Petroleum Science 23 (2026) 6941–6959

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

# Fault and fracture distribution prediction in continental shale based on graph neural network

Guo-Qing Lu [a,b], Lian-Bo Zeng [a,*], Guo-Ping Liu [c], Zhen-Lin Wang [d], Yu-Xiao Wu [d], Shao-Qun Dong [e], José Luis Drummond Alves [b], Jun-Feng Zhao [f], Mehdi Ostadhassan [g], Jian Su [h]

[a] College of Geosciences, China University of Petroleum (Beijing), Beijing, 102249, China
[b] Civil Engineering Department, Federal University of Rio de Janeiro, Rio de Janeiro, 21941-853, Brazil
[c] School of Energy Resources, China University of Geosciences (Beijing), Beijing, 100083, China
[d] PetroChina Xinjiang Oilfield Company, Karamay, 834000, Xinjiang, China
[e] College of Science, China University of Petroleum (Beijing), Beijing, 102249, China
[f] CNPC Brasil, Rio de Janeiro, 22250-206, Brazil
[g] Institute of Geosciences, Christian-Albrechts-Universität, Kiel, 24118, Germany
[h] Nuclear Engineering Department, Federal University of Rio de Janeiro, Rio de Janeiro, 21941-972, Brazil

ARTICLE INFO

*Article history:*
Received 8 April 2025
Received in revised form
23 January 2026
Accepted 1 July 2026
Available online 4 July 2026

Edited by Xiu-Fang Hu

*Keywords:*
Fault–fracture system
Graph neural network
Graph construction
Continental shale

ABSTRACT

The fault–fracture system plays a crucial role in the migration, accumulation, and preservation of shale oil. However, weak geophysical response of these fractures and the ambiguity in picking faults and natural fractures make their identification and recognition challenging. To address these issues, this paper proposes a set of methods for predicting the distribution of fault–fracture systems based on graph neural networks. The method set consists of three parts: (1) a logging-based fracture identification method via graph pooling and global graph construction, which uses logging data sampling points to create small-scale graphs that simulate fracture distribution in the wellbore and integrates lithology, stratigraphy, and mechanical stratigraphy information into the global graph structure and fracture identification model through depth sequences, spatial distance calculations, and global graph convolution; (2) a method combining semantic segmentation with graph convolution, which adds fault graph structures and graph computation modules to the U-Net network to expand the receptive field and clarify segmentation boundaries of the model, and constructs a joint loss function to simultaneously train and optimize the fault segmentation and fault continuity analysis modules; (3) a fracture intensity prediction method combining fault damage zone graph construction with ensemble learning, which effectively integrates the controlling effect of faults on fractures into the inter-well fracture distribution prediction model through information transfer between vertices. The proposed method set was applied to the Fengcheng Formation in the Mabei area of the Junggar Basin, achieving fracture identification accuracy of 91.96% through conventional logging data, a fault detection accuracy of 97.38%, and a mean absolute error of tectonic fracture distribution prediction of 9.38%, representing a 22.47% improvement over existing technologies. Overall, the research findings provide a new approach for predicting the distribution of fault–fracture systems in shale oil reservoirs.
© 2026 Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

With the shift in oil and gas exploration and development from conventional to unconventional resources, shale oil and gas have received a significant research attention while they have become a major contributor to the growth in hydrocarbon production globally (Ghosh et al., 2018; Gong et al., 2021, 2025; Gou et al., 2021a; Lu et al., 2024a). Their large-scale commercial development has transformed the global energy supply structure as well. In China, major breakthroughs have been achieved in basins such as the Junggar, Ordos, Santanghu, Songliao, Bohai Bay, and Qaidam, with the proven and predicted reserves of shale oil steadily increasing year by year (Gou et al., 2019, 2021b; Wang et al., 2022;

* Corresponding author.
*E-mail address:* lbzeng@sina.com (L.-B. Zeng).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2026.07.001
1995-8226/© 2026 Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Du et al., 2023; Liu et al., 2023a, 2023b; Su et al., 2025). However, shale oil reservoirs typically exhibit low porosity and low permeability characteristics, where fault–fracture systems play a crucial role in the migration, accumulation, and preservation of shale oil (Zeng et al., 2012, 2021; Liu et al., 2023b, 2025; Lu et al., 2024b; Du et al., 2025). In this regard, understanding the distribution of fault–fracture systems is essential for the effective exploration and development of shale oil reservoirs.

Faults can cause the repetition of formations in the wellbore whereby, on seismic images, they are characterized by discontinuous reflections (Cohen et al., 2006; Alonaizi et al., 2013; Hale, 2013). By interpreting the offset features of reflections on seismic profiles or extracting attributes that reflect the continuity of the reflectors, the spatial distribution characteristics of faults can be described in three dimensions (Wu et al., 2019; Robledo Carvajal et al., 2023; Alves et al., 2025; Lu et al., 2025a). Fault activity often leads to the development of tectonic fractures, which typically exhibit features such as hydrocarbon traces, or mineral fillings on their surfaces (Ma et al., 2023; Yao et al., 2023). In image logs, these fractures usually appear as high-resistance or low-resistance sinusoidal curves (Ja'fari et al., 2012; Dong et al., 2020; Lu et al., 2024a). In conventional logging, they can cause cycle skipping in the acoustic log and anomalous values in the density log (Tokhmechi et al., 2009; Aghli et al., 2016; Lu et al., 2024a). Fracture zones composed of multiple tectonic fractures can also alter the travel time, velocity, amplitude, frequency, and azimuth of seismic waves in anisotropic media (Chen et al., 2021; Shang et al., 2021). Therefore, through the joint training of multi-source heterogeneous data, the characteristics and distribution of fault–fracture system can be described in detail. Inversion of microseismic source mechanisms represents an additional valuable approach for fault and fracture characterization (Hui et al., 2023a, 2023b). For instance, the Unconventional Fracture Model (UFM) has been employed to simulate the dynamic propagation of hydraulic fractures and their complex interactions with natural fracture networks (Hui et al., 2024).

In recent years, with the continuous advancement of artificial intelligence, the use of machine learning methods for big data analysis and prediction in well log interpretation and seismic inversion has gradually become a mainstream trend (Wu et al., 2019; Dong et al., 2020, 2023; Shi et al., 2024; Lu et al., 2025b). Compared to traditional methods, these approaches have shown improvements in both interpretation efficiency and prediction accuracy (Lu et al., 2023). Currently, several major machine learning methods are being applied for fault and fracture prediction, such as Convolutional Neural Networks (CNNs) (Gu et al., 2018), Support Vector Machines (Cortes and Vapnik, 1995), Kernel Fisher Discriminant Analysis (Mika et al., 1999), Bayesian Networks (Cooper and Herskovits, 1992), K-Nearest Neighbors (Zhang et al., 2017), and Random Forests (Breiman, 2001). However, these methods were not specifically developed for fault–fracture system studies, leading to uncertainties when applied to fracture interpretation. For example, the limited receptive field of CNNs makes it difficult to extract global features of faults, and the models often fail to consider geological factors influencing fracture development. Additionally, they struggle to effectively address data imbalance, resulting in low precision and recall for fracture prediction. To address these shortcomings, it is essential to develop new methods for fault–fracture system prediction, expand the receptive field of fault segmentation techniques, and effectively integrate geological knowledge into intelligent fracture identification models, all of which are crucial for enhancing prediction accuracy.

Graph Neural Networks (GNNs) are a type of neural network designed for processing graph-structured data (Scarselli et al., 2009). By establishing a fault–fracture prediction system based on GNNs and incorporating geological factors influencing fault–fracture system development into the graph structure, geological entities can be abstracted as graphs, making it easier to capture global features of these inputs and effectively enhance model prediction performance and generalization capability of the model. Therefore, this study utilizes multi-source heterogeneous data, including core observations, image logs, conventional logging data, and 3D seismic data, to study fault–fracture system distribution prediction methods based on GNNs. This approach is divided into three parts: conventional logging fracture identification, seismic fault detection, and inter-well fracture distribution prediction.

First and foremost, conventional logging fracture identification is the foundation for fault–fracture system evaluation and prediction. This paper proposes a conventional logging fracture identification method based on the GNNs. The method uses core and image log interpretation results as labels, and conventional logging as original data for model training. By constructing small-scale graphs from conventional logging data sampling points, and through deep sequence analysis, spatial distance calculation, and global graph convolution, lithology and mechanical stratigraphy information are integrated into the global graph structure and fracture identification model.

Secondly, seismic fault detection results are critical constraints for fault–fracture system evaluation and prediction. This paper introduces a method combining semantic segmentation with graph convolution. By incorporating fault graph structures and graph computation modules into the U-Net method, the receptive field of the model is expanded, and segmentation boundaries are clarified. A joint loss function is constructed to simultaneously train and optimize the fault segmentation and fault continuity analysis modules.

Finally, building on the conventional logging fracture identification approach and seismic fault detection, this paper proposes an inter-well fracture intensity prediction method that integrates fault damage zone graph construction with ensemble learning. The method transforms fault and damage zone into graph data, using an unsupervised graph computation method to generate embedding matrices as input data for ensemble learning. The controlling effect of faults on fractures is effectively incorporated into the model through information propagation between vertices.

This methodology holds significant theoretical and practical importance for improving fault and tectonic fracture prediction technologies, clarifying the vertical and horizontal distribution of fault–fracture systems, to guide effective exploration and development of shale oil in the future. To validate the accuracy of the method, the model was applied to the data that was collected from the shale oil reservoirs of the Fengcheng Formation in the Mahu Sag, Junggar Basin, northwestern China.

## 2. Dataset and methodology

### 2.1. Basic idea of the fault–fracture system distribution prediction method

Conventional logging–based fracture identification provides a detailed characterization of the vertical distribution of tectonic fractures within each well, and these well-level identification results are also used to label seismic data for subsequent inter-well fracture prediction. Seismic fault-detection results, in turn, delineate the spatial distribution of faults across the study area. Although the seismic fault detection workflow is not directly coupled with the well-log-based fracture identification, its outputs serve as structural constraints for the inter-well fracture prediction process. Consequently, the inter-well prediction integrates information derived from both log-based fracture identification and seismic fault detection, enabling a more comprehensive delineation of the fracture distribution between wells. The

workflow for fault–fracture system distribution prediction can be divided into the following steps: (1) fracture identification in key wells. Based on core, image logs, and conventional logging data, the GNN-based fracture identification method is applied to study the vertical distribution of fractures in key wells within the study area. Using synthetic seismograms, these key wells are converted from the depth domain to the time domain. Fracture intensity is calculated based on a predefined sampling interval thickness and is used to label the 3D seismic data. (2) Fault detection and gridding: the study area is gridded, and a GNN-based fault detection method is applied to analyze the distribution of faults using 3D seismic data. Herein, detected faults are then used as constraints in the tectonic fracture distribution prediction model. (3) Fault damage zone analysis: width of the fault damage zones is calculated using the cumulative frequency method of variance attributes. Connections are established between fault vertices, damage zone vertices, and between damage zone vertices and fault vertices based on geodesic distance, with different weights assigned to these connections. (4) Inter-well fracture distribution prediction: seismic attributes that reflect fracture development are employed as a feature matrix, which is matched with the fault damage zone graph structure. Through unsupervised graph computation, the generated embedding matrix is used as input data to train the fracture distribution prediction model.

### 2.2. Conventional logging fracture identification

The conventional fracture identification method consists of three main modules: small-scale graph pooling, global graph construction, and vertex classification (Fig. 1). Small-scale graph pooling module constructs small-scale graphs by sliding a window along the depth axis of the well log. Within each window, five consecutive well log sampling points are selected as vertices and connected sequentially based on their depth. Convolution and pooling operations are performed on these small-scale graphs to filter and integrate well log information. The graph pooling layer employs the gPool layer (Ronneberger et al., 2015), which retains vertices containing important information while discarding others, achieving information integration and downsampling. After pooling, each small-scale graph is represented by a new vertex and a one-dimensional feature vector.

These new vertices and feature vectors are used to construct a global graph. Assuming a single lithology layer contains $K+1$ vertices at consecutive depths, the set of vertices in that lithology layer $\mathscr{L}_{mn}$ is defined as:

$$\mathscr{L}_{mn} = \{v_P, v_{P+1}, ..., v_{P+K}\}, \tag{1}$$

where $v_P, v_{P+1}, ..., v_{P+K}$ represents the vertices within a single lithology layer, and the set of vertices with the same lithology is given by

$$\mathscr{L}_m = \bigcup_{n=1}^{G} \mathscr{L}_{mn}, \tag{2}$$

where $G$ denotes the number of single lithology layers within each lithology type. The complete set of lithology vertices $\mathscr{L}$ is thus given by

$$\mathscr{L} = \bigcup_{m=1}^{H} \bigcup_{n=1}^{G} \mathscr{L}_{mn}, \tag{3}$$

where $H$ denotes the total number of lithology categories.

First, the vertices are sorted sequentially by depth, and connections are established between adjacent vertices, with the element $a_{ij}^1$ given by:

$$a_{ij}^1 = a_{ji}^1 = \begin{cases} 1, & \text{if } j = i-1 \text{ or } j = i+1 \\ 0, & \text{if } j < i-1 \text{ or } j > i+1 \end{cases}, \quad \forall i,j \in \mathscr{L}, \tag{4}$$

when $a_{ij}^1$ equals 1, it indicates the presence of an edge between vertex $i$ and vertex $j$; when $a_{ij}^1$ equals 0, it specifies the absence of an edge between vertex $i$ and vertex $j$.

Next, based on the Euclidean distance between vertices within each lithology category, an appropriate quantile $\delta_d$ is selected. Connections between vertices are determined in ascending order of Euclidean distance, with the matrix element $a_{ij}^2$ given by:

$$a_{ij}^2 = a_{ji}^2 = \begin{cases} 1, & \text{if } d_{ij} < \delta_d \\ 0, & \text{if } d_{ij} \geq \delta_d \end{cases}, \quad \forall i,j \in \mathscr{L}_m, \tag{5}$$

when $a_{ij}^2$ equals 1, the presence of an edge between vertex $i$ and vertex $j$ is observed and when $a_{ij}^2$ equals 0, the absence of an edge between vertex $i$ and vertex $j$ is recognized.

