Machine-learning-based automatic well-log completion and generation: Examples from the Ordos Basin, China

Ziheng Guan[1], Xuan Tang[1], Bo Ran[2], Shaobin Guo[1], Jinchuan Zhang[1], Kefeng Du[1], and Ting Jia[3]

Abstract

Oilfields have large amounts of old well-logging data, some of which were possibly lost or distorted for borehole situation, limiting the use of well-logging in formation evaluation. Machine-learning algorithms provide possibility to complete or correct bad quality logging, even to generate new loggings. We took 50 wells in the Ordos Basin, a prolific hydrocarbon production basin, as an example to complete and generate well- loggings. We applied three algorithms, such as random forest (RF), extreme gradient boosting (XGBoost), and deep neural network (DNN) algorithm, for well-logging curve completion experiments. We generated resistivity loggings including deep investigate lateral resistivity log (RILD) and medium investigate lateral resistivity log (RILM) using four loggings, e.g. the spontaneous potential (SP), gamma ray (GR), acoustic log (AC), and electrical resistivity log (R4). After data preprocessing, we used training data sets and validation data sets, accounting for 90% and 10% of all database, respectively, to complete and generate well-logs. The results reveal that the XGBoost algorithm has a better effect on well-log completion if the parameters used are sufficiently optimized with experience, whereas the DNN algorithm has great advantages if large sufficient amounts of well-log data sets are available in the training sets. In this experiment, the accuracy of results by RF algorithm is better than those by XGBoost algorithm because the optimized parameters are difficult to guarantee without experience, and better than that, by DNN algorithms in which the input number of wells is less than 300 and may not be sufficient. In addition, RF algorithm has wider expansibility, higher efficiency, lower computation requirements, and better generalization ability. Our work provides a better understanding of the conditions and function of the application of different machine-learning algorithms to well-logging completion and generation.

Introduction

In recent years, many great breakthroughs in oil and gas exploration have been achieved in the Ordos Basin. Some large oil and gas fields have been discovered in the Carboniferous-Permian sandstone reservoirs (Liu et al., 2015; Wu et al., 2017), Lower Ordovician carbonate reservoirs (Liu et al., 2012), and Chang 7 shales (Tang et al., 2014; Li et al., 2019b; Liu et al., 2022). Well-logging provides important information for the reservoir evaluation (Chang et al., 1997; Avseth et al., 2013; Feng et al., 2018, 2021; Shan et al., 2021). With the development of unconventional oil and gas in the Ordos Basin, the demands for reservoir description and directional wells are increasing because of the large thickness of surface loess, significant landform height difference, low seismic resolution, and the high dependence of formation property analysis of well-logging. Under this circumstance, the richness and integrity of well logging play more and more important role in optimum well drilling, completion, and production in the oilfield (Ellis and Singer, 2007; Obiora et al., 2016). However, after 50 years of exploration and development, the Ordos Basin accumulates a large number of well-logging data, in which the deep investigate lateral resistivity log (RILD) and medium investigate lateral resistivity log (RILM) logging have not been well measured due to poor well completion quality or kept in every drilled well due to various reasons. In some cases, we need to recover or correct well-logging data in a fast and cheaper way. The usually used methods mainly include statistical fitting or calculation with theoretical formula according to the geologic and mechanic properties via certain physical models (Bateman, 1986; Asquith et al., 2004). Statistical fitting methods (Zhang et al., 2012; Yang et al., 2013; Luo et al., 2016) were previously used to predict unknown

[1]China University of Geosciences, Ministry of Natural Resources, Key Laboratory of Shale Gas Exploration and Evaluation, Beijing, China. E-mail: 949819850@qq.com; tangxuan@cugb.edu.cn (corresponding author); guosb58@126.com; zhangjc@cugb.edu.cn; 418883164@qq.com.
[2]Research Institute of Exploration and Development of Petro China Liaohe Oilfield Company, Panjin, China. E-mail: ranb@petrochina.com.cn.
[3]Wuji Data Technology (Beijing) Co. Ltd., Beijing, China. E-mail: 174592821@qq.com.

Manuscript received by the Editor 31 October 2021; revised manuscript received 20 March 2022; published ahead of production 13 May 2022; published online 20 July 2022. This paper appears in *Interpretation*, Vol. 10, No. 3 (August 2022); p. SJ91–SJ99, 6 FIGS., 1 TABLE.
http://dx.doi.org/10.1190/INT-2021-0228.1.

S-wave logging. Crossplot and histogram statistics combined with physical model were applied to obtain theoretical formula for logging data prediction (Li et al., 2014; Liu et al., 2016). However, manual identification of these properties and parameters requires professional knowledge and experience that might lead uncertainty (Theisges Dos Santos et al., 2021). Machine-learning algorithms provide a method to complete and predict well loggings. Artificial neural network (ANN) algorithm was usually used to generate new well logs using available loggings data (Rolon et al., 2009; Alizadeh et al., 2012; Mo et al., 2015; Long et al., 2016; Apurwa et al., 2021), and applied to porosity inversion through seismic attributes (Gholami and Ansari, 2017). The mean-square error was introduced into the objective function and proposed a logging curve prediction method based on generative adversarial network (Hu, 2020). Wang et al. (2020) use long and short-term memory (LSTM) network to predict predrilling well- logging, which has advantages in mining correlation information before and after time series. Full-connection deep neural network (DNN) model, characterized by strong nonlinear mapping ability and fast learning speed, also was used to generate logging curves (Zhang et al., 2020). Other work combined the standard LSTM with cascade system and proposed a cascade LSTM neural network to generate logging curves, which had higher prediction accuracy and less uncertainty (Zhang et al., 2018; Chen, 2020; Chen and Zhang, 2020; Yang et al., 2021a).

