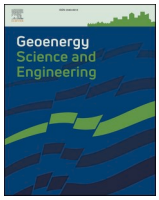
230 (2023) 212165

Contents lists available at ScienceDirect

Geoenergy Science and Engineering

journal homepage: www.sciencedirect.com/journal/geoenergy-science-and-engineering



Fracture identification in reservoirs using well log data by window sliding recurrent neural network

Shaoqun Dong [a,b], Leting Wang [a,b], Lianbo Zeng [a,c,*], Xiangyi Du [a,d], Chunqiu Ji [a,c], Jingru Hao [a,b], Xu Yang [a,b], Haiming Li [e]

[a] State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing, 102249, China
[b] College of Science, China University of Petroleum, Beijing, 102249, China
[c] College of Geoscience, China University of Petroleum, Beijing, 102249, China
[d] Bohai Oilfield Research Institute, Tianjin Branch of CNOOC China Co., Ltd., Tianjin, 300459, China
[e] PetroChina Tarim Oilfield Company, Korla, 841000, China

ARTICLE INFO

Keywords:
Fracture identification
Recurrent neural network
Well logs
Window sliding
Carbonate reservoir

ABSTRACT

Detecting fractures using well logs can be difficult due to the complex response of conventional logs. To address this issue, a novel method called **F**racture **I**dentification by window **s**liding and **r**ecurrent neural network (FIsr) is proposed. FIsr uses window sliding to generate sequence image data for training a bidirectional recurrent neural network (BiLSTM) classifier, with columns selected from both conventional and reconstructed logs. Undersampling is applied to balance the data, as the number of fracture samples is much smaller than nonfracture samples. BiLSTM extracts features from the sequence data in two directions, considering label correlations and detecting local log anomalies caused by fractures. The prediction for each sample is based on multiple overlapping sequence images to reduce uncertainties. The proposed method is validated using a dataset from carbonate reservoirs of the Asmari Formation in the Middle East, with an accuracy of 95% and recall and precision metrics exceeding 90%. A blind well test shows that FIsr can detect all fracture zones, and the distribution of fractures along the well trajectory confirms previous knowledge of the area. The study also discusses the influence of factors in FIsr.

1. Introduction

Fractures are not only reservoir space but also crucial seepage channels for oil and gas in carbonate reservoirs (Laubach et al., 2019). Identification of fractures in carbonate reservoirs holds paramount significance in the domain of oil exploration and developments (Rashid et al., 2020; Samarkin et al., 2021).

The task of identifying fractures along a well trajectory poses challenges due to the limited availability of valuable data, such as rock cores, borehole imaging logs and so on. These techniques, although effective, are costly, resulting in a scarcity of such data for numerous wells. Consequently, understanding the distribution of a subsurface fracture network system becomes increasingly difficult (Dong et al., 2020a). Nevertheless, there is a silver lining as conventional logs, which are readily accessible in nearly every well, provide an alternative source of information (Nouri-Taleghani et al., 2015; Yuan et al., 2021). Meanwhile, fractures in subsurface reservoirs can induce anomalies in geophysical properties, such as increased density (DEN), great separation between deep resistivity (RD) and shallow resistivity (RS), and elevated acoustic time difference (AC or DT), and so on (Dong et al., 2020b; Mazhari et al., 2018). Hence, the utilization of conventional logs offers a practical approach to comprehending the distribution of subsurface fractures. However, it is imperative to acknowledge that geophysical anomalies observed in these logs can be attributed to various other factors, including fluid types, lithological variations, and more (Dong et al., 2020c). For example, the presence of a gas-bearing reservoir can potentially result in a high amplitude of AC (Aghli et al., 2020). Furthermore, the high heterogeneity of a reservoir can lead to logging responses that make it different to distinguish fractures from other geological factors (Shalaby, Islam, 2017). Such challenges pose a significant hindrance to the identification of fractures (Gamal et al., 2022).

* Corresponding author. tate Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing, 102249, China.
E-mail addresses: dshaoqun@163.com (S. Dong), wang_leting@126.com (L. Wang), lbzeng@sina.com (L. Zeng), dxycup@126.com (X. Du), 2289343691@qq.com (C. Ji), 3112902624@qq.com (J. Hao), yangxucup@163.com (X. Yang), lihaim-tlm@petrochina.com.cn (H. Li).

https://doi.org/10.1016/j.geoen.2023.212165
Received 6 April 2023; Received in revised form 30 June 2023; Accepted 17 July 2023
Available online 23 July 2023

The advancement of artificial intelligence (AI) technology has brought about a remarkable revolution in handling geoscience problems (Desouky et al., 2021a; Liu, Wang, 2022). Recently, AI algorithms (e.g., deep artificial neural networks) have shown significant potentials to enhance the accuracy of fracture identification through the deep mining of fracture identification information within logging data (Bahramali Asadi Kelishami et al., 2022; Tariq et al., 2022a). Commonly used AI methods include conventional machine learning methods (e.g., LDA, Naive Bayes, KNN, Decision tree, Bayesian network) (Bhattacharya, Mishra, 2018), kernel methods (e.g., SVM, KFD, MKFD) (Dong et al., 2016, 2020a; Shi, 2008), ensemble learning methods (e.g., random forest, AdaBoost, XG-Boost, LightGBM) (Dong et al., 2023a), and artificial neural network algorithms (e.g., MLP, BP, CNN, RNN) (Desouky et al., 2020; Tian et al., 2021; Xue et al., 2014; Zazoun, 2013).

In the majority of fracture identification cases, a vector sample comprising well logs at a specific depth is selected as the input data for the AI model identifying fractures. In the training process, the AI model undergoes a supervised learning approach wherein it is trained using labeled samples derived from rock core observations or image log interpretation. This process aims to augment the fracture identification performance of the AI model and enhance its generalization capabilities (Dong et al., 2022a). Log responses of fractures exhibit anomalies not only in absolute values but also in relative changes when compared to the background responses of neighboring layers (Dong et al., 2023b). The prevailing way to identify fractures is to treat each sample, associated with a specific depth, as an independent and discrete entity. Nevertheless, this approach can pose difficulties in effectively extracting relative changes or anomalies within the data. To deal with this issue (Problem 1), a sequence data sample will be constructed instead of a discrete entity by a window sliding method. Based on conventional logs, some logs indicating fractures will also be reconstructed to amplify the logging response characteristics associated with fractures. Then the relative changes and the spatial location correlation can be detected by feature extracting in a deep artificial neural network. In this work, a recurrent neural network will be used to classify fracture labels according to the sequence data.

The opposite log response sequences resulting from positive and reverse sedimentary cycles can hinder fracture development. To address this issue (Problem 2), a bidirectional Long Short-Term Memory (BiLSTM) network is employed for feature extraction and utilization. The implementation of Bidirectional Long Short-Term Memory (BiLSTM) addresses the difficulties presented by sedimentary cycles effectively. Moreover, as an extension of Recurrent Neural Network (RNN), BiLSTM leverages the inherent advantages suitable for processing sequential data, enabling it to extract classification features from the constructed sequence data sample proficiently. Furthermore, BiLSTM retains the benefits of LSTM, which can resolve the challenges associated with gradient vanishing or exploding gradients encountered when dealing with lengthy sequential data in conventional RNN models (Zeng et al., 2022).

In addition, the identification of fractures is hindered by the problem of data imbalance (Problem 3), where the number of nonfracture samples greatly surpasses that of fracture samples by a ratio of at least 10:1, and in some cases even up to 50:1. Consequently, conventional artificial intelligence models encounter difficulties in effectively distinguishing between these two categories, namely fractures and nonfractures. In such instances, there is a tendency for these models to erroneously classify all samples as belonging to the majority class, as the high ratio of the major class corresponds to the high classification accuracy (Dong et al., 2016; Li et al., 2022). To address this issue, this study employs the technique of random undersampling to balance the dataset and reduce the disparity in sample proportions.

The lithological complexity and heterogeneity of carbonate reservoirs pose challenges in accurately identifying fractures through logging responses. This issue turns fracture identification using logs into a complex nonlinear classification problem, resulting in weaker and more ambiguous logging responses for fractures. Introducing Recurrent Neural Networks (RNN) in this study offers a promising solution, as RNN is a deep learning artificial neural network method that inherits the advantages of handling nonlinear classification in artificial neural networks.

In general, to improve fracture identification, a novel approach, termed **F**racture **I**dentification by window **s**liding and **r**ecurrent neural network (FIsr), is proposed, which combines the aforementioned techniques to address the three problems associated with fracture identification. To evaluate the effectiveness of the proposed FIsr method, a dataset from the carbonate reservoirs of the Asmari Formation in the Middle East is utilized, where numerous fractures are known to occur in the target formations.

2. Geological settings

The dataset is from A Oilfield located in the Zagros Mountain folded belt, adjacent to Mesopotamia plain in the west and Iranian plateau in the east (Fig. 1(a)) (Al-Banna, 2008; Ghanadian et al., 2017; Le Garzic et al., 2019). The target reservoirs are carbonates of Palaeogene Oligocene-Neogene Miocene Asmari Formation. Due to the Himalayan Movement, Zagros Mountain was formed by rapid uplift under the action of NE-SW huge horizontal compression tectonic stress caused by continuous collision between Arabian plate and Eurasian plate (Ala et al., 1980). Affected by the compressive stress, the oilfield is characterized by a NW-SE anticlinal structure. Several NW-NE trending strike-slip faults are developed in the west, while NNW-SSE trending normal faults are prevalent in the central region. The target formation is the Asmari member of Tertiary Oligocene-Lower Miocene Asmari Formation (Fig. 1(b)). The lithology is mainly dolomite and anhydrite mass dolomite, and anhydrite mostly appears in clumps (Luo et al., 2019; Wang et al., 2018). Asmari Formation is deposited to semi-confined platform deposition, mainly carbonate deposition, with less terrigenous clastic deposits. The upper layer is overlain by evaporites, which are extensively distributed in Gachsaran formation (Lower Fars). The sedimentary facies of member A are confined platform, and the abundance of evaporites (generally anhydrite masses) in dolomite generally shows an upward trend. The reservoir has an average porosity of 8.7% and average permeability of 10.6mD. There are few primary pores, and pore types are mainly secondary pores produced by dissolution during diagenesis. Moreover, fractures develop widely within the targeted reservoirs significantly impacting the oil production (Dong et al., 2022b).

