

Rock Thickness Identification by XGBoost with Logging Data

Weikai LIU

Northeast Petroleum University

Yanbin ZHAO (✉ 844756885@qq.com)

Northeast Petroleum University

Mei YANG

No.6 Oil Provide Factory, Daqing Oilfield Limited Company

Yueqing XU

Institute of Drilling Engineering Technology

Guangming LI

CNPC Logging Daqing Branch

Research Article

Keywords: formation identification, reservoir identification, logging curves, XGBoost algorithm, machine learning, class imbalance learning

Posted Date: December 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1133777/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Rock thickness identification by XGBoost with logging data

Weikai LIU¹, Yanbin ZHAO^{1*}, Mei YANG², Yueqing XU³, Guangming LI⁴

(1. *Northeast Petroleum University, Daqing, 163318, China;*

2. *No.6 Oil Provide Factory, Daqing Oilfield Limited Company, Daqing 163000, China;*

3. *Institute of Drilling Engineering Technology, Daqing 163000, China;*

4. *CNPC Logging Daqing Branch, Daqing 163000, China)*

Abstract: Based on research on the response mechanism of rock formations and reservoirs to logging curves, 12 logging curves selected by combining the depth characteristics of formations are proposed to identify rock formations and reservoirs using four algorithms: logistic regression (LR), support vector machine (SVM), random forest (RF) and XGBoost. Out of 60 wells in the study block, 57 wells were selected for training and learning, and the remaining 3 wells were used as prediction samples for testing the algorithm. The recognition of rock formations and reservoirs is performed by each of these four machine learning algorithms, and predictive knowledge is obtained separately. It was found that the accuracy of the 4 algorithms for rock formation and reservoir layer identification reached over 90%, but the XGBoost algorithm was found to be the best in terms of the 4 scoring criteria of F1-score, precision, recall and accuracy. The accuracy of rock formation identification could reach over 95%, and the correlation analysis between the logging curve and rock formation could be performed on this basis. The results show that the RMN, RLLD and RLLS have the most obvious responses to the sandstone layer, off-surface reservoir and effective thickness layer, and the CAL has the least effect on the formation and reservoir identification, which can provide an effective reference for the selection and dimensionality reduction of the subsequent logging curves.

Keywords: formation identification; reservoir identification; logging curves; XGBoost algorithm; machine learning; class imbalance learning

1 Introduction

At present, the methods of logging lithology identification are mainly conventional logging identification methods, conventional logging rendezvous map identification methods, imaging logging identification methods, principal component analysis methods, neural network methods, and transverse wave information rendezvous identification methods. In recent years, the use of machine learning algorithms to predict rock formation and reservoir layer methods has also gradually diversified, among which He Ping¹ proposed using the fuzzy ISODATA algorithm to optimize the prediction of rock formation. Song Jianguo et al.² used the nonlinear prediction feature of the random forest algorithm to predict reservoir layers. Wang Zhong et al.³ used a support vector machine to eliminate multiple correlations between variables to study rock layer identification. Liu Ke et al.⁴ used a mixed density network to establish a theoretical model to achieve effective prediction of sand thickness. Shan Jingfu⁵ used a BP neural network to identify gas reservoir layers with complex rock layers. Lei Huang et al.⁶ used a deep learning approach for rock formation identification and correlation analysis. In addition to the use of machine learning algorithms to establish a mathematical model, the reasonable choice of logging curves and features will also have a large impact on the accuracy of rock layer identification⁷. For example, Mou Dan et al.⁸ used five logging curves of deep lateral, acoustic time difference, compensated neutron, density and natural gamma to identify basal rocks in the Liaohe Basin. Wang Zehua et al.⁹ used five curves of resistivity, acoustic time difference, natural gamma, density and acoustic impedance to identify basal rocks in the Liaohe Basin. Logging curves were used to identify the 4 major rock phases in the Jungar Basin; Zhu Yixiang and Shi Guangren¹⁰ used 15 logging curves, such as acoustic time difference, bulk density, photoelectric absorption cross-sectional index, natural gamma, compensated neutron, deep lateral, and shallow lateral, as multidimensional geological characterization parameters to identify the coal seams; Guo Xu¹¹ used five logging curves, such as lateral resistivity, density, natural gamma, sound speed, and natural potential, to identify the coal seams. The positioning

was interpreted qualitatively; Yu et al.¹² used 4 logging curves of acoustic time difference, density, natural potential and natural gamma to select appropriate nuclear functions to predict the three major types of rock formations of siltstone, mudstone and conglomerate in an oil field.

However, all of the above articles have some problems, such as the high model complexity cannot obtain the trained and mature model quickly, the formation identification results are easily affected by the noise of the raw data, and the clarity of the formation data obtained from the logging curve is not high enough to support a higher formation identification efficiency¹³. Based on the above problems, this paper classifies the formation into sandstone layer, off-surface reservoir and effective thickness layer from the perspective of reservoir evaluation, and tries to identify the rock layer in this block by XGBoost, Random Forest, Support Vector Machine and Logistic Regression. Through the comparison of 4 algorithms, the XGBoost algorithm can effectively solve the problem of accurate identification of rock layers and reservoirs, which proves the universality of the method.

2 Description of the study block

Based on the field logging data and actual conditions, the working wells in this study block are basically shallow wells with depths ranging from 1,260 to 1,314 m. The logging data obtained from the field have well depths ranging from 1,035 to 1,300 m. The well section in this study block passes through the Sa Zero, Sa One, Sa Two and Sa Three formations, which contain three to four large interbedded distribution formations, all of which are at depths below 1200 meters.