Fracture identification is achieved by classifying vertices in the global graph using GNNs (Hamilton et al., 2017; Kipf and Welling, 2017). Macroscopically, graph convolution integrates information from the graph structure and the feature matrix, while microscopically, it represents an information propagation process between adjacent vertices. The process begins with the first information integration, where the graph convolutional layer combines the graph structure and feature matrix $\boldsymbol{X}$ to produce the embedding matrix $\boldsymbol{H}^1$. The activated $\boldsymbol{H}^1$ is then used as input for the second integration, generating the embedding matrix $\boldsymbol{H}^2$. This iterative process continues through multiple graph convolutions, resulting in the final embedding matrix $\boldsymbol{H}^n$. The classification result is obtained by applying the softmax function to $\boldsymbol{H}^n$. The graph convolution process can be expressed by the following equation:

$$\boldsymbol{h}_v^{(k+1)} = \text{Dropout}\left(\left(\sigma\left(\boldsymbol{W}^{(k)} \cdot \text{CONCAT}\left(\boldsymbol{h}_v^{(k)}, \text{AGG}_{u \in \mathscr{N}(v)}\left(\boldsymbol{h}_u^{(k)}\right)\right)\right) + \boldsymbol{b}^{(k)}\right), p_{\text{drop}}\right), \tag{6}$$

where $\mathscr{N}(v)$ represents the neighboring vertices of the root vertex $v$. $\boldsymbol{h}_u^{(k)}$ is the representation of a neighboring vertex $u$ at the $(k)$-th layer. $\boldsymbol{h}_v^{(k)}$ is the representation of the root vertex itself at the $(k)$-th layer. $\boldsymbol{W}^{(k)}$ and $\boldsymbol{b}^{(k)}$ are the weight and bias of the $k$-th layer, respectively. $\sigma$ is the activation function. AGG $(\cdot)$ is the aggregation function and CONCAT$(\cdot)$ is the concatenation function. $p_{\text{drop}}$ is the dropout rate. This equation represents the message passing process of neighboring vertices and the updated representation process of root vertex information.

The conventional logging fracture identification results are used to calculate the fracture intensity for each segment based on a uniform depth sampling interval, evaluating the degree of fracture development. The fracture intensity $Q$ is defined as follows:

$$Q = \frac{B_f}{B}, \tag{7}$$

where $B$ denotes the length of a continuous depth interval in the wellbore, and $B_f$ is the length of the fracture segment developed within the continuous depth interval $B$.

### 2.3. Seismic fault detection

The fault detection method primarily consists of two modules: fault segmentation and continuity analysis (Fig. 2). The fault

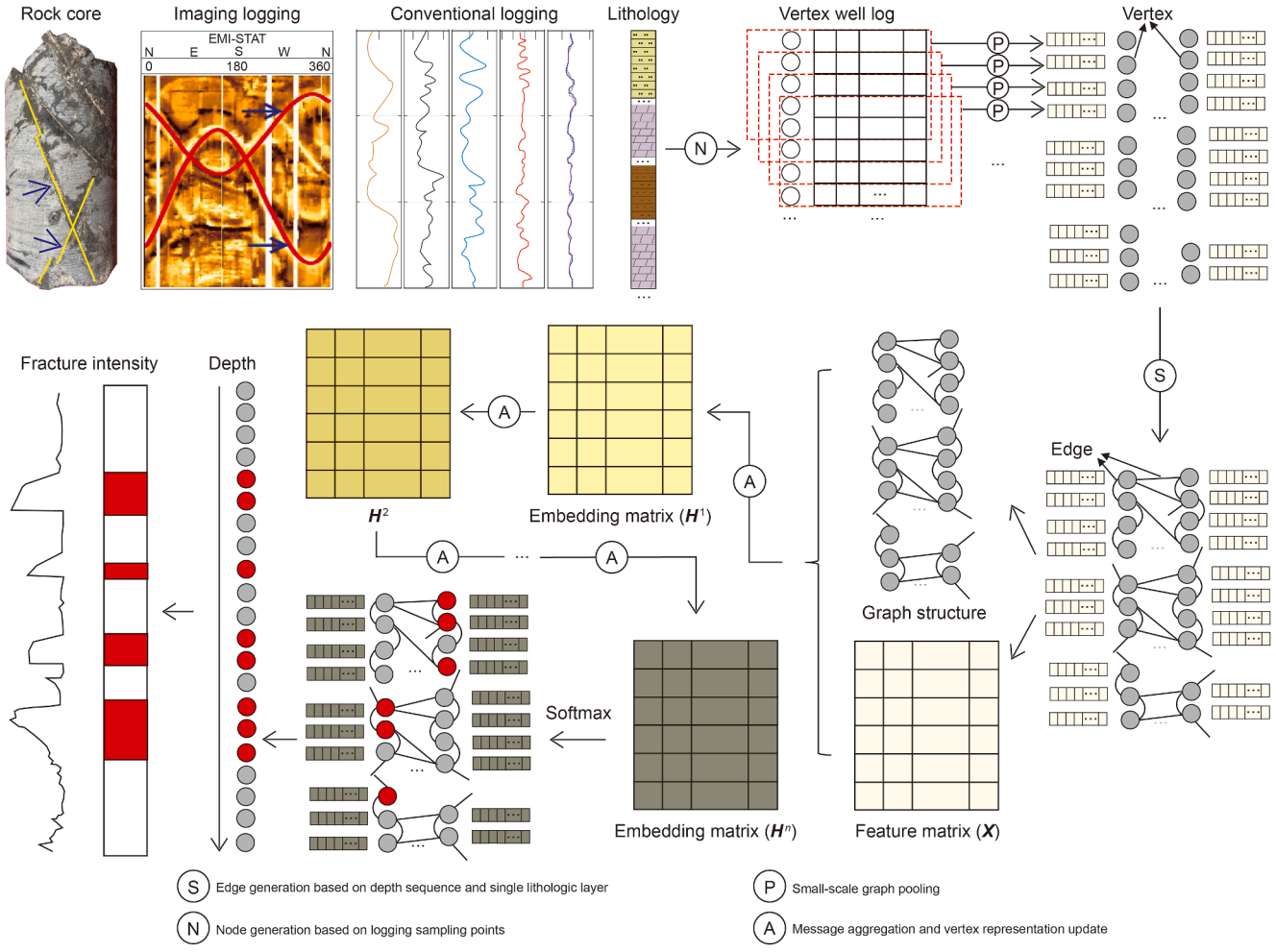



**Fig. 1.** The process of conventional logging fracture identification.

segmentation module employs a simplified U-Net, which includes an encoder–decoder network structure (Ronneberger et al., 2015). The encoding process can be expressed as:

$$\boldsymbol{X}_{n+1} = \mathrm{MaxPool3D}(\mathrm{Conv3D}_{x2}(\boldsymbol{X}_n)), \tag{8}$$

where $\boldsymbol{X}_n$ represents the input features of the $n$-th layer, $\mathrm{Conv3D}_{x2}$ denotes the two consecutive 3D convolutional layers and activation function, and MaxPool3D represents the 3D max-pooling layer.

The feature map after decoding at the $n$-th layer $\boldsymbol{E}_n$ during the decoding process can be expressed as:

$$\boldsymbol{E}_n = \mathrm{Conv3D}_{x2}(\mathrm{UpSample}(\boldsymbol{E}_{n+1}) \oplus \boldsymbol{X}_n), \tag{9}$$

where UpSample denotes the upsampling operation, and $\oplus$ signifies the feature concatenation operation.

Fault continuity analysis module calculates the fault graph to extract global fault features and analyze fault continuity. A coordinate system is established for the 3D fault labels, and the data is gridded. Given the 3D fault label function $f : \mathbb{Z}^3 \to \mathbb{R}$, each grid cube $f : \mathbb{Z}^3 \to \mathbb{R}$ is defined, and the gridded data volume has dimensions of $16 \times 16 \times 16$. The coordinates of each vertex are selected as follows:

$$v_{\mathrm{pos}} = \begin{cases} \left(x + \dfrac{w_x}{2}, y + \dfrac{w_y}{2}, z + \dfrac{w_z}{2}\right), & if \sum_{(x_i,y_j,z_k)\in B} f\left(x_i, y_j, z_k\right) = 0 \\ (x + \Delta x_{\max}, y + \Delta y_{\max}, z + \Delta z_{\max}), & if \sum_{(x_i,y_j,z_k)\in B} f\left(x_i, y_j, z_k\right) > 0 \end{cases}, \tag{10}$$

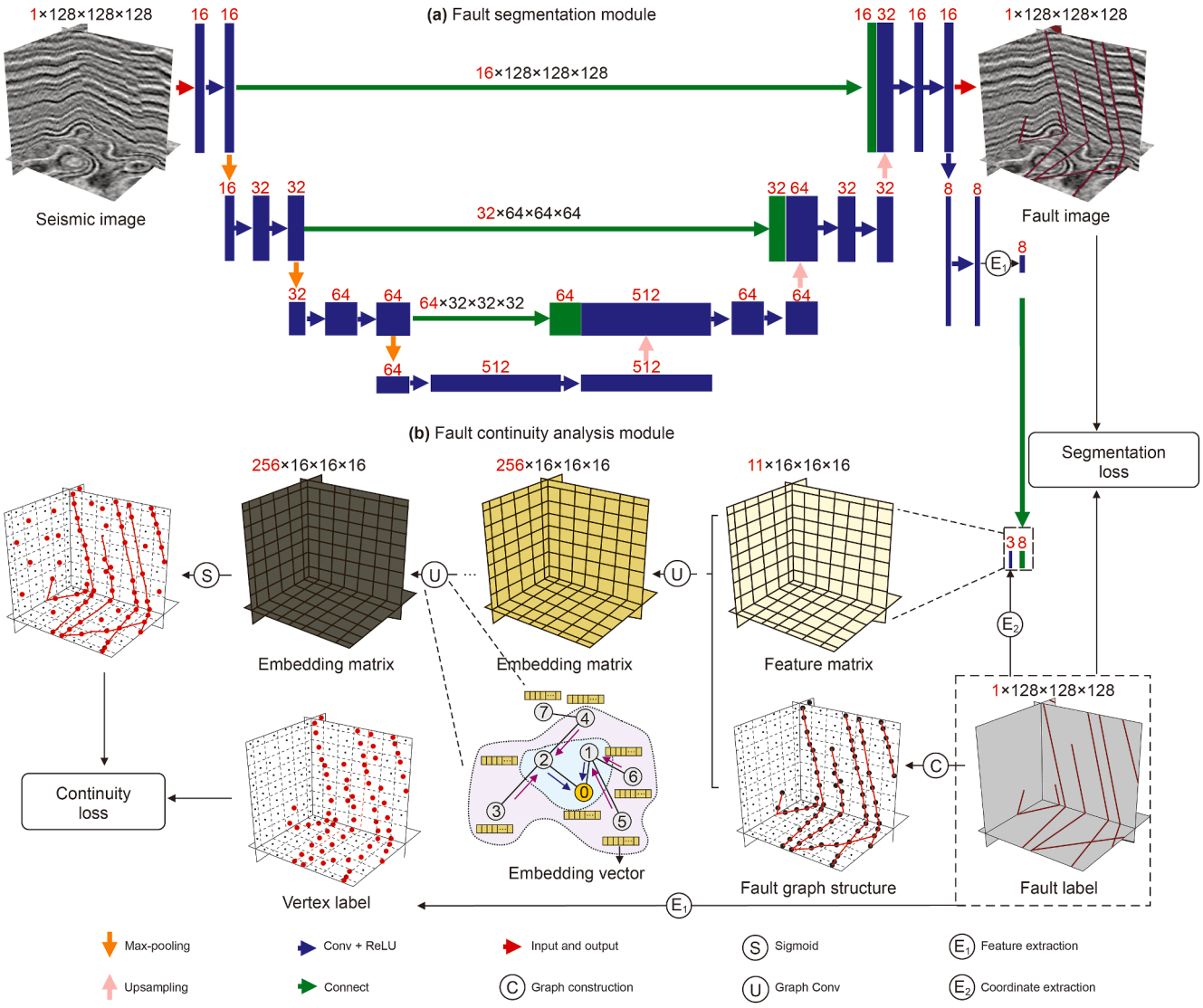



**Fig. 2.** The process of seismic fault detection (Lu et al., 2025a).

where $(x, y, z)$ represents the starting coordinates of the cube in the 3D array, $w_x, w_y, w_z$ denotes the width of the cube along each axis, $B$ signifies the non-empty cube, and $(\Delta x_{\max}, \Delta y_{\max}, \Delta z_{\max})$ is the relative positional offset that achieves the maximum value within $B$.

The dilation function is applied to expand the fault voxels, and a velocity field $V : \Omega \to \mathbb{R}^+$ is constructed:

$$V(x,y,z) = V_{\text{nofault}} + \left(V_{\text{fault}} - V_{\text{nofault}}\right) \cdot \chi_{\text{fault}}(x,y,z), \tag{11}$$

where $V_{\text{fault}}$ represents the velocity within the fault region, and $V_{\text{nofault}}$ denotes the velocity in the non-fault region. In this velocity field, the relationship $V_{\text{nofault}} < V_{\text{fault}}$ holds.

Define the source point $\boldsymbol{s} \in \Omega$ in the three-dimensional space $\Omega$, where $\Omega \subseteq \mathbb{R}^3$. The travel time $T$ from the source point $\boldsymbol{s}$ to any point $\boldsymbol{p} \in \Omega$ is calculated by solving the Eikonal equation:

$$\|\nabla T(\boldsymbol{p})\| = \frac{1}{V(\boldsymbol{p})}, \quad \boldsymbol{p} \in \Omega, \tag{12}$$

where the initial condition is set as $T(\boldsymbol{s}) = 0$. Within the fault region, $V = V_{\text{fault}}$, and in the non-fault region, $V = V_{\text{nofault}}$.