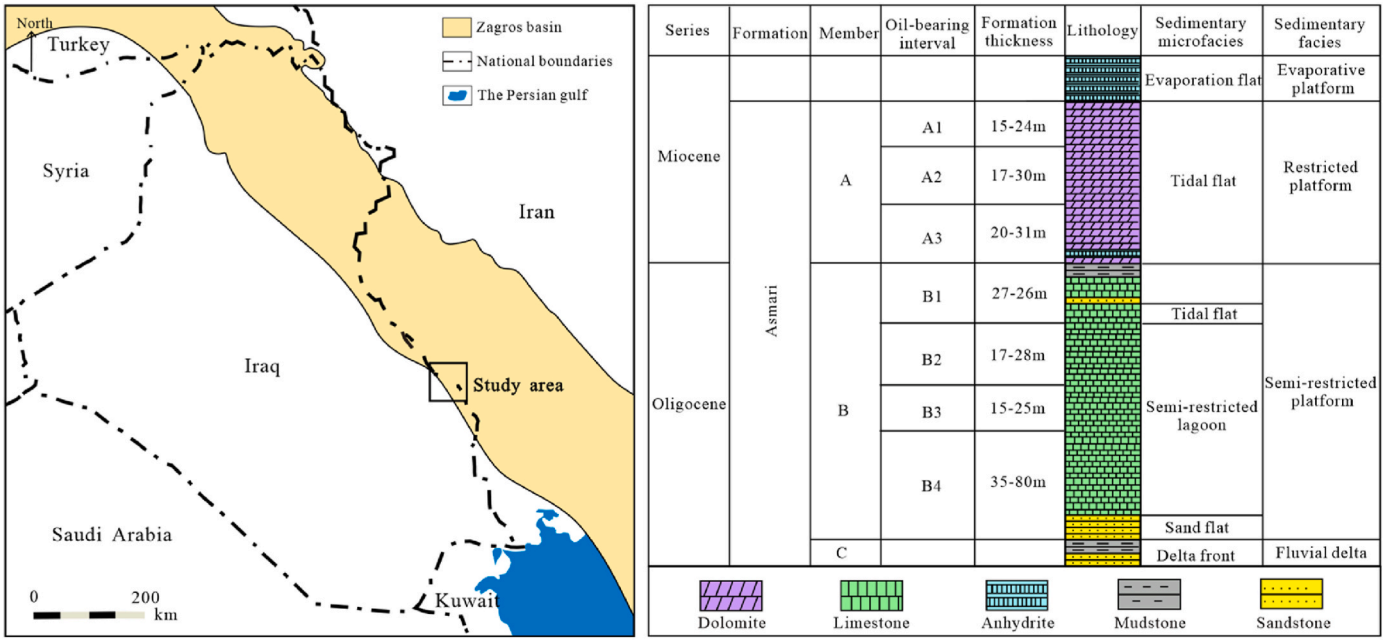
3. Method of fracture identification

The proposed fracture identification method, FIsr, incorporates three distinct components as described in Section 3.1. These components are: (1) the reconstruction and selection of input logs for fracture identification (Section 3.2) corresponding to Fig. 2(a), (2) the training and prediction process utilizing window sliding in FIsr (Section 3.3) corresponding to Fig. 2(b)–(c), and (3) the recurrent neural network classifier employed in FIsr (Section 3.4) corresponding to Fig. 2(b).

3.1. Workflow of FIsr for fracture identification

The workflow for utilizing FIsr to identify fractures can be summarized into three steps, as depicted in Fig. 2.

The first step involves the reconstruction of fracture-indicating logs based on theoretical geophysical responses and the selection of input logs from conventional and reconstructed logs for fracture identification, as illustrated in Fig. 2(a). Once the appropriate logs are determined, the window sliding method is employed to scan the well logs and corresponding labels in a top-down fashion. This scanning process generates sequence data samples that serve as inputs for the RNN. Furthermore, a random undersampling method is utilized to ensure data



(a) Location of A Oilfield (b) Target formation

Fig. 1. Location of A Oilfield and the target formation (Dong et al., 2022a).

balance is achieved.

The second step entails training a BiLSTM classifier to differentiate sequence samples' labels, as depicted in Fig. 2(b). The training dataset, comprising approximately 80% of the data, will be utilized to train the BiLSTM model. The remaining data will serve as the test dataset to assess the efficacy of the developed BiLSTM classifier. Upon satisfactory evaluation, the model will undergo blind well tests, where fracture interpretation from rock core observations or image logging interpretation will be employed. If this evaluation yields satisfactory results, the BiLSTM model can then be employed to predict fracture labels for new wells.

In the third step, the developed model in Fig. 2(c) is utilized to predict fracture labels for a new well. Similar to the previous training step, log reconstruction and window sliding techniques are applied to generate sequence data samples for this particular well. Subsequently, the BiLSTM classifier is employed to predict the label for each individual sample. By aggregating the labels of various sequence data points, the probability curve illustrating the development of fractures in this new well is generated.

This comprehensive workflow provides a systematic approach to fracture identification using FIsr, encompassing data reconstruction, RNN-based classification, model training, evaluation, and prediction for new wells.

3.2. Reconstruction and selection of logs for input of FIsr

The objective of log reconstruction is to enhance the log responses induced by fractures. Consequently, logs exhibiting geophysical responses are theoretically constructed to serve as potential inputs for FIsr (as discussed in Section 3.2.1). Subsequently, a comprehensive selection method (as detailed in Section 3.2.2) can be employed to choose the input logs. This selection process aims to prioritize logs that not only contribute significantly to the accuracy of fracture identification but also demonstrate strong correlations with each other. This ensures that both conventional logs and reconstructed logs are considered during the selection procedure.

3.2.1. Log reconstruction for distinguishing fractures

Table 1 displays several reconstructed logs as examples. To account for the abnormal log responses caused by fractures, change rates of logs are constructed, wherein *Slope* represents a specific log at the i-th depth. Here, v_i is a well log. As opened fractures can result in a disparity between deep and shallow resistivity, the ratio of deep and shallow resistivity (RSD) is constructed. Fractures can lead to high AC, high CNL, and low DEN values, hence the ratio of AC and DEN (ADR) and the ratio of CNL and DEN (CDR) are constructed.

3.2.2. Logs selection for fracture identification

The combination of feature contributions in the random forest method, hierarchical clustering, and correlation heat map is employed to select logs for fracture identification.

(1) Random forest is an ensemble learning method, in which the randomness means randomly sampling data and randomly selecting features (Dong et al., 2023a; Tariq et al., 2022b). It employs the Gini index metric to calculate the weight of each feature value, determining their contribution towards building decision trees (). This analysis enables comparison of the degree of importance of each feature value, which in turn facilitates feature ranking based on their relative contributions. It aims to generate trees in which nods have high purity (namely low Gini value). The contribution degree of the j-th feature is the change of Gini index on all nodes of a tree as shown in Eq. (1).

$$VIM_j = \sum_{i=1}^{n} \sum_{m \in M} VIM_{jm} \tag{1}$$

where n is the number of decision trees in the random forest, and M is the set of all nodes on a decision tree; $VIM_{jm} = GI_m\text{-}GI_l\text{-}GI_r$, VIM_{jm} represents the Gini index change of feature j on node m, reflecting the weight of feature j in the CART tree. GI_m represents the Gini index before node branch, and GI_l, GI_r represents the Gini index of left node and right node after node branch respectively.

As an example, the process of calculating the weight of feature x_2 using dataset D is depicted in Fig. 3. The dataset D comprises two