Lots of work have been done with well-logging completion and prediction using machine-learning algorithms (Wang et al., 2021; Yadav et al., 2021), most of them choose more layers of DNNs based on deep learning and their improved algorithms, which requires very large amounts of data for training to obtain reasonable algorithm parameters. But actually, we do not always have sufficient data for training, particularly in the Ordos Basin or other basins.

To test the availability of different machine-learning algorithms applied to the well- logging, extreme gradient boosting (XGBoost) and random forest (RF) methods were tried for synthetic well-log completion and generation compared with the result by DNN algorithm. In this experiment, RILD and RILM were generated by spontaneous potential (SP), gamma ray (GR), acoustic log (AC), and electrical resistivity log (R4), with RF, XGBoost, and DNN algorithms. The algorithm efficiency and suitability were discussed in this study.

Geologic settings

The Ordos Basin is located in Northwest China. The basin could be divided into six tectonic units, namely, Western Margin Thrust Belt, Tianhuan Depression, Yishan Slope, Jinxi Flexural Fold Belt, Weibei Uplift, and Yimeng Uplift. The research area, in the east part of Yimeng Uplift, is characterized by gentle angled strata and low inclination. The used logging curves in this study, only cover part of the Triassic Yan'an and Yanchang Formations, range from 1600 to 2200 m in depth, with a total of 600 m length.

Principle and methods

This experiment was based on the basic principle of well-logging theory, including AC, GR, R4, SP, RILM, and RILD, which were according to different layers in the formation period of lithology, physical property, and oiliness difference caused by the change of value. There were two sets of data which must be associated, while correlated with regular expression of mathematics and physics model. Therefore, the experiment attempted to use machine learning for finding the connection between two of them.

RF

The RF algorithm, as shown in Figure 1, is a combination prediction algorithm proposed by Breiman (1996, 2001), after bagging algorithm, which is based on decision tree. Multiple decision trees were constructed by random repeated sampling technology (Kushary, 2000) and node random splitting technology (He et al., 2020). To classify a new object from a new sample, we separated the sample vector from each of the trees in the forest. Each tree gives a classification, and the majority vote of trees in the forest is used to determine the final class. The forest chooses the classification having the most votes (Ai et al., 2014).

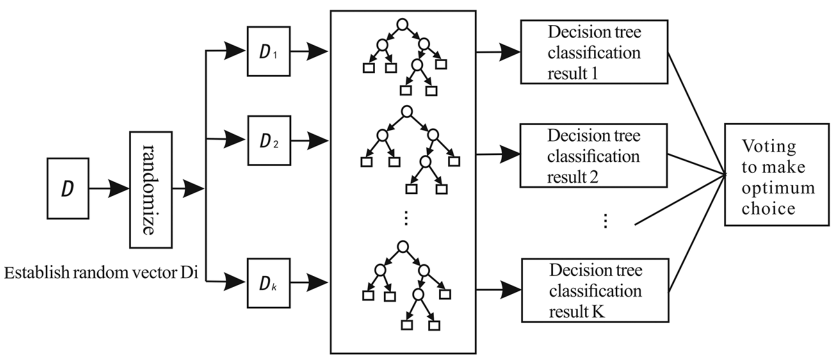

Figure 1. Concept model of RF algorithm.

DNN

DNN algorithm, as shown in Figure 2, is a kind of deep learning algorithm. The concept of deep learning has attracted wide attention since Hinton et al. (2006) proposed it in 2006. This algorithm can extract data features, and then the features are abstracted, which are represented in the form of hidden layers. Multilayer perceptron is obtained by using different kernel functions, and then learning outputs of unknown inputs are obtained (Moyano, 2017; Fu et al., 2018). Compared with traditional machine learning algorithms, DNN algorithm contains many hidden layers, which make up for the shortcomings of single-layer neural networks, and has very strong autonomous learning abilities to obtain essential characteristics of data (Yang et al., 2021b). But making full use of this model often requires higher dimensional and larger volumes of databases, which is not always having such large amounts of data available at hand in reality. The CNN algorithm also has limitations, such as slow converging, long-time computing, and easy overfitting while in the process of calculating (Chen et al., 2019). When running in the process of deep-level network, advanced graphics processing unit (GPU) are essential for training in a large amount of data within a reasonable time. Fast central processing unit (CPU), solid state disk storage, and large-capacity RAM also are required when such advanced GPUs are applied, which is not conducive to large-scale promotion.

XGBoost

XGBoost algorithm, as shown in Figure 3, was proposed by Chen and Guestrin (2016), which combines multiple weak classifiers into a strong classifier (Zhang et al., 2019). It is based on the gradient lifting decision tree algorithm, proposed by Friedman (2001), which consists of two parts: decision tree and gradient lifting (Li et al., 2019a). However, compared with the gradient lifting decision tree algorithm, XGBoost algorithm has higher prediction accuracy and stronger generalization performance. This algorithm can use CPU to carry out multithread parallel computing automatically, improve the computing speed and work efficiency (Shen et al., 2019; Sun et al., 2020). Because of its fast calculation speed, good model performance, excellent effect, and efficiency in practice it is highly respected by the academic and industrial circles (Chen et al., 2019) and used in some areas of the petroleum industry (Gumus and Kiran, 2017; Yan et al., 2019; Chen et al., 2020, 2021; Meng et al., 2020; Mousavi et al., 2020; Zhong et al., 2020a, 2020b; Gu et al., 2021) at present. However, the XGBoost algorithm without parameter optimization has a low fit with the existing dataset, which leads to its poor generalization performance and adaptability.