The set of vertices whose coordinates fall within the fault region is defined as $\mathscr{T}$. The travel time between vertex $m$ and vertex $n$ within $\mathscr{T}$ is denoted as $T_{mn}$. A threshold $\gamma_{FF}$ is defined, and the connection relationship between vertices is determined by leveraging the velocity difference inside and outside the velocity field:

$$a_{mn}^{FF} = a_{nm}^{FF} = \begin{cases} 1, & \text{if } T_{mn} < \gamma_{FF} \\ 0, & \text{if } T_{mn} \geq \gamma_{FF} \end{cases}, \quad \forall m, n \in \mathscr{T}, \tag{13}$$

where $a_{mn}^{FF}$ represents the element in the $m$-th row and $n$-th column of the adjacency matrix. When the travel time difference between two vertices is less than the threshold $\gamma_{FF}$, a connection is established. If $a_{mn}^{FF} = 1$, it indicates that a connection has been established between vertex $m$ and vertex $n$; if $a_{mn}^{FF} = 0$, it means no connection exists between vertex $m$ and vertex $n$.

The vertices and edges form the fault graph structure $G_{\text{fault}} = (V_{\text{fault}}, E_{\text{fault}})$. To construct a complete fault graph, a corresponding feature matrix must also be matched. The feature vector $f_i$ is

extracted from the coordinates of each vertex. To emphasize the spatial attributes of the fault, the spatial coordinates $(x_i, y_i, z_i)$ of the vertex are concatenated with the feature vector $f_i$, forming the feature representation of the vertex:

$$h_i = [f_i, x_i, y_i, z_i]. \tag{14}$$

The feature vectors of all vertices form the feature matrix. Using GNNs (Hamilton et al., 2017), computations are performed on the complete fault graph, which consists of the feature matrix and the graph structure, to achieve information aggregation from neighboring vertices. The graph convolution follows the same message-passing process defined in Eq. (6).

### 2.4. Inter-well fracture distribution prediction

The process of inter-well fracture distribution prediction is illustrated in Fig. 3. The fault zone consists of a fault core and the surrounding damage zone (Katz et al., 2003; Savage and Brodsky, 2011). The fault damage zone is a structural deformation zone where secondary faults and fractures develop intensively due to fault interactions (Choi et al., 2016; Yao et al., 2023). The variance value is similar to the degree of fracture development. A higher variance value indicates a stronger structural discontinuity. The boundaries and width of the fault damage zone can be estimated using the cumulative frequency curve of variance attributes (Ma et al., 2023). The average variance per unit distance is represented by histograms, while the cumulative variance is represented by discrete points. By analyzing the characteristics of the discrete points of the cumulative variance, the points of slope change in the line formed by these discrete points can be clearly identified, which represents the outer boundary of the damage zone (Fig. 4).

The essence of constructing the fault damage zone graph structure lies in using vertices and edges to describe the fault's orientation and the extent of the damage zone. A three-dimensional coordinate system is established for the seismic data volume, and the data is gridded. The positions where the fault intersects the grid are selected as vertices for each grid. The set of vertices belonging to the fault is defined as $\mathscr{F}$, and the set of vertices belonging to the damage zone is defined as $\mathscr{D}$.

A threshold $\delta_{FD}$ is set, and connections are established between damage zone vertices and fault vertices where the Euclidean distance is less than $\delta_{FD}$. The element in the $m$-th row and $n$-th column of the adjacency matrix, $a_{mn}^{FD}$, is defined as:

$$a_{mn}^{FD} = a_{nm}^{FD} = \begin{cases} 1, & \text{if } d_{mn} < \delta_{FD} \\ 0, & \text{if } d_{mn} \geq \delta_{FD} \end{cases}, \quad \forall m \in \mathscr{F} \text{ and } \forall n \in \mathscr{D}, \tag{15}$$

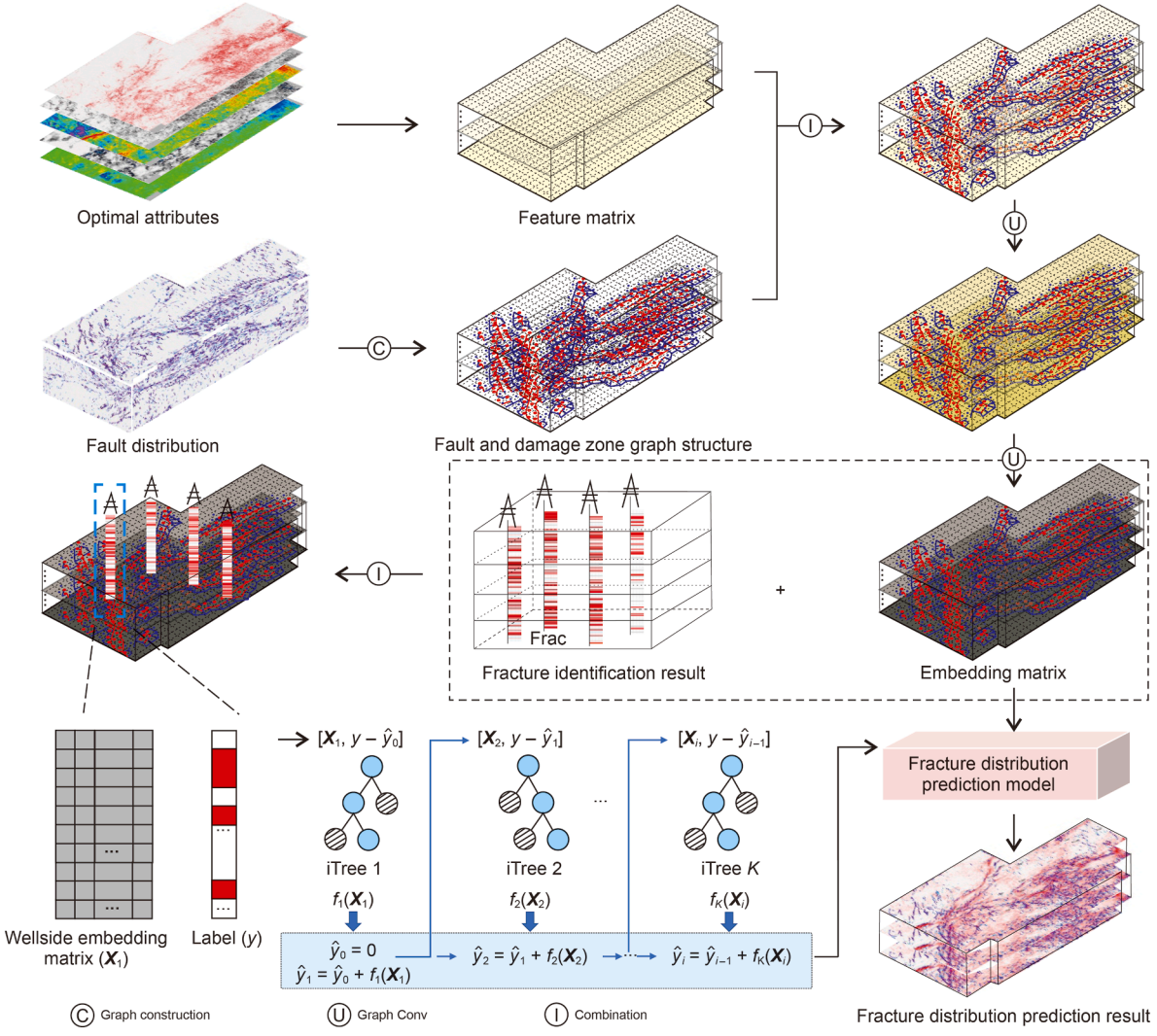



**Fig. 3.** The process of seismic fracture distribution prediction.

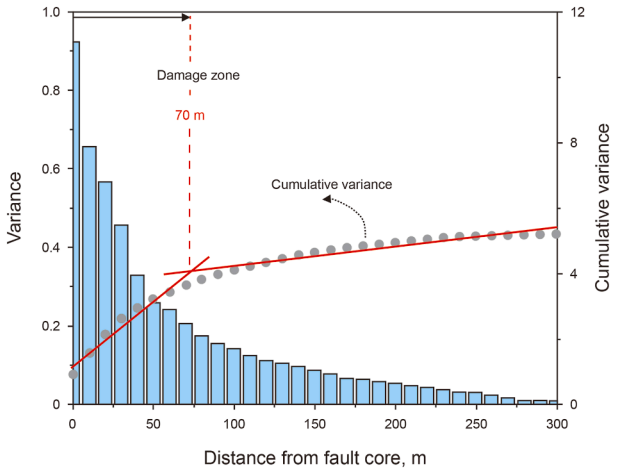



**Fig. 4.** The cumulative frequency method for the estimation of the damage zone width (Ma et al., 2023).

when $a_{mn}^{FD}=1$, it indicates the presence of an edge between vertex $m$ and vertex $n$; when $a_{mn}^{FD}=0$, it indicates the absence of an edge between vertex $m$ and vertex $n$. The set of fault vertices is denoted as $\mathscr{F}$, and the set of damage zone vertices is denoted as $\mathscr{D}$.

A threshold $\delta_{DD}$ is set, and connections are established between damage zone vertices where the Euclidean distance is less than $\delta_{DD}$. The element $a_{mn}^{DD}$ of the adjacency matrix is given as follows:

$$a_{mn}^{DD}=a_{nm}^{DD}=\begin{cases}1, & \text{if } d_{mn}<\delta_{DD}\\ 0, & \text{if } d_{mn}\geq\delta_{DD}\end{cases},\quad \forall m,n\in\mathscr{D}, \tag{16}$$

when $a_{mn}^{DD}=1$, means the presence of an edge between vertex $m$ and vertex $n$; while $a_{mn}^{DD}=0$, indicates the absence of an edge between vertex $m$ and vertex $n$. The set of damage zone vertices is denoted as $\mathscr{D}$.

The computation of the fault and damage zone graph is a process of information propagation between vertices. In the vertex aggregation process proposed in this study, edge weight information is incorporated. The global graph structure includes both the fault core and the damage zone, with different weights assigned to edges between fault vertices, edges between fault and damage zone vertices, and edges between damage zone vertices. These weights control the directional bias and the amount of information propagated during the aggregation process. The new method incorporates edge weight information into the vertex aggregation process. For the aggregation of the root vertex "1" in the graph, the $(k+1)$-th layer can be expressed as follows:

$$\boldsymbol{h}_1^{(k+1)}=\sigma\left(\sum_{u\in\mathscr{N}_1}\omega_{1,u}\frac{\tilde{A}_{1,u}}{\sqrt{\tilde{D}_{1,1}\tilde{D}_{u,u}}}\boldsymbol{h}_u^{(k)}\right), \tag{17}$$

where $\omega_{1,u}$ represents the edge weight between the root vertex "1" and its neighboring vertex $u$. $\mathscr{N}_1$ represents the set of neighboring vertices of the root vertex "1", and $u$ is one of the neighboring vertices of the root vertex "1". $\tilde{A}_{1,u}$ refers to the element in the first row and the $u$-th column of the modified adjacency matrix. $\tilde{D}_{u,u}$ represents the degree of the neighboring vertex $u$, $\tilde{D}_{1,1}$ represents the degree of the root vertex "1", and $\boldsymbol{h}_u^{(k)}$ represents the feature representation of the $(k)$-th layer neighboring vertices.

The calculation of the fault damage zone graph aims to make the vertex feature representations at different fault locations, such as the fault core and various parts of the damage zone, become more consistent. At the same time, the inclusion of edge weights also takes into account the fault's influence on the fracture strength within the damage zone. In this section, the embedding matrix obtained after graph convolution is used as input data for fracture strength prediction, and an Extreme Gradient Boosting (XGBoost) model (Chen and Guestrin, 2016) is used to establish the fracture distribution prediction model.

The distribution prediction results $\hat{y}_i$ can be expressed as:

$$\hat{y}_i=\hat{y}_{i-1}+f_i(\boldsymbol{X}), \tag{18}$$

where $\boldsymbol{X}$ represents the input features and $f_i(\boldsymbol{X})$ represents the output of the $i$-th regression tree.

### 2.5. Performance evaluation indicators

Conventional logging fracture identification is a vertex classification task, evaluated using accuracy (Acc), precision (Pr), and recall (Re). Accuracy is defined as the ratio of correctly classified samples to the total number of samples (Bressan et al., 2020). High accuracy indicates good overall classification performance. However, since fracture identification involves imbalanced data classification, precision and recall are introduced to further evaluate the model's ability to distinguish between fracture and non-fracture samples (Lu et al., 2024a). Precision measures the proportion of correctly identified fractures among all predicted fractures, with higher precision indicating fewer non-fracture samples misclassified as fractures (Bressan et al., 2020; Lu et al., 2024a). Recall measures the proportion of correctly identified fractures among all actual fractures, with higher recall indicating fewer fracture samples misclassified as non-fractures (Bressan et al., 2020; Lu et al., 2024a). These metrics collectively provide a comprehensive assessment of the model's performance in fracture identification.

Fault detection combines image segmentation and vertex classification, evaluated using accuracy (Acc), Dice coefficient (DSC), specificity (Spe), and area under the curve (AUC). The Dice coefficient measures the similarity between the predicted fault regions and the actual fault regions, with a higher value indicating closer alignment between prediction and ground truth (Lu et al., 2025a). Specificity represents the ratio of correctly predicted non-fault pixels to the total number of actual non-fault pixels, reflecting the model's ability to avoid misclassifying non-fault regions as faults (Lu et al., 2025a). AUC evaluates the overall performance of the classification model, with a higher value indicating better ability to distinguish between fault and non-fault regions. These metrics ensure a robust evaluation of the model's fault detection capabilities (Lu et al., 2025a).

Inter-well fracture distribution prediction is a regression task, evaluated using mean squared error (MSE), coefficient of determination ($R^2$), and mean absolute error (MAE). MSE measures the average error in predicting fracture intensity, with lower values indicating higher prediction accuracy. The coefficient of determination ($R^2$) represents how well the model captures the variability in the data, with values closer to one suggesting that the predicted fracture distribution closely matches the actual distribution. MAE, which is less sensitive to outliers compared to MSE, measures the precision of fracture intensity predictions, with lower values indicating more accurate results. These metrics provide a comprehensive evaluation of the model's performance in predicting fracture distribution.

