



Reservoir Parameter Prediction Based on the Neural Random Forest Model

Mingchuan Wang*, Dongjun Feng, Donghui Li and Jiwei Wang

SINOPEC Petroleum Exploration and Production Research Institute, Beijing, China

Porosity and saturation are the basis for describing reservoir properties and formation characteristics. The traditional, empirical, and formulaic methods are unable to accurately capture the nonlinear mapping relationship between log data and reservoir physical parameters. To solve this problem, in this study, a novel hybrid model (NRF) combining neural network (NN) and random forest (RF) was proposed based on well logging data to predict the porosity and saturation of shale gas reservoirs. The database includes six horizontal wells, and the input logs include borehole diameter, neutron, density, gamma-ray, and acoustic and deep investigate double lateral resistivity log. The porosity and saturation were chosen as outputs. The NRF model with independent and joint training was designed to extract key features from well log data and physical parameters. It provides a promising method for forecasting the porosity and saturation with R^2 above 0.94 and 0.82 separately. Compared with baseline models (NN and RF), the NRF model with joint training obtains the unsurpassed performance to predict porosity with R^2 above 0.95, which is 1.1% higher than that of the NRF model with independent training, 3.9% higher than RF, and superiorly greater than NN. For the prediction of saturation, the NRF model with joint training is still superior to other algorithms, with R^2 above 0.84, which is 2.1% higher than that of the NRF model with independent training and 7.0% higher than RF. Furthermore, the NRF model has a similar data distribution with measured porosity and saturation, which demonstrates the NRF model can achieve greater stability. It was proven that the proposed NRF model can capture the complex relationship between the logging data and physical parameters more accurately, and can serve as an economical and reliable alternative tool to give a reliable prediction.

Keywords: logging interpretation, machine learning, reservoir parameter estimation, neural random forest, well logs

OPEN ACCESS

Edited by:

Kai Zhang,
China University of Petroleum, China

Reviewed by:

Olabode Ijasan,
ExxonMobil, United States
Leila Aliouane,
University of Boumerdés, Algeria

*Correspondence:

Mingchuan Wang
wangmc.syky@sinopec.com

Specialty section:

This article was submitted to
Environmental Informatics and Remote
Sensing,
a section of the journal
Frontiers in Earth Science

Received: 03 March 2022

Accepted: 11 April 2022

Published: 13 May 2022

Citation:

Wang M, Feng D, Li D and Wang J
(2022) Reservoir Parameter Prediction
Based on the Neural Random
Forest Model.
Front. Earth Sci. 10:888933.
doi: 10.3389/feart.2022.888933

1 INTRODUCTION

Logging curves can reflect different lithology and formation characteristics. In recent years, the prediction of reservoir parameters using log curve data has become a focus of research.

Reservoir physical parameters, which mainly include porosity, permeability, water saturation, and oil saturation, are the basis for describing reservoir properties and reservoir modeling (Li et al., 2016; Wang et al., 2019; Song et al., 2021a). Specifically, the pore space of a reservoir is an important space for the accumulation and transport of hydrocarbons. It is also a fundamental requirement for the formation of hydrocarbon reservoirs. The size of porosity directly reflects the ability of the rock to store hydrocarbons. It is an important part of reservoir evaluation and occupies a very important position in the exploration and development of oil and gas fields. The oil, gas, and water saturation in

a reservoir is also a fundamental parameter for estimating oil reserves and judging the characteristics of the reservoir.

At present, there are two major approaches used for the measurement of porosity and saturation (Song et al., 2020; Wang et al., 2020; Song et al., 2021b). The first method is a direct estimation, namely, obtaining the actual physical parameter data using rock slices or cores. This method is often performed in the laboratory and is more accurate, but it is time-consuming and costly. The second method is the indirect measurement, namely, estimating the porosity and saturation based on the geological and statistical methods as well as function approximation (Wyllie et al., 1956; Raymer et al., 1980). Early logging methods for predicting reservoir parameters are mainly aimed at linear data. The physical parameters are calculated using linear equations and empirical formulas, which is a purely mathematical method without considering the actual reservoirs' environment.

However, some reservoirs are highly heterogeneous, and the geological environment is complex. In addition, the relationship between the logging data and reservoir parameters is nonlinear. The traditional regression analysis methods are difficult to achieve satisfactory results. Therefore, exploring a novel method for reservoir parameter prediction is particularly necessary for the development of unconventional and complex oil and gas fields.

With the rise of the emergence of big data and artificial intelligence, machine learning has been rapidly developed and applied. Some researchers have obtained the reservoir parameters such as permeability, porosity, and saturation. Akande et al. (2015) proposed an artificial neural network (ANN) based on the correlation feature selection to predict permeability. The results show that this method can predict permeability with fewer features. Komarialaei and Salahshoor (2012) also used the ANN model with the principal component analysis (PCA) to predict permeability. The experimental results show that the method has certain practicality. Hadi and Sadegh (2016) predicted the porosity using an intelligent method based on seismic attribute data. J. Song et al. (2016) introduced the random forest method to predict seismic reservoirs, and it is found that the method is less affected by noisy data and has certain stability and accuracy.

Predicting the porosity and saturation based on well logs using machine learning algorithms is a feasible and alternative method. However, the accuracy cannot fully meet the requirement. Therefore, the complex nonlinear relationship is still needed to be further explored.

2 METHODOLOGY

In this section, the methodologies of the neural random forest algorithm are introduced systematically. As the basis of the neural random forests, the principle of neural network and random forests are presented at the first, and the ensemble of neural network and random forests is introduced later. Finally, the evaluation criteria of the machine learning model are presented.

2.1 Model Establishing

2.1.1 Neural Network

The neural network (NN) is a robust and effective computational tool for establishing nonlinear patterns between the complex nonlinear data. In particular, supervised learning is adopted for most applications (Rolon et al., 2009; Khandelwal and Singh, 2010; Saputro et al., 2016). The typical NN contains an input layer, an output layer, and more than one hidden layer (Gardner and Dorling, 1998; Basheer and Hajmeer, 2000; Schmidhuber, 2015; Prieto et al., 2016).

In the training process, the weight and threshold between each neuron of the neural network are adjusted continuously.

Suppose the training data are $D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(r)}, y^{(r)}), \dots, (x^{(m)}, y^{(m)})\}$, $w^{(l)}$ is the weight matrix from the $l-1$ layer to l , $w_{jk}^{(l)}$ is the weight from the k -th neuron in the $l-1$ layer to the j -th neuron in the l layer. $b_j^{(l)}$ is the bias of the j -th neuron in the l layer. $z_j^{(l)}$ is the input of the j -th neuron in the l layer. Then the input of each neuron in each layer is expressed as Eq. 1.

$$\begin{bmatrix} z_1^{(l)} \\ z_2^{(l)} \\ \vdots \\ z_{N^{(l)}}^{(l)} \end{bmatrix} = \begin{bmatrix} w_{11}^{(l)} & w_{12}^{(l)} & \cdots & w_{1N^{(l-1)}}^{(l)} \\ w_{21}^{(l)} & w_{22}^{(l)} & \cdots & w_{2N^{(l-1)}}^{(l)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N^{(l)}1}^{(l)} & w_{N^{(l)}2}^{(l)} & \cdots & w_{N^{(l)}N^{(l-1)}}^{(l)} \end{bmatrix} \begin{bmatrix} a_1^{(l-1)} \\ a_2^{(l-1)} \\ \vdots \\ a_{N^{(l-1)}}^{(l-1)} \end{bmatrix} + \begin{bmatrix} b_1^{(l-1)} \\ b_2^{(l-1)} \\ \vdots \\ b_{N^{(l-1)}}^{(l-1)} \end{bmatrix} \quad (1)$$