In the practical well-logging curve prediction job, the algorithms that could be popularized in oilfields need to have higher prediction accuracy, faster working efficiency, and lower cost. Compared with deep learning, machine-learning algorithms have the characteristics of lower hardware cost and faster running speed. Compared with RF algorithm, XGBoost algorithm is more accurate and easy to overfit, requiring more parameter settings. If the parameters are not properly set, the accuracy advantage of XGBoost algorithm cannot be fully taken, and even the accuracy is not as good as that of RF. Therefore, it is not suitable for XGBoost algorithm to be widespread promoted. RF algorithm has the characteristics of no easy overfitting, fewer parameter adjustments, and stronger generalization ability, which make it easier in practical application. Next, three algorithm models were chosen for well-logging curve completion experiment, then the effects of operating results were evaluated, and some suggestions were made in the end.

Experiment

To compare the effects of RF, XGBoost, and DNN algorithms on well-logging curve completion, this research used 10 exploration wells and 50 production wells with completed well-log curve as well-log data sets and applied three algorithms in the well-logging curve completion experiment. Then, this research compared and discussed the results of this experiment.

Data source

Six conventional logging curves of 10 exploration wells and 50 oil wells were chosen for this experiment, including GR, SP, AC, R4, RILD, and RILM. The sounding depth of target layer ranges from 1600 to 2200 m, a total of 29,160 well-log data sets in a single well. Because RILD and RILM appeared later than GR, SP, AC, and R4, some old wells lacked these two logs. Four ancient well logs, which are SP, GR, AC, and R4, are used for predicting the RILD and RILM well logs. In the experiment, to verify the curve generation effect, the well-log data of 2000–2050 m were used as the verification set to simulate resistivity logging and as the verification algorithm effect, and the remaining data were label sets. In well-logging

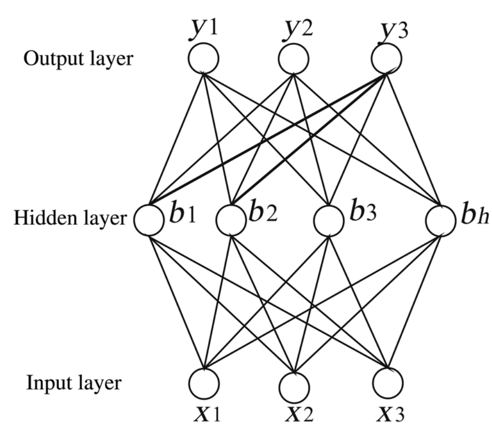

Figure 2. Concept model of DNN algorithm.

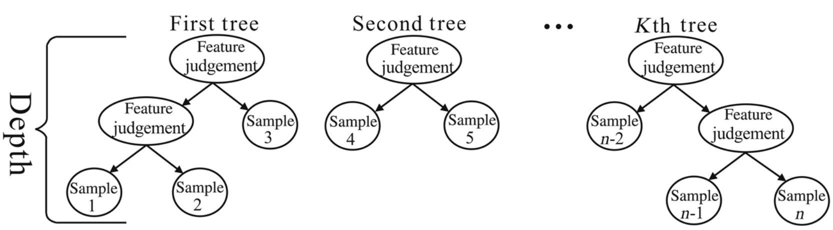

Figure 3. Concept model of XGBoost algorithm.

completion, four logging data, GR, SP, R4, and AC, were applied as input, and two logging data, RILD and RILM, were applied as output.

Experimental process and result

During the experiment, the MaxAI platform of Wuji Data Technology (Beijing) Co. Ltd., was implemented to establish the algorithm model for well-logging curve completion and evaluation. The operation steps were as follows. First, we filter the well-log curves and remove singular values, as shown in Figure 4. Then, the algorithm model was established after the label set and validation set being input. Finally, the algorithm model was evaluated, completed, and predicted according to the validation set.

It was essential to opt for the proper parameter of model, which might reflect on the results in the end. To avoid overfitting, 26,910 well-log data sets in a single well were used for training, and 2400 well-log data sets were used for testing. The Gaussian distribution and 50 trees were chosen in the distribution function and the number of trees of the RF algorithm. In the XGBoost algorithm, it turned out that the model did not appear overfitting when the number of trees was six after some attempts. Therefore, the Gaussian distribution and six trees were chosen in XGBoost algorithm. The DNN algorithm used Tanh as the activation function, Huber as the loss function, the distribution function as Gaussian, and the hidden layer size as [200, 200]. Feature importance was a technique that assigns scores to the input features in prediction model and represents the relative importance of each feature when making predictions. It played an important role in the prediction model. First, it provided the insight into the data, highlighted the most important relevant features, and was used as the basis for collecting more data. Second, it provided the insight into the model and made the machine understood the most important features of the model during predicting. Third, it provided the foundation for dimension reduction and feature selection, thereby improving the efficiency and effectiveness of the prediction model. This paper scored and compared the importance of features, and then selected the chosen well-logs with strong correlation between RILD and RILM predicting for experiment, simplifies modeling problems, speeds up the modeling process, and achieves dimension reduction. According to the result of feature importance analysis, as shown in Figure 5, well logs with an importance of more than 50% were regarded as the high-correlation well logs and selected for experiment. The results showed that the importance of R4, DT, and SP in the RILD and RILM logs predictions of the three algorithm models is higher than that of GR. Among them, DT, SP, and R4 were more closely related to RILD and RILM in the RF algorithm model. The high degree importance in the XGBoost algorithm model was R4. And, in the DNN algorithm, they were SP, R4, and DT in order.

Table 1. Evaluation indicators of different models.

	RF		DNN		XGBoost	
	RILD	RILM	RILD	RILM	RILD	RILM
R^2	0.810	0.580	0.306	0.474	0.711	0.484
RE	0.352	0.223	0.646	0.361	0.402	0.257

Figure 6 shows the result of RILD and the RILM curves' comparison between the original and the generated curve predicted by RF, XGBoost, and DNN algorithms of the research well in the target zone (2000–2050 m). The dotted curve is the hidden original one, and the solid curve is the output result of these models. It is clear that the well-logging curves generated by three models can capture the true curve trend, but the logging curve predicted by different models has different fitness, and the same model has different predicted accuracies for RILD and RILM.