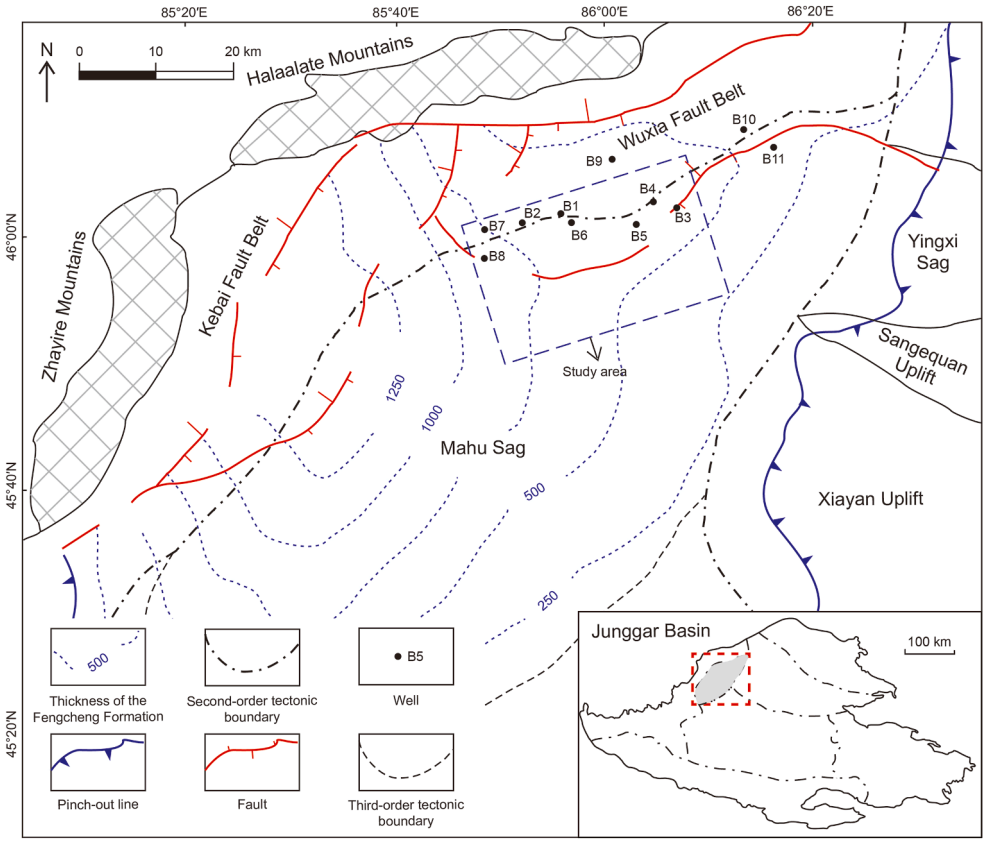



**Fig. 5.** Tectonic setting and well distribution of the study area in the Mahu Sag (Zhi et al., 2021).

## 3. Examples

### 3.1. Geological setting

The study area is situated on the northern slope of the Mahu Sag, located in the western Junggar Basin, China (Fig. 5). It is structurally influenced by the Kobei and Wuxia fault zones and is characterized by the presence of mudstone source rocks and diverse reservoir lithologies. These conditions have facilitated the formation of an integrated petroleum system, encompassing shale oil, tight oil, and conventional oil reservoirs (Wang et al., 2022). This area exemplifies a comprehensive hydrocarbon system, combining both conventional and unconventional oil and gas resources, and holds substantial potential for future exploration and development (Tang et al., 2021; Zhi et al., 2021). The Mahu Sag is a Carboniferous-Quaternary depression formed on a pre-Carboniferous folded basement, controlled by multi-stage peripheral thrust activities. In the Early Permian, under north–south extension, the main oil-generating formations of the Lower Permian in the northwestern margin of the Junggar Basin were deposited, accompanied by extensive basic volcanic activities. In the Late Permian, the compression and collision between the western Junggar Ocean and the Kazakhstan plate led to the formation of large-scale thrust structures along the basin's edge.

The study target is the Lower Permian Fengcheng Formation, which represents a mixed sedimentary environment from multiple source areas. The sources include volcanic materials, endogenous carbonates, and terrigenous clastics, leading to frequent interlayered structures (Tang et al., 2021; Zhi et al., 2021; Wang et al., 2022). From bottom to top, it is divided into three members: $P_1f^1$, $P_1f^2$ and $P_1f^3$. The $P_1f^1$ is characterized by shore-shallow lake and semi-deep lake facies sediments, the $P_1f^2$ is composed of semi-deep lake to shallow lake sediments, and the $P_1f^3$ is mainly dominated by semi-deep lake sediments.

### 3.2. Characteristics of fault–fracture system

The fault–fracture system consists of a fault core and the surrounding damage zone (Katz et al., 2003; Savage and Brodsky, 2011). The fault damage zone is a structural deformation zone where secondary faults and fractures develop intensively due to fault interactions (Choi et al., 2016; Yao et al., 2023). The width of the fault damage zone refers to the extent of the damage structures on both sides of the fault core, defined by the frequency

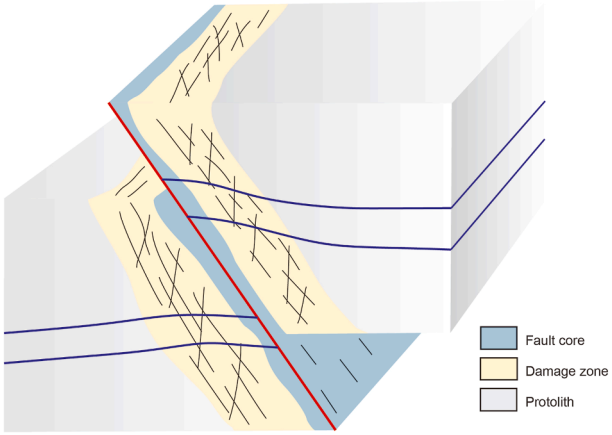



**Fig. 6.** Illustration of a reverse fault and damage zone.

distribution of damage structures per unit length. Typically, this width decreases with increasing distance from the fault core (Berg and Skar, 2005; Rempe et al., 2013; Schueller et al., 2013) (Fig. 6).

The study area exhibits frequent tectonic activity, with primary deformation styles categorized into extension, compression, transpression, and transtension. The main faults and secondary faults display multi-phase superposition characteristics, and the structural features vary at different depths. Shallower zones are predominantly characterized by left-lateral strike-slip, while deeper zones are dominated by right-lateral strike-slip. Within the fault damage zones, high-angle tectonic fractures, induced by fault activity, are commonly developed. These fractures typically occur in mudstone layers rich in brittle minerals and are characterized by a wide range of orientations. Step-like features are often observed on the fracture surfaces, which are either unfilled or filled with alkaline minerals.

### 3.3. Training process and parameter setting

(1) Conventional logging fracture identification

A detailed core description was conducted on a 258 m cored section of Well B1 in the study area. The fracture description results and image logs interpretation were used as training labels for the model, while the corresponding conventional logging data from the cored section were utilized for model training. All well log curves were reconstructed and standardized. The dataset was randomly split, with 70% used as the training set and the remaining 30% as the testing set. Additionally, Well B1 in the study area was selected as a blind well to validate the reliability of the model.

The model parameters are set as shown in Table 1. In the fracture identification model, the number of vertices in each small-scale graph was set to 5, and the number of small-scale graph convolutional layers was set to 1. The quantile used to construct edges between vertices of the same lithology was set to 1%. The model depth was set to 2, with full sampling of adjacent vertices. Based on previous studies, a total of 10 lithologies were identified in the study area: basalt, dolomite, dolomitic mudstone, ignimbrite, mudstone, silicalite, siltstone, silty mudstone, tuffaceous fine sandstone, and sedimentary tuff (Lu et al., 2023). The visualized global graph is an undirected graph, drawn using the Force-Directed Layout Algorithm, which simulates physical "forces" to arrange vertex distribution. In the graph, the size of the vertices represents their degree, with larger radius indicating higher degrees (i.e., more connected edges). Vertices of the same lithology are relatively concentrated, while vertices of different lithologies are more dispersed (Fig. 7).

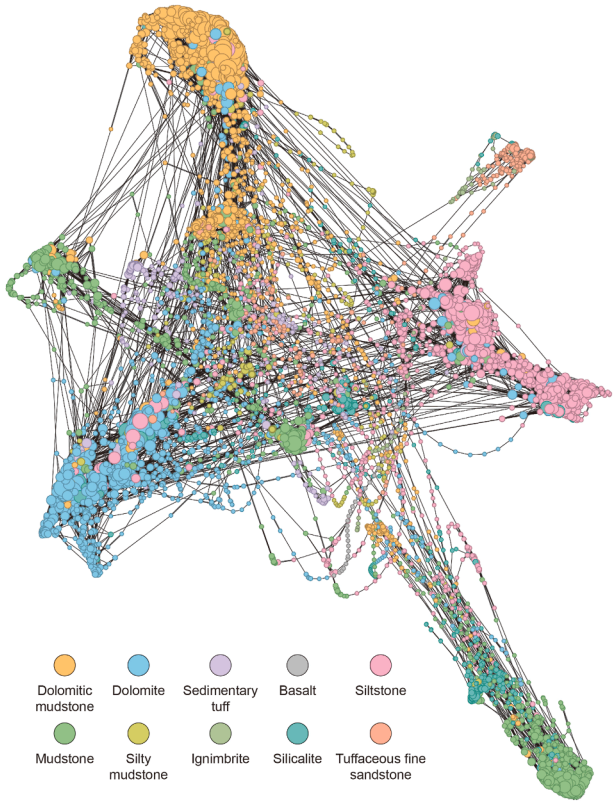



**Fig. 7.** Global graph visualization of conventional logging fracture identification model.

**Table 1**
Model parameter settings for the GNN-based improvement methods.

| Module | Parameter | Setting |
|---|---|---|
| Conventional logging fracture identification | Number of vertices in the small-scale graph | 5 |
| | Number of small-scale graph convolutional layers | 1 |
| | Quantile ($\delta_d$) | 1% |
| | Number of global graph convolutional layers | 2 |
| | Sampling rate | 100% |
| Seismic fault detection | Grid window size | 8 |
| | Convolution kernel size | 3×3×3 |
| | Pooling kernel size | 2×2×2 |
| | Stride of the convolution kernel | 1 |
| | Stride of the pooling kernel | 2 |
| | Number of graph convolution layers | 3 |
| | Threshold ($\gamma_{FF}$) | 10 |
| Inter-well fracture prediction | Equal-spacing sampling thickness | 12.5 m |
| | Number of regression trees | 500 |
| | Learning rate | 0.5 |
| | Max tree depth | 20 |

(2) Seismic fault detection

The data used for training the fault detection model are sourced from a publicly available dataset, which consists of multiple 3D seismic images and their corresponding fault labels. Each 3D seismic image has dimensions of 128×128×128 and a single channel. The dataset was randomly split, with 90% used as the training set and the remaining 10% as the testing set. Data augmentation is particularly important in model training as it significantly increases the variability of the training dataset. In this study, the data augmentation method involves rotating the 3D seismic images and their labels counterclockwise along the *Z*-axis by 90° increments, resulting in four augmented data volumes at 0°, 90°, 180°, and 270°. The fault graphs corresponding to each 3D seismic image are also rotated synchronously. The method employed here incorporates vertex coordinates into the fault graph as a feature matrix, enabling the transformation of vertex coordinates in the fault graph to achieve the rotation of the graph structure (Fig. 8).

The optimal parameter settings for the fault detection model are shown in Table 1. The gridding window is set to 8, the convolutional layer uses Conv3D with a kernel size of 3×3×3, the pooling kernel size is 2×2×2 with a stride of 2, and the graph convolutional layer employs SAGEConv (Hamilton et al., 2017)

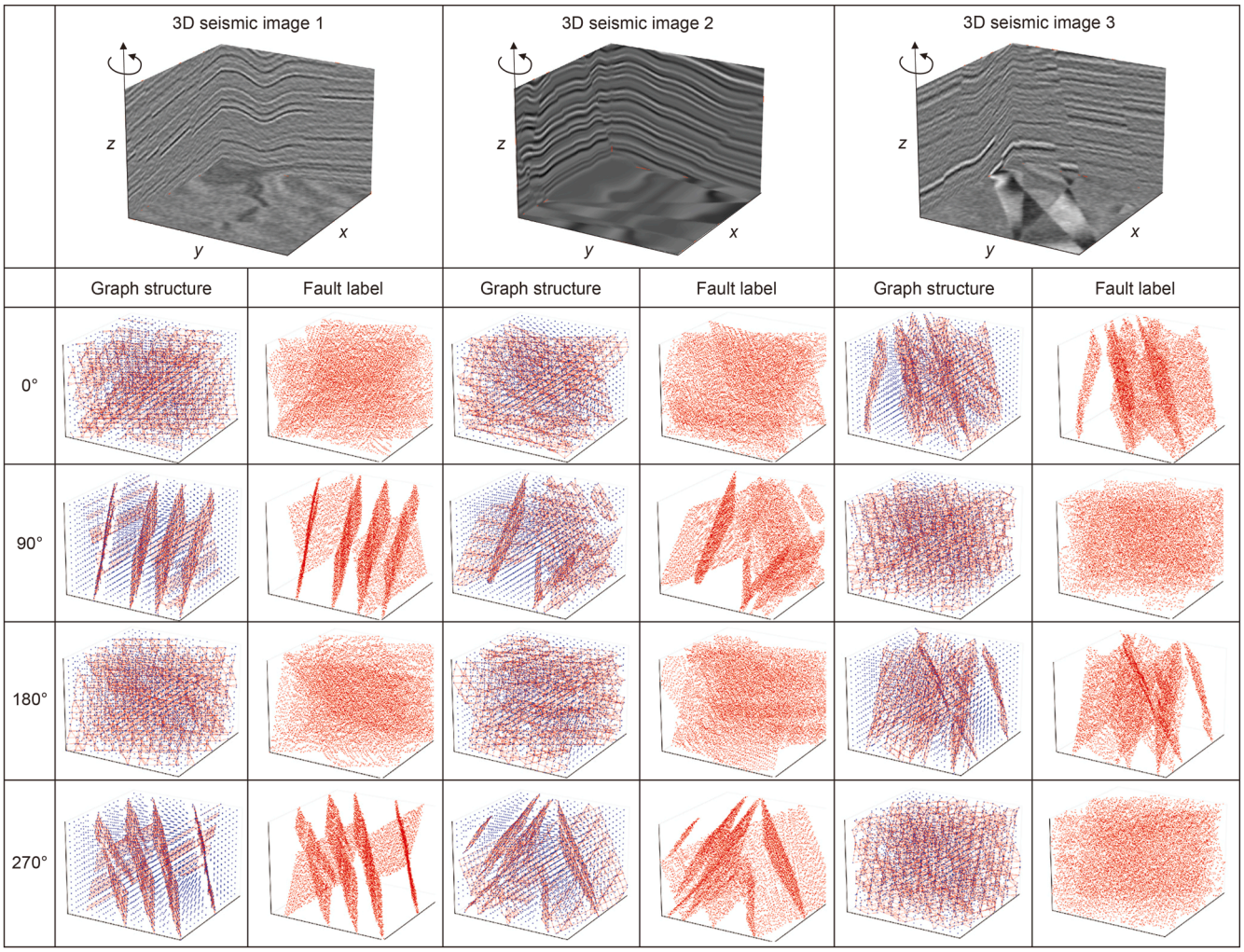



**Fig. 8.** Illustration of data augmentation for the fault label and graph structure.

with a total of 3 graph convolutional layers. The threshold parameter ($\gamma_{FF}$) is set to 10.