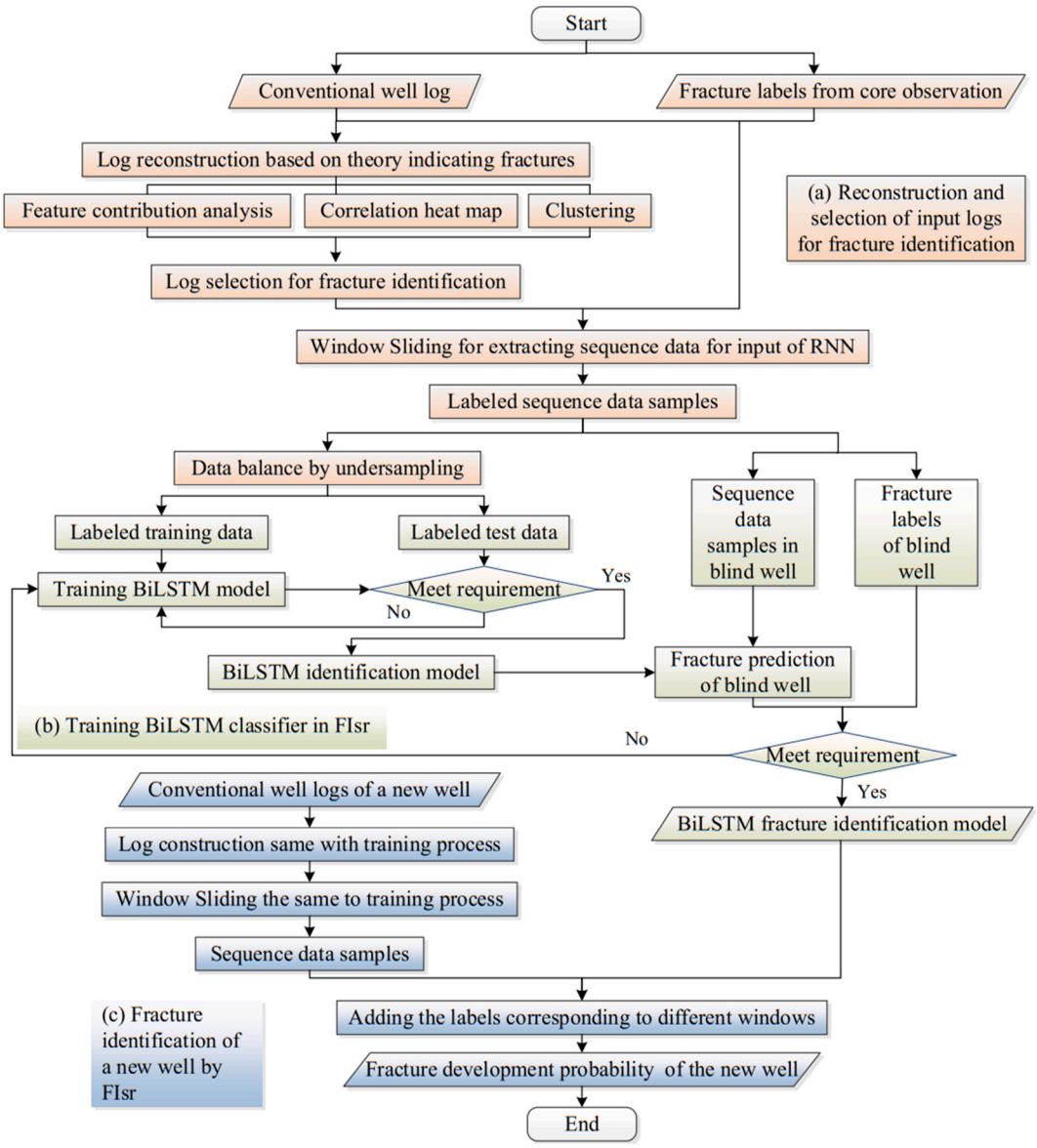


Fig. 2. The workflow of FIsr.

Table 1
Reconstructed curves.

Reconstructed logs	Equation
Change rates of logs	$Slope = \frac{(\lvert v_i - v_{i-1}\rvert + \lvert v_{i+1} - v_i\rvert)}{2}$
Ratio of deep and shallow resistivity	$RSD = \frac{RD}{RS}$
Ratio of AC and DEN	$ADR = \frac{AC}{DEN}$
Ratio of CNL and DEN	$CDR = \frac{CNL}{DEN}$
Ratio of the difference in micro resistivity	$MRR = \frac{R_{max} - R_{XO}}{R_{max} - R_{min}}$
Difference in relative magnitude of resistivity	$RDS = \frac{\lvert RD - RS\rvert}{RS}$

features, x_1 and x_2, and five data points. Firstly, the Gini index of the two features is computed, with the feature having the lower Gini value chosen as the primary node in the decision tree. This process leads to the generation of decision tree 1 and decision tree 2. The contribution or the weight of feature x_2 in these two decision trees is then determined by calculating the Gini index of feature x_2 before and after the branch at node m_1 by using Eq. (1). The sum of the weights of feature x_2 in all decision trees, which is 0.33, represents the weight of feature x_2 in dataset D.

(2) Correlation heat map method

The correlation coefficient is usually used to describe the correlation degree between two variables (Eq. (2)) (Tariq et al., 2022b). Logs that exhibit a high correlation coefficient can be consolidated into a single representative log.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2(y-\bar{y})^2} \quad (2)$$

where the absolute value of r is between 0 and 1. Let x and y represent the sequence data of the logging curves. $\bar{x}$ and $\bar{y}$ represent the weighted arithmetic means of two variables. The higher r, the higher correlation of x with y.

(3) Hierarchical clustering

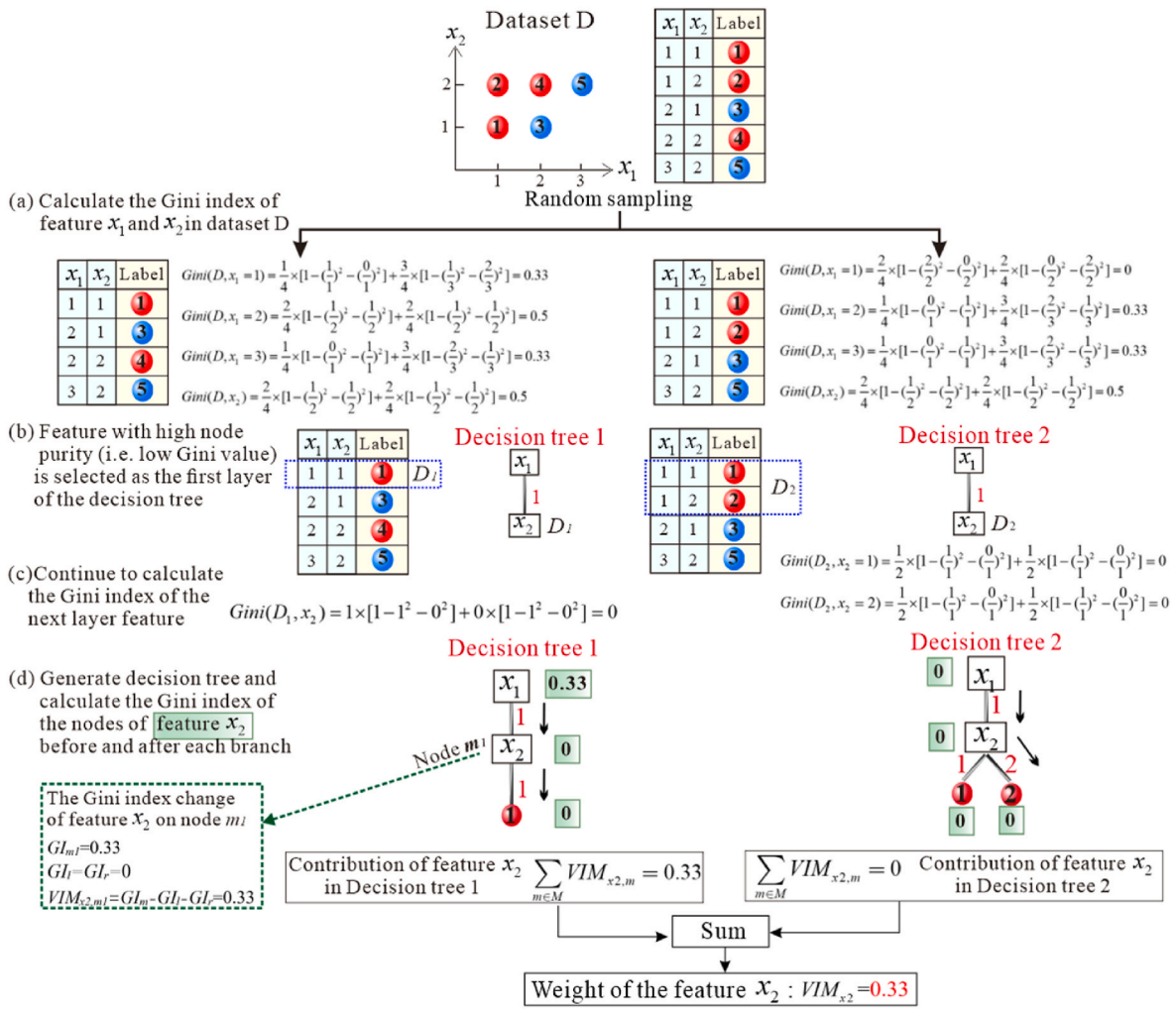


Fig. 3. Schematic diagram of contributions of features for classification.

Hierarchical clustering is a process of categorizing analyzed data based on similarity using distance as a metric. Initially, each sample is treated as a unique cluster. The algorithm then iteratively performs two key steps: first, it identifies the two clusters with the highest similarity, and second, it merges the most similar clusters. This iterative process continues until all clusters are merged, resulting in a complete hierarchical clustering structure.

The visualisation of the hierarchical clustering dendrogram with four features is shown in Fig. 4(b), from which we can see that Feature 2 and Feature 3 are easily clustered together, indicating a high correlation. By using random forest to select the top four features with high weights among the seven features (Fig. 4(c)), it can further improve the accuracy of the removal of the highly linearly correlated Feature 3 by performing correlation heat map analysis and hierarchical clustering.

3.3. *Window sliding in the training and prediction processes in the FIsr*

In this study, the utilization of the "segment" by window sliding as a replacement for "point" is employed within the conventional fracture identification method. Similarly, the window sliding approach is

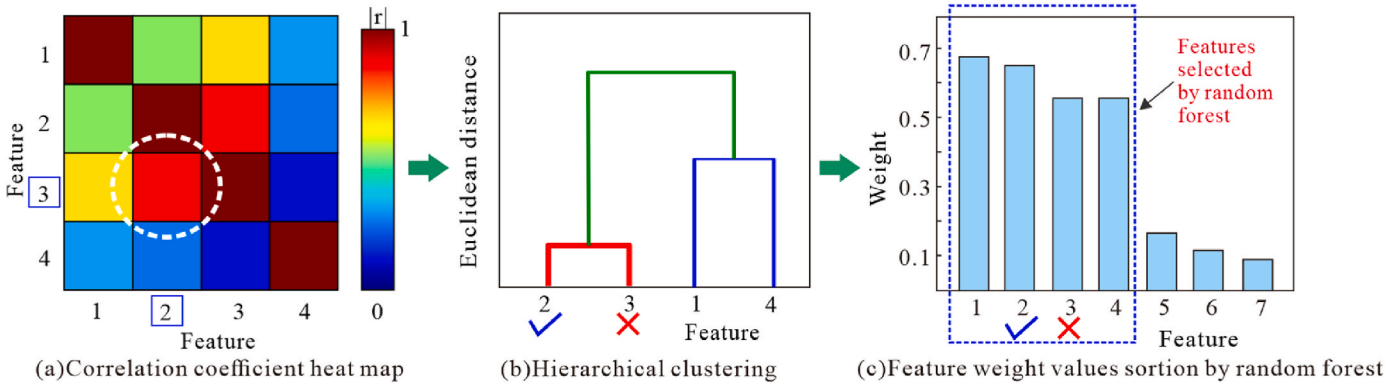


Fig. 4. Schematic diagram of features selection.

adopted during the training of FIsr and in the prediction of fractures for a new well using FIsr.

