Research Article

Wang Wenhua, Wang Zhuwen*, Han Ruiyi, Xu Fanghui, Qi Xinghua, and Cui Yitong

Lithology classification of volcanic rocks based on conventional logging data of machine learning: A case study of the eastern depression of Liaohe oil field

https://doi.org/10.1515/geo-2020-0300 received March 05, 2021; accepted September 14, 2021

Abstract: The reservoirs in the eastern depression of Liaohe basin are formed by multistage igneous eruption. The lithofacies and lithology are complex, and the lithology is mainly intermediate and basic igneous rocks. Based on the integration of debris data of igneous rocks and logging data, this article selected 6,462 continuous logging data with complete cuttings data and five conventional logging curves (RLLD, AC, DEN, GR, and CNL) from four wells in the eastern depression of Liaohe basin as the training set. A variety of lithologic identification schemes based on support vector machine and random forest are established to classify the pure igneous strata and actual strata. By comparing the classification results with the identification data of core slice and debris slice, a practical lithologic classification scheme for igneous rocks in the eastern depression of Liaohe basin is obtained, and the classification accuracy reaches 97.46%.

Keywords: igneous rock, machine learning, lithology classification

Wang Wenhua: Key Laboratory for Evolution of Past Life and Environment in Northeast Asia (Jilin University), Ministry of Education, Changchun 130026, China; College of Earth Sciences, Jilin University, Changchun, Jilin 130021, China

Han Ruiyi, Xu Fanghui, Qi Xinghua, Cui Yitong: College of Geo-Exploration Science and Technology, Jilin University, Changchun, Jilin 130026, China

1 Introduction

From Mesozoic to Cenozoic, the eastern depression has been regarded as the magmatic activity center of Liaohe Depression, where igneous rocks are widely developed [1]. The main lithologies in the eastern depression are trachyte, basalt, and diabase [2]. The igneous rock exploration with trachyte as the main reservoir in this area has realized the continuous oil and gas bearing situation in many areas. The practice of igneous oil and gas exploration shows that the diagenetic mode of igneous rocks has a direct impact on its reservoir properties [3,4]. The different chemical compositions of igneous rocks with similar diagenetic patterns lead to great differences in pore types and spatial distribution after diagenesis. Therefore, it is of practical significance to distinguish igneous rocks according to their chemical composition [5].

The types of igneous rocks are complex. Different eruption modes and periods and different regions and horizons have different lithologies. Not only are the mineral compositions of rocks in the same region or horizon different, but the mineral compositions of the same lithology are also very different [6]. Due to the limitation of quantity, the classification of igneous rocks cannot be evaluated systematically according to the actual core data. The conventional logging data are relatively easy to obtain, rich, and comprehensive, and can reflect the formation characteristics in the vertical direction. It is worth noting that different types of igneous rocks have great differences in conventional logging data, especially natural gamma, acoustic, compensated neutron, and resistivity curves logging data, which provide a basis for us to use conventional logging data for igneous rock lithology identification. There are many methods to identify igneous rocks by using logging data, such as cross-plot method, imaging logging method, and formation element logging method. Among them, the cross-plot method is the simplest.

^{*} Corresponding author: Wang Zhuwen, College of Geo-Exploration Science and Technology, Jilin University, Changchun, Jilin 130026, China, e-mail: wangzw@jlu.edu.cn

a Open Access. © 2021 Wang Wenhua et al., published by De Gruyter. © This work is licensed under the Creative Commons Attribution 4.0 International License.

First, select the standard igneous rock data and cross-plot the conventional logging curves in pairs. Different types of igneous rocks will generally fall in different areas in the cross-plot, which can effectively distinguish the types of igneous rocks. However, the cross-plot method needs a large number of accurate standard lithologic data, and is limited by the geological region. At the same time, due to the complexity and particularity of the igneous rock, different types of igneous rock data may overlap in the cross-plot, so the cross-plot method cannot be used as a separate lithology identification method, but needs to be combined with other identification methods. Imaging logging method and formation element logging method can identify the lithology of volcanic rocks from the structure and chemical composition, respectively, but these two methods are expensive, and the cost of large-scale application is too high.

Machine learning is one of the most advanced research fields in artificial intelligence. As a way to realize artificial intelligence [7–10], machine learning has aroused wide interest in the field of artificial intelligence. Machine learning is not only applied in knowledge-based systems, but also widely used in natural language processing [11,12], machine vision, pattern recognition, and many other fields [13,14]. Support vector machine (SVM) is a widely used kernel machine learning algorithm [15,16], which has excellent network generalization performance in solving small sample learning and high-dimensional pattern recognition [17,18]. Random forest (RF) is a kind of statistical learning theory [19]. It uses bootstrap resampling method to extract multiple samples from the original samples, models the decision tree for each bootstrap sample and then combines the prediction of multiple decision trees to get the final prediction result by voting.

According to the conventional logging data of igneous rock area in Liaohe east depression, this article uses SVM and RF methods to classify the lithology under the condition of pure igneous rock and then designs different schemes to classify the lithology under the stratum condition, and divides the lithology of actual well section [20–22]. Finally, the correctness of identification results is verified by combining with actual core data and lithology section.

2 Method, classifiers, and data sets

2.1 Method

Methodology of this work is divided into three steps:

Step 1: Suppose a stratum has six kinds of igneous rock lithology under the condition of pure igneous rock, design two schemes for classification, and determine the better scheme.

Step 2: Sedimentary rock and coal are added to the formation to construct the formation under actual conditions. Based on the previous optimization scheme, four classification schemes are designed to determine an optimal scheme suitable for the actual formation conditions.

