energies

Article

A Novel Method for Predicting Oil and Gas Resource Potential Based on Ensemble Learning BP-Neural Network: Application to Dongpu Depression, Bohai Bay Basin, China

Zijie Yang [1,2,3], Dongxia Chen [1,2,3,*], Qiaochu Wang [1,2,3], Sha Li [1,2,3], Fuwei Wang [1,2,3], Shumin Chen [1,2,3], Wanrong Zhang [1,2,3], Dongsheng Yao [1,2,3], Yuchao Wang [1,2,3] and Han Wang [1,2,3]

[1] State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing 102249, China; yangzj2834@163.com (Z.Y.)
[2] College of Geosciences, China University of Petroleum (Beijing), Beijing 102249, China
[3] Hainan Institute of China University of Petroleum (Beiijing), Sanya 572025, China
* Correspondence: lindachen@cup.edu.cn

Abstract

Assessing and forecasting hydrocarbon resource potential (HRP) is of great significance. However, due to the complexity and uncertainty of geological conditions during hydrocarbon accumulation, it is challenging to accurately establish HRP models. This study employs machine learning methods to construct a HRP assessment model. First, nine primary controlling factors were selected from the five key conditions for HRP: source rock, reservoir, trap, migration, and accumulation. Subsequently, three prediction models were developed based on the backpropagation (BP) neural network, BP-Bagging algorithm, and BP-AdaBoost algorithm, with hydrocarbon resources abundance as the output metric. These models were applied to the Dongpu Depression in the Bohai Bay Basin for performance evaluation and optimization. Finally, this study examined the importance of various variables in predicting HRP and analyzed model uncertainty. The results indicate that the BP-AdaBoost model outperforms the others. On the test dataset, the BP-AdaBoost model achieved an R^2 value of 0.77, compared to 0.73 for the BP-Bagging model and only 0.64 for the standard BP model. Variable importance analysis revealed that trap area, sandstone thickness, sedimentary facies type, and distance to faults significantly contribute to HRP. Furthermore, model accuracy is influenced by multiple factors, including the selection and quantification of geological parameters, dataset size and distribution characteristics, and the choice of machine learning algorithm models. In summary, machine learning provides a reliable method for assessing HRP, offering new insights for identifying high-quality exploration blocks and optimizing development strategies.

Keywords: machine learning; boosting algorithm; prediction model; Dongpu depression; petroleum resources

check for updates

Academic Editor: Albert Ratner

Received: 3 September 2025
Revised: 17 October 2025
Accepted: 18 October 2025
Published: 22 October 2025

Citation: Yang, Z.; Chen, D.; Wang, Q.; Li, S.; Wang, F.; Chen, S.; Zhang, W.; Yao, D.; Wang, Y.; Wang, H. A Novel Method for Predicting Oil and Gas Resource Potential Based on Ensemble Learning BP-Neural Network: Application to Dongpu Depression, Bohai Bay Basin, China. *Energies* **2025**, *18*, 5562. https://doi.org/10.3390/en18215562

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the continuous growth in energy demand, the ability to accurately assess and predict oil and gas resources has become more critical than ever before [1–3]. Traditional oil and gas resources, as vital sources of global energy supply, are constrained by multiple complex geological factors, including reservoir characteristics and hydrocarbon migration models. Integrating geophysical techniques, geochemical laboratory data, and geological

theoretical analysis is crucial for enhancing the accuracy of oil and gas resource potential assessments. By employing advanced geological modeling techniques and multidisciplinary integration approaches, exploration companies can more effectively address uncertainties in hydrocarbon exploration, ensuring the sustainable exploration and development of these resources. Therefore, a comprehensive assessment of hydrocarbon resources must holistically consider these multidimensional parameters and data, rather than relying on overly simplified or isolated applications of these data.

In the early stages of oil and gas exploration, potential hydrocarbon accumulations are often identified by superimposing multiple geological parameters [4–6]. However, as exploration and development activities advance, the applicability and reliability of traditional geological evaluation models become increasingly limited. It is necessary to establish a comprehensive evaluation system for oil and gas resources to address the precision requirements of modern exploration tasks. [7–11]. In recent years, the advent of high-performance computing and advancements in computer technology have spawned novel simulation tools and expert decision systems, enabling the integration of statistical techniques, mathematical theories, and geological sciences [12–18]. This convergence has given rise to exploration methodologies grounded in mathematical geology. Although exploration and development technologies have achieved breakthrough progress, the geological data collected during exploration and development still face practical challenges such as fragmented sources and scarcity in sample types and quantities [19]. These factors collectively lead to traditional geological models and mathematical geological models being susceptible to subjective bias when addressing complex geological challenges. Conventional resource evaluation methods are increasingly inadequate for meeting the objectives and tasks of digital exploration in modern oilfields. Therefore, there is an urgent need to leverage advanced computational techniques and statistical methods to comprehensively enhance exploration and development efficiency while improving resource evaluation accuracy.

Artificial intelligence (AI) technology has achieved significant breakthroughs across multiple industries [20–22], yet its application in oil and gas exploration and production remains in its infancy. Nevertheless, machine learning (ML) methods demonstrate considerable potential in oil and gas energy sectors, including resource potential prediction, exploration, and production optimization [23–26]. Algorithms like Backpropagation (BP) Neural Networks, Multi-Layer Perceptron (MLP), Support Vector Machines (SVM), Extreme Gradient Boosting (XGBoost), and Random Forests excel at solving regression and classification problems. Consequently, they are widely applied to predict shale organic carbon content and lithofacies types [27,28]. These algorithms are also crucial for forecasting unconventional shale gas production and shale fluidity [29,30]. Furthermore, ensemble methods that integrate multiple base learners can effectively enhance predictive performance, addressing the inherent limitations of individual models in data quality, feature selection, and algorithmic constraints. Among these, ensemble approaches based on Bagging and Boosting have been successfully applied in the oil and gas field sector, specifically covering scenarios such as crude oil production forecasting, drilling rate assessment, and reservoir property analysis. The successful implementation of ensemble learning principles in petroleum geology further validates its capability to address complex problems involving multiple interrelated factors [31–34]. Compared to traditional methods, artificial intelligence demonstrates superior adaptability and higher precision in handling highly nonlinear and multivariate relationships through its advanced data processing and analytical capabilities [35–37].

In the field of oil and gas exploration, the complexity of geological processes and hydrocarbon accumulation has led to the development trend of “one model per oilfield,”

marking a significant advancement in the industry [38]. To address this challenge, this study proposes a resource potential assessment method based on ensemble learning, using the Dongpu Depression in the Bohai Bay Basin as a case study [39,40]. Due to the complex tectonic evolution and sedimentary stratigraphy of the study area [41–43], existing oil and gas resource evaluation models are difficult to apply to the Dongpu Depression. Therefore, this study first constructs a BP neural network model to predict hydrocarbon resources. Subsequently, BP-Adaboost and BP-Bagging ensemble learning algorithms are employed to enhance model performance. Finally, the optimal ensemble model is utilized for hydrocarbon resource prediction. This method provides an effective tool for oil and gas exploration companies to address the challenge of assessing resource potential in complex geological environments, simultaneously reducing drilling failure risks and costs while enhancing exploration efficiency and economic benefits.

2. Geological Setting

The Dongpu Sag, located on the southwestern margin of the Bohai Bay Basin in China, is a typical Neogene faulted basin (Figure 1a) [38]. From east to west, the Dongpu Sag consists of five major secondary structural units: the Lanliao fault zone, eastern sub-sag zone, central low uplift zone, western sub-sag zone and western slope zone (Figure 1b). The Paleogene oil and gas reservoirs in the Dongpu Sag are primarily concentrated in the Northern Central Uplift Zone and the Western Slope Zone [39]. The Wenliu area, studied in this research, is located in the northern part of the central low uplift zone, which is one of the most important oil and gas accumulation regions of the Dongpu Sag (Figure 1c). The Dongpu Sag is characterized by well-developed faulting and complex structures, with a profile exhibiting the feature of 'two sags, one uplift, and one slope' (Figure 1d).

The Dongpu Sag experienced three stages of evolution, Specifically, it can be divided into [40]: (1) The initial rifting period (depositional period of the fourth member of the Shahejie Formation (Es_4)). (2) The main rifting periods. (3) The inherited development period. (4) The rifting decay period [41–43]. The Dongpu Sag possesses favorable source rock conditions and complex tectonics, which have resulted in the formation of numerous fault-block traps of varying sizes. Although the number of oil and gas reservoirs is more, the scale of individual fault-block traps is relatively small, and resource distribution is uneven, has led to a widespread distribution of oil and gas reservoirs in various spatial locations [44–47].

From the base to the top of the Cenozoic in the Dongpu Depression, the Paleogene Shahejie Formation (Es), Dongying Formation (Ed), Neogene Guantao Formation (Ng), Minghuazhen Formation (Nm), and Quaternary Plain Formation (Qp) were successively deposited [48,49] (Figure 2). The Shahejie Formation is the primary hydrocarbon source rock and reservoir development layer, serving as the target for oil and gas exploration in the region. The third member of the Shahejie Formation (Es_3) is the most widespread and thickest, deposited in a semi-deep to deep lake environment. It consists mainly of gray shale, gray siltstone, fine sandstone, and salt gypsum [50,51]. The Es_3 can be subdivided into three sub-members: upper (Es_3^U), middle (Es_3^M) and lower (Es_3^L). The primary stratum for this study is: ES_3^M Formation. In the Es_3^M Formation, the top of Wenliu area develops thick gypsum salt rock, and the middle and lower parts are mainly a set of dark gray mudstone with oil shale, which is interbedded with sandstone. Es_3^M formation represents the primary reservoir development horizon in the Wenliu region. The study area is characterized by the presence of four gypsum-salt rock formations, three of which are found within Es_3 formation, while the fourth is located in Es_1 formation. These gypsum-salt layers are crucial in the region's geological framework, acting as key cap rocks that play an essential role in the accumulation and trapping of hydrocarbons [52–54]. Research

indicates that several geological factors control the distribution of oil and gas resources in the Dongpu Depression, with many uncertainties in resource evaluation [55–57].

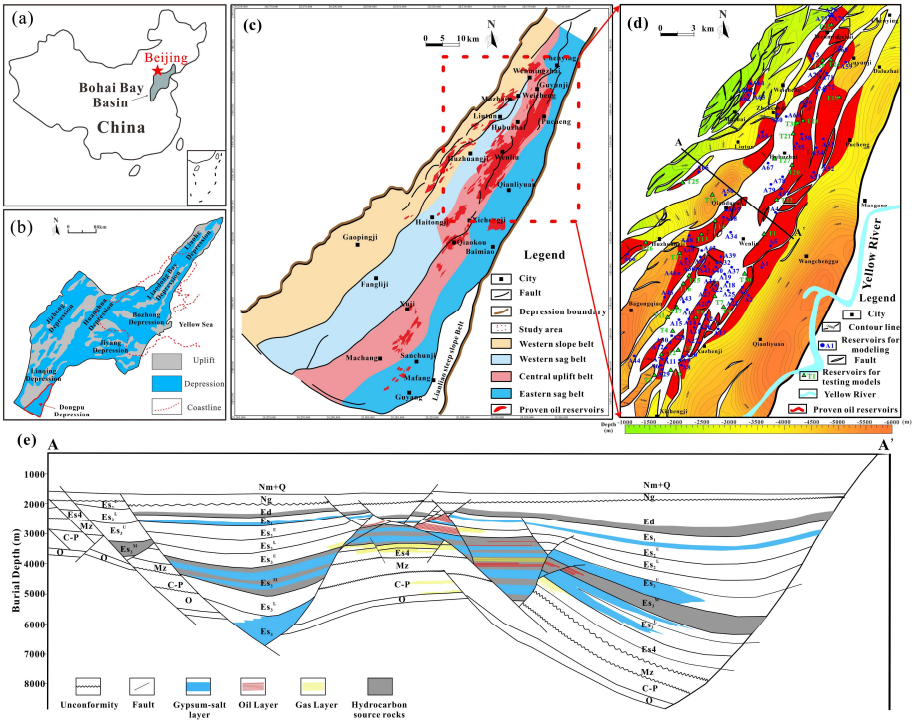


Figure 1. (**a**,**b**) Location of the Bohai Bay Basin, China; (**c**) General structural settings of the Dongpu Depression; (**d**) Structural topographic map of northern Dongpu depression and the location of sampling well; (**e**) Typical cross section of the AA'line (the abbreviations of the stratigraphic symbols in the figure are provided in Figure 2 for reference).