(3) Inter-well fracture distribution prediction

The Isolation Forest method (Liu et al., 2008) was employed to detect and remove outliers from the dataset, thereby reducing noise interference. Additionally, this study adopts a feature-selection method based on XGBoost, which analyzes the weight of each feature and thus determines its contribution within the decision trees. By comparing these contributions, a feature-importance ranking was obtained. From highest to lowest, the selected features for model training are: edge detection, curvature, dip, dip deviation, and variance. The labels were derived from the fracture identification results of eight key wells (B1, B2, B3, B4, B5, B6, B7, and B8) within the study area (Fig. 5).

To achieve scale alignment between heterogeneous datasets, it is necessary to upscale fracture labels identified at the well-log scale to the seismic scale. This process requires consideration of both the seismic resolution and detection capability, and the determination of an equal-interval sampling thickness between near-well seismic traces and fracture labels. Based on the fracture identification results from well logs, the average fracture intensity is calculated within each equal-interval sampling window, thereby enabling the upscaling of well-log-scale fracture labels and their assignment to near-well seismic traces. The detection limit of three-dimensional seismic data is commonly considered to be $\lambda/8$, which represents the minimum thickness that can be detected but not resolved (Zeng, 2013, 2018; Zeng and Kerans, 2003). In the study area, $\lambda/8$ corresponds to a thickness of approximately 12.5 m, and this value is therefore selected as the equal-interval sampling thickness. Given that the horizontal extent of fracture zones is typically much larger than the vertical resolution thickness, even when fractures can only be detected but not resolved in the vertical direction by seismic data, their planar distribution characteristics can still be delineated with reasonable clarity. Based on the above equal-interval fracture intensity calculation, fractures identified in key wells within the study area are transformed into fracture intensity indicators. Using the relatively accurate time–depth relationships provided by the operating oil company, the fracture intensity values from the key wells are subsequently calibrated to the seismic attributes.

The extracted well-side seismic traces were randomly split into training and testing sets in a 7:3 ratio. All parameters were globally optimized using a grid search method, and the testing set was used to evaluate the predictive performance of the model. The optimal parameter settings for the model are summarized in Table 1. Based on the seismic detection limit ($\lambda/8$), an optimal sampling thickness of 12.5 m was selected, with 500 regression trees and a learning rate of 0.5.

To demonstrate the improved performance of the proposed methods compared to the original models without GNN integration, the Multilayer Perceptron (MLP), U-Net, and non-GNN XGBoost models were selected for training and result comparison with the improved methods. For the conventional logging–based fracture identification, the comparative model uses exactly the same parameter settings as the improved method, except that no edges are constructed between vertices—making it equivalent to an MLP. For the seismic fault detection, the comparative model is a U-Net with identical parameter settings to the improved method (e.g., convolution kernel size, pooling kernel size, and number of graph convolution layers), except that it does not include the fault-graph computation module. For the inter-well fracture prediction, the comparative model is the original XGBoost, which excludes the computation of the fault damage zone graph. Other parameter settings of XGBoost, such as the number of regression trees and maximum tree depth, remain consistent with those used in the improved method.

### 3.4. Results

Identification of fractures via conventional logging data achieves an accuracy of 91.96% on the testing set, with precision and recall for fracture samples in the testing set at 92.35% and 82.44%, respectively. In contrast, the model trained using MLP without graph construction has a testing set accuracy, fracture sample precision, and recall of 72.03%, 8.43%, and 57.69%, respectively, which is much lower than that of the comparison model using GNNs (Table 2). Based on this model, fractures were identified vertically in the remaining wells of the study area. In the B1 testing well, the GNN model also achieved better results and showed a higher agreement with core and image logging interpretation results. For example, from 4563.125 m to 4565.25 m, the GNN model correctly identified all fractures within this interval, while the MLP did not fully identify them. From 4565.875 m to 4569.25 m, the GNN model identified most fracture samples in the segment, whereas the MLP recognized only a few fractures. From 4574.25 m to 4576.75 m, the GNN model found most fractures, while the MLP classified this segment as non-fractured (Fig. 9). Based on the above results, the GNN-based fracture identification model not only achieves higher accuracy but also demonstrates better recognition ability for fracture samples. It effectively addresses the issue of low precision and recall for fractures caused by data imbalance.

The fault detection results on the testing set are shown in Table 3. The accuracy (Acc) is 97.39%, which is an improvement of 3.97% compared to the U-Net method without the GNN fault continuity analysis module. The Dice coefficient (DSC) reached 83.37%, an increase of 15.07%; specificity (Spe) is 98.02%, up by 4.85%; and AUC is 99.12%, an improvement of 0.58%. Overall, the fault detection method with the GNN continuity analysis module outperforms the original U-Net in all four evaluation metrics on the testing set, with significant advantages in DSC and Spe. The trained model was applied to fault detection in the study area, and the plane fault distribution results for each target layer were obtained (Fig. 10).

**Table 2**
Conventional logging fracture identification results on testing sets.

| Method | Accuracy, % | Precision, % | Recall, % |
|---|---|---|---|
| Non-GNN | 72.03 | 8.43 | 57.69 |
| GNN | 91.96 | 92.35 | 82.44 |

The fracture distribution prediction results for the testing set are shown in Table 4. The mean squared error (MSE) is $1.26\times10^{-2}$, the coefficient of determination ($R^2$) reaches $7.40\times10^{-1}$, and the mean absolute error (MAE) is only $9.38\times10^{-2}$. Compared to methods that do not incorporate the construction of fault damage zone graphs using GNNs, there is a certain degree of improvement, signifying that the enhanced model has a good fitting capability for the data, and the prediction results exhibit a relatively high degree of consistency. This model was applied to the task of predicting fracture distribution in the study area, and the planar distribution results of the target formation were extracted (Fig. 11(a)). The predicted profiles adjacent to wells B2 and B3 in the study area were compared with conventional logging fracture identification results. The overall consistency between the predicted profiles and the conventional logging identification results is relatively high (Fig. 11(b) and (c)). In Fig. 11, the fracture labels represent the logging fracture identification results, while the predicted profiles represent the predicted fracture intensity, which is depicted by darker shades of red.

Based on the stratified data, time windows were applied to the predicted fault and fracture intensity data volumes, resulting in distribution of faults and fractures for the target formation. To better illustrate the relationship between fractures and faults, the fault detection results were overlaid with the fracture prediction results (Fig. 12(a)), where purple labels represent faults and red labels represent fracture intensity. Using the GNN fracture identification model, fractures were identified in multiple key wells of the Fengcheng Formation in the Mabei area. Profiles were established for wells B1, B2, and B3 (Fig. 12(b)). The overall development of faults and fractures in the Fengcheng Formation of the study area is relatively high, with faults mainly distributed in the southern and eastern parts of the study area, trending primarily in near-E–W and near-S–N directions. Fracture development is more intense near faults, while it decreases with distance from the faults.

## 4. Discussion

### 4.1. Verification of prediction results from an engineering perspective

To further verify the reliability of the proposed method, production data from several key wells in the study area were statistically analyzed. The statistical results were then compared with the spatial distribution patterns of faults and fractures in the study area, providing an engineering-based validation of the method's effectiveness. Natural fractures not only provide migration pathways and storage space for shale oil, but also play a controlling role during hydraulic fracturing. Fracturing tends to propagate along mechanically weak planes in the formation, which commonly correspond to the extension directions of natural fractures, thereby generating induced hydraulic fractures. As a result, the development degree of natural fractures exerts a significant influence on the daily oil production of wells, generally characterized by higher production rates in areas with more well-developed natural fractures. In this study, the development degree of natural fractures is characterized by two parameters: fracture intensity and cumulative thickness of fracture-developed intervals. Both parameters are predicted using the method proposed in this work.

To verify the reliability of the vertical well-log-based fracture identification results, well B6 in the study area was selected for detailed analysis. Compared with other wells, well B6 has more complete production data, including multiple oil-producing intervals as well as non-productive layers. After excluding all non-productive layers, a statistical analysis was conducted on the

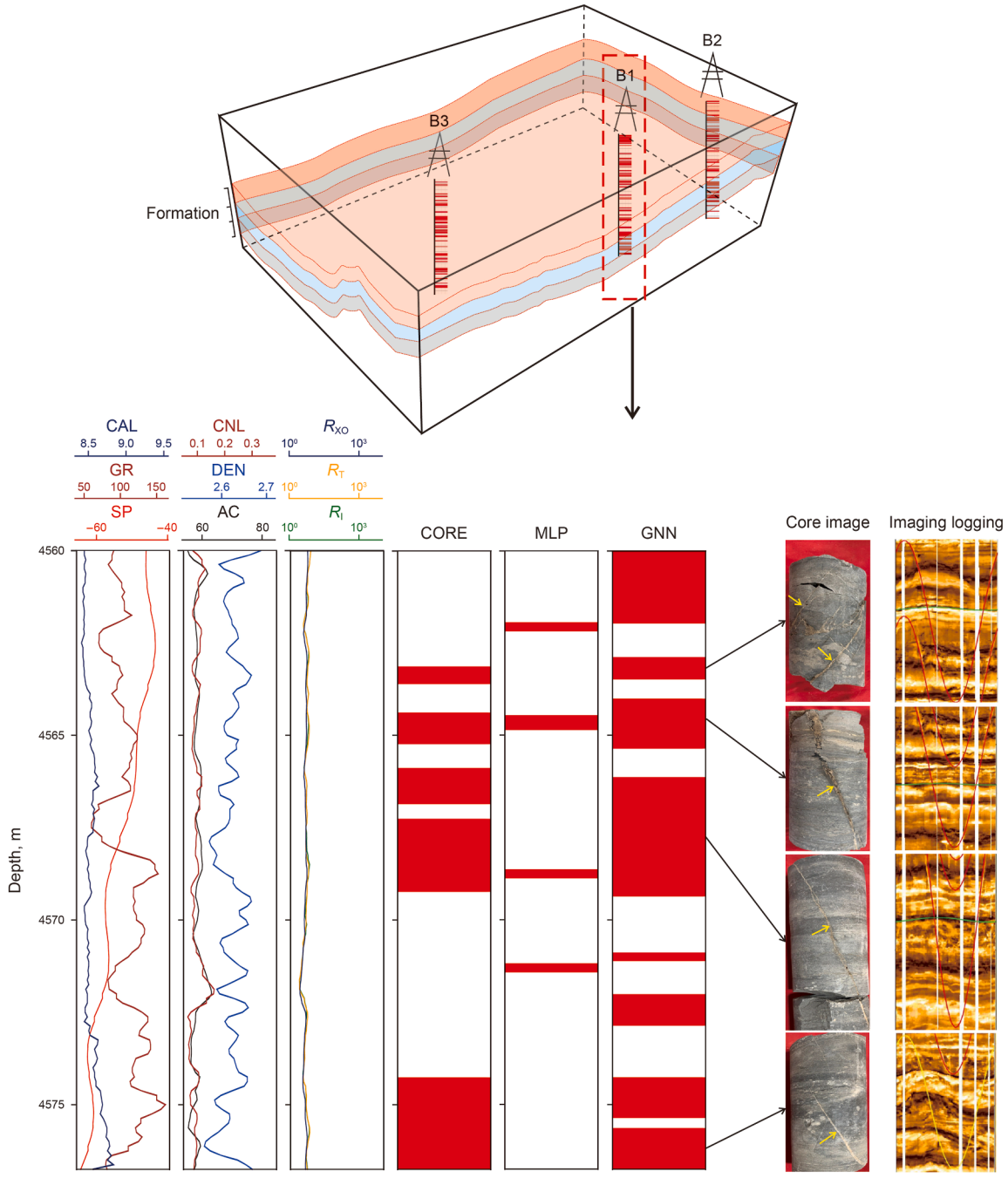



**Fig. 9.** Comparison of blind well identification results and core samples (modified from Lu et al., 2024a).

**Table 3**
Seismic fault detection results on testing sets.

| Method | DSC, % | Accuracy, % | Specificity, % | AUC, % |
|---|---|---|---|---|
| Non-GNN | 68.30 | 93.42 | 93.17 | 98.54 |
| GNN | 83.37 | 97.39 | 98.02 | 99.12 |

relationship between daily oil production and the cumulative fracture length of the oil-producing intervals. The results show a robust positive correlation between daily oil production and cumulative fracture length, indicating that the well-log-based fracture identification results are consistent with the observed production statistics (Fig. 13).