Step 3: The optimal scheme is applied to four wells for lithologic identification, and compared with lithologic section and thin section, so as to verify the accuracy of the scheme.

2.2 Classifiers

2.2.1 SVM

SVM is a common discriminant method. It is a supervised learning model in the field of machine learning, which is usually used for pattern recognition, classification, and regression analysis [23,24]. The SVM method maps the sample space to a high-dimensional or even infinite dimensional feature space (Hilbert space) through a nonlinear mapping P, so that the nonlinear separable problem in the original sample space can be transformed into a linear separable problem in the feature space. In the classification problem, for the sample set that may not be linearly processed in the low dimensional sample space, SVM first completes the calculation in the low dimensional space, then maps the input space to the high dimensional feature space through the kernel function, and finally constructs the optimal separation hyperplane in the high dimensional feature space, so as to separate the nonlinear data which is not easy to separate on the plane [25–27]. The process of finding hyperplane can be transformed into solving a quadratic programming problem [28-30].

SVM classifier is implemented by SVM module in scikit-learn library of Python programming language. SVM in scikit-learn library has four kernel functions: linear kernel function, polynomial kernel function, radial basis function (RBF), and sigmoid kernel function. In this article, the radial basis kernel function is used for the following reasons:

- There are many parameters of polynomial kernel function, and the process of parameter optimization is complex.
- (2) Linear kernel function can only be used to deal with linear problems.

- (3) The sigmoid kernel function and the kernel function of RBF are similar under some conditions and can replace each other.
- (4) The five conventional logging data sets used in this article are not high-dimensional data sets, so it is more appropriate to use radial basis kernel function. There are two important parameters of RBF kernel function, which are penalty factor C and kernel parameter gamma.

In this article, 30% of the training set is taken as ten times cross-validation of the back-judgment test set, and the penalty factor C and the parameter gamma are optimized as the optimal values. The Z-score normalization of all the training set and test set data of SVM is carried out.

2.2.2 RF

RF is a machine learning algorithm based on decision tree, which combines the decision tree into a RF, that is, randomize the use of variables and data, generate many decision trees and then summarize the classification results. Like other classification methods, RF can explain the effect of several independent variables on the dependent variable y. If the dependent variable y has n observations and its k features are always related, the RF will put back *n* observations in the data set when constructing the decision tree, which is called bootstrap resampling method. At the same time, RF selects a part of k features randomly to construct the decision tree nodes, which can ensure that the decision tree generated each time is inconsistent. In general, a RF will generate several decision trees randomly for classification and then summarize the classification results of each tree and decide the classification results by voting.

RF classifier is implemented by RF classifier module in scikit-learn library of Python programming language. It needs to set three main parameters: the number of decision trees (*n*_estimators), the maximum depth of decision trees (max_depth), and the maximum number of features of decision trees (max_features). The optimization process of the three main parameters is divided into two parts:

- (1) Using the out-of-bag data generated by bootstrap resampling to estimate the internal error, and the out-of-bag error must be controlled to be low enough and stable.
- (2) Taking 30% of the training set as ten times of crossvalidation to ensure that the cross-validation results are high enough.

RF classifier does not need to preprocess the data.

2.3 Selection of data sets

This article mainly studies the eastern depression of Liaohe basin, where the igneous rocks are mainly intermediate basic rocks [31,32]. According to the lithologic classification scheme of deep igneous rocks in Song Liao basin, combined with the feasibility of logging lithologic identification and the actual production demand, the igneous rocks in the study area are divided into six kinds of lithology: compact trachyte, non-compact trachyte, compact basalt, non-compact basalt, diabase, and gabbro. There are many kinds of mineral compositions in igneous reservoir rocks, and the pore and reservoir types are complex. The logging response characteristics are more complex. It is difficult to identify the lithology of igneous rocks effectively by using conventional logging interpretation methods.

Igneous rocks, coal, and sedimentary rocks coexist in the eastern depression of Liaohe basin. There are obvious differences among the three kinds of lithology in the characteristics of Genesis, structure, and mineral composition, which are difficult to be accurately distinguished by conventional logging interpretation. According to the actual needs of producers, igneous reservoirs in this area are mainly concentrated in non-compact trachyte and non-compact basalt, so the key to lithology classification is to accurately select non-compact trachyte and non-compact basalt. To identify the above two kinds of lithology as accurately as possible, this article divides the lithology classification of the area into two stages, first for the pure igneous strata and second for the actual strata.

In the conventional logging methods, RLLD, AC, CNL, DEN, and GR curves reflect the petrological characteristics of conductivity, sound velocity, porosity, density, and radioactivity, respectively, (Table 1). In this article, 6,462 continuous logging data of complete cuttings data from four wells in the eastern depression of Liaohe basin are selected as the training set, including 2,202 sedimentary rock data and 4,162 igneous rock data as well as continuous rock data of six lithology, compact trachyte, non-compact trachyte, compact basalt, noncompact basalt, diabase, and gabbro, and 98 coal seam data. It should be noted that because the actual selection of data is difficult, the data quantity of different lithologies in the training set varies greatly, that is, the training set used in this article is a typical unbalanced data set.