2.1 Geological characteristics of the study block

According to the logging data of 60 wells obtained from an oilfield block in Daqing and related articles, the development area has poor oil content, physical properties and reservoir lithology, the stratum has a more severe sensitivity, and the mobilization of crude oil reserves is low. From the perspective of the geological

structure, the Songliao Basin is divided into three developmental stages: early, middle and late stages. The study area experienced tectonic movement in the Songliao Basin in the early stage, which was mainly fault subsidence. The middle stage experienced subsidence of the land mass, and the late stage lifted the land mass again¹⁴. The average content of clay minerals is 27.3%; kaolinite, montmorillonite and illite constitute the majority of clay minerals, and the cementation type of rock is mainly pore cementation, with contact cementation accounting for a smaller proportion¹⁵. The reservoir physical characteristics of the study block can be obtained from an analysis of the field logging curves (Table 1). From the porosity distribution characteristics in Table 1, the block can be classified into medium- and low-porosity reservoirs.

Table 1 Reservoir physical properties in the study area

Reservoir Physical Characteristics	Porosity %	Bound Water Saturation %	Water Saturation %	Mud Content %	Permeability $10^{-3}\mu m^2$
Average	23.73	42.11	61.14	14.96	85.37
Maximum	31.6	75.69	97.8	35.0	3932.8
Minimum	15.4	8.6	1.7	3.0	4

There is a certain relationship between the reservoir physical properties of the study block. For example, there is an obvious correlation between porosity and permeability (Fig. 1). It can be seen from the figure that with increasing permeability, the porosity of a rock layer also increases. At the same time, it can be seen that the porosity of the vast majority of stratigraphy measured is greater than 18%, which verifies that the block belongs to medium- and low-porosity reservoirs.

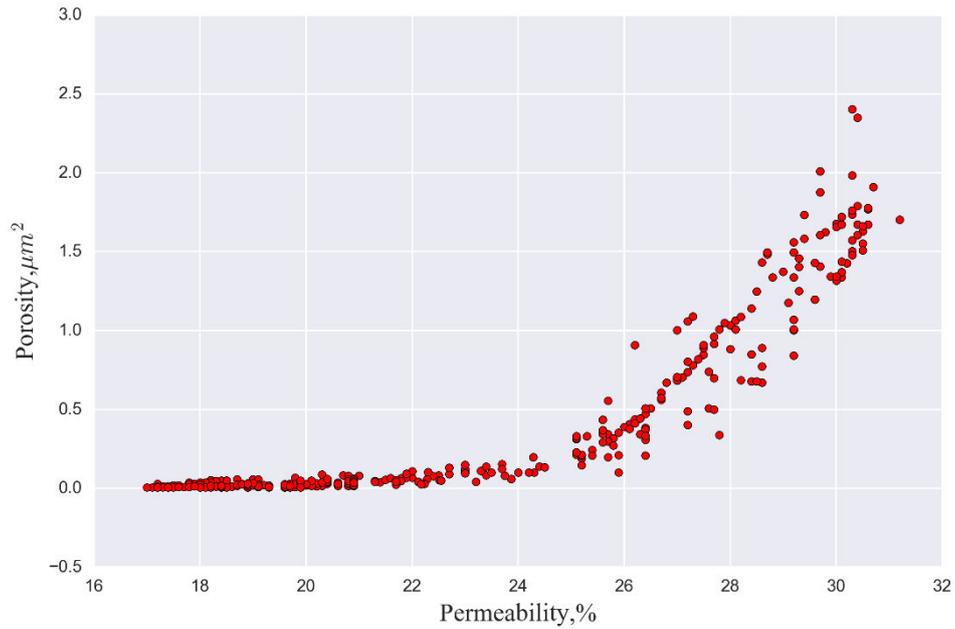


Fig. 1 Relationship between porosity and permeability in the study area

2.2 Selection of logging curves for the study block

In identifying formations and reservoirs, it is important to select the correct and reasonable parameters. It is not the case that a larger number of input features is better; a certain degree of negative feedback occurs when too many logging curves are selected, leading to a decrease in identification accuracy. To select more appropriate logging curves, this paper investigates the response characteristics of sandstone formations and reservoirs.

Mudstone and sandstone, which are electrically distinct and easily identified, dominated the sandstone layers of this formation. For the mudstone, the corresponding characteristic curves are RLLD and RLLS below $5 \Omega \cdot \text{m}$, GR and DEN high, and SP at the baseline position. For sandstone, the corresponding characteristic curves are SP negative anomaly amplitudes greater than 2 mV and RLLDs greater than $5.5 \Omega \cdot \text{m}$. The sandstone layer also contains some nonmudstone units, and the large set of sand bodies contains a small amount of mud interlayer constituting nonmudstone units. The corresponding characteristic curves are high resistance, high

return rate, low GR and DEN, and large SP negative anomaly amplitude.

Off-surface reservoirs are reservoirs that still have oil-bearing characteristics under the physical criteria of effective thickness reservoirs, and thus far, there are no unified physical criteria for off-surface reservoirs. The off-surface reservoir is the transition zone and variation zone of the effective thickness reservoir, which interacts with the effective thickness layer in the longitudinal direction and bridges with the effective thickness layer in the lateral direction. The physical, electrical, oil-bearing and lithological properties of off-surface reservoirs are consistent with each other. Performance is the reservoir with good sorting and coarse lithology has greater porosity and permeability and higher oil-bearing grade, which is expressed in the logging curve as a higher apparent resistivity value, smaller DEN, larger BHC and larger SP anomaly.

Based on the above analysis and combined with the petrophysical logging response mechanism¹⁶, this paper selected the logging depth and 12 logging curves with sensitive response to the rock layer, namely, micronormal curve RMN, microinverse RMG, high-resolution deep lateral resistivity devices curve RLLD, high-resolution shallow sensors curve RLLS, the resistivity of the flushed zone RXO, high-resolution acoustic curve HAC, 2.5 m resistivity R25, caliper curve CAL, gamma-ray curve GR, density curve DEN, borehole compensated curve BHC, self-potential curve SP, as a multidimensional geological parameters characterization parameters for intelligent identification of sandstone layers, off-surface reservoir layers and effective thickness layers (Table 2).