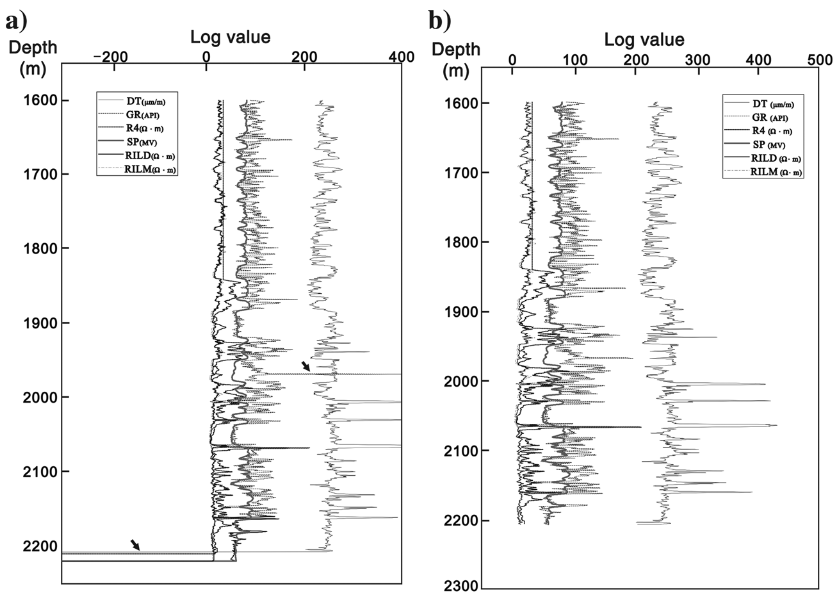

Figure 4. Comparison chart of well-logging curves (a) before and (b) after data cleaning, The arrow points to the cleaning section.

Results comparison

To reflect the fitness degree of different models quantitatively, the determination coefficient (R^2) and relative error (RE) were selected as the evaluation indexes of the model, as shown in Table 1. The R^2 reflected the fitness degree of the predicted value to the real value. The closer the value was to one, the better the fitting degree. The RE was the percentage of the absolute difference between the measured value and the real value. The smaller the RE was, the higher the prediction accuracy.

Combined with the prediction and original values comparison diagram and the model evaluation index table, it could be seen that RF and XGBoost algorithms had better performance in high and low values in RILD prediction, whereas DNN algorithm had worse performance in high values than the other two algorithms. The performance of three algorithms in high value prediction of RILM was poor, and the analysis reason might be that the RILM high-value training samples were few and the prediction effect was insufficient. In RILD and RILM, RF and XGBoost algorithms had more advantages than DNN algorithm in detailed coincidence degree. In RILD prediction, R^2 (RF) > R^2 (XGBoost) > R^2 (DNN), it could be seen that the effect of RF algorithm was slightly better than that of XGBoost algorithm, and XGBoost algorithm was better than that of DNN algorithm, RE (DNN) > RE (XGBoost) > RE (RF) was consistent with R^2. In RILM prediction, R^2 (RF) > R^2 (XGBoost) > R^2 (DNN), it could be seen that the RF algorithm was superior to the XGBoost algorithm, and XGBoost algorithm was superior to the DNN algorithm, RE (DNN) > RE (XGBoost) > RE (RF), whose result was consistent with R^2.

The fitness of XGBoost algorithm results changed greatly with parameters varying. For example, when the eta value was 0.2 and the max depth value was seven, the fitness of RILD well-log obtained by XGBoost algorithm was very high, R^2 (XGBoost) > R^2 (RF) > R^2 (DNN), suggesting that when the parameters were sufficiently optimized, the prediction accuracy of XGBoost algorithm would be higher than that of RF algorithm.

With the increase of well-logs in training sets, the prediction accuracy of RILD well-logs obtained from DNN algorithm improves a lot. For example, as well number in training data sets increases from 60 to 300, R^2 of DNN algorithm changes from 0.306 to 0.839. The well number of training sets has a great influence on the prediction accuracy of DNN algorithm. The more the training sets are, the higher the prediction accuracy.

Based on the comparison and analysis of the preceding well-log completion and generation experiment, results could be made next. When the parameters were sufficiently optimized, XGBoost algorithm had better completion effect. When large amounts of well-log data sets were put into the training sets, DNN algorithm had great advantages on well-log completion. When the number of wells was less than 300 and simpler parameter adjustment was considered, RF algorithm had better prediction effect than XGBoost algorithm and DNN algorithm in the well-log completion and generation experiment.