In addition, since the focus of this article is to classify the igneous rocks in the target area, all types of sedimentary rocks in this area are collected as sedimentary rock data set. It includes mudstone, carbonaceous mudstone, basaltic mudstone, silty mudstone, glutenite, argillaceous

Lithology	CNL (%)	DEN (g cm ⁻³)	GR (API)	AC (μs m ⁻¹)	RLLD (Ω m)
Non-compact trachyte	16-24	2.1-2.3	152-160	223-263	42-80
Compact trachyte	4-8	2.3-2.6	140-155	171-197	44-2,000
Non-compact basalt	24-36	2.3-2.6	40-56	197-263	10-20
Compact basalt	24-29	2.6-2.8	28-40	184-236	8-12
Diabase	12-18	2.6-2.7	30-45	171-210	128-800
Gabbro	13-16	2.6-2.7	60-80	158-171	698-8,000

Table 1: Logging response of igneous rock in eastern depression, Liaohe oil field [33]

glutenite, argillaceous siltstone, gravelly sandstone, fine sandstone, and other common sedimentary rock data in the eastern depression area of Liaohe basin.

In this article, 5,070 continuous logging data from seven wells in the eastern depression of Liaohe basin, which are different from the training set, are selected as the test set, including 801 sedimentary rock data as well as common sedimentary rock types of the same type as the training set. There are 4,158 igneous rock data, including continuous strata data of six kinds of lithology, compact trachyte, non-compact trachyte, compact basalt, non-compact basalt, diabase, and gabbro, and 111 coal seam logging data.

3 Experimental verification

3.1 Lithologic classification of pure igneous strata

In this stage, it is assumed that there is an ideal pure igneous rock stratum, that is, it is assumed that sedimentary rock and coal in the stratum have been screened by other means, and only igneous rock exists. To get more accurate classification results, two sets of lithologic classification schemes are designed as follows (Figure 1):

Scheme A: Direct use of SVM classifier and RF classifier to classify six lithologies of compact trachyte, non-compact trachyte, compact basalt, non-compact basalt, diabase, and gabbro.

Scheme B: Four lithologies of trachyte, basalt, diabase, and gabbro in igneous formation are classified by SVM classifier and RF classifier and then the compactness of basalt and trachyte is classified by the same classifier (Figure 2).

Lithological classification of pure igneous formation is carried out according to the above two schemes (Table 2). It can be found that out of the four classification results of the two schemes mentioned above, the RF method used in scheme A, which directly classifies the six lithologies, has only 10.25% correct classification rate for non-compact trachyte and 25.6% correct comprehensive classification rate, which is obviously not competent for classification requirements. The classification accuracy of scheme A is 97.14%, but the classification accuracy of compact trachyte and non-compact basalt is 58.33

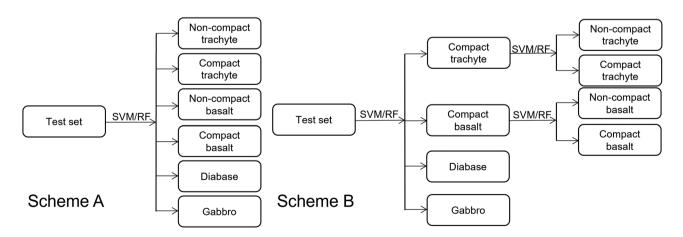


Figure 1: Flow chart of lithology classification of pure igneous strata.



Figure 2: Results of lithology classification of pure igneous strata.

and 58.17%, respectively, which cannot meet the actual production needs.

Two classification methods are used in scheme B. The SVM method has a higher accuracy for non-compact trachyte and gabbro. For compact trachyte, non-compact

basalt, compact basalt, and diabase, the classification accuracy is lower than that of RF method and the total classification accuracy is slightly lower than that of RF method. The RF classification method used in scheme B is more accurate for each lithology classification.

Table 2: Statistics of lithological classification results of pure igneous formation

Lithology	Data and results	Scheme A SVM	Scheme A RF	Scheme B SVM	Scheme B RF
Non-compact trachyte	Return correct rate (%)	97.88	100	97.99	99.59
	Number of test samples	3,179	3,179	3,179	3,179
	Number of correct	3,179	326	3,179	3,166
	Correctness (%)	100	10.25	100	99.59
Compact trachyte	Return correct rate (%)	97.88	100	97.99	99.59
	Number of test samples	48	48	48	48
	Number of correct	28	48	48	48
	Correctness (%)	58.33	100	100	100
Non-compact basalt	Return correct rate (%)	97.88	100	97.99	99.59
	Number of test samples	153	153	153	153
	Number of correct	89	147	123	153
	Correctness (%)	58.17	96.08	80.39	100
Compact basalt	Return correct rate (%)	97.88	100	97.99	99.59
	Number of test samples	178	178	178	178
	Number of correct	178	177	170	178
	Correctness (%)	100	99.44	95.51	100
Diabase	Return correct rate (%)	97.88	100	99.64	99.59
	Number of test samples	300	300	300	300
	Number of correct	265	300	264	294
	Correctness (%)	88.33	100	88	98
Gabbro	Return correct rate (%)	97.88	100	99.64	99.59
	Number of test samples	300	300	300	300
	Number of correct	300	300	288	278
	Correctness (%)	100	100	96	92.67
Total	Return correct rate (%)	97.88	100	_	_
	Number of test samples	4,158	4,158	4,158	4,158
	Number of correct	4,039	1,298	4,072	4,117
	Correctness (%)	97.14	25.60	97.93	99.01