A lateral comparison of the relationship between production performance and fracture development among different wells also provides an effective means to validate the accuracy of the predicted inter-well fracture distribution. Because the thicknesses of oil-producing intervals vary from well to well, the daily oil production per unit thickness was calculated by compiling statistics on the production interval thickness of each well, allowing for a more meaningful comparison with fracture intensity. A comparative analysis was conducted between the daily oil production per unit thickness of productive intervals in key wells (B2, B6, B9, B10, and B11) and the predicted occurrence of fracture planes. The results indicate that wells with higher daily oil production per unit thickness are predominantly located in areas characterized by high

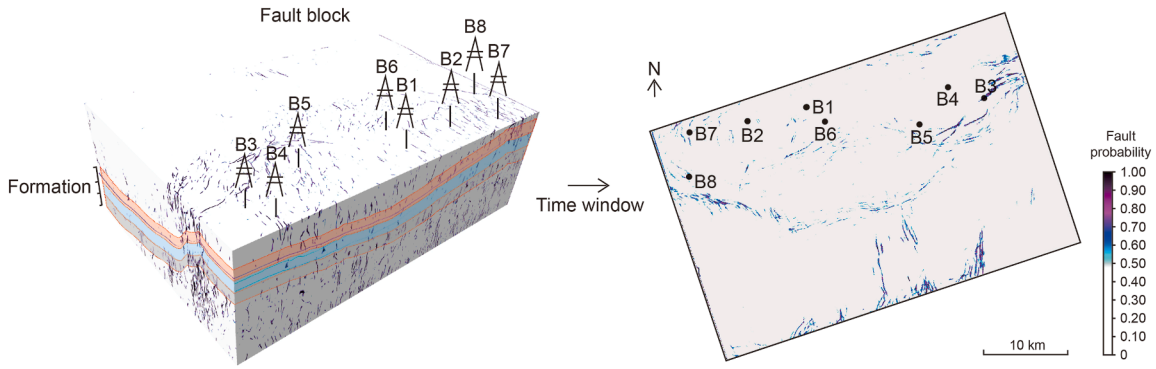



**Fig. 10.** Seismic detection results of the study area.

**Table 4**
Inter-well fracture distribution predication results on testing sets.

| Method | $R^2$ | MSE | MAE |
|---|---|---|---|
| Non-GNN | $2.15\times10^{-1}$ | $2.10\times10^{-2}$ | $1.21\times10^{-1}$ |
| GNN | $7.40\times10^{-1}$ | $1.26\times10^{-2}$ | $9.38\times10^{-2}$ |

fracture intensity. For example, Wells B9 and B10 exhibit high fracture intensity within their production intervals, exceeding 0.8, and correspondingly show relatively high daily oil production per unit thickness. In contrast, wells with relatively low production performance, such as Wells B2 and B11, are associated with lower fracture intensity within their production intervals, both below 0.25 (Table 5).

Considering the statistical results summarized in Table 5, a strong positive correlation is observed between daily oil production per unit thickness of the entire oil-producing formation and fracture intensity within the oil-producing formation. An increase in fracture intensity corresponds to higher daily oil production per unit thickness, indicating that the predicted inter-well fracture plane distribution is consistent with the production statistics and further validates the reliability of the proposed method from an engineering perspective (Fig. 14).

### *4.2. Interpretation of important parameter settings based on geological theory*

The number of vertices included in each small-scale graph pooling unit is discussed. In the proposed model, discrete sampling points within a fixed depth interval are treated as an integrated entity during training to simulate the continuous distribution characteristics of fractures along the borehole. At the same time, the connections between fracture segments at different depths are incorporated into the data structure, which is crucial for enhancing the generalization capability of the model. In the fracture identification model based on conventional well logs, the purpose of small-scale graph pooling is to extract and fuse information from discrete logging samples. Therefore, the number of vertices in each small-scale graph should be determined according to the fracture heights observed from core data. Since the small-scale graph is constructed along the depth sequence and the interval between adjacent vertices is consistent with the logging sampling interval (0.125 m), the number of vertices directly corresponds to the vertical interval represented. According to core observations and fracture statistics from the cored sections in the study area, more than 90% of the fractures have heights less than 0.5 m. This height corresponds to four adjacent vertex intervals, that is, five vertices connected sequentially along the depth direction. An additional and important consideration is maintaining consistency with the aggregation range defined by the graph convolution process. Because the global graph is also constructed along the depth sequence and the GCN employs two convolutional layers, the corresponding aggregation range (0.5 m) yields the highest test accuracy of 91.96%. When the number of convolutional layers increases, the model aggregates excessive irrelevant information, whereas a single convolutional layer leads to underfitting. The accuracies obtained with other layer configurations are all lower than that achieved with two layers. Because a small-scale graph composed of five vertices still incurs non-negligible computational cost during graph operations, only a single graph convolution layer and a single graph pooling layer are adopted in the small-scale graph computation to improve training efficiency. This design significantly reduces computational burden while preserving the fracture identification performance of the model.

In this study, all computations are based on GNNs, which require operations to be defined on graph structures. Incorporating established geological knowledge into the graph structure can effectively enhance the prediction accuracy of the model. However, an excessively complex graph structure may substantially reduce computational efficiency and introduce unnecessary information propagation. Therefore, it is necessary to manually control the number of generated edges by setting appropriate thresholds to constrain the complexity of the graph topology. In the well-log fracture identification model, for Eq. (5), the threshold is set as $\delta_d = 1\%$, meaning that connections are established only between vertices whose Euclidean distance falls within the nearest 1%. In the fault-graph construction, the velocity field within fault zones is set significantly higher than that outside the faults, allowing vertices within the fault zone to connect and describe the spatial extent of faults, while preventing connections between vertices outside the fault. For Eq. (11), the parameters are set as $V_{\text{fault}} = 1.0$ and $V_{\text{nofault}} = 0.01$, indicating that the velocity in fault zones is 100 times greater than that in non-fault zones. Here, the velocity represents a dimensionless travel-time parameter. Edge generation is based on the travel time computed from the Eikonal equation, representing the time required for movement between two vertices. For Eq. (13), $\gamma_{FF} = 10$ is adopted, which corresponds to approximately a voxel distance of 10 when the velocity is equal to 1. This threshold depends on both the velocity of the corresponding field and the voxel resolution. When the travel time between two vertices—computed through the geodesic distance within the given velocity field—is less than the threshold $\gamma_{FF}$, an edge is established between them. By adjusting the grid window size, the number of grids can be modified, thereby controlling the number of vertices and the overall size of the fault graph. In this

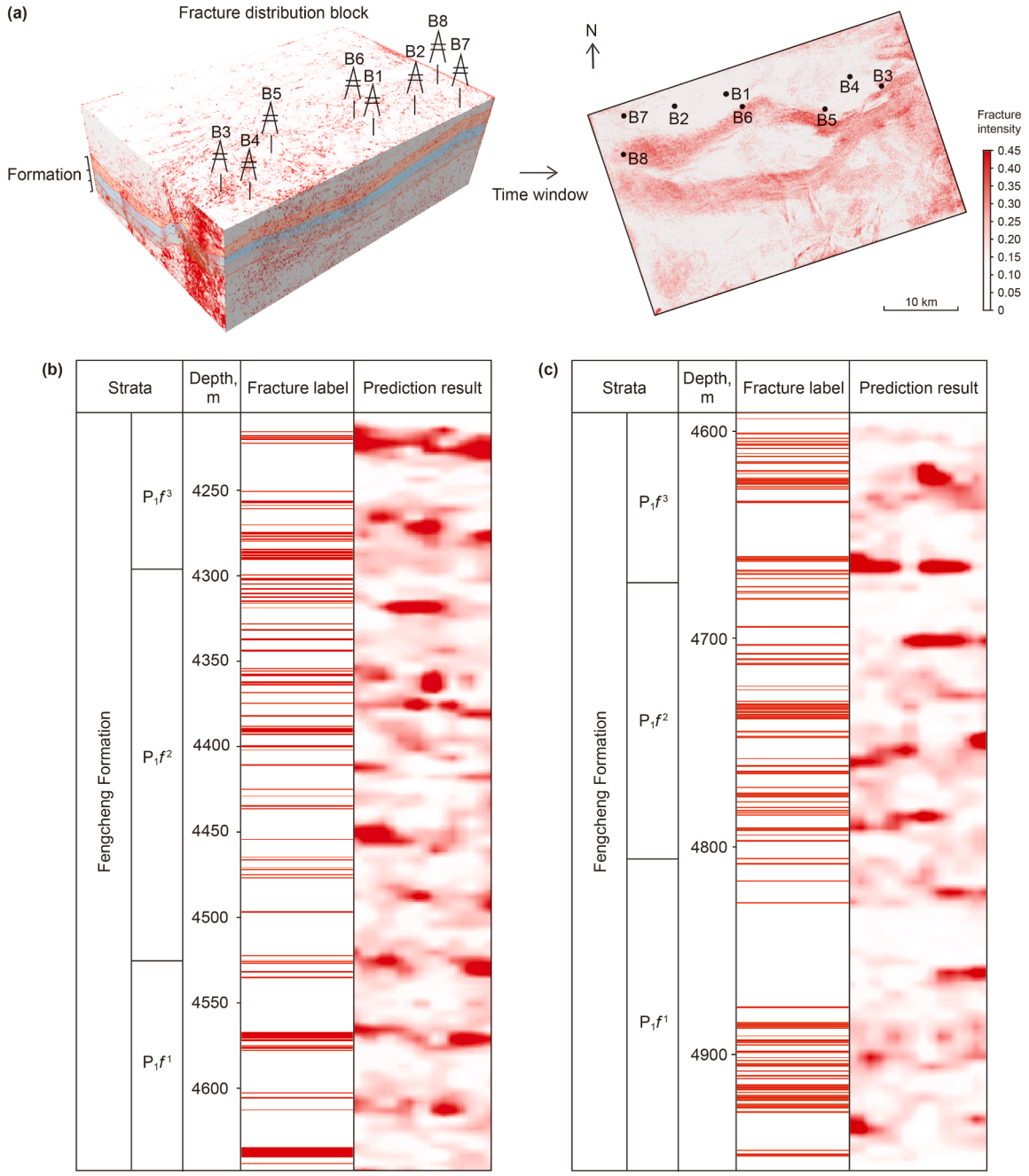



**Fig. 11.** Inter-well fracture distribution prediction results of the study area. **(a)** Fracture distribution block, **(b)** Well B2, **(c)** Well B3.

study, the grid window size is set to 8, which corresponds to transforming a 128×128×128 three-dimensional image into a 16×16×16 three-dimensional fault-graph structure.

### 4.3. Uncertainty analysis by Monte Carlo simulation

To quantitatively assess the reliability of the fracture identification results, an uncertainty analysis based on Monte Carlo simulation was performed for the well-logging fracture identification model. This model is selected for uncertainty analysis because it directly outputs node-level fracture classification results and serves as the fundamental input for inter-well fracture prediction. Therefore, quantifying its predictive uncertainty is particularly important.

Monte Carlo dropout is adopted to approximate the predictive uncertainty of the GNN. Specifically, dropout layers are kept active during the inference stage, and multiple stochastic forward passes are performed for the same input well-logging data. For each depth node, the model prediction is repeated several times, resulting in a distribution of class probabilities.

The final fracture prediction uncertainty is quantified using the predictive entropy of the averaged probability distribution. A lower entropy value indicates a more confident and stable prediction, whereas a higher entropy reflects increased uncertainty. Based on the spatial distribution of predictive entropy, regions

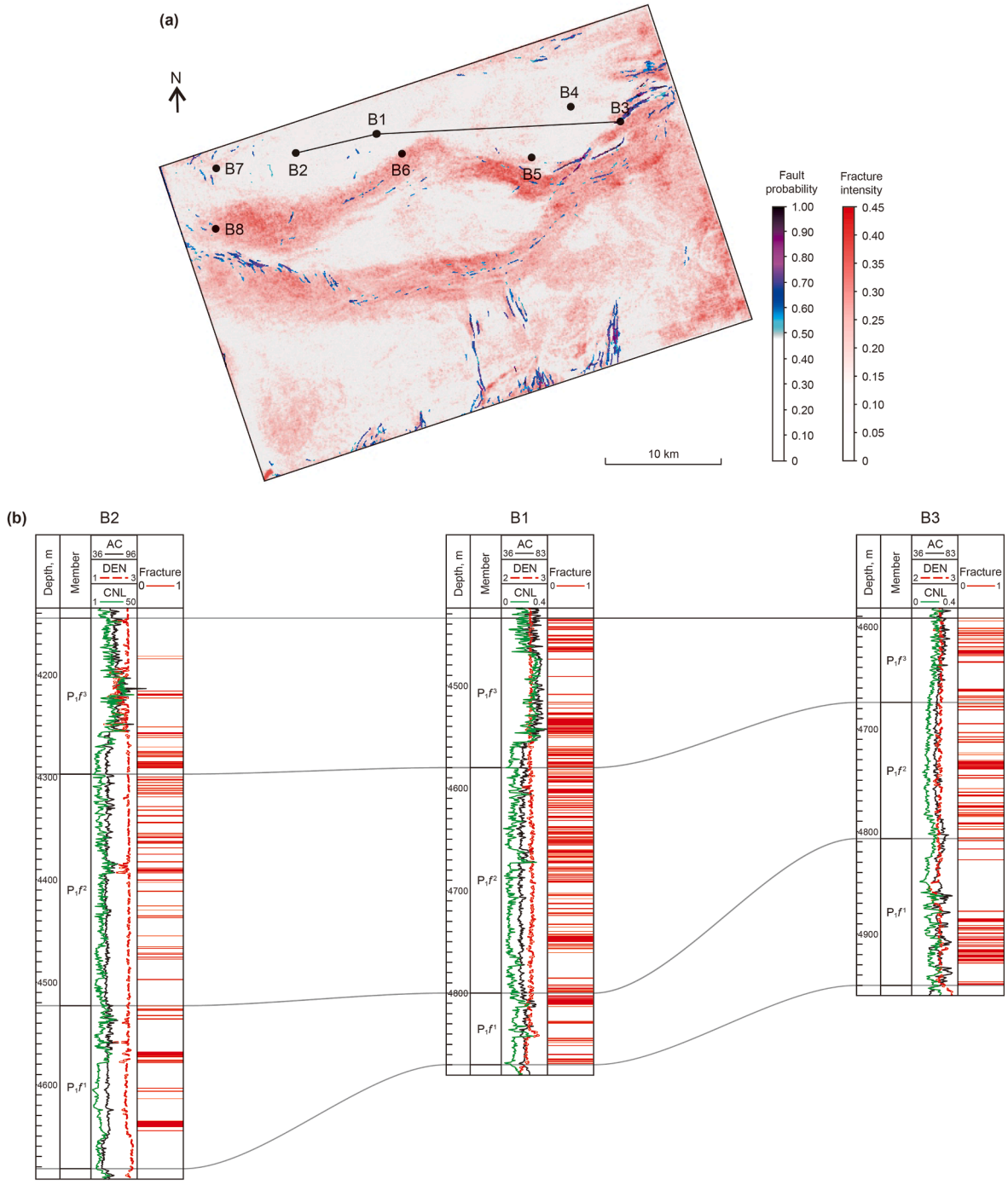



**Fig. 12.** **(a)** Fault–fracture system distribution prediction results in the study area, **(b)** fracture identification results of the B2–B1–B3 section.

with low uncertainty are identified as high-confidence zones, and regions with high uncertainty are interpreted as low-confidence zones.