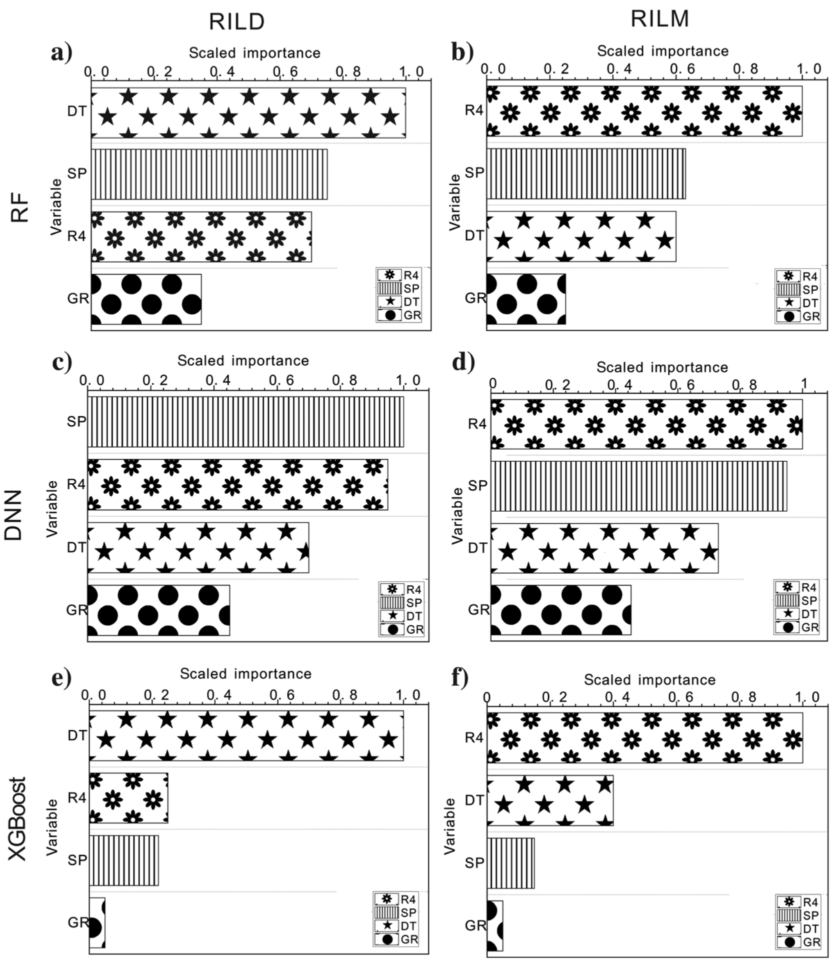

Figure 5. Feature importance comparison bar chart. (a and b) The RF, (c and d) DNN, and (e and f) XGBoost.

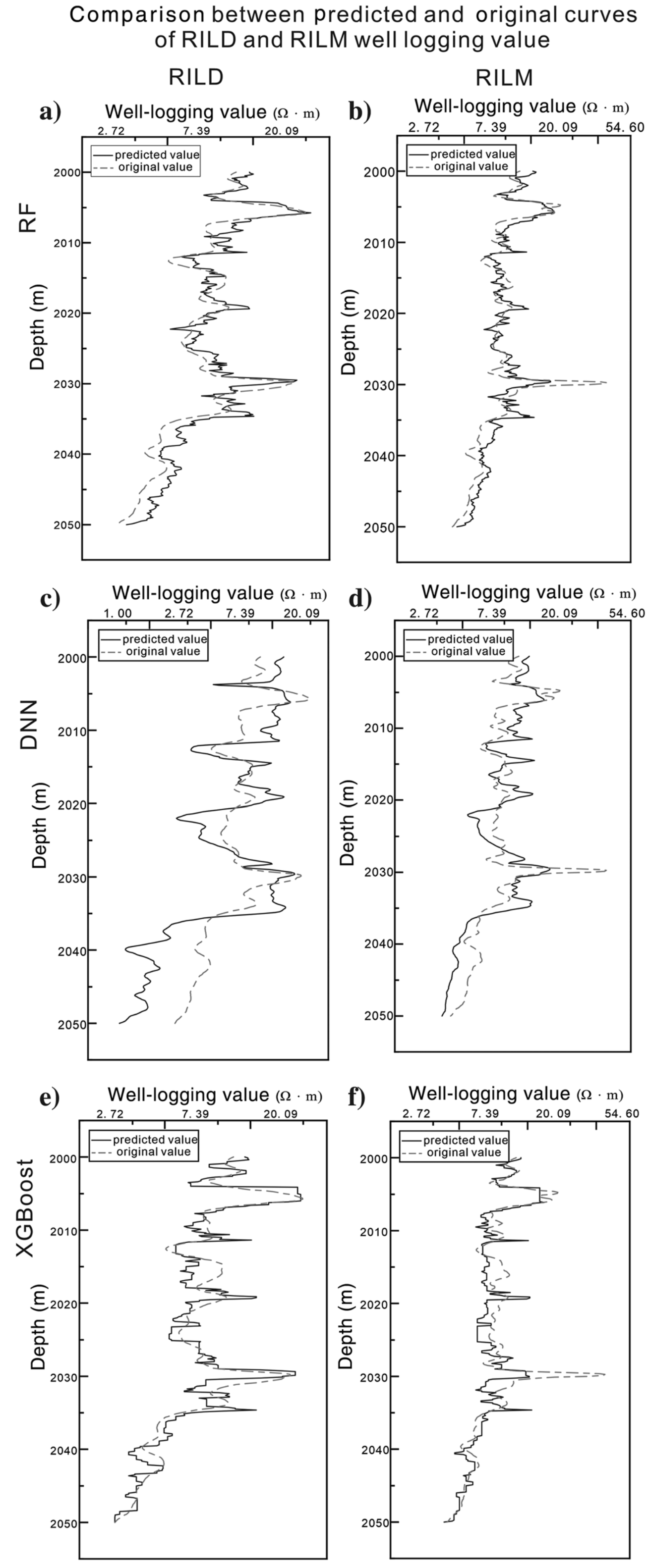

Figure 6. Comparison of the predicted and original values of (a, c, and e) RILD and (b, d, and f) RILM well-logging curves by (a and b) RF, (c and d) DNN, and (e and f) XGBoost.

Conclusion and recommendation

In this study, 10 exploration wells and 50 oil wells were taken as the research examples in a block of the Ordos Basin in the experiment. This paper initiated the use of XGBoost- and RF-based ensemble learning methods, for prediction of the lost well-logs. Two advanced tree-based machine-learning algorithms (XGBoost and RF) and DNN were applied on these well-log data sets. The unknown RILD and RILM were predicted by the four well-logs what had been known before, GR, AC, R4, and SP, and some important conclusions were drawn.

Well-logging, generated by the three algorithms of RF, DNN, and XGBoost, could capture the true curve trend, but the fitness of the predicted well-logging curves generated by different algorithms was different. At the same time, the logging curve fitness was different when the same algorithm predicts different types of well logs. In this experiment, the RF algorithm had greater advantages than XGBoost trees and DNNs. In the comparison of feature importance, the importance of R4 and AC in RILD and RILM logging prediction of the three algorithms was higher than GR, and they were the best choice for predicting RILD and RILM.

