



ARTICLE

Formation Pore Pressure Detection Using a Hybrid Model of Convolutional Neural Network and Machine Learning under Physical Constraints

Xinniu Xu¹, Jinde Li², Hong Huang¹, Guangfu Cao¹, Yuhe Shi², Biao Ruan¹, Daojin Ge¹ and Hu Yang^{2,*}

¹Exploration Division, Xinjiang Oil Field Company, Karamay, China

²Petroleum School, China University of Petroleum (Beijing) at Karamay, Karamay, China

*Corresponding Author: Hu Yang. Email: yanghu@cupk.edu.cn

Received: 18 December 2025; Accepted: 24 March 2026; Published: 18 October 2026

ABSTRACT: Accurate detection of formation pore pressure is a critical technology for ensuring the safety of oil and gas drilling and reservoir evaluation. Due to the heterogeneity of geological structures and the nonlinearity of stress-strain relationships, traditional empirical formulas and numerical simulation methods often fail to meet the precision requirements when detecting formation pore pressure, constrained by the limited amount of logging data. The physically-constrained convolutional neural network-Transformer (Phys-CNN-Transformer) hybrid architecture proposed in this paper enhances the accuracy and generalization capability of formation pore pressure detection. The convolutional operations in CNN learn the mapping relationships of local features in well-logging parameters, effectively capturing local correlations within well-logging curves. The Transformer model establishes global temporal and spatial connections among eight formation parameters—*Depth*, acoustic interval transit time (*AC*), bulk density (*DEN*), gamma ray (*GR*), compensated neutron log (*CNL*), borehole diameter (*CAL*), flushed zone resistivity (*RXO*), and spontaneous potential (*SP*)—across different formation depths, thereby improving the generalization ability of the pore pressure detection model. Further physical function constraints were applied to the CNN-Transformer model: utilizing the Ldata primary loss function and Eaton's method, which achieved an R^2 result of 0.968. This study utilized actual logging data from eight wells in the Mahu Oilfield (Well Mahu 1, Well Mahu 2, Well Mahu 8, Well Mahu 6, Well Mahu 15, Well Mahu 16, Well Mahu 25, and Well Mahu 26) as training and testing sets. During the training process, hyperparameter tuning and ablation comparison tests were employed to enhance the accuracy of the physically constrained CNN-Transformer model. The Phys-CNN-Transformer model was compared with other convolutional neural network models, including LSTM and BP. The results demonstrate that this approach offers a new perspective for advancing logging technology and improving the detection accuracy of formation pore pressure.

KEYWORDS: Convolutional neural network; transformer; physical constraint function; Mahu oilfield; formation pore pressure detection

1 Introduction

Formation pore pressure, as a core parameter in drilling engineering design, directly impacts drilling safety and economic efficiency. Furthermore, the precise detection of formation pore pressure enhances our understanding of the sedimentary evolution history and reveals the patterns of hydrocarbon migration and trap mechanisms [1]. During the drilling process, real-time and accurate detection of formation pore pressure values can reduce downhole accidents such as blowouts and drilling fluid loss, indirectly lowering drilling costs [2,3]. Traditional empirical formulas (e.g., Eaton's method, Bowers' method) rely on linear

simplification assumptions and struggle to address the nonlinear coupling issues among parameters such as wellbore enlargement and abnormal acoustic transit time. Although numerical simulation methods can depict complex geological structures, they rely on preset boundaries and expert experience, which limits their effectiveness in detecting formation pore pressure within concealed structures, such as unknown sealed high-pressure zones and isolated fracture zones. Additionally, 3D simulations incur high costs. Both approaches have significant shortcomings in nonlinear modeling and engineering practicality. In contrast, machine learning models, particularly deep learning models, can autonomously learn representative features from data, adapt to diverse geological scenarios, and improve the accuracy and transferability of formation pore pressure detection.

Prior to build a formation pore pressure detection model, standardized data preprocessing is essential. For data quality optimization, Patel et al. [4] systematically evaluated the effectiveness of Data Envelopment Analysis (DEA) and Principal Component Analysis (PCA) in the energy sector. Regarding incomplete data, Salem et al. [5] established a hybrid framework incorporating imputation and outlier tolerance, demonstrating its applicability to large-scale oil well datasets. He et al. [6] developed a comprehensive preprocessing solution for complex geological environments, encompassing data screening, cleaning, and processing to enhance data reliability. The current technical framework primarily comprises core modules such as data cleaning, augmentation, and resampling [7], with the choice of methods depending on the characteristics of specific well datasets.

In terms of AI applications, Farsi et al. [8] developed a pore pressure detection model for the Malun Oilfield by integrating extreme learning machines with a particle swarm optimization algorithm, which achieved higher accuracy than traditional methods. Li et al. [9] constructed a system for abnormal pressure identification based on K-Nearest Neighbors (KNN), Random Forest (RF), Extremely Randomized Trees (ET), and Light Gradient Boosting Machine (Light GBM). Abdulmalek et al. [10] applied fuzzy logic to drilling parameters, confirming its advantages over empirical models. Research indicates that through innovations in feature engineering and adaptive hyperparameter tuning, machine learning algorithms often surpass traditional methods in detection performance.

In the application of deep learning models, the Back Propagation (BP) neural network remains dominant due to its simple structure and strong interpretability. For instance, Xu et al. [11] employed support vector machines and BP networks to effectively reduce pressure detection errors in the Yingqiong Basin of the South China Sea; Rashidi and Asadi [12] validated the applicability of the BP model in Iranian sandstone formations based on two key drilling parameters—mechanical specific energy and drilling efficiency; Rahmanifard and plaksina [13] proposed integrating neural networks with reservoir simulation software to achieve tasks such as porosity detection and drilling fluid design; Kuang's team [14] demonstrated that computer vision techniques can be transferred to seismic fault identification, while data mining techniques are suitable for well-log curve reconstruction; Otchere et al. [15] compared mainstream models in the petroleum field and found that hybrid models outperform single models in parameter optimization and weight adjustment.

Sun et al. [16] proposed a CNN-Transformer hybrid model for porosity prediction using 1524 sets of logging data (AC, GR, CNL, DEN), achieving excellent results on the test set with $R^2 = 0.95647$, RMSE = 0.76363, and MAPE = 5.87%, thereby validating the superiority of the CNN-Transformer hybrid architecture in well-log parameter sequence modeling. Meng et al. [17] developed a CNN-Transformer hybrid neural network (with 8.59M parameters) integrated with physical information matrices for permeability prediction on the DeePore porous media dataset (6250 samples), attaining accuracies of $R^2 = 0.9857$ (CNN-Transformer) and $R^2 = 0.9907$ (PhyCNN-Transformer) on the validation set. Furthermore, through transfer learning, their model achieved $R^2 = 0.9563$ on differently shaped porous media with only 300 training samples, significantly outperforming 3D CNN (20.2M parameters, $R^2 = 0.5484$). Cheng and Zhang [18]

proposed, for the first time, a CNN-Transformer model incorporating the S-ReLU activation function for formation density prediction in tight sandstone reservoirs in the western Ordos Basin. Utilizing detector count rate data from nine wells (1680 samples) as the training set and data from three wells (210 samples) as the test set, their results demonstrated that the CNN-Transformer model achieved superior prediction accuracy ($R^2 = 0.958$, RMSE = 0.049) and robustness compared to traditional inversion models, pure CNN models, and pure Transformer models, providing a reliable and rapid new method for formation density prediction. Although deep learning techniques have achieved success in numerous tasks with similar data types in petroleum geology, their application in formation pore pressure detection remains in the exploratory stage.

Zhang et al. [19] selected the central anticline belt of the Xihu Sag in the East China Sea Shelf Basin as their study area and constructed a deep learning-based anomaly formation pore pressure prediction model. By comparing and analyzing different input parameter combinations (GR, AC, Density, RXO) and various deep learning algorithms (BP, CNN, LSTM), the optimally selected CNN model achieved precise prediction of abnormal overpressure, with the R^2 between predicted and measured pressure coefficients exceeding 0.90 and an average absolute error of less than 0.05. Huang et al. [1] utilized 2900 sets of logging data (90% as training set, 10% as test set) from the 3726–4015.9 m interval of exploration well LD10-A in the South China Sea, provided by the China offshore petroleum company, with P-wave velocity, shale content, density, and porosity as inputs. They pioneered the integration of machine learning techniques with the effective stress theorem for formation pore pressure prediction. By comparing five algorithms (Multilayer Perceptron, Radial Basis Function, Support Vector Machine, Random Forest, Gradient Boosting Machine), they found that the Multilayer Perceptron (MLP) neural network performed optimally, achieving $R^2 = 0.9976$ and RMSE = 0.00788 g/cm³ on the test set, and $R^2 = 0.9981$, RMSE = 0.00718 g/cm³ on the overall dataset, significantly outperforming traditional parametric models (RMSE = 0.34053 g/cm³). Zhang et al. [20] proposed an adaptive physics-informed deep learning model named CGP-NN (combining 1DCPP feature extraction, multi-layer GRU, and an adaptive physics-informed loss function). By integrating seismic data with Eaton/Bowers physical constraints, their case study on complex oil and gas reservoirs in the Tarim Basin validated the high accuracy ($R^2 = 0.94$) and strong generalization capability of pore pressure prediction in new wells, with the optimal generalization achieved at a physical constraint weight of $\lambda = 0.5$.