3.3.1. Window sliding for obtaining training and test data

Fig. 5 illustrates the application of the window sliding methodology to obtain a series of sequential data samples. The majority of these samples are utilized as training data for constructing a fracture identification model, while the remaining samples are used as test data to evaluate the performance of the developed model. The window height and sliding stride are set as k and s, respectively. If this window touches a fracture zone, the label of this sequential data sample will be set as fracture ("1″), otherwise nonfracture ("0″).

It should be noted the problem of data imbalance still exists in the sequential data samples. To address the issue of data imbalance in the labeled sequence data samples, where there is a significant disparity between the number of fracture and nonfracture samples, it becomes necessary to employ data balancing techniques. Data balance refers to the process of equalizing the distribution of each group by reducing the difference in sample numbers. In this work, the Random Undersampling (RUS) method is utilized to achieve data balance, as depicted in Fig. 6. This approach involves randomly selecting a certain number of samples from the class with an abundance of examples and removing them from the dataset until the desired distribution is attained. By adopting RUS, the data is effectively rebalanced, enabling a more equitable representation of both fracture and nonfracture samples within the dataset.

3.3.2. Window sliding for predicting labels of data samples

In the prediction of the fracture identification by FIsr, window sliding is a necessary step (Fig. 7). Typically, the sliding step size (or stride) is set as one, which results in each point being overlapped by k windows. The predicted labels are aggregated, and then divided by k to obtain the fracture development probability in Eq. (3) at a depth point. By sliding the window from top to bottom, a curve representing the probability of fracture development can be generated. The fracture intensity score (*FIsr*), ranging from 0 to 1. A higher value on the curve signifies a greater probability of fracture development.

$$FIsr = \frac{\sum_1^k y_i}{k} \tag{3}$$

where y_i is the predicted result of the i-th sequence data sample.

As illustrated in Fig. 7, at the depth point x, three ($k = 3$) windows traverse through this specific depth. In the context where two sequence data samples corresponding to the first two windows are identified as fractures ("1″), and one data sample corresponding to the third window is recognized as a nonfracture ("0″). Then the FIsr will be calculated as 2/3.

3.4. Recurrent neural network as a classifier in FIsr

The structure of the BiLSTM is shown in Fig. 8. Compared with the hidden state of front-to-back transmission of traditional recurrent neural network, the hidden layer of back-to-front transmission of information is added, making it more flexible in dealing with longer sequence data (Zhou et al., 2021). Assume that the state of the forward hiding layer is $\overrightarrow{h}_t$ (Eq. (4)), and the reverse hidden state be $\overleftarrow{h}_t$ (Eq. (5)). The calculation formula is as follows:

$$\overrightarrow{h}_t = \varphi\left(X_t W_{xh}^{(f)} + \overrightarrow{h}_{t-1} W_{hh}^{(f)} + b_h^{(f)}\right) \tag{4}$$

$$\overleftarrow{h}_t = \varphi\left(X_t W_{xh}^{(b)} + \overrightarrow{h}_{t+1} W_{hh}^{(b)} + b_h^{(b)}\right) \tag{5}$$

where the weights $W_{xh}^{(f)}$, $W_{hh}^{(f)}$, $W_{xh}^{(b)}$, $W_{hh}^{(b)}$ and deviation $b_h^{(f)}$, $b_h^{(b)}$ are the model parameters and φ are activation function, X_t is the sequence data extracted from the logging curve after window sliding.

Then the hidden state $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ are merged into the hidden state h_t and inputs it to the output layer. Output layer calculation output O_t (Eq. (6)):

$$O_t = h_t W_{hq} + b_q \tag{6}$$

where the weight W_{hq} and deviation b_q are the model parameters of the output layer, and the number of hidden layer units in different directions can also be different.

4. Fracture identification in tight carbonate reservoirs of Zagros basin

4.1. Reconstruction and selection of input logs for fracture identification

The set of 17 candidate variables comprises the 8 original conventional logs (GR, CAL, CNL, DEN, AC, logRD, logRS, and logRXO) and the 9 reconstructed logs (SlopeAC, SlopeCAL, SlopeGR, RSD, MRR, ACR, ADR, CDR, and RDS). Their contributions to fracture identification, as determined by the random forest approach, are illustrated in Fig. 9(a). Variables that exhibit high contribution weights should be given priority

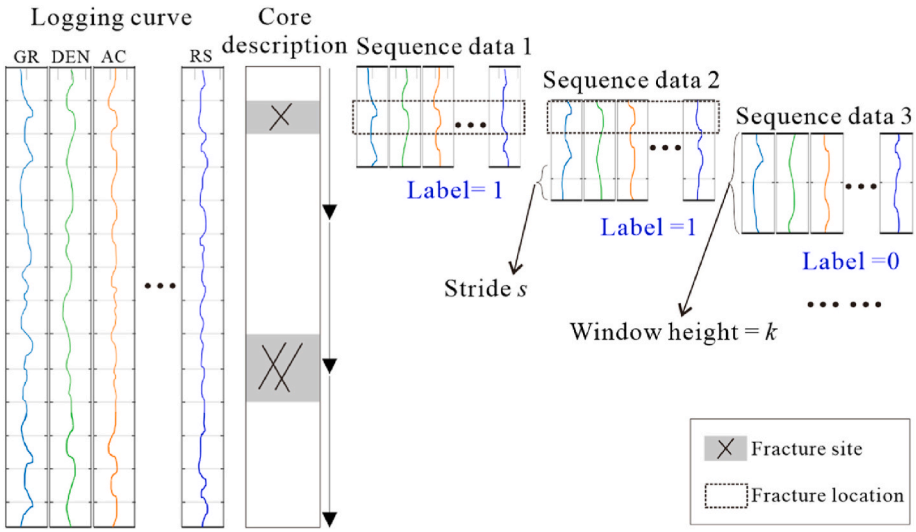


Fig. 5. The training process of window sliding for logs.

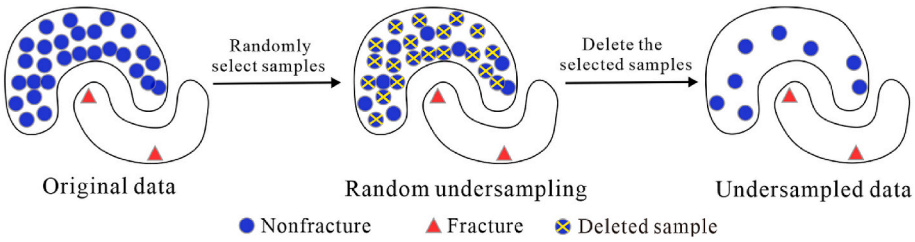


Fig. 6. Schematic diagram of data balance by the random undersampling method.

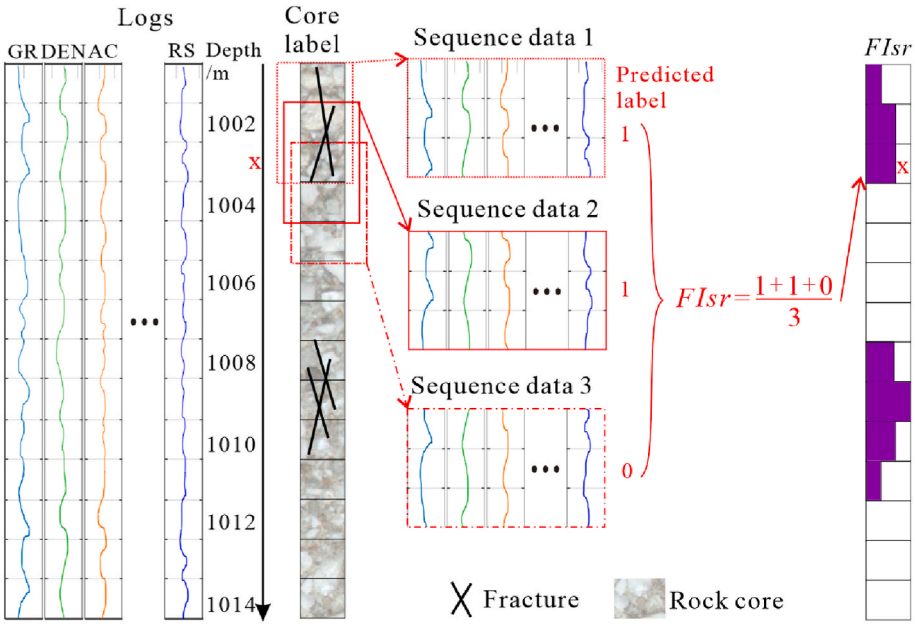


Fig. 7. Fracture prediction process of window sliding for logs.