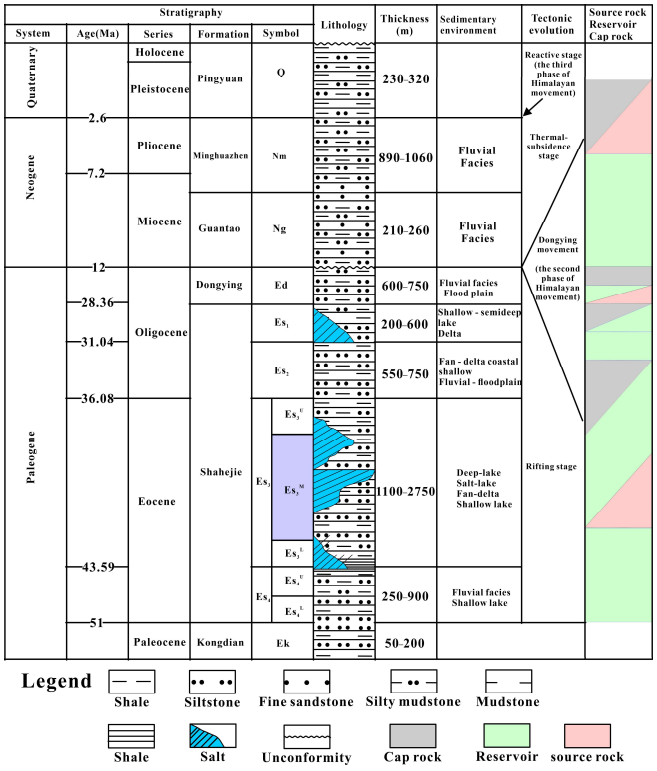


Figure 2. Generalized Paleogene stratigraphy of the Dongpu Depression (the stratigraphic layers highlighted in blue correspond to the levels examined in this study; [43,54]).

3. Material

3.1. Data Sources

The data for this study were provided by the Zhongyuan Oilfield Branch of the China Petroleum & Chemical Corporation, including sedimentary facies maps, reservoir profiles, scientific research illustrations, and logging, oil and gas testing, production, and experimental analysis data from 110 drilled wells. These data were used to quantitatively characterize the oil and gas resource potential and its associated geological factors.

The geological parameters influencing oil and gas resources in the Es_3^M section of the Wenliu area in the Dongpu sag are complex, encompassing five key hydrocarbon accumulation conditions: source rock, reservoir, trap, migration, and accumulation. Based on geological theory, exploration practice, and prior research [58,59], seven quantitative geological parameters were identified: sand thickness (ST), average porosity (Apor), average permeability (Aper), trap area (TA), surface crude oil density (SCOD), pressure coefficient (PC), and distance from fault (DF). Additionally, two qualitative geological parameters—sedimentary facies (SF) and hydrocarbon generation center (HGC) [55,60]—were considered. Reserve abundance in a hydrocarbon reservoir is defined as the ratio of reserves to reservoir area. This metric mitigates the impact of field size variations on reserve estimates, providing a clearer indication of the reservoir's richness. The measured oil and gas resource data for the target parameters were derived from reservoir reserve data as recorded in the reserve report provided by the oilfield. In this study, the whole dataset is divided into two parts: one part is used for modeling, light blue labeled reservoirs shown in Figure 1, and is divided into training set data and validation set data according to the ratio of 70%:30%; the other part is used for testing the model and is not involved in modeling.

3.2. Data Standardization

When applying machine learning to build models, data of different magnitudes and orders of magnitude can have an impact on the modeling results. It is therefore necessary to first perform factorless quantization of the raw data to minimize the effects caused by differences in orders of magnitude and to make the input parameters comparable [61]. Considering the above factors and the characteristics of geological data distribution, the Min-max data standardization method is chosen for data normalization. The specific formula is Equation (1)

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X is the original data, X' is the normalized data, X_{min} is the minimum value in the data set, and X_{max} is the maximum value in the data set.

3.3. Quantification of Contribution Characteristics of Hydrocarbon Generation Center

Continuing to extend this concept on the idea of source-control theory, while considering the controlling role of hydrocarbon source conditions on hydrocarbon distribution, a probabilistic model of hydrocarbon formation was established [62–65] (Figure 3). We can use the probability of oil and gas generation to quantitatively represent the contribution of source rock hydrocarbon generation center to HGC. The HGC reflects the hydrocarbon production capacity per unit area, integrating factors such as organic matter type, organic carbon content, and the thermal evolution degree of the source rock, thereby reducing the total number of parameters for analysis [9]. The probability quantitative equation of oil and gas reservoir distribution is as follows:

$$\begin{gathered} F_e = 0.046 \cdot e^{0.12 \cdot q_e} - 0.16 \cdot ln(L) + 0.65 \cdot e^{-8.2357 \cdot (l+0.1)^2} + 0.1345 \\ R^2 = 0.8825 \end{gathered} \tag{2}$$

where F_e is the probability of hydrocarbon accumulation, dimensionless; L is the standardized distance from the reservoir to the hydrocarbon expulsion center, dimensionless; l is the distance from the standardized reservoir to the hydrocarbon expulsion boundary, dimensionless. Additionally, q_e is maximum hydrocarbon expulsion intensity of the hydrocarbon kitchen, in 10^6 t/km^2.

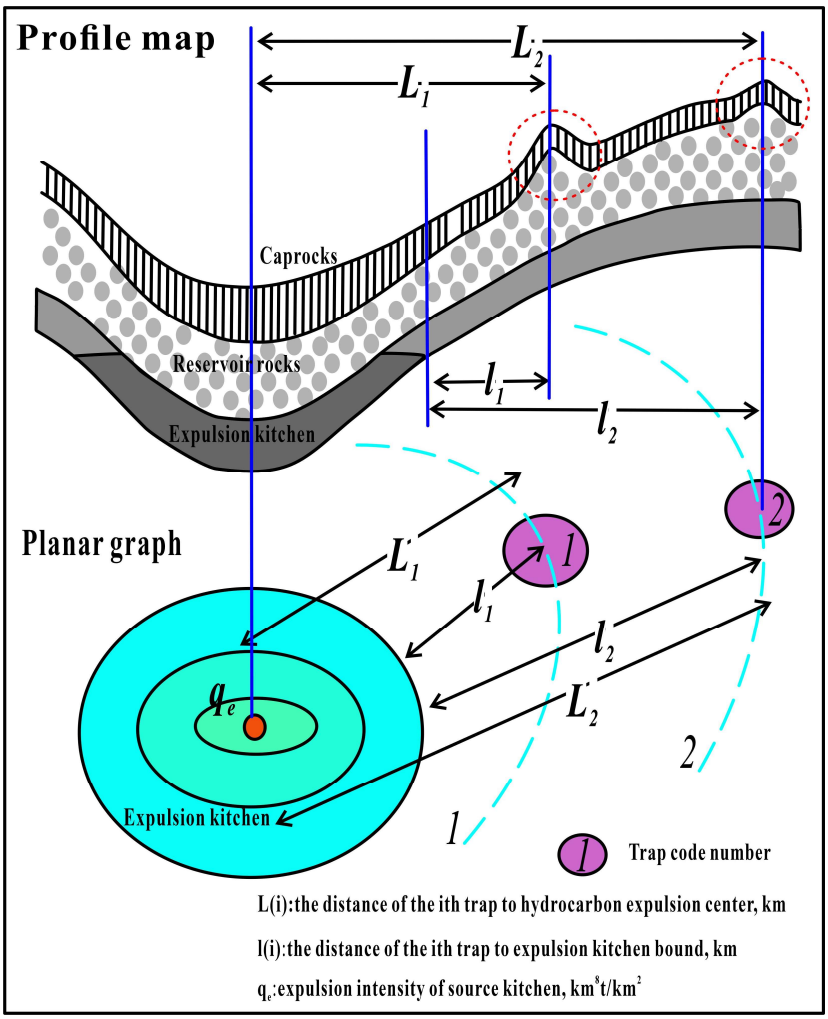


Figure 3. Hydrocarbon distribution threshold concept model of Multiple Source kitchens [64].

Here, $L = \frac{L_a}{L_0}$, $l = \frac{l_a}{L_0}$, where L_a is the actual distance from the reservoir to the hydrocarbon expulsion center (km), and l_a is the actual distance from the reservoir to the hydrocarbon expulsion boundary, km. L_0 is distance between hydrocarbon expulsion boundary and hydrocarbon expulsion center of source rocks (km).

3.4. Quantification of Sedimentary Facies Characteristics

The dominant sedimentary facies have the characteristics of high porosity and high permeability, which is conducive to the accumulation and accumulation of oil and gas in the dominant sedimentary facies area, and then controls the distribution and range of favorable accumulation zones. In this study, by quantifying different sedimentary facies

types, the control degree of different sedimentary facies on petroleum accumulation is quantitatively characterized (SF). The specific formula is Equation (3).

$$F_a = \begin{cases} 1, Delta\ front\ subfacies \\ \dfrac{Q_{facies}}{Q_{total\ facies}}, Other\ sedimentary\ subfacies \end{cases} \quad (3)$$

where F_a is the control probability of sedimentary relative hydrocarbon accumulation; Q_{facies} is the amount of oil and gas resources in a sedimentary subfacies; $Q_{total\ facies}$ is the amount of oil and gas resources in total sedimentary subfacies.

4. Methodology

4.1. Back Propagation Neural Network (BP)

The Back Propagation (BP) neural network, proposed in 1986 by Rumelhart and McClelland [66], is a multi-layer feedforward neural network trained using the error backpropagation algorithm. It is one of the most widely adopted neural network models. The core principle involves applying the gradient descent method to iteratively adjust the network's weights and thresholds through backpropagation, minimizing the mean squared error between the actual and expected output values [67–69]. The BP neural network consists of three layers: input, hidden, and output. The structure of the BP neural network is illustrated in Figure 4.

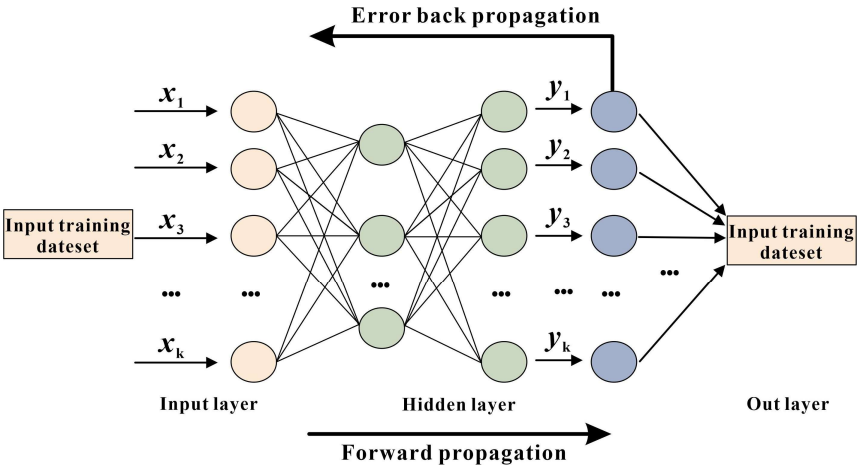


Figure 4. BP neural network structure diagram.

4.2. Ensemble Learning Methods

4.2.1. BP-Adaboost Algorithm (BP-Adaboost)

AdaBoost is an iterative algorithm within the Boosting framework. The core principle of AdaBoost is to train multiple weak classifiers on the same training set and combine them to form a stronger classifier. A key advantage of AdaBoost is its use of weighted training data rather than random sampling, along with a weighted voting mechanism instead of average voting [70,71]. In BP-AdaBoost, the BP neural network serves as a weak learner, iteratively trained to predict sample outputs. This process results in a more accurate and robust strong learner model composed of multiple BP neural network weak learners (Figure 5).

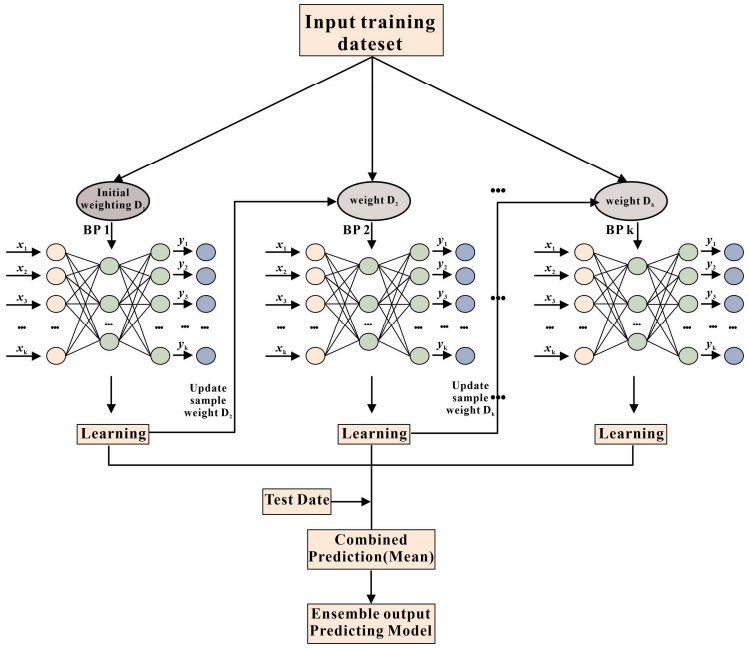


Figure 5. Workflow diagram of BP-AdaBoost model.

