employed to assess the feature contribution and weight between input indicators and output indicators, and this then serves as the basis for evaluating feature correlation. In this article, we utilize a method proposed by Oxford University and Google, one based on Gated Residual Networks (GRN) and Variable Selection Networks (VSN), to perform feature correlation analysis.

The process of using neural networks to conduct feature correlation analysis is similar to the process of predicting key development indicators. First, the dataset needs to be preprocessed. The processed data then enter the Variable Selection Network (VSN), where variables are selected, and multiple groups of training subsets are created. These subsets are then put into the Gated Residual Network (GRN) for training, and finally, the importance of each feature is judged based on the accuracy of all subsets during verification (Figure 1).

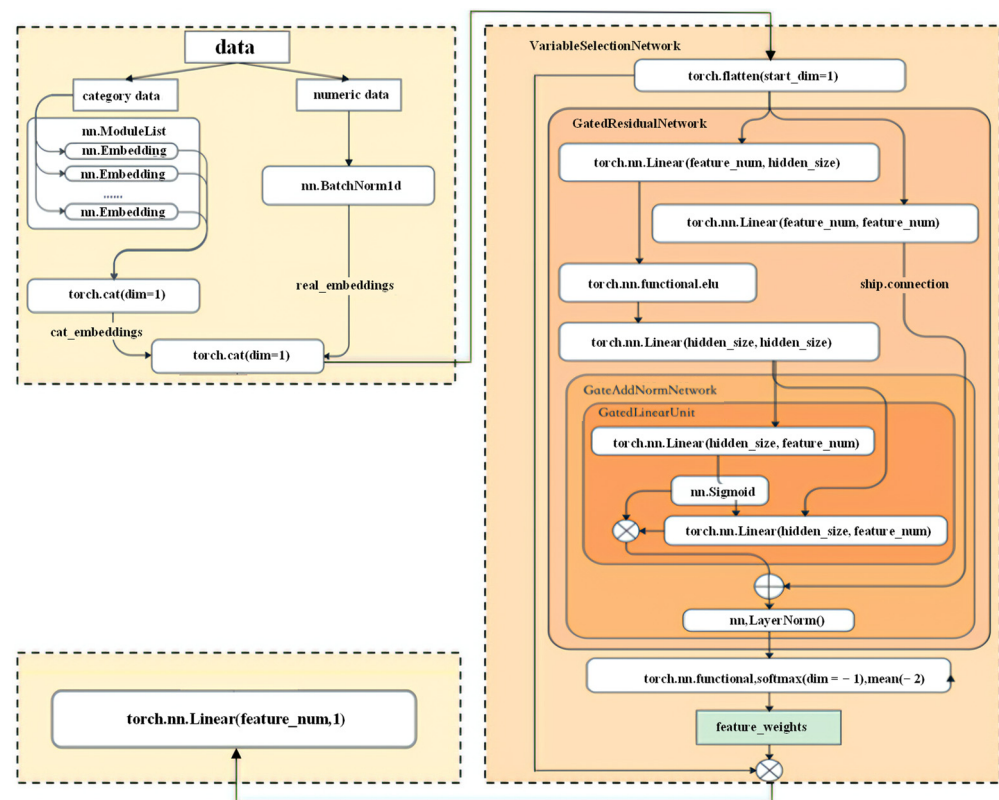


Figure 1. GRN-VSN algorithm flow.

The GRN controls the residual module by introducing the concept of a gate, which can effectively prevent the problems of gradient descent stagnation and gradient explosion, ensuring the accuracy of the constructed function. Additionally, it plays a precise role in constructing complex functions. Typically, the gate is composed of a Sigmoid function.

3.2. Key Development Indicator Prediction Algorithm

3.2.1. Residual Network (Res Net)

Res Net was proposed by Microsoft Research in 2015. The algorithm was initially used for classification and object detection. Compared with the traditional convolutional neural network (CNN), it introduces a residual structure module and batch normalization to address the problem of model degradation. The residual is the difference between the observed value and the predicted value. The residual neural network primarily utilizes CNN to extract data features. Compared with the traditional fully connected neural network, the convolutional neural network introduces the concepts of a convolutional layer and a pooling layer, which can accelerate the convergence speed of the network.

In the early days, it was widely believed that the more convolutional and pooling layers a CNN had, the better the model would perform. However, in actual research, it was discovered that an increase in the number of layers did not necessarily improve the accuracy, and could instead degrade the overall performance of the model. This phenomenon is known as model degradation. The reason for this is that as the number of layers increases, the gradient can gradually diminish during back-propagation, rendering the model unable to effectively adjust its weights. In response to this phenomenon, Res Net introduced short-circuit connections to solve the problem of model degradation, as shown in Figure 2, below.

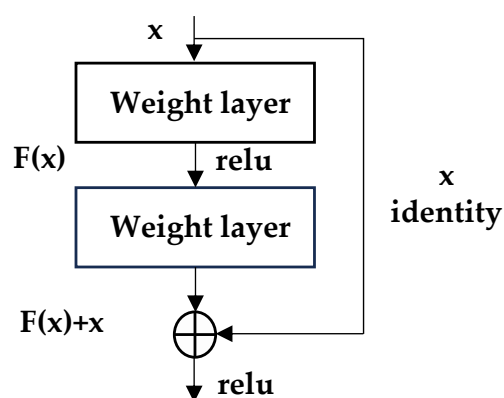


Figure 2. Res Net residual module.

In the two hidden layers depicted in the figure, Res Net adds a shortcut connection before the activation function in the second hidden layer, causing the input value of the activation function to change from the original $F(x)$ to $F(x) + x$. This ensures that the network can continue to learn effectively, even when $F(x)$ approaches zero. This setting helps to avoid the problem of network degradation. Secondly, Res Net also employs batch normalization to replace the original global normalization algorithm, normalizing the same batch of data input into the network for each training iteration. This can mitigate the issues of gradient vanishing or gradient explosion. The introduction of shortcut connections prevents Res Net from degrading in performance when stacking more layers, allowing Res Net to have ultra-deep architectures which are unmatched by other networks.

Residual networks are able to construct a very deep network structure by stacking multiple residual blocks to learn richer feature representations. These feature representations can better reflect the complex relationship between oil recovery and various influencing factors. Residual networks can adjust the network structure and parameters according to the needs of specific tasks in order to adapt to different datasets and prediction goals. This flexibility makes residual networks more advantageous in complex tasks such as oil recovery rate prediction.

3.2.2. Long Short-Term Memory (LSTM)

LSTM is based on the recurrent neural network (RNN) and adds memory units to each hidden layer neural unit to achieve controllable memory of information in a time series. It is suitable for processing and predicting important events with relatively long intervals and delays in time series [39,40].

The LSTM neural network features a repeating chain structure. The hidden neurons in LSTM differ from those in the single neural network layer of the RNN chain structure, as LSTM incorporates four distinct neural network layers. The relationship between these four neural network layers is intricate. The specific internal structure is illustrated in Figure 3. The red part represents the neural network layers, the yellow part represents operational symbols, the plus sign module indicates vector addition, and the arrow lines depict the transfer of vector information.

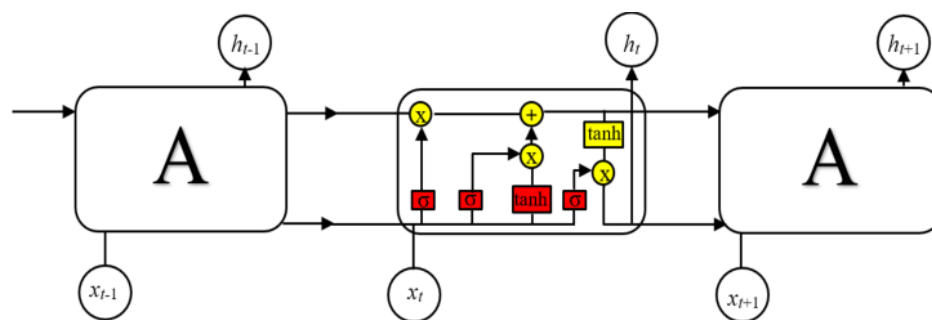


Figure 3. LSTM repeated chain structure diagram.