The uncertainty analysis results show that high confidence fracture identification is mainly concentrated in depth intervals with strong logging responses and sufficient training constraints, where the model has learned stable and discriminative feature representations. In contrast, higher uncertainty tends to occur in intervals characterized by ambiguous logging responses or sparse representative samples, in which the input features provide weaker constraints for fracture discrimination. This indicates that the predictive uncertainty is closely related to both the fracture response characteristics of logging curves and the local data coverage in the training set.

### 4.4. Small sample performance of relevant models

In practical applications, it is often difficult to obtain a sufficient number of fully labeled data samples due to high acquisition and interpretation costs. To investigate the robustness of the proposed models under small-sample conditions, a series of experiments were conducted by progressively reducing the number of training

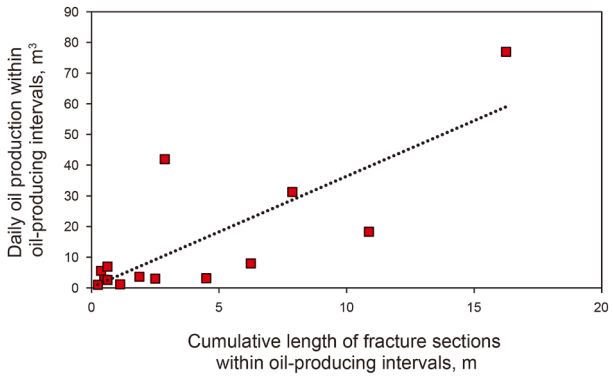



**Fig. 13.** The relationship between the cumulative length of fractured sections and the daily oil production in the oil-production intervals.

**Table 5**
Comparison of daily oil production per unit thickness with fracture intensity in key wells.

| Well | Fracture intensity of the entire oil-producing intervals | Daily oil production of the entire oil-producing intervals, $m^3$ | Cumulative length of the entire oil-producing intervals, m | Daily oil production per unit thickness of the entire oil-producing intervals, $m^3$ |
|---|---|---|---|---|
| B2 | 0.23 | 46.10 | 91 | 0.51 |
| B6 | 0.75 | 86.06 | 88 | 0.98 |
| B9 | 0.83 | 58.65 | 40 | 1.47 |
| B10 | 0.91 | 18.33 | 12 | 1.53 |
| B11 | 0.21 | 3.04 | 12 | 0.25 |

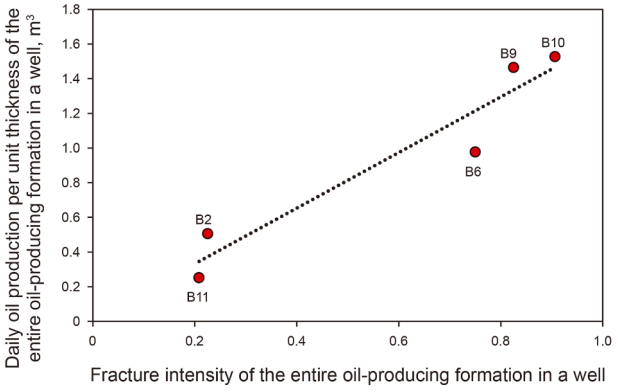



**Fig. 14.** Relationship between daily oil production and fracture intensity.

data samples while keeping the testing set fixed. Specifically, the size of the training set was reduced to 70%, 50%, 30%, 20%, and 10% of the full training dataset.

The performance of the well-log fracture identification model was evaluated under different training data ratios while maintaining the same testing set. When the full training dataset was used, the model achieved a high prediction accuracy of approximately 91.96%. As the number of training samples decreased, a gradual degradation in performance was observed. When more than 50% of the training samples were retained, the model maintained relatively stable performance, with accuracy remaining above 83.32%, indicating that representative fracture-related logging patterns can still be effectively learned from a limited amount of data. However, when the training set was reduced to 30%, the prediction accuracy dropped to approximately 72.24%, and further decreased to below 70% when fewer training samples were available. These results suggest that although the proposed well-log fracture identification model exhibits a certain degree of robustness under limited training data, an excessively small training dataset is insufficient to fully capture the diversity of fracture-related logging responses.

A small-sample sensitivity analysis was further conducted for the inter-well fracture prediction model by progressively reducing the number of training data samples while keeping the testing data unchanged. When all available training samples were used, the model achieved a coefficient of determination ($R^2$) of approximately 0.74, indicating good agreement between the predicted and observed fracture attributes. As the number of training samples decreased, the prediction performance exhibited a clear degradation trend. When more than 50% of the training samples were retained, the model still maintained relatively stable performance, with an $R^2$ value of approximately 0.66. However, when the training set was reduced to 30%, the $R^2$ value dropped to around 0.48. With fewer training samples available, the reliability and spatial continuity of the predicted inter-well fracture distributions tended to deteriorate significantly, particularly in structurally complex zones. These results indicate that although the proposed model remains applicable under moderately limited data conditions, an excessively small number of training samples is insufficient to effectively constrain inter-well fracture prediction.

Overall, these experiments demonstrate that the proposed models maintain acceptable performance under reduced training data conditions, which is highly relevant for real-field applications where labeled data samples are scarce. The well-log fracture identification model shows relatively stable performance with a moderate number of training samples, suggesting that fracture-related logging responses share common patterns across different samples. In contrast, the inter-well fracture prediction model is more sensitive to training data availability, as it relies on data-driven spatial constraints to infer fracture distributions between locations. From an engineering perspective, these results indicate that a minimum amount of training data is required to ensure reliable inter-well fracture prediction, while additional training samples contribute diminishing performance improvements beyond this threshold. Furthermore, the observed performance degradation under extreme small-sample conditions suggests that incorporating prior geological knowledge, transfer learning from neighboring blocks, or semi-supervised learning strategies may further enhance model robustness when labeled data are severely limited.

### 4.5. Application prospects and future work

Although core data and image logs can effectively indicate fracture distribution patterns, their availability is usually limited and insufficient to achieve comprehensive coverage of the study area and target formations. In contrast, conventional well logs and seismic data are more readily available and provide extensive spatial coverage. The proposed method requires only a limited amount of image log and core data to provide fracture labels. By sequentially applying three sub-methods for fracture identification and prediction, reliable fracture distribution patterns at both well and inter-well scales, as well as fault distribution characteristics, can be efficiently obtained even with a small number of labeled samples. Compared with conventional artificial intelligence approaches, the proposed framework achieves significantly higher prediction accuracy. Moreover, the three sub-methods form an integrated and mutually constrained system, which offers clear

advantages over single-method solutions in terms of robustness and practical applicability.

But some limitations of the method have been identified and need further research. First, due to the incorporation of graph construction and graph computation, the training process is more complicated compared to traditional intelligent fracture identification and prediction methods. The numerous parameters result in a high model complexity and longer training time. Second, this method is currently applicable only to the prediction of fault–fracture systems in the Mahu Sag. To extend its application to other regions, different graph-construction schemes and edge-weight assignment strategies must be designed according to the reservoir characteristics and the geological factors controlling fracture development in the target formation of each basin. In addition, the conventional well logs, seismic data, and corresponding labels from the new region must be incorporated into the model for retraining, as each study area exhibits its own distinct fault and fracture signatures in geophysical datasets. Lastly, this method is only applicable to high-angle tectonic fractures controlled by faults, which can be interpreted through image logs, and cannot identify or predict the bedding-parallel fractures commonly found in shale oil reservoirs.

Future improvements should focus on simplifying the training process of small-scale graph pooling and global graph construction, reducing model complexity, shortening training time, improving prediction accuracy, and enhancing the model's generalization ability. At the same time, considering the critical role that bedding-parallel fractures play in the accumulation, preservation, and migration of shale oil, further research on the identification and prediction methods for bedding-parallel fractures should be pursued in subsequent studies.

Bedding-parallel fractures are usually parallel to the strata, have small apertures, short extension distances, and are mostly filled with organic-rich mud, making it difficult to accurately identify individual bedding-parallel fractures in either image or conventional logging surveys. Moreover, since they might not have any impact on seismic wave propagation they could not become visible in 3D seismic data. Therefore, the identification of bedding-parallel fractures is more challenging than tectonic fractures and cannot be identified by conventional logging as individual fractures. However, lithofacies and lamination have a significant control over bedding-parallel fractures which might become useful to this end. By integrating sedimentary, lithofacies, and lamination information with conventional and image logging data, it is possible to predict the fracture density at each depth section, rather than identifying the exact development locations of individual fractures as in the tectonic fracture identification process that can be the focus of future studies.

## 5. Conclusions

This paper proposes a set of methods for predicting the distribution of fault–fracture systems based on graph neural networks, comprising three sub-methods that are inherently interconnected. Conventional logging-based fracture identification provides a detailed characterization of the vertical distribution of tectonic fractures within each well, and these well-level identification results are also used to label seismic data for subsequent inter-well fracture prediction. Seismic fault-detection results, in turn, delineate the spatial distribution of faults across the study area. Although the seismic fault-detection workflow is not directly coupled with the well-bore fracture identification, its outputs serve as structural constraints for the inter-well fracture prediction process. Consequently, the inter-well prediction integrates information derived from both logging-based fracture identification and seismic fault detection, enabling a more comprehensive delineation of fracture distributions between wells. Together, these integrated training outcomes constitute a coherent and comprehensive framework for evaluating and predicting the vertical and horizontal distributions of faults and fractures in shale oil reservoirs.

The intelligent method of identifying fractures from conventional logging data integrates lithology, stratigraphy, and mechanical stratigraphy information from the graph structure into the identification model, effectively addressing the issue of low precision and recall in fracture identification caused by data imbalance. Compared to non-GNN methods, this approach achieves higher accuracy in fracture identification from logging data, reaching 91.96%. The seismic fault detection method constructs a joint loss function to jointly train and optimize the fault segmentation and fault continuity analysis modules, effectively resolving the issue of poor fault continuity detection caused by the limited receptive field of the U-Net. Compared to U-Net, the improved method achieves a fault detection accuracy of 97.39%, representing a 3.96% improvement over the previous method. The inter-well fracture distribution prediction model effectively incorporates the controlling effect of faults on fractures into the fracture distribution prediction model through information transfer between vertices, significantly improving the prediction accuracy of inter-well tectonic fracture distribution. The average error in tectonic fracture prediction is 9.38%, a 24% reduction compared to existing methods.

However, this set of methods still has some limitations. Due to the inclusion of graph construction and graph computation, the training process is more cumbersome, and the numerous input parameters result in higher model complexity and longer training times. Additionally, if the model is applied to other regions, it is necessary to readjust the number of vertices in small-scale graphs, the constraints of the global graph, and the percentile for edge generation. Future improvements should focus on simplifying the training process for small-scale graph pooling and global graph construction, reducing model complexity, shortening training times, improving prediction accuracy, and further enhancing the generalization capabilities of the present model.

## CRediT authorship contribution statement

**Guo-Qing Lu:** Writing – review & editing, Writing – original draft, Methodology, Data curation. **Lian-Bo Zeng:** Writing – review & editing, Visualization, Funding acquisition. **Guo-Ping Liu:** Writing – review & editing, Validation. **Zhen-Lin Wang:** Resources, Data curation. **Yu-Xiao Wu:** Resources. **Shao-Qun Dong:** Writing – review & editing, Validation. **José Luis Drummond Alves:** Writing – review & editing. **Jun-Feng Zhao:** Writing – review & editing. **Mehdi Ostadhassan:** Writing – review & editing, Visualization. **Jian Su:** Writing – review & editing, Visualization.

## Declaration of interest statement

No conflict of interest exists in the submission of this manuscript, and all authors approve this manuscript for publication. We would like to declare on behalf of my co-authors that the work described was original research that has not been published previously and is not under consideration for publication elsewhere, in whole or in part.

## Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (No.42302148, No. 42090025).

## References


Aghli, G., Soleimani, B., Moussavi-Harami, R., et al., 2016. Fractured zones detection using conventional petrophysical logs by differentiation method and its correlation with image logs. J. Petrol. Sci. Eng. 142, 152–162. https://doi.org/10.1016/j.petrol.2016.02.002.

Alonaizi, F., Pevzner, R., Bóna, A., et al., 2013. 3D diffraction imaging of linear features and its application to seismic monitoring. Geophys. Prospect. 61 (6), 1206–1217. https://doi.org/10.1111/1365-2478.12063.

Alves, T.M., Taylor, J., Corcorant, P., et al., 2025. Data sampling strategies for accurate fault analyses: A scale-independent test based on a machine learning approach. J. Struct. Geol. 191, 105342. https://doi.org/10.1016/j.jsg.2025.105342.