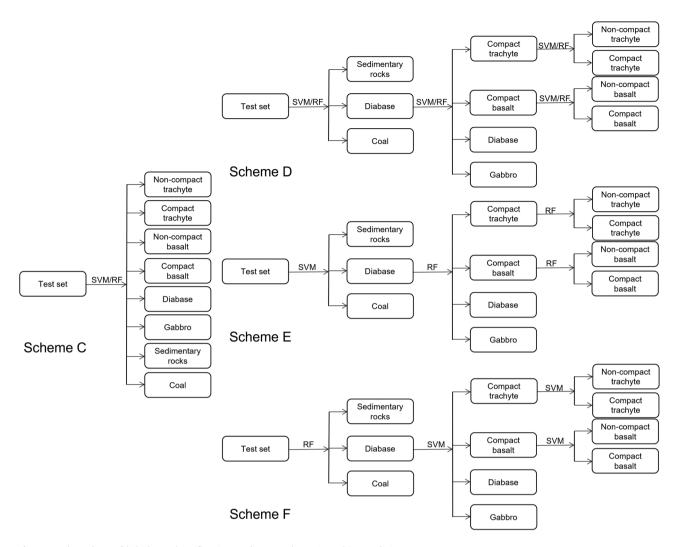


Figure 3: Flow chart of lithology classification under actual stratigraphic conditions.

3.2 Lithological classification under actual stratigraphic conditions

This stage is lithological classification under actual formation conditions, i.e., formation is divided into sedimentary, coal, and igneous rocks according to the feasibility of logging lithological identification and actual production requirements. Because the conventional logging curves of some sedimentary rocks and igneous rocks are similar in characteristics, it is difficult to distinguish lithology effectively by one multi-classification, so four lithology classification schemes are designed in this stage, as follows (Figure 3):

Scheme C: Sedimentary rocks, coal seams, dense trachyte, non-dense trachyte, dense basalt, non-dense basalt, diabase, gabbro, and other eight lithologies are directly classified by SVM and RF classifier.

Scheme D: By SVM and RF classifier, the strata are first divided into sedimentary rocks, igneous rocks, and

coal and then the igneous rocks are classified by Scheme B using the same classifier.

Scheme E: Strata are divided into sedimentary rocks, igneous rocks, and coal by SVM classifier and then lithological classification is carried out by RF classifier of scheme B.

Scheme F: Strata are divided into sedimentary rocks, igneous rocks, and coal by RF classifier and then lithological classification is carried out by SVM classifier of scheme B (Figure 4).

Lithological classification of actual strata is carried out according to the above four schemes (Table 3). It can be found that scheme C, which directly uses the classifier for one-off multi-classification, has poor classification results for both classifiers and cannot reach the classification accuracy required in actual production.

Schemes D and F for classification of coal and igneous rocks using RFs have low accuracy in overall classification of igneous rocks, which indicates that RFs cannot effectively

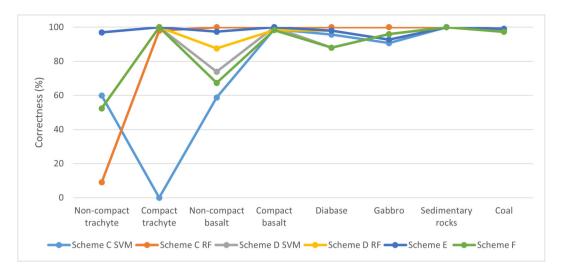


Figure 4: Results of lithology classification of actual strata.

 Table 3: Statistics of lithological classification results of actual strata

Lithology	Data and results	Scheme C SVM	Scheme C RF	Scheme D SVM	Scheme D RF	Scheme E	Scheme F
Non-compact	Return correct rate (%)	93.74	95.19	95.65	93.55	97.21	92.05
trachyte	Number of test samples	3,179	3,179	3,179	3,179	3,179	3,179
	Number of correct	1,906	287	3,083	1,665	3,083	1,665
	Correctness (%)	59.96	9.03	96.98	52.37	96.98	52.37
Compact trachyte	Return correct rate (%)	93.74	95.19	95.65	93.55	97.21	92.05
	Number of test samples	48	48	48	48	48	48
	Number of correct	0	47	48	48	48	48
	Correctness (%)	0	97.92	100	100	100	100
Non-compact basalt	Return correct rate (%)	93.74	95.19	95.65	93.55	97.21	92.05
	Number of test samples	153	153	153	153	153	153
	Number of correct	90	153	113	134	149	103
	Correctness (%)	58.82	100	73.86	87.58	97.39	67.32
Compact basalt	Return correct rate (%)	93.74	95.19	95.65	100	97.21	92.05
	Number of test samples	178	178	178	178	178	178
	Number of correct	176	178	178	175	178	175
	Correctness (%)	98.88	100	100	98.31	100	98.31
Diabase	Return correct rate (%)	93.74	95.19	95.65	100	97.21	93.60
	Number of test samples	300	300	300	300	300	300
	Number of correct	287	300	264	294	294	264
	Correctness (%)	95.67	100	88.00	98.00	98.00	88.00
Gabbro	Return correct rate (%)	93.74	95.19	95.65	93.55	97.21	93.60
	Number of test samples	300	300	300	300	300	300
	Number of correct	272	300	288	278	278	288
	Correctness (%)	90.67	100	96.00	92.67	92.67	96.00
Sedimentary rocks	Return correct rate (%)	93.74	95.19	97.61	93.94	97.61	93.94
	Number of test samples	801	801	801	801	801	801
	Number of correct	801	801	801	801	801	801
	Correctness (%)	100	100	100	100	100	100
Coal	Return correct rate (%)	93.74	95.19	97.61	93.94	97.61	93.94
	Number of test samples	111	111	111	111	111	111
	Number of correct	109	110	110	108	110	108
	Correctness (%)	98.20	99.10	99.10	97.30	99.10	97.30
Total	Return correct rate (%)	93.74	95.19	_	_		
	Number of test samples	5,070	5,070	5,070	5,070	5,070	5,070
	Number of correct	3,641	2,176	4,885	3,503	4,941	3,552
	Correctness (%)	71.81	42.92	96.35	69.09	97.46	70.06