Namely,

$$z^{(l)} = w^{(l)} a^{(l-1)} + b^{(l)}. \quad (2)$$

Among them, $a^{(l)} = \sigma(z^{(l)})$, where $\sigma(x)$ is the activation function. Then,

$$a^{(l)} = \sigma(w^{(l)} a^{(l-1)} + b^{(l)}). \quad (3)$$

Finally, in the forward propagation, the output of the neural network is expressed as

$$f(x; \theta) = \sigma(w^{(L)} \cdots \sigma(w^{(2)} \sigma(w^{(1)} x + b^{(1)}) + b^{(2)}) \cdots + b^{(L)}), \quad (4)$$

where L represents the output layer of the neural network, and

$$\theta = \{w^{(1)}, b^{(1)}, w^{(2)}, b^{(2)}, \dots, w^{(L)}, b^{(L)}\}. \quad (5)$$

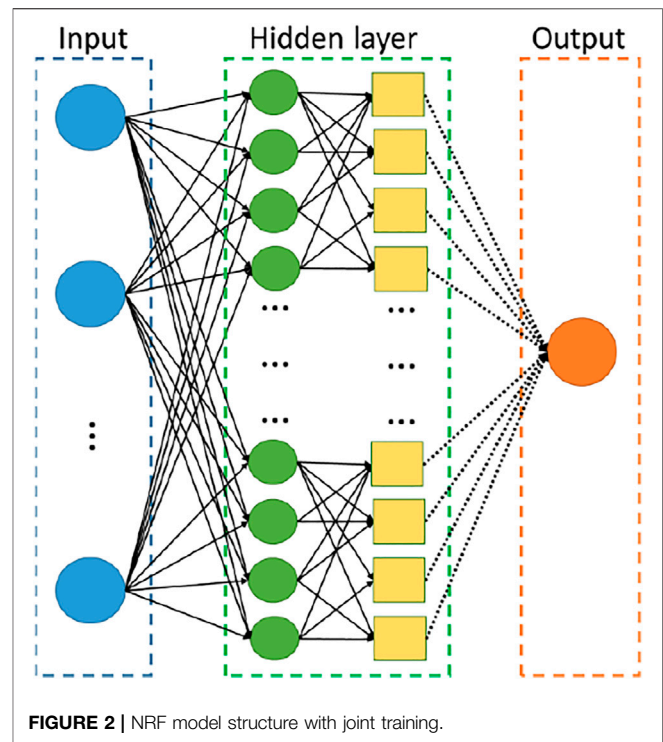
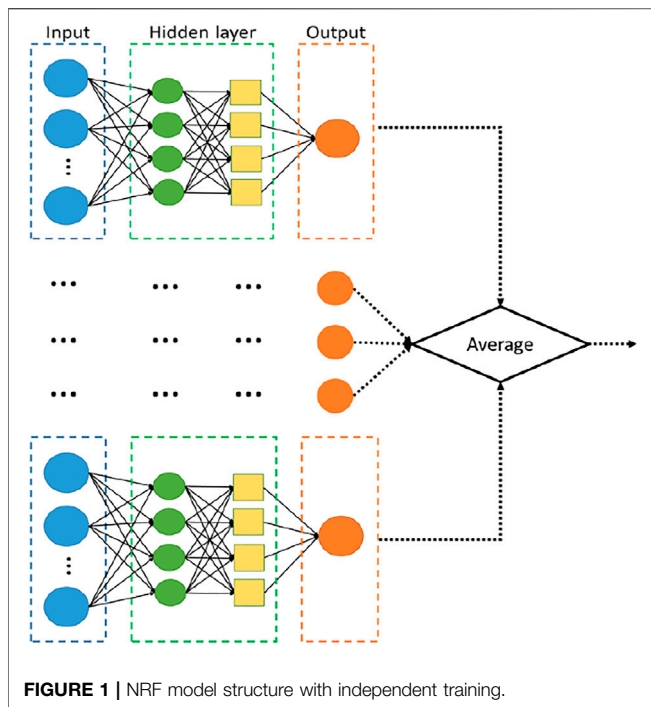
The loss function $C(\theta)$ is defined as

$$C^{(i)}(\theta) = \frac{1}{m} \sum_{r=1}^m \|f(x^{(r)}; \theta) - y^{(r)}\|, \quad (6)$$

$$C(\theta) = \frac{1}{N^{(L)}} \sum_{r=1}^{N^{(L)}} C^{(i)}(\theta). \quad (7)$$

In addition, in this article, we adopt the particle swarm optimization (PSO) to optimize the weight and bias. The PSO is performed to solve the global optimization problems by simulating the biological population (Kennedy and Eberhart, 1995; Eberhart and Shi, 2001). It has received extensive attention because of its few parameters and easy implementation.

It is noticed that the interpretability of the machine learning model is necessary to assist to make decisions. Although the



neural network can obtain nonlinear relationships among data, it is uneasy to clarify how it works, which is called the “black box.” In addition, the neural network has plenty of hyperparameters to tune, which is time-consuming.

2.1.2 Random Forest Algorithm

Random forest (RF) is an ensemble machine learning approach proposed by Breiman (2001), which has the advantages of interpretability, convenience, and fast calculating speed. RF is independently built by several decision trees. Decision trees are the basic classifier in the RF algorithm. Compared with the neural networks, the decision trees are interpretable. It starts from the root node and constructs the tree nodes one by one based on the rule for prediction or classification.

The standard RF is built upon the bootstrap datasets and splitting with the classification and regression tree (CART) (Breiman et al., 1984) methodology. This method works on the bagging principle. Bagging algorithms randomly select samples from the raw data so that the training of each basic classifier in the ensemble is independent of the others. Specifically, the flow of the construction of RF is as follows:

First, at each node of the decision tree, the predictor variables are sampled randomly. Then, the algorithm finds the minimal residual sum of squares (RSS) for regression or categorization. Furthermore, data are divided into the “in-bag” subset for training and the “out-of-bag (OOB)” subset for validation. Finally, the decision trees are combined through the majority (categorization) or the average (regression) vote to form the final prediction result (Despoina et al., 2021).

Compared with the neural network, RF has fewer parameters to tune. However, on some classification or regression problems, it is prone to overfit with the noisy data.

2.1.3 Proposed Neural Random Forests

With respect to the shortcomings of both the neural network framework and random forests, Welbl (2014) and Richmond et al. (2015) had demonstrated the importance of casting the RF algorithm into a neural network framework. To exploit the benefits of both algorithms, Biau et al. (2019) proposed two new hybrid procedures to reformulate the RF method into a neural network which is called neural random forests (NRF). The NRF method exploits prior knowledge of regression trees and provides interpretability. In addition, this method has more excellent performance than the RF method and neural networks.

NRF has two different ways to combine the individual networks: one is the independent training and the other is the joint training.

Assume a random forest is a predictor consisting of a collection of M (large) regression trees: the training sample is $D_n = ((X_1, Y_1), \dots, (X_n, Y_n))$, $n \geq 2$, where X is the feature data and Y is the corresponding label.

For independent training, the parameters of each tree-type network are fitted network by network. The prediction value is defined as

$$r(x; \theta_1, \dots, \theta_M, D_n) = \frac{1}{M} \sum_{m=1}^M r(x; \theta_m, D_n), \quad (8)$$

where $\theta_1, \dots, \theta_M$ are random variables and $r(x; \theta_1, \dots, \theta_M, D_n)$ is the predicted value at the point x for the m -th tree. The structure of NRF with independent training is shown in **Figure 1**.

For the joint training, the individual tree networks are concatenated into one network and then fitted based on the

whole model. The final estimate is obtained by minimizing the empirical error.

$$J_n(f) = \frac{1}{n} \sum_{i=1}^n |Y_i - f(X_i)|^2, \quad (9)$$

where $f(X_i)$ is the neural network implementing functions, Y_i is the actual data, and n is the number of samples.