The average time spent for duty by each algorithm was used as criteria for comparison of running speed, which were 0.46 s for RF, 0.68 s for XGBoost, and 1.76 s for DNN. Although the gaps of running time between different algorithms were not huge, the efficiency of the algorithm would be significant when thousands of well-logging data needed to be operated in the practical oilfield. Therefore, RF took the least time for well-logging prediction.

When performing single-well-logging curve prediction, the prediction accuracy of shallow machine-learning algorithms, such as RF and XGBoost trees, was higher than that of deep learning represented by DNNs significantly. When choosing the algorithm that could be applied to the practical oilfield, the complexity of the algorithm, which was to improve accuracy, should not be pursued excessively.

In the comparison of RILD and RILM well-logging prediction, we find that the prediction accuracy of RILD was generally better than that of RILM. In this RILD and RILM well-logging prediction experiment, the RF algorithm was the optimal one. According to the analysis of the characteristics of RF and XGBoost algorithm, RF algorithm was not easy to overfit and less sensitive to noise data, and the tree-based bagging algorithm did not need to be normalized and has strong adaptability. XGBoost algorithm needed to select multiple parameters during the use process. If the parameters were not set properly, it was difficult to achieve the best effect of the algorithm. The overfitting phenomenon caused by the pursuit of accuracy made the algorithm insufficient in generalization ability relative to the RF, which made it difficult to scale up.

The RF algorithm needed smaller amounts of memory, computation, fewer setting parameters, strong generalization ability, and high accuracy compared with DNN and XGBoost algorithm, which made RF algorithm more suitable for practical application in well-log curve completion and generation as accuracy, generalization, cost, and parameter adjustment needed to be considered comprehensively.

Suggestions for further study

At present, the application of artificial intelligence algorithm in well-logging prediction is still at the exploratory experimental stage, and there is no large number of oilfield production. How to integrate the intelligent algorithm in the laboratory into software to improve work efficiency and reduce production cost is the future issue that needs to be worked out. In this experiment, RF, XGBoost, and DNN algorithms reveal that they reflect the well-log curve trend, but the predicted accuracy needs to be improved. With the continuous optimization of machine learning and deep learning algorithm, the algorithm more suitable for practical oilfield production still needs continuous exploration.

Due to the characteristics of geologic data itself, the application of artificial intelligence in the oil and gas geology field is faced with many difficulties such as small-size sample tag and strong multisolution. Therefore, in the intelligent exploration of oil and gas exploration and development, it is necessary to consider the algorithm model more in line with the practical oilfield. In the future, machine-learning algorithms represented by shallow machine learning and deep learning will shine in the fields of stratigraphic division, sedimentary facies recognition, interlayer analysis, reservoir architecture analysis, and scanning electron microscope image recognition.

Acknowledgments

Thanks to the Wuji Data Technology (Beijing) Co. Ltd., for providing the MaxAI platform and data process tutoring. This work was supported by the National Natural Science Foundation of China (grant nos. 41730421 and 41972132) and the Fundamental Research Funds for the Central Universities. We thank Z. Zheng at Brandeis University for his help with English and constructive discussions.

Data and materials availability

Data associated with this research are confidential and cannot be released.

References

Ai, F., J. Bin, Z. Zhang, J. Huang, J. Wang, Y. Liang, L. Yu, and Z. Yang, 2014, Application of random forests to select premium quality vegetable oils by their fatty acid composition: Food Chemistry, **143**, 472–478, doi: 10.1016/j.foodchem.2013.08.013.

Alizadeh, B., S. Najjari, and A. Kadkhodaie-Ilkhchi, 2012, Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the South Pars Gas Field, Persian Gulf, Iran: Computers & Geosciences, **45**, 261–269, doi: 10.1016/j.cageo.2011.11.024.

Apurwa, Y., Y. Kriti, and S. Anirbid, 2021, Feedforward neural network for joint inversion of geophysical data to identify geothermal sweet spots in Gandhar, Gujarat, India: Energy Geoscience, **2**, 189–200, doi: 10.1016/j.engeos.2021.01.001.

Asquith, G. B., D. Krygowski, and C. R. Gibson, 2004, Basic well log analysis: AAPG.

Avseth, P., N. Skjei, and A. Skalnes, 2013, Rock physics modelling of 4D time-shifts and time-shift derivatives using well log data — A North Sea demonstration: Geophysical Prospecting, **61**, 380–390, doi: 10.1111/j.1365-2478.2012.01134.x.

Bateman, R. M., 1986, Openhole log analysis and formation evaluation: SPE.

Breiman, L., 1996, Bagging predictors: Machine Learning, **24**, 123–140, doi: 10.1007/BF00058655.

Breiman, L., 2001, Random forests: Machine Learning, **45**, 5–32, doi: 10.1023/A:1010933404324.

Chang, H., H. Chen, and J. Fang, 1997, Lithology determination from well logs with fuzzy associative memory neural network: IEEE Transactions on Geoscience and Remote Sensing, **35**, 773–780, doi: 10.1109/36.582000.

Chen, J., J. Yang, and Z. Lou, 2019, Research on a new short-term load forecasting model based on XGBoost algorithm: Electrical Measurement & Instrumentation, **56**, 23–29.

Chen, J., F. Zhao, Y. Sun, and Y. Yin, 2020, Improved XGBoost model based on genetic algorithm: International Journal of Computer Applications in Technology, **62**, 240–245, doi: 10.1504/IJCAT.2020.106571.

Chen, L., X. Gao, and X. Li, 2021, Using the motor power and XGBoost to diagnose working states of a sucker rod pump: Journal of Petroleum Science and Engineering, **199**, 108329, doi: 10.1016/j.petrol.2020.108329.

