258 (2026) 214326

Contents lists available at ScienceDirect

Geoenergy Science and Engineering

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

The application of generative adversarial networks (GANs) for petroleum exploration in deep strata—A case study on overpressure prediction

Qiaochu Wang [a,b], Dongxia Chen [a,b,*], Meijun Li [a,b], Fuwei Wang [a,b], Zijie Yang [a,b], Zaiquan Yang [a,b], Sha Li [a,b], Yuqi Wang [a,b]

[a] State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
[b] College of Geosciences, China University of Petroleum (Beijing), Beijing, 102249, China

ARTICLE INFO

Keywords:
Petroleum and natural gas
Deep learning
Generative adversarial networks
Deep strata
Overpressure prediction

ABSTRACT

Petroleum resources are among the most important fossil fuels in the world. With the development of field knowledge and exploration technology, the giant potential of petroleum resources in deep strata deeper than 5000 m has been recognized. However, the complex geological environment characterized by extremely high pressure seriously constrains deep-strata petroleum resource exploration and exploitation in terms of technology, engineering and economics. In this study, the generative adversarial network (GAN) deep learning algorithm is first applied to construct a virtual exploration well dataset to compensate for the lack of wells in deep strata and then to enlarge the well log dataset. Then, a boosted back propagation neural network (BPNN) is used to predict the complex pore pressure based on the enlarged GAN dataset. Finally, the gradient descent method is utilized for the calibration of the whole model to provide a precise deep-strata pore pressure prediction model. The final results show that with the enlarged GAN dataset, the pore pressure prediction model performs well during training and validation, with the highest correlation coefficients, 0.821 and 0.755, respectively. The test results without loss function calibration show a slightly lower accuracy, with correlation coefficients ranging from 0.584 to 0.606. The cross-validation also indicates similar accuracy, with correlation coefficients ranging from 0.589 to 0.606. By applying the data-driven and physical informed loss functions, both the GAN model and BPNN model are optimized. The loss of the GAN model shows a steady status after the 2000th calibration, which indicates that the best model is constructed. The final correlation coefficient between the actual and predicted pressures also reaches 0.98 as the loss of the BPNN model decreases from 4.0 to 1.1. This study not only introduces a novel method for pore pressure prediction for deeply buried petroleum reservoirs by using the deep learning algorithms of GANs and BPNN, which is the first use in the petroleum industry field, but also provides significant guidance for the utilization of artificial intelligence in petroleum exploration and exploitation, thus further extending the application of AI to small-dataset situations.

1. Introduction

As a nonrenewable fossil energy source, petroleum and natural gas are highly important for the development of industry and people's lives. With the end of the COVID-19 pandemic and the continuation of the Russia–Ukraine conflict, the demand for petroleum and natural gas has increased rapidly, revealing the irreplaceable nature of petroleum at present (EIA, 2022). However, after more than 100 years of exploration and exploitation, conventional petroleum and natural gas resources in shallowly buried strata have decreased, especially in the last 10 years (EIA, 2021). Moreover, deeply buried unconventional resources, including shale oil, shale gas, tight oil and tight gas, have played increasingly important roles in worldwide petroleum consumption (Song et al., 2017; Merriam et al., 2018; Zou et al., 2022). Along with improvements in petroleum exploration technology and the recognition of the great potential of petroleum resources in deeply buried strata, deeply buried strata has attracted increasing attention from the petroleum industry and is becoming one of the most important parts of the whole petroleum system (Zhang et al., 2021; Snee and Zoback, 2022; Bashkova et al., 2022).

The petroleum resources in deep strata greatly differ from those of shallowly buried conventional petroleum resources. Owing to multiple

* Corresponding author. State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China.
E-mail addresses: Wangqc@cup.edu.cn (Q. Wang), lindachen@cup.edu.cn (D. Chen).

https://doi.org/10.1016/j.geoen.2025.214326
Received 3 March 2025; Received in revised form 1 November 2025; Accepted 7 December 2025
Available online 12 December 2025

periods of tectonic movement and the depositional processes of deep strata, deeply buried oil and gas reservoirs are characterized by extremely high temperatures, high pressures, and even intensive overpressure (Fig. 1) (Sun et al., 2019; Zhu et al., 2020; Jia et al., 2023). Petroleum exploration under such special conditions faces two main problems. First, exploitation strategies and drilling processes in the high temperatures and high pressures area are much more difficult owing to the complex temperature environment and the potential risk of well blowouts, gas explosions and hydrogen sulfide leakage, which are always caused by overpressure (Lim et al., 2021; Snee and Zoback, 2022). Second, the data from deep strata are insufficient because of technical difficulties and large expenses in drilling engineering and well logging (Van Balen et al., 2022; Xu et al., 2024). Petroleum exploration and exploitation in deep strata require precise evaluation and prediction of geological conditions, especially for the development of overpressure, which is highly important for the whole petroleum industry, including determining favourable zones for petroleum exploration and exploitation and avoiding risks during drilling and petroleum production evaluation and prediction (Skvortsov, 2020).

Overpressure in petroliferous basins was first discovered in the 1960s. With the development of the petroleum industry, overpressure prediction has become recognized as an important component of drill planning and geological analyses. The traditional pore pressure evaluation and prediction methods are mostly based on well log data. Hottman and Johnson first applied acoustic travel time well log data to calculate the pore pressure in upper Texas (Hottman and Johnson, 1965). In the 1970s, Eaton introduced Eaton's method for pore pressure prediction in the Gulf Coast Basin based on resistivity well log data (Eaton, 1975). In the 1990s, Bowers proposed his improved pore pressure prediction method by calculating the effective stress from sonic interval velocity well log data (Bowers, 1995). Subsequently, many improved methods based on different kinds of well log data were introduced for different petroliferous basins in the 2000s (Swarbrick and Osborne, 1998; Tingay et al., 2009). However, for deep strata, the pore pressure distribution becomes complex, and the present pore pressure prediction methods based on one or two kinds of well log data with a single mechanism exhibit unstable performance (Liu et al., 2021b; Wei et al., 2023). Furthermore, the present methods require a large amount of tested pore pressure data for calibration, which is difficult because of the lack of exploration wells in deep strata. In fact, the lack of data is always a key problem in petroleum exploration, exploitation, drilling and production of deep strata.

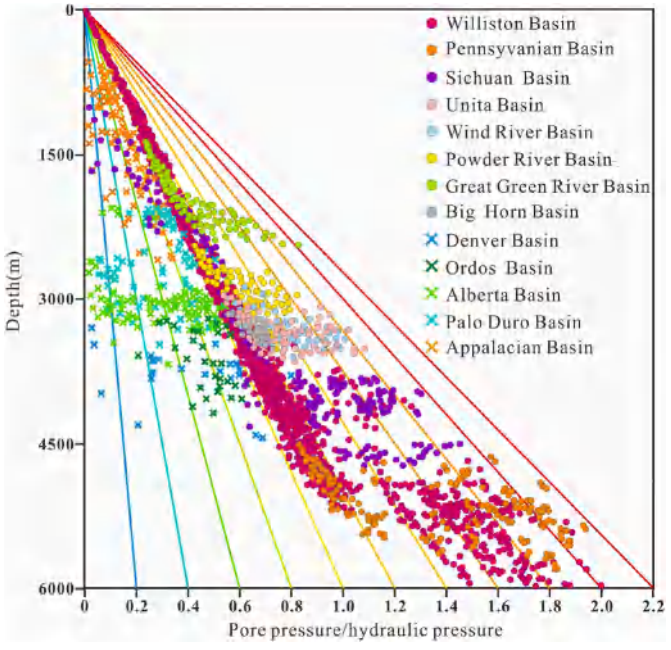


Fig. 1. Tested pressure of the main petroleum and natural gas reservoirs around the world.

Machine learning (ML) methods have been widely used in education, transportation, manufacturing, medicine and other fields (Zamo et al., 2014; Wang et al., 2017; Gafni et al., 2024) along with the information technology revolution. ML methods are characterized by priority in data management because they can extract the relationships among multiple dimensions of data and therefore show good performance with complex nonlinear problems (Pradhan, 2013). With the integration of theoretical knowledge, methods based on ML can both objectively reduce human subjectivity and rationally fit the cognition of theoretical laws (Hastie et al., 2008; Kreimeyer et al., 2021; Dorotic et al., 2024).

In recent years, ML methods have been applied in the 1D petroleum industry for geological feature extraction and classification, petroleum production evaluation and prediction, 2D seismic interpretation and 3D sedimentary model construction. The geological feature classification by ML methods is mainly based on the well log dataset. Multiple well log data with numerous wells can provide sufficient data for ML algorithms to construct classification models for lithologies and geofluids with high prediction accuracy (Imamverdiyev and Sukhostat, 2019; Alzubaidi et al., 2021; Shehata et al., 2021). Petroleum production prediction by ML methods has been widely used in many petroliferous districts. Due to long-term production processes, which can often last for several years, the datasets available for petroleum production evaluation and prediction are always large and therefore suitable for model construction via ML (Shi et al., 2021; Alharbi and Alarifi, 2023; Kim et al., 2023). Advanced algorithms such as gradient boosting decision tree (GBDT), Xgboost and support vector machine (SVM) models can always provide good performance with high prediction accuracy (Liu et al., 2021a; Rahmanifard et al., 2020). With the development of artificial neural networks (ANNs), seismic interpretation via ML methods has become popular. Seismic message identification and seismic image interpretation by back propagation neural networks (BPNNs) and convolutional neural networks (CNNs) have shown high efficiency and accuracy in the North Sea Basin and Gulf of Mexico (Colombera and Mountney, 2020 Chaki et al., 2022; Mousavi and Gregory, 2023; Chitkeshwar, 2024). Furthermore, with the powerful capability of generative adversarial networks (GANs), 3D palaeosedimentary models can be reconstructed to show the status of basins millions and even billions of years ago (Creswell et al., 2018; Saxena and Cao, 2021; Puzyrev et al., 2022; Song et al., 2022a,b). Although ML methods have been applied in many fields of the petroleum industry, the utilization of ML in deep-strata petroleum exploration is rare. Some researchers have found that the performance of ML regression models (such as petroleum production prediction) is worse in deep strata than in shallower strata because in the background of the more complex geological conditions in deep strata, the calibration of regression models requires more tested data, which is difficult (Cheng et al., 2022). Moreover, constrained by the lack of exploration wells and hence the small data size, classification models (such as geological feature classification) based on ML methods have also shown unstable performance in deep strata (Ma et al., 2024).

To conclude, the dominant problem for the application of ML methods in the petroleum industry in deep strata is the conflict between the requirement of a much larger data size and the actual situation of the data provided. Considering the requirements of petroleum exploration, exploitation, drilling and production, more well log and test data are required for precise evaluation and prediction of complex geological conditions in deep strata. However, constrained by the technology, equipment and extremely high costs in the deep-strata petroleum industry, the sizes of well log and test data are usually insufficient.

To solve this problem, this study used overpressure evaluation and prediction as an example to provide an intelligent method for the application of ML algorithms in the deep-strata petroleum industry. The study area was the Jurassic formation (burial depth of 5300~6500 m) in the Central Depression of the Junngar Basin in western China, which is

characterized by widely developed overpressure with complex lateral and vertical distributions (Lu et al., 2022; Gao et al., 2023). To address the dominant problem, this study used a method including two ANN models to precisely predict the overpressure. First, the GANs were applied to generate a virtual well log dataset constrained by real well log data to increase the size of the deep strata data. Second, the enlarged virtual well log dataset was used for constructing a pore pressure prediction model via boosted BPNN. This study is the first to introduce an ML method system that can be applied in the deep-strata petroleum industry with limited data. The method system can also be used to obtain other important information on petroleum exploration and production. Furthermore, the method provides a practical way for the utilization of ML algorithms in a data-limited situation is also significant in practice.

2. Geological setting

The Junggar Basin is one of the largest petroliferous basins in western China, with an area of approximately 1.36×10^5 km^2 (Fig. 2a). The whole basin is divided into 6 tectonic belts with an overthrust belt, 3 uplifts and 2 depressions, named the Wulugu Depression and Central Depression, which are the study areas of this research (Fig. 2a). The Central Depression of the Junggar Basin is characterized by deeply buried hydrocarbon accumulation systems in which almost all the discovered petroleum and gas resources are from Jurassic, Triassic and even Permian strata with burial depths ranging from approximately 4500 m–6000 m (Fig. 2b) (Zhi et al., 2022; Zhang et al., 2024a). The deeply buried strata have shown great resource potential with more than 30 oil reservoirs, which are characterized by high production from 10.04 to 39.3 tons/day and widely developed and intensive overpressure with formation pressures from approximately 75 MPa–102 MPa, which are approximately 1.7–2.2 times the hydraulic pressure (Fig. 2c) (Wang et al., 2023).

There are 24 formations in the Central Depression of the Junggar Basin with total burial depths greater than 10,000 m. The dominant source rocks consist of dark mudstones in the Permian Fengcheng Formation (P_1f Fm) at depths of 8000–9000 m and the Lower and Upper Uerhe Formations (P_2w and P_3w Fms) at depths of 7100–7600 m (Tang et al., 2021). The main hydrocarbon reservoirs are the tight sandstones in the Jurassic Xishanyao Formation (J_2x Fm) and Jurassic Sangonghe Fm (J_2s Fm). The deeply buried oil and gas reservoirs with burial depths of 4700–6500 m are characterized by large disparities in physical properties due to long-term compaction processes, multiple tectonic movements and several sedimentary events. The main caprocks are the low-permeability mudstones in the Eogene and Neozoic Formations.