4.2.2. BP-Bagging Algorithm (BP-Bagging)

Bagging (Bootstrap aggregating) is another ensemble learning method that generates diverse learners through Bootstrap sampling [72,73]. The main idea of Bagging is to combine multiple base learners, which are trained in parallel, and then average their outputs. In this study, the BP neural network is selected as the base learner for the Bagging ensemble model. For numerical regression tasks, Bagging typically averages the results of multiple base learners to produce the final prediction [74,75] (Figure 6).

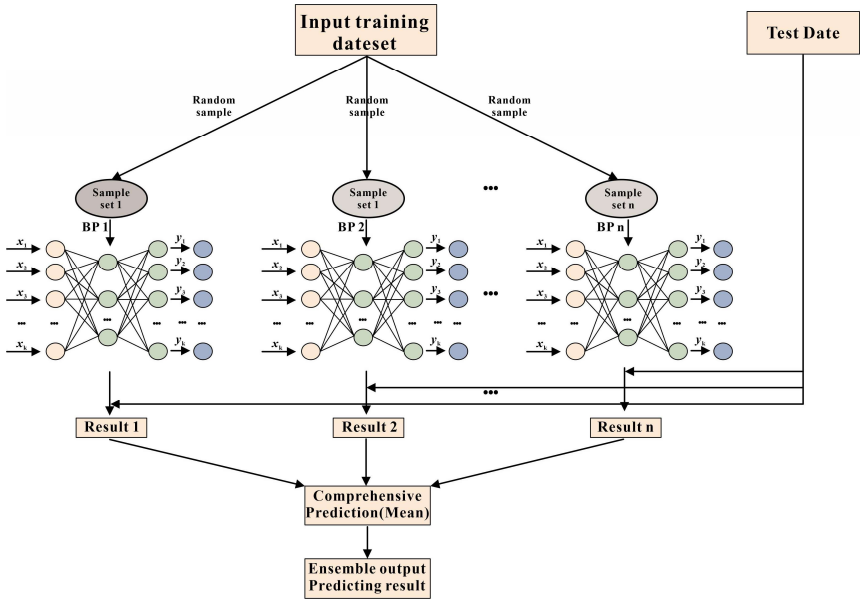


Figure 6. Workflow diagram of BP-Bagging model [75].

4.3. Prediction Performance Evaluation Metrics

To assess the prediction accuracy of these four algorithmic models, three metrics were selected to evaluate the performance of various models, including correlation coefficient

(R^2), root mean square error (RMSE), and mean absolute error (MAE), which are calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n}} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}| \quad (6)$$

where y_i is the actual value, $\hat{y}$ is the predicted value, $\overline{y}$ is the mean of the actual value, and n is the number of samples. The value of the correlation coefficient R^2 ranges from [0–1], and as R^2 becomes closer to 1, and as the values of RMSE and MAE decreases, it indicates a more accurate model and a better fit.

4.4. Workflow Diagram for Predicting Hydrocarbon Resource Abundance

The BP neural network, BP-AdaBoost, and BP-Bagging models were applied to predict the abundance of oil and gas resources. Model performance was assessed using reliability evaluation parameters, including R^2, MAE, RMSE, to select the most representative model for predicting resource abundance in other blocks within the study area. All models were implemented using SPSS Modeler (v18.0). The dataset was randomly split into a training set (70%) and a validation set (30%), which were used for model training and performance validation, respectively. This partitioning enhances the model's generalization ability and reduces the risk of overfitting (Figure 7). After comprehensive evaluation, the optimal model was applied to previously unmodeled blocks to assess its practical application.

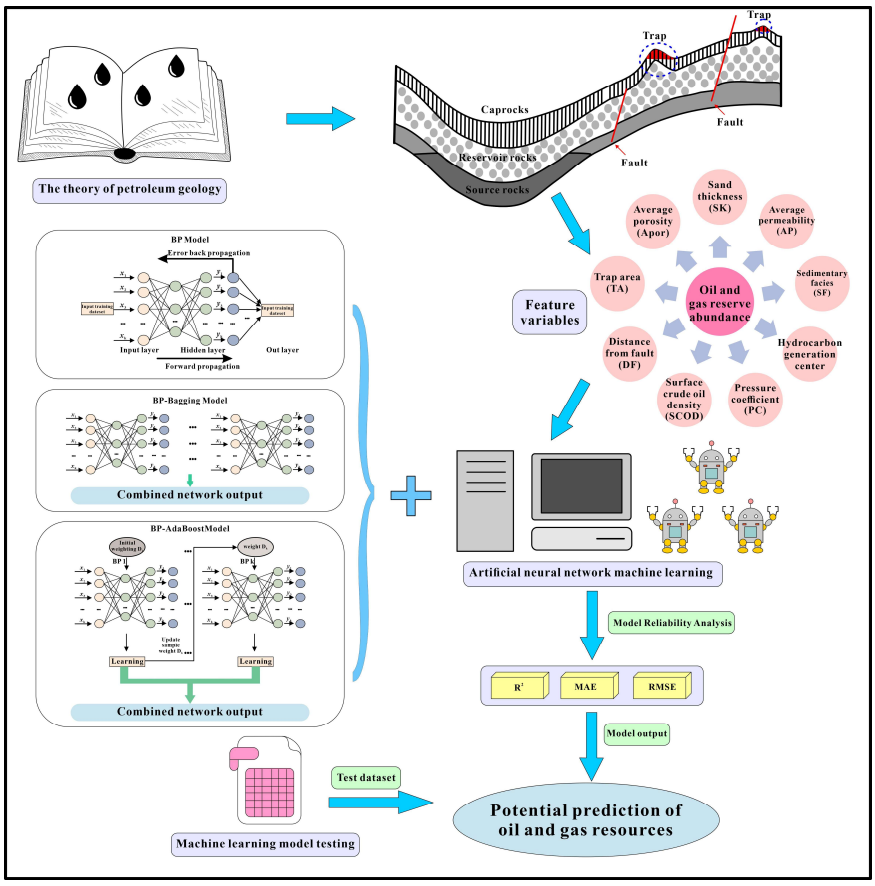


Figure 7. Workflow diagram for predicting hydrocarbon resource abundance in this study.

5. Results

5.1. Data Processing

5.1.1. Quantified Data

The contribution of hydrocarbon centers to hydrocarbon accumulation in different reservoirs was quantitatively calculated using Equation (2). The results show that the HGC index distribution is highest in the southeastern part of the study area, followed by the northern and southwestern regions. Three hydrocarbon centers were identified, which have a very good match with the hydrocarbon expulsion centers of the hydrocarbon source rocks (Figure 8).

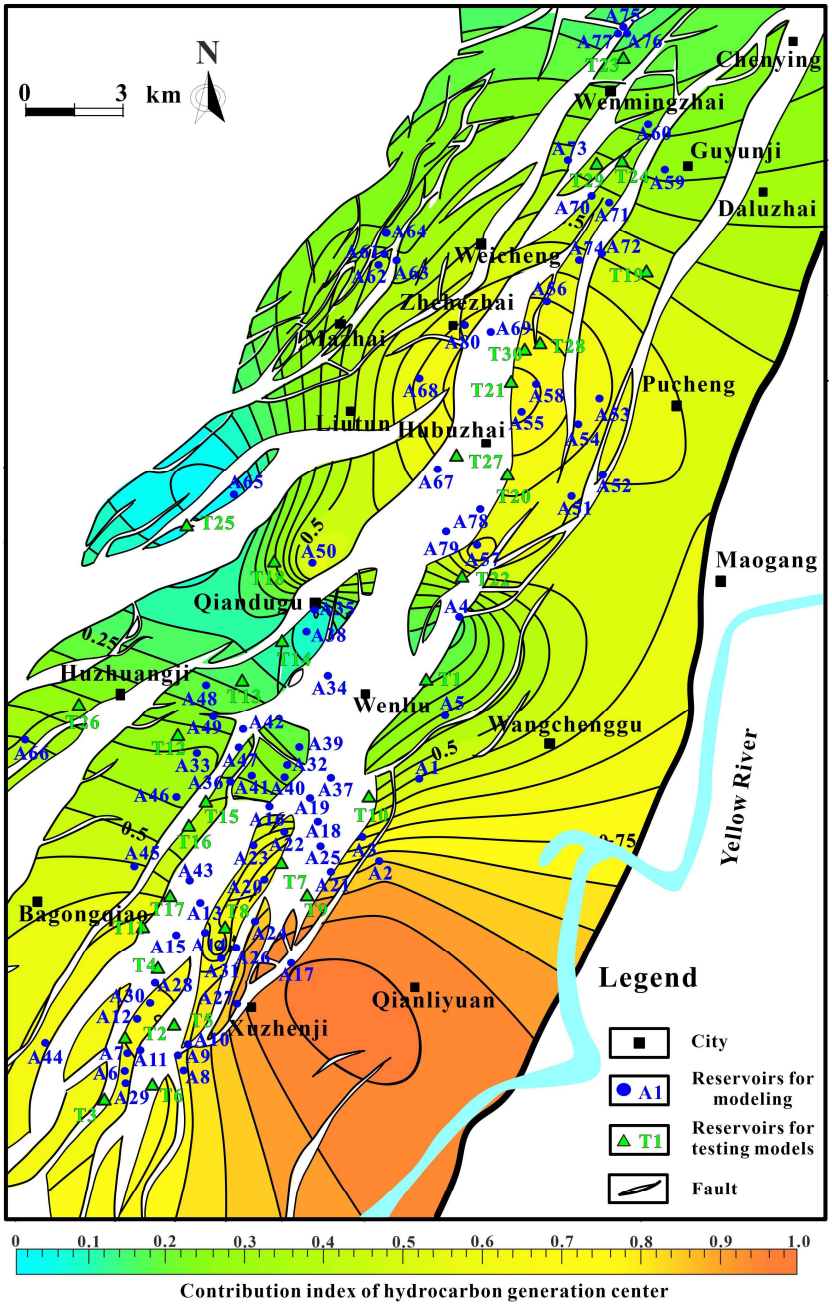


Figure 8. The contribution index plan of hydrocarbon generation center for the Es_3^M Formation in the study area.

Previous studies have identified a fan-lacustrine mud-beach bar sedimentary system in the Dongpu sag [76,77]. Based on these findings, six sedimentary facies were delineated,

including Shore-shallow Lake, Shore shallow lake-Delta front, Delta front facies, Semi-deep lake—Deep Lake facies, Distributary channel and Shallow water delta (Figure 9a). More than 50% of the oil and gas reservoirs in the study area are distributed in the front subfacies, which exhibit favorable physical properties, contributing 46.6% to the proven geological reserves. Therefore, the front subfacies is most favorable for hydrocarbon accumulation is assigned a value of 1 (Figure 9b). The values for other sedimentary facies are assigned based on the ratio of proven geological reserves within each subfacies relative to the total reserves, with values ranging from 0 to 1, reflecting the control probability of each facies on hydrocarbon accumulation (Figure 10).

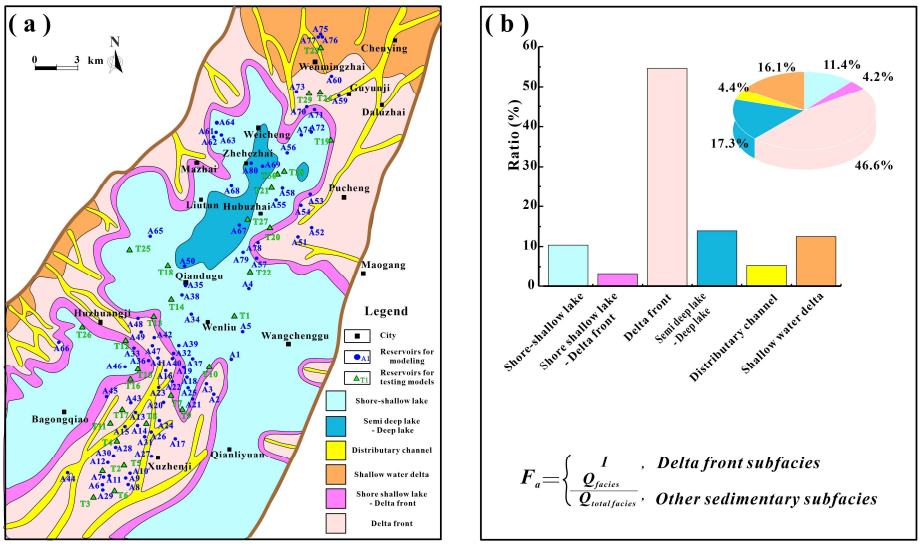


Figure 9. (**a**) The sedimentary facies type display map for the Es_3^M Formation of the study area; (**b**) Proportion of hydrocarbon resource quantities in different sedimentary facies types.