In summary, existing research either relies solely on statistical learning without incorporating physical law constraints, or fails to achieve deep integration of geological features with network frame. The CNN-Transformer hybrid architecture captures the local details of thin interbedded layers with small convolution kernels and extracts the macroscopic deposition patterns of thick layers with large convolution kernels through parallel dilated convolution design. In the feature extraction stage, multi-scale geological information fusion is realized, which lays a foundation for physical constraints. Based on this, this paper takes the physical law of Eaton's method as the loss function to construct the Phys-CNN-Transformer model. The local abnormal features of logging data are extracted by CNN, and the feature dimension is reduced by pooling down sampling. Transformer alleviates the problem of long-sequence gradient explosion, dynamically fuses thin-layer and thick-layer features through parallel dilated convolution, and combines electrical gradient attention weight and lithology sequence position coding to achieve deep fusion of geological features and network architecture. Furthermore, through the dual physical constraint loss function, the physical relationship between vertical stress, normal pore pressure and acoustic time difference anomaly is coupled to the training target in the form of residual, and the physical law is used to guide the parameter update in the back propagation-to realize the data and physical collaborative optimization, and improve the detection accuracy.

2 CNN-Transformer Theory

The original Transformer model was proposed by Li et al. [21] in 2017. The Transformer model is a deep learning architecture based on the self-attention mechanism, which abandons traditional recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), providing a new paradigm for sequence data processing. The core advantages of the Transformer model lie in its global dependency modeling and adaptive extraction of multi-scale features. Through the self-attention mechanism, the model can capture the pressure transmission relationships across entire well sections and identify the spatial distribution patterns of fault sealing and overpressure systems. Leveraging the multi-head attention architecture, it automatically learns multi-scale pressure features ranging from thin-layer interactions to sedimentary cycles, overcoming the limitations of traditional methods under complex geological conditions and improving pressure detection accuracy and geological applicability. Fig. 1 illustrates the basic structure of the Transformer module.

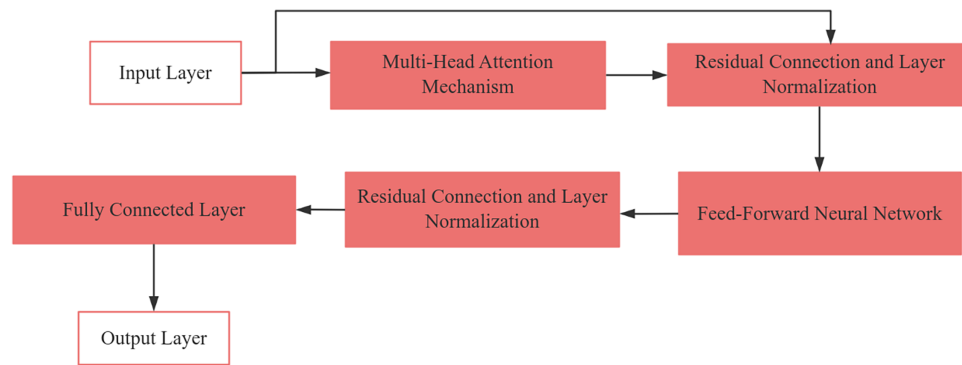


Figure 1: Lightweight transformer network architecture.

In the task of formation pore pressure detection, the Transformer module offers the following advantages:

(1) Through the self-attention mechanism, the model can apply global attention to each position in the sequence. For well-log data, this means that when detecting formation pore pressure at each depth point, information across the entire depth range can be comprehensively considered, effectively capturing global dependencies among porosity values.

(2) The multi-head self-attention mechanism (MHSA) enables the model to learn features at different scales, thereby adapting to information representation at various levels. Well-log data typically contains multi-scale features, such as local porosity fluctuations at small scales and formation trend variations at large scales.

(3) Residual connections allow the input of a neural network layer to be directly added to its output, forcing the network to learn residual mappings between inputs and outputs. This ensures that gradients during back-propagation of formation information can flow directly back to shallow layers, guaranteeing that the performance of deep models is no worse than that of shallow benchmarks. It effectively mitigates gradient vanishing or explosion issues and facilitates the construction of ultra-deep networks.

(4) The feed-forward neural network (FFN) module introduces nonlinear transformations to enhance feature representation, further improving the detection accuracy of formation pore pressure in complex geological formations.

To further enhance the model's ability to extract local features of formation parameters, the Convolutional Neural Network (CNN) is introduced. CNN was first proposed by LeCun et al. [22] in 1989. This architecture helps more effectively capture local patterns in input data, such as feature variations within specific depth ranges, thereby improving the accuracy of porosity detection. Additionally, formation structures often exhibit local characteristics in the vertical direction, such as thin layers or local lithological variations. The introduction of local information layers enhances the model's sensitivity to such structures. The pooling layer performs downsampling on feature maps, retaining significant features like maximum values to reduce dimensionality and suppress noise, while also improving the model's robustness against minor disturbances, providing efficient and stable feature representations for subsequent modules.

The local sensitivity of CNN ensures precise identification of microscopic anomalies, while the global perspective of Transformers enables the inference of spatial propagation paths of pressure anomalies. The complementary nature of these two approaches avoids the limitations of traditional methods that rely on empirical formulas or the restricted perspective of a single model, significantly enhancing the accuracy and robustness of pressure detection under complex formation conditions. The entire CNN-Transformer model combines the advantages of CNN in local feature extraction with the global dependency modeling capability of Transformers, thereby achieving precise detection of formation pore pressure.

3 Methodology

3.1 Data Preprocessing

The data was sourced from eight wells in the Mahu Oilfield of the Junggar Basin, Xinjiang, China, namely Mahu 1, well Mahu 2, well Mahu 8, well Mahu 6, well Mahu 15, well Mahu 16, well Mahu 25, and well Mahu 26, comprising a total of 48,118 sets of logging data as experimental data. First, the random forest importance coefficient method was employed to screen the eight most influential parameters for formation pore pressure calculation: *Depth*, *AC*, *DEN*, *GR*, *CNL*, *CAL*, *R XO*, *SP*. The random forest importance coefficient method is a non-parametric model suitable for handling non-normal and non-linear data. When the cumulative contribution of formation-related parameters to formation pore pressure exceeds 85%, it can be concluded that the selected parameters encompass the majority of significant influence information on formation pore pressure within the data, making them core driving factors for the detection model.

This method uses the node splitting criterion of decision tree in random forest to calculate the feature contribution, counts the cumulative contribution of features in all trees, and selects the key features. The core idea is: If some of the 12 logging and rock mechanics parameters significantly reduce the prediction error in the splitting process of the majority tree, then they are more important. The detailed steps are as follows:

(1) Constructing the Random Forest Model

Using the Self-sampling method, N subsets containing N well depth data points (N is the number of decision trees) are extracted with replacement from the original dataset. For each decision tree, m parameters are randomly selected from the 12 logging and rock mechanics parameters as candidate splitting features.

(2) Calculating Feature Importance for a Single Tree

For each decision tree, traversing all node splitting processes, the importance of feature f in a single tree is the cumulative sum of $\Delta G(f, t)$ during the splitting of all nodes. The calculation of $\Delta G(f, t)$ is shown in Eqs. (1)–(3):

$$\Delta G(f, t) = G_{\text{parent}}(t) - \left(\frac{n_{\text{left}}}{n_{\text{parent}}} G(t_{\text{left}}) + \frac{n_{\text{right}}}{n_{\text{parent}}} G(t_{\text{right}}) \right) \quad (1)$$

$$G(t) = 1 - \sum_{i=1}^C p\left(\frac{i}{t}\right)^2 \quad (2)$$

$$I_f = \sum_{t \in \text{Split}(f)} \Delta G(f, t) \quad (3)$$

where, $G(t)$ is the Gini impurity of node t ; C is the number of classification categories; $p\left(\frac{i}{t}\right)$ is the sample proportion of class i in node t ; $\Delta G(f, t)$ is the reduction in Gini impurity when node t splits using feature f ; $G_{\text{parent}}(t)$ is the Gini impurity of the parent node; n_{parent} is the number of samples in the parent node; n_{left} , n_{right} are the number of samples in the left and right child nodes; I_f is the total contribution of feature f in reducing classification uncertainty during the splitting process.

(3) Aggregating Global Feature Importance

A higher global importance coefficient indicates a greater influence of the feature on predicting the target variable. By averaging the feature importance across all trees in the random forest, the global importance coefficient I_f^{global} is obtained, as shown in Eq. (4):

$$I_f^{\text{global}} = \frac{1}{N} \sum_{k=1}^N I_{fk} \quad (4)$$

The results are shown in Table 1, and Fig. 2 is derived from the data in Table 1. It can be observed that the importance of AC is 21.38%, the importance of DEN is 18.72%, and the importance of CAL is 13.15%. The importances of GR, CNL, RXO, SP, and well Depth are 9.89%, 8.24%, 7.56%, and 5.76%, respectively, with one value at 6.31%. The cumulative contribution of the first eight formation parameters reaches 91.01%, which already covers the main information of the original data. In contrast, the contribution of the ninth parameter, true formation resistivity, is only 2.85%, and the information gain from other parameters is also limited. Therefore, a truncation measure can be adopted. Consequently, this paper selects AC, DEN, GR, CNL, RXO, SP, and well Depth as the input feature parameters for the subsequent neural network.