The Junngar Basin has experienced 5 main tectonic movements from the Carboniferous to the present. The first stage was a rapid depositional process controlled by normal faults from the Carboniferous to the Permian. The second stage was a continuous subsidence process from the Permian to the Middle Jurassic. Afterwards, the third stage was characterized by large-scale tectonic uplift and erosion of deposits during the Middle to Late Jurassic. The fourth stage had more subsidence involving hydrocarbon charging and migration, which led to complex geofluid interactions and activity. Furthermore, overpressure in deeply buried Jurassic sandstone reservoirs was generated by these geological processes (Lu et al., 2022; Zhang et al., 2023b, 2024b). The last stage was characterized by slow subsidence (Fig. 3). To conclude, the tectonic movements in the study area were complex, which led to multistage transport and interactions of geofluids and therefore the complexity and particularity of the formation pressure distribution in the deep strata (Liu et al., 2019; Yang et al., 2023).

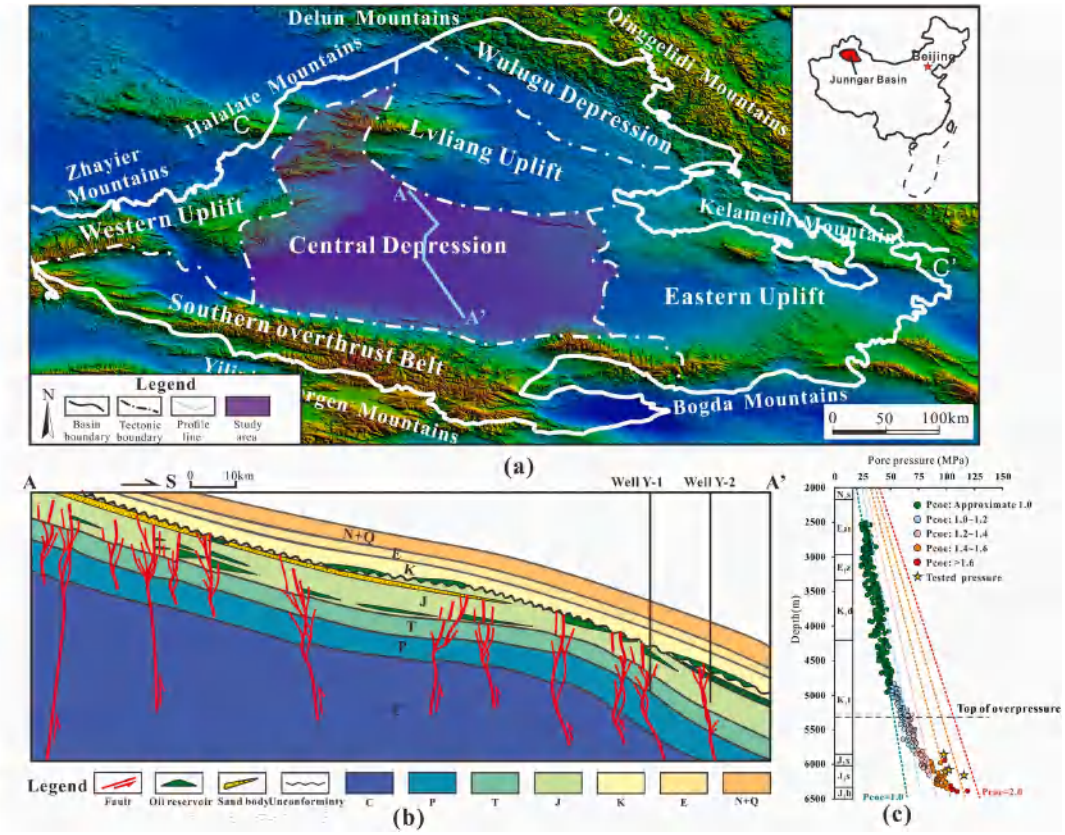


Fig. 2. Location, tectonic belts, hydrocarbon reservoir distribution and formation pressure characteristics in the Central Depression of the Junggar Basin, western China. (a) Tectonic belts of the Junggar Basin. (b) Cross-section of the Junngar Basin showing the distribution of hydrocarbon reservoirs. (c) Formation pressure characteristics of the deep strata (Pcoe: Pressure coefficient, which is the ratio between pore pressure and hydraulic pressure at the same depth).

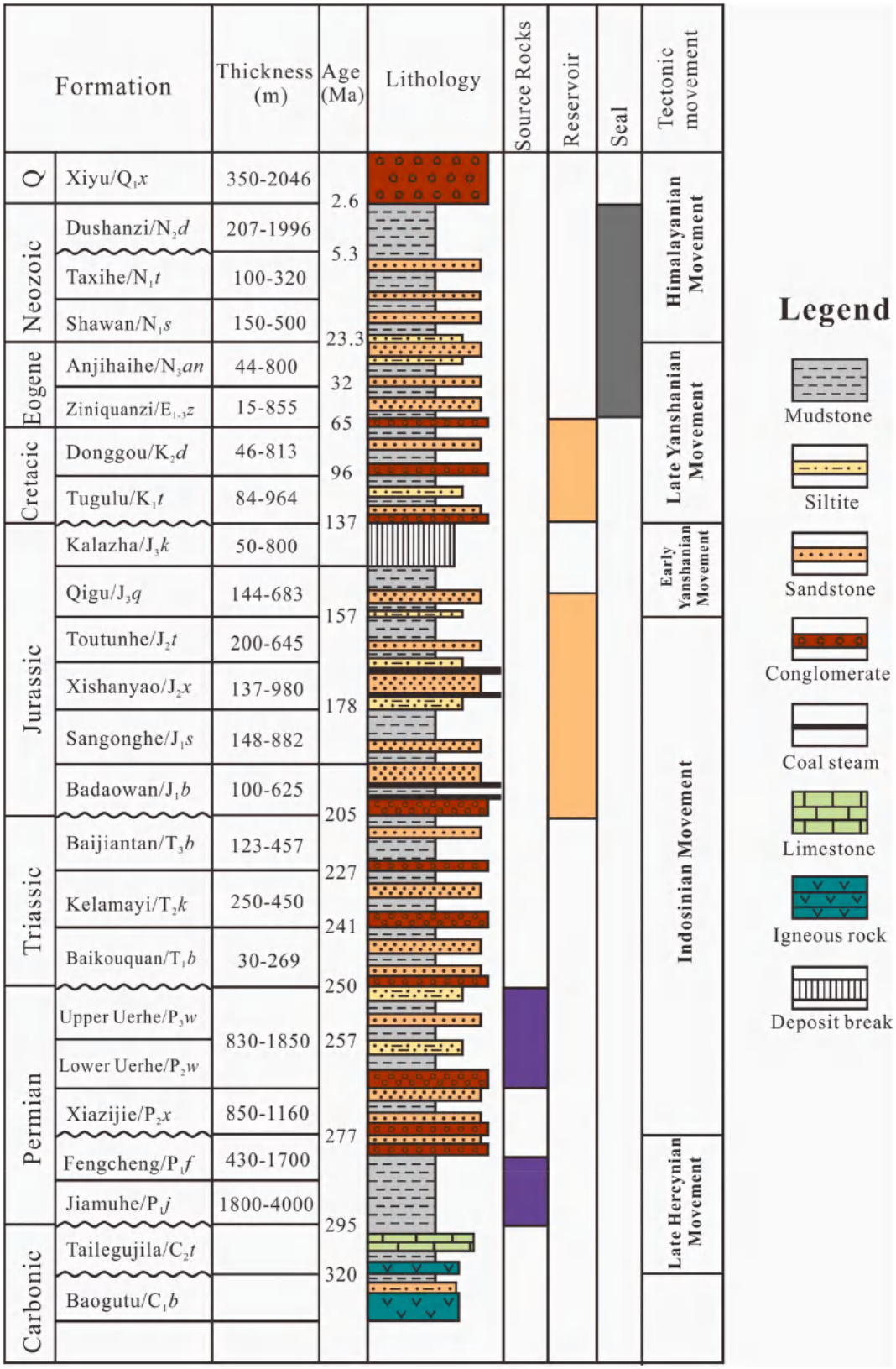


Fig. 3. Generalized stratigraphic column for the Junggar Basin, western China (Edited by Wang et al., 2023).

3. Methods and data

There are 4 steps for the construction of the deep-strata pore pressure prediction model. First, the data processing process is required for the original seismic and well log data collected in different data types, colour types, sizes, and resolutions. Second, the processed seismic data are input into the generator of the GANs to construct random well log data. In addition, the processed well log data are input into the

discriminator of the GANs to distinguish the fake image of the well log data generated by the generator. When this process is finished, the virtual well log data generation process is complete, and the well log curves generated by the GANs are considered real well log data in the deep strata. Third, the generated virtual well log data are used as the original data to construct the deep-strata pore pressure prediction model via the BPNN. To enhance the performance of pore pressure prediction, the ensemble learning boosting method is applied to make precise pore pressure predictions based on multiple BPNN models. Finally, the whole prediction model is calibrated by the tested pore pressure in the deep strata. The calibration process is based on loss function construction and random vector searching methods, which change the imputing random vector of the GAN model and hence regenerate the virtual well log data. Therefore, the prediction results of the pore pressure will change and gradually approach the tested pressure until the error is low enough to meet the prediction requirements (Fig. 4).

3.1. Data processing

The principal data used in the pore pressure prediction model are seismic data and real well log data of the deep strata. For the utilization of the GANs, we applied 2D vector images of the seismic section and well log curves, which are of totally different sizes, types, resolutions and even colour styles, which all have a significant influence on the performance of the GANs and hence the BPNN models (De Souza et al., 2023). Therefore, data processing involving the unification of the original data while considering multiple factors is needed.

3.1.1. Dataset of the GANs for virtual well construction

3.1.1.1. Seismic dataset for the generator. The seismic data are prepared for the GANs to guide the generator to generate random well log curve images that aim to replace the real well log curve. Therefore, the type of seismic data is a 2D seismic attribute profile image in the depth domain. To enhance the performance of the GAN generator, the amount and type of seismic images need to be considered. First, the through-well seismic section can be obtained by slice visualization from the 3D seismic data volume, which means that the number of seismic images should be sufficiently large to fulfil the requirements of model construction. Note that we collected seismic data from multiple basins with deeply buried strata, including the Tarim Basin, Sichuan Basin and Bohai Bay Basin, to increase the size of the dataset. Second, we tended to choose the seismic inversion profile because it is related to the well log data and therefore can provide better guidance for the generator (Fig. 5). All the seismic

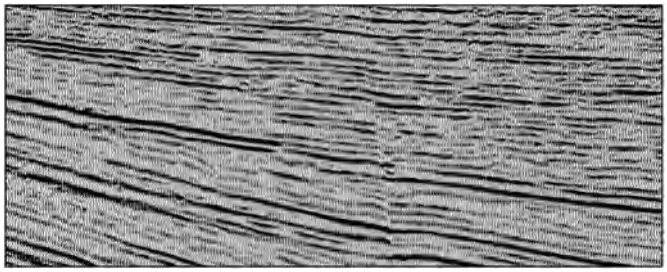


Fig. 5. 2D prestack seismic inversion profile grayscale image for the GAN generator.