The structure of NRF with joint training is shown in **Figure 2**.

The activation function in this model is the hyperbolic tangent activation function. This function can provide better generalization and favor smoother decision boundaries. The hyperbolic tangent activation function is defined as

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}. \quad (10)$$

2.2 Model Evaluation Criteria

In our experiments, four predictive metrics are calculated to assess the performance of the prediction results of well logs, namely, the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), and root mean square error (RMSE).

The R^2 demonstrates the prediction accuracy of the proposed method. The MAE describes an average difference between the predicted and actual measurements. RMSE denotes the standard deviation between the predictions of the model and actual data.

The three criteria are defined below:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (11)$$

$$MAE = \frac{1}{N} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (13)$$

where n is the number of samples, y_i is the i -th actual value, \hat{y}_i is the i -th predicted value, and \bar{y} is the average of n .

3 EXPERIMENTS AND DISCUSSION

In this section, the experiments are conducted for shale gas reservoirs. The proposed model is to search the complex nonlinear relationship between the well logs and physical parameters, and efficiently predict the porosity and saturation of blind wells.

All experimental studies were carried out in the Python 3.6 compiling environment using Anaconda. All artificial intelligence models were developed using TensorFlow (Abadi et al., 2016).

3.1 Dataset

To evaluate the performance of the proposed approach, we selected the logging data of six classic wells (Well-1, Well-2, Well-3, Well-4,

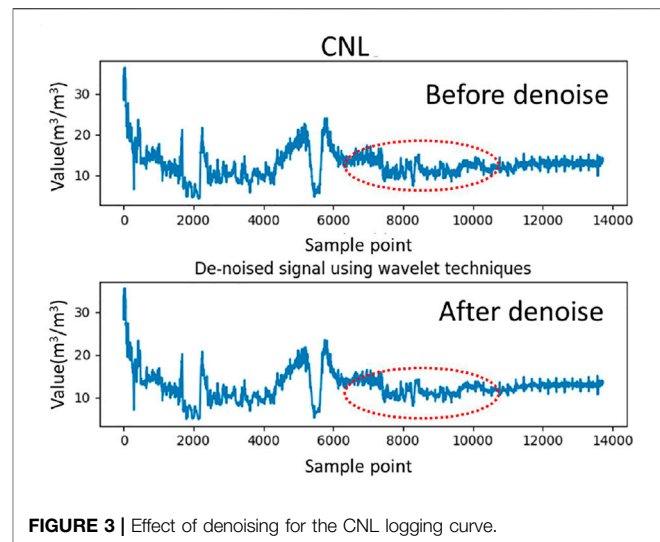


FIGURE 3 | Effect of denoising for the CNL logging curve.

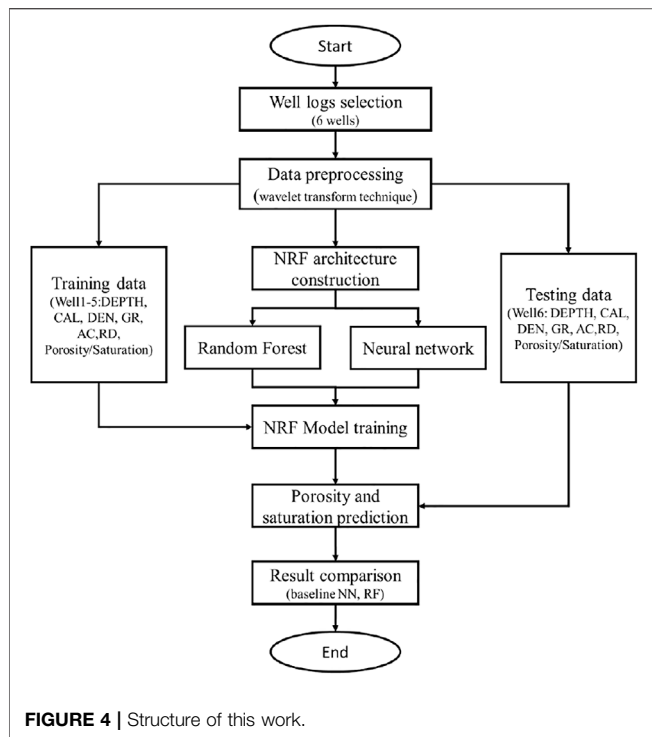
Well-5, and Well-6) from the study area, and the shale gas reservoirs at the Chongqing oilfield in China comprise the dataset for the prediction of porosity and saturation. We selected six logging curves from each well, namely depth (DEPTH), borehole diameter (CAL), neutron (CNL), gamma ray (GR), density (DEN), acoustic (AC), and deep investigate double lateral resistivity log (RD). The depth of six wells is 2000–3000 m. The sampling interval is 0.125 m. In addition, the porosity and saturation data are calculated using the petrophysical volume model. To validate the accuracy of the calculated results, we compared them with the core test results, and they matched very well. Therefore, we regard the calculated porosity and saturation results as the “actual” data. For the machine learning model to learn the best permutation sequence between the input and target vectors, the well log data are subdivided into a training set and a test set. To improve the accuracy of prediction, it is necessary to carry out data preprocessing.

The research area of six wells is first determined by manual selection. Then, the null data from six wells are discarded. Furthermore, the noise data are filtered using the wavelet transform technique. The wavelet transform technique can decompose a signal into multiple lower resolution levels by controlling the scaling and shifting factors of a single wavelet function (Foufoula-Georgiou et al., 1994; Lau and Weng, 1995; Torrence and Compo, 1998; Percival, 2000). After the wavelet transform technique, the noise data are filtered, and high quality data are obtained.

In our experiments, CAL, CNL, GR, DEN AC, and RD logging curves are denoised using the wavelet transform technique taking the CNL logging curve from Well-1 as an example. The comparisons of the logging curves before and after denoising are depicted in **Figure 3**. It can be seen that the data are smoother after wavelet denoising and the denoising effect is more pronounced in the red circle marked in the picture.

3.2 NRF Architecture and Procedure

In this study, we built a porosity and saturation prediction model separately based on the NRF algorithm and designed the model’s



framework. The overall structure of this work is illustrated in **Figure 4**.

The neural network in the NRF model has four layers which use a backpropagation algorithm. The hyperbolic tangent function is used on the hidden layers, and Adam optimizer is used during the training process to avoid overfitting. In addition, the trees number of RF included in the NRF model is 30 and the max depth is 6.

The steps used to develop and implement the proposed model are summarized in the following text:

Step 1. Well log selection: The input and output vector for the proposed model are determined and specified.

Step 2. Data preprocessing: This involves discarding invalid data and denoising data using the wavelet transform technique.

Step 3. Data division: The processed data are split into the training set and testing set.

Step 4. NRF architecture construction: This consists of collaborating the neural network and RF. The related parameters should also be set at this stage.

Step 5. NRF model training: The relationship between the input and output vectors is obtained through the NRF model.

Step 6. Physical parameters prediction: Based on the trained NRF model in Step 5, the porosity and saturation of the blind well are predicted.

TABLE 1 | Comparison of predictive results of saturation before and after denoising using wavelet transforms.

Method	MAE	MSE	RMSE
NRF1-before-denoise	19.754	2.280	4.444
NRF1-after-denoise	9.068	2.195	3.011
NRF2-before-denoise	17.522	1.902	4.185
NRF2-after-denoise	7.470	1.836	2.733

Step 7. Result comparison: To evaluate the performance of the proposed model, the baseline models, NN and RF, were compared.

3.3 Experiment Results and Discussion

In our experiments, the NRF model was trained in two ways: independent training and joint training. For simplicity, the NRF model with independent training is named NRF1, and the NRF model with joint training is named NRF2. We first validate the performance of denoising based on the wavelet transforms. The predictive performance of saturation is taken as an example, and the results are shown in **Table 1**.