Oil recovery prediction involves a large amount of time-series data, such as oil well production, water injection, formation pressure, etc. These data vary over time and there may be dependencies between the data at different time points. An LSTM network, as a special recurrent neural network, is particularly good at dealing with this kind of time-series data, and is able to capture the long-term dependencies in the data, thus predicting the future crude-oil recovery rate more accurately. There are complex nonlinear relationships between oil recovery rates and a variety of geological, engineering-based, and economic factors, and LSTM networks, through their internal memory units and gating mechanisms, can learn and model these nonlinear relationships in order to more accurately reflect the actual situation.

3.2.3. Back-Propagation Neural Network (Back-Propagation, BP)

The BP neural network (back-propagation) [41], also known as the back-propagation neural network, is named for its use of the gradient descent method to modify the weights and biases of each node based on the error computed after each training iteration, until an optimal result is obtained. The BP neural network has a simple structure and consists of three main components: the input layer, the hidden layer(s), and the output layer. It was first proposed in 1986 and has since achieved good results in many fields (Figure 4).

BP neural networks are known for their ability to handle complex nonlinear relationships by means of powerful nonlinear mapping. In crude-oil recovery prediction, there are often complex nonlinear relationships among geological, engineering-based, and economic factors, and BP neural networks can effectively capture these relationships in order to predict crude-oil recovery more accurately. The BP neural network has the abilities of self-learning and self-adaptation, and it can automatically adjust the network parameters by learning the laws in the training samples, so it can quickly adapt and give accurate prediction results when facing a new oilfield or new production conditions.

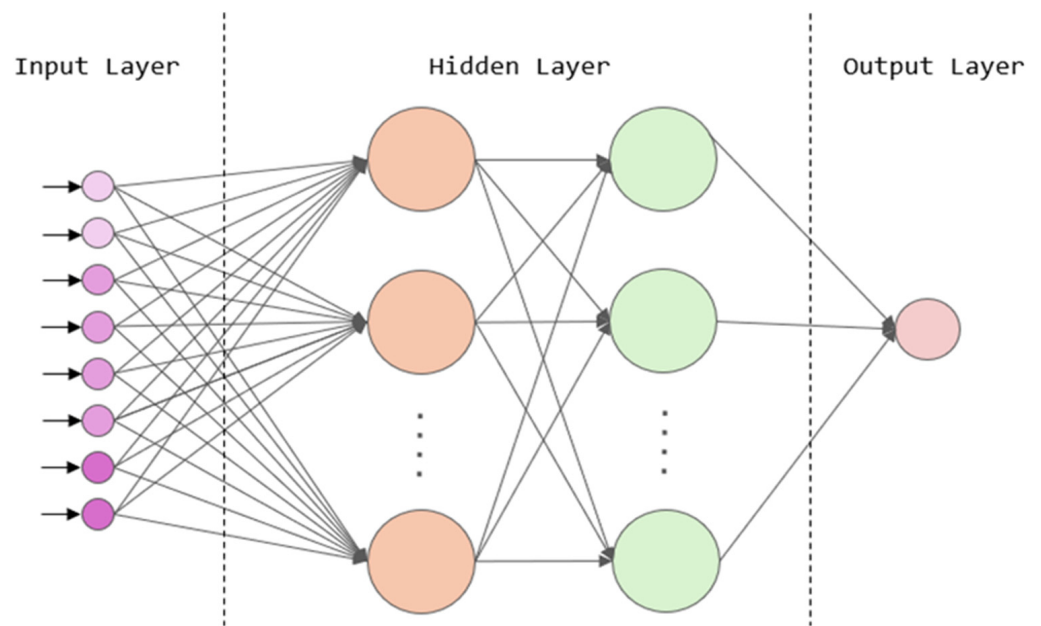


Figure 4. BP neural network architecture.

In order to prevent the BP neural network model from falling into a local optimal solution instead of the global optimal solution, and to enhance the accuracy and robustness of the particle swarm model, the genetic algorithm (GA) and optimization (PSO) algorithm were employed to optimize the BP neural network (Figure 5). As a result, the GA-BP and PSO-BP models were developed and subsequently applied to the prediction of key development indicators (Figure 6).

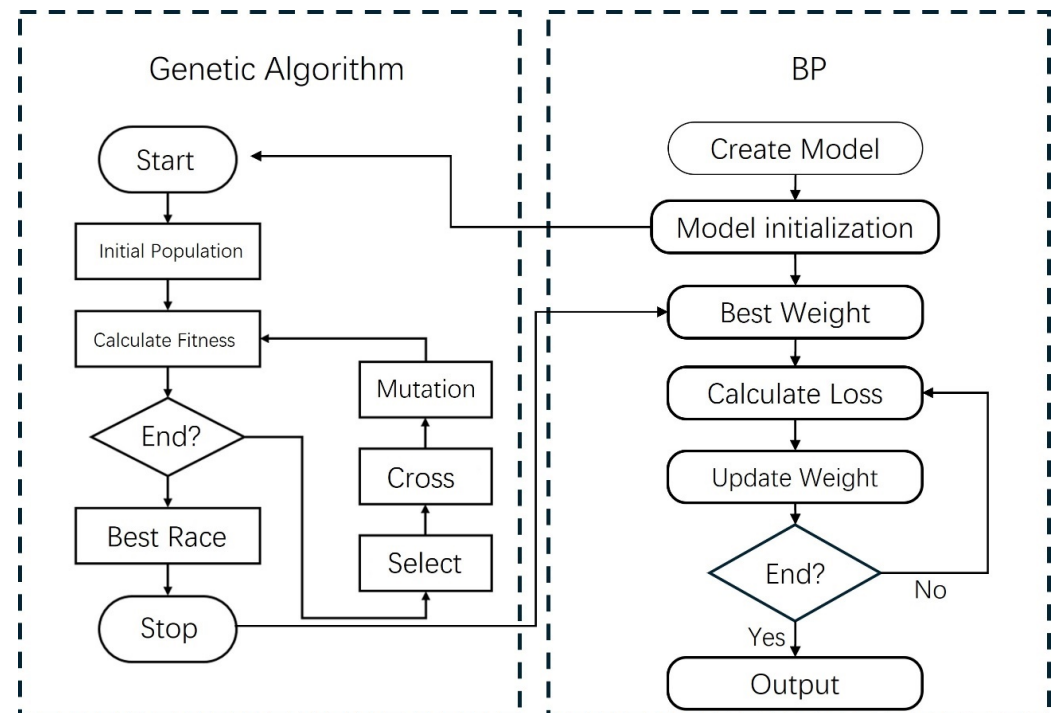


Figure 5. Genetic algorithm optimization neural network process.