Table 1: Feature importance coefficients.

Input Feature	Importance Coefficient
AC	0.2138
DEN	0.1872
CAL	0.1315
GR	0.0989
Single axis compressive strength	0.0212
CNL	0.0824
R XO	0.0756
Well depth	0.0631
SP	0.0576
True resistivity	0.0285
Coefficient of internal friction	0.0253
Tensile strength	0.0149

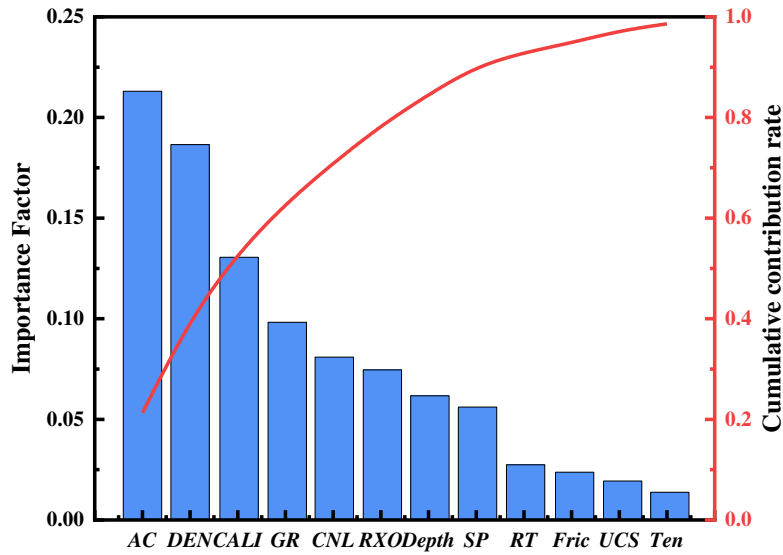


Figure 2: Cumulative contribution rate of features.

Parameters with a high proportion of missing data were removed, while missing values were filled using mean imputation [23]. Finally, the Min-Max method was applied to normalize the original data, scaling values such as acoustic travel time, formation density, AC, and CNL, etc. to a range of 0 to 1 to eliminate the influence of varying units across different features. The specific steps of the Min-Max method are as follows:

$$Y_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)} \quad (5)$$

where X_i is the original data, $X_{\min}$ is the minimum value of the feature in the dataset, $X_{\max}$ is the maximum value of the feature in the dataset, and Y_i is the normalized data.

These data are used to obtain the physical properties of subsurface rocks to interpret formation structure and characteristics. The collected formation pore pressure data are derived from the commercial, actual production tests, well tests, overflow well control, and lost circulation measurements. The process involves first calculating initial pressure labels based on theoretical models, constructing a pressure-depth reference curve, and guiding drilling operations. The formation pore pressure curve is then adjusted according to field conditions. In cases of abnormal conditions, the pressure values are correspondingly modified, and the pore pressure curve is corrected by comparing the measured pore pressure coefficient values (the pore pressure coefficient: The formation pressure coefficient is the ratio of formation pressure to hydrostatic pressure, which is used to judge whether the formation pressure is abnormal and to guide drilling and oil and gas development). The corrected pore pressure curve closely aligns with and is slightly lower than the drilling fluid density, ensuring safe guidance in actual drilling operations. After review and approval by field engineers, the corrected pore pressure coefficient is considered valid and reliable. In subsequent studies, the calibrated pore pressure coefficient values are used as labels for formation pore pressure detection. The dataset is then divided into training and testing sets. Logging data from six wells, including well Mahu 2, well Mahu 8, and well Mahu 16, is used as the training set (33,244 samples), while data from well Mahu 6 serves as the validation set for model hyperparameter tuning (6882 samples). Logging data from well Mahu 1 is used as the test set to evaluate model accuracy (7992 samples). To ensure sufficient model training and accurate evaluation, both the training and test sets exclusively use data from the Mahu.

3.2 Phys-CNN-Transformer Model Architecture

The concept of fine-tuning deep learning models using optimization algorithms has demonstrated significant success across various domains, such as enhancing photovoltaic panel diagnosis through fine-tuned convolutional neural networks. These approaches typically employ metaheuristic algorithms—including Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and Bayesian Optimization—to automatically search for optimal hyperparameters or network architectures, thereby improving model accuracy and generalization. In the context of formation pore pressure detection, such optimization techniques offer promising potential. For instance, optimization algorithms could automatically determine the optimal dilation rates for the parallel convolutional branches in our CNN module or the optimal number of attention heads in the Transformer encoder, potentially yielding even higher accuracy than manually tuned configurations. However, the integration of optimization algorithms with physics-constrained models like Phys-CNN-Transformer remains an underexplored area. While the current study employs manual hyperparameter tuning via grid search, future work could leverage advanced optimization algorithms to achieve more efficient and adaptive model fine-tuning, particularly when dealing with diverse geological conditions and limited training data.

Fig. 3 illustrates the flowchart of the CNN-Transformer-based formation pore pressure detection algorithm incorporating physical constraints.

First, input the preprocessed formation pore pressure parameters into the CNN. Upon detection, the CNN captures local rock physical properties along the well depth direction and local morphological features of logging curves through multi-scale convolutional kernels, identifying microscopic anomaly signals such as sudden porosity increases and borehole enlargement. It automatically learns the local synergistic relationships among multiple parameters. Considering the significant vertical heterogeneity of the Mahu formation, a 3×1 dilated convolutional kernel ($Q = 4$) is employed here to expand the receptive field to 3 m, covering thin interbedded sequences. Multi-channel input is also used to suppress borehole collapse noise interference, thereby equipping the model with high-frequency signal capture capability. Fig. 4 illustrates the dilated convolutional structure.

Dilated convolution inserts fixed intervals of voids between elements of the standard convolution kernel, skipping certain positions, thereby expanding the receptive field without increasing the number of parameters. The interval of inserted voids is represented by the dilation rate. By adjusting the dilation rate, dilated convolution can flexibly control the feature extraction range, making it suitable for analyzing cross-depth parametric correlations in stratigraphic data. Large-kernel convolution employs a 30×1 kernel, covering a 15 m range, and down-samples to a 3 m resolution, effectively extracting the dominant frequency of cycles.