It is demonstrated in **Table 1** that the NRF models can achieve better results with lower error after denoising technique. Therefore, it is necessary and significant to denoise to obtain high-quality data.

Based on the processed data, the performance of different prediction models are compared. As the proposed NRF model makes up of the neural network and random forest, the standard neural network and random forest are compared together. These four models (NRF1, NRF2, NN, and RF) are used for forecasting porosity and saturation of the study reservoir. To represent the fit of the logging curve more visually, the predicted and actual logging data with depth are depicted in **Figure 5**. The red line denotes the measured data, the blue line means the predicted data, and the dots indicate the core test values.

It was found that most of the predicted values using the NN model did not fall on the perfect linear trend line and have the worst performance, while other three methods fit well.

To compare the results further, part of the well logs is selected to be analyzed specifically (the red box in **Figure 5**). The selected part is from the middle of the logging with a depth of 3210–3330 m (**Figure 6**). This part is 120 m, which contains 960 samples.

From **Figure 6**, it is demonstrated that the forecast accuracy of porosity obtained by the NN has a large error zone, while the prediction using the RF model has been enhanced. Using the combination of the NN and RF algorithms, the prediction accuracy has been further improved evidently. Therefore, the NRF model can be a powerful tool for predicting the porosity based on well logging data.

The predicted results for saturation are also depicted in **Figure 7**. Similarly, the red line denotes the measured saturation value, and the blue line means the predicted data. It was found that the predictions of four models fluctuate greatly around the standard measured values, and the prediction of saturation is not as good as the prediction of porosity.

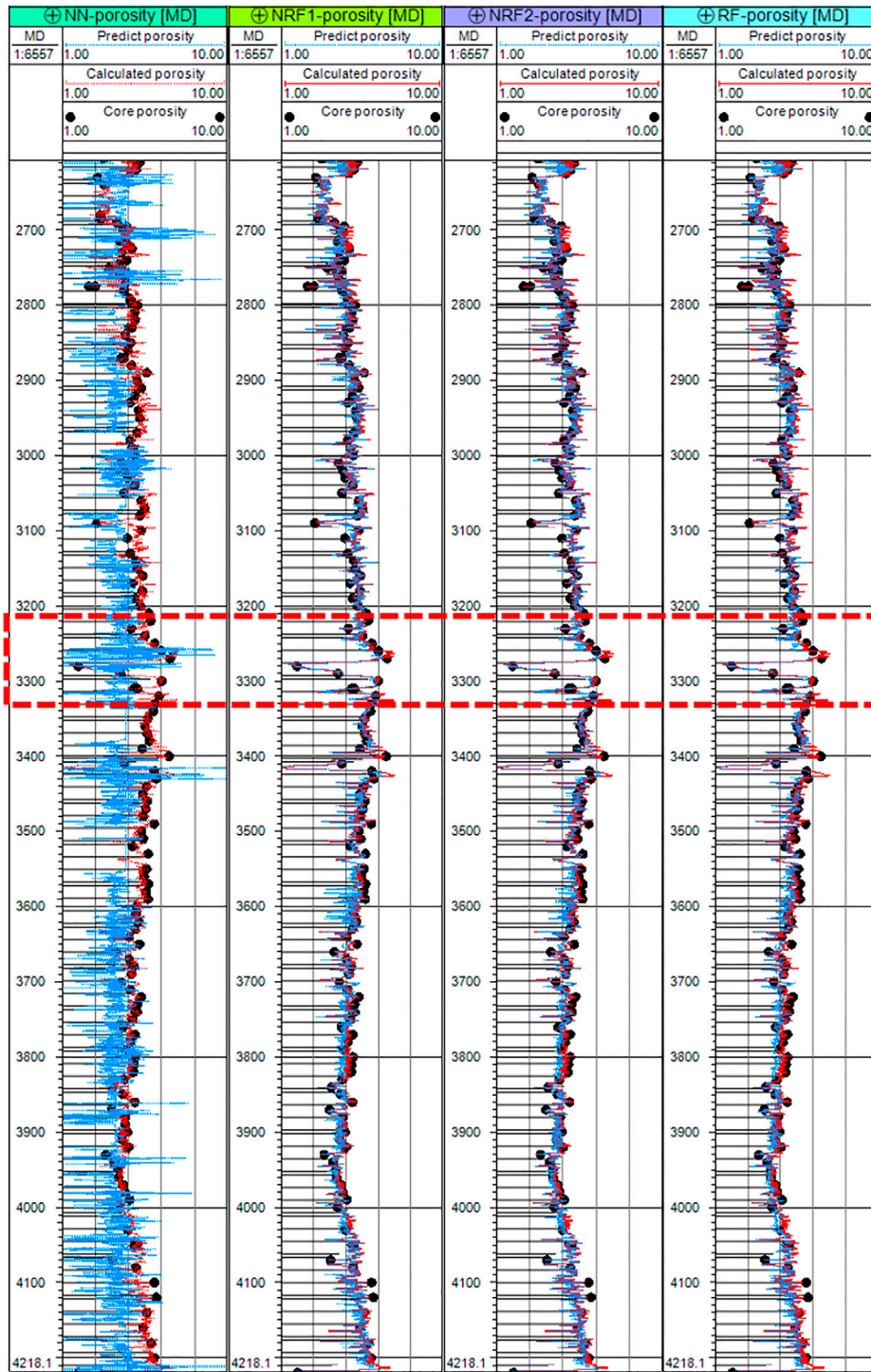


FIGURE 5 | Comparison between actual and predicted porosity data in different models (Well-6).

To compare the results further, the well logging curves with a depth of 3650–4100 m are selected (the red box in **Figure 7**). This part is 450 m, which contains 3,600 sample. The details are depicted in **Figure 8**.

It can be drawn from **Figure 8** that the forecast accuracy of saturation obtained by the NN has a large error zone. The prediction using the RF model is competitive to the NRF1 model, while the prediction accuracy has been further