To study the implied relationship between the 12 logging curves in Table 2, the relationship curves can be obtained by correlating the potential, gradient, and resistivity data (Figure 2) to understand that there is a certain trend relationship between the logging curves in this test area. The implied relationship between the logging curve data can be accurately identified, and the corresponding algorithmic parameter equations can be established by machine learning.

Table 2 statistical summary of 12 logging curves

Curve Names	RMN	RMG	RLLD	RLLS	RXO	HAC	R25	CAL	GR	DEN	BHC	SP
Mean	4.33	3.86	6.02	5.15	9.35	339.38	5.61	22.17	95.85	2.30	339.38	-2.71
Std	3.10	2.49	4.46	3.03	13.51	53.52	4.39	1.90	22.42	0.15	52.61	27.17
25%	2.19	2.19	3.44	3.40	4.47	305.14	2.90	21.78	89.11	2.26	305.10	-2.89
50%	3.72	3.50	4.85	4.29	6.22	326.96	4.60	22.20	95.98	2.31	327.17	3.86
75%	5.55	4.92	6.75	5.71	9.05	384.60	6.80	22.65	102.19	2.36	383.55	7.22
Maximum	90.46	81.30	144.76	50.03	893.39	550.00	84.04	50.00	561.44	2.63	474.37	144.17

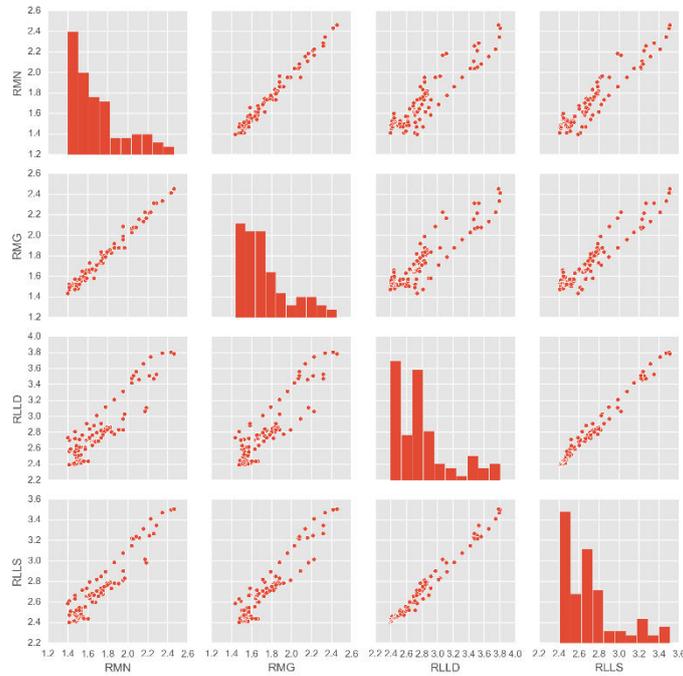


Fig. 2 Scatter plot of correlation among RMN, RMG, RLLD and RLLS

3 Brief description of the 4 algorithms

This paper attempts to use the following 4 algorithms: logistic regression (LR), support vector machines (SVM), random forests (RF) and XGBoost. These 4 algorithms (LR, SVM, RF, XGBoost) are used in this paper. XGBoost is selected when the parameters overlap and the input values are the same.

Logistic regression is a simple classification algorithm commonly used in machine learning to find the optimal partition boundaries by iterating over the data (Hobza T 2020), completing the classification of the data with the following assumption functions.

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \quad (1)$$

where $h_{\theta}(x)$ is the hypothetical function and θ is the weight corresponding to the attribute.

The support vector machine is a linear classifier defined to have the largest spacing on the feature space¹⁷. The SVM also includes different kernel functions, which makes the SVM both a linear and a nonlinear classifier, solving for the largest feature space spacing by the following dual optimization equation. The optimization function for solving the maximum feature space interval is as follows:

$$\max_{\alpha} \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n [\alpha_i \alpha_j y_i y_j \exp(-\gamma \|x_i - x_j\|^2)] \right\} \quad (2)$$

where α is the Lagrange multiplier vector; x is the current data; y is the linear function of the data; and γ is the adjustment function.

Random forest is an ensemble learning algorithm based on decision trees. Random forest uses CART decision trees as a weak learner, using a method that randomly selects a small number of features, and the number of selected features defaults to the square root of the total number of features¹⁸. Assuming that the given dataset is $D = \{X_i, Y_i\}, X_i \in R^k, Y_i \in \{1, 2, \dots, C\}$, the random forest uses M decision trees $D = \{X_i, Y_i\}, X_i \in R^k, Y_i \in \{1, 2, \dots, C\}$ as base classifiers on this dataset, the combinatorial classifier obtained after ensemble learning.

XGBoost is an optimally distributed decision gradient boosting model with low computational complexity, flexibility, high accuracy and portability, which is 10 times faster than existing popular solutions on a single machine¹⁹. A regularization term is added to reduce the complexity of the model and prevent overfitting. The overall function of the model is:

$$\hat{y}_i = \phi(X_i) = \sum_{k=1}^K f_k(X_i) \quad (3)$$

where ensemble learning y_i is the predicted value of the output, k denotes the number of trees in the tree model, and f_k denotes the model of the k th tree, which is trained to combine each tree in an additive manner.

4 Application of the 4 algorithms

4.1 Data preprocessing

Data from a total of 60 wells were obtained from the study block, and as much data as possible should be used in the model learning process. Therefore, 57 of them were randomly selected for training and learning, and three of them (N1-10, N1-20 and N1-31) were used as a validation set to verify the actual effect of each algorithm. By aggregating all the training data and observing the distribution of the training data, it can be seen that the sandstone layer, off-surface reservoir and effective thickness layers account for a tiny percentage of the total formation data; that is, there is a certain class of samples in the dataset that is much more or much less than the other classes. Therefore, the problem is transformed into an imbalance problem at this point (Fig. 3). Under these conditions, according to the characteristics of each algorithm, traditional machine learning methods will fail to face the problem, leading to a reduction in the accuracy of rock formation identification. Therefore, the SMOTETomek algorithm is used for sampling to produce new data sufficient to ensure the rationality of the decision space in a few classes, balancing the data classes on the premise of ensuring the original distribution (Fig. 4).