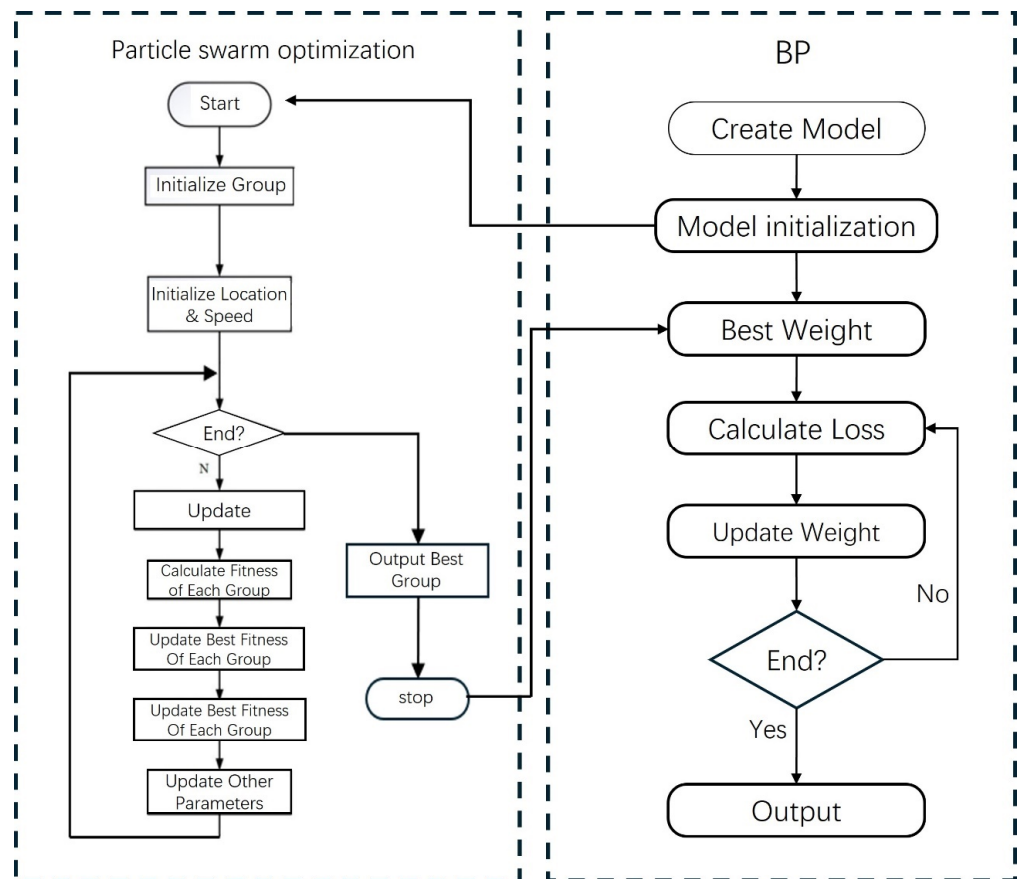


Figure 6. Particle swarm optimization algorithm process.

4. Forecast of Key Development Indicators

The degree of crude-oil recovery serves as one of the crucial indicators for assessing the effectiveness of oilfield development. Predicting this value can provide insights into both the current development level of the oilfield and its subsequent development potential.

4.1. Data Preprocessing

The selected dataset needs to be preprocessed first; the specific process includes data interpolation, data cleaning, and discretization.

4.1.1. Data Interpolation

The purpose of data interpolation is to eliminate the possibility of accidental missing values in the data. There is no need to interpolate the missing data before the first valid data point or after the last valid data point. A scientific and reliable missing-data interpolation method not only does not affect the overall accuracy of the data, but also significantly improves the learning accuracy of the deep learning model. Currently, the commonly used data interpolation methods include linear interpolation, least squares interpolation, and inverse distance interpolation.

In this paper, inverse distance interpolation is employed to interpolate missing data. This method utilizes the inverse of the distance between a known point and an unknown point as a factor that influences the weights assigned to the known points. The data value at the unknown point is then determined by computing a weighted sum of the data values at the known points, where the weights are determined as described below:

$$\omega_i = \frac{\frac{1}{d_i}}{\sum_{i=1}^n \frac{1}{d_i}} \quad (11)$$

$$y = \sum_{i=1}^n \omega_i x_i \quad (12)$$

In the above equation, d_i is the distance from the known point to the unknown point and x_i is the actual value of the known point.

4.1.2. Data Cleaning

In the deep learning process, the input model receives data consisting of one or more input indicators. The deep learning model then studies the relationship between these data and the indicator data to be predicted. The data cleaning process primarily focuses on identifying and handling invalid data and outliers. Invalid data typically refers to data with too many missing values within a single piece of data, or which cannot effectively reflect the correlation between the input indicators and the indicators to be predicted due to the limited amount of valid data. When such data is input into the model for learning, it can negatively impact the model's accuracy. Outliers, on the other hand, are values that significantly deviate from the normal pattern of change in the data, and their inclusion in the model can also hinder its learning ability.

4.1.3. Discretization

The purpose of de-scaling (or normalization) is to eliminate the magnitude differences between different data points, as the order-of-magnitude differences between the data can directly affect the weight allocation in deep learning, thereby seriously impacting the accuracy and robustness of the deep model. Data de-scaling methods typically involve normalization or homogenization.

The purpose of data normalization is to uniformly scale each data point to the range [0, 1], thereby unifying the influence size of each indicator at the initial stage to a common level. This helps the algorithm converge to the optimal solution more quickly. At the same time, it is necessary to record the maximum and minimum values of different indicators for a subsequent anti-normalization operation. This process, known as anti-normalization, aims to restore the prediction results to their original scale, enabling better comparison with real data and thereby assessing the model's prediction accuracy.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (13)$$

4.2. Calculation of Indicator Correlation

The selected indicators and crude-oil recovery degree data are analyzed using two correlation detection algorithms: grey correlation analysis and the GRN-VSN neural network. The correlations calculated by the two methods are combined to determine the indicators used for oil recovery prediction. The optimization is performed, and the indicators with low correlation are discarded, with the following results (Figure 7).

After performing grey correlation analysis, the correlations between dynamic indicators, static indicators, and management indicators, as well as the degree of crude-oil production, were obtained. Among the dynamic indicators, A7 and A5 are the two most-correlated indicators, namely, oil production speed and the number of adjusted wells, with correlation degrees of 0.8197 and 0.8212, respectively. However, the correlations of other indicators, excluding water content, are also strong. Among the static indicators, B9, B12, B3, and B4 constitute the top 30% of the correlated indicators, and correspond to reserve abundance, formation type, viscosity, and effective thickness. Among these, viscosity has the highest correlation, at 0.791. For the management indicators, the two indicators with the highest correlation are C6 and C3, namely, the completion rate of the dynamic detection plan and the injection qualification rate of the sub-injection well section, with correlation degrees of 0.812 and 0.804, respectively (Figure 8).

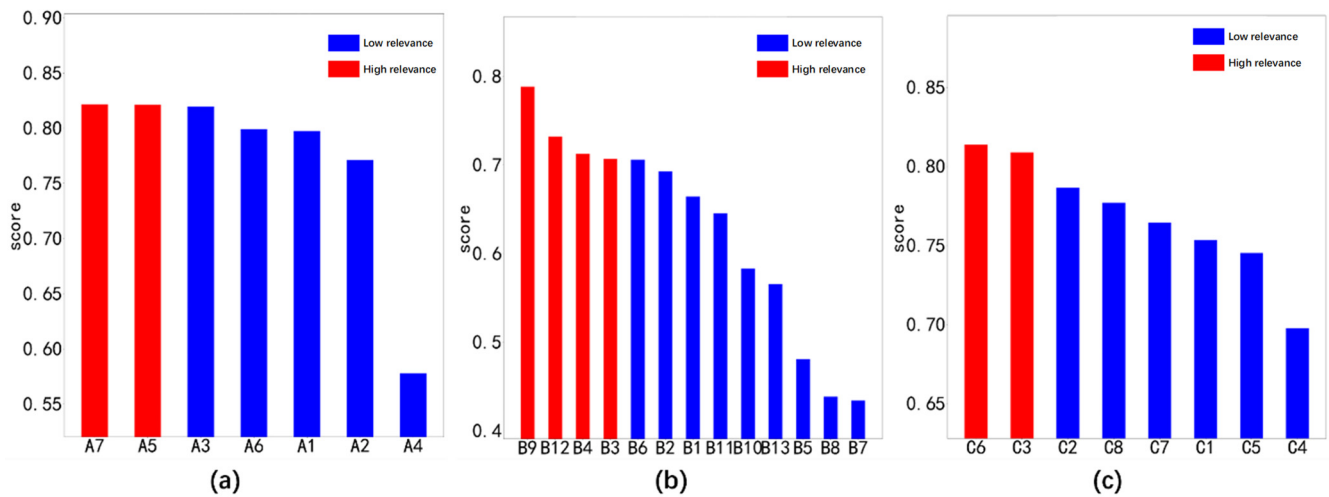


Figure 7. Grey correlation algorithm correlation calculation results of crude-oil recovery degree; (a): dynamic indicators, (b): static indicators, and (c): management indicators.