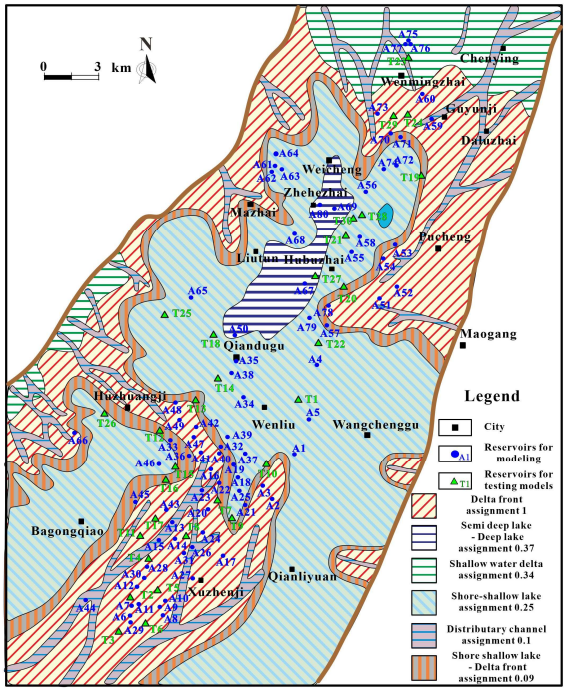


Figure 10. Schematic diagram of the sedimentary facies assignment results for the Es_3^M Formation in the study area.

5.1.2. Data Characteristics

Figure 11 illustrates the Pearson correlation coefficients between characteristic and target parameters. The closer the absolute value of the Pearson correlation coefficient is to 1, the stronger the linear relationship between reserve abundance and the corresponding characteristic parameters. For instance, the absolute values of the Pearson correlation coefficients for pressure coefficient, SF, sand thickness, surface crude oil density, trap area, and HGC are all greater than 0.3, suggesting a strong linear correlation with reserve abundance. Interestingly, while porosity and permeability show a strong linear relationship with each other, their Pearson correlation coefficients with reserve abundance are below 0.5, indicating a weak linear correlation. However, based on geological knowledge, porosity and permeability are known to have significant effects on reserve abundance, implying a strong nonlinear correlation with it.

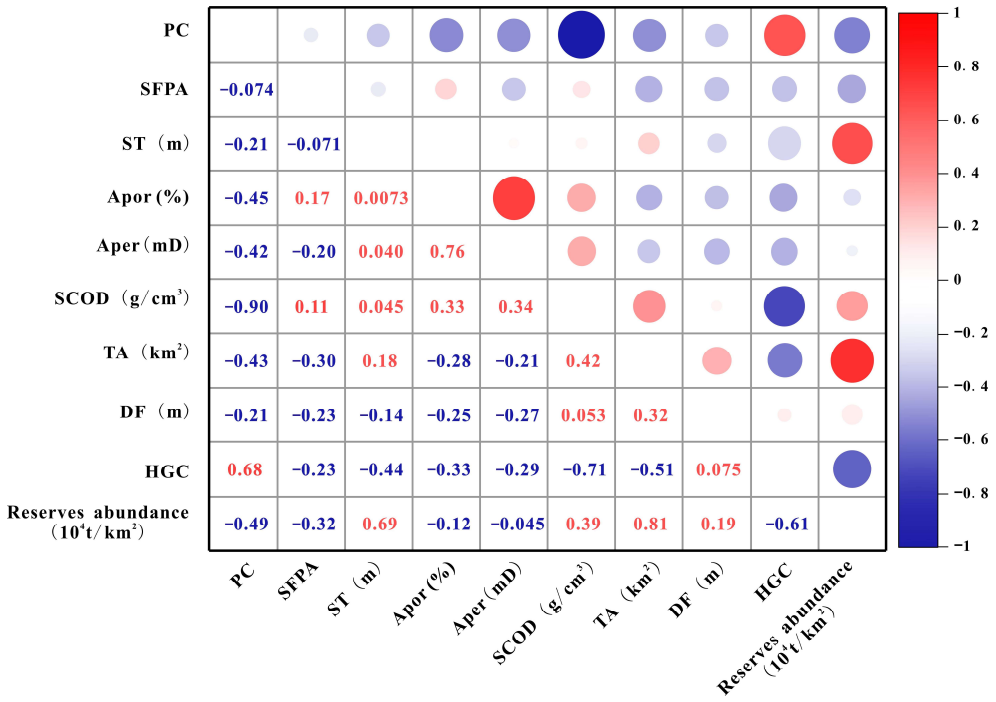


Figure 11. Matrix plot of Pearson correlation coefficients of characteristic parameters. (PC: pressure coefficient; SF: sedimentary facies; ST: sand thickness, m; Apor: average porosity, %; Aper: average permeability, mD; SCOD: surface crude oil density, g/cm^3; TA: trap area, km^2; DF: distance from fault, m; HGC: hydrocarbon generation center).

This study uses data from 80 discovered oil and gas reservoirs in the northern Dongpu Depression oil and gas aggregation area for modeling. The feature variables exhibit distinct distribution characteristics (Figure 12). The pressure coefficient, sandstone thickness, and surface crude oil density follow an approximately normal distribution, while the average porosity, permeability, and trap area are concentrated in narrower ranges. Specifically, trap area is mostly between 0.01 and 0.8 km^2, the average porosity is between 10% and 18%, permeability primarily ranges from 1 to 10 mD, and (Figure 12a,d,e). The distributions of SF and HGC show little variation (Figure 12b,i), with HGC values primarily between 0.4 and 0.7. To test the evaluation models of BP, BP-Adaboost, and BP-Bagging, data from 30 neighboring reservoirs, not included in the modeling, were selected. These reservoirs exhibit distribution characteristics similar to those in the training and test sets (Figure 12). In summary, neither the training data, nor the test data have exactly the same distributional characteristics, in this case it is undoubtedly a challenge to the traditional method of

assessing the abundance of geologic resources, and there is an urgent need to explore one or more evaluation models based on artificial intelligence methods.

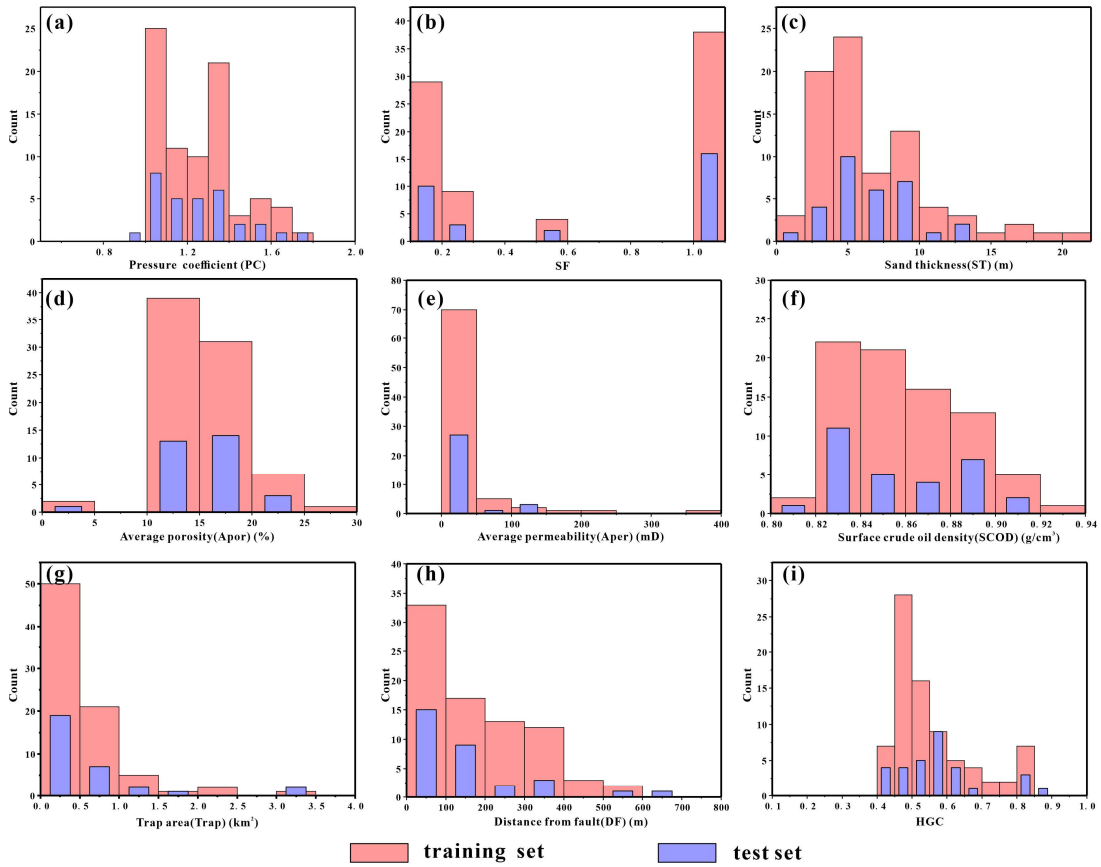


Figure 12. Gaussian distribution of training and test data sets. (**a**) The distribution characteristics of the pressure coefficient (PC) data; (**b**) The distribution characteristics of the sedimentary facies (SF) data; (**c**) The distribution characteristics of the sand thickness (ST, m) data; (**d**) The distribution characteristics of the average porosity (Apor, %) data; (**e**) The distribution characteristics of the average permeability (mD) data; (**f**) The distribution characteristics of the surface crude oil density (SCOD, g/cm^3) data; (**g**) The distribution characteristics of the trap area (TA, km^2) data; (**h**) The distribution characteristics of the distance from fault (DF, m) data; (**i**) The distribution characteristics of the hydrocarbon generation center (HGC) data.

5.2. Model Prediction Results of Different Machine Learning Algorithms

5.2.1. BP Model

This study focuses on 80 reservoirs within the research area. Given the limited data and large number of input parameters, a neural network model with one hidden layer was employed to prevent overfitting. The learning rate was set to 0.001, and the optimal number of neurons was determined using a search method. Figure 13 presents the MAE, RMSE, and R^2 values of the BP model with varying numbers of neurons in the hidden layer. The combination of evaluation metrics indicates that the model performs best with 4 neurons, yielding the lowest MAE and RMSE values and the highest correlation coefficient. Consequently, a BP model with a $9 \times 4 \times 1$ structure was selected.

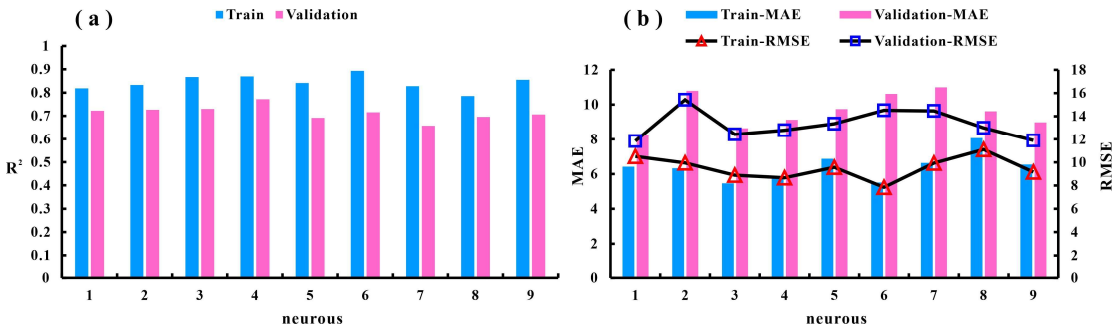


Figure 13. Performance of single-layer BP model with different number of neurons. (**a**) The correlation coefficients (R^2) of the model between the training and validation datasets under different neuron counts.; (**b**) The mean absolute error (MAE) and root mean square error (RMSE) of the model for the training and validation datasets under different neuron counts.