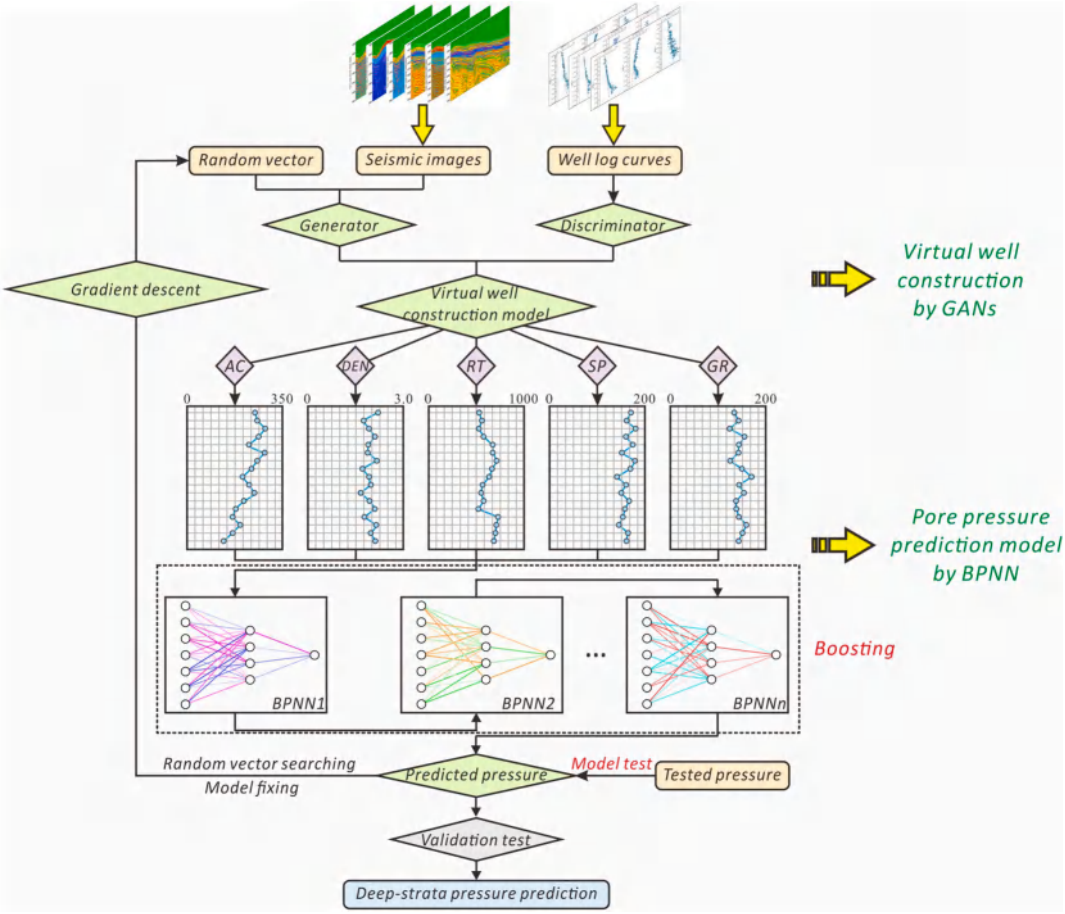


Fig. 4. Workflow of the pore pressure prediction in the deep strata by the GANs and BPNN.

images are converted to depth domain to match with the corresponding well log data in depth range. Furthermore, among several kinds of seismic inversion profiles, we selected the prestack seismic inversion profile without too much additional interpretation from the human mind to provide objective data for the generator of the GANs. In this research, a total of 3043 2D prestack seismic inversion profile grayscale images were prepared. Furthermore, to maintain data integrity and provide much more possibility in virtual well data construction for complete drilling wells, we have applied the seismic and well log data covering all depth range from shallower to deep strata, although this research focuses on deep strata in application.

3.1.1.2. Well log curve images for the discriminator. The well log data are prepared for the discriminator to distinguish the fake well log curve images generated by the generator. Therefore, the types of well log curve data generated by the GANs depend on the well log data applied for the discriminator. Owing to the final goal of the model, which is to predict the pore pressure, the selected well log curves must be related to the pore pressure in porous media. Previous studies have shown that spontaneous potential logging (SP), gamma-ray logging (GR), true formation resistivity logging (RT), density logging (DEN) and acoustic logging (AC) are the dominant well log methods for pore pressure prediction (Eaton, 1975; Law and Dickinson, 1985; Osborne and Swarbrick, 1997; Bowers, 2002). Therefore, these 5 types of well log data are selected as the targets for providing the identification basis of the discriminator. In this research, a total of 1054 pairs of well log curve images are collected for the GAN discriminator (Fig. 6).

3.1.2. Dataset of the BPNN for pore pressure prediction

The data used for pore pressure prediction in the BPNN model are the virtual well log data generated by the GANs. In this research, a total of 1000 pairs of well log data with 5 types of well log data are used as the dataset for the BPNN model, among which 700 pairs are used as training data and 300 couples are used as test data.

3.2. Feature engineering

To further improve the quality of the data and therefore enhance the performance of both the GAN and BPNN models, feature engineering processes are highly important. Considering that the dataset used in this research is dominated by 2D images, the image profiling processes for the GANs and the image vectoring and data extraction processes for the BPNN are applied.

3.2.1. Image profiling for GANs

(1) Image size unification

The aim of GANs is to generate virtual well log data for deep strata via the guidance of seismic images and the examination of real well log curve images. Therefore, the through-well seismic section and the well log curve images are exported at the same resolution and cut into the same sizes. With unification, the seismic images and the well log curves are in the same depth range (2000–6500m) with the same width, which is convenient for enhancing the ability of the GANs to generate well log curves. In this research, all the images used for GAN model construction are cut to the same size (5 cm wide and 20 cm high) (Fig. 7).

(2) Image gridding

Image gridding is a decisive factor for the final effect of the GANs. With different gridding strategies, the features of the images learned by the generator and distinguished by the discriminator will significantly change (Fig. 8). The different levels of features for seismic sections provide different types of guidance for virtual well log curve image generation. Correspondingly, the different levels of features for well log curve images indicate different messages for distinguishing fake well log curves. Considering the convolutional processes that will be applied in the GANs, the gridding strategy must be fit with the size of the

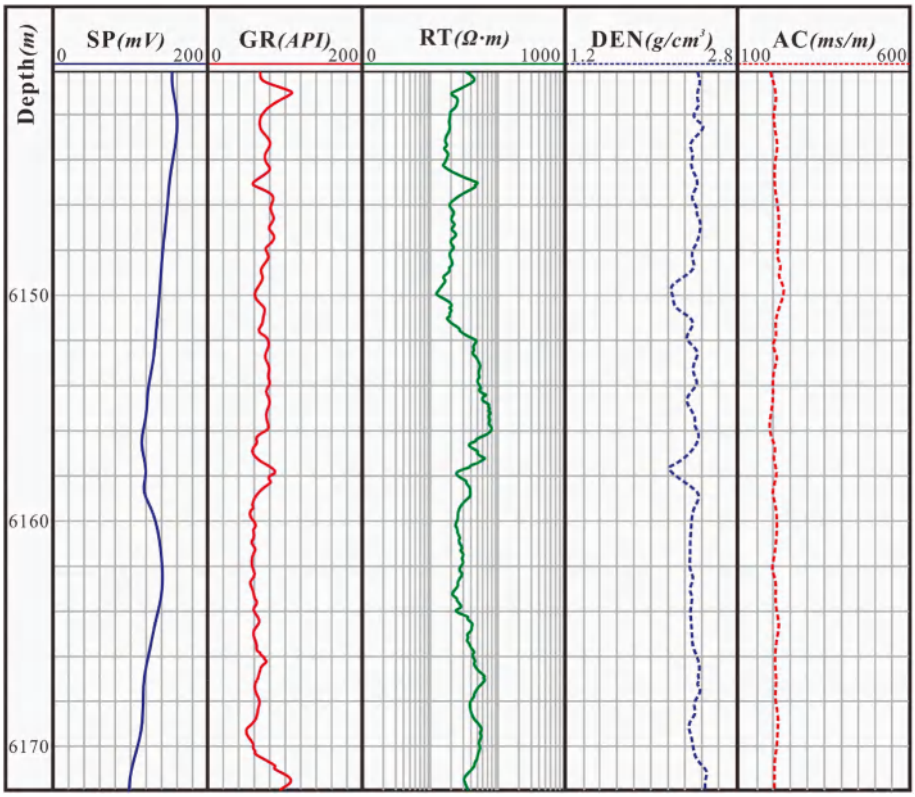


Fig. 6. Well log curves for the GAN discriminator.

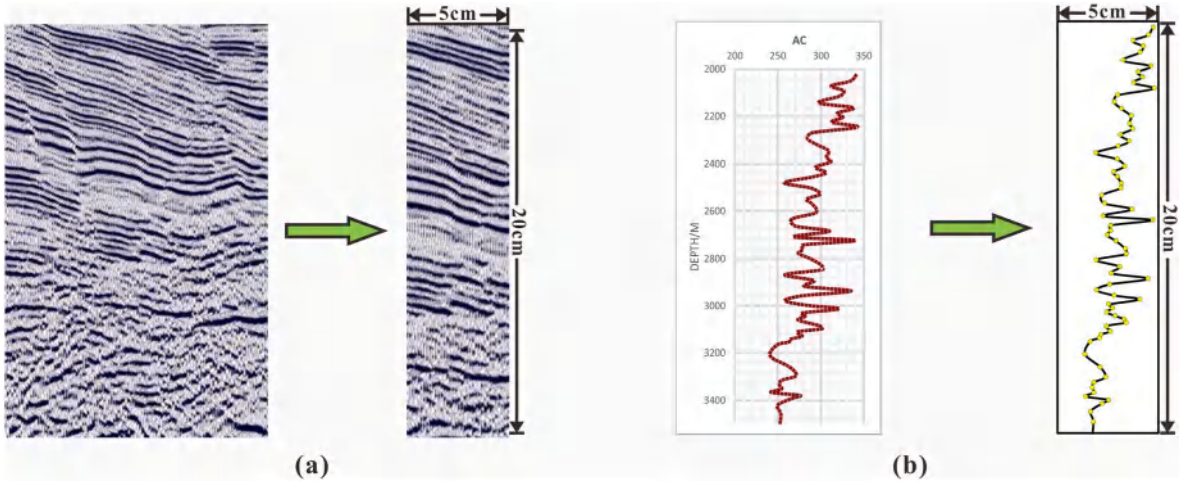


Fig. 7. Image size unification process for (a) the seismic profile image and (b) the well log curve image.

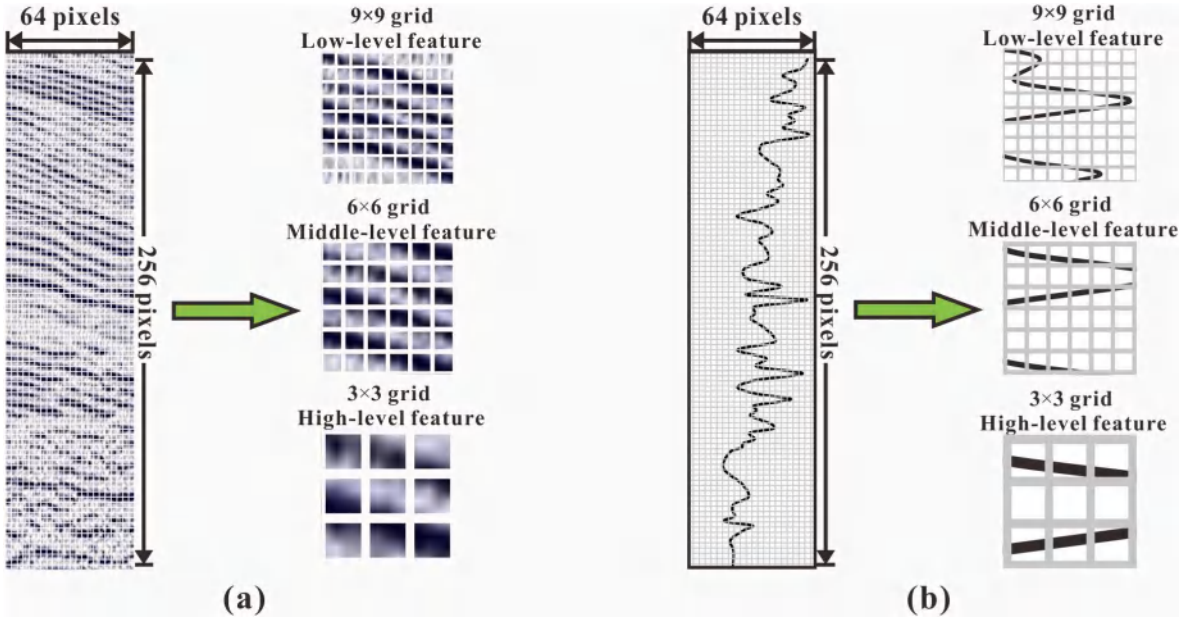


Fig. 8. Image gridding process and different levels of features by gridding for (a) the seismic profile image and (b) the well log curve image.

convolutional core, which is always square (Gene et al., 2024; Eshun et al., 2024; Zhang et al., 2023a). Therefore, the seismic and well log curve images with a size of 5 × 20 cm are gridded into 64 × 256 grids (Fig. 8).

3.2.2. *Image vectoring for BPNN*

The dataset of the BPNN model is the well log data generated by the GAN model. Therefore, the 2D well log curve images must be vectored and transferred to the data message. The former progress of image profiling shows that all the generated 2D well log curve images are of the same size with the same resolution, which is convenient for image vectoring processes. In this research, illustrator software is applied for image vectoring, which shows good performance in vectoring grayscale images (He et al., 2023; Puhachov et al., 2021; Bessmeltsev and Solomon, 2019).

3.3. *Virtual well construction based on GANs*

The generative adversarial network is a framework for estimating generative models via an adversarial process and was first introduced in 2014. This framework constructs a self-supervised learning system for generating images when initially introduced (Goodfellow et al., 2018; Zhang et al., 2019). After rapid development, GANs have shown potential. Deep convolutional GANs (DCGANs) based on convolutional neural networks have shown better performance in image generation. Pix2pix GANs achieve picture translation, while Cycle-GANs are characterized by their ability to transfer figure styles (Yilmaz and Korn, 2024; Vrettos et al., 2024). Moreover, GANs can generate sounds, words and videos as expected or as required with their own ability to learn, which means that it is not necessary to provide label or target data, and GANs can generate items that do not exist in the world (Goodfellow et al., 2020; Aldhubri et al., 2024). Thus, the ability of the GANs fits our requirements to generate virtual well datasets in deep strata.

3.3.1. *Well log curve generation principle and processes*

The basic generative adversarial network is a model in which two neural network models are trained simultaneously. The generative model (abbreviated as G) is applied to capture the data distribution and generate the images, and the discriminative model (abbreviated as D) is constructed for estimating the probability that a sample comes from the

real training samples rather than G (Goodfellow et al., 2018). In this case, the purpose of G is to maximize the probability of D making a mistake in judging the source of the imputing images through the training processes. The well log curve image generation process can be divided into the following steps.