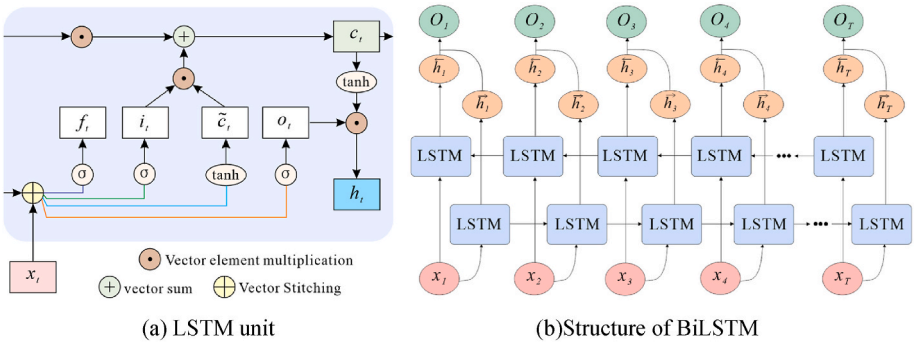


Fig. 8. Schematic diagram of BiLSTM.

in the selection process, namely RSD, CNL, AC, DEN, ADR, CAL, CDR, logRS and GR. When applying random forests for curve selection on the 17 logs, the time consumed was 57.9s.

As shown in Fig. 9(b) and (c), a strong positive correlation exists between AC and CDR, as well as between ADR and logRS. Due to the relatively higher contributions of AC and ADR compared to CDR and logRS, the latter two variables (CDR and logRS) are eliminated. Therefore, a subset of seven logs (RSD, CNL, AC, DEN, CAL, RS, and GR) is selected as input variables for fracture identification.

The original conventional well logs and the selected three constructed logs (RSD, ADR, and CDR) are shown in Fig. 10. Fracture description from rock observation is shown in the last column of Fig. 10, which is used to generate fracture labels of well logs. The red lines indicate the zones with fractures. As shown in Fig. 10, fractures are associated with higher values of ADR and CDR, whereas nonfracture regions exhibit lower RSD. By examining the sections demarcated by the blue rectangle in the illustration, it becomes apparent that the reconstructed curves possess the ability to enhance the conventional logging response characteristics at locations where fractures have developed.

4.2. Development of fractures identification model

Three wells containing rock cores were observed, with a combined length of 528 m. The logging data sampling interval is 0.125 m, meaning that there are 8 sampling points for every 1-m depth. The initial dataset consists of 5243 discrete samples, comprising 4969 fracture

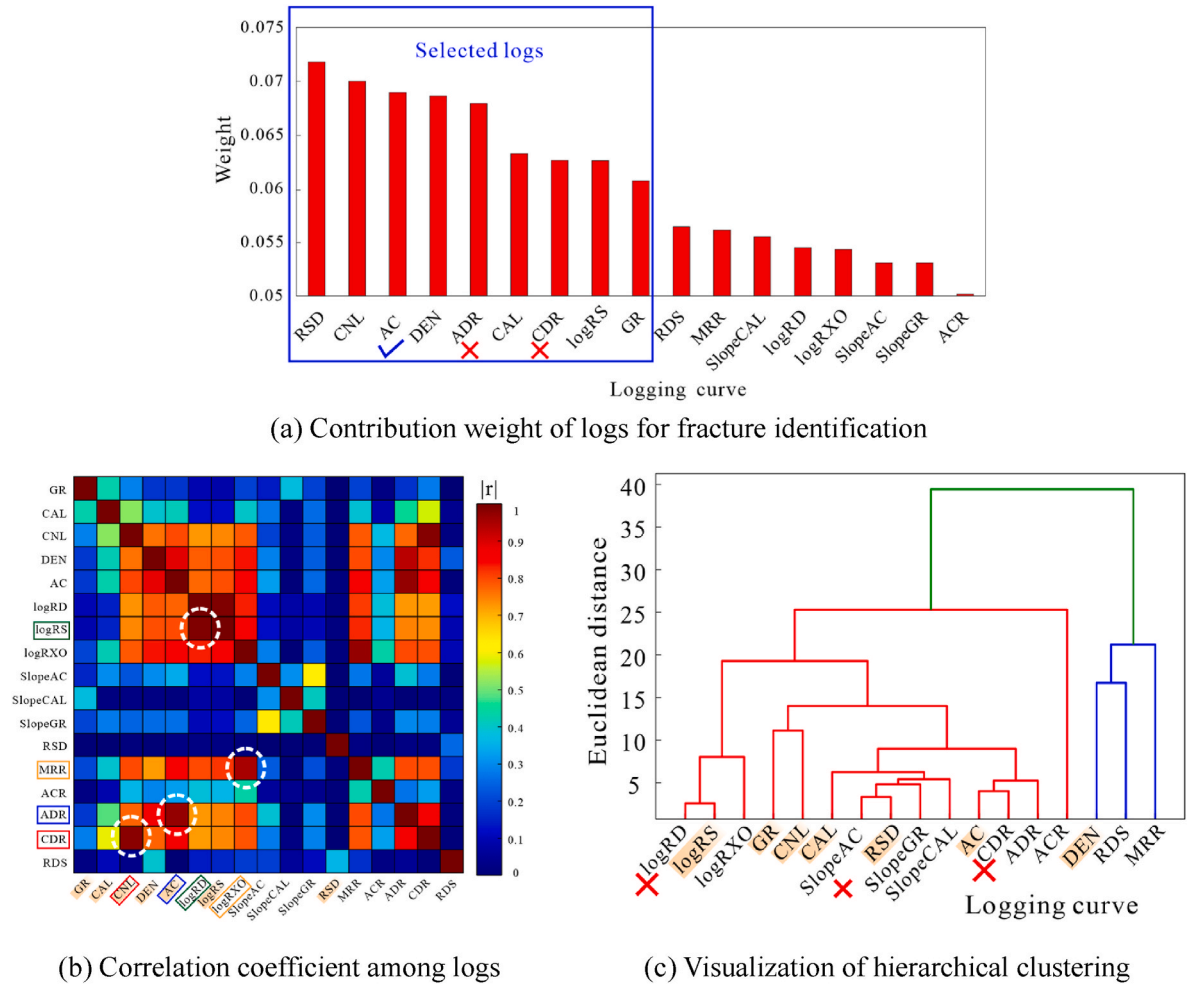


Fig. 9. Logging curve optimization results.

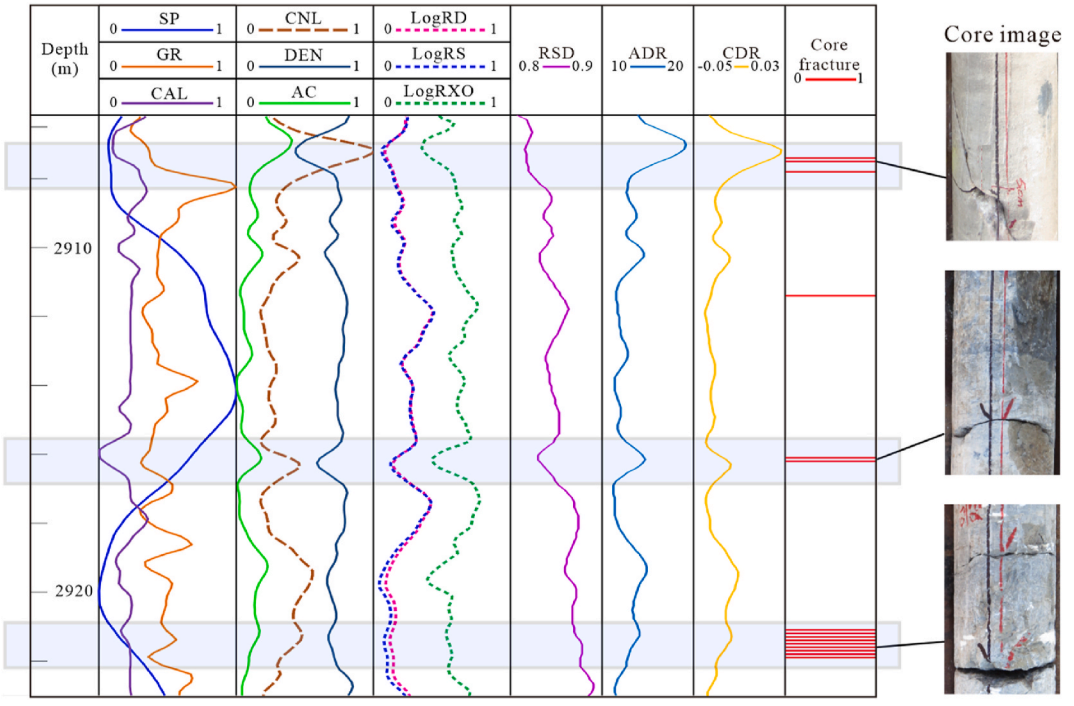


Fig. 10. The original conventional logs and the selected reconstructed curves.

observations and 274 nonfracture observations. After applying a window sliding approach with a window height of 5 and a sliding stride of 1, a total of 5239 sequential data samples were generated. Among these samples, there were 572 samples corresponding to fractures and 4667 samples representing nonfracture. Then the random undersampling of the nonfracture data was performed, with the sampling coefficient determining the proportion of the data. In the original data, fracture data accounted for 11%, while nonfracture data accounted for 89%. When the undersampling rate was 0.35, the proportions of fracture data and nonfracture data were 26% and 74%, respectively, and the processing time was 3.1s. Then there left 2200 sequential data samples (572 fracture and 1628 nonfracture). The 80% of data will randomly selected as training data to generate a BiLSTM model while the rest 20% will be as the test to evaluate this model.

Besides another well will be as a blind well to assess the efficiency of the built model for fracture identification. Noted the blind well data will not undergo the undersampling process.

The FIsr method classifier, as shown in Fig. 11(a), includes a sequence data input layer, two BiLSTM layers, two Batch Normalization layers, a Dropout layer, a fully connected layer, a Softmax layer, and a classification output layer. The hidden layer contains 500 neural units in the BiLSTM with two layers. The method predicts training data with an accuracy of 95% after 400 iterations, taking a total time of 11 min. The technologies, corresponding parameters, and CPU time involved in the FIsr fracture identification process using all the provided fracture data are shown in Table 2. The training and validation process of the model is depicted in Fig. 11(b). The accuracy of both the training set and validation set gradually increased and ultimately stabilized with the increase in the number of iterations. The final training data accuracy is 98%, and the validation data accuracy is 95%.