5.2.2. BP-AdaBoost Model

To ensure model comparability, the same dataset was used, with 70% allocated for training and 30% for validation, selected randomly. The weak learner in the BP-AdaBoost model is the BP model with the $9 \times 4 \times 1$ structure established earlier. The number of weak learners is determined via a search method. Figure 14 presents the MAE, RMSE, and R^2 for different numbers of weak learners in the ensemble. The results show that MAE, RMSE, and R^2 remain relatively stable when the number of weak learners exceeds seven. To avoid overfitting, particularly given the small sample size, seven weak learners were selected. The BP-AdaBoost model was then constructed, with iterative updates to the training data weight distribution based on previous BP model results, and the final prediction obtained by combining the weighted outputs of all weak classifiers.

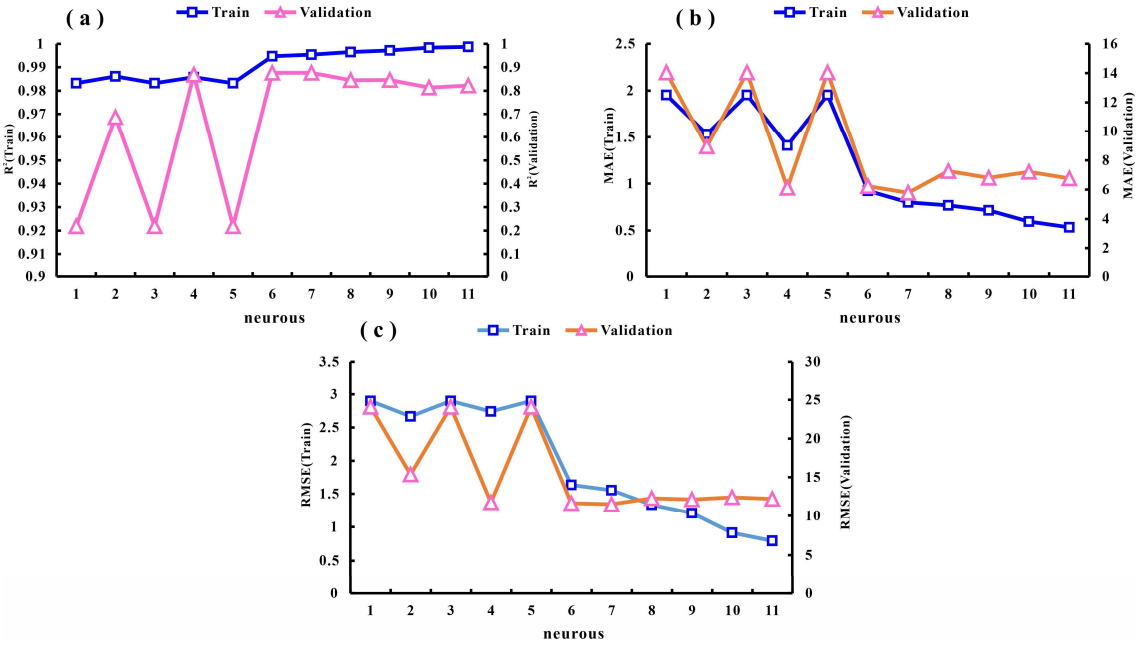


Figure 14. Performance of different base learners for BP-AdaBoost models. (**a**) The correlation coefficients (R^2) of the model between the training and validation datasets under different neuron counts.; (**b**) The mean absolute error (MAE) of the model for the training and validation datasets under different neuron counts. (**c**) The root mean square error (RMSE) of the model for the training and validation datasets under different neuron counts.

5.2.3. BP-Bagging Model

The bagging ensemble algorithm is applied to optimize the BP neural network model. The same dataset is used, with 70% for training and 30% for validation, selected randomly. The base learner in the BP-Bagging model is the BP model with the $9 \times 4 \times 1$ structure described earlier. The number of base learners is the same as in the BP-AdaBoost model, with seven BP models used as base learners. The final prediction is obtained by averaging the results of the seven base learners.

5.2.4. Model Reliability Analysis

Figure 15 compares the performance of the BP, BP-AdaBoost, and BP-Bagging models for oil and gas resource abundance prediction. Both the AdaBoost and Bagging algorithms significantly improve model performance, as evidenced by higher R^2 values and lower MAE and RMSE on the training and validation sets. However, the two ensemble methods exhibit different enhancement effects. While the BP-AdaBoost and BP-Bagging models achieve high R^2 values on the training set (0.9947 and 0.9773, respectively), the R^2 for the validation set is lower (0.876 and 0.8058). The BP-AdaBoost model demonstrates better stability on the validation set, with a smaller gap between the training and validation results (Figure 15). Overall, the AdaBoost method outperforms Bagging in optimizing model performance.

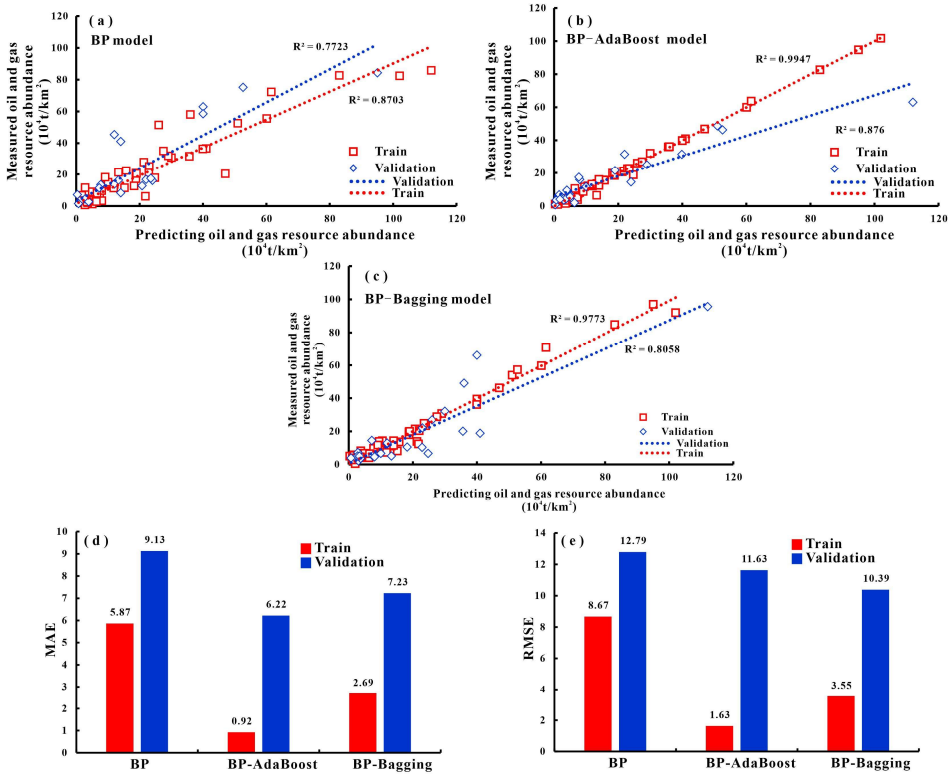


Figure 15. Performance comparison of three models. (**a**) show the correlation (R^2) analysis between the predicted oil and gas resource abundances and the measured abundances for Models BP, (**b**) show the correlation (R^2) analysis between the predicted oil and gas resource abundances and the measured abundances for Models BP-AdaBoost. (**c**) show the correlation (R^2) analysis between the predicted oil and gas resource abundances and the measured abundances for Models BP-Bagging. (**d**) presents the mean absolute error (MAE) of the predictions for Models BP, BP-AdaBoost, and BP-Bagging. (**e**) illustrates the root mean square error (RMSE) of the predictions for Models BP, BP-AdaBoost, and BP-Bagging.

5.3. Testing of Machine Learning Models

Thirty exploratory reservoirs from a neighboring area were selected to evaluate the BP, BP-AdaBoost, and BP-Bagging models. A total of 300 data points (30 reservoirs × 10 data points each) were collected and processed using the same preprocessing method. The combined evaluation metrics indicated the following test accuracies (R^2) for the models: 0.64 for BP, 0.77 for BP-AdaBoost, and 0.73 for BP-Bagging (Table 1). Among these, the BP-AdaBoost model demonstrated the best performance in assessing oil and gas resource abundance, with an R^2 of 0.77, MAE of 10.18, and RMSE of 20.86 (Table 1). Oil and gas resources are primarily concentrated in the structural high points of the central region of the study area, with higher resource potential in the northern region compared to the southern region. Furthermore, the predictions made by the BP-AdaBoost model, based on the test set, corroborate this distribution pattern (Figure 16). The BP-AdaBoost model effectively estimates the spatial distribution of oil and gas resource abundance, particularly in areas with resource abundance below 50 ($\times 10^4$ t/km^2), where its predictions are notably accurate, with observed values closely aligning with the predicted results (Figure 16). Despite training on a limited dataset, the results indicate that the BP-AdaBoost model maintains high reliability in predicting hydrocarbon resource potential in neighboring regions.

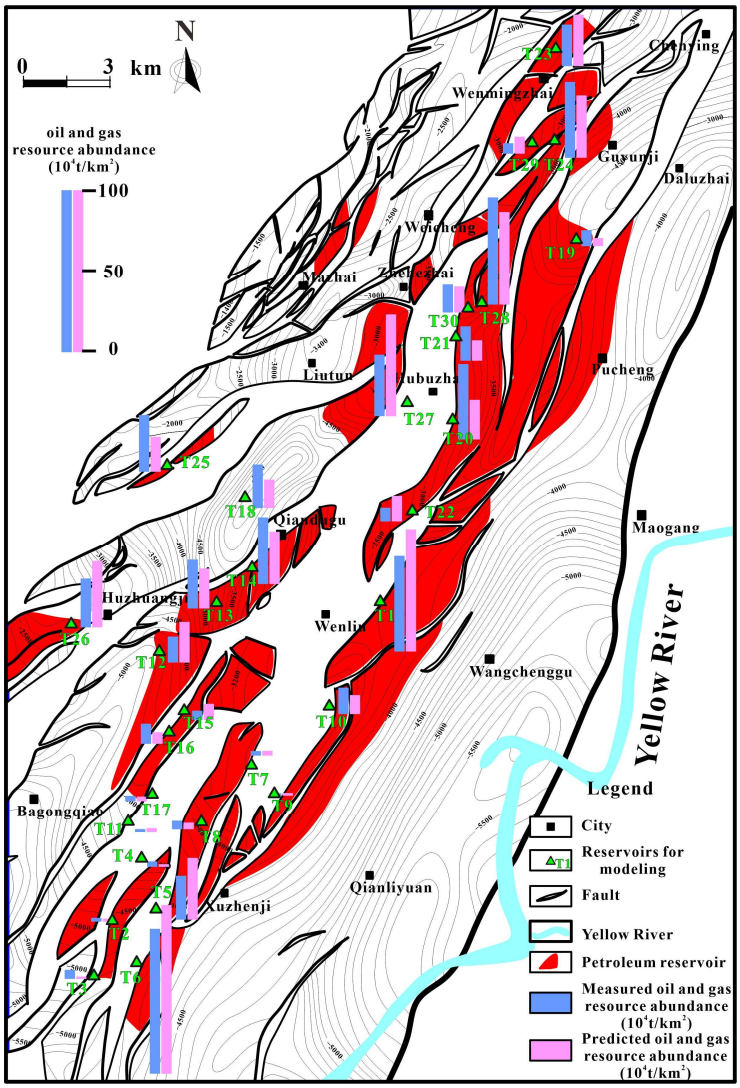


Figure 16. Comparison of actual hydrocarbon resource abundance with those predicted by BP-AdaBoost modeling.

Table 1. Comparison of performance evaluation indices of different neural network prediction models.

	R^2	MAE	RMSE
BP model	0.64	10.98	25.81
BP-AdaBoost model	0.77	10.18	20.86
BP-Bagging model	0.73	8.19	24.14

6. Discussion

6.1. Comparison with Other Prediction Models

In this study, an optimized BP neural network model based on an integrated learning strategy has been employed to predict the abundance of oil and gas resources. While the model achieves satisfactory results, it is important to recognize that other typical algorithms exist for regression problems in this area [78–80]. For comparison, several other models, including Classification and Regression Trees (CART), Random Forests (RF), Linear Support Vector Machines (LSVM), and Linear Regression (LR), were constructed to predict oil and gas resource abundance. These models were evaluated using the same training set and performance metrics. The results, presented in Table 2, reveal significant performance differences among the models. The BP-AdaBoost model outperforms all others, showing superior performance across all evaluated parameters. The RF model follows closely, exhibiting higher R^2 and lower MAE and RMSE values compared to the remaining models. The CART model ranks third, but while it performs better on the training set (higher R^2), it suffers from overfitting, as evidenced by lower R^2 and higher MAE and RMSE on the validation set. Both LR and LSVM demonstrate better stability with smaller discrepancies between training and validation sets, despite having lower accuracy. If the sample dataset is expanded, both models could offer improved accuracy and remain viable options for future research.