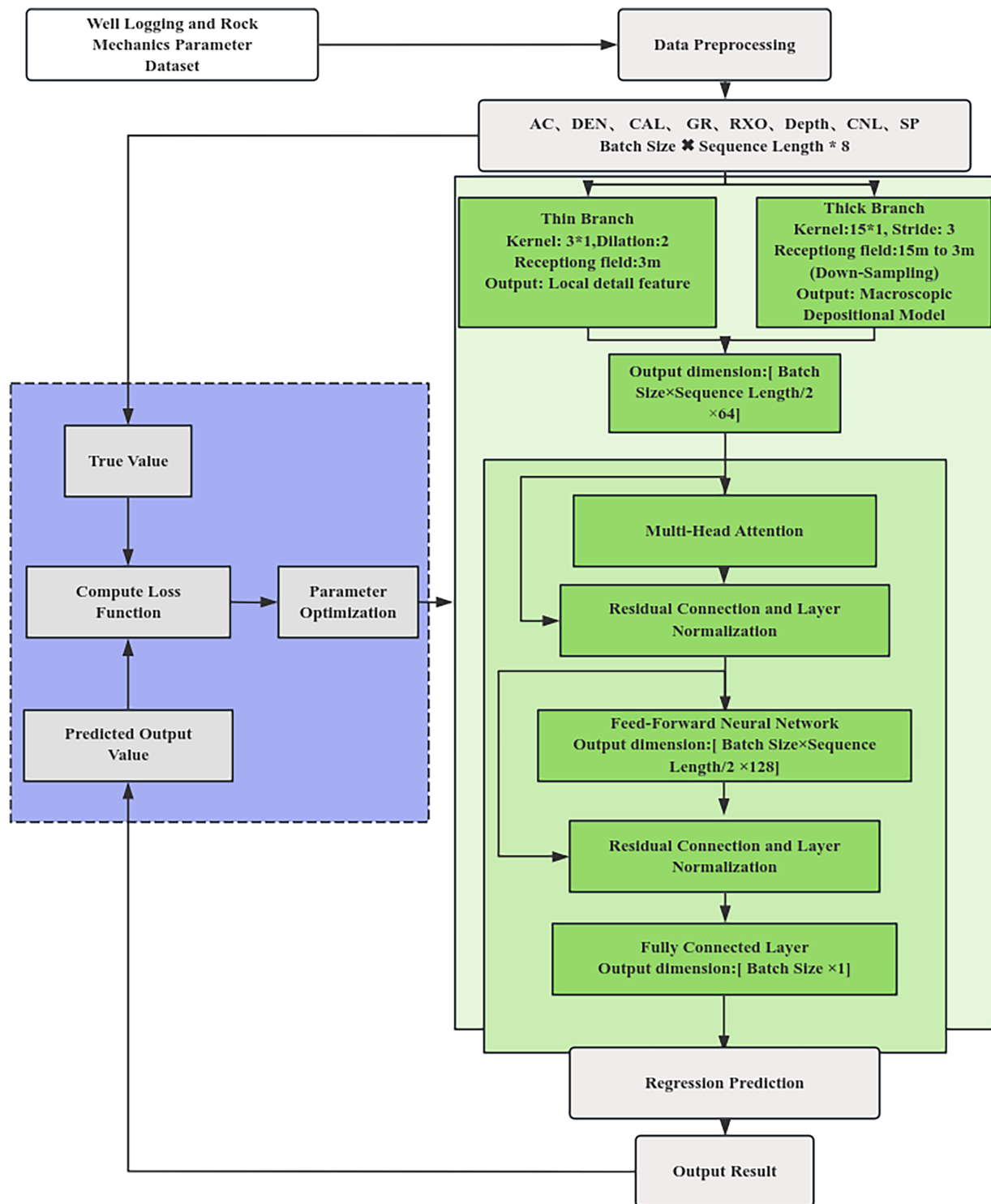


Figure 3: Workflow diagram of formation pressure prediction algorithm based on CNN-transformer network.

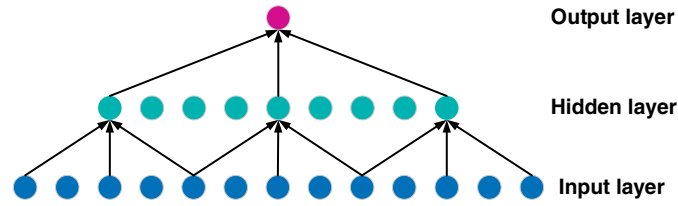


Figure 4: Dilated convolution structure.

To balance local details and global trends, a parallel architecture of thin-layer and thick-layer branches is constructed, and weight coefficients are calculated to achieve dynamic weighted fusion, as shown in Eq. (6):

$$F_{out} = w_{thin} \cdot F_{thin} + w_{thick} \cdot Upsample(F_{thick}) \quad (6)$$

where F_{thin} is the output feature map of the thin branch, F_{thick} is the output feature map of the thick branch, and w_{thin} and w_{thick} are the weight coefficients.

The weight coefficients are dynamically adjusted based on the vertical proportion of the stratum [24], as shown in Eq. (7):

$$w_{thin} = \frac{N_{thin}}{N_{total}}, w_{thick} = 1 - w_{thin} \quad (7)$$

where N_{thin} is the number of sampling points in the thin interbedded section, and N_{total} is the total number of samples.

After dataset training and hyperparameter optimization, CNN adopts a thin-layer kernel size of 3×1 , a thin-layer dilation rate of 2, a thick-layer kernel size of 15×1 , and a stride of 3. The processed data is then fed into the Transformer encoder module, where the self-attention mechanism dynamically analyzes long-range mechanical correlations across the entire well section. Automatic well-depth positional encoding injects depth sequence and distance information, enabling the model to comprehend the normal compaction trend of formation pore pressure and abrupt changes caused by faults. Additionally, the self-attention mechanism dynamically adjusts parameter weights to analyze correlations between parameters at any depth point within the well section. The hyperparameter-tuned model configuration includes 4 encoder layers, 8 attention heads, and a feedforward network hidden dimension of 512. The CNN-processed parameters—*Depth*, *AC*, *DEN*, *GR*, *CNL*, *CAL*, *R XO* and *SP*—are mapped into a 128-dimensional high-dimensional space, as shown in Eq. (8):

$$h_i = \text{ReLU}(W_e x_i + b_e) \quad (8)$$

where h_i is the high-dimensional embedded representation of the input features after nonlinear transformation, ReLU is the activation function, which sets negative values to zero and retains positive values, W_e is the trainable weight matrix that maps the original features to a 128-dimensional space, x_i is the input sequence, and b_e is the bias term.

Next, a hybrid positional encoding is employed to integrate well-depth vertical information, as shown in Eq. (9):

$$PE = \text{Concat}(P_{\text{linear}}, P_{\text{sin}}) \quad (9)$$

where PE is the mixed positional encoding, P_{linear} is the linear encoding, preserving absolute depth information, P_{sin} is the sinusoidal encoding, generating periodic positional signals to capture local correlations of depth intervals.

The linear encoding $Linear(d_i)$ directly maps the well depth to 64 dimensions, as shown in Eq. (10):

$$P_{\text{linear}} = Linear(d_i) = W_l d_i + b_l \quad (10)$$

where d_i is the normalized well depth.

The sinusoidal encoding generates periodic positional signals, as shown in Eq. (11):

$$\begin{aligned} P_{\sin}^{(d_i, 2k)} &= \sin\left(\frac{d_i}{10000^{2k/64}}\right) \\ P_{\sin}^{(d_i, 2k+1)} &= \cos\left(\frac{d_i}{10000^{2k/64}}\right) \end{aligned} \quad (11)$$

Concatenate the linear and sinusoidal encodings into a 128-dimensional positional encoding, then add it to the input features, as shown in Eq. (12):

$$z_i = h_i + PE \quad (12)$$

Next, the input is fed into the Transformer encoder, which employs a 4-layer multi-head self-attention structure with 8 attention heads. Each layer consists of dual modules: a self-attention mechanism and a feed-forward neural network (FFN). The multi-head attention layer dynamically calculates the feature correlation between any well-depth points using learnable weight matrices, establishing global dependencies across depths and cross-property parameters. The calculation formula is as follows:

$$MultiHead(Z^{(l)}) = \text{Concat}(head_1, \dots, head_8) W_O \quad (13)$$

Among them, for the input $Z^{(l)}$ at the l -th layer, the single-head attention calculation is:

$$head_i = \text{Attention}(Q_i, K_i, V_i) = \text{Softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (14)$$

here $Q = Z^{(l)} W_q$, $K = Z^{(l)} W_k$, $V = Z^{(l)} W_v$, where W_q , W_k , and W_v are weight matrices.

The feed-forward neural network (FFN) module introduces nonlinear transformations to enhance feature representation capability, as shown in Eq. (15). Each sublayer employs residual connections and layer normalization to ensure stable gradient propagation, with the computation expressed as:

$$FFN(Z^{(l)}) = \text{Dropout}(\text{ReLU}(W_1 Z^{(l)} + b_1)) \cdot W_2 + b_2 \quad (15)$$

For the l -th layer, the dropout rate is set to 0.2, and LayerNorm performs normalization on the feature dimension.

$$Z^{(l+1)} = \text{LayerNorm}(Z^{(l)} + \text{Dropout}(MultiHead(Z^{(l)}))) \quad (16)$$

Then, the fully connected layer maps the 128-dimensional encoding to the detected pore pressure value of the well depth:

$$P_{\text{pred}} = W_p Z^{(l)} + b_p \quad (17)$$

Finally, the detected pore pressure values are input into the physically constrained loss function. The primary loss function is the mean squared error, which directly constrains the deviation between the detected pore pressure and the measured values. The secondary loss function couples the Eaton method's physical constraints with vertical stress, normal pore pressure, and sonic time difference anomalies, ensuring that the pore pressure detection aligns with sedimentary formation laws and balances data accuracy with geological credibility.

(1) The primary loss function, L_{data} , directly constrains the error between the detected pore pressure and the measured pore pressure, with the specific as follows:

$$L_{data} = \frac{1}{N} \sum_{i=1}^N (P_{pred}^{(i)} - P_{true}^{(i)})^2 \quad (18)$$

where P_{pred} is the detected value, MPa; P_{true} is the true value, MPa.

(2) The secondary loss function L_{Eaton} represents the physical constraint of the Eaton method.

$$L_{Eaton} = \frac{1}{N} \sum_{i=1}^N (P_{pred}^{(i)} - [S_v^{(i)} - (S_v^{(i)} - P_n^{(i)}) \cdot \left(\frac{\Delta t_n^{(i)}}{AC_i}\right)^3])^2 \quad (19)$$

where S_v is the vertical stress, MPa; P_n is the normal pore pressure, MPa.

The calculation of vertical stress S_v and normal pore pressure P_n is as follows:

$$S_v^{(i)} = \sum_{k=0}^i DEN_k \cdot 9.81 \cdot 0.5 \quad (20)$$

$$P_n^{(i)} = 1.05 \cdot 9.81 \cdot \frac{Depth_i}{1000} \quad (21)$$

(3) The total loss function is:

$$L = L_{Data} + \lambda \cdot L_{Eaton} \quad (22)$$

where λ is a hyperparameter representing physics-related metrics.

When $L > 0.0025$, the phys-CNN-Transform model will receive parameter optimisation instructions to train the model structure until $L < 0.0025$, then continue to reuse the current phys-CNN-Transform model structure and monitor the pore pressure in various strata until λ stabilises.