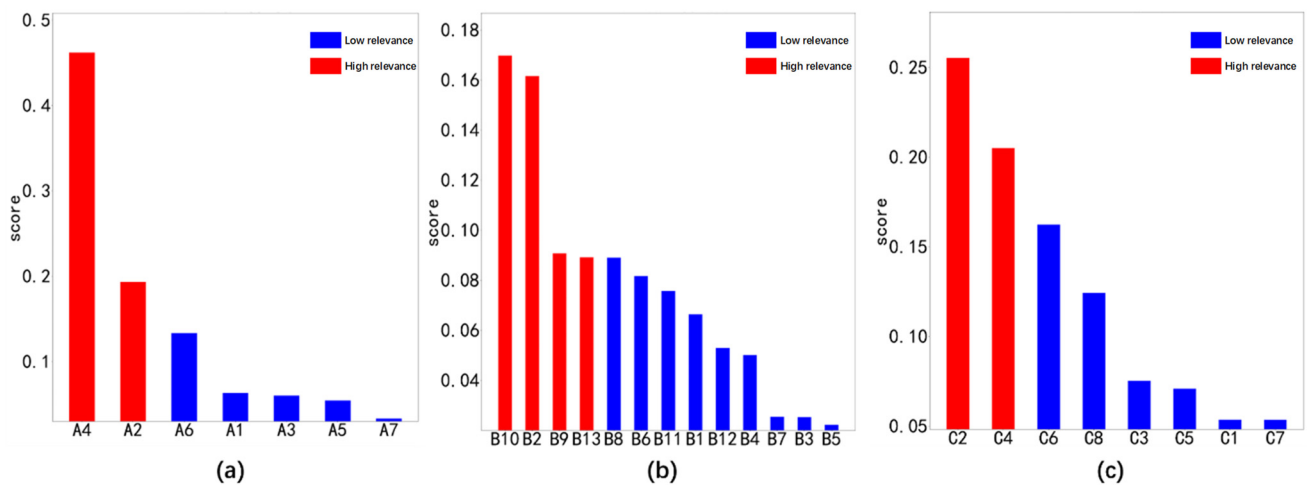


Figure 8. GRN-VSN neural network crude-oil recovery degree correlation calculation results; (a): dynamic indicators, (b): static indicators, and (c): management indicators.

The correlation results obtained through GRN-VSN neural network analysis differ significantly from those derived using the grey correlation algorithm. Through GRN-VSN neural network analysis, it is found that the dynamic indicators with the highest correlation are water content and water consumption rate. The static indicators with the strongest correlation are sedimentary facies type, viscosity, reservoir type, and reservoir depth. The management indicators with the greatest correlation are the water injection well-injection rate and the oil and water well comprehensive hourly rate.

Since the two algorithms have very different analytical results of indicator correlations, the two algorithms are combined, and the one or more indicators with the greatest correlation are used to predict the degree of crude-oil production. The results are as follows (Figure 9):

In the dynamic indicators, there are seven kinds of indicators: gas–oil ratio (A1), water consumption rate (A2), production measures to increase the amount of oil (A3), water content (A4), the number of wells to be adjusted (A5), annual oil production (A6), and recovery speed (A7). From among these, the water content (A4) and annual oil production (A6) were selected as input indicators.

There are thirteen static indicators, including field classification (B1), sedimentary phase type (B2), viscosity (B3), effective thickness (B4), porosity (B5), permeability (B6), saturation (B7), surface crude-oil density (B8), reserve abundance (B9), surface crude-oil

density (B10), drive type (B11), stratigraphic layer (B12), medium-depth reservoir (B13), and effective thickness (B14). From among these, viscosity (B3), reserve abundance (B9), mid-depth reservoir (B13), and effective thickness (B14) are selected as input indicators.

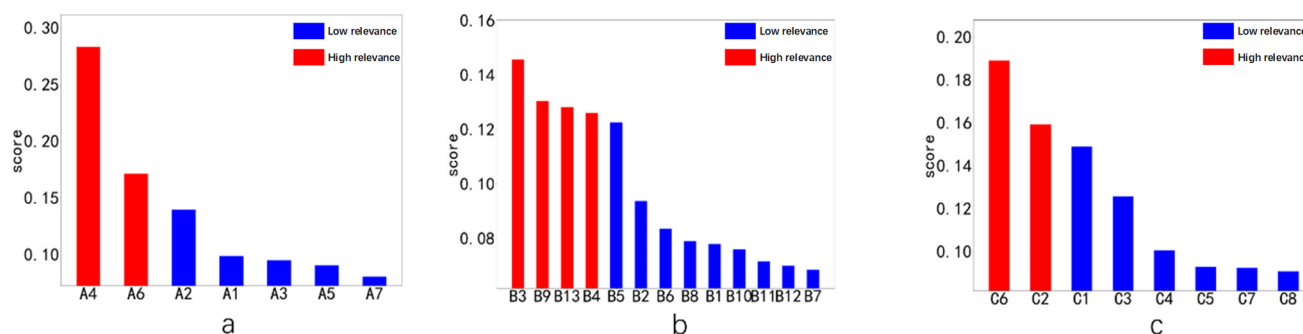


Figure 9. Correlations between post-mean indicators and crude-oil production degree; (a): dynamic indicators, (b): static indicators, (c): management indicators.

There are eight kinds of management indices, including the effective rate of well measures (C1), the injection rate of water injection wells (C2), the qualified rate of injection in the layer section of water injection wells (C3), the integrated time rate of oil and water wells (C4), the energy retention level (C5), the rate of completion of the dynamic testing program (C6), the degree of control of water-driven reserves (C7), and the degree of water-driven reserves utilization (C8); from among these, the injection rate of water injection wells (C2) and the completion rate of dynamic testing program (C6) are selected as the input indices. The completion rate of the dynamic testing program (C6) was also selected as an input indicator.

Based on the average results, the final indicators for the degree of crude-oil recovery are selected, as shown in Table 5:

Table 5. Selection of prediction indicators for crude-oil recovery degree.

Indicator Type	Select Indicator
Dynamic indicators	Water content (A4), annual oil production (A6)
Static indicators	Viscosity (B3), reserve abundance (B9), medium-depth reservoir (B13), effective thickness (B14)
Management indicators	Dynamic detection plan completion rate (C2), water injection well-injection rate (C6)

4.3. Index Prediction Results

The optimized indicator data are divided into a training set and a prediction set, with the training set comprising 70% of the total data volume. Different network models are trained, and the efficiency and accuracy of each network are statistically analyzed (Table 6).