Table 2. Performance comparison of the BP-AdaBoost model with three classical machine learning models and a linear regression model.

Model	Train			Validation		
	R^2	MAE	RMSE	R^2	MAE	RMSE
BP-AdaBoost	0.99	0.92	1.63	0.88	6.22	11.63
RF	0.82	5.84	10.52	0.63	8.54	18.26
CART	0.86	4.34	7.88	0.53	10.67	21.48
LSVM	0.77	7.48	11.39	0.68	9.48	12.39
LR	0.80	6.28	9.37	0.72	10.74	14.73

In summary, based on R^2, MAE, and RMSE metrics, the models are ranked as follows: BP-AdaBoost > RF > CART > LR > LSVM. The results clearly indicate that the integrated learning algorithm outperforms single base-learners in establishing complex oil and gas resource abundance prediction models.

Although ensemble learning strategies have enhanced prediction accuracy, the influence of target parameter distribution on model performance remains significant. Specifically, the BP-AdaBoost model encounters difficulties in accurately predicting high oil and gas potential reservoirs, primarily due to the skewed distribution of the target parameter values. As illustrated in Figure 17, the majority of oil and gas resource abundance values fall within the range below 50 ($\times 10^4$ t/km^2). This non-normal distribution creates an imbalanced dataset, which may diminish the model's ability to learn from the high-abundance

samples that occur less frequently. Consequently, the model's applicability is reduced, with the model being more suitable for regions with oil and gas resource abundance values below 50 ($\times 10^4$ t/km^2).

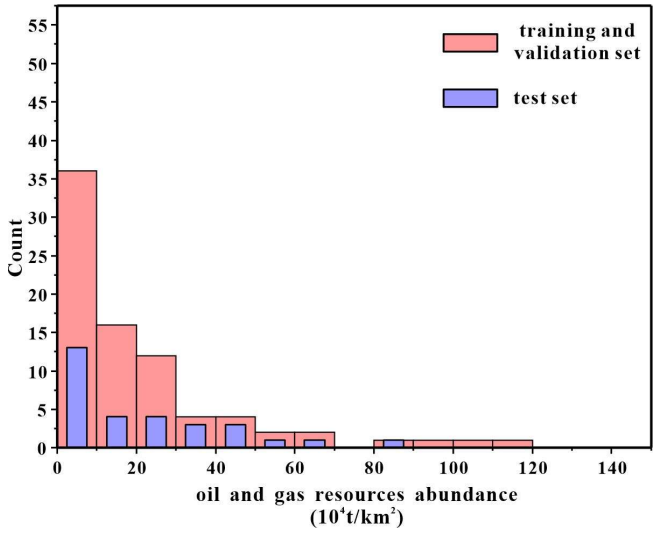


Figure 17. Distribution of oil and gas resource abundance.

6.2. Importance Analysis of Geological Parameters

The BP, BP-AdaBoost, and BP-Bagging models were used to assess the contribution of various geologic parameters to oil and gas resource abundance (Figure 18). All three models indicate that TA and ST are the most significant contributors to oil and gas resource abundance. The relative importance of the other parameters varied across the models. These findings are consistent with previous studies on factors influencing hydrocarbon aggregation [81–83]. Parameters such as SF, ST, Apor, and Aper are essential for evaluating reservoir quality and significantly influence oil and gas accumulation and reservoir capacity. Additionally, SF and faults control the primary migration pathways for oil and gas. Variations in ST, Apor, and Aper across layers affect migration capacity, while larger trap areas tend to correlate with better oil and gas accumulation, as larger traps enhance both accumulation and preservation [84]. The distance between the reservoir and the closest gully source fault serves as a key indicator of fault influence on oil and gas reservoirs [82].

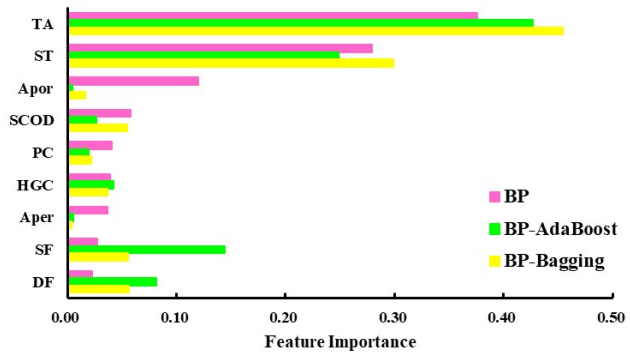


Figure 18. Importance analysis of nine characteristic parameters on oil and gas resource abundance based on neural network modeling. (TA: trap area, km^2; ST: sand thickness, m; Apor: average porosity, %; SCOD: surface crude oil density, g/cm^3; PC: pressure coefficient; HGC: hydrocarbon generation center; Aper: average permeability, mD; SF: sedimentary facies; DF: distance from fault, m).

This study identifies the ST, TA, SF and DF as the primary geologic parameters influencing the abundance of oil and gas resources. Both TA and ST provide essential storage

space for hydrocarbons, as illustrated in Figure 19a,b, where the abundance of hydrocarbon resources increases with both TA and ST. Larger trap sizes and thicker sand layers enhance the capture of hydrocarbons. Reservoir quality is further controlled by sedimentary facies, with oil and gas primarily distributed within shore-shallow lake facies (Figure 19c). Although porosity and permeability are also important factors influencing storage quality, they contribute less significantly to resource abundance in this study area [84]. Although porosity and permeability are crucial factors influencing reservoir quality, their contribution to the resource abundance in the study area is relatively minimal. As depicted in Figure 19d, oil and gas resource abundance exhibits an initial increase followed by a decrease with respect to porosity and permeability. This suggests that increases in porosity (10–20%) and permeability (1 mD–50 mD) within a certain range facilitate large-scale oil and gas accumulation. However, this also highlights the concentrated distribution of porosity and permeability values within a narrow range, which results in an imbalanced sample distribution. The HGC is another crucial factor for hydrocarbon accumulation, increasing HGC is associated with higher oil and gas resource abundance, as stronger discharge intensities from source rocks facilitate larger-scale accumulation. Hydrocarbon migration conditions also play a critical role in oil and gas aggregation. A negative correlation exists between the distance to faults connecting source rocks and the oil and gas abundance (Figure 19g), with greater distances corresponding to lower resource abundance [83]. Overpressure is a primary driving force for hydrocarbon migration. The pressure coefficient in the study area ranges from 1.0 to 1.3, indicating the presence of overpressure, which enhances migration and accumulation efficiency [85,86] (Figure 19h). Additionally, crude oil density, which ranges from 0.8 to 0.9 g/cm^3, falls within the light-medium oil category and supports large-scale migration and accumulation [60] (Figure 19i).

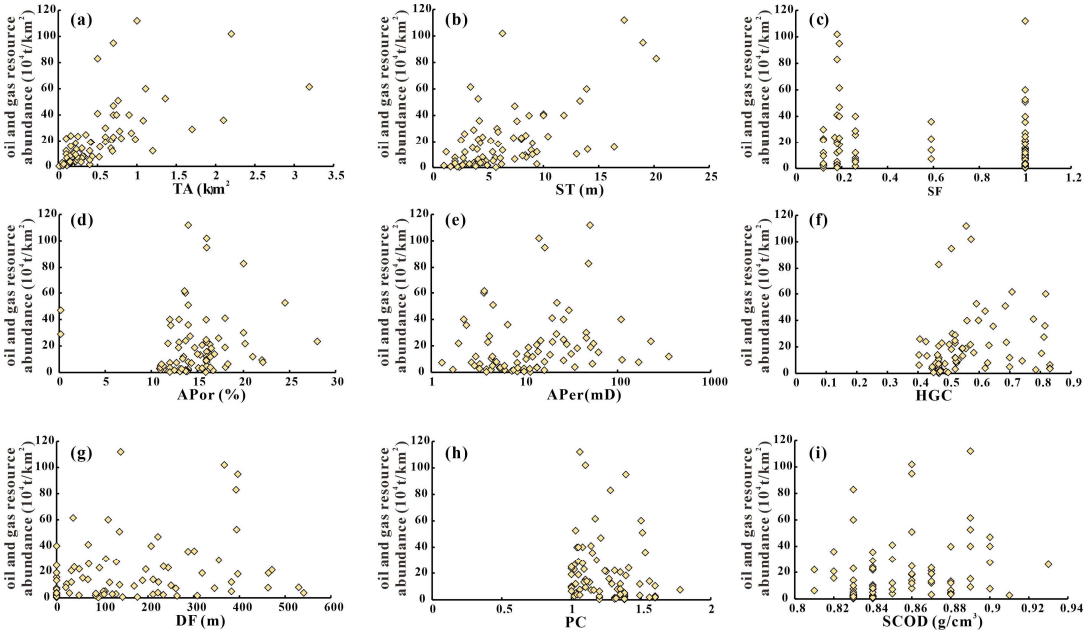


Figure 19. Scatterplot of nine geologic parameters in relation to hydrocarbon resource abundance. (**a**) trap area (TA), km^2; (**b**) sand thickness (ST), m; (**c**) sedimentary facies (SF); (**d**) average porosity (Apor), %; (**e**) average permeability (Aper), mD (**f**) hydrocarbon generation center (HGC); (**g**) distance from fault (DF), m; (**h**) pressure coefficient (PC); (**i**) surface crude oil density (SCOD), g/cm^3.

6.3. Attentions About Improving the Accuracy of Machine Learning Models

The proposed BP-AdaBoost method effectively predicts hydrocarbon resource abundance with superior model performance. However, given the complexity of geological

conditions, uncertainty remains in hydrocarbon resource abundance prediction. Therefore, we suggest the following optimization directions for future research:

(1) Optimizing Geological Parameters and Data Quality: Our focus is on improving dataset dimensionality and quality by incorporating a broader range of geological parameters. Additionally, we apply principal component analysis (PCA) to reduce the dimensionality of geological parameters, enabling the effective representation of multi-factor, multi-parameter, and low-dimensional data. This approach aims to mitigate the impact of data features on the prediction results. Furthermore, incorporating uncertainty analysis into the dataset, particularly for parameters with high variability, would provide a more comprehensive understanding of the data's inherent uncertainty and help improve the predictive capabilities of the model.

(2) Optimizing Prediction Models: Geological data in oil and gas exploration inherently carries uncertainty. Beyond enhancing prediction accuracy through ensemble learning strategies, swarm intelligence algorithms can be introduced to optimize models. This approach assists machine learning models in identifying optimal parameter configurations, thereby improving both accuracy and generalization capabilities.

(3) Quantitative Analysis of Uncertainty: The model developed in this study is applicable to adjacent blocks within sag areas, where the reservoirs are expected to exhibit conventional characteristics, including porosity ranging from 10% to 20% and permeability between 1 mD and 50 mD. Notably, the range of the target parameter—oil and gas resource abundance—is of paramount importance. The dataset collected for this study shows that the distribution of oil and gas resource abundance mainly spans from 0 to 50 ($\times 10^4$ t/km^2), with a lack of high-oil and gas resource abundance data, which limits the model's ability to effectively predict high-oil and gas resource abundance regions. To enhance the model's performance and address high-resource abundance areas, future research should incorporate uncertainty quantification techniques (such as Bayesian methods) to better capture the variability in these regions and optimize the predictive model accordingly. This would strengthen the machine learning model's ability to handle high-resource abundance areas. For other regions, a new database should be established based on local geological characteristics, and the machine learning algorithms proposed in this study should be employed to reconstruct evaluation and prediction models.

7. Conclusions

This study employed three distinct neural network models to predict hydrocarbon resource abundance. The results indicate that, compared to using the BP neural network alone, the two ensemble models show significant performance improvement in predicting oil and gas resource abundance. Specifically, the R^2 value for the BP neural network is 0.64, while BP-AdaBoost and BP-Bagging yield values of 0.77 and 0.73, respectively, demonstrating the strong predictive capability of the ensemble models. Among the two ensemble models, BP-AdaBoost achieves the highest overall prediction accuracy, with an RMSE value of 20.86, lower than the 24.14 obtained by BP-Bagging. Based on comprehensive evaluation metrics, BP-AdaBoost is recommended as the optimal model for predicting oil and gas resource abundance. The interpretability analysis of the three neural network models reveals that the trap area and sand thickness are the primary factors influencing oil and gas resource abundance. Machine learning methods have demonstrated high efficiency and effectiveness in assessing oil and gas resource potential. Future research integrating petroleum geology with machine learning algorithms can further enhance model accuracy by improving data quality and dimensionality, as well as optimizing algorithms. This

will enable machine learning approaches to deliver superior solutions for oil and gas exploration, development planning, and well location deployment.