(1) The training data of G are formed:

$$z = (x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)$$

where z is the random noise that obeys the probability distribution of $p_z(z)$. (x_n, y_n) is the n-dimensional vector.

(2) The expectations of G and D are determined:

In a normal GAN model, G is a multilayer perceptron with the parameters θ_g, while the output images of G represent a mapping to the data space as $G(z, \theta_g)$. Similarly, D is also a multilayer perception $D(x, \theta_d)$ with the parameter θ_d. The output of D is $D(x)$, which is a single scalar ranging from 0 to 1 and represents the probability that the input image x comes from the real well log curve instead of the fake images from G.

In the real situation, when the real well log curve image is put into D, the expectation of D is (Goodfellow et al., 2018):

$$E_{x\sim p_{data(x)}} = \log D(x) \tag{1}$$

When the fake well log curve image is put into D, the expectation of D is:

$$E_{z\sim p_{z(x)}} = \log(1 - D(G(z))) \tag{2}$$

(3) The best GAN model is determined by the determination of G and D:

To obtain the best generative model, D is trained to maximize the probability of assigning the true label to both the well log curve images from G and the real well log curves. G is simultaneously trained to minimize $E_{z\sim p_{z(x)}}$, which indicates that we want to maximize $D(G(z))$, which is the probability that the well log curve image generated by G is assigned as true by D. Therefore, D and G play a two-player minimax game with the following function (Goodfellow et al., 2020):

$$\min_G \left| \max_D V(D, G) = E_{x\sim p_{data(x)}} + E_{z\sim p_{z(x)}} \right. \tag{3}$$

For a fixed G, the optimal D is:

$$D^*_G(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \tag{4}$$

In Equation (4), we can conclude that only if $p_{data}(x) = p_g(x)$, $D^*_G(x) = 1/2$, which means that for every input image, D can only reach 50 % accuracy, which means that neither G nor D can improve performance and reach Nash equilibrium (Bhatti and Broadwater, 2020; Timm et al., 2023; Pfeifer et al., 2023). In this case, the best GAN model is determined.

3.3.2. Optimization of GANs

Normal GANs can be optimized in many ways, which enhances the performance and efficiency of the GAN model. The optimization methods of the GAN model applied in this research are as follows.

(1) Generating process with the constraint of seismic data

The G of the normal GANs generates the well log curve images randomly by the random vector. This process is unstable with large uncertainty (Song et al., 2021). However, when the dataset is sufficiently large, G and D can provide a GAN model with demonstrated convergence, and we cannot obtain countless data for the training process (Goodfellow et al., 2020). Therefore, the GANSim framework is applied to guide G to generate the images. GANSim is a GAN framework that sets boundary conditions to help G generate targets with high efficiency with less training time and limited samples based on professional knowledge (Song et al., 2022, 2023). In the actual situation of deep-strata petroleum exploration, relatively high-quality seismic data can be obtained (Ahmad and Tsuji, 2021; Reda et al., 2024). Theoretically, either seismic or well log data reflect the characteristics of the geological body underground, and therefore, they are genetically related (Ulfers et al., 2022). Seismic inversion data are suitable for providing guidance for G because they can be directly compared with well log data by converting conventional interface profiles to rock logging profiles with the properties of rocks, including rock density and wave velocity (Fig. 9) (Ebraheem et al., 2022; Zrelli et al., 2024). In this case, the 2D seismic inversion profiles with rock property distributions are used as the boundary conditions for G to generate well log curve images.

In this study, we use 2D seismic inversion profiles as constraining conditions by transferring the inversion profiles to a map of the probability distribution and then construct a conditional loss function to push G to map the output figure that conforms to the probability distribution of the seismic data, which is called GANSim (Song et al., 2022a,b). In this study, the seismic inversion image with the fixed size is transferred to the probability maps by setting the black line as a probability value of 1, and then the other pixel units are assigned by a decreasing tendency along with the distance to the pixel with the probability of 1. The generated well log curve image is constrained by the seismic data by a dot product process between the seismic and well log image for the corresponding positions. Then a loss function is defined as:

$$Loss = 256 - S(x) \cdot G(x) \tag{5}$$

Where $S(x)$ is the probability distribution of seismic inversion map.

The allowed range of the loss function is 0–100, which means we hope that the generated well log curve is only partly constrained by the seismic inversion results instead of completely coinciding with the seismic inversion results.

Without the guidance of the GANSim framework, G can only generate the images randomly (Fig. 10a), while with the help of the seismic images, the mapping to the data space generated by G will be constrained to fulfil the requirements of the seismic data (Fig. 10b). Therefore, G can generate well log curve images that cannot be distinguished by D faster with fewer training epochs.

CNNs have outstanding abilities in image management, including image recognition, image creation and image transfer (Fan et al., 2023; Lyra et al., 2024; Shafee and Awaad, 2021). Thus, the DCGAN is also applied in this research to replace the normal GAN by using the CNN to replace the multilayer perception of D and G (Sim et al., 2021; Radford et al., 2015; Izumi et al., 2022; Mohammadjafari et al., 2021). The LeNet-5 CNN structure is applied in this research for both G and D (Fig. 11). The G of the GAN is composed by …, the main hyperparameters applied for this research is listed in Table 1.

3.4. Pore pressure prediction model by the BPNN

BPNNs are among the most important neural networks for addressing both classification and regression models via deep learning (Saengrung et al., 2007; Cui et al., 2022). For every output by the neural network, a loss function is constructed to record the difference between the output and the expected real value. Then, the difference is regarded as the error and propagated back to the neural network. The weights and biases of the neural network will be adjusted to minimize the error (The framework of BPNN is in the Appendix) (Xue et al., 2022). However, the BPNN easily overfits without controlling the back propagation processes. Therefore, the ensemble learning of the boosting method is also used to decrease the influence of overfitting by a single BPNN model (Fig. 12). In this research, a total of 10 BPNN models are constructed for the boosting

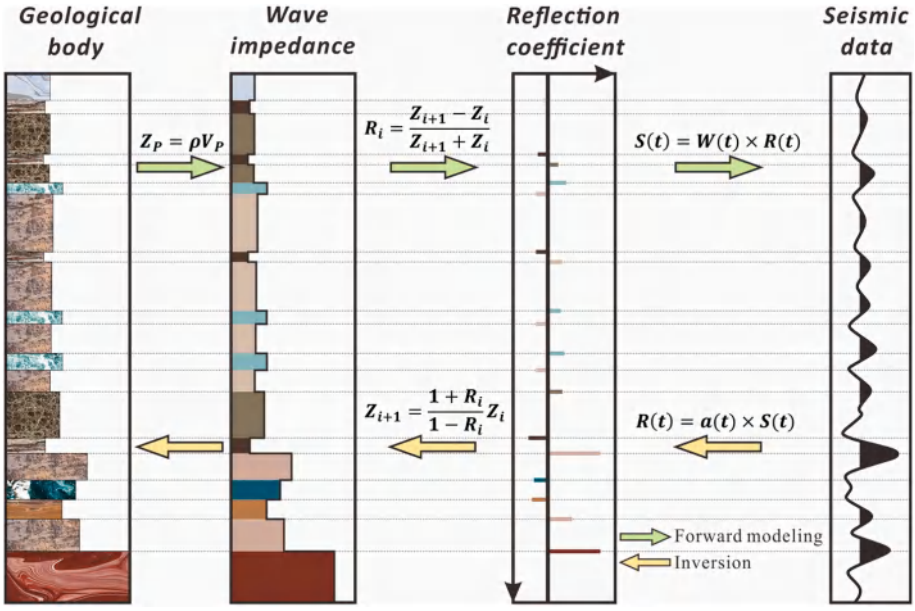


Fig. 9. Seismic inversion processes to reflect the characteristics of the geological body.

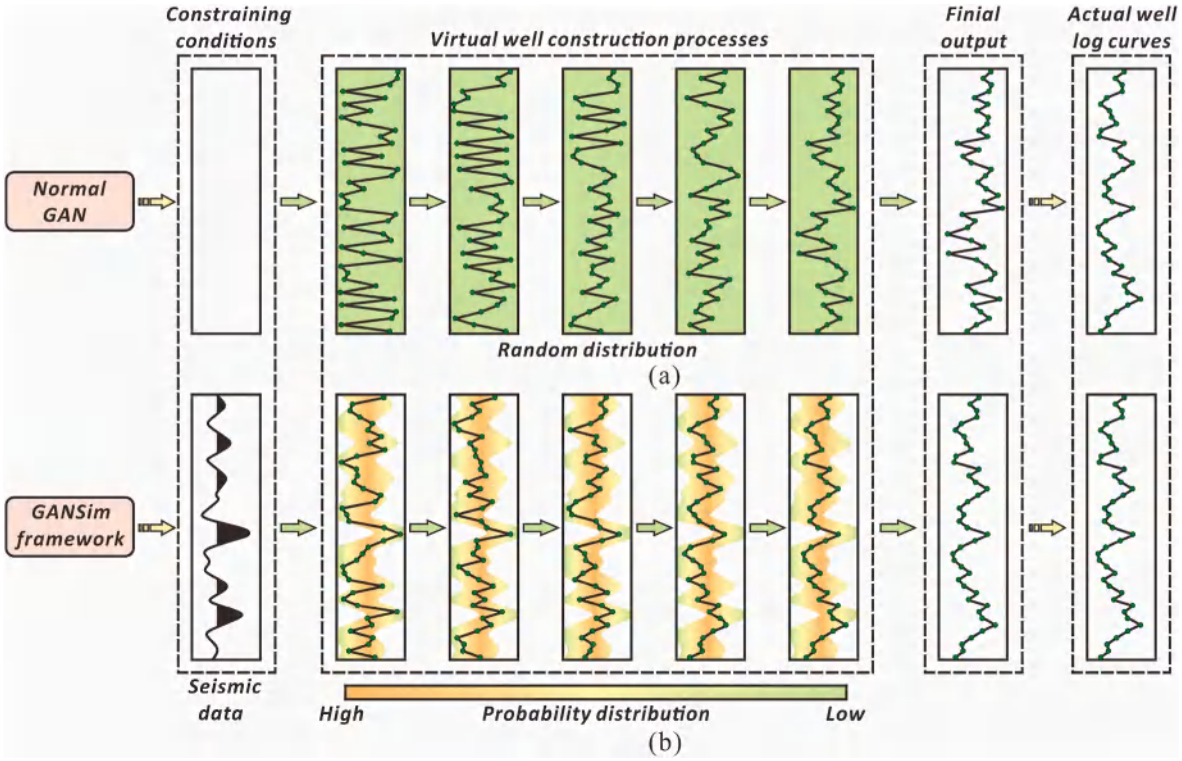


Fig. 10. Comparison of performance between the normal GANs and the GANs with GANSim. (a) Well log curve image generated by normal GANs; and (b) Well log curve image generated by GANSim.

(2) Convolutional neural network used in G and D.

process by AdaBoost framework. The model training procedure is as follows.

(1) Data preprocessing and initial sample weights setting: The data applied for BPNN model training are preprocessed by a standard normalization procedure. Then all the samples are assumed as equally important initially with the same sample weights:

$$w_{1i} = \frac{1}{m}, i = 1, 2, \ldots, m. \quad (6)$$

(2) Boosting Loop: The initial training dataset D with sample weight distribution w_1 is used to train the first BPNN h_1, then a new training dataset via weighted random sampling is constructed for training the second BPNN h_2. During this process, samples with higher weights (large prediction errors in the first training round)