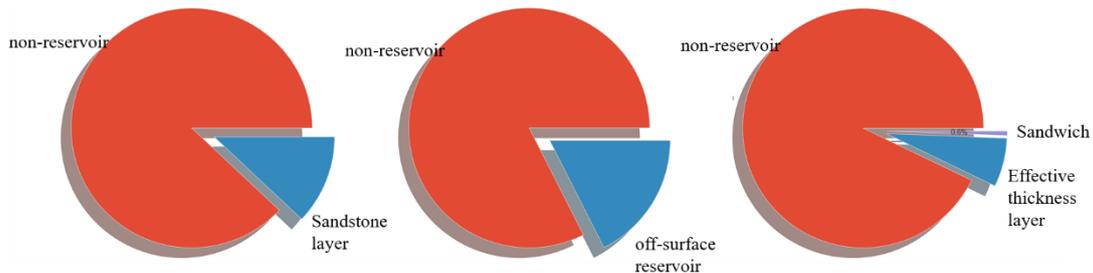


Fig. 3 Distribution ratio of sandstone and reservoir in original data

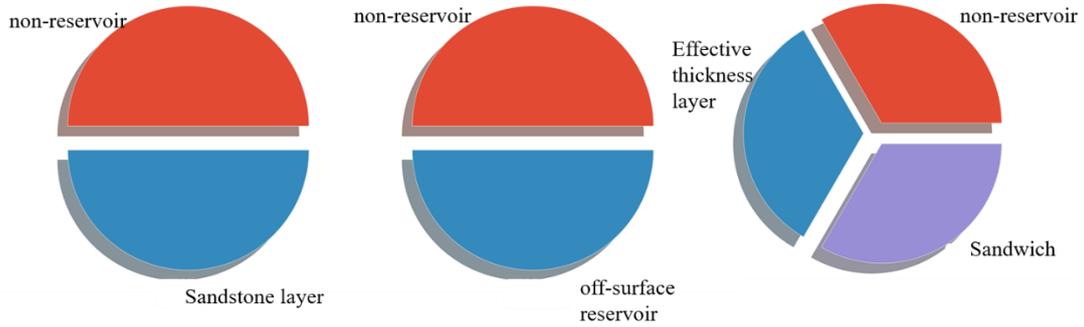


Fig. 4 Distribution ratio of sandstone and reservoir after SMOTETomek sampling

The equilibrium dataset obtained after sampling by the SMOTETomek algorithm was standardized to obtain training data with a normal data distribution, and the test well section data were fitted according to the distribution of the training well section to obtain test well section data conforming to the training sample. The final valid input data were 246051 training data and 13972 test data with a normal distribution of well depth and 12 logging curves.

4.2 Rock formation identification results of the 4 algorithms

In this paper, the functions used to measure the effect of rock layer and reservoir identification are F1-score, precision, recall and accuracy²⁰. The F1-score is a measure of the classification problem; it is the harmonic average of precision and recall, the maximum is 1, and the minimum is 0. Precision refers to the proportion of positive samples among the positive cases judged by the classifier²¹; recall refers to the proportion of positive cases among the total positive cases predicted by the classifier; accuracy represents the proportion of the correct judgment of the whole sample by the classifier, and the formula is as follows:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (5)$$

$$recall = \frac{TP}{TP + FN} \quad (6)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

where TP is the correct predicted answer; FP is the incorrect prediction of other classes as this class; FN is the prediction of this class label as another class label; and TN is the incorrect predicted answer.

Four algorithms were used to perform the formation and reservoir identification process based on the study block logging data (Fig. 5). Valid data after data preprocessing were entered into the LR, SVM, RF, and XGBoost algorithms. Due to the unbalanced logging data in this classification, there are limited valid data for the rock and reservoir layers, so recall and accuracy should be considered in the results.