The learning rate is 10×10^{-4} , the batch size is 16, and the optimizer used is Adam.

Table 6. Comparison of algorithm running time and accuracy at different crude-oil recovery levels.

Model Name	Iterations	Run Time	Error (RMSE)
LSTM	500	600 s	0.003
Res Net	2000	550 s	0.015
BP	2000	400 s	0.028
GA-BP	2000	650 s	0.024
PSO-BP	2000	630 s	0.036

The crude-oil recovery rate of the oilfield is predicted, and the results are shown in the figure below (Figure 10).

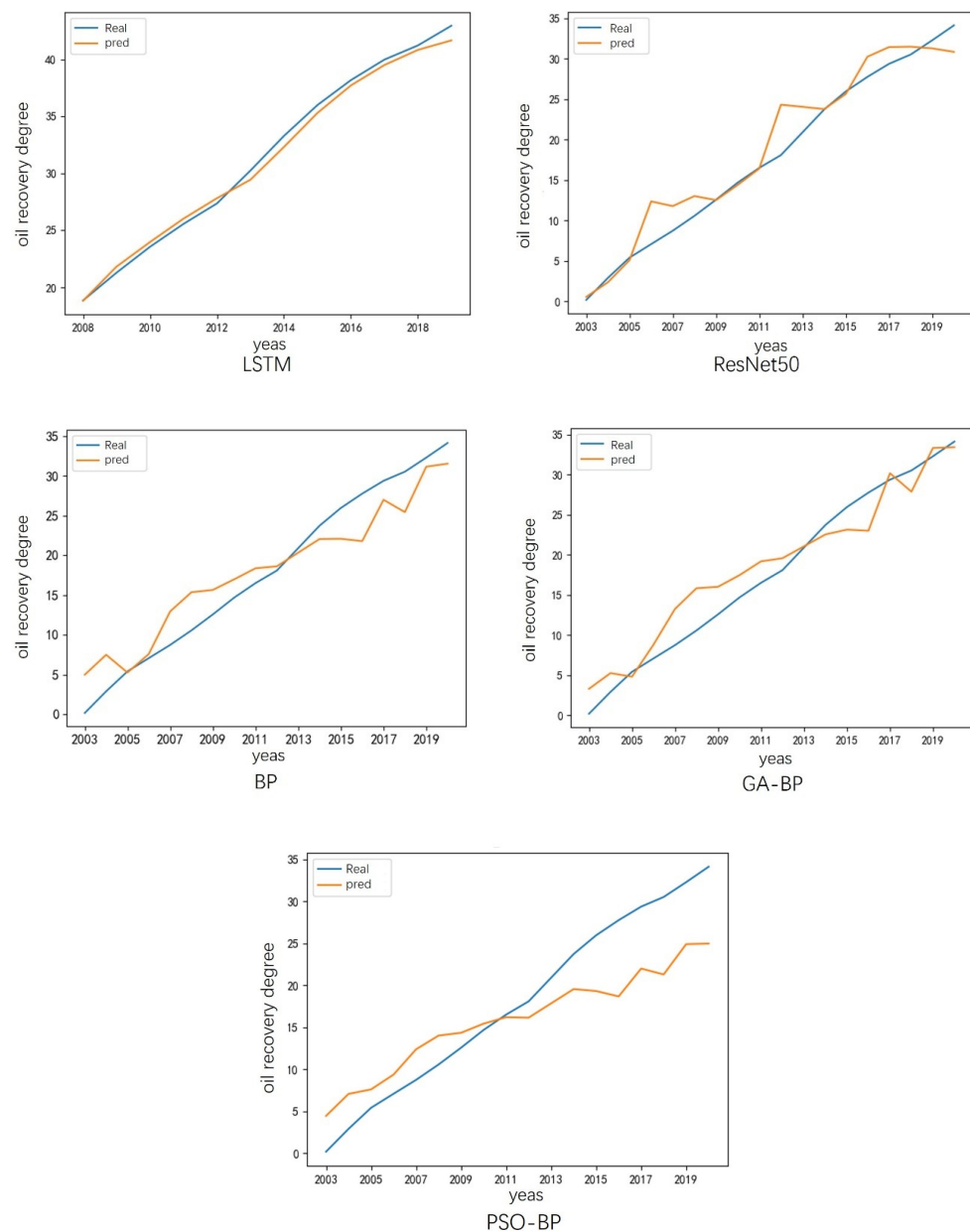


Figure 10. Prediction results of crude oil recovery degree of different models.

For the prediction of crude-oil production levels, the LSTM neural network exhibits significant advantages. This is due to the fact that crude-oil production levels are monotonically increasing. For LSTM, the learning of the index change pattern is relatively straightforward, enabling it to achieve better prediction results. The accuracy levels of ResNet-50 and the BP neural network are similar, but ResNet-50 achieves slightly higher accuracy. By applying genetic algorithm optimization and particle swarm optimization to the BP neural network, the accuracy of the BP neural network optimized using the genetic algorithm has been significantly improved. However, the accuracy of the BP neural network based on the particle swarm optimization algorithm has decreased. This may be related to the particle swarm algorithm's tendency to fall into a local optimal solution.

4.4. Model Interpretability Analysis Based on the SHAP Algorithm

First, each model needs to be explained using the SHAP explanation tool. However, during the actual research process, it was found that some models have a greater impact on the interpretability of features due to their inherent learning mechanisms. The LSTM

model, while having the best prediction effect, poses challenges due to its loop structure and gate settings in the training process. The functional relationship corresponding to the model's internal workings is extremely complex, and the relationship between features cannot be well-captured when interpreting the model. ResNet-50, which has the second-highest prediction accuracy, requires convolution and pooling of features in its training process, merging the features into new features and continuing its learning. During the pooling process, since the selection of features is random and nonlinear (e.g., max pooling, min pooling, median pooling), this down sampling method can significantly affect the reconstruction of the original indicator characteristics. This, in turn, impacts the correlation between the original input indicators and the predictive indicators (Figure 11). Therefore, in the actual relationship function fitting process, the BP neural network is selected for function fitting. In the learning process, the BP neural network has clear formula methods for feature conversion and new feature generation, making it more mathematically rigorous and theoretically supported in the explanation process.

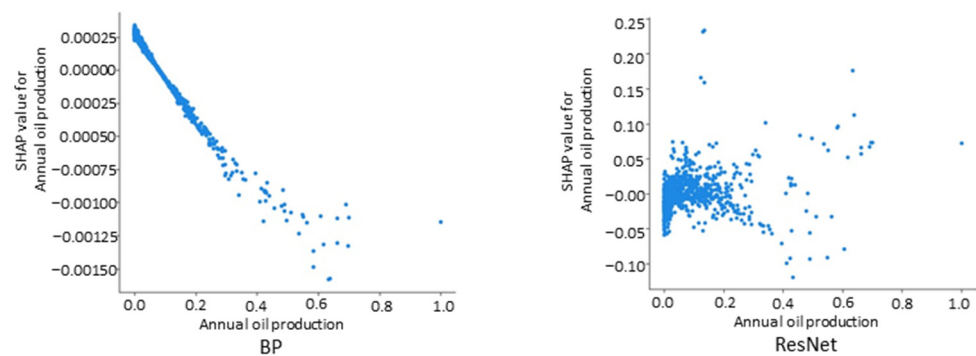


Figure 11. Comparison of the correlation between the Res Net model and BP model as to total annual oil production.

This article uses the GA-BP model, optimized by genetic algorithm, to explain the degree of crude-oil recovery. GA-BP also has high accuracy in prediction accuracy. First, the contribution proportion of each indicator in the overall prediction process is obtained, as shown in the figure below (Figure 12).