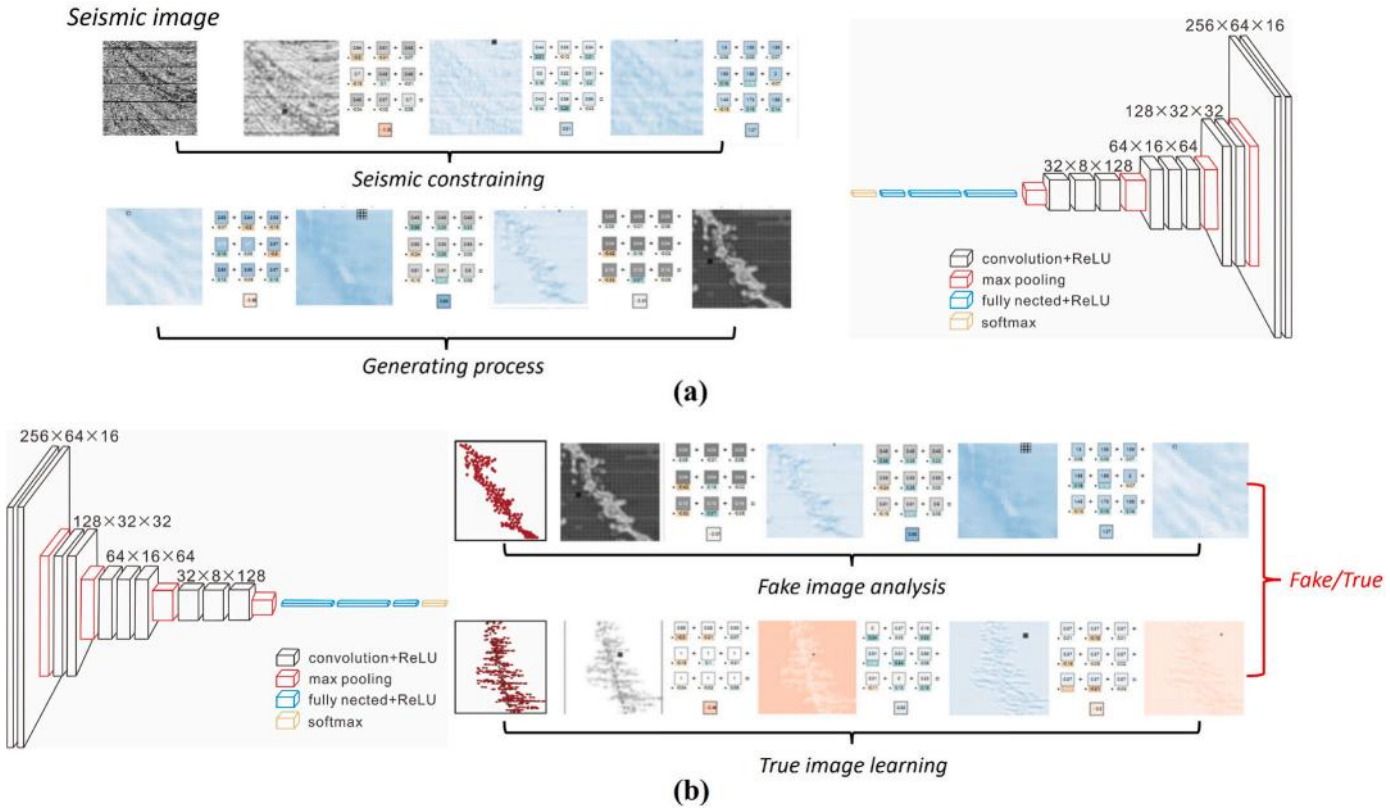


Fig. 11. Virtual well construction model of GANs in this research. (a) G model with GANSim and CNN; and (b) D model with CNN.

Table 1
Architecture of the GAN structure utilized in this study.

	Generator	Discriminator
Input size	256 × 64 × 16	256 × 64 × 16
Output size	256 × 64 × 16	Probability scalar
Convolutional layers	2	2
Max-pooling layers	2	2
Full connected layer	3	3
Softmax layers	1	1

have a higher probability of being selected, forcing the new network h_2 to focus on correcting the precious "large error" samples. Then the model h_2 is used to predict pore pressure on the original training dataset D and obtaining the predictions. The Squared loss L_i for each sample will also be calculated. Then the weighted average loss is calculated:

$$\overline{L_t} = \sum_{i=1}^{m} w_{ti} \cdot L_i \tag{7}$$

This procedure will be iterated by 10 times to obtain 10 different weak BPNN learners.

(3) The calculation of weak learner's weight: The calculation formula is used to assign higher weight to models with lower error:

$$\alpha_t = \ln \frac{1}{\overline{L_t}} \tag{8}$$

(4) Sample weights updating: The goal of the sample weights updating is to increase the weights of samples that were poorly predicted by the current model according to the formula:

$$w_{t+1,i} = w_{t,i} \cdot \beta_t^{L_i} / Z_t \tag{9}$$

Where $\beta_t = e^{\alpha_t}$, Z_t is the loss for sample i; Z_t is the normalization factor ensuring $\sum_{i=1}^{m} w_{t+1,i} = 1$.

With this procedure, the Boosted-BPNN is constructed by 10 single BPNN learners. The final output of the ensemble model is the weight average of all weak learner's predictions:

$$H(x) = \frac{\sum_{t=1}^{T} \alpha_t \cdot h_x(x)}{\sum_{t=1}^{T} \alpha_t} \tag{10}$$

3.5. Model calibration and improvement

In addition to the optimization methods of the GAN and BPNN models, the determination of the hyperparameters is also highly important for model performance. Furthermore, we use limited test data on the pore pressure in the deep strata to calibrate and further improve the prediction model by constructing a loss function.

3.5.1. Hyperparameter determination

The hyperparameters of the GANs are dominated by the functions of the convolutional kernel and the learning processes of G and D, which largely influence the convolutional and deconvolutional processes (Byeon et al., 2024; Rafiei et al., 2024). In this research, the determination of the hyperparameters is based on previous studies (Abdellatif et al., 2022; Huang et al., 2024; Ye et al., 2024). The definitions of the key hyperparameters and their values are listed in Table 2.

The hyperparameters of the BPNN are mainly used to constrain the structure of the neural network and find a balance between training accuracy and training efficiency (Sun et al., 2019; Bai et al., 2023; Chen et al., 2022). In this study, the hidden layer size, activation function, learning rate and maximum number of iterations are set as the dominant hyperparameters (Table 3).

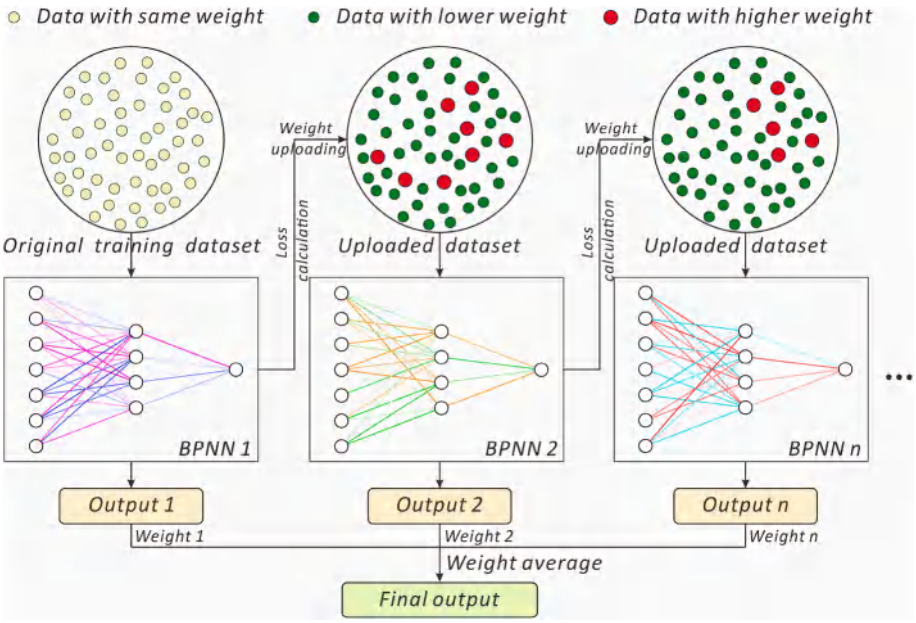


Fig. 12. Ensemble learning of the boosting BPNN.

Table 2
Definition and determination of the key hyperparameters for the GAN model.

Hyperparameter	Definition/Usage	Value
Kernel size	The size of the convolutional core which decides the shape of the output images	3 × 3
Kernel stride	The displacement of the convolutional kernel during the convolution process	1
Kernel padding	The boundary filling process that can adjust the shape of the output images (usually needed for the nonsquare image)	0
Kernel types	The composition of the convolutional kernel that determines the effects of the convolutional process	Max pooled kernel
Learning rate of G	The adjusting frequency of G	0.0001
Learning rate of D	The adjusting frequency of D	0.001

Table 3
Definition and determination of the key hyperparameters for the BPNN model.

Hyperparameter	Definition/Usage
Hidden layer sizes	The amount of the hidden layer
Activation function	Enhance the ability of the model to address nonlinear problems
Learning rate	Determine the variation of the weights during the training process
Maximum of iteration	Determine the maximum of training time to avoid overfitting

3.5.2. Loss function construction and random vector searching

Although testing the pore pressure in the deep strata is difficult, we collected a total of 128 tested pore pressure data points for the calibration of the whole pore pressure prediction model. The calibration is achieved by constructing a loss function to characterize the difference between the pore pressure predicted by the BPNN model and the actual test pressure. In this research, the cross-entropy loss function is applied (Semenov et al., 2019; Sangari and Sethares, 2016; Amaya et al., 2021):

$$Loss_{data} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{q} y_j^{(i)} \log \hat{y}_j^{(i)} \tag{11}$$

where $Loss_{data}$ is the error of the pore pressure prediction model; n is the size of the sample data; q is the number of categories; y_j is the result of the predicted probability distribution for item j, which represents the pore pressure predicted by the model; and $\hat{y}_j$ is the actual probability distribution for item j, which represents the tested pore pressure.

For further enhancing the model training efficiency, the physical-based guidance is applied in this model by the loss function construction. We have added two parts of losses for applying professional knowledge of geophysics and hydrodynamics. The Eaton's method, which is wildly used for pore pressure prediction based on acoustic velocity in porous media, is selected as one of the physical informed conditions to participate in model training by defining Eaton loss:

$$Loss_{Eaton} = \frac{1}{N}\sum_{i=1}^{N}\left|P_p(z_i, V_i) - \left[P_0(z_i) - (P_0(z_i) - P_n(z_i))\left(\frac{V_i}{V_n(z_i)}\right)^x\right]\right|^2 \tag{12}$$

Where P_0 is the overburden pressure; P_n is hydrostatic pressure; V_i is the acoustic velocity obtained from the generated well log curve; V_n is the acoustic velocity under the condition of normal compaction; x is Eaton index, which is 3.0 in this research. According to the definition of the Eaton method, the P_0, P_n and V_n are functions of depth, which can be calculated (Eaton, 1975). The Eaton loss is applied for provide some guidance for pore pressure prediction based on the wildly used geophysical knowledge.

While considering the Eaton's method may be less effective in deep-strata, we have also added another loss function based on Darcy's law to make the prediction result fulfill basic stable fluid flow principle:

$$Loss_{Darcy} = \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{k}{u}\frac{\partial^2 P_p}{\partial z_i^2}\right\|^2 \tag{13}$$

Where k is the permeability of the porous media, u is the velocity of the pore fluids. The target of the Darcy loss is to force the predicted pore pressure fulfill basic stable fluid flow principle, which means the Darcy loss should be 0 in an ideal situation. Therefore we use the residual sum of squares to apply Darcy law to guide the predicted results.

Therefore the final loss function is defined as:

$$Loss = Loss_{data} + \lambda_1 Loss_{Eaton} + \lambda_2 Loss_{Darcy}$$

Where λ_1 and λ_2 are weight of loss, which are determined by adaptive weights algorithm (Raissi et al., 2019; Hou et al., 2024).

The error of the model is decreased by the gradient descent method through changing the random vectors put into G in the GAN model. When the error of the predicted results is sufficiently low, the calibrating process is accomplished, which indicates that the whole model can show good performance for pore pressure prediction in the study area.

4. Results

4.1. Virtual well log curve constructed by GANs

4.1.1. Original well log curve images generated by G

The original well log curve image generated by G has an evolutionary process, which is a random distribution of dots that initially and gradually shows the shape and characteristics of curves. In this research, we used a GPU to train the GAN model, which allows a maximum of over 4000 training epochs each time.

The results show that for the first epoch, the generated images are indeed a mess of dots without any shapes or messages (Fig. 13a). For the 500th training epoch, the images generated by G start to show some changes compared to the initial status, which means that G starts to learn the characteristics of the well log curves in the deep strata (Fig. 13b). In the 1000th training epoch, the images by G show obvious characteristics of curves with different shapes (Fig. 13c). However, in the 2000th epoch, the characteristics of the curve are more obvious in the generated images, and the whole images clearly lack many noisy points (Fig. 13d). Although the final output images after the 2000th training process are also characterized by some disparities from the real ones, they can be used for optimization and to construct a virtual well log dataset.

4.1.2. Construction of the virtual well log dataset

The original 2D well log curve images generated by the GANs cannot be directly used as a dataset; therefore, optimization and vectoring are needed. The unification of the image sizes for both the seismic and actual well log curves provide a way to optimize and vector G images.

4.1.2.1. Optimization of well log curve images. As previously stated, the generated virtual well log curves also have some disparities from the actual curves. In the images generated by the GANs, the curves are not clear as a single line but rather as a band with a certain width (Fig. 14a). This shape of the well log curve may provide more than one value for a certain burial depth during the image vectoring processes, which may lead to confusion in the construction of the virtual well log dataset. In this research, the convolutional method is applied again for the sharpening of the generated well log images to specify the distribution of the generated well log maps (Fig. 14b). After the sharpening process, the well log curve images are optimized for image vectoring.

4.1.2.2. Image vectoring. Image vectoring is the final step for virtual well log dataset construction. The generated and optimized 2D well log curve images have the same sizes as the original seismic images and actual well log images because we consider the convenience of image vectoring and do not change the size of the generated images during the convolutional process by adjusting the hyperparameters of the GANs (Huang et al., 2023; Sapijaszko and Mikhael, 2018). Therefore, the axis of the well log and depth can be determined, and the value of the well log in the continuous depth range can be determined according to the shape of the generated well log curves (Fig. 15).