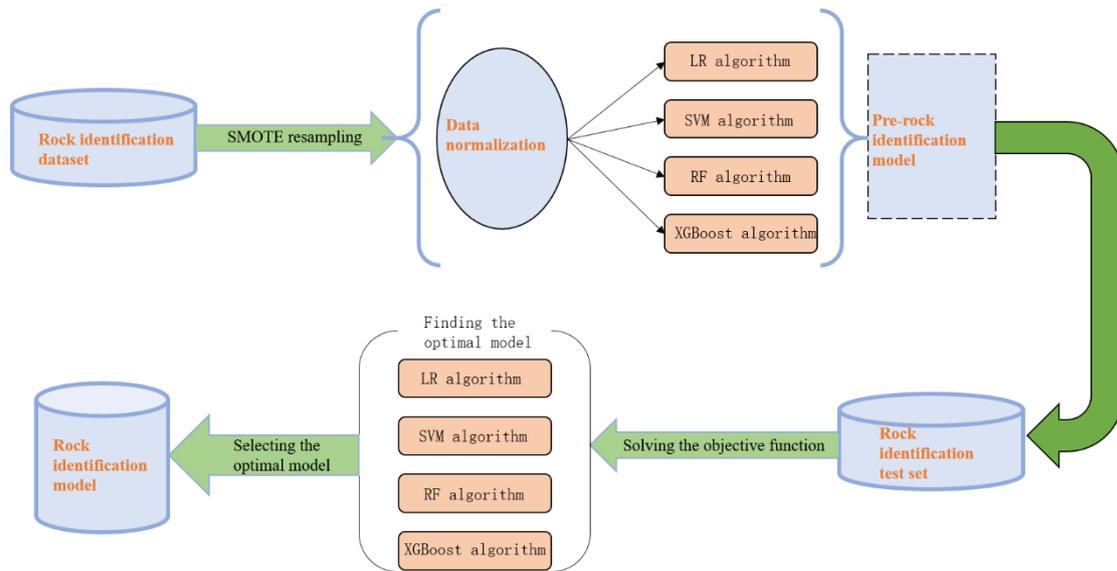


Fig. 5 Identification process of strata and reservoirs

The configuration used by the paper in training the model is as follows: AMD Ryzen 5 CPU, NVIDIA GeForce GTX 1660 GPU, and 16G running memory (Tang, J. et al., 2019). For the sandstone layer, XGBoost has an 86% recall with 96% accuracy; for the off-surface reservoir layer, XGBoost has an 85% recall with 94% accuracy; for the effective thickness reservoir layer, XGBoost has an 81% recall with accuracy remaining above 95%, while the LR, SVM and RF algorithms also have an accuracy of over 90%. It can be seen in the overall F1-score that these 3 algorithms are not as

effective as XGBoost for integrated rock layer identification; in the process of effective thickness layer identification, due to the distribution ratio of the three types of interlayers in the stratum and the characteristics of the algorithms of LR and SVM²², the 2 algorithms cannot converge, so they are not discussed here. From the time point of view, the training time of the XGBoost algorithm for logging curve data is also the least among the 4 algorithms, which takes only 3 seconds, indicating that the XGBoost algorithm is more popular in the field of rock formation and reservoir identification. (Table 3). On this basis (Fig. 6), correlation analysis of logging curve characteristics, the logging curve with the highest sensitivity to the effect of rock formation and reservoir identification was calculated by the analysis²³. The results showed that the response of the RMN, RLLD and RLLS is the most obvious for the sandstone layer, off-surface reservoir layer and effective thickness layer, and the well diameter curve has the least effect on the rock and reservoir layer identification. These results can provide an effective reference for the selection and dimensionality reduction of subsequent logging curves.

Table 3 Comparison of LR, SVM, RF and XGBoost algorithms

Algorithms		Well No.			Operating Time
		N1-10	N1-20	N1-31	
LR	Recall	0.84	0.87	0.78	5s
	Accuracy	0.90	0.92	0.91	
SVM	Recall	0.81	0.85	0.78	13s
	Accuracy	0.90	0.92	0.91	
RF	Recall	0.83	0.88	0.82	9s
	Accuracy	0.90	0.90	0.91	
XGBoost	Recall	0.88	0.92	0.92	3s
	Accuracy	0.95	0.94	0.97	

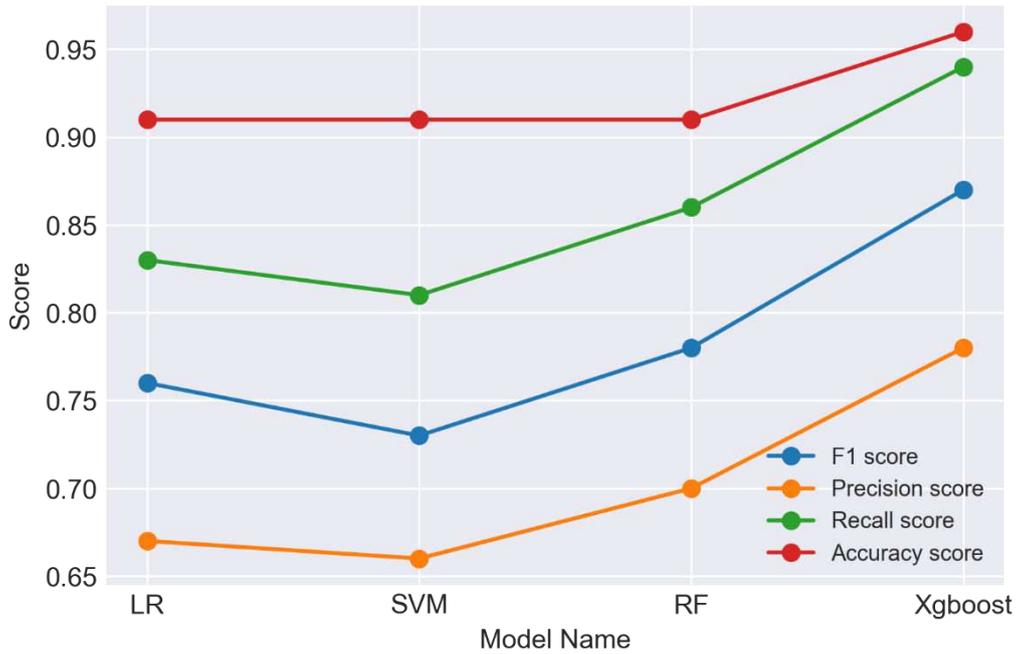
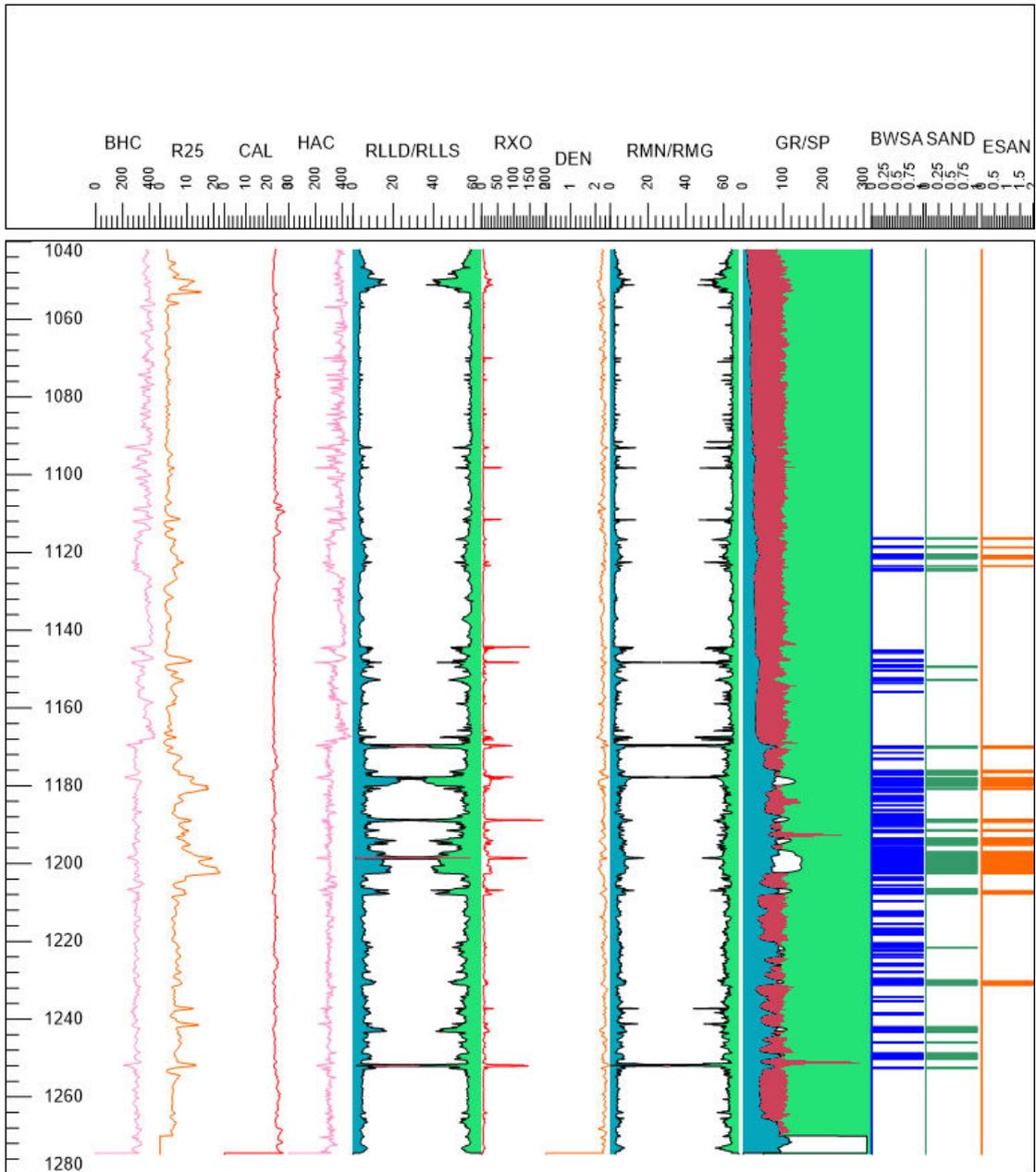