4.3. Validation test of fracture identification model

(1) Validation from statistical aspect

A confusion matrix of a classifier for test data is the common way to evaluate the built classifier (Desouky et al., 2021b). The confusion matrix of the built fracture identification model for test data is shown in Fig. 12(a). Rows represent predicted classes, while columns represent actual classes. The overall accuracy is 95%. The recall and precision of fracture category are 90.2% and 94.8%, respectively, while those of the nonfracture are 97.4% and 95.1%, both above 90%. Accuracy provides a general measure of a prediction's performance. Precision focuses on the accuracy of the identified results, while recall leans towards the comprehensiveness of the identified results. In general, the built fracture identification model performs well from the statistical perspective.

(2) Blind well test

Table 2
Parameter table of FIsr method.

Technique	Name	Parameter	CPU Time
RNN classifier	Input dimensions of the sequence	7	15min
	BiLSTM layer	2	
	Batch normalization layer	2	
	Dropout layer	1	
	Dropout rate	50%	
	Full connected layer	2	
	Softmax layer	1	
	Classification layer	1	
	MaxEpochs	9500	
	MiniBatchSize	400	
	L2 regularization	0.001	
	Learn rate	0.01	
Random Undersampling	N_samples(fracture)	4969	3.1s
	N_samples(nonfracture)	274	
	Undersampling rates	0.1–0.8	
Window sliding	Window height	5	2.6s
	Sliding stride	1	
Random Forest	Number of CART trees	500	57.9s
	Min_samples_leaf	5	
Hierarchical clustering	Number of clusters	2	2.3s

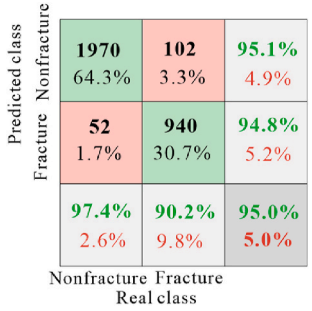


Fig. 12. Confusion matrix of test data.

A blind well test is implemented, and the result is shown in Fig. 13, in which the logs are data normalized. A comparison between conventional well logging curves and reconstructed well logging curves is performed, while rock core images and imaging logs are employed to display the fractures serving as the input for the study. The middle row in the figure represents the position of core fracture development, and it can be seen that there are 15 fracture development sections. The probabilistic interpretation result of the FIsr bidirectional neural network fracture recognition model is on the right. It can be seen that all fracture zones can be recognized by FIsr. The more densely developed the core

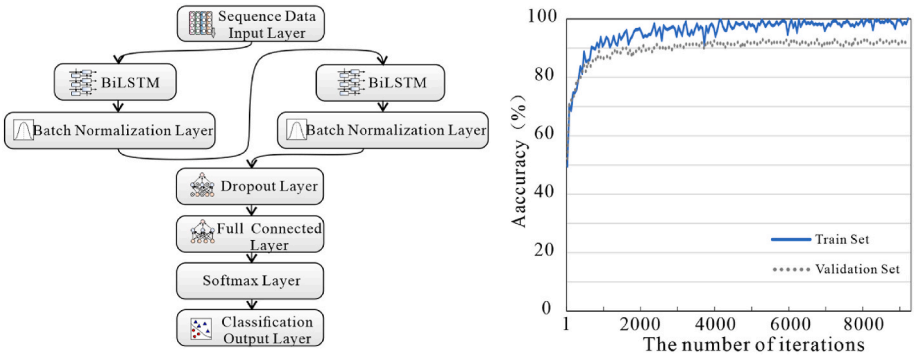


Fig. 11. The classifier of RNN model.

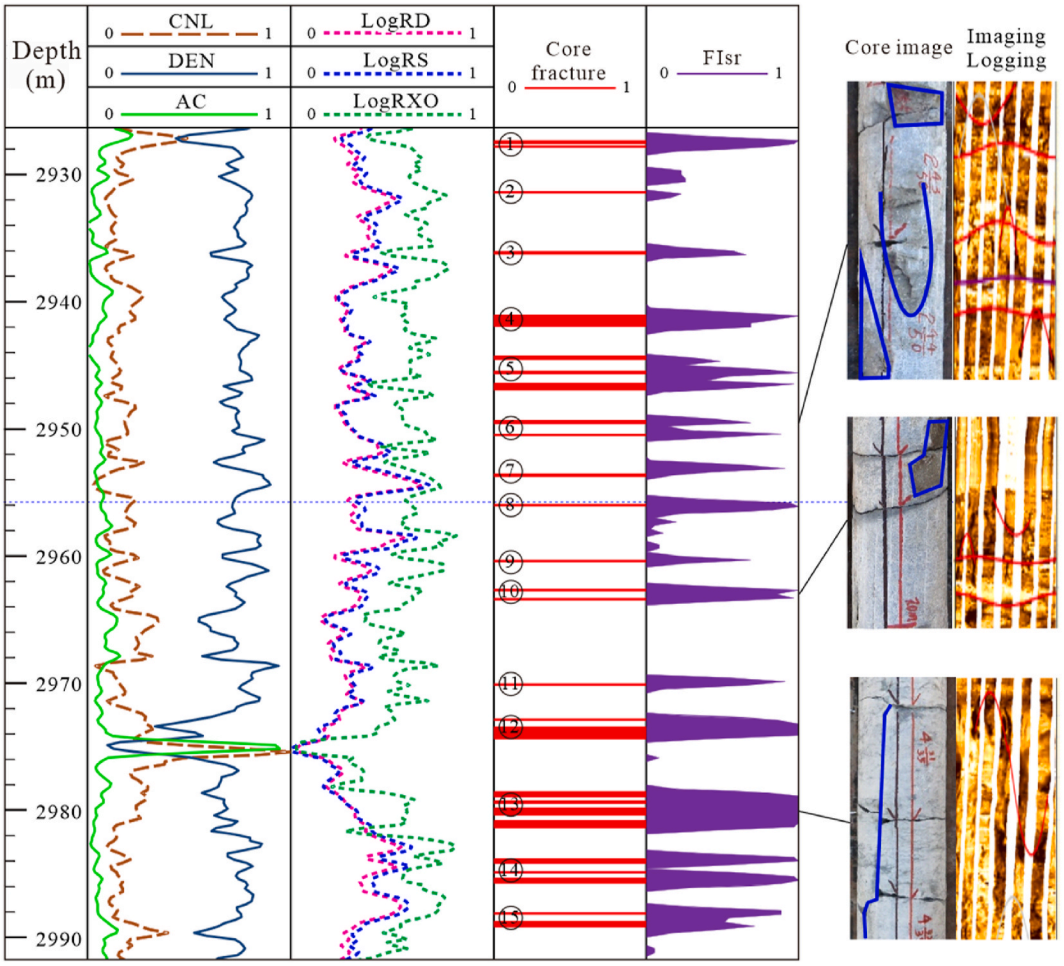


Fig. 13. Blind well test.

fractures, the greater the probability of FIsr predicting the fractures. However, near the ③ and ⑧ fractures, FIsr identifies fractures with a certain probability but actually has no fractures in the cores. This is because fractures develop around these locations. Although window sliding reduces the uncertainty, it also decreases the vertical resolution.

5. Discussions

5.1. Influence of height of window sliding on fracture identification

The window height can affect the performance of the FIsr. Small window may not capture the influence of fractures on the adjacent logging curves. Large window contains curves of long depth, but it may amplify the response of other factors, leading to reduced accuracy.

The study evaluated the RNN performance under different window heights for logs with a sampling interval of 0.125m. The recall for samples with fractures (R_{frac}) and with nonfracture (R_{non}), as well as the accuracy for samples with fractures (A_{frac}) and the overall accuracy of the model were analyzed. Results in Fig. 14 show that window heights between 2 and 5 intervals exhibit a significant increase in R_{frac} and A_{frac} as the window height increases. Additionally, R_{non} and overall accuracy rate also increase slightly. However, window heights between 5 and 10 intervals show a significant decrease in R_{frac} and A_{frac}, indicating inaccurate prediction of fracture location due to excessively large window heights. Based on the above evaluations, the best prediction corresponds to the window height of 5 intervals (about 0.625m).

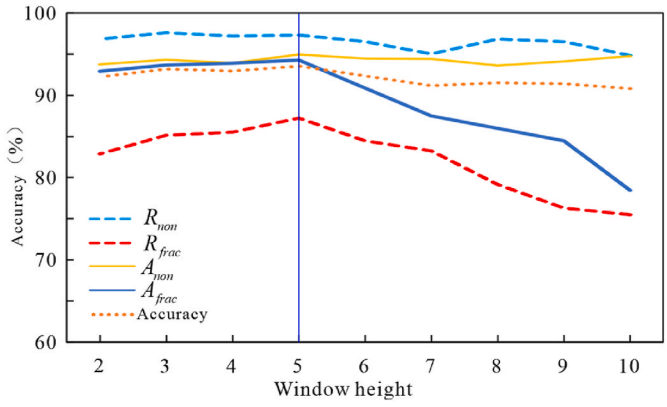


Fig. 14. Performance comparison of RNN corresponding to different window heights.