4.2. Deep-strata pore pressure prediction model using the BPNN

4.2.1. Dataset construction for the BPNN model

The whole dataset of the BPNN model contains the well log data provided by the GAN model and the actual pore pressure data from testing. The well log data are set as feature variables, and the tested pressure is matched with the corresponding seismic data. Note that all of the tested pressure are deep-strata pressure points with the depth from 5500m to 6500m. By random division, 70 % of the data from the dataset are set as the training data, 20 % of the data are set as the validation data, and 10 % of the data are set as the test data.

4.2.2. Training and validating the BPNN model

The model training and validation processes are performed by the Jupyter laboratory on the Anaconda platform. With random division, a total of 2000 ensemble BPNN models are constructed and validated to determine the best hyperparameters and therefore the best pore pressure

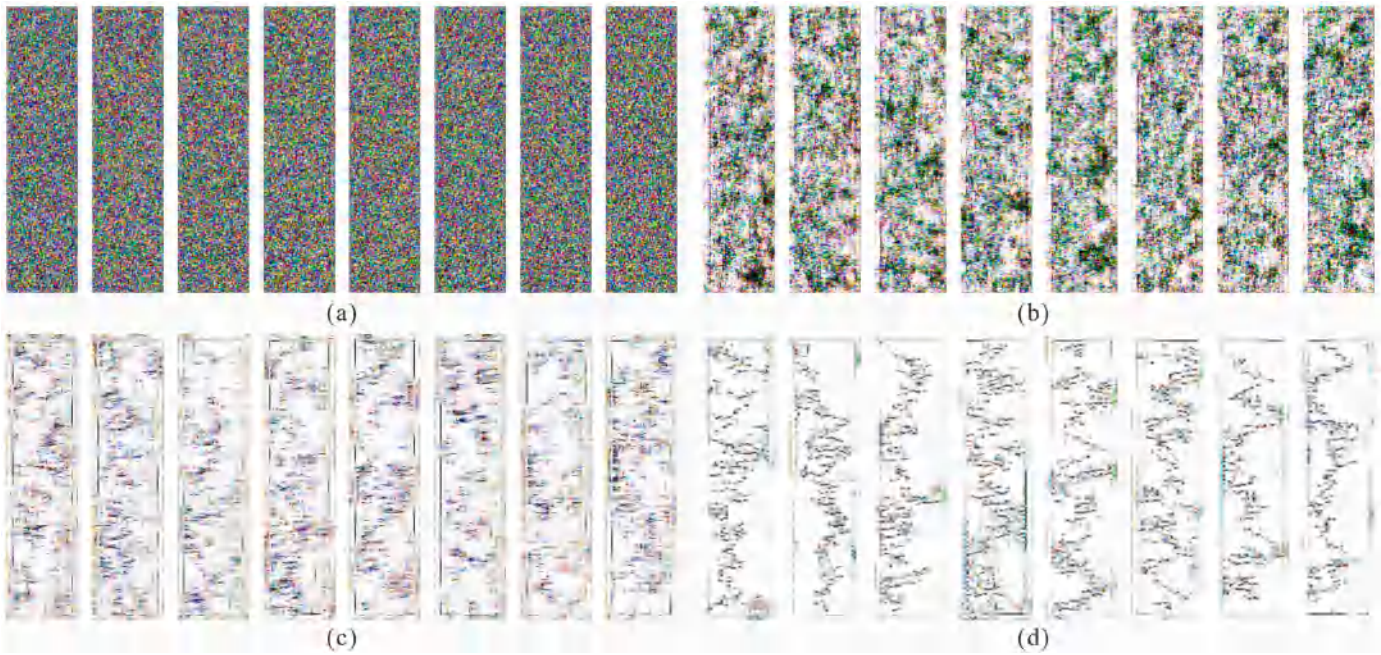


Fig. 13. Original well log curve images generated by the GANs: (a) G images for the 1st epoch; (b) G images for the 500th epoch; (c) G images for the 1000th epoch; and (d) G images for the 2000th epoch.

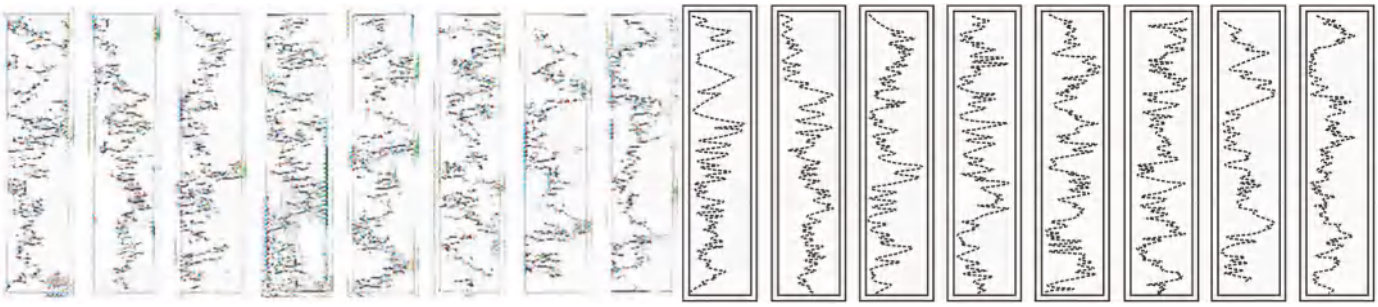


Fig. 14. Well log image optimization via the convolutional process of sharpening.

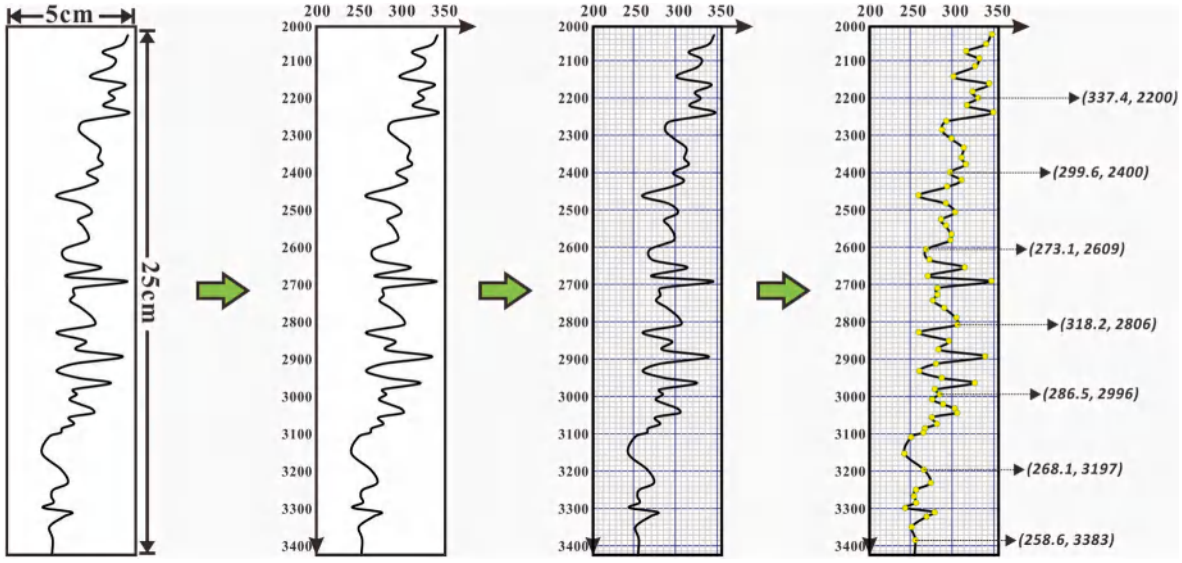


Fig. 15. Image vectoring process for virtual well dataset construction.

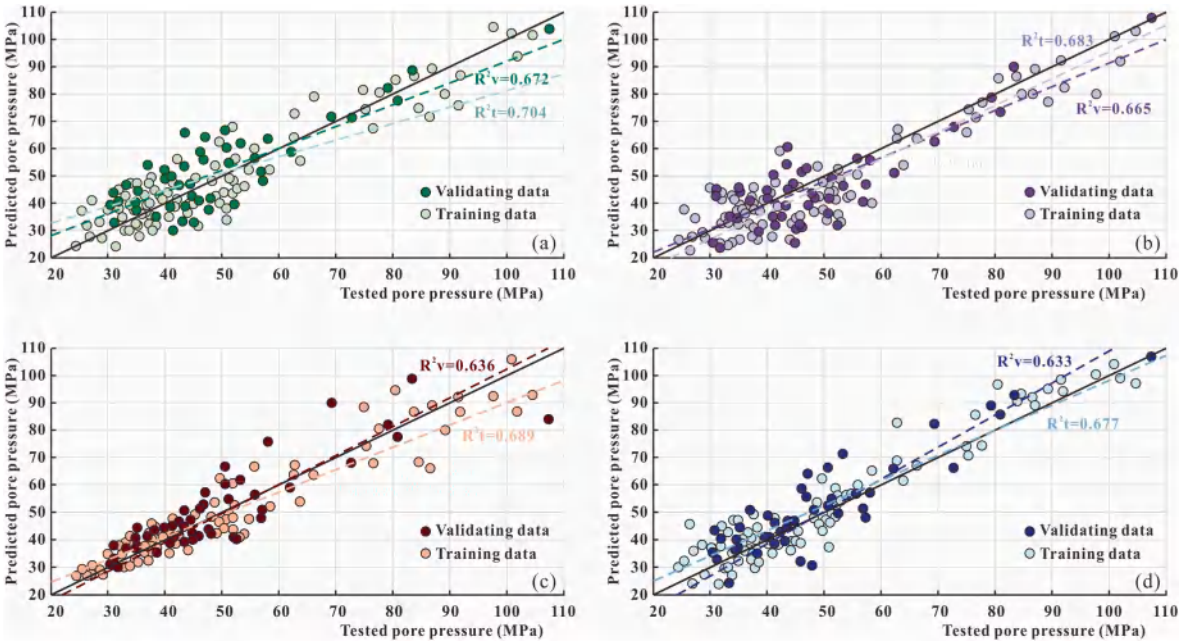


Fig. 16. Pore pressure prediction models with the highest validation accuracy. (a) Model 19 with a training Ccoe of 0.704 and a validating Ccoe of 0.672; (b) Model 371 with a training Ccoe of 0.683 and a validating Ccoe of 0.665; (c) Model 1265 with a training Ccoe of 0.689 and a validating Ccoe of 0.636; and (d) Model 1798 with a training Ccoe of 0.677 and a validating Ccoe of 0.633.

prediction model. The hyperparameters are determined by random searching. All the models are training by 1 RTX GeForce 3060 GPU with 12.7 TFLOPS and 12 GB memory. The average training time is 1.54 h. The training accuracy shows a correlation coefficient (Ccoe) between the predicted pore pressure and the actual test pressure ranging from 0.654 to 0.821, with an average of 0.701. The training results indicate the instability of the training models with large differences. For the 4 best models with the highest training accuracies, Model 19 shows similar performance in terms of the prediction of small and large values (Fig. 16a), while Model 371 and Model 1798 show better performance in terms of the prediction of large values (Fig. 16b and d). Model 371 shows better performance in terms of the prediction of small values (Fig. 16c). With the determination of the hyperparameters, the validation accuracy shows that the Ccoe range from 0.585 to 0.755, with an average of 0.674 (Table 4 and Fig. 16). The validation performances of the 4 best models are similar to the corresponding training performances. Note that the training and validation processes are based on one series of datasets and are not calibrated by the whole model optimization and improvement processes, which will change the original BPNN dataset and are introduced in Section 4.3.

4.2.3. Testing of the BPNN model

The BPNN model is tested in two ways: on the one hand, the test dataset containing 10 % of the data in the original dataset are used for testing the generalization ability of the constructed models; on the other hand, 5-fold cross-validation are applied considering the relatively small dataset used in this study.

4.2.3.1. Model testing by using test data. The models used for testing are the best 4 models with the highest validation accuracy. The testing results indicate that the 4 models exhibit similar pore pressure prediction performances, which are lower than the training and calibration accuracies (Fig. 17). Model 19 has the best Ccoe, 0.606, for the predicted pressure, while Models 371, 1265 and 1798 have Ccoe values of 0.592, 0.584 and 0.602, respectively. Note that these 4 models also show disparities: Model 19 has higher accuracy in predicting low values, while Model 371 and Model 1265 are good at predicting high values, especially for pressures larger than 80 MPa. Model 1798 has a balanced performance in predicting low values and high values.

4.2.3.2. Model testing by 4-fold cross-validation. Considering that the size of the test data is relatively small, through which the testing accuracy may not reflect the actual performance of the model, cross-validation is applied to eliminate the bias of the testing accuracy (Malakouti et al., 2023; Roy and Saha, 2022; Saud et al., 2020). In this research, 4-fold cross-validation is applied. The results show that the testing Ccoe of the cross-validation ranges from 0.589 to 0.606 with an average of 0.598, which is similar to that of the testing of Ccoe by the test dataset (Fig. 18). Overall, the BPNN pore pressure prediction model can predict deep-strata pore pressure with a Ccoe of approximately 0.6, which also has great potential for improvement.

4.3. Model calibration and improvement

Owing to the large potential for improvement, the BPNN model and the GAN model are calibrated by the constructed loss function and the random vector searching process, the principles of which are shown in Section 3.5. The calibrating process first changes the original dataset of the BPNN model and then constructs the new BPNN model again for pore pressure prediction. In other words, the loss function helps the GAN model find a better way to generate a virtual well log dataset that can be used as a real dataset. For the cycle of dataset generation, pore pressure prediction and model adjustment guided by the loss function, the performance of the whole model can be enhanced.