Fig. 6 Comparison of average scores of 4 models

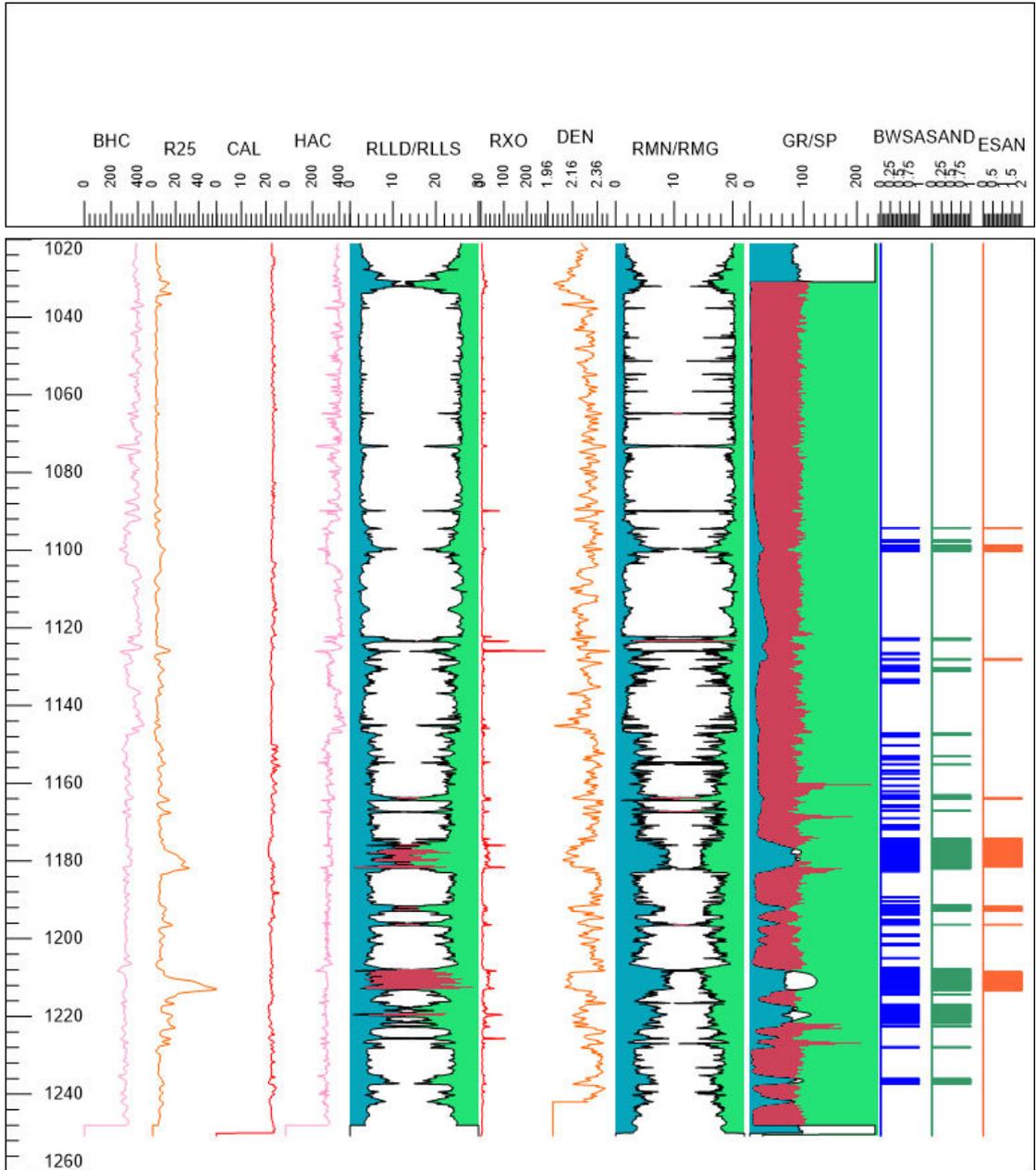
4.3 Analysis of the applicability of the method