Berg, S.S., Skar, T., 2005. Controls on damage zone asymmetry of a normal fault zone: Outcrop analyses of a segment of the Moab fault, SE Utah. J. Struct. Geol. 27 (10), 1803–1822. https://doi.org/10.1016/j.jsg.2005.04.012.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Bressan, T.S., Kehl de Souza, M., Girelli, T.J., et al., 2020. Evaluation of machine learning methods for lithology classification using geophysical data. Comput. Geosci. 139, 104475. https://doi.org/10.1016/j.cageo.2020.104475.

Chen, S., Wang, Y., Guo, J., et al., 2021. Multi-scale evaluation of fractured carbonate reservoir and its implication to sweet-spot optimization: A case study of Tazhong oilfield, Central Tarim Basin, China. Energy Rep. 7, 2976–2988. https://doi.org/10.1016/j.egyr.2021.05.017.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, Ithaca, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Choi, J.H., Edwards, P., Ko, K., et al., 2016. Definition and classification of fault damage zones: A review and a new methodological approach. Earth Sci. Rev. 152, 70–87. https://doi.org/10.1016/j.earscirev.2015.11.006.

Cohen, I., Coult, N., Vassiliou, A.A., 2006. Detection and extraction of fault surfaces in 3D seismic data. Geophysics 71 (4), P21–P27. https://doi.org/10.1190/1.2215357.

Cooper, G.F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Mach. Learn. 9 (4), 309–347. https://doi.org/10.1007/BF00994110.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297. https://doi.org/10.1023/A:1022627411411.

Dong, S., Zeng, L., Lyu, W., et al., 2020. 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., Zhong, Z.H., Cui, X.H., et al., 2023. A deep kernel method for lithofacies identification using conventional well logs. Petrol. Sci. 20 (3), 1411–1428. https://doi.org/10.1016/j.petsci.2022.11.027.

Du, X., Jin, Z., Zeng, L., et al., 2023. Formation of natural fractures and their impact on shale oil accumulation in the Mahu Sag, Junggar Basin, NW China. Int. J. Coal Geol. 279, 104385. https://doi.org/10.1016/j.coal.2023.104385.

Du, X.Y., Jin, Z.J., Zeng, L.B., et al., 2025. Characteristics, controlling factors and mechanisms of natural fractures formation in lacustrine shale oil reservoirs: The Chang 7 member in Ordos Basin, China. Petrol. Sci. 22 (4), 1391–1406. https://doi.org/10.1016/j.petsci.2025.03.005.

Ghosh, S., Galvis-Portilla, H.A., Klockow, C.M., et al., 2018. An application of outcrop analogues to understanding the origin and abundance of natural fractures in the Woodford Shale. J. Petrol. Sci. Eng. 164, 623–639. https://doi.org/10.1016/j.petrol.2017.11.073.

Gong, L., Liu, G., Wang, R., et al., 2025. Editorial: Distribution and development of faults and fractures in shales. Minerals 15 (11), 1154. https://doi.org/10.3390/min15111154.

Gong, L., Wang, J., Gao, S., et al., 2021. Characterization, controlling factors and evolution of fracture effectiveness in shale oil reservoirs. J. Petrol. Sci. Eng. 203, 108655. https://doi.org/10.1016/j.petrol.2021.108655.

Gou, Q., Xu, S., Hao, F., et al., 2019. Full-scale pores and micro-fractures characterization using FE-SEM, gas adsorption, nano-CT and micro-CT: A case study of the Silurian Longmaxi Formation shale in the Fuling area, Sichuan Basin, China. Fuel 253, 167–179. https://doi.org/10.1016/j.fuel.2019.04.116.

Gou, Q., Xu, S., Hao, F., et al., 2021b. The effect of tectonic deformation and preservation condition on the shale pore structure using adsorption-based textural quantification and 3D image observation. Energy 219, 119579. https://doi.org/10.1016/j.energy.2020.119579.

Gou, Q., Xu, S., Hao, F., et al., 2021a. Making sense of micro-fractures to the Longmaxi shale reservoir quality in the Jiaoshiba area, Sichuan Basin, China: implications for the accumulation of shale gas. J. Nat. Gas Sci. Eng. 94, 104107. https://doi.org/10.1016/j.jngse.2021.104107.

Gu, J., Wang, Z., Kuen, J., et al., 2018. Recent advances in convolutional neural networks. Pattern Recognit 77, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013.

Hale, D., 2013. Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images. Geophysics 78 (2), O33–O43. https://doi.org/10.1190/geo2012-0331.1.

Hamilton, W.L., Ying, R., Leskovec, J., 2017. Inductive representation learning on large graphs. In: 31st International Conference on Neural Information Processing Systems, pp. 1025–1035. https://doi.org/10.48550/arXiv.1706.02216. Long Beach, CA, USA.

Hui, G., Chen, S., Gu, F., 2024. Strike–slip fault reactivation triggered by hydraulic-natural fracture propagation during fracturing stimulations near Clark Lake, Alberta. Energy Fuels 38 (19), 18547–18555. https://doi.org/10.1021/acs.energyfuels.4c02894.

Hui, G., Chen, Z., Schultz, R., et al., 2023a. Intricate unconventional fracture networks provide fluid diffusion pathways to reactivate pre-existing faults in unconventional reservoirs. Energy 282, 128803. https://doi.org/10.1016/j.energy.2023.128803.

Hui, G., Chen, Z.X., Lei, Z.D., et al., 2023b. A synthetical geoengineering approach to evaluate the largest hydraulic fracturing-induced earthquake in the East Shale Basin, Alberta. Petrol. Sci. 20 (1), 460–473. https://doi.org/10.1016/j.petsci.2023.01.006.

Ja'fari, A., Kadkhodaie-Ilkhchi, A., Sharghi, Y., et al., 2012. Fracture density estimation from petrophysical log data using the adaptive neuro-fuzzy inference system. J. Geophys. Eng. 9 (1), 105–114. https://doi.org/10.1088/1742-2132/9/1/013.

Katz, O., Reches, Z., Baer, G., 2003. Faults and their associated host rock deformation: Part I. Structure of small faults in a quartz–syenite body, southern Israel. J. Struct. Geol. 25 (10), 1675–1689. https://doi.org/10.1016/S0191-8141(03)00011-7.

Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1609.02907.

Liu, F.T., Ting, K.M., Zhou, Z.H., 2008. Isolation forest. In: Eighth IEEE International Conference on Data Mining, Pisa, Italy, pp. 413–422. https://doi.org/10.1109/ICDM.2008.17.

Liu, G., Jin, Z., Yang, S., et al., 2023a. Distribution pattern of natural fractures in lacustrine shales: a case study of the Fengcheng formation in the Mahu Sag of the Junggar Basin, China. Front. Earth Sci. 11, 1207033. https://doi.org/10.3389/feart.2023.1207033.

Liu, G., Jin, Z., Zeng, L., et al., 2023b. Natural fractures in deep continental shale oil reservoirs: a case study from the Permian Lucaogou Formation in the Eastern Junggar Basin, Northwest China. J. Struct. Geol. 174, 104913. https://doi.org/10.1016/j.jsg.2023.104913.

Liu, G., Jin, Z., Zeng, L., et al., 2025. Laminar controls on bedding-parallel fractures in Permian lacustrine shales, Junggar Basin, northwestern China. Geol. Soc. Am. Bull. 137 (7–8), 3512–3526. https://doi.org/10.1130/b37792.1.

Lu, G., Zeng, L., Dong, S., et al., 2023. Lithology identification using graph neural network in continental shale oil reservoirs: A case study in Mahu Sag, Junggar Basin, Western China. Mar. Petrol. Geol. 150, 106168. https://doi.org/10.1016/j.marpetgeo.2023.106168.

Lu, G., Zeng, L., Liu, G., et al., 2024a. Fracture identification based on graph pooling and graph construction in continental shale. IEEE Trans. Geosci. Remote Sens. 62, 4506020. https://doi.org/10.1109/TGRS.2024.3397861.

Lu, G., Zeng, L., Liu, G., et al., 2024b. Bedding-parallel fracture density prediction using graph convolutional network in continental shale oil reservoirs: a case study in Mahu Sag, Junggar Basin. China. Mar. Petrol. Geol. 167, 106992. https://doi.org/10.1016/j.marpetgeo.2024.106992.

Lu, G., Zeng, L., Liu, G., et al., 2025a. Fault detection based on U-Net and GNN integration. J. Struct. Geol. 197, 105426. https://doi.org/10.1016/j.jsg.2025.105426.

Lu, G., Zeng, L., Liu, G., et al., 2025b. Subsurface natural fracture identification using an integrated ensemble learning method. IEEE Trans. Geosci. Remote Sens. 63, 4700716. https://doi.org/10.1109/TGRS.2024.3525191.

Ma, S., Zeng, L., Tian, H., et al., 2023. Fault damage zone and its effect on deep shale gas: Insights from 3D seismic interpretation in the southern Sichuan Basin, China. J. Struct. Geol. 170, 104848. https://doi.org/10.1016/j.jsg.2023.104848.

Mika, S., Ratsch, G., Weston, J., et al., 1999. Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No. 98TH8468), pp. 41–48.

Rempe, M., Mitchell, T., Renner, J., et al., 2013. Damage and seismic velocity structure of pulverized rocks near the San Andreas Fault. J. Geophys. Res. Solid Earth 118 (6), 2813–2831. https://doi.org/10.1002/jgrb.50184.

Robledo Carvajal, F., Butler, R.W.H., Bond, C.E., 2023. Mapping faults in 3D seismic data–why the method matters. J. Struct. Geol. 177, 104976. https://doi.org/10.1016/j.jsg.2023.104976.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Savage, H.M., Brodsky, E.E., 2011. Collateral damage: Evolution with displacement of fracture distribution and secondary fault strands in fault damage zones. J. Geophys. Res. 116 (B3), B03405. https://doi.org/10.1029/2010JB007665.

Scarselli, F., Gori, M., Tsoi, A.C., et al., 2009. The graph neural network model. IEEE Trans. Neural Netw. 20 (1), 61–80. https://doi.org/10.1109/TNN.2008.2005605.

Schueller, S., Braathen, A., Fossen, H., et al., 2013. Spatial distribution of deformation bands in damage zones of extensional faults in porous sandstones: Statistical analysis of field data. J. Struct. Geol. 52, 148–162. https://doi.org/10.1016/j.jsg.2013.03.013.

Shang, X., Long, S., Duan, T., 2021. Fracture system in shale gas reservoir: Prospect of characterization and modeling techniques. J. Nat. Gas Geosci. 6 (3), 157–172. https://doi.org/10.1016/j.jnggs.2021.06.001.

Shi, J.X., Zhao, X.Y., Zeng, L.B., et al., 2024. Identification of reservoir types in deep carbonates based on mixed-kernel machine learning using geophysical logging data. Petrol. Sci. 21 (3), 1632–1648. https://doi.org/10.1016/j.petsci.2023.12.016.

Su, X., Zhu, R., Zhang, J., et al., 2025. Multi-scale characterization and control factors of bedding-parallel fractures in continental shale reservoirs: Insights from the Qingshankou Formation, Songliao Basin. China. Mar. Petrol. Geol. 182, 107580. https://doi.org/10.1016/j.marpetgeo.2025.107580.

Tang, Y., Cao, J., He, W.J., et al., 2021. Discovery of shale oil in alkaline lacustrine basins: The Late Paleozoic Fengcheng Formation, Mahu Sag, Junggar Basin, China. Petrol. Sci. 18 (5), 1281–1293. https://doi.org/10.1016/j.petsci.2021.04.001.

Tokhmechi, B., Memarian, H., Noubari, H.A., et al., 2009. A novel approach proposed for fractured zone detection using petrophysical logs. J. Geophys. Eng. 6 (4), 365–373. https://doi.org/10.1088/1742-2132/6/4/004.

Wang, X., Jin, Z., Chen, G., et al., 2022. Multi-scale natural fracture prediction in continental shale oil reservoirs: A case study of the Fengcheng Formation in the Mahu Sag, Junggar Basin, China. Front. Earth Sci. 10, 929467. https://doi.org/10.3389/feart.2022.929467.

Wu, X., Liang, L., Shi, Y., et al., 2019. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics 84 (3), IM35–IM45. https://doi.org/10.1190/geo2018-0646.1.

Yao, Y., Zeng, L., Mao, Z., et al., 2023. Differential deformation of a strike-slip fault in the Paleozoic carbonate reservoirs of the Tarim Basin, China. J. Struct. Geol. 173, 104908. https://doi.org/10.1016/j.jsg.2023.104908.

Zeng, H., 2013. Frequency-dependent seismic-stratigraphic and facies interpretation. AAPG Bull. 97 (2), 201–221. https://doi.org/10.1306/06011212029.

Zeng, H., 2018. What is seismic sedimentology? A tutorial. Interpretation 6 (2), SD1–SD12. https://doi.org/10.1190/INT-2017-0145.1.

Zeng, H., Kerans, C., 2003. Seismic frequency control on carbonate seismic stratigraphy: A case study of the Kingdom Abo sequence, west Texas. AAPG Bull. 87 (2), 273–293. https://doi.org/10.1306/08270201023.

Zeng, L., Shu, Z., Lyu, W., et al., 2021. Lamellation fractures in the Paleogene continental shale oil reservoirs in the Qianjiang Depression, Jianghan Basin, China. Geofluids 2021, 6653299. https://doi.org/10.1155/2021/6653299.

Zeng, L., Tang, X., Wang, T., et al., 2012. The influence of fracture cements in tight Paleogene saline lacustrine carbonate reservoirs, western Qaidam Basin, northwest China. AAPG Bull. 96 (11), 2003–2017. https://doi.org/10.1306/04181211090.

Zhang, S., Li, X., Zong, M., et al., 2017. Learning *k* for kNN classification. ACM Trans. Intell. Syst. Technol. 8 (3), 1–19. https://doi.org/10.1145/2990508.

Zhi, D., Tang, Y., He, W., et al., 2021. Orderly coexistence and accumulation models of conventional and unconventional hydrocarbons in Lower Permian Fengcheng Formation, Mahu sag, Junggar Basin. Petrol. Explor. Dev. 48 (1), 43–59. https://doi.org/10.1016/S1876-3804(21)60004-6.