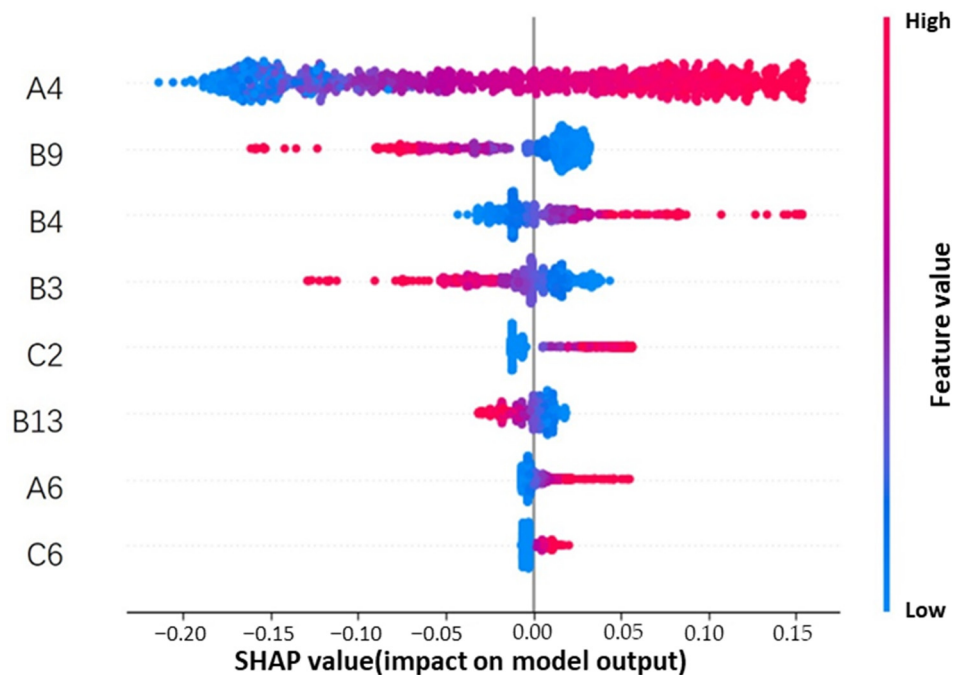


Figure 12. Contribution of various indicators in predicting crude-oil recovery level.

From the figure below (Figure 13), we can clearly see the impacts of different indicators on the prediction results when they change. The specific changes obtained by analyzing the correlation function of each indicator are as follows:

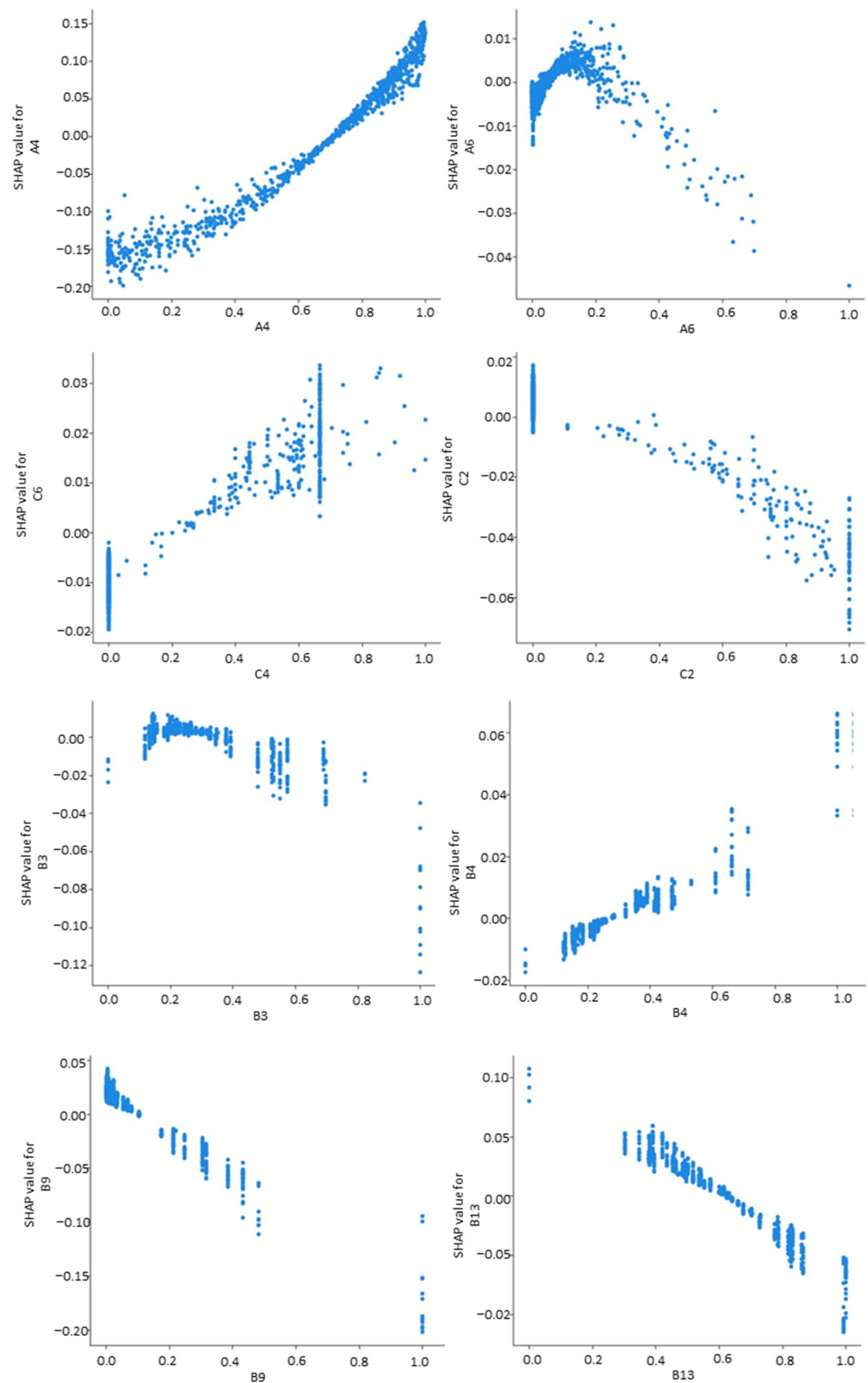


Figure 13. The functional relationship between various indicators and the degree of crude-oil recovery.

Based on the outcomes of the correlation analysis, it is evident that there exists a quadratic relationship between the total annual oil production and the degree of crude-oil recovery. Notably, this relationship is asymmetrical with respect to the Y-axis, suggesting that its functional relationship can be approximately modeled as

$$y = a(x - b)^2 + c \tag{14}$$

The relationship between water cut and crude-oil recovery degree is exponential, and it deviates from the standard exponential function. Therefore, the direct relationship between water cut and crude-oil recovery degree can be roughly fitted as

$$y = e^{ax+b} + b \tag{15}$$

The remaining indicators have a linear relationship with the degree of crude-oil recovery:

$$y = kx + b \tag{16}$$

These relationships are fitted using real oilfield data. Since all data are normalized during the model training process, a rash return to the original dimensions may weaken the functional relationship between indicators. Therefore, this article derives a specific empirical formula by fitting between the normalized indicators. When using this formula, all data must be normalized prior to calculation, and the calculation results are then de-normalized back to the original dimensions. The obtained empirical formula and prediction results are as follows:

$$\begin{aligned} \text{Oil Recovery factor} &= 0.0365 \times e^{1.1745 \times A4 + 1.6471} - 0.3449 \times B9 \\ &- 0.2728 \times (A6 - 0.5356)^2 - 0.0952 \times B13 + 0.2422 \times B4 \\ &- 0.2078 \times B3 + 0.0036 \times C6 + 0.0583 \times C2 - 0.0302 \end{aligned} \tag{17}$$

The correlation coefficient between the results fitted by the empirical formula and the real value can reach 0.884 (Figure 14), and the correlation coefficient can also reach 0.816 when single-well data are used for verification (Figure 15).