3.3 Phys-CNN-Transformer Principle

To address how geological principles are integrated into the data-driven learning process, this section elaborates on the mechanism of the physics-constrained loss function and the specific implementation of Eaton's method within the Phys-CNN-Transformer framework.

3.3.1 Mechanism of Physics-Constrained Learning

Traditional pure data-driven models only learn from the statistical model of logging data, which may lead to mathematically accurate but geologically unreasonable predictions (e.g., violation of the effective stress principle or prediction of negative pore pressure). In order to overcome this limitation, we embed the laws of physics directly into the training target through a two-component loss function, as shown in [Formula \(22\)](#).

The integration of geological principles is achieved by:

L_{Data} : This is the mean square error between the pore pressure predicted by the model ($P_{pred}^{(i)}$) and the measured pore pressure from well logging and production data ($P_{true}^{(i)}$). This component ensures that the model accurately fits the actual field observation data.

L_{Eaton} (Physics-Constrained Component): This secondary loss function quantifies the difference between the predicted value of the model and the pore pressure value calculated by the Eaton method. Eaton method is a mature empirical relationship in rock physics, which describes the physical relationship between pore pressure and vertical stress, normal pressure and acoustic time difference anomaly. By minimizing this term, we force the prediction results of the model to be consistent with the physical laws that control the generation of sedimentary compaction and overpressure.

λ : This hyperparameter controls the trade-off between the fitted data and compliance with physical constraints. In the Mahu area, we set λ to 0.5 according to the empirical optimization, and reached an optimal equilibrium point, so that the model can not only respect the data fidelity, but also conform to the geological rationality.

The key innovation is that physical constraints act as a regularizer in the training process. When the prediction of the model deviates from the Eaton method, the L_{Eaton} term will produce a gradient, guiding the model to return to the geologically consistent solution space, especially in sparse data areas or complex overpressure zones. Pure data-driven learning may over-fit or produce physically unreasonable outputs.

3.3.2 Implementation of Eaton's Method

The Eaton method is directly implemented in the calculation of the secondary loss function. For each training sample, the model calculates a pore pressure reference value (L_{Eaton}) based on the Eaton method as follows:

(1) Calculation of vertical stress (S_v): The vertical stress represents the total weight of the overlying strata. Through the integral calculation of the formation density (ρ_b) from the surface to the target depth:

$$S_v = \int_0^Z \rho_b(z) g dz \quad (23)$$

In practice, this is done by numerical integration of discrete density logging data. This step integrates the basic geological principle of gravity loading, which is the main driving force of compaction and pore pressure evolution in sedimentary basins.

(2) Calculation of normal pore pressure (P_n): Normal pore pressure is assumed to be hydrostatic pressure, which represents the pressure exerted by continuous formation water column:

$$P_n = \rho_w g z \quad (24)$$

This step integrates the geological principle of hydrostatic balance in normally compacted and hydraulically connected strata.

(3) Determination of normal compaction trend line (∇t_n)

The normal compaction trend line represents the variation function of the expected acoustic time difference with depth in the normal pressure mudstone. For the Mahu area, we established this trend line by analyzing the acoustic logging response of the shallow normal pressure layer. This reflects the geological principle of normal compaction, that is, due to mechanical compaction, porosity decreases exponentially with depth.

(4) Calculate the residual based on Eaton's method

$$L_{Eaton} = \frac{1}{N} \sum_{i=1}^N (P_{pred,i} - P_{Eaton,i})^2 \quad (25)$$

In the back propagation process, the gradient of the loss term updates the weight of the model to minimize the difference between the data-driven predicted value and the physically derived Eaton estimated value.

4 Test Results and Analysis

Table 1 presents the values of four evaluation metrics for various neural network learning models. Four key metrics—*MSE*, *RMSE*, *MAE*, and R^2 [25]—were used to assess the models' performance in detecting formation pore pressure of the wells in Mahu. The *MSE*, *RMSE*, and *MAE* values reflect the discrepancy between the detected and actual formation pore pressure, indicating the precision of the model's detection. Lower values for these metrics signify better detection performance. The coefficient of determination (R^2) evaluates the overall fitting effectiveness of the model, representing its explanatory power. R^2 value closer to 1 indicates superior model performance, as expressed by the formula:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (26)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (27)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (28)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (29)$$

where $\bar{y}$ is the mean of the true values of formation pore pressure, $\hat{y}$ is the true value of formation pore pressure, and Y is the detected value of formation pore pressure, N is the number of sample points.

Fig. 5 presents the crossplots of each model. The crossplots verify the accuracy of the models by comparing the detected results with the actual values and analyzing the errors. Each plot includes a diagonal reference line (ideal detection line), where a lower degree of deviation indicates better model performance. From this perspective, the concentration of scatter points is ranked as follows: Phys-CNN-Transformer > CNN-Transformer > Transformer > CNN-LSTM > CNN-BP > LSTM > BP. The CNN-Transformer model combines the advantages of convolutional neural networks (CNN) and Transformers, making it suitable for processing high-dimensional data with complex spatiotemporal features, such as logging data. CNN excel at capturing local features, while Transformers handle global correlations, enabling the model to better adapt to the characteristics of logging data. A pure Transformer model may outperform CNN when processing spatiotemporal data, but could be slightly less effective in handling local features.

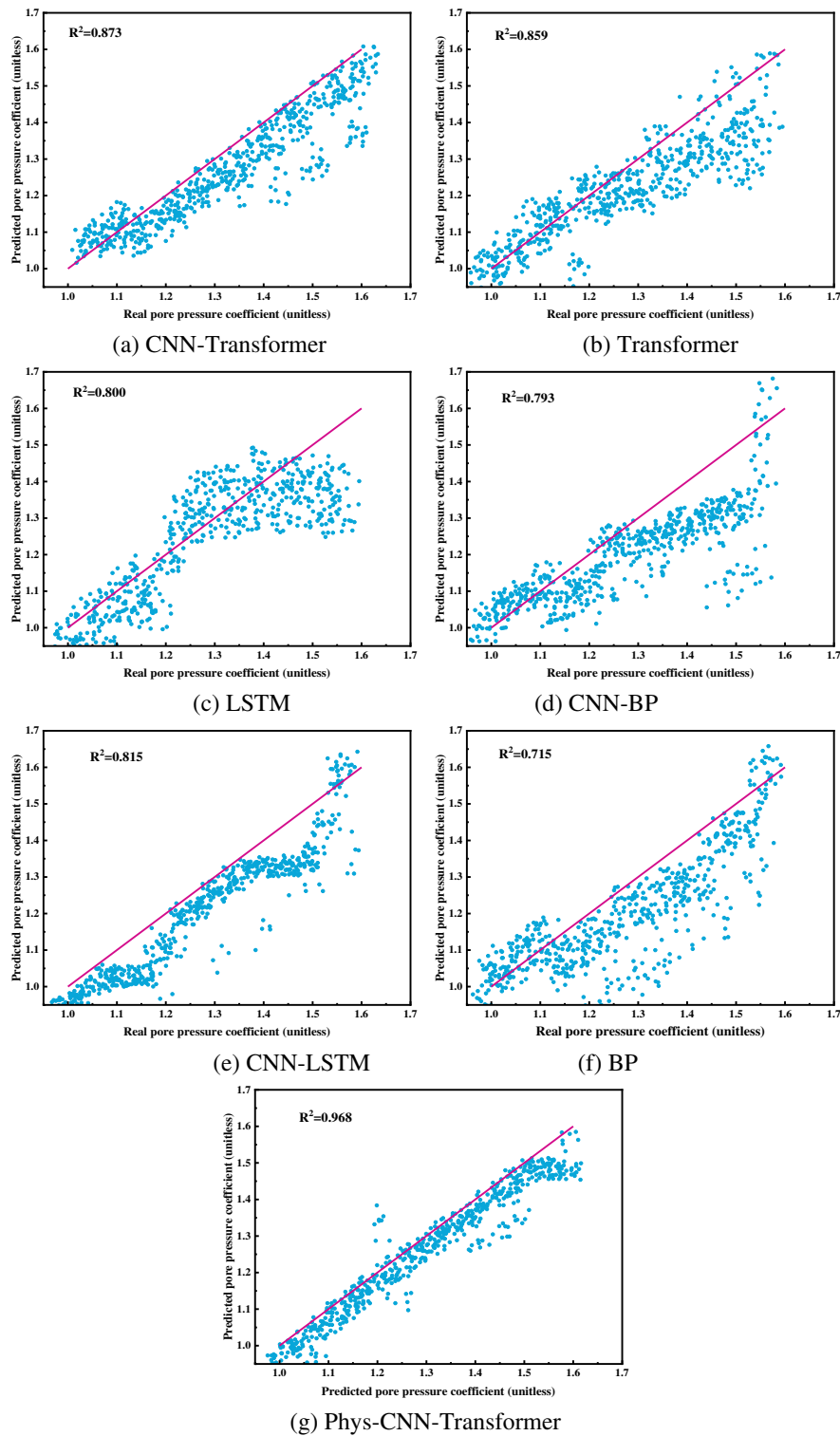


Figure 5: Intersection diagram of the neural network model.