Author Contributions: Data curation, Z.Y. and S.L.; Formal analysis, Q.W. and F.W.; Funding acquisition, D.C.; Investigation, S.L., S.C. and W.Z.; Methodology, Z.Y.; Project administration, D.C.; Software, F.W. and H.W.; Supervision, D.C.; Validation, S.C., W.Z., D.Y. and Y.W.; Visualization, D.Y., Y.W. and H.W.; Writing—original draft, Z.Y.; Writing—review and editing, Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (Grant No. 41972124). We gratefully acknowledge the Zhongyuan Oilfield Branch of the China Petroleum & Chemical Corporation for providing field test data and reservoir data.

Data Availability Statement: The original contributions presented in this study are included in the article material. Further inquiries can be directed to the corresponding author.

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

References

1. Zhao, W.Z.; Liang, K.; Wang, K. Hydrocarbon Security Strategy and Carbon Peak and Carbon Neutral Strategy: Relationship and Path. *Bull. Chin. Acad. Sci.* **2023**, *38*, 1–10. (In Chinese) [CrossRef]
2. Suslick, S.B.; Schiozer, D.J. Risk analysis applied to petroleum exploration and production: An overview. *J. Pet. Sci. Eng.* **2004**, *44*, 1–9. [CrossRef]
3. Sun, L.; Feng, Z.; Jiang, H.; Jiang, T. Responsibilities of petroleum prospectors: Discussions on dual logic and development trend of hydrocarbon exploration. *Pet. Explor. Dev.* **2021**, *48*, 999–1006. (In Chinese) [CrossRef]
4. White, D. Oil and Gas Play Maps in Exploration and Assessment: GEOLOGIC NOTE. *AAPG Bull.* **1988**, *72*, 944–949. [CrossRef]
5. Grant, S.M.; Milton, N.J.; Thompson, M. Play fairway analysis and risk mapping: An example using the Middle Jurassic Brent Group in the northern North Sea. *Nor. Pet. Soc. Spec. Publ.* **1996**, *6*, 167–181. [CrossRef]
6. Huang, X.H.; Bao, S.J.; Fu, Z.H.; You, Y.; Cheng, K.F.; Xu, Y.Q.; Mo, J.H.; Zheng, Z.H.; Zhu, C.G. Discussion on economic evaluation method of oil and gas exploration. *Pet. Explor. Dev.* **2000**, 9–13+109–110+118. (In Chinese)
7. Yan, Q.; Zhou, Z.Y. Establishment on Oil Resources Abundance Statistical Model in East China Rift Basins. *Pet. Geol. Exp.* **2009**, *31*, 292–295+306. (In Chinese)
8. Zhao, W.; Hu, S.; Wang, H.; Bian, C.; Wang, Z.C.; Wang, Z.Y. Large-scale accumulation and distribution of medium-low abundance hydrocarbon resources in China. *Pet. Explor. Dev.* **2013**, *40*, 1–13. (In Chinese) [CrossRef]
9. Zhang, W.; Liu, C.L.; Wu, X.Z.; Zheng, M.; Guo, Q.L.; Zhang, D.Y.; Chen, X.M.; Li, B. Statistical Characteristics and Prediction Models for Oil and Gas Resources Abundance in Different Types of Chinese Basins. *Geol. Explor.* **2019**, *55*, 1518–1527. (In Chinese)
10. Bédir, M.; El Asmi, A.M. New insights into Upper Cretaceous hydrocarbon traps of platform-basin flanks in the Sahel Eastern Tunisian petroleum province: Inferred optimal hydrocarbon reserves accumulations. *J. Pet. Sci. Eng.* **2022**, *220*, 111232. [CrossRef]
11. Akram, R.K.; Naqshabandi, S.F.; Sherwani, G. Integration of petrophysical and geochemical approaches for lower Miocene Euphrates and Jeribe Formations in selected wells in Tawke Oil Field, Northwestern Iraq, Zagros-fold belt. *J. Afr. Earth Sci.* **2023**, *205*, 105000. [CrossRef]
12. Lee, P.J.; Wang, P.C. Prediction of oil or gas pool sizes when discovery record is available. *Math. Geosci.* **1985**, *17*, 95–113. [CrossRef]
13. Liu, C.Z. Main Advances and Development Trends in Mathematicalgeology. *Geol. Rev.* **1996**, *4*, 364–368. (In Chinese) [CrossRef]
14. Otis, R.M.; Schneiderm, N. A Process for Evaluating Exploration Prospects. *AAPG Bull.* **1997**, *81*, 1087–1109. [CrossRef]
15. Zhou, Z.Y.; Bai, S.S.; He, H. Comparison of Genetic and Statistical Methods for Petroleum Resource Assessment. *Pet. Geol. Exp.* **2005**, 67–73. (In Chinese)
16. Komlosi, Z.; Komlosi, J. Application of the Monte Carlo Simulation in Calculating HC-Reserves. In Proceedings of the EUROPEC/EAGE Conference and Exhibition, Amsterdam, The Netherlands, 8–11 June 2009. [CrossRef]
17. Thander, B.; Sircar, A. Hydrocarbon Resource Estimation: Application of Monte Carlo Simulation. *IJLTEMAS* **2014**, *III*, 30–47.
18. Livshits, V.R.; Kontorovich, A.E. Distribution of Hydrocarbon Resources by Fields of Different Sizes and by the Number of Pools in Each Field. *Russ. Geol. Geophys.* **2022**, *63*, 1313–1319. [CrossRef]
19. Kuang, L.; Liu, H.; Ren, Y.; Luo, K.; Shi, M.; Su, J.; Li, X. Application and development trend of artificial intelligence in petroleum exploration and development. *Pet. Explor. Dev.* **2021**, *48*, 1–14. (In Chinese) [CrossRef]
20. Mutawa, A.M.; Hassouneh, A. Multimodal Real-Time Patient Emotion Recognition System Using Facial Expressions and Brain Eeg Signals Based on Machine Learning and Log-Sync Methods. *SSRN Electron. J.* **2024**, *91*, 105942. [CrossRef]

21. Olisah, C.C.; Trewhella, B.; Li, B.; Smith, M.L.; Winstone, B.; Whitfield, E.C.; Fern'andez, F.F.; Duncalfe, H. Convolutional Neural Network Ensemble Learning for Hyperspectral Imaging-based Blackberry Fruit Ripeness Detection in Uncontrolled Farm Environment. *Eng. Appl. Artif. Intell.* **2024**, *132*, 107945. [CrossRef]
22. Kuppenheimer, G.; Shelly, S.; Strauss, J. Can Machine Learning Identify Sector-Level Financial Ratios that Predict Sector Returns? *Financ. Res. Lett.* **2023**, *57*, 104241. [CrossRef]
23. Guo, Q.; Ren, H.; Yu, J.; Wang, J.; Liu, J.; Chen, N. A method of predicting oil and gas resource spatial distribution based on Bayesian network and its application. *J. Pet. Sci. Eng.* **2021**, *208*, 109267. [CrossRef]
24. Alolayan, O.S.; Raymond, S.J.; Montgomery, J.B.; Williams, J.R. Towards better shale gas production forecasting using transfer learning. *Upstream Oil Gas Technol.* **2022**, *9*, 100072. [CrossRef]
25. Wang, M.; Hui, G.; Pang, Y.; Wang, S.; Chen, S. Optimization of machine learning approaches for shale gas production forecast. *Geoenergy Sci. Eng.* **2023**, *226*, 211719. [CrossRef]
26. Ren, H.J.; Guo, Q.L.; Cao, Z.; Ren, H.B. Risk prediction for petroleum exploration based on Bayesian network classifier. *Geoenergy Sci. Eng.* **2023**, *228*, 211924. [CrossRef]
27. Hou, M.; Xiao, Y.; Lei, Z.; Yang, Z.; Lou, Y.; Liu, Y. Machine learning algorithms for lithofacies classification of the gulong shale from the Songliao Basin, China. *Energies* **2023**, *16*, 2581. [CrossRef]
28. Sun, J.; Dang, W.; Wang, F.; Nie, H.; Wei, X.; Li, P.; Zhang, S.; Feng, Y.; Li, F. Prediction of TOC content in organic-rich shale using machine learning algorithms: Comparative study of random forest, support vector machine, and XGBoost. *Energies* **2023**, *16*, 4159. [CrossRef]
29. Wang, H.; Guo, Z.; Kong, X.; Zhang, X.; Wang, P.; Shan, Y. Application of Machine Learning for Shale Oil and Gas "Sweet Spots" Prediction. *Energies* **2024**, *17*, 2191. [CrossRef]
30. Wang, E.; Fu, Y.; Guo, T.; Li, M. A new approach for predicting oil mobilities and unveiling their controlling factors in a lacustrine shale system: Insights from interpretable machine learning model. *Fuel* **2025**, *379*, 132958. [CrossRef]
31. Fan, Z.; Liu, X.; Wang, Z.; Liu, P.; Wang, Y. A novel ensemble machine learning model for oil production prediction with two-stage data preprocessing. *Processes* **2024**, *12*, 587. [CrossRef]
32. Jiao, S.; Li, W.; Li, Z.; Gai, J.; Zou, L.; Su, Y. Hybrid physics-machine learning models for predicting rate of penetration in the Halahatang oil field, Tarim Basin. *Sci. Rep.* **2024**, *14*, 5957. [CrossRef]
33. Delavar, M.R.; Ramezanzadeh, A. Machine learning classification approaches to optimize ROP and TOB using drilling and geomechanical parameters in a carbonate reservoir. *J. Pet. Explor. Prod. Technol.* **2024**, *14*, 1–26. [CrossRef]
34. Kumar, J.; Mukherjee, B.; Sain, K. Porosity prediction using ensemble machine learning approaches: A case study from Upper Assam basin. *J. Earth Syst. Sci.* **2024**, *133*, 99. [CrossRef]
35. Meng, J.; Zhou, Y.J.; Ye, T.R.; Xiao, Y.T.; Lu, Y.Q.; Zheng, A.W.; Liang, B. Hybrid data-driven framework for shale gas production performance analysis via game theory, machine learning, and optimization approaches. *Pet. Sci.* **2023**, *20*, 277–294. [CrossRef]
36. Latrach, A.; Malki, M.L.; Morales, M.; Mehana, M.; Rabiei, M. A critical review of physics-informed machine learning applications in subsurface energy systems. *Geoenergy Sci. Eng.* **2024**, *239*, 212938. [CrossRef]
37. Chi, P.; Sun, J.; Zhang, R.; Luo, X.; Yan, W. Reconstruction of large-scale anisotropic 3D digital rocks from 2D shale images using generative adversarial network. *Mar. Pet. Geol.* **2024**, *170*, 107065. [CrossRef]
38. Niu, W.; Sun, Y.; Yang, X.; Lu, J.; Zhao, S.; Yu, R.; Liang, P.; Zhang, J. Toward Production Forecasting for Shale Gas Wells Using Transfer Learning. *Energy Fuels* **2023**, *37*, 5130–5142. [CrossRef]
39. Yu, H.B.; Cheng, X.J.; Qi, J.F.; Tan, Y.M.; Xu, T.W. Effects of Paleogene faulting on the subsag evolution and hydrocarbon generation in Dongpu Sag. *Pet. Geol. Recovery Effic.* **2018**, *25*, 24–31. (In Chinese) [CrossRef]
40. Jiang, Y.L.; Fang, L.; Liu, J.D.; Hu, H.J.; Xu, T.W. Hydrocarbon charge history of the Paleogene reservoir in the northern Dongpu Depression, Bohai Bay Basin, China. *Pet. Sci.* **2016**, *13*, 625–641. [CrossRef]
41. Zhang, K.X.; Qi, J.F.; Zhao, Y.B.; Chen, S.P. Structure and Evolution of Cenozoic in Dongpu Sag. *Xinjiang Pet. Geol.* **2007**, 714–717. (In Chinese)
42. Xu, H.; Wang, X.W.; Yan, D.P.; Qiu, L. Subsidence transition during the post-rift stage of the Dongpu Sag, Bohai Bay Basin, NE China: A new geodynamic model. *J. Asian Earth Sci.* **2018**, *158*, 186–199. [CrossRef]
43. Zhu, C.; Jiang, F.; Zhang, P.; Hu, T.; Liu, Y.; Xu, T.; Zhang, Y.; Deng, Q.; Zhou, Y.; Xiong, H. Identification of effective source rocks in different sedimentary environments and evaluation of hydrocarbon resources potential: A case study of paleogene source rocks in the Dongpu Depression, Bohai Bay Basin. *J. Pet. Sci. Eng.* **2021**, *201*, 108477. [CrossRef]
44. Liu, Q.; He, L.; Yi, Z.; Zhang, L. Anomalous Post-Rift Subsidence in the Bohai Bay Basin, Eastern China: Contributions From Mantle Process and Fault Activity. *Tectonics* **2022**, *41*, e2021TC006748. [CrossRef]
45. Tan, Y.M.; Chen, X.J.; Chen, S.P.; He, F. Complex fault-block groups in Dongpu Sag and their exploration potential. *Oil Gas Geol.* **2011**, *32*, 584–592. (In Chinese) [CrossRef]
46. Shang, M.H. Relationship between structural-depositional evolution and oil-gas accumulation in Dongpu sag. *Pet. Geol. Recovery Effic.* **2014**, *1*, 50–53+57+114. (In Chinese) [CrossRef]