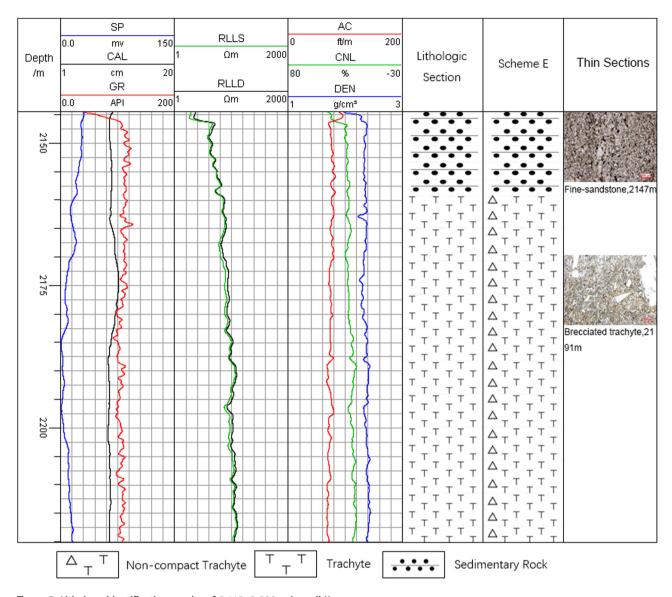


Figure 5: Lithology identification results of 2,145-2,220 m in well X.

distinguish sedimentary rocks from igneous rocks. Compared with scheme E, which uses SVM to classify sedimentary rocks, coal, and igneous rocks, and scheme E, which uses RF to classify igneous rocks, the classification accuracy of scheme E for gabbro is slightly lower than that of scheme D, 92.67%, and the classification accuracy of other lithologies is not lower than that of scheme D, and the comprehensive classification accuracy reaches 97.46%.

4 Results and discussion

To validate the accuracy of the method described in scheme E, four wells in the east depression of Liaohe oil field are taken as examples to illustrate the igneous lithological characteristics and logging lithological identification results.

Figure 5 is the result map of lithological identification of Well X from 2,145 to 2,220 m. It can be seen from the figure that the classification method described in Scheme E can distinguish the formation boundary between sedimentary rock and trachyte in Well X according to the difference in logging responses. Based on the lithological profile, trachyte in target formation can be identified as non-compact trachyte and the thickness of the formation can be determined.

The depth of well section is 2,158-2,220 m, in which 2,200 m cuttings specimen is described as trachyte, of which 70% of the cuttings are trachyte (Figure 6a) with

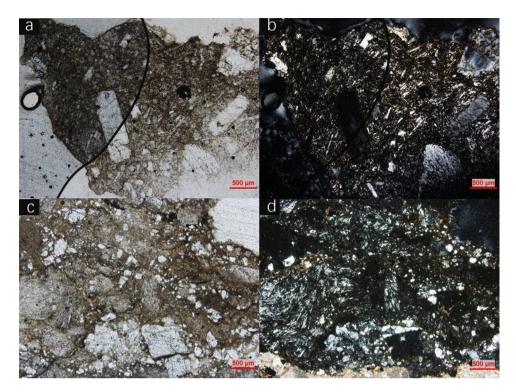


Figure 6: Well X, 2,200 m, cuttings slice. (a) Trachyte (monopolar), (b) trachyte (orthogonally polarized), (c) trachyte igneous breccia (single polarized light), and (d) trachyte igneous breccia (orthogonal polarized light).

porphyry structure, block structure, and local polymorphic structure, mainly alkaline feldspar; and a few plagioclase phenocrysts are also observed. The matrix is of trachytic texture, alkaline feldspar microcrystal and glassy are found in the matrix. Feldspar clay mineralization is developing generally. 20% of them are trachyte igneous breccia (Figure 6c). Igneous breccia structure is a massive structure. 50% of the breccia content is trachyte and basalt clasts are also found. Igneous breccia is poorly sorted and rounded-subprismatic; another 10% is mudstone.

Core specimens at a depth of 2133.3 m are described: Lithology is trachyte, phenocryst is potassium feldspar, strongly dissolved, reaction edge is visible at the edge, matrix is directed by feldspar microcrystals of potassium feldspar, trachytic in texture, and a small amount of pyroxene is visible. The fracture is self-breaking seam and unfilled (Figure 7). The results show that the lithologic classification results of scheme E are consistent with the core cuttings data.

Figure 8 shows the lithologic identification results of well O 3,320–3,380 m. It can be seen from the figure that scheme E can distinguish the boundary between basalt and trachyte in Well O according to the difference in the logging response. Based on the lithologic profile, the

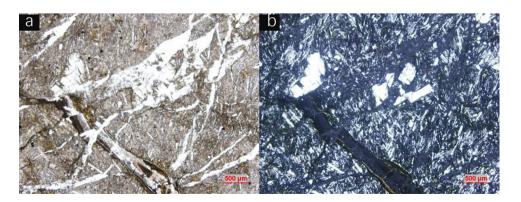


Figure 7: Well X, 2193.3 m, trachyte. (a) Core slice (single polarization) and (b) core slice (orthogonal polarization).