5.2. Influence of data balance on fracture identification

The random undersampling method is used to address this problem of data imbalance, as explained in Section 3.3.1. The ratio of random undersampling is an important parameter affecting fracture identification. The original data had 4969 fractures and 274 nonfracture data. Fig. 15 shows the accuracy of RNN under different undersampling ratios. When the undersampling rate is high (0.6–1), the accuracy of fractures is very low, although the overall accuracy of the model can be high (90%–

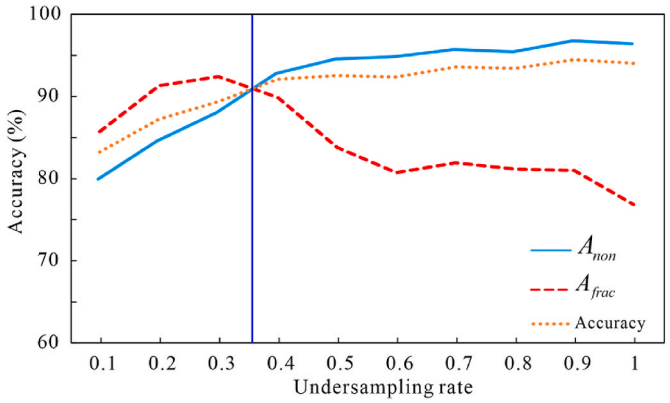


Fig. 15. Accuracy against different undersampling ratios.

95%). As the undersampling rate decreases, the accuracy gradually increases. However, when the undersampling rate falls below 0.3, the overall accuracy and fracture precision of the model begin to decline due to too few nonfracture labels. Based on the above analysis, it is concluded that the appropriate undersampling rate is 0.35, which yields an accuracy rate and precision rate ranging from 90% to 95%.

5.3. Distribution of fractures along well trajectory

Fig. 16 provides a graphical representation of the vertical distribution of fractures identified by the FIsr method. The plot displays the FIsr and the fracture density of each layer (FFI) (Dong et al., 2020b). The amplitude reflects the magnitude of the probability of fracture development.

The plot clearly shows that Section A has a higher probability of fracture development than Section B. This observation is indicated by the black arrow in the plot, which highlights the difference in fracture probability between the two sections. These findings align with previous knowledge of the area, suggesting that the FIsr method is a reliable approach for identifying and characterizing fractures in logging data.

5.4. Application of FIsr fracture identification method

The FIsr fracture prediction method, developed in this study, not only serves as a valuable tool for predicting fractures in reservoirs but also plays a crucial role in fracture network modeling, oil or gas development adjustment, and hydraulic fracturing design.

(1) This method can effectively reveal the vertical distribution of fractures within the study area, enabling analysis of key factors influencing fracture development, such as sedimentation and tectonic influences.
(2) Establishing a subsurface natural fracture network model necessitates incorporating spatial constraints on fracture density. By interpolating FIsr values obtained from multiple wells, a 3D constraint model for fracture network modeling can be created and refined using image logging interpretation.
(3) The distribution of the natural fracture network holds significant importance for oil and gas development and adjustment. For instance, well productivity often correlates with fracture density. Consequently, identifying well locations with higher fracture connectivity to oil and gas can guide the placement of new wells.
(4) The methodology presented in this paper offers a comprehensive interpretation of individual wells, enabling analysis of fracture network connectivity in subsurface reservoirs. This knowledge helps address various development challenges. For example, water invasion is commonly encountered during the development of many gas fields, particularly in tight reservoirs. Understanding the connectivity of the fracture network allows for proactive deployment of appropriate development strategies to mitigate water invasion risks. For wells connected to water, it may be advisable to decrease gas production or drill additional drainage wells.
(5) The well-established knowledge of subsurface natural fracture networks also proves valuable in hydraulic fracturing operations, facilitating the creation of more efficient fluid channels by leveraging the existing fracture network.

Moreover, this method exhibits potential for application in reservoir fracture identification across different lithologies.

6. Conclusions

The identification of fractures using conventional logging responses is a challenging task, as these responses are often intricate and feeble. To address this issue, a novel method for intelligent fracture identification,

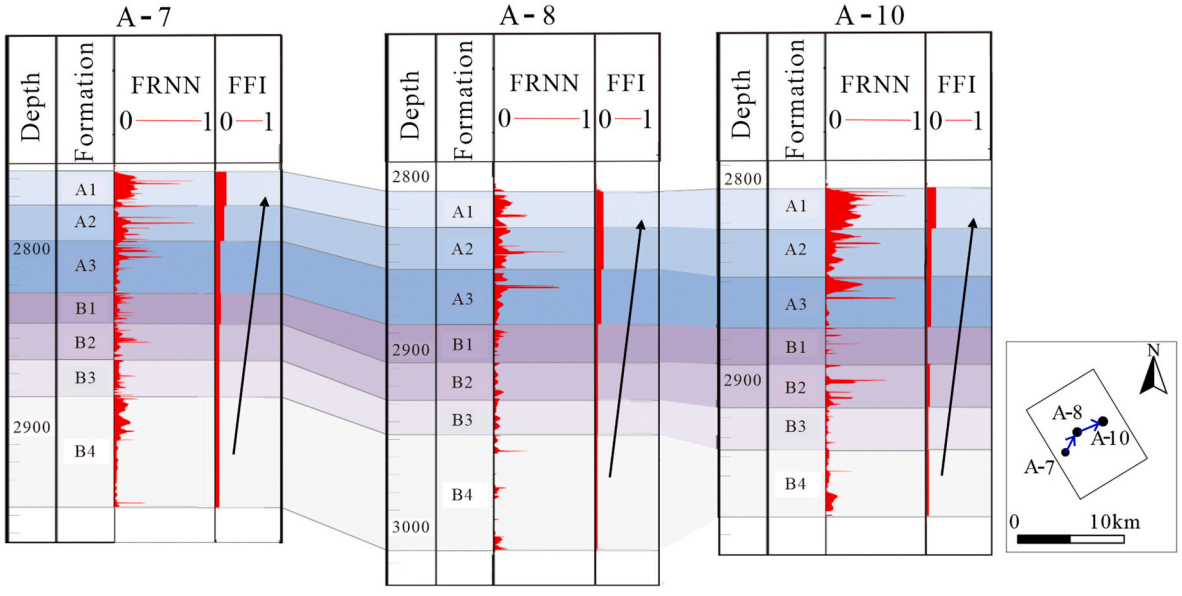


Fig. 16. Longitudinal distribution of fractures.

known as FIsr, is proposed. The efficacy of this method is validated using a dataset from carbonate reservoirs in the A Oilfield of the Middle East. The following conclusions are obtained:

(1) Through the utilization of a combination of random forest, correlation coefficient heat map, and hierarchical clustering methods, a total of seven logs were selected as inputs for FIsr from both the original and reconstructed logs. These logs include RSD, CNL, AC, DEN, CAL, RS, and GR. The statistical validation of the FIsr method revealed a remarkably high accuracy of 95% and impressive recall and precision metrics, both exceeding 90%. Moreover, a blind well test demonstrated that FIsr was capable of detecting all fracture zones. The distribution of fractures along the well trajectory revealed that Section A was more likely to develop fractures than Section B, a finding consistent with prior knowledge of the area. Overall, the validation results indicate that the FIsr method is a reliable approach for the identification and characterization of fractures in logging data.
(2) The influence of key factors in FIsr method on prediction results is clarified. The height of the window sliding affects the accuracy of fracture identification, with window heights of 2–5 intervals (0.25m–0.625m) showing the best performance. The data imbalance issue was addressed by undersampling, and an undersampling rate of 0.35 was found to be appropriate for achieving high accuracy rates and precision rates in fracture identification.

Enough labeled data is the basis of good performance of FIsr. However, labeled data is costly. Hence, in future work, semi-supervised learning will be considered into improving FIsr.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (Grant No. 42002134), China Postdoctoral Science Foundation (Grant No. 2021T140735).

References

Aghli, G., Moussavi-Harami, R., Tokhmechi, B., 2020. Integration of sonic and resistivity conventional logs for identification of fracture parameters in the carbonate reservoirs (A case study, Carbonate Asmari Formation, Zagros Basin, SW Iran). J. Petrol. Sci. Eng. 186, 106728 https://doi.org/10.1016/j.petrol.2019.106728.

Ala, M.A., Kinghorn, R.R.F., Rahman, M., 1980. Organic geochemistry and source rock characteristics of the Zagros petroleum province, southwest Iran. J. Petrol. Geol. 3 (1), 61–89, 10./j.1747-5457.1980.tb01004.x.

Al-Banna, N.Y., 2008. Oligocene/Miocene boundary in northern Iraq. GeoArabia 13 (2), 187–190.

Bahramali Asadi Kelishami, S., Mohebian, R., Salmian, O., 2022. A comprehensive perspective on pore connectivity and natural fracture analysis in Oligo-Miocene heterogeneous carbonates, southern Iran. J. Petrol. Sci. Eng. 208, 109199 https://doi.org/10.1016/j.petrol.2021.109199.

Bhattacharya, S., Mishra, S., 2018. Applications of machine learning for facies and fracture prediction using Bayesian network theory and random forest: case studies from the Appalachian Basin, USA. J. Petrol. Sci. Eng. 170, 1005–1017. https://doi.org/10.1016/j.petrol.2018.06.075.

Desouky, M., Tariq, Z., Aljawad, M.S., Alhoori, H., Mahmoud, M., Abdulraheem, A., 2021a. Machine learning-based propped fracture conductivity correlations of several shale formations. ACS Omega 6 (29), 18782–18792. https://doi.org/10.1021/acsomega.1c01919.

Desouky, M., Tariq, Z., Aljawad, M.S., Alhoori, H., Mahmoud, M., AlShehri, D., 2020. Data-Driven acid fracture conductivity correlations honoring different mineralogy and etching patterns. ACS Omega 5 (27), 16919–16931. https://doi.org/10.1021/acsomega.0c02123.

Desouky, M., Tariq, Z., Jawad, M.A., Alhoori, H., Mahmoud, M., Abdulraheem, A., 2021b. Development of machine learning based propped fracture conductivity correlations in shale formations. In: SPE Middle East Oil & Gas Show and Conference. https://doi.org/10.2118/204606-MS.

Dong, S.Q., Wang, Z.Z., Zeng, L.B., 2016. Lithology identification using kernel Fisher discriminant analysis with well logs. J. Petrol. Sci. Eng. 143, 95–102. https://doi.org/10.1016/j.petrol.2016.02.017.