Chen, T., and C. Guestrin, 2016, XGBoost: A scalable tree boosting system: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Chen, Y., 2020, Research on well log completion and generation based on machine learning: Ph.D. thesis, Peking University.

Chen, Y., and D. Zhang, 2020, Well log generation via ensemble long short-term memory (EnLSTM) network: Geophysical Research Letters, **47**, e2020GL087685, doi: 10.1029/2020GL087685.

Ellis, D. V., and J. M. Singer, 2007, Well logging for earth scientists: Springer.

Feng, R., D. Grana, and N. Balling, 2021, Imputation of missing well log data by random forest and its uncertainty analysis: Computers & Geosciences, **152**, 104763, doi: 10.1016/j.cageo.2021.104763.

Feng, R., S. M. Luthi, D. Gisolf, and E. Angerer, 2018, Reservoir lithology determination by hidden Markov random fields based on a Gaussian mixture model: IEEE Transactions on Geoscience and Remote Sensing, **56**, 6663–6673, doi: 10.1109/TGRS.2018.2841059.

Friedman, J. H., 2001, Greedy function approximation: A gradient boosting machine: The Annals of Statistics, **29**, 1189–1232, doi: 10.1214/aos/1013203451.

Fu, C., N. Lin, D. Zhang, B. Wen, Q. Wei, and K. Zhang, 2018, Prediction of reservoirs using multi-component seismic data and the deep learning method: Chinese Journal of Geophysics, **61**, 293–303, doi: 10.6038/cjg2018L0193.

Gholami, A., and H. R. Ansari, 2017, Estimation of porosity from seismic attributes using a committee model with bat-inspired optimization algorithm: Journal of Petroleum Science and Engineering, **152**, 238–249, doi: 10.1016/j.petrol.2017.03.013.

Gu, Y. F., D. Y. Zhang, and Z. D. Bao, 2021, Lithology identification in tight sandstone reservoirs using CRBM-PSO-XGBoost: Oil & Gas Geology, **42**, 1210–1222, doi: 10.11743/ogg20210518.

Gumus, M., and M. S. Kiran, 2017, Crude oil price forecasting using XGBoost: International Conference on Computer Science and Engineering, 1100–1103.

He, J., X. Wen, W. Nie, L. Li, and J. Yang, 2020, Prediction of fracture zone by random forest algorithm: Geophysical Petroleum Exploration, **55**, 161–166.

Hinton, G. E., S. Osindero, and Y. Teh, 2006, A fast learning algorithm for deep belief nets: Neural Computation, **18**, 1527–1554, doi: 10.1162/neco.2006.18.7.1527.

Hu, J., 2020, Logging curve prediction and reservoir identification based on deep learning: M.S. thesis, Shaanxi University of Science and Technology.

Kushary, D., 2000, Bootstrap methods and their application: Technometrics, **42**, 216–217, doi: 10.1080/00401706.2000.10486018.

Li, C., W. Zhang, and J. Lin, 2019a, Study on star/galaxy classification based on XGBoost algorithm: Chinese Astronomy and Astrophysics, **43**, 539–548, doi: 10.1016/j.chinastron.2019.11.005.

Li, P., C. Jia, and Z. Jin, 2019b, The characteristics of movable fluid in the Triassic lacustrine tight oil reservoir: A case study of the Chang 7 member of Xin'anbian Block, Ordos Basin, China: Marine and Petroleum Geology, **102**, 126–137, doi: 10.1016/j.marpetgeo.2018.11.019.

Li, W., C. Peng, and H. Yang, 2014, The application of S-wave prediction technology in YB area: Progress in Geophysics, **4**, 1695–1700, doi: 10.6038/pg20140427.

Liu, C. Y., H. R. Zheng, and Z. Q. Hu, 2012, Characteristics of carbonate cementation in clastic rocks from the Chang 6 sandbody of Yanchang Formation, southern Ordos Basin: Science China Earth Sciences, **55**, 58–66, doi: 10.1007/s11430-011-4352-5.

Liu, Q., Z. Jin, Q. Meng, X. Wu, and H. Jia, 2015, Genetic types of natural gas and filling patterns in Daniudi gas field, Ordos Basin, China: Journal of Asian Earth Sciences, **107**, 1–11, doi: 10.1016/j.jseaes.2015.04.001.

Liu, Q., P. Li, Z. Jin, Y. Sun, G. Hu, D. Zhu, Z. Huang, X. Liang, R. Zhang, and J. Liu, 2022, Organic-rich formation and hydrocarbon enrichment of lacustrine shale strata: A case study of Chang 7 Member: Science China Earth Sciences, **65**, 118–138, doi: 10.1007/s11430-021-9819-y.

Liu, Y., S. Li, Y. Wang, and Y. Xia, 2016, Reservoir prediction based on shear wave in Sulige gas field: Oil Geophysical Prospecting, **51**, 165–172, doi: 10.13810/j.cnki.issn.1000-7210.2016.01.021.

Long, W., D. Chai, and F. Aminzadeh, 2016, Pseudo density log generation using artificial neural network: Presented at the Western Regional Meeting, SPE.

Luo, S., P. Yang, G. Hu, and S. Liu, 2016, S-wave velocity prediction based on the modified P-L model and matrix equation iteration: Chinese Journal of Geophysics, **59**, 1839–1848, doi: 10.6038/cjg20160527.

Meng, M., R. Zhong, and Z. Wei, 2020, Prediction of methane adsorption in shale: Classical models and machine learning based models: Fuel, **278**, 118358, doi: 10.1016/j.fuel.2020.118358.

Mo, X., Q. Zhang, and X. Li, 2015, Well logging curve reconstruction based on genetic neural networks: 12th International Conference on Fuzzy Systems and Knowledge Discovery, 1015–1021.

Mousavi, S. M., H. Jabbari, M. Darab, M. Nourani, and S. Sadeghnejad, 2020, Optimal well placement using machine learning methods: Multiple reservoir scenarios: Presented at the Norway Subsurface Conference, SPE.