In this research, Model 19, which has the highest training, validating and testing accuracy, is set as the initial model for calibration (Fig. 19a). With the constraint of the loss function and random vector searching, the predicted pore pressure gradually approaches the actual tested pressure. For the 1000th calibrating epoch, the model shows a training Ccoe of 0.755 and a testing Ccoe of 0.705 (Fig. 19b). For the 1500th calibrating epoch, the model shows a training Ccoe of 0.904 and a testing Ccoe of 0.823, which is an obvious improvement compared to the initial epoch (Fig. 19c). Finally, the model for the 2000th calibrating epoch is characterized by a training Ccoe of 0.988 and a testing Ccoe of 0.894, which indicates that the model can predict the pore pressure in the deep strata with an error of less than 1 MPa (Fig. 19d). During the calibration and improvement process, the loss of the model decreases from 4.1 to 1.9 (Fig. 20). Therefore, the final model can absolutely fulfil the requirements of pore pressure prediction in practice.

Table 4
Hyperparameter and training time of the 4 best training models.

Hyperparameter	Model 19	Model371	Model 1265	Model 1798
Kernel stride	2	2	2	2
Kernel padding	1	1	1	1
Learning rate of G	0.0001	0.001	0.0001	0.00001
Learning rate of D	0.001	0.01	0.0001	0.0001
Training time	1.2 h	0.7 h	1.7 h	2.1 h

4.4. Uncertainty analysis

For further testing the reliability and the practicality of the pore pressure prediction model, the uncertainty analysis is required. Since the ensemble BPNN is applied in this research, the uncertainty of pore pressure predicting is evaluated by a comparison between a couple of models in the ensemble learners. The uncertainty evaluation result and metrics are listed in Table 5. The result shows that the mean uncertainty of the 10 models is 3.40 with the root mean square error (RMSE) of 2.73. Approximate 84.6 % predicted results locate in the confidence interval of 95 % with roughly stable width of the confidence interval, which indicate that the model has a balanced performance for pore pressure prediction in deep strata (Fig. 21a). For further evaluating the reliability of the uncertainty analysis in this research, a correlation analysis between pore pressure predicting error and uncertainty is constructed. The result shows a positive relation between error and uncertainty with a correlation coefficient of 0.775, which indicate high reliability of the uncertainty analysis in this research (Fig. 21b).

4.5. Prediction of deep-strata pore pressure by the GANs and BPNN

The constructed model can be applied for obtaining the continuous distribution in deep strata, which are characterized by fewer exploration wells. In this research, the pore pressure prediction models of GANs and BPNNs provide a total of 27 virtual well log datasets for the deeply buried Jurassic strata of the Central Junggar Basin for pore pressure prediction profile construction. Without the pore pressure prediction model, there are only 3 exploration wells with 5 tested pressure data points, which limits the difficulty and uncertainty of determining the pore pressure distribution in the deep strata (Fig. 22a). With the application of the prediction model, 27 virtual wells with over 300 predicted pore pressure data points are used for the construction of the pore pressure distribution profile (Fig. 22b). Therefore, the deep-strata pore pressure distribution is easy to obtain (Fig. 22c).

5. Discussion

5.1. Effectiveness evaluation of the GANSim framework

In this research, the GANSim framework is applied to constrain G to generate a well log dataset with high efficiency. One of the most important reasons for the utilization of the GANSim algorithms is that

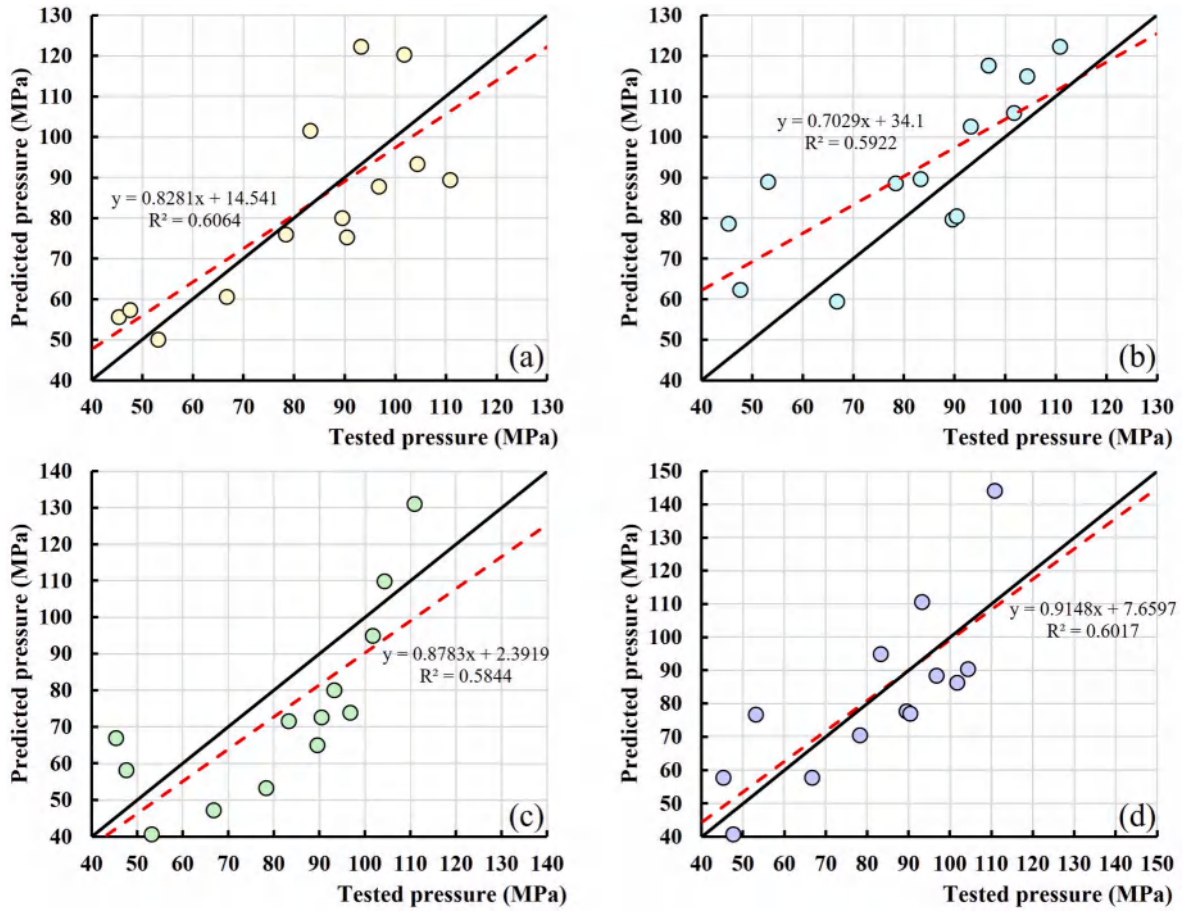


Fig. 17. Testing performance of the 4 models with the best validation accuracy (the dotted line in red is the trendline). (a) Model 19 with a Ccoe of 0.606; (b) Model 371 with a Ccoe of 0.592; (c) Model 1265 with a Ccoe of 0.584; and (d) Model 1798 with a Ccoe of 0.602. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the virtual well construction model based on single GANs always shows instability and low efficiency, which results in long training times and relatively high losses for both D and G in the GANs. The seismic data used in GANSim can provide guidance for the construction of well log curves by constraining the outline for the generated images of G; therefore, the randomly distributed dots are constrained in a limited space to form the images, which will obviously enhance the efficiency of G in finding ways to generate the expected well log curve images (Fig. 23a).

To prove the effectiveness of the GANsim framework, a virtual well construction model with the single DCGAN model is also constructed. The results of the single DCGAN model show that in the 2000 training epochs, the model can only generate well log images with a shadowy outline, which is similar to the image generated by GANSim for the 500th epoch (Fig. 23b). The loss of D and G for the GAN models also shows the effectiveness of the GANSim framework. The loss of DCGAN with the GANSim framework shows a gradually decreasing or increasing trend for D and G. Along with the training procedure, the loss of D and G becomes stable without obvious change, which indicates that they reach the Nash equilibrium and that model training is accomplished. The loss of the single DCGAN indicates that during the whole training process, the loss of D and G is not stable without an obvious increasing or decreasing trend, which indicates that the model training process is still not accomplished and that the model cannot provide the well log curves as expected (Fig. 23c and d).

5.2. Comparison between multiple GANs for virtual well construction

The performance of the virtual well construction largely depends on the selection of the GANs. Different types of GANs have their own advantages and disadvantages for different requirements or targets. In this research, the normal GAN, Cycle-GAN, Style GAN and Self-attention GAN (SAGAN) models for virtual well construction are established for comparison. For each pair of comparisons, the initial imputing data for D are unique, and 10 different G models using different GANs are constructed. The performances of the virtual well construction models are shown by the generated images and the recoded loss of D and G.

5.2.1. Performance of normal GANs

Although the normal GAN is the initial GAN model based on fully connected neural networks, it shows good performance and fast model training speed for certain problems (Kuntalp and Düzyel, 2024; Lataifeh et al., 2024; Farhadinia et al., 2024). By imputing actual well log curve images, the G of the normal GANs is trained for 2000 epochs. The final result of the generated images is characterized by randomly distributed dots without obvious shapes of the well log curves (Fig. 24a). From the first to the last epoch, the generated images undergo few changes, which indicates that the G of the normal GANs may not be effective at generating well log curves in the deep strata. The record of losses of D and G show that for all 10 models, the loss of D and G is characterized by increasing or decreasing trends instead of a stable status, which indicates that the model is not sufficiently good for meeting the requirements (Fig. 24b).

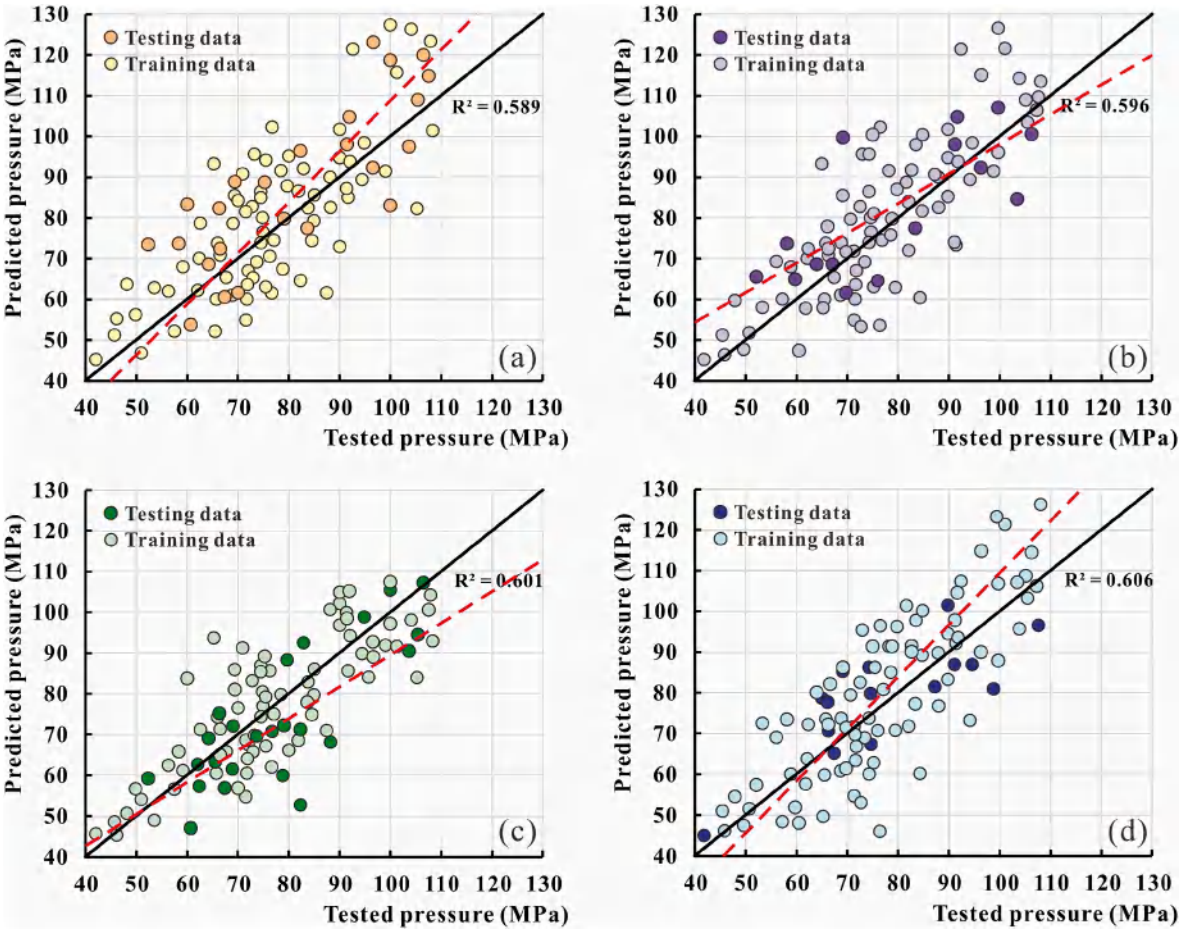


Fig. 18. 4-fold cross-validation test results for the reserve abundance prediction models using SVR. Figures (a)–(d) show the 4 SVR models of the 4-fold cross-validation.