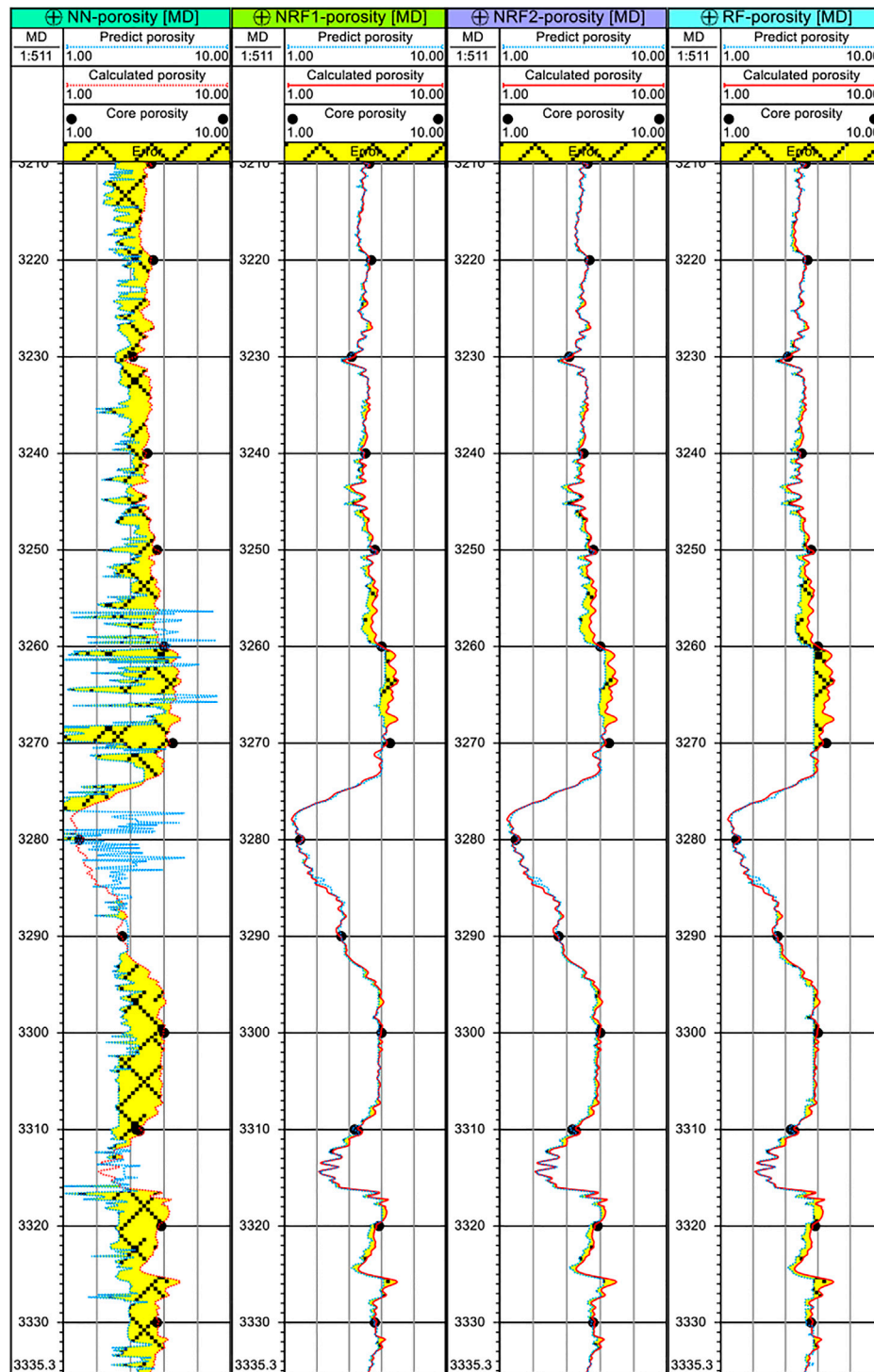


FIGURE 6 | Comparison results of different models for porosity (depth: 3210–3330 m).

improved using the NRF2 model. Therefore, the NRF2 model is chosen to predict saturation based on well logging data.

To provide a qualitative analysis of the model prediction, the R^2 , MAE, MSE, and RMSE are calculated for porosity and saturation prediction models, respectively.

The qualitative analysis results of the predicted porosity and saturation from four models are shown in **Table 2**. For the prediction of porosity, it is demonstrated that among the classic machine learning algorithms (NN and RF), RF works much better due to the majority voting mechanism, with the R^2 of

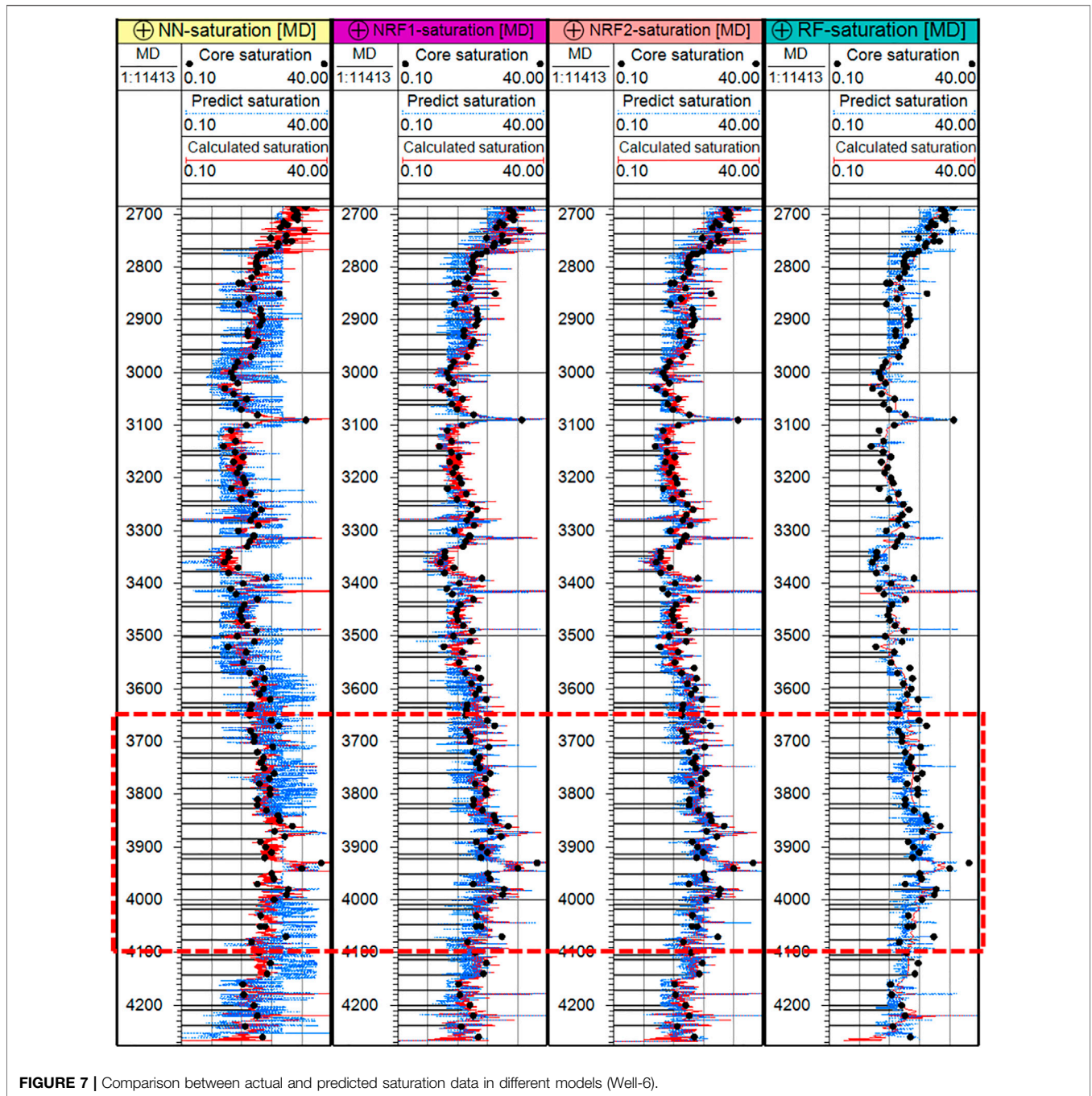


FIGURE 7 | Comparison between actual and predicted saturation data in different models (Well-6).

0.913, MAE of 0.066, MSE of 0.201, and RMSE of 0.226. The NRF1 model whose R^2 is 0.941 performs very close to the NRF2 model. Comparingly, the NRF2 model is superior, and it yields the best R^2 , which is 1.1% higher than the NRF1 model, 3.9% higher than the RF, and greatly higher than the NN. It is worth noting that our proposed NRF model significantly outperforms the baseline models, resulting in at least 3% improvement in the prediction accuracy.

It can be noticed that all the four models have a poorer capability of predicting saturation than porosity. This is

attributed to the strong correlation of the logging curves applied with porosity and the weak mapping to saturation. Nevertheless, the proposed NRF model, especially the NRF2 model, provides high accuracy prediction for saturation, with R^2 above 80%.