Dong, S.Q., Zeng, L.B., Lyu, W.Y., Xu, C.S., Liu, J.J., Mao, Z., Tian, H., Sun, F.W., 2020a. Fracture identification by semi-supervised learning using conventional logs in tight sandstones of Ordos Basin, China. J. Nat. Gas Sci. Eng. 76, 103131. https://doi.org/10.1016/j.jngse.2019.103131.

Dong, S.Q., Zeng, L.B., Lyu, W.Y., Xia, D.L., Liu, G.P., Wu, Y., Du, X.Y., 2020b. Fracture identification and evaluation using conventional logs in tight sandstones: A case study in the Ordos Basin, China. Energy Geoscience 1 (3–4), 115–123. https://doi.org/10.1016/j.engeos.2020.06.003.

Dong, S.Q., Zeng, L.B., Liu, J.J., Gao, A., Lyu, W.Y., Du, X.Y., Yang, K.Y., Bao, M.Y., 2020c. Fracture identification in tight reservoirs by multiple kernel Fisher discriminant analysis using conventional logs. Interpretation 8 (4), 215–225. https://doi.org/10.1190/INT-2020-0048.1.

Dong, S.Q., Zeng, L.B., Du, X.Y., Bao, M.Y., Lyu, W.Y., Ji, C.Q., Hao, J.R., 2022a. An intelligent prediction method of fractures in tight carbonate reservoirs. Pet. Explor. Dev. 49 (6), 1364–1376. https://doi.org/10.1016/S1876-3804(23)60355-6.

Dong, S.Q., Sun, F.W., He, J., Sun, F.T., Zeng, L.B., 2022b. Development characteristics and main controlling factors of fractures in the carbonate reservoirs of Asmari Formation of A Oilfield, Iraq. Journal of Xi'an Shiyou University (Natural Science Edition) 37 (3), 1–19.

Dong, S.Q., Sun, Y.M., Xu, T., Zeng, L.B., Du, X.Y., Yang, X., Liang, Y., 2023a. How to improve machine learning models for lithofacies identification by practical and novel ensemble strategy and principles. Pet. Sci. 20 (2), 733–752. https://doi.org/10.1016/j.petsci.2022.09.006.

Dong, S.Q., Zhong, Z.H., Cui, X.H., Zeng, L.B., Sun, Y.M., Hao, J.R., 2023b. A deep kernel method for lithofacies identification using conventional well logs. Pet. Sci. 20 (3), 1411–1428. https://doi.org/10.1016/j.petsci.2022.11.027.

Gamal, M., El-Araby, A.A., El-Barkooky, A.N., Hassan, A., 2022. Detection and characterization of fractures in the Eocene Thebes formation using conventional well logs in October field, Gulf of Suez, Egypt. Egypt. J. Pet. 31 (3), 1–9. https://doi.org/10.1016/j.ejpe.2022.06.001.

Ghanadian, M., Faghih, A., Fard, I.A., Grasemann, B., Soleimany, B., Maleki, M., 2017. Tectonic constraints for hydrocarbon targets in the dezful embayment, Zagros fold and thrust belt, SW Iran. J. Petrol. Sci. Eng. 157, 1220–1228. https://doi.org/10.1016/j.petrol.2017.02.004.

Laubach, S.E., Lander, R.H., Criscenti, L.J., Anovitz, L.M., Urai, J.L., Pollyea, R.M., Hooker, J.N., Narr, W., Evans, M.A., Kerisit, S.N., Olson, J.E., Dewers, T., Fisher, D., Bodnar, R., Evans, B., Dove, P., Bonnell, L.M., Marder, M.P., Pyrak Nolte, L., 2019. The role of chemistry in fracture pattern development and opportunities to advance interpretations of geological materials. Rev. Geophys. 57 (3), 1065–1111. https://doi.org/10.1029/2019RG000671.

Le Garzic, E., Vergés, J., Sapin, F., Saura, E., Meresse, F., Ringenbach, J.C., 2019. Evolution of the NW Zagros Fold-and-Thrust Belt in Kurdistan Region of Iraq from balanced and restored crustal-scale sections and forward modeling. J. Struct. Geol. 124, 51–69. https://doi.org/10.1016/j.jsg.2019.04.006.

Li, Y.J., Song, L.H., Tang, Y.J., Zuo, J.P., Xue, D.J., 2022. Evaluating the mechanical properties of anisotropic shale containing bedding and natural fractures with discrete element modeling. Int. J. Coal Sci. Technol. 9 (1) https://doi.org/10.1007/s40789-022-00473-5.

Liu, S.S., Wang, Z.M., 2022. Reservoir grain size profile prediction of multiple sampling points based on a machine learning method. Petroluem Science Bulletin 7 (1), 93–105. https://doi.org/10.3969/j.issn.2096-1693.2022.01.009.

Luo, S.C., Tan, X.C., Chen, L., Li, F., Chen, P.Y., Xiao, D., 2019. Dense brine refluxing: a new genetic interpretation of widespread anhydrite lumps in the Oligocene–Lower Miocene Asmari Formation of the Zagros foreland basin, NE Iraq. Mar. Pet. Geol. 101, 373–388. https://doi.org/10.1016/j.marpetgeo.2018.12.005.

Mazhari, S.M., Memarian, H., Tokhmechi, B., 2018. A hybrid learning automata and case-based reasoning for fractured zone detection using petrophysical logs. Arab. J. Geosci. 11 (19), 577. https://doi.org/10.1007/s12517-018-3934-3.

Nouri-Taleghani, M., Mahmoudifar, M., Shokrollahi, A., Tatar, A., Karimi-Khaledi, M., 2015. Fracture density determination using a novel hybrid computational scheme: a case study on an Iranian Marun oil field reservoir. J. Geophys. Eng. 12 (2), 188–198. https://doi.org/10.1088/1742-2132/12/2/188.

Rashid, F., Hussein, D., Lawrence, J.A., Khanaqa, P., 2020. Characterization and impact on reservoir quality of fractures in the Cretaceous Qamchuqa Formation, Zagros folded belt. Mar. Pet. Geol. 113, 104117 https://doi.org/10.1016/j.marpetgeo.2019.104117.

Samarkin, Y., Aslanov, E., Tariq, Z., Al-Jawad, M., Alafnan, S., 2021. Extension of unified fracture design (UFD) concept to naturally fractured formations. EAGE. 6(82nd EAGE Conference and Exhibition 2021). https://doi.org/10.3997/2214-4609.202112933.

Shalaby, M.R., Islam, M.A., 2017. Fracture detection using conventional well logging in carbonate Matulla Formation, Geisum oil field, southern Gulf of Suez, Egypt. J. Pet. Explor. Prod. Technol. 7 (4), 977–989. https://doi.org/10.1007/s13202-017-0343-1.

Shi, G.R., 2008. Superiorities of support vector machine in fracture prediction and gassiness evaluation. Pet. Explor. Dev. 35 (5), 588–594. https://doi.org/10.1016/S1876-3804(09)60091-4.

Tariq, Z., Yan, B., Sun, S., Gudala, M., Aljawad, M.S., Murtaza, M., Mahmoud, M., 2022a. Machine learning-based accelerated approaches to infer breakdown pressure of several unconventional rock types. ACS Omega 7 (45), 41314–41330. https://doi.org/10.1021/acsomega.2c05066.

Tariq, Z., Murtaza, M., Mahmoud, M., Aljawad, M.S., Kamal, M.S., 2022b. Machine learning approach to predict the dynamic linear swelling of shales treated with different waterbased drilling fluids. Fuel 315, 123282–j.fuel.2022.123282.

Tian, M., Li, B.T., Xu, H.M., Yan, D.Z., Gao, Y.N., Lang, X.Z., 2021. Deep learning assisted well log inversion for fracture identification. Geophys. Prospect. 69 (2), 419–433. https://doi.org/10.1111/1365-2478.13054.

Wang, H.Q., Qi, M.M., Sun, F.T., Xie, H.L., 2018. The study of characteristic and genetic model for the anhydrite in Asmari Formation in A oilfield, Iraq. Adv. Geosci. 8 (4), 748–755. https://doi.org/10.12677/AG.2018.84080.

Xue, Y.C., Cheng, L.S., Mou, J.Y., Zhao, W.Q., 2014. A new fracture prediction method by combining genetic algorithm with neural network in low-permeability reservoirs. J. Petrol. Sci. Eng. 121, 159–166. https://doi.org/10.1016/j.petrol.2014.06.033.

Yuan, L., Xin, Y., Wu, S.Y., Yu, B., Luo, D.T., Ma, X.D., 2021. Research on qualitative identification, parameter modeling and control factors of cracks in deep Crelaceoustight sandstone: taking the Cretaceous Bashijiqike Formation reservoir in Keshen Area, Kuqa Depression.Tarim Basin as an example. J. Northeast Petrol. Univ. 45 (1), 20–31. https://doi.org/10.3969/j.issn.2059-4107.2021.01.003.

Zazoun, R.S., 2013. Fracture density estimation from core and conventional well logs data using artificial neural networks: the Cambro-Ordovician reservoir of Mesdar oil field, Algeria. J. Afr. Earth Sci. 83, 55–73. https://doi.org/10.1016/j.jafrearsci.2013.03.003.

Zeng, L.L., Ren, W.J., Shan, L.Q., Huo, F.C., 2022. Well logging prediction and uncertainty analysis based on recurrent neural network with attention mechanism and Bayesian theory. J. Petrol. Sci. Eng. 208, 109458 https://doi.org/10.1016/j.petrol.2021.109458.

Zhou, X.Q., Zhang, Z.S., Zhang, C.M., 2021. Bi-LSTM deep neural network reservoir classification model based on the innovative input of logging curve response sequences. IEEE Access 9, 19902–19915. https://doi.org/10.1109/ACCESS.2021.3053289.