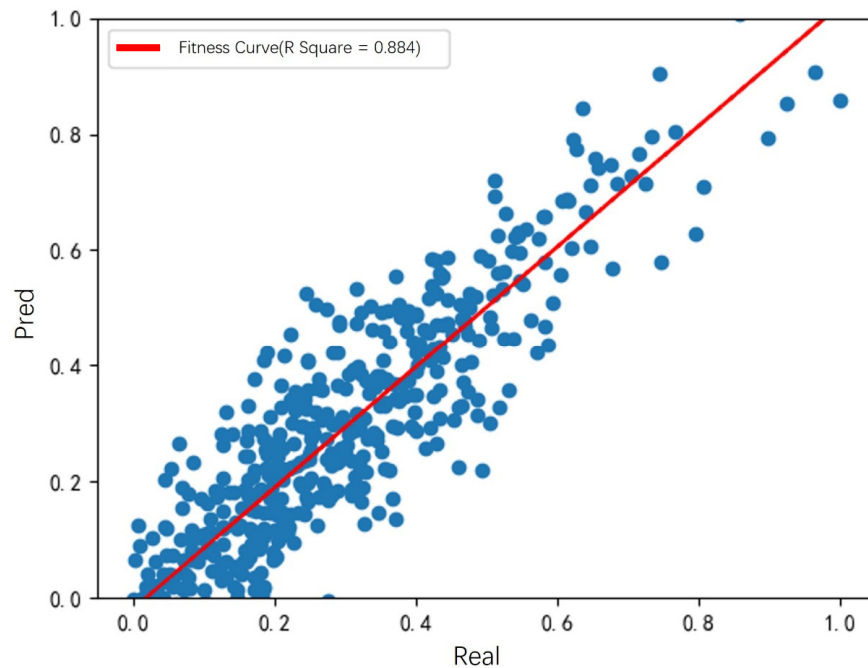


Figure 14. Scatter plot of crude-oil recovery degree calculated by empirical formula and real value.

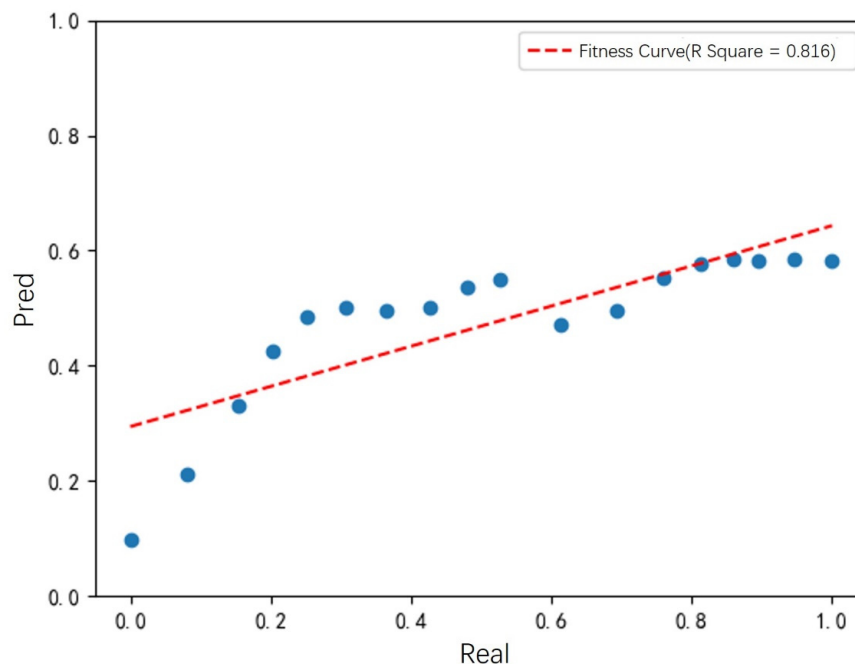


Figure 15. Validation effect of the empirical formula for crude-oil recovery degree.

5. Conclusions

The prediction of oilfield development indicators plays a crucial role in the overall development process of the oilfield. Accurate prediction of these indicators can not only enhance the overall economic benefits that the oilfield can generate, but also enable proactive adjustments to be made to oilfield development methods and means based on the prediction results. This, in turn, can help to prevent in advance the slowing down of overall oilfield production progress and the deterioration of production quality.

In the process of predicting key development indicators, this study first used the grey correlation theory and GRN-VSN algorithm to optimize the selection of input data and reduce the dimensions of input features.

In this study, different artificial intelligence algorithms are used to predict the degree of oilfield recovery, and the predicted results from the different algorithms are compared in order to select the optimal prediction model for key development indicators.

A fitting method for the corresponding empirical formula of the oilfield was constructed using SHAP. SHAP was employed to analyze the correlation between the input index and the output index. Given the high prediction-accuracy of the model, the accuracy of this corresponding relationship can also be ensured, which significantly reduces the difficulty of manual analysis. This method is, furthermore, applicable to other oilfields as well.

Author Contributions: Conceptualization, K.L. and K.W.; methodology, C.T.; software, Y.H.; validation, Y.Z. and S.C. (Suidong Chen); formal analysis, K.L.; investigation, K.W.; resources, K.W. and Y.P.; data curation, C.T. and S.C. (Shaobin Cai); writing—original draft preparation, K.L.; writing—review and editing, C.T.; visualization, K.W.; supervision, K.L.; project administration, K.L.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Tianshan Innovation Team Plan of Xinjiang Uygur Autonomous Region (Grant number: 2023D14011), the National Natural Science Foundation of China (Grant number: 52274030), the Key Research and Development Program Project of Karamay (Grant number: 2024jjldsqld0001) and “Tianchi Talent” Introduction Plan of Xinjiang Uygur Autonomous Region (2022).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Authors Ke Li, Kai Wang, Chenyang Tang, Yue Pan, Yufei He and Shaobin Cai were employed by the company China National Offshore Oil Corporation. 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.