Moreover, we generated the histograms of porosity and saturation in the target well and predicted the values using four algorithms (Figures 9, 10). The horizontal coordinates denote the range of data and the longitudinal coordinates means the amount of data in different range, presented as a

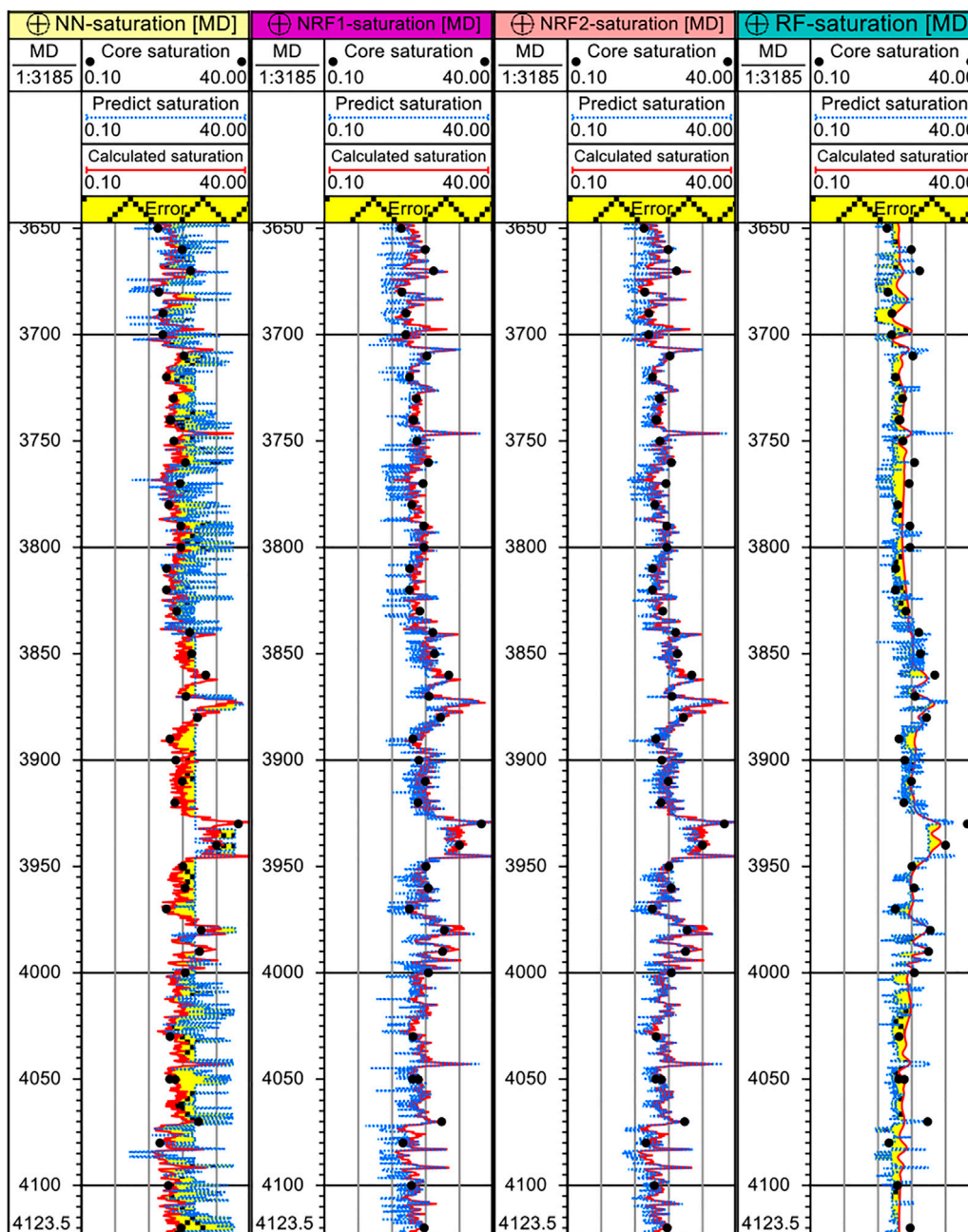


FIGURE 8 | Comparison results of different models for saturation (depth: 3650–4100 m).

percent format. The histograms can obviously reflect the distribution, the center, and dispersion of data.

It can be drawn from **Figure 9** that the porosity data of Well-6 are consistent with normal distribution and the center of data is 5, which is named the standard center point. As for the predicted results from the neural network, the data are roughly in accordance with normal distribution, but the center of data is about 3, which is far away from the standard center point. The predicted results of the RF model have a large dispersion and the center point is around 4.5, which is still far away from the standard center point. In comparison, the NRF with

the independent and joint training is similar with the distribution of actual porosity data, and the center of the data is near the standard center point.

As can be seen in **Figure 10**, the saturation value for Well-6 conforms to a normal distribution, while the distribution of predicted values for the NN, RF, and NRF1 models does not fit. In comparison, the results from the NRF2 model have a normal distribution, and the center point is the same with the standard center point. Consequently, the NRF2 model can be employed to predict the saturation of shale gas reservoirs.

TABLE 2 | Qualitative analysis of predicted porosity and saturation results from four models.

Method for porosity	R ²	MAE	MSE	RMSE
NN	0.256	2.096	0.985	1.462
RF	0.913	0.066	0.201	0.226
NRF1	0.941	0.047	0.153	0.186
NRF2	0.952	0.039	0.114	0.162
Method for saturation				
NN	0.363	24.821	3.069	4.865
RF	0.774	9.040	2.193	3.006
NRF1	0.823	8.845	1.902	2.942
NRF2	0.844	7.169	1.836	2.135

A qualitative analysis of the results from four models is also provided by calculating the minimum (Min), maximum (Max), mean of all data in each model (Mean), standard deviation (Std Dev), and Variance (Var).

The Var is expressed as

$$Var = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2. \tag{14}$$

The Std Dev is calculated by

$$StdDev = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2}, \tag{15}$$

where n is the number of samples, y_i denotes the data to be assessed, and \bar{y}_i is the mean of all samples.

The consequences are shown in **Table 3**.

For the prediction of porosity and saturation, it can be concluded from **Table 3** that the predicted results from the

NRF2 model have the closest value of mean, standard deviation, and variance with the actual data, which means this model has a similar distribution with the actual measured porosity and saturation data. In addition, it also demonstrates greater stability and better performance. In this case, the NRF1 method is slightly inferior to the NRF2 method. It is noted that for the prediction of saturation, the NN model obtains close qualitative values with the actual data. However, it does not imply that this algorithm has better performance according to **Figures 7, 8**.

Therefore, from the perspective of data distribution, the superiority of the algorithm proposed in this article is further illustrated.

It is crucial to notice that the models created in the present research are suitable only for the shale gas reservoir. When the same methodologies used are presented in this research, it is necessary to retrain the model with the data in the respective area.

4 CONCLUSION

In this study, a novel hybrid model combining the neural network and random forest is proposed for predicting the porosity and saturation based on well logging of the shale gas reservoir. The ability of the proposed model to forecast porosity and saturation is discussed. Meanwhile, the proposed model is also compared with the baseline methods, namely, NN and RF. The main conclusions are as follows:

1. The proposed NRF model provided a promising method for predicting the porosity and saturation, as evidenced by the

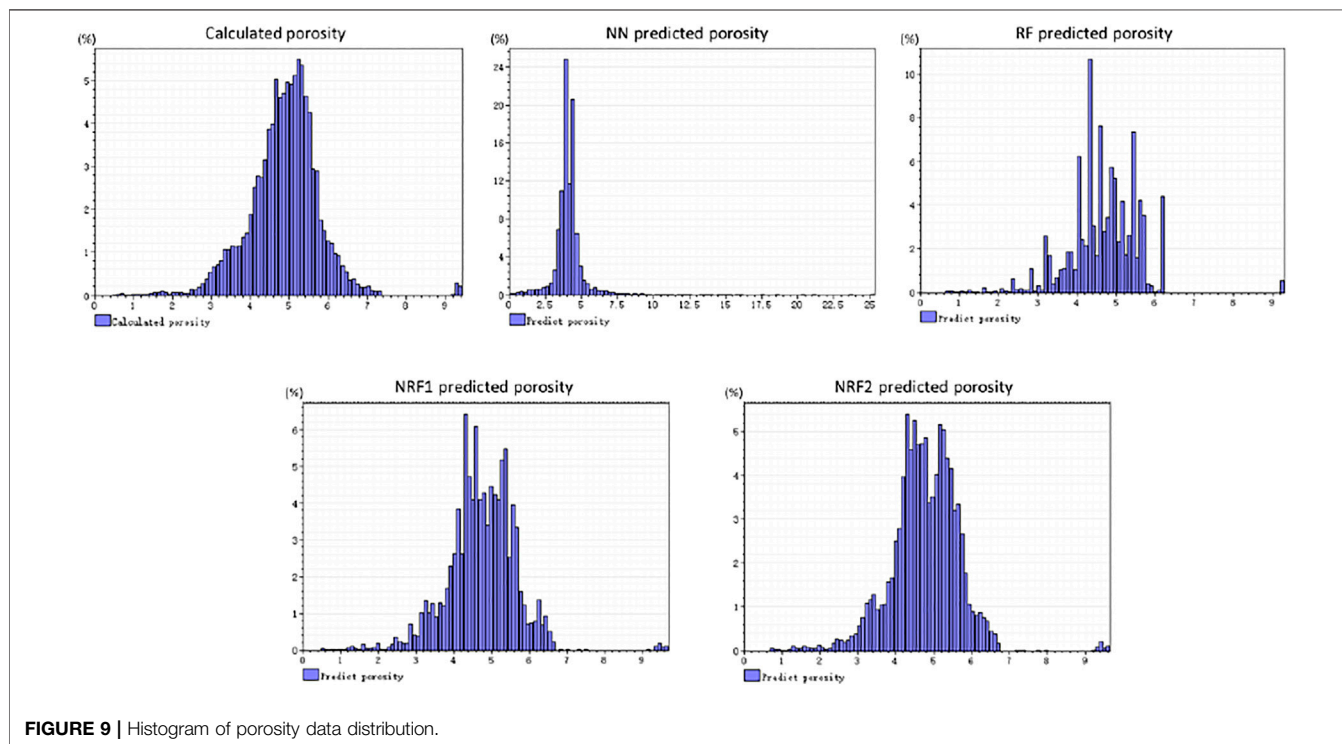


FIGURE 9 | Histogram of porosity data distribution.

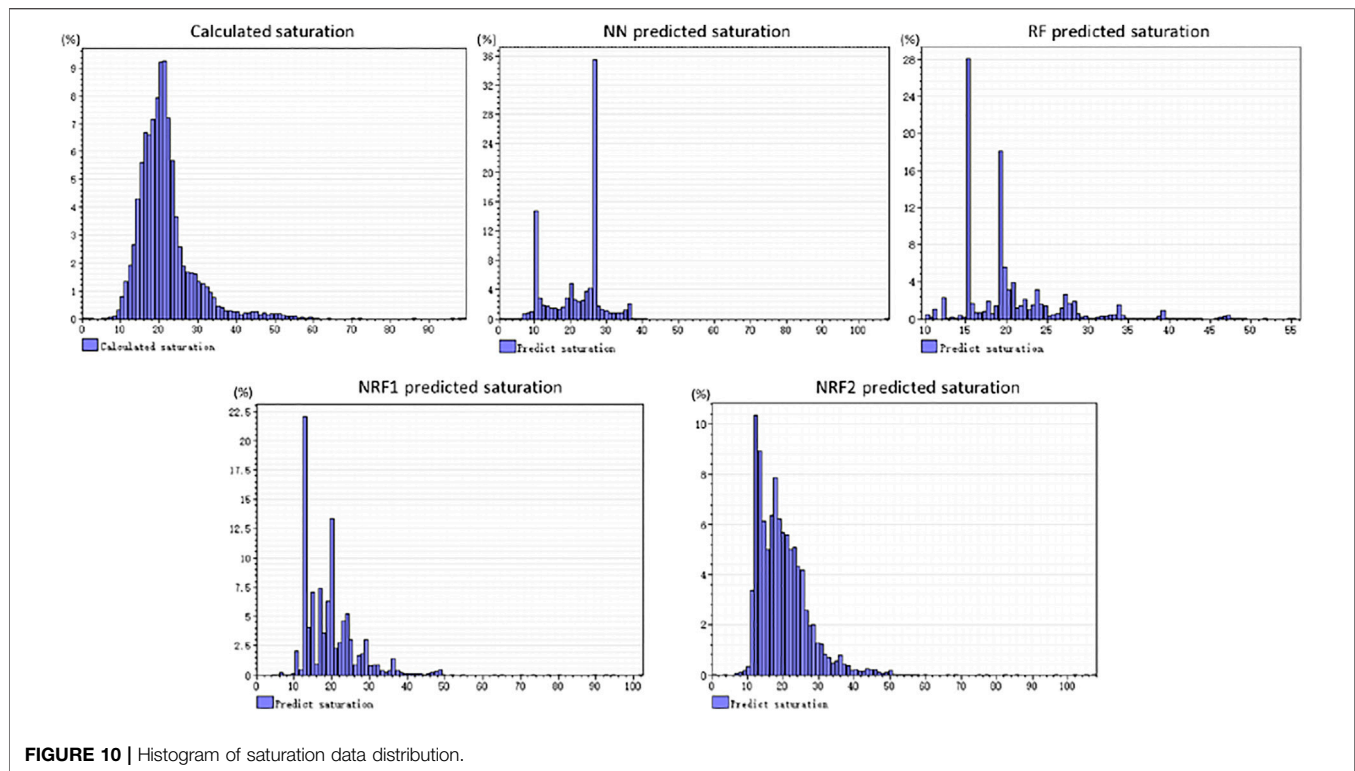


FIGURE 10 | Histogram of saturation data distribution.

TABLE 3 | Qualitative analysis of data distribution for porosity and saturation from the actual data and predicted results.

Method for porosity	Min	Max	Mean	Var	Std. Dev.
Actual	0.53	9.43	4.88	0.82	0.91
NN	-3.45	20.13	4.12	1.09	1.04
RF	0.78	9.34	4.49	0.82	0.90
NRF1	0.51	9.68	4.78	0.83	0.91
NRF2	0.66	9.58	4.80	0.81	0.90
Method for saturation					
Actual	3.80	78.35	21.43	40.14	6.34
NN	3.12	63.45	22.34	44.96	6.70
RF	10.25	55.21	20.17	35.58	5.96
NRF1	10.02	56.41	21.64	34.86	5.90
NRF2	9.31	58.02	21.27	32.04	5.66

satisfactory performance on porosity data with R^2 above 0.94 and saturation data with R^2 above 0.82. It can serve as an alternative tool to give a reliable prediction.

- The proposed NRF model outperformed the classical neural networks and random forest models. In comparison, for the prediction of porosity, the RF algorithm worked much better than the NN model with R^2 above 0.91. The NRF2 model obtained the unsurpassed performance with R^2 above 0.95, which is 1.1% higher than that of the NRF1 model, 3.9% higher than RF, and greatly higher than the NN. For the prediction of saturation, the NRF2 model with R^2 above 0.84 is also superior than other algorithms, which is 2.1% higher than the NRF1 model and 7.0% higher than the RF model. It has been proven

that the NRF2 model can more accurately capture the complex relationship between the logging data and physical parameters.

- In terms of the histogram of data distribution, the NRF2 method demonstrated greater stability, while the NRF1 method was slightly inferior in the study case. Therefore, the superiority of the algorithm proposed in this article is further illustrated.