The sandstones in the study block of this paper mostly develop trough-like interlacing laminations but also block formations, deformation formations, drainage formations, oblique laminations and horizontal laminations²⁴. The mudstone is a rock dominated by clay minerals. According to the silt content, there are two types of siltstone and mudstone. Siltstone is a type of mudstone with 10% to 50% silt content, and mudstone is a type of mudstone with <10% silt content. Therefore, the research method proposed in this paper can be well applied to strata with the above characteristics²⁵. The intelligent method proposed in this paper is a pioneering experiment in this field because few researchers have studied the problem of rock stratification and others have done so on the basis of experiments.

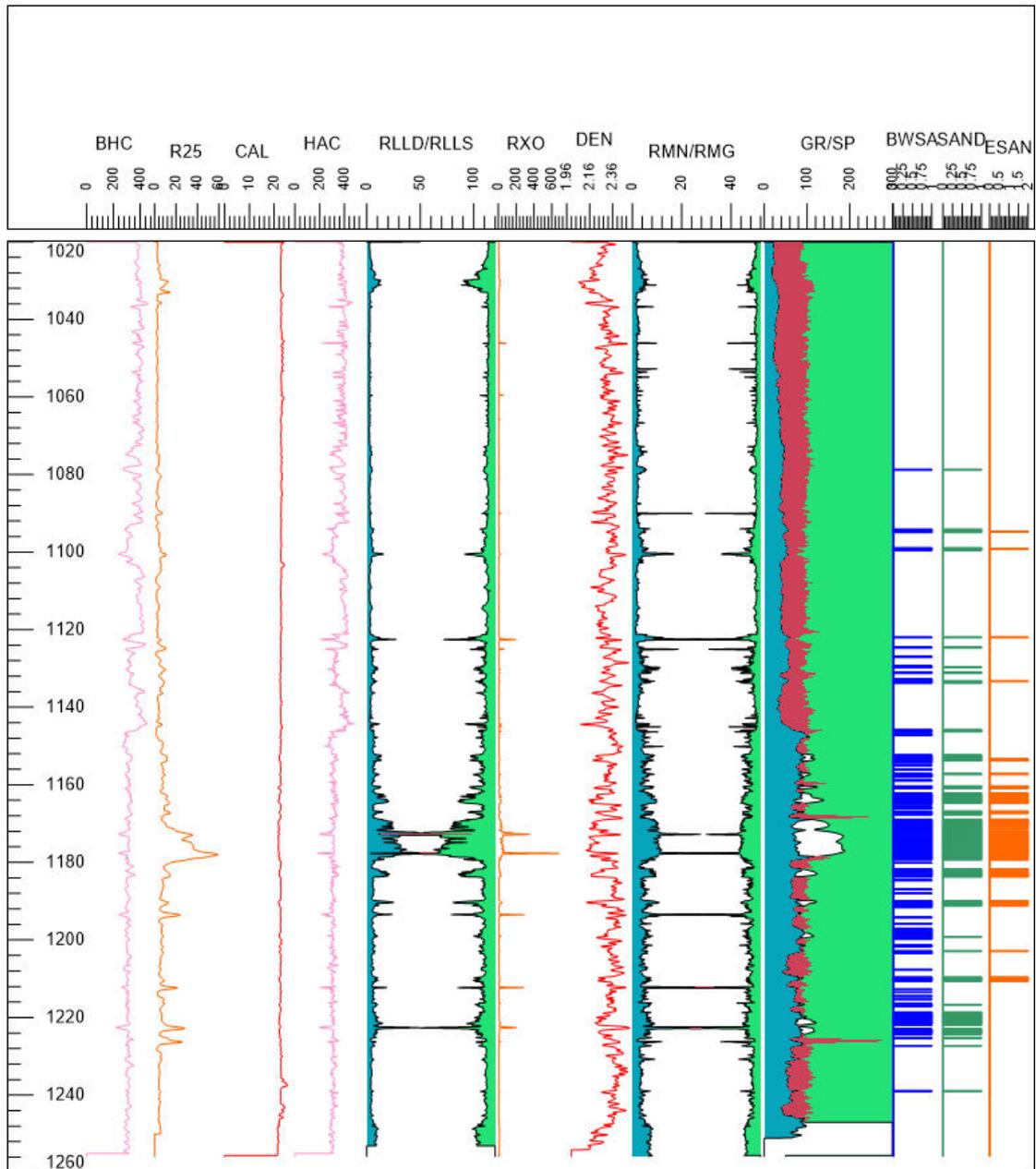
The specific interpretation results for the three wells are shown in Figure 7.



(a)



(b)



(c)

Fig. 7 Interpretation results of three wells (a, b, c are N1-10, N1-20 and N1-31)

5 Conclusions

(1) In the identification of rock formations and reservoirs, through field logging data, the selected logging depth and RMN, RMG, RLLD, RLLS, RXO, HAC, R25, CAL, GR, DEN, BHC, SP, a total of 13 response features, can effectively identify

reactive rock and reservoir layers intelligently by their location in the ground.