CNN perform relatively well in capturing local features but may not be as effective as Transformers in modeling global dependencies. Deep learning models like CNN-Transformer and Transformer have a large

number of parameters and high model capacity, enabling better adaptation to complex data distributions. This helps improve generalization when sufficient data is available. BP neural network and LSTM were introduced to better highlight the advantages of hybrid Transformer-CNN models. The LSTM model can effectively capture the deep sequential dynamic features of formation pore pressure evolution, but local deviations of $\pm 10\%$ still exist between detected and true values, especially in high-value regions where the pore pressure coefficient exceeds 1.3. Data points consistently fall below the solid red line, indicating that the generalization capability of the LSTM architecture in modeling complex geological parameters does not fully meet the precision requirements of practical engineering. CNN-BP combines convolutional structures with BP neural networks but exhibit limited learning ability for long-distance vertical dependencies (e.g., overpressure systems spanning hundreds of meters). Comparative results demonstrate that the Transformer's unique multi-head attention mechanism effectively addresses the bottleneck of traditional sequence models in extracting long-range pore pressure transmission features through parallelized global dependency modeling. The standard deviation of detection errors for pore pressure anomaly fluctuation zones is further reduced compared to LSTM and BP neural network, confirming the superiority of the self-attention weight allocation mechanism in processing non-stationary geological signals. Based on [Table 2](#) and [Fig. 5](#), the phys-CNN-Transformer hybrid model more accurately captures the variation trends of formation pore pressure, reduces detection errors, and achieves the best detection performance.

Table 2: Comparison of model evaluation metrics.

Methods	Metrics			
	MSE	RMSE	MAE	R ²
BP	0.0051	0.0711	0.0571	0.715
LSTM	0.0041	0.0642	0.0374	0.800
Transformer	0.0034	0.0583	0.0337	0.859
CNN-BP	0.0044	0.0663	0.0486	0.793
CNN-LSTM	0.0039	0.0624	0.0367	0.815
CNN-Transformer	0.0027	0.0520	0.0288	0.873
Phys-CNN-Transformer	0.0017	0.0412	0.0269	0.968

To demonstrate the necessity of each submodule design in the physically constrained CNN-Transformer model, the independent contributions and synergistic effects of the following core optimization modules for pore pressure detection in the Mahu area will be validated:

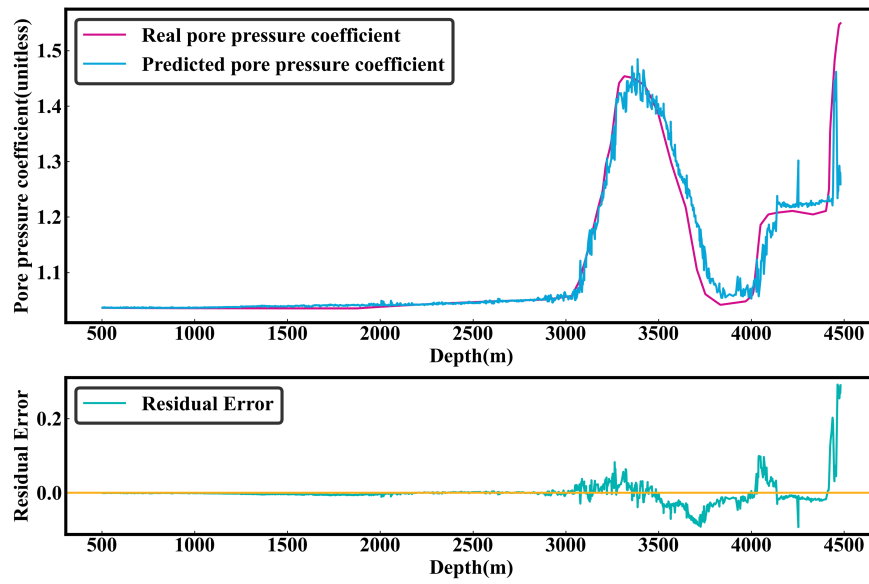
- (1) Parallel dilated convolution architecture: Thin-layer branch with 3×1 dilated convolution + thick-layer branch with 15×1 large-kernel convolution + dynamic weighted fusion;
- (2) Electrical gradient attention weights + dynamic lithology sequence positional encoding: Attention enhancement based on the vertical gradient of resistivity and frequency adjustment of positional encoding driven by GR values;
- (3) Physically constrained loss function: MSE + Eaton constraint. Thus, ablation tests are designed, with the test group configurations shown in [Table 3](#), where “-” indicates the removal of the corresponding module.

Table 3: Ablation testing design.

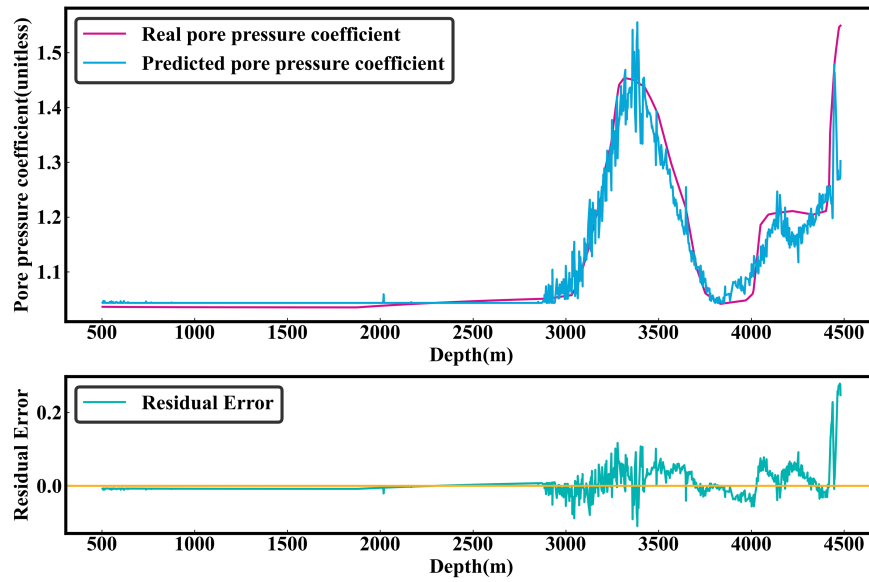
Run	(Network) Architecture Design	Objective
1	CNN-Transformer (Complete module)	Baseline performance
2	Parallel dilated convolutions replaced with a single-branch 3×1 convolution	Validate the effectiveness of multi-scale feature extraction
3	Electrical-lithology-guided Transformer replaced with standard attention + fixed positional encoding	Validate the effectiveness of the Transformer optimization
4	Hybrid physics-informed loss function replaced with <i>MSE</i> -only loss function	Validate the necessity of physics-constrained embedding

To ensure the comparability of ablation experiment results, all comparative experiments adopted the following unified training configuration: the ablation tests in this section uniformly used the Adam optimizer with an initial learning rate of 0.0001, a batch size of 32, and a maximum training epoch of 300. By employing the control variable method to replace core model modules, the independent contribution of each component to performance was validated. For each group of experiments, only a single variable was modified while keeping all other parameters identical to the baseline model. Each experiment was repeated three times, and the average was taken to eliminate the impact of randomness.

Fig. 6 presents the fitted detection curves from four sets of tests. In the well section from 500 to 3000 m, the detection values of all models coincide with the blue curve, indicating good performance. However, in the abnormally high-pressure section beyond 3000 m, deviations begin to emerge among the models. When replacing the parallel dilated convolution with a single-branch 3×1 convolution, the model exhibits smoother processing at the abrupt pore pressure transition interface but lags in capturing abrupt details. Substituting the electro-lithology-guided Transformer optimization module with standard attention and fixed positional encoding results in phase delay and amplitude attenuation at the pore pressure transition between 3700 and 4200 m, leading to significant errors. Replacing the loss function with only *MSE* introduces high-frequency oscillations, and the overall detection values fall below the actual values. In contrast, the complete model's detection values nearly overlap with the actual values, demonstrating excellent detection performance.