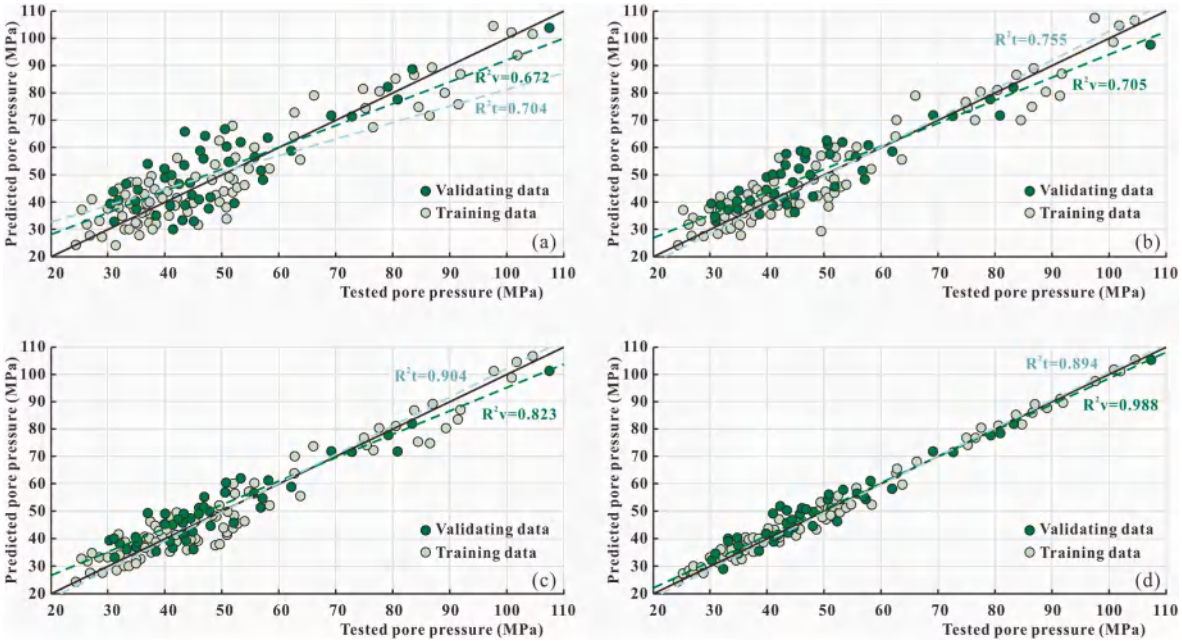


Fig. 19. Training and testing accuracies of the calibrated pore pressure prediction model. (a) 1st epoch with a training Ccoe of 0.704 and testing Ccoe of 0.605; (b) 1000th epoch with a training Ccoe of 0.755 and testing Ccoe of 0.705; (c) 1500th epoch with a training Ccoe of 0.904 and testing Ccoe of 0.823; and (d) 2000th epoch with a training Ccoe of 0.988 and testing Ccoe of 0.894.

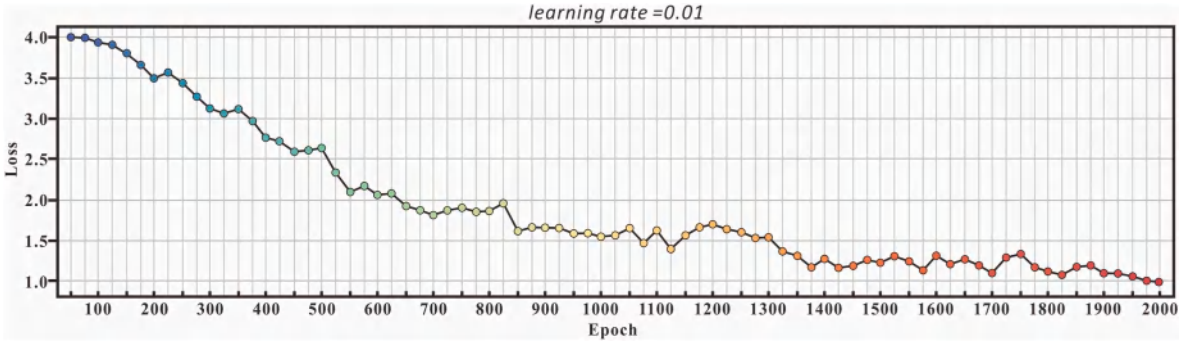


Fig. 20. Loss of the pore pressure prediction model.

Table 5
Uncertainty analysis result and metrics of pore pressure prediction model.

Metric	Value
Confidence Level	0.95
Coverage	0.846
Uncertainty-Error Correlation	0.77
Mean Uncertainty	3.40
RMSE	2.73
Number of Groups	10
Number of Samples	39

5.2.2. Performance of Cycle-GAN

Cycle-GAN is an innovative GAN model for image style transfer that can change the styles of images without converting the outlines of the original images (Fukuda et al., 2024; Chen et al., 2023; Verma et al., 2024). In this research, the seismic images are set as the imputing images, and they are expected to be transferred to the styles of well log curves. However, the output images do not show the ideal results: from the first to the 2000th epoch, the predicted images are characterized by a random distribution without effective messages (Fig. 25a). In fact, the Cycle-GAN framework may not be suitable for virtual well construction because the outlines of the seismic images and well log curves are different. Even though the Cycle-GAN model can learn the style of well log curve images, it cannot change the outline of the inputted seismic images. The loss of D and G for the 10 generated Cycle-GANs indicates that the loss of both D and G tends to decrease or increase instead of stabilize, which proves that the Cycle-GAN model is still not sufficiently good to provide the expected output images (Fig. 25b).

5.3. Effectiveness and generalization of the pore pressure by GAN-BPNN

5.3.1. Comparison between multiple pore pressure prediction methods

For further testing the performance of the BPNN-GAN model in pore pressure prediction, a comparison between different pore pressure prediction methods is raised. Beside our model, there are several dominated pore pressure prediction methods are chosen, including Eaton's method, Bowers' method and Fillippone method. Take the measured pressure data showed in Fig. 23 as the test data, the pore pressure predicting results and accuracy analysis are listed in Table 6. The result for pore pressure prediction in Junngar basin showed that the BPNN-GAN based model provide the best predicting performance with the highest coefficient correlation, the lowest mean absolute error (MAE), mean square error (MSE) and RMSE. Comparing to all other methods, the BPNN-GAN model has obvious advantage in predicting ultra-high pore pressure with a decreasing in MSE. While the other ML methods also show better performance than that of the traditional pore pressure prediction method based on physical mechanism and experimental parameters. This result indicates that the current physical-driven methods is not suitable for pore pressure prediction in deep strata.

5.3.2. Generalization of the BPNN-GAN model in pore pressure prediction

The generalization of the model is important for the utilization of the methods. In this study, all the seismic and drilling data is from the Junngar basin. For testing the generalization of the model in pore pressure prediction, we have collected other 32 pore pressure data measured in deep-strata (with the depth deeper than 6000m) from Sichuan Basin, which is another petroliferous basin in China. The final result shows an acceptable generalization performance of our model with the coefficient correlation of 0.90 in pore pressure prediction of Sichuan Basin. Comparing to other ML methods, the GAN-BPNN shows better performance pore prediction results of Junngar Basin which indicates the good generalization of our model. Furthermore, the predicting performance can be further enhanced by a re-training process

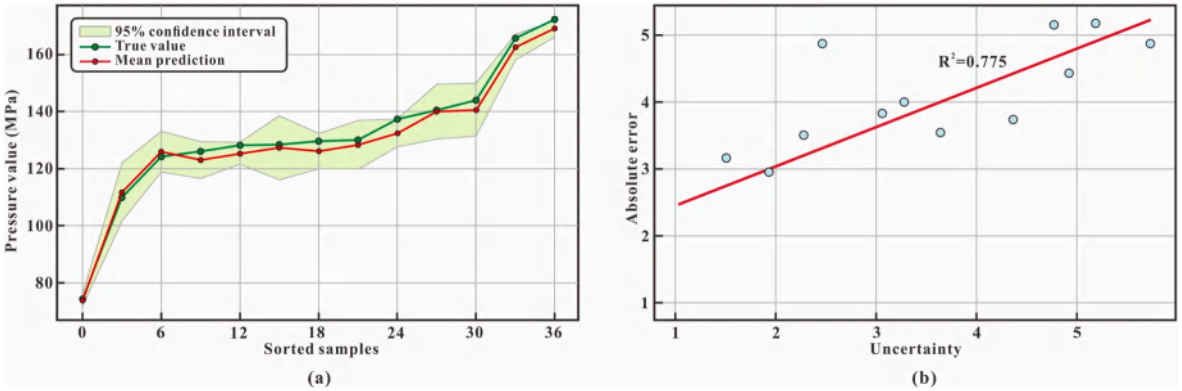


Fig. 21. Uncertainty evaluation of pore pressure prediction. (a) Distribution of 95 % confidence interval. (b) Corelation between predicting error and uncertainty.

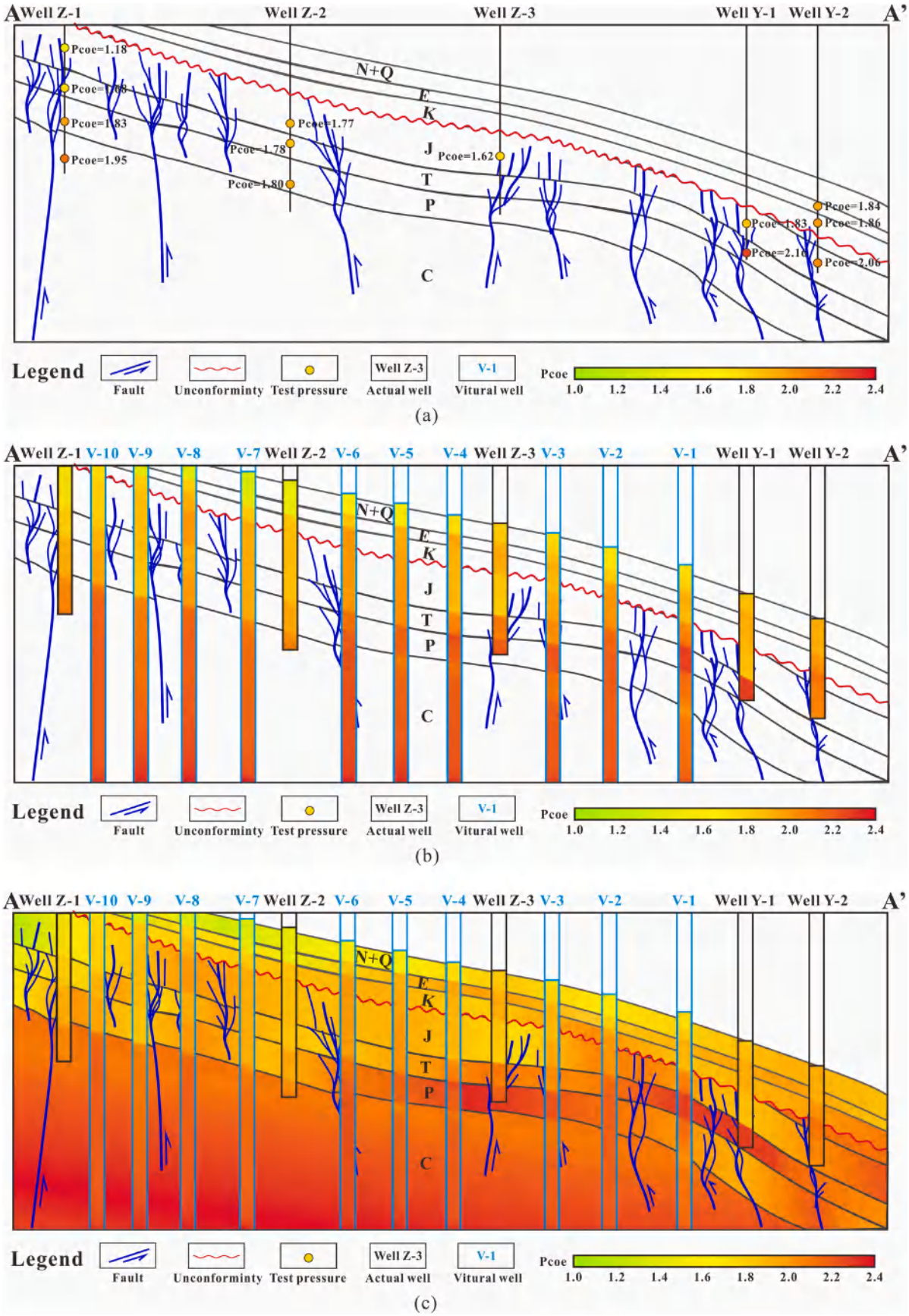


Fig. 22. Prediction of the deep-strata pore pressure in the Central Junggar Basin. (a) Tested pressure distribution. (b) Virtual well construction and pore pressure prediction for virtual wells. (c) Profile overpressure distribution prediction in the deep strata. (d) Consistence between tested pressure and predicted pressure by GANs and BPNNs.