In the future, we will investigate the logging data and other physical properties.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MW contributed to the conception and design of the study, and wrote the first draft of the manuscript. DF supervised this investigation. DL wrote sections of the manuscript. JW contributed to manuscript revision.

FUNDING

This work was supported by the Sinopec Science and Technology Tackle Project (P19017-2).

REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). *TensorFlow: A System for Large-Scale Machine Learning*.
- Akande, K., Owolabi, T., and Olatunji, S. (2015). Investigating the Effect of Correlation-Based Feature Selection on the Performance of Neural Network in Reservoir Characterization[J]. *J. Nat. Gas Sci. Eng.* 27, S1875510015301074. doi:10.1016/j.jngse.2015.08.042
- Basheer, I. A., and Hajmeer, M. (2000). Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *J. Microbiol. Methods* 43, 3–31. doi:10.1016/s0167-7012(00)00201-3
- Biau, G., Scornet, E., and Welbl, J. (2019). Neural Random Forests. *Sankhya A* 81, 347–386. doi:10.1007/s13171-018-0133-y
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth and Brooks/Cole Advanced Books and Software.
- Breiman, L. (2001). Random forest. *Mach. Learn.* 45, 5–32. doi:10.1023/a:1010933404324
- Despoina, M., Pauline, B., and Yining, C. (2021). A Random forest-based Approach for Predicting Spreads in the Primary Catastrophe Bond Market. *Insurance: Maths. Econ.* 101, 140–162. doi:10.1016/j.insmatheco.2021.07.003
- Eberhart, R., and Shi, Y. (2001). “Particle Swarm Optimization: Developments, Applications and Resources,” in Proceedings of the 2001 Congress on Evolutionary Computation, Seoul, Korea, May 2001 (IEEE), 81–86.
- Foufoula-Georgiou, E., Kumar, P., and Chui, C. K. (Editors) (1994). *Wavelets in Geophysics*. Academic Press, Vol. 4.
- Gardner, M. W., and Dorling, S. R. (1998). Artificial Neural Networks (The Multilayer Perceptron)-A Review of Applications in the Atmospheric Sciences. *Atmos. Environ.* 32, 2627–2636. doi:10.1016/s1352-2310(97)00447-0
- Hadi, F., and Sadegh, K. (2016). Prediction of Porosity and Water Saturation Using Pre-stack Seismic Attributes: a Comparison of Bayesian Inversion and Computational Intelligence Methods[J]. *Comput. Geosciences* 20 (5), 1075–1094. doi:10.1007/s10596-016-9577-0
- Kennedy, J., and Eberhart, R. (1995). “Particle Swarm Optimization,” in Proc. Int. Conf. Neural Networks, Perth, WA, Australia, 27 Nov.-1 Dec. 1995 (IEEE), 1942–1948. doi:10.1109/ICNN.1995.488968
- Khandelwal, M., and Singh, T. N. (2010). Artificial Neural Networks as a Valuable Tool for Well Log Interpretation. *Pet. Sci. Tech.* 28, 1381–1393. doi:10.1080/10916460903030482
- Komarialaei, H., and Salahshoor, K. (2012). The Design of New Soft Sensors Based on PCA and a Neural Network for Parameters Estimation of a Petroleum Reservoir[J]. *Liquid Fuels Tech.* 30 (22), 12. doi:10.1080/10916466.2010.512899
- Lau, K.-M., and Weng, H. (1995). Climate Signal Detection Using Wavelet Transform: How to Make a Time Series Sing. *Bull. Amer. Meteorol. Soc.* 76, 2391–2402. doi:10.1175/1520-0477(1995)076<2391:csduwt>2.0.co;2
- Li, T., Song, H., Wang, J., Wang, Y., and Killough, J. (2016). An Analytical Method for Modeling and Analysis Gas-Water Relative Permeability in Nanoscale Pores with Interfacial Effects. *Int. J. Coal Geology.* 159, 71–81. doi:10.1016/j.coal.2016.03.018
- Percival, W. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge, UK: Cambridge Univ. Press.
- Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., et al. (2016). Neural Networks: an Overview of Early Research, Current Frameworks and New Challenges. *Neurocomputing* 214, 242–268. doi:10.1016/j.neucom.2016.06.014
- Raymer, L. L., Hunt, E. R., and Gardner, J. S. (1980). “An Improved Sonic Transit Time-to-Porosity Transform,” in SPWLA 21st Annual Logging Symposium, Lafayette, Louisiana, July, 1980 (OnePetro).
- Richmond, D. L., Kainmueller, D., Yang, M. Y., Myers, E., and Rother, C. (2015). Relating Cascaded Random Forests to Deep Convolutional Neural Networks for Semantic Segmentation. arXiv:1507.07583.
- Rolon, L., Mohaghegh, S. D., Ameri, S., Gaskari, R., and McDaniel, B. (2009). Using Artificial Neural Networks to Generate Synthetic Well Logs. *J. Nat. Gas Sci. Eng.* 1, 118–133. doi:10.1016/j.jngse.2009.08.003
- Saputro, O. D., Maulana, Z. L., and Latief, F. D. E. (2016). Porosity Log Prediction Using Artificial Neural Network. *J. Phys. Conf. Ser.* 739, 012092. doi:10.1088/1742-6596/739/1/012092
- Schmidhuber, J. (2015). Deep Learning in Neural Networks: an Overview. *Neural Networks* 61, 85–117. doi:10.1016/j.neunet.2014.09.003
- Song, J., Gao, Q., and Li, Z. (2016). Application of Random Forests for Regression to Seismic Reservoir Prediction[J]. *Oil Geophys. Prospect.* 51 (6), 1202–1211. doi:10.13810/j.cnki.issn.1000-7210.2016.06.021
- Song, H., Xu, J., Fang, J., Cao, Z., Yang, L., and Li, T. (2020). Potential for Mine Water Disposal in Coal Seam Goaf: Investigation of Storage Coefficients in the Shendong Mining Area. *J. Clean. Prod.* 244, 118646. doi:10.1016/j.jclepro.2019.118646
- Song, H., Liu, C., Lao, J., Wang, J., Du, S., and Yu, M. (2021). Intelligent Microfluidics Research on Relative Permeability Measurement and Prediction of Two-phase Flow in Micropores[J]. *Geofluids* 2021, 1–12. doi:10.1155/2021/1194186
- Song, H., Zhang, J., Ni, D., Sun, Y., Zheng, Y., Kou, J., et al. (2021). Investigation on In-Situ Water Ice Recovery Considering Energy Efficiency at the Lunar South Pole. *Appl. Energ.* 298, 117136. doi:10.1016/j.apenergy.2021.117136
- Torrence, C., and Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bull. Amer. Meteorol. Soc.* 79, 61–78. doi:10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2
- Wang, J., Song, H., Rasouli, V., and Killough, J. (2019). An Integrated Approach for Gas-Water Relative Permeability Determination in Nanoscale Porous media. *J. Pet. Sci. Eng.* 173, 237–245. doi:10.1016/j.petrol.2018.10.017
- Wang, J., Song, H., and Wang, Y. (2020). Investigation on the Micro-flow Mechanism of Enhanced Oil Recovery by Low-Salinity Water Flooding in Carbonate Reservoir. *Fuel* 266, 117156. doi:10.1016/j.fuel.2020.117156
- Welbl, J. (2014). “Casting Random Forests as Artificial Neural Networks (And Profiting from it),” in *Pattern Recognition* (Springer), 765–771. doi:10.1007/978-3-319-11752-2_66
- Wyllie, M. R. J., Gregory, A. R., and Gardner, L. W. (1956). Elastic Wave Velocities in Heterogeneous and Porous Media. *Geophysics* 21 (1), 41–70. doi:10.1190/1.1438217

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Feng, Li and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.