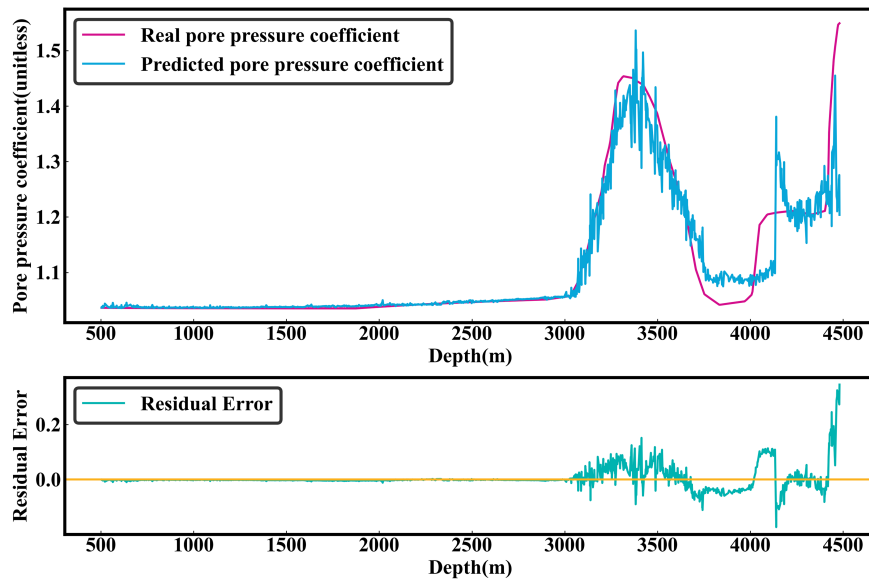


(a) Phys-CNN-Transformer

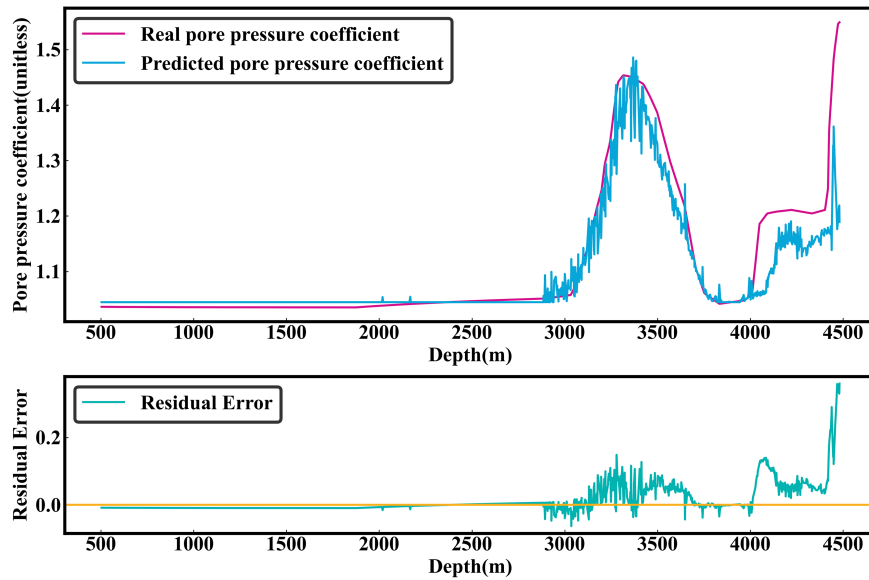


(b) MSE-only loss function

Figure 6: (Continued)



(c) single-branch convolution model



(d) fixed positional encoding model

Figure 6: Prediction fitting plots of different models.

Fig. 7 displays the crossplots of the four test models. The results indicate that removing any submodule from the complete CNN-Transformer model affects detection accuracy. The distribution of data points reveals that replacing the parallel dilated convolution architecture with a single-branch convolution has the most substantial impact, followed by using only *MSE* as the loss function, which causes noticeable dispersion and numerous off-diagonal data points. Replacing the optimized Transformer with traditional attention mechanisms and fixed positional encoding introduces significant deviations in processing pore pressure coefficient values between 1.4 and 1.5.

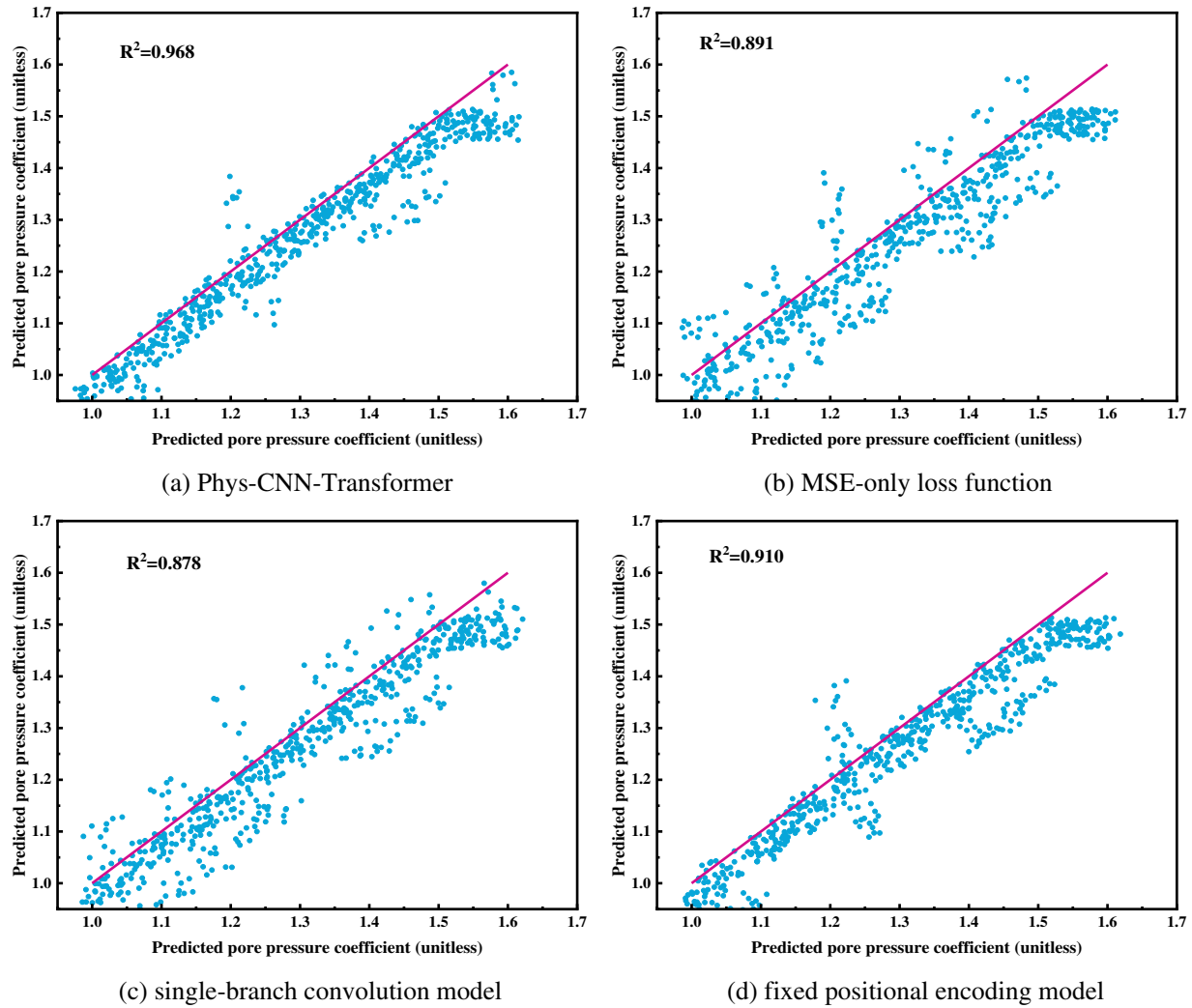


Figure 7: Intersection diagram of different models.

As shown in Table 4, by comparing the data from the three ablation tests with the complete CNN-Transformer model, it can be observed that each optimization component has a certain impact on the model. The CNN-Transformer model with multi-module collaboration achieves the lowest *MSE*, *RMSE*, and *MAE* values, along with the highest R^2 value of 0.968. When the parallel convolutional architecture is removed, R^2 decreases by 9.30%, indicating that for high-frequency oscillation features of thin layers, small-kernel dilated convolutions can enhance the capture of local details, while for meter-scale periodicities of thick-layer cycles, large-kernel strided convolutions extract macroscopic sedimentary patterns. Combined with a dynamic weight fusion mechanism, the model achieves adaptive adjustment to different stratigraphic structures, significantly influencing its performance. Removing the lithology-electrical property-guided Transformer optimization module reduces R^2 by 5.99%, demonstrating that the dual optimization of electrical gradient attention weights and lithology sequence positional encoding is crucial for improving model performance. Eliminating the physics-constrained loss function decreases R^2 by 7.95%, indicating that the designed hybrid loss function not only enhances detection accuracy but also enforces adherence to geological domain knowledge, achieving “physics-data” dual-driven optimization for better detection results.

Table 4: Comparison of model evaluation metrics.

Run	<i>MSE</i>	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²
Phys-CNN-Transformer model	0.0017	0.0412	0.0269	0.968
Parallel Dilated Convolutions	0.0026	0.0510	0.0263	0.878
-Electro-Lithology Guided Transformer	0.0023	0.0480	0.0333	0.910
Physics-Constrained Loss Function	0.0022	0.0510	0.0276	0.891
CNN-Transformer model (Vanilla)	0.0027	0.0520	0.0288	0.873

Compared to the CNN-Transformer model without any optimizations, the optimized model improves R^2 by 10.88%. Therefore, tailoring the model to geological conditions effectively enhances detection accuracy.