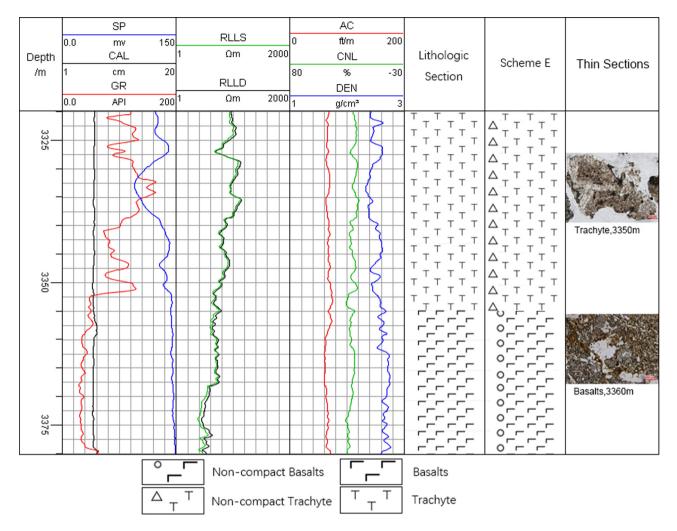


Figure 8: Lithology identification results of Well O 3,320-3,380 m.

trachyte in the target formation is identified as non-compact trachyte and basalt is non-compact basalt, and the thickness of the formation is determined.

The depth is 3,320–3,352 m, in which 3316.5 m cuttings are described as follows: The lithology is comprehensively

named trachyte, in which 80% of the cuttings are trachyte (Figure 9), with porphyry structure, massive structure, and local porphyry structure. The phenocrysts are mainly alkaline feldspar (5%). A small amount of plagioclase phenocrysts can also be seen, and the matrix is of coarse texture.



Figure 9: Well O, 3316.5 m, trachyte. (a) Trachyte (monopolar) and (b) trachyte (orthogonally polarized).

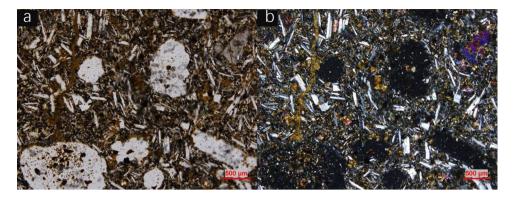


Figure 10: Well O, 3,360 m, basalt. (a) Basalt (single polarization) and (b) basalt (orthogonal polarization).

Alkali feldspar microcrystalline and glassy can be seen in the matrix, and feldspar clay mineralization is generally developed. Another 20% is mudstone.

Description of core specimen at depth of 3,360 m: The lithology is basalt (Figure 10), porphyritic texture, phenocryst is augite, and matrix is intergranular cryptic

texture, and the plagioclase microcrystals are directionally distributed in the brown black igneous glass. The whole rock has developed dissolution pores, unfilled, and moderately altered. The results show that the lithologic classification results of scheme E are consistent with the core cuttings data.

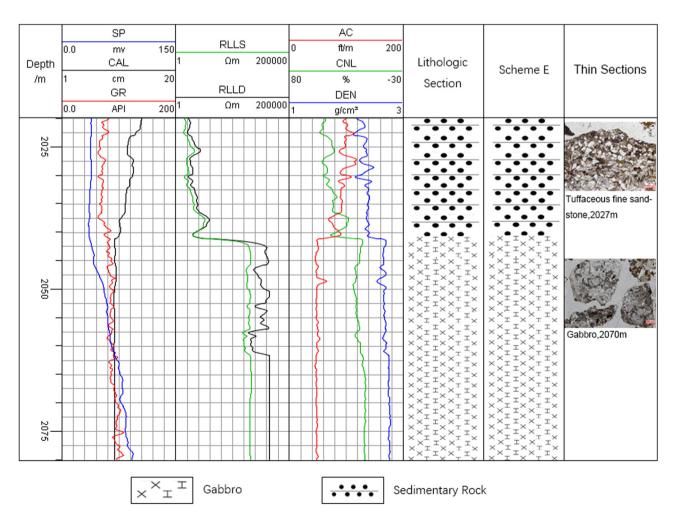


Figure 11: Lithological identification results of Well J from 2,020 to 2,080 m.

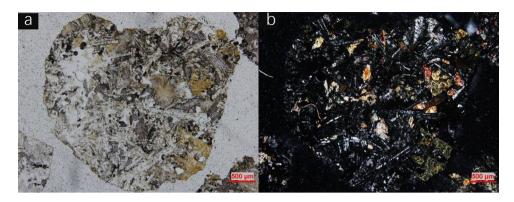


Figure 12: Well J, 2,050 m gabbro. (a) gabbro (single polarized) and (b) gabbro (orthogonal polarized).

Figure 11 shows the lithologic identification results of J Well from 2,020 to 2,080 m. It can be seen from the figure that scheme E can distinguish the boundary between the sedimentary rock and gabbro in Well J according to the difference in logging response and determine the thickness of the formation at the same time.

Depth of well section 2,041–2,080 m, including 2,050 m cuttings specimen description: the lithology is comprehensively named gabbro, 90% of which is gabbro structure (Figure 12). The main mineral compositions are basic plagioclase (50%), clinopyroxene (30%) and olivine (20%). Particle diameter is 1–2 mm with high degree of crystallization.