References

1. Li, B.; Bi, Y.-B.; Pan, H.; Wang, Z.-K.; Zhang, S.-Z. Combination Method for Selecting Comprehensive Oilfield Development Effect Evaluation Targets. *Pet. Sci. Technol. Forum* **2012**, *31*, 38–41, 50.
2. Xu, H. Technical limits of water flooding development index. *Petrochem. Ind. Technol.* **2017**, *24*, 127.
3. Yang, T. Study on Variation of Development Index and Reasonable Production Allocation of Putaohua Oilfield. Master's Thesis, Northeast Petroleum University, Daqing, China, 2021. [\[CrossRef\]](#)
4. SY/T 6219-2023; Oilfield development level classification. Petroleum Industry Press: Beijing, China, 2024.
5. Xu, Y. The Suitable Conditions and Application of Water Drive Characteristic Curves. *China Sci. Technol. Inf.* **2009**, *21*, 32–33.
6. Li, Z.; Sun, L.; Deng, H.; Zhang, J.; Li, Y. The research of suitable conditions to water drive characteristic curve. *Comput. Tech. Geophys. Geochem. Explor.* **2012**, *34*, 143–146.
7. Zhu, M.; Shi, L.; Xue, Y.; Wen, J.; Li, S.; Liu, M.; Xin, C. Study and application of new water drive characteristic curve. *Unconv. Oil Gas* **2022**, *9*, 65–70.
8. Zhu, L. New Chemical Flooding Tracking Evaluation and Water Flooding Characteristic Curve Study in Offshore Oil Field. Master's Thesis, China University of Petroleum, Beijing, China, 2022. [\[CrossRef\]](#)
9. Deng, J.; Wu, X.; Zhu, Z.; Zhang, L.; Gao, Y. Study on the prediction method of horizontal well development index in the S Oilfield of Bohai Sea. *Complex Hydrocarb. Reserv.* **2023**, *16*, 211–214.
10. Zhang, J. Application Research of New Reservoir Engineering Method in the Late Stage of Ultra High Water Cut in Xingbei Development Area. Master's Thesis, Northeast Petroleum University, Daqing, China, 2024. [\[CrossRef\]](#)
11. Ma, C.; Chen, K.; Zhang, H. Application research on the evaluation system for the development effect of a certain offshore oilfield. *West-China Explor. Eng.* **2022**, *34*, 83–86.
12. Zong, H. Research and application of methods for improving oil recovery in high water cut oil reservoirs. *Inn. Mong. Petrochem. Ind.* **2007**, 198–201. [\[CrossRef\]](#)
13. Guo, W. Hydrodynamics Intensity Method of Well Patterns and Rate Adjustment in Ultra—High Water Cut State. Ph.D. Thesis, China University of Geosciences (Beijing), Beijing, China, 2016. [\[CrossRef\]](#)
14. Gao, M. Study on Optimization of Hydrodynamic Development Mode at High Water-Cut Stage in Fault Block Reservoirs. Master's Thesis, China University of Petroleum (East China), Dongying, China, 2019. [\[CrossRef\]](#)
15. Wu, N.; Shi, S.; Zheng, S.; Zhao, H.; Wang, H. Formation pressure calculation of tight sandstone gas reservoir based on material balance inversion method. *Coal Geol. Explor.* **2022**, *50*, 115–121.
16. Wang, D.; Jiang, Y.; Huang, L.; Wu, Y. Research on oil ring determination and dynamic reserve calculation method of buried hill condensate gas reservoir. *Petrochem. Appl.* **2021**, *40*, 26–30+34. [\[CrossRef\]](#)
17. Gu, H.; Zheng, S.; Zhang, D.; Yang, Y. Modification and application of material balance equation for ultra-deep reservoirs. *Acta Pet. Sin.* **2022**, *43*, 1623–1631.
18. Ma, Q.; Yang, Z.D.; Zheng, P.Y.; Yu, C.; Guo, Q. Design and application of water injection development adjustment scheme for complex faultblock reservoir. *Pet. Plan. Des.* **2016**, *27*, 19–23.
19. Han, R.; Qi, D.; Wu, Z.; Yan, G. Application of BP neural network in predicting production changes of Shinan 31 oilfield. *Inn. Mong. Petrochem.* **2010**, *36*, 170–172.
20. Ren, B.; Zhao, M.; Liu, Z.; Wang, J. Support vector machine prediction of oilfield development dynamic indicators. *Pet. Plan. Des.* **2008**, 12–15+48. [\[CrossRef\]](#)
21. Ma, L.; Li, D.; Guo, H.; Li, W. Application of BP neural network optimized by genetic algorithm in crude oil production prediction: A case study of BED test area in Daqing Oilfield. *Math. Pract. Theory* **2015**, *45*, 117–128.
22. Zhao, L.; Li, X.; Xu, S.; Xia, H. Oilfield development index prediction model based on process support vector regression machine. *Math. Pract. Theory* **2018**, *48*, 83–88.
23. Zhang, Y. Oilfield Development Index Prediction and Status Assessment Method Based on Reservoir Modeling Results. Master's Thesis, Northeast Petroleum University, Daqing, China, 2016.
24. Chen, C. Research on Oilfield Development Data Analysis and Prediction Based on Artificial Neural Network. Master's Thesis, Northeast Petroleum University, Daqing, China, 2022. [\[CrossRef\]](#)
25. Zhong, Y.; Wang, S.; Luo, L.; Yang, J.; Yue, Y. Using deep learning to mine knowledge of oilfield development index prediction model. *J. Southwest Pet. Univ.* **2020**, *42*, 63–74.
26. Li, T. Research on Oilfield Development Data Analysis and Prediction Model Based on Dynamic Neural Network. Master's Thesis, Northeast Petroleum University, Daqing, China, 2016.
27. Nguyen, H.H.; Chan, C.W.; Wilson, M. Prediction of oil well production using multi-neural networks. In Proceedings of the IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No.02CH37373), Winnipeg, MB, Canada, 12–15 May 2002; pp. 798–802.

28. Hu, H.; Feng, J.; Guan, X. A Method of Oil Well Production Prediction Based on PCA-GRU. In Proceedings of the 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 18–20 October 2019; pp. 710–713.
29. Dang, C. Research on Theapplication of Deep Learning in Oilfield Development Indicator Early Warning. Master's Thesis, Xi'an Shiyou University, Xi'an, China, 2024. [[CrossRef](#)]
30. Zhu, B. Research on Injection and Production Control Model and Algorithm of Water Drive Reservoir Based on Computational Intelligence. Master's Thesis, Northeast Petroleum University, Daqing, China, 2024. [[CrossRef](#)]
31. Qu, Q. Design and Development of Intelligent Analysis System for Dynamic Indicators of Reservoir Polymer Flooding Development. Master's Thesis, Northeast Petroleum University, Daqing, China, 2024. [[CrossRef](#)]
32. Sun, D. Research Onmultidisciplinary Data Analysis Methods and Applications for Oilfield Development Evaluation. Master's Thesis, Shandong University of Science and Technology, Qingdao, China, 2021. [[CrossRef](#)]
33. Fan, S. Layered Water Injection Prediction and Downhole Injection System Design Based on CNN-LSTM. Master's Thesis, Harbin University of Science and Technology, Harbin, China, 2024. [[CrossRef](#)]
34. Liu, S.; Forrest, J.Y.-L. *Grey System Theory and Its Application*, 5th ed.; Springer: Berlin/Heidelberg, Germany, 2010.
35. Chen, H.; Zhang, Y. Application of Grey Correlation Analysis Method in Bayan Chagan Reservoir Evaluation. *Pet. Geol. Eng.* **2023**, *37*, 45–51.
36. Chen, N.; Wang, H.; Guo, P.; Li, Y.; Zhang, B. Research on evaluation of factors affecting oil well production based on grey correlation analysis. *Petrochem. Appl.* **2023**, *42*, 87–91+113.
37. Wang, C.; Du, H.; Sun, X.; Dai, C.; Yang, J.; Chen, R. Comprehensive evaluation method of shale oil sweet spot based on grey correlation analysis: A case study of Bonan Sag in Bohai Bay Basin. *Pet. Drill. Technol.* **2023**, *51*, 130–138.
38. Liang, Y.; Li, N.; Liu, L.; Han, J.; He, P.; Ai, X. Evaluation of tight gas field gathering and transportation technology based on multi-level grey correlation analysis method. *Nat. Gas Explor. Dev.* **2024**, *47*, 104–111.
39. Hou, C. Oil production prediction method for new wells in oil fields based on longshort-term memory neural network. *Oil Gas Geol. Recovery Effic.* **2019**, *26*, 105–110.
40. Wang, H.; Lin, X.; Jiang, L.; Liu, Z. Oilfield production prediction based on clustering and long short-term memory neural network. *Pet. Sci. Bull.* **2024**, *9*, 62–72.
41. Huang, Q. Research on the Improvement and Application of BP Algorithm. Master's Thesis, Southwest Jiaotong University, Chengdu, China, 2010. [[CrossRef](#)]

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