Moyano, L. G., 2017, Learning network representations: European Physical Journal-Special Topics, **226**, 499–518, doi: 10.1140/epjst/e2016-60266-2.

Obiora, D. N., D. Gbenga, and G. Ogobiri, 2016, Reservoir characterization and formation evaluation of a “Royal onshore field”, Southern Niger Delta using geophysical well log data: Journal of the Geological Society of India, **87**, 591–600, doi: 10.1007/s12594-016-0433-6.

Rolon, L., S. D. Mohaghegh, S. Ameri, R. Gaskari, and B. McDaniel, 2009, Using artificial neural networks to generate synthetic well logs: Journal of Natural Gas Science and Engineering, **1**, 118–133, doi: 10.1016/j.jngse.2009.08.003.

Shan, L., Y. Liu, M. Tang, M. Yang, and X. Bai, 2021, CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction: Journal of Petroleum Science and Engineering, **205**, 108838, doi: 10.1016/j.petrol.2021.108838.

Shen, Q., F. Shao, and R. Sun, 2019, Prediction model of breast cancer based on XGBoost: Journal of Qingdao University (Nature Science Edition), **32**, 95–100.

Sun, Y., Y. Huang, T. Liang, H. Ji, P. Xiang, and X. Xu, 2020, Identification of complex carbonate lithology by logging based on XGBoost algorithm: Lithologic Reservoirs, **32**, 98–106, doi: 10.12108/yxyqc.20200410.

Tang, X., J. Zhang, X. Wang, B. Yu, W. Ding, J. Xiong, Y. Yang, L. Wang, and C. Yang, 2014, Shale characteristics in the southeastern Ordos Basin, China: Implications for hydrocarbon accumulation conditions and the potential of continental shales: International Journal of Coal Geology, **128**, 32–46, doi: 10.1016/j.coal.2014.03.005.

Theisges Dos Santos, D., M. Roisenberg, and M. Dos Santos Nascimento, 2021, Deep recurrent neural networks approach to sedimentary facies classification using well logs: IEEE Geoscience and Remote Sensing Letters, **19**, 1–5, doi: 10.1109/LGRS.2021.3053383.

Wang, J., J. Cao, Z. Liu, X. Zhou, and X. Lei, 2020, Method of well logging prediction prior to well drilling based on long short-term memory recurrent neural network: Journal of Chengdu University of Technology, **47**, 227–236.

Wang, J., J. Cao, J. You, M. Cheng, and P. Zhou, 2021, A method for well log data generation based on a spatio-temporal neural network: Journal of Geophysics and Engineering, **18**, 700–711, doi: 10.1093/jge/gxab046.

Wu, X., Q. Liu, J. Zhu, K. Li, G. Liu, and Y. Chen, 2017, Geochemical characteristics of tight gas and gas-source correlation in the Daniudi gas field, the Ordos Basin, China: Marine and Petroleum Geology, **79**, 412–425, doi: 10.1016/j.marpetgeo.2016.10.022.

Yadav, A., K. Yadav, and A. Sircar, 2021, Feedforward neural network for joint inversion of geophysical data to identify geothermal sweet spots in Gandhar, Gujarat, India: Energy Geoscience, **2**, 189–200, doi: 10.1016/j.engeos.2021.01.001.

Yan, X., H. Gu, Y. Xiao, H. Ren, and J. Ni 2019, XGBoost algorithm applied in the interpretation of tight-sand gas reservoir on well logging data: Oil Geophysical Prospecting, **54**, 447–455.

Yang, J., Y. Chen, and C. Jiang, 2021a, Machine learning modeling paradigm for log curve generation problems: Taking shale gas wells in Changning-Weiyuan area as examples: China Offshore Oil and Gas, **33**, 76–84.

Yang, J., N. Lin, K. Zhang, C. Zhang, C. Fu, G. Tian, and C. Song, 2021b, Reservoir characterization using multicomponent seismic data in a novel hybrid model based on clustering and deep neural network: Natural Resources Research, **30**, 3429–3454, doi: 10.1007/s11053-021-09863-z.

Yang, Y., C. Huang, and Y. Wang, 2013, Study of shear wave prediction for Xujiahe formations in Q area of West Sichuan Basin: Site Investigation Science and Technology, **2**, 49–52.

Zhang, B., D. Luo, P. Hu, J. Fan, and C. Jing, 2020, A well log curve generation method based on deep neural network: Electronic Measurement Technology, **43**, 107–111.

Zhang, D., Y. Chen, and J. Meng, 2018, Synthetic well logs generation via recurrent neural networks: Petroleum Exploration and Development, **45**, 629–639, doi: 10.1016/S1876-3804(18)30068-5.

Zhang, J., H. Li, and F. Yao, 2012, Rock critical porosity inversion and S-wave velocity prediction: Applied Geophysics, **9**, 57–64, doi: 10.1007/s11770-012-0314-8.

Zhang, J., D. Mucs, U. Norinder, and F. Svensson, 2019, LightGBM: An effective and scalable algorithm for prediction of chemical toxicity — Application to the Tox21 and mutagenicity data sets: Journal of Chemical Information and Modeling, **59**, 4150–4158, doi: 10.1021/acs.jcim.9b00633.

Zhong, R., R. Johnson, Jr., and Z. Chen, 2020a, Generating pseudo density log from drilling and logging-while-drilling data using extreme gradient boosting (XGBoost): International Journal of Coal Geology, **220**, 103416, doi: 10.1016/j.coal.2020.103416.

Zhong, R., R. L. Johnson, and Z. Chen, 2020b, Using machine learning methods to identify coal pay zones from drilling and logging-while-drilling (LWD) data: SPE Journal, **25**, 1241–1258, doi: 10.2118/198288-PA.

Biographies and photographs of the authors are not available.