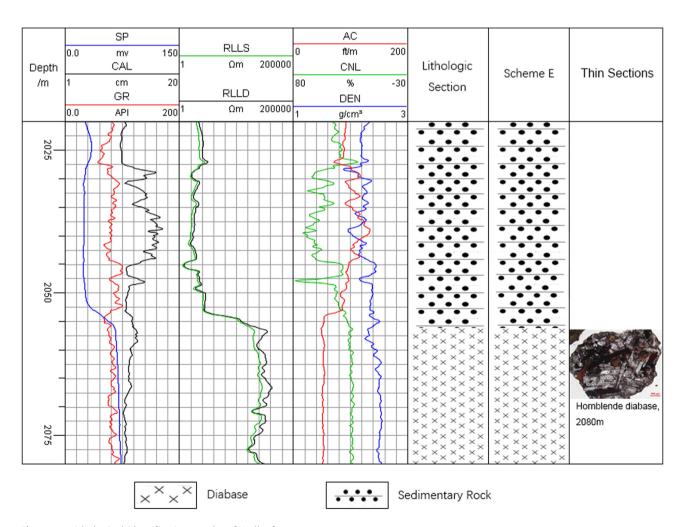


Figure 13: Lithological identification results of Well L from 2,020 to 2,080 m.

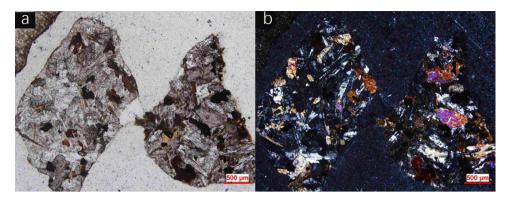


Figure 14: L Well, 2,058 m, amphibolite-diabase. (a) diabase (single polarization) and (b) diabase (orthogonal polarization).

Another 10% is mudstone. The result shows that the lithological classification result of scheme E is consistent with the data of core and cuttings.

Figure 13 shows the lithological identification results of Well L from 2,020 to 2,080 m. As can be seen from the figure, Scheme E can distinguish the boundary between the sedimentary rocks and diabase in Well L and determine the thickness of the formation according to the difference in logging responses.

The depth of the well section is 2,056-2,080 m, of which 2,058 m cuttings are described as follows: The lithology is comprehensively named amphibolite-diabase, 80% of which is amphibolite-diabase structure (Figure 14). The minerals are mainly pyroxene and plagioclase with a few biotite and amphibole and another 20% is mudstone. The result shows that the lithological classification result of scheme E is consistent with the data of core and cuttings.

5 Conclusion

In this article, the igneous rock types are summarized from the geological coring and mineral identification data, and the logging response combination characteristics of logging curves are analyzed from the logging data. Six igneous lithologies of basalt, non-compact basalt, trachyte, non-compact trachyte, diabase, and gabbro were effectively identified by SVM and RF methods. Finally, mutual proof between geological data and logging data is achieved, and the purpose of igneous lithology identification is achieved, and the following conclusions are drawn:

(1) A classification scheme of igneous rocks is determined, which can be used in actual stratigraphic conditions, and the highest classification accuracy is 97.46%.

- (2) For lithological classification of pure igneous formation, the method of RF sectional classification used in scheme B has the highest classification accuracy, reaching 99.01%.
- (3) When the training set is unbalanced data, the RF classifier cannot effectively distinguish sedimentary rocks from igneous rocks, i.e., under unbalanced classification, the effect of RF classifier is significantly lower than that of SVM.

Funding information: The work described in this article is supported by the National Natural Science Foundation of China (No. 41874135 and No. 41790453) and the National Key R&D Program of China (2019YFC0605402).

Conflict of interest: Authors state no conflict of interest.

References

- Zongli L, Zhuwen W, Dapeng Z, Shuqin Z, Min X. Pore distribution characteristics of the igneous reservoirs in the eastern sag of the Liaohe depression. Open Geosci. 2017;9(1):161-73. doi: 10.1515/geo-2017-0014.
- Mou D, Wang ZW, Huang YL, Xu S, Zhou DP. Lithological identification of volcanic rocks from SVM well logging data: case study in the eastern depression of Liaohe Basin. Chin J Geophys (Acta Geophysica Sin). 2015;58(5):1785-93. doi: 10.6038/cjg20150528.
- Zou C, Zhao W, Jia C, Zhu R, Zhang G, Zhao X, et al. Formation and distribution of volcanic hydrocarbon reservoirs in sedimentary basins of China. Pet Expl Dev. 2008;35(3):257-71. doi: 10.1016/S1876-3804(08)60071-3.
- Jin C, Pan W, Qiao D. Volcanic facies and their reservoir char-[4] acteristics in Eastern China Basins. J Earth Sci (Wuhan, China). 2013;24(6):935-46. doi: 10.1007/s12583-013-0380-8.
- Zhao W. Identification of the lithology of igneous rocks in central of the Junggar Basin. Nat Gas Ind. 2010;2(2):21-5. doi: 10.1016/S1876-3804(11)60008-6.