(2) The study selected 4 machine learning algorithms, LR, SVM, RF and XGBoost, and compared the F1-scores, precision, recall, accuracy and computation time of the 4 algorithms. Through a comprehensive comparison of the five discriminatory methods, the study shows that the XGBoost algorithm is the most effective for rock and reservoir layer identification, with an average accuracy of more than 95%.

(3) When faced with an uneven distribution of logging data, such as the three types of interlayer interference and nonreservoir interference, the SMOTETomek algorithm is selected to interpolate the interlayer and reservoir layers, which can effectively balance the data and improve the accuracy of formation and reservoir identification.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Youth Project) (Grant no. 52004064) and Postdoctoral Projects in Heilongjiang Province (Grant no. LBH-Z20044).

References

1. HE Ping, PAN Guo-feng. The Improvement of Fuzzy Isodata and Its Application to Reservoir Prediction[J]. Journal of Hebei University of Technology, 2002(02):101-105.
2. Song Jianguo, Gao Qiangshan, Li Zhe. Application of random forests for regression to seismic reservoir prediction[J]. Oil Geophysical Prospecting
3. WANG Zhong, ZHANG Shao-nan, KUANG Jian-chao. A PCA-SVM MODEL FOR RESERVOIR IDENTIFICATION[J]. Computing Techniques for Geophysical and Geochemical Exploration, 2010, 32(06):636-640+568.
4. LIU Ke, YIN Cheng, ZHAO Hu, et al. Prediction of sand body thickness based on a mixed density network constrained by a spatially approximated

- probability[J]. *Geophysical Prospecting for Petroleum*,2020,59(04):596-606+615.
5. SHAN Jing-fu, CHEN Xin-xin, ZHAO Zhong-jun, et al, Identification of complex lithology for tight sandstone gas reservoirs based on BP neural network[J]. *Progress in Geophysics*,2015,30(03):1257-1263.
 6. DU Chen-yu. Research on Reservoir Identification Method Based on Convolutional Neural Networks[D]. Chengdu University of Technology,2018
 7. ZHANG Jin-feng,LIU Yan-hong,WAN Wen-sheng. LITHOLOGY IDENTIFICATION AND RESERVOIR PREDICTION OF VOLCANIC RESERVOIR IN XIQUAN AREA[J]. *Xinjiang Oil & Gas*,2020,16(03):6-10+1.
 8. MOU Dan,WANG Zhu-Wen,HUANG Yu-Long,et al. Lithological identification of volcanic rocks from SVM well logging data:Case study in the eastern depression of Liaohe Basin[J].*Chinese Journal of Geophysics*,2015,58(05):1785-1793.
 9. WANG Zehua, ZHU Xiaomin, SUN Zhongchun,et al. Igneous lithology identification and lithofacies classification in the basin using logging data:Taking Junggar Basin as an example[J]. *Earth Science Frontiers*,2015,22(03):254-268.
 10. ZHU Yixiang, SHI Guangren. Identification of lithologic characteristics of volcanic rocks by support vector machine[J]. *Acta Petrolei Sinica*,2013,34(02):312-322.
 11. GUO Xu. Identification of coal reservoir by logging curves -- a case study of CK1 well[J] *Energy Technology and Management*,2020,45(05):166-168.
 12. Yu Daiguo, Sun Jianmeng, Wang Huanzeng, et al.A new method for logging lithology identification:SVM[J].*Petroleum Geology&Oilfield Development in Daqing*, 2005, 24 (5) :93-95.
 13. HAN Weixue, ZHOU Yatong, CHI Yue. Deep learning convolutional neural networks for random noise attenuation in seismic data[J]. *Geophysical Prospecting for Petroleum*,2018,57(06):862-869+877.
 14. Deng Mingzhe, Fang Chengmin, Deng Peng,et al. Origin of strike-slip through

- st structure in Lishu area,southern Songliao Basin:a case study of Xiaokuan fault belt[J]. *Acta Petrolei Sinica*,2020,41(09):1089-1099.
15. ZHAO Yingnan. Study on reservoir geological characteristics of SA0 formation in Sabei development area of Daqing Oilfield[J] *Chemical Enterprise Management*,2018(16):220.
 16. CHEN Tong. The Study of Logging Response Characteristics and Mechanism of Ultra-Long Borehole Cores Based on the Scientific Drilling in Songliao Basin[D].Jilin University,2018
 17. Hobza T , Pardo L , Vajda I . Robust median estimator in logistic regression[J]. *Journal of Statal Planning & Inference*, 2020, 138(12):3822-3840.
 18. Burges C J C . A Tutorial on Support Vector Machines for Pattern Recognition[J]. *Data Mining and Knowledge Discovery*, 1998, 2(2):121-167.
 19. Lin, Zhao-Mu, Guo,等. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach[J]. *Geoderma: A Global Journal of Soil Science*, 2015.
 20. Tianqi Chen, Tong He, Michael Benesty. xgboost: Extreme Gradient Boosting[J]. 2016.
 21. Tharwat A , Hassanien A E , Elnaghi B E . A BA-based algorithm for parameter optimization of Support Vector Machine[J]. *Pattern Recognition Letters*, 2017, 93(jul.1):13-22.
 22. Gong Qingshun, Liu Zhanguo, Song Guangyong,et al. Inter-layers in alluvial fan thick sandy conglomerate reservoir of Kunbei oilfield, Qaidam Basin[J] *Acta Petrolei Sinica*,2019,40(02):152-164.
 23. Ryu, B., Yoo, S., Kim, S. et al. Thirty-day hospital readmission prediction model based on common data model with weather and air quality data. *Sci Rep* 11, 23313 (2021).
 24. Candefjord, S., Nguyen, L., Buendia, R. et al. A wearable microwave instrument can detect and monitor traumatic abdominal injuries in a porcine model. *Sci Rep* 11, 23220 (2021).
 25. Li, J., Yu, Z., Zhou, Z. et al. Mechanical analysis and failure modes prediction

ction of composite rock under uniaxial compression. *Sci Rep* 11, 22826 (2021).