5 Conclusion

In the study, addressing issues such as significant errors in traditional numerical simulation methods for detecting formation pore pressure in oil wells and the complexity of multi-physical parameter modeling, convolutional neural networks (CNN) and Transformer models were established. A physically constrained loss function was proposed and successfully integrated, thereby constructing a higher-precision formation pore pressure detection model. By innovating the detection algorithm architecture and optimizing the data governance system, the detection accuracy and engineering applicability were synergistically improved. The main research achievements include:

(1) A Transformer-based fusion model for formation pore pressure detection was constructed. By continuously optimizing the structure with hyperparameters, the model not only captures local features of parameters, such as acoustic time difference and formation density, to a certain extent, but also handles long-range dependencies across the entire well section.

(2) Building on the Transformer-based formation pore pressure detection model, CNN was introduced as a feature extractor to capture spatial features. An optimized structure employing 3×1 dilated convolution kernels, 30×1 large-kernel convolutions, downsampled to 3 m resolution, was used to process sequence relationships, further enhancing detection accuracy.

(3) To adapt to a wider range of geological structures and different types of logging data, geological principles were innovatively integrated into the neural network model architecture. Experimental results showed that the *MSE* decreased to 0.0020, the *MAE* to 0.044, and the R^2 improved to 0.968, achieving lower detection errors. Visualization comparisons between detected and actual values demonstrated the model's high accuracy even at abrupt changes in formation pore pressure. Finally, comparative tests and ablation experiments verified the effectiveness of the three optimized components of the CNN-Transformer fusion model: the optimization of the convolutional neural network reduced model error by 9.5%, the physically constrained Transformer architecture reduced error by 15.3%, and the loss function optimization reduced error by 2.5%.

Due to the complexity of formation pore pressure detection tasks, current research still has shortcomings and areas requiring further improvement:

(1) The number of feature parameters used is relatively limited. Without compromising the accuracy of formation pore pressure detection, additional types of feature parameters, such as 3D seismic data and mud logging data, could be incorporated.

(2) The proposed physics-driven CNN-Transformer fusion model for formation pore pressure detection has improved detection accuracy to some extent. However, the integration between the method and physical

knowledge remains insufficient. Future work could explore more effective ways to combine geological knowledge with neural network architectures.

Acknowledgement: The authors gratefully acknowledge the financial support of the key research project of Science and Technology Department of Xinjiang Uygur Autonomous Region, China [Grant Number: 2024B01014].

Funding Statement: The key research project of Science and Technology Department of Xinjiang Uygur Autonomous Region, China [Grant Number: 2024B01014].

Author Contributions: Conceptualization: Xinniu Xu, Jinde Li, Hu Yang; Methodology: Hong Huang, Guangfu Cao, Yuhe Shi; Formal analysis: Yuhe Shi, Biao Ruan; Investigation: Hong Huang, Daojin Ge; Writing—Original Draft: Xinniu Xu, Jinde Li, Yuhe Shi; Writing—Review & Editing: Xinniu Xu, Jinde Li, Hong Huang, Guangfu Cao, Yuhe Shi, Biao Ruan, Daojin Ge, Hu Yang; Funding acquisition: Hu Yang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Huang H, Li J, Yang H, Wang B, Gao R, Luo M, et al. Research on prediction methods of formation pore pressure based on machine learning. *Energy Sci Eng.* 2022;10(6):1886–901. doi:10.1002/ese3.1112.
2. McPherson BJOL. Hydrodynamics and overpressure mechanisms in the Sacramento Basin. *California Am J Sci.* 1999;299(6):429–66. doi:10.2475/ajs.299.6.429.
3. Muggeridge A, Mamode H. Hydrodynamic aquifer or reservoir compartmentalization? *Bulletin.* 2012;96(2):315–36. doi:10.1306/06141110169.
4. Patel H, Prajapati D, Mahida D, Shah M. Transforming petroleum downstream sector through big data: a holistic review. *J Petrol Explor Prod Technol.* 2020;10(6):2601–11. doi:10.1007/s13202-020-00889-2.
5. Salem AM, Yakoot MS, Mahmoud O. Addressing diverse petroleum industry problems using machine learning techniques: literary methodology-spotlight on predicting well integrity failures. *ACS Omega.* 2022;7(3):2504–19. doi:10.1021/acsomega.1c05658.
6. He K, Jia X, Cheng Z, Wang X. TBM construction data mining and cloud computing integrated management platform based on complex geological environment. *J Saf Environ Eng.* 2023;30(4):10,46–55. (In Chinese). doi:10.13578/j.cnki.issn.1671-1556.20220367.
7. Mishra P, Biancolillo A, Roger JM, Marini F, Rutledge DN. New data preprocessing trends based on ensemble of multiple preprocessing techniques. *Trac Trends Anal Chem.* 2020;132(3):116045. doi:10.1016/j.trac.2020.116045.
8. Farsi M, Mohamadian N, Ghorbani H, Wood DA, Davoodi S, Moghadasi J, et al. Predicting formation pore-pressure from well-log data with hybrid machine-learning optimization algorithms. *Nat Resour Res.* 2021;30(5):3455–81. doi:10.1007/s11053-021-09852-2.
9. Li H, Tan Q, Deng J, Dong B, Li B, Guo J, et al. A comprehensive prediction method for pore pressure in abnormally high-pressure blocks based on machine learning. *Processes.* 2023;11(9):2603. doi:10.3390/pr11092603.
10. Abdulmalek AS, Elkatatny S, Abdulaheem A, Mahmoud M, Ali Abdulwahab Z, Mohamed IM. Pore pressure prediction while drilling using fuzzy logic. In: *Proceedings of the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition; 2018 Apr 23–26; Dammam, Saudi Arabia.* doi:10.2118/192318-ms.
11. Xu D, Yang J, Zhao Y, Fan J, Li Y, Liu X, et al. Research on pore pressure prediction technology of HTHP wells in South China Sea based on machine learning. In: *Proceedings of the Gas & Oil Technology Showcase and Conference; 2023 Mar 13–15; Dubai, United Arab Emirates.* 1 p. doi:10.2118/214059-ms.

12. Rashidi M, Asadi A. An artificial intelligence approach in estimation of formation pore pressure by critical drilling data. In: Proceedings of the 52nd U.S. Rock Mechanics/Geomechanics Symposium; 2018 Jun 17–20; Seattle, WA, USA.
13. Rahmanifard H, Plaksina T. Application of artificial intelligence techniques in the petroleum industry: a review. *Artif Intell Rev.* 2019;52(4):2295–318. doi:10.1007/s10462-018-9612-8.
14. Kuang L, Liu H, Ren Y, Luo K, Shi M, Su J, et al. Application and development trend of artificial intelligence in petroleum exploration and development. *Petrol Explor Dev.* 2021;48(1):1–14. doi:10.1016/S1876-3804(21)60001-0.
15. Otchere DA, Arbi Ganat TO, Gholami R, Ridha S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: comparative analysis of ANN and SVM models. *J Petrol Sci Eng.* 2021;200(10):108182. doi:10.1016/j.petrol.2020.108182.
16. Sun Y, Pang S, Zhang J, Zhang Y. Porosity prediction through well logging data: a combined approach of convolutional neural network and transformer model (CNN-transformer). *Phys Fluids.* 2024;36(2):026604. doi:10.1063/5.0190078.
17. Meng Y, Jiang J, Wu J, Wang D. Transformer-based deep learning models for predicting permeability of porous media. *Adv Water Resour.* 2023;179(1):104520. doi:10.1016/j.advwatres.2023.104520.
18. Cheng X, Zhang H. Forecasting formation density from well logging data based on machine learning model. *Front Earth Sci.* 2025;13:1530234. doi:10.3389/feart.2025.1530234.
19. Zhang Y, Fan C, Qin D, Wang X, Jiang B. Deep-learning-based prediction of mutant formation pore pressure: a case study from the Xihu Sag in the East China Sea. *Gas Sci Eng.* 2026;147(2):205846. doi:10.1016/j.jgsce.2026.205846.
20. Zhang X, Lu YH, Jin Y, Chen M, Zhou B. An adaptive physics-informed deep learning method for pore pressure prediction using seismic data. *Petrol Sci.* 2024;21(2):885–902. doi:10.1016/j.petsci.2023.11.006.
21. Li C, Yu Q, Sun B, Li H, Li Y. Analysis of low-temperature thermochemical heat transformer system based on hydrated salt: dynamic modelling and performance evaluation. *J Energy Storage.* 2025;117:116166. doi:10.1016/j.est.2025.116166.
22. LeCun Y, Boser B, Denker JS. Handwritten digit recognition with a back-propagation network. *Adv Neural Inf Process Syst.* 1989;2:396–404.
23. Fu P, Chu L, Li J, Guo Z, Hu J, Hou Z. State of health prediction of lithium-ion battery based on deep dilated convolution. *Sensors.* 2022;22(23):9435. doi:10.3390/s22239435.
24. Wang X, Wang K, Liu X, Hao X. Expression recognition based on adaptive reweighting pooling deep multi task learning. *Comput Eng Des.* 2024;43(4):1111–20. (In Chinese). doi:10.16208/j.issn1000-7024.2022.04.029.
25. Ferdinand R, Malanchini M, Rinfeld K. Mathematics interest, self-efficacy, and anxiety predict STEM career choice in emerging adulthood. *npj Sci Learn.* 2024;9(1):66. doi:10.1038/s41539-024-00275-1.