- Yujiao H, Chao Y, Yiren F, Xinmin G, Zhuoying F, Wenchao Y. Identification of igneous reservoir lithology based on empirical mode decomposition and energy entropy classification: a case study of Carboniferous igneous reservoir in Chunfeng oil field. Shíyóu Yǔ Tiānránqì Dìzhì. 2018;39(4):759-65. doi: 10.11743/
- Saeed U, Jan SU, Lee YD, Koo I. Fault diagnosis based on extremely randomized trees in wireless sensor networks. Reliab Eng Syst Saf. 2021;205:107284. doi: 10.1016/j.ress.2020.107284.
- [8] Onan A, KorukoGlu S. A feature selection model based on genetic rank aggregation for text sentiment classification. J Inf Sci. 2017;43(1):25-38. doi: 10.1177/0165551515613226.
- Yan D, Yasin Q, Cui M. A novel neural network for seismic anisotropy and fracture porosity measurements in carbonate reservoirs. Arab J Sci Eng. 2021;46:1-23. doi: 10.1007/s13369-021-05970-4.
- [10] Yasin Q, Yan D, Ismail A, Du Q. Estimation of petrophysical parameters from seismic inversion by combining particle swarm optimization and multilayer linear calculator. Nat Resour Res. 2020;29:3291-317. doi: 10.1007/s11053-020-09641-3.
- [11] Onan A. Classifier and feature set ensembles for web page classification. J Inf Sci. 2016;42(2):150-65. doi: 10.1177/ 0165551515591724.
- [12] Onan A. Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. IEEE Access. 2019;7:145614-33. doi: 10.1109/ACCESS.2019.2945911.
- [13] Onan A. An ensemble scheme based on language function analysis and feature engineering for text genre classification. J Inf Sci. 2018;44(1):28-47. doi: 10.1177/0165551516677911.
- [14] Onan A, Tocoglu MA. A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. IEEE Access. 2021;9:1-56. doi: 10.1109/ ACCESS.2021.3049734.
- [15] Tsoupos A, Khadkikar V. A novel SVM technique with enhanced output voltage quality for indirect matrix converters. IEEE Trans Ind Electr(1982). 2019;66(2):832-41. doi: 10.1109/ TIE.2018.2835404.
- [16] Al-Anazi A, Gates ID. On the capability of support vector machines to classify lithology from well logs. Nat Resour Res (N York, NY). 2010;19(2):125-39. doi: 10.1007/s11053-010-9118-9.
- [17] Hsu C, Lin C. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw. 2002;13(2):415-25. doi: 10.1109/72.991427.
- [18] Opitz D, Maclin R. Popular ensemble methods: an empirical study. J Artif Intell Res. 1999;11:169-98. doi: 10.1613/jair.614.
- [19] Breiman L. Bagging predictors. Mach Learn. 1996;24(2):123-40. doi: 10.1007/BF00058655.
- [20] Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett. 1999;9(3):293-300. doi: 10.1023/A:1018628609742.

- [21] Nigam K, Mccallum AK, Thrun S, Mitchell T. Text classification from labeled and unlabeled documents using EM. Mach Learn. 2000;39(2):103-34. doi: 10.1023/A:1007692713085.
- [22] Abedi M, Norouzi G, Bahroudi A. Support vector machine for multi-classification of mineral prospectivity areas. Comput Geosci. 2012;46:272-83. doi: 10.1016/j.cageo.2011.12.014.
- [23] Nguyen-Sy T, To QD, Vu MN, Nguyen TD, Nguyen TT. Predicting the electrical conductivity of brine-saturated rocks using machine learning methods. J Appl Geophys. 2021;184:104238. doi: 10.1016/j.jappgeo.2020.104238.
- [24] Vapnik VN. An overview of statistical learning theory. IEEE Trans Neural Netw. 1999;10(5):988-99. doi: 10.1109/
- [25] Li X, Wang L, Wang J, Zhang X. Multi-focus image fusion algorithm based on multilevel morphological component analysis and support vector machine. IET Image Process. 2017;11(10):919-26. doi: 10.1049/iet-ipr.2016.0661.
- [26] Min X, Pengbo Q, Fengwei Z. Research and application of logging lithology identification for igneous reservoirs based on deep learning. J Appl Geophys. 2020;173:103929. doi: 10.1016/j.jappgeo.2019.103929.
- [27] Chen Y, Wang G, Dong S. Learning with progressive transductive support vector machine. Pattern Recognit Lett. 2003;24(12):1845-55. doi: 10.1016/S0167-8655(03)00008-4.
- [28] Yasin Q, Khalid P, Du Q. Application of machine learning tool to predict the porosity of clastic depositional system Indus Basin, Pakistan. J Pet Sci Eng. 2020;197:107975. doi: 10.1016/ j.petrol.2020.107975.
- [29] Qiang Z, Yasin Q, Du Q. Prediction of reservoir quality from logcore and seismic inversion analysis with an artificial neural network: a case study from the Sawan Gas Field, Pakistan. Energies. 2020;13(2):486. doi: 10.3390/en13020486.
- [30] Du Q, Yasin Q. Combining classification and regression for improving shear wave velocity estimation in a highly heterogeneous reservoir from well logs data. J Pet Sci Eng. 2019;182:106260. doi: 10.1016/j.petrol.2019.106260.
- [31] Feng Z. Volcanic rocks as prolific gas reservoir: a case study from the Qingshen gas field in the Songliao Basin, NE China. Mar Pet Geol. 2008;25(4):416-32. doi: 10.1016/ j.marpetgeo.2008.01.008.
- [32] Mao ZG, Zhu RK, Luo JL, Wang JH, Du ZH, Su L, et al. Reservoir characteristics, formation mechanisms and petroleum exploration potential of volcanic rocks in China. Pet Sci. 2015;12(1):54-66. doi: 10.1007/s12182-014-0013-6.
- [33] Liu YL, Saraf A, Catanese B, Lee SM, Zhang Y, Connolly EP, et al. A comparison of binary and multiclass support vector machine models for volcanic lithology estimation using geophysical log data from Liaohe Basin, China. Explor Geophys (Melb). 2016;47(2):145-9. doi: 10.1071/EG14114.