47. Jiang, Y.L.; Sun, S.M.; Xin, F.L.; Tan, Y.M.; Liu, J.D. Heterogeneity of hydrocarbon distribution and its main controlling factors in oil-rich depression. *J. China Univ. Pet. (Ed. Nat. Sci.)* **2019**, *43*, 34–43. (In Chinese)
48. Wang, K.; Pang, X.; Zhang, H.; Hu, T.; Xu, T.; Zheng, T.; Zhang, X. Organic geochemical and petrophysical characteristics of saline lacustrine shale in the Dongpu Depression, Bohai Bay Basin, China: Implications for Es3 hydrocarbon exploration. *J. Pet. Sci. Eng.* **2020**, *184*, 106546. [CrossRef]
49. Leng, Y.Y.; Qian, M.H.; Lu, K.; Xu, E.S.; Zhou, Y.S.; Bao, Y.J.; Li, Z.M.; Jiang, Q.G. Enrichment types and hydrocarbon composition characteristics of shale oil in the northern part of Dongpu Sag, Bohai Bay Basin: A case study of the third member of Paleogene Shahejie Formation of well Wen 410. *Pet. Geol. Exp.* **2022**, *44*, 1028–1036. (In Chinese)
50. Huang, J.J.; Ji, Y.L.; Wang, G.W.; Chen, F.L.; Han, F.M. Sequence characteristics and genesis of the Eogene salt-bearing formation in Dongpu depression. *Oil Gas Geol.* **2007**, *28*, 479–484. (In Chinese) [CrossRef]
51. Tang, L.; Song, Y.; Pang, X.; Jiang, Z.; Guo, Y.; Zhang, H.; Pan, Z.; Jiang, H. Effects of paleo sedimentary environment in saline lacustrine basin on organic matter accumulation and preservation: A case study from the Dongpu Depression, Bohai Bay Basin, China. *J. Pet. Sci. Eng.* **2020**, *185*, 106669. [CrossRef]
52. Ji, Y.L.; Feng, J.H.; Wang, S.L.; Tan, Y.M.; Zhang, H.A.; Wang, D.R. Origin of Salt and Gypsum Rock in the Third Member of Shahe Jie Formation of Lower Tertiary in Dongpu Depression. *Acta Sedimentol. Sin.* **2005**, 225–231. (In Chinese)
53. Hu, T.; Pang, X.Q.; Jiang, F.J.; Wang, Q.F.; Xu, T.W.; Wu, G.Y.; Cai, Z.; Yu, J.W. Factors Controlling Differential Enrichment of Organic Matter in Saline Lacustrine Rift Basin: A case study of third member Shahejie Fm in Dongpu Depression. *Acta Sedimentol. Sin.* **2021**, *39*, 140–152. (In Chinese) [CrossRef]
54. Sun, N.; Chen, T.; Gao, J.; Zhong, J.; Huo, Z.; Qu, J. Lithofacies and reservoir characteristics of saline lacustrine fine-grained sedimentary rocks in the northern Dongpu Sag, Bohai Bay Basin: Implications for shale oil exploration. *J. Asian Earth Sci.* **2023**, *252*, 105686. [CrossRef]
55. Yang, S.; Yan, X.B.; Cai, L.X.; Liu, Z.P.; Hang, X.X. Parameter definition in resource calculation of a trap: A fault block example. *Pet. Geol. Exp.* **2015**, *37*, 530–534. (In Chinese) [CrossRef]
56. Yang, S.; Yan, X.B.; Liu, Z.P.; Cai, L.X.; Ma, X.J. Resource calculation methods for multilayer traps: Case study of two-layer traps. *Pet. Geol. Exp.* **2016**, *38*, 698–702. (In Chinese) [CrossRef]
57. Li, S.; Ji, H.; Wan, Z.; Pang, X.; Zhang, H.; Xu, T.; Zhou, Y. Geochemical characteristics and factors controlling the deep lithologic reservoirs in Puwei Sag, Dongpu Depression—A Case Study of Well PS20. *J. Pet. Sci. Eng.* **2021**, *203*, 108669. [CrossRef]
58. Tan, Y.M.; Li, H.L.; Zhang, Y.X.; Xu, T.W.; Zhou, Y.S.; Gong, R.C. Analysis to high quality source rock characteristics and residual resource potential in Dongpu Sag in Paleogene. *Fault-Block Oil Gas Field* **2020**, *27*, 551–555+572. (In Chinese)
59. Li, H.; Wang, B.H.; Lu, J.L.; Lu, K.; Wang, M.; Zhou, Y.; Zhao, L.J. Geological characteristics and exploration prospects of Paleogene continental shale oil accumulation in Dongpu Sag, Bohai Bay Basin. *J. China Univ. Pet. (Ed. Nat. Sci.)* **2021**, *45*, 33–41. (In Chinese)
60. Guo, W.; Zhang, X.; Sun, Y.; Li, Q.; Liu, Z. Migration mechanism of pyrolysis oil during oil shale in situ pyrolysis exploitation. *Energy* **2023**, *285*, 128769. [CrossRef]
61. Luor, D.C. A comparative assessment of data standardization on support vector machine for classification problems. *Intell. Data Anal.* **2015**, *19*, 529–546. [CrossRef]
62. Hu, C.Y. Research on the Appliance Extent of "Source Control Theory" by Semi Quantitative Statistics Characteristics of Oil and Gas Migration Distance. *Nat. Gas Ind.* **2005**, *25*, 1–3. (In Chinese) [CrossRef]
63. Xia, Q.L.; Pang, X.Q.; Jiang, F.J.; Ma, X.Z. Control of source rock on hydrocarbon accumulation and prediction of favorable plays in the Bozhong Depression of the Bohai Sea waters. *Oil Gas Geol.* **2009**, *30*, 398–404+411. (In Chinese)
64. Pang, X.Q.; Yu, Q.H.; Guan, X.Y.; Li, S.M.; Jiang, F.J. Evolution and Movement of Source Kitchens and Their Control of Oil and Gas in the Tarim Cratonic Basin, China. *Energy Explor. Exploit.* **2012**, *30*, 239–272. [CrossRef]
65. Pang, X.Q.; Huo, Z.P.; Fan, B.J.; Dong, Y.X.; Jiang, T. Control of source rocks on hydrocarbon accumulation and assessment of gas pools in the Nanpu Sag, Bohai Bay Basin. *Nat. Gas Ind.* **2014**, *34*, 28–36. (In Chinese)
66. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
67. Song, S.; Xiong, X.; Wu, X.; Xue, Z. Modeling the SOFC by BP neural network algorithm. *Int. J. Hydrogen Energy* **2021**, *46*, 20065–20077. [CrossRef]
68. Tian, J.; Liu, Y.; Zheng, W.; Yin, L. Smog prediction based on the deep belief—BP neural network model (DBN-BP). *Urban Clim.* **2022**, *41*, 101078. [CrossRef]
69. Ma, X.Y.; Ma, X.J.; Wang, Z.; Song, S.; Sheng, Y. Investigation of changing SARA and fatigue properties of asphalt bitumen under ageing and analysis of their relation based upon the BP neural network. *Constr. Build. Mater.* **2023**, *394*, 132163. [CrossRef]
70. Liu, Y.; Zhao, C.; Liang, H.; Lu, H.; Cui, N.; Bao, K. A rotor fault diagnosis method based on BP-Adaboost weighted by non-fuzzy solution coefficients. *Measurement* **2022**, *196*, 111280. [CrossRef]
71. Yan, W.; Yang, D.; Zhang, Y.; Li, B. Twin-array capacitance sensor for multi-parameter measurement of gas-solid particle flow based on BP- Adaboost. *Flow Meas. Instrum.* **2023**, *94*, 102445. [CrossRef]

72. Kotsiantis, S.B.; Kanellopoulos, D.; Zaharakis, I.D. Bagged Averaging of Regression Models. In *IFIP International Federation for Information Processing*; Springer: Boston, MA, USA, 2006; Volume 204. [CrossRef]
73. Khwaja, A.S.; Naeem, M.; Anpalagan, A.; Venetsanopoulos, A.; Venkatesh, B. Improved short-term load forecasting using bagged neural networks. *Electr. Power Syst. Res.* **2015**, *125*, 109–115. [CrossRef]
74. Ha, K.; Cho, S.; MacLachlan, D. Response models based on bagging neural networks. *J. Interact. Mark.* **2005**, *19*, 17–30. [CrossRef]
75. Li, H.W.; Liu, J.N.; Yang, Y.; Lu, G.L.; Qiao, B.X. Coupling flow channel optimization and Bagging neural network to achieve performance prediction for proton exchange membrane fuel cells with varying imitated water-drop block channel. *Int. J. Hydrogen Energy* **2022**, *47*, 39987–40007. [CrossRef]
76. Wan, T.; Tan, Y.M.; Su, H.; Wang, X.W.; Ni, J.F.; Liu, H.M.; Wan, J. Palaeogeomorphology of Middle Es3 Formation in Pucheng area of Dongpu depression and its relationship with sedimentary facies. *Geol. China* **2014**, *41*, 206–214. (In Chinese) [CrossRef]
77. Li, X.; Zhang, J.; Xie, J.; Li, C.; Dai, Y.; Li, W.; Zhang, Y.; Li, S. Sedimentary and sequence-stratigraphic characteristics of the lower second submember, Shahejie formation, M1 block, Wenmingzhai oilfield, Dongpu depression, China. *Arab. J. Geosci.* **2015**, *8*, 5397–5406. [CrossRef]
78. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
79. Kori, G.S.; Kakkasageri, M.S. Classification and Regression Tree (CART) based resource allocation scheme for Wireless Sensor Networks. *Comput. Commun.* **2023**, *197*, 242–254. [CrossRef]
80. Subhashini, R.; Amudha, V. Evaluation of High Reliability in 5G Network using Novel SVM Comparison with LSVM To Improve Accuracy. In Proceedings of the 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 6–7 April 2023. [CrossRef]
81. Mi, L.; Zhang, Z.; Pang, X.; Liu, J.; Zhang, B.; Zhao, Q.; Feng, X. Main controlling factors of hydrocarbon accumulation in Baiyun Sag at northern continental margin of South China Sea. *Pet. Explor. Dev.* **2018**, *45*, 963–973. (In Chinese) [CrossRef]
82. Wang, F.; Chen, D.; Wang, Q.; Shi, X.; Xie, G.; Wang, Z.; Li, J.; Liao, W. Evolution characteristics of transtensional faults and their impacts on hydrocarbon migration and accumulation: A case study from the Huimin Depression, Bohai Bay Basin, eastern China. *Mar. Pet. Geol.* **2020**, *120*, 104507. [CrossRef]
83. Perkins, J.R.; Fraser, A.J.; Muxworthy, A.R.; Neumaier, M.; Schenk, O. Basin and petroleum systems modelling to characterise multi-source hydrocarbon generation: A case study on the inner Moray Firth, UK North Sea. *Mar. Pet. Geol.* **2023**, *151*, 106180. [CrossRef]
84. Li, W. Characteristics and exploration prospects of low-maturity oil reserves in the Bohai Bay Basin: A case study on the southern sag of the Liaoxi Depression. *Unconv. Resour.* **2022**, *2*, 116–123. [CrossRef]
85. Marghani, M.M.A.; Zairi, M.; Radwan, A.E. Facies analysis, diagenesis, and petrophysical controls on the reservoir quality of the low porosity fluvial sandstone of the Nubian formation, east Sirt Basin, Libya: Insights into the role of fractures in fluid migration, fluid flow, and enhancing the permeability of low porous reservoirs. *Mar. Pet. Geol.* **2023**, *147*, 105986. [CrossRef]
86. Li, C.; Zhang, L.; Luo, X.; Lei, Y.; Yu, L.; Cheng, M.; Wang, Y.; Wang, Z. Overpressure generation by disequilibrium compaction or hydrocarbon generation in the Paleocene Shahejie Formation in the Chezhen Depression: Insights from logging responses and basin modeling. *Mar. Pet. Geol.* **2021**, *133*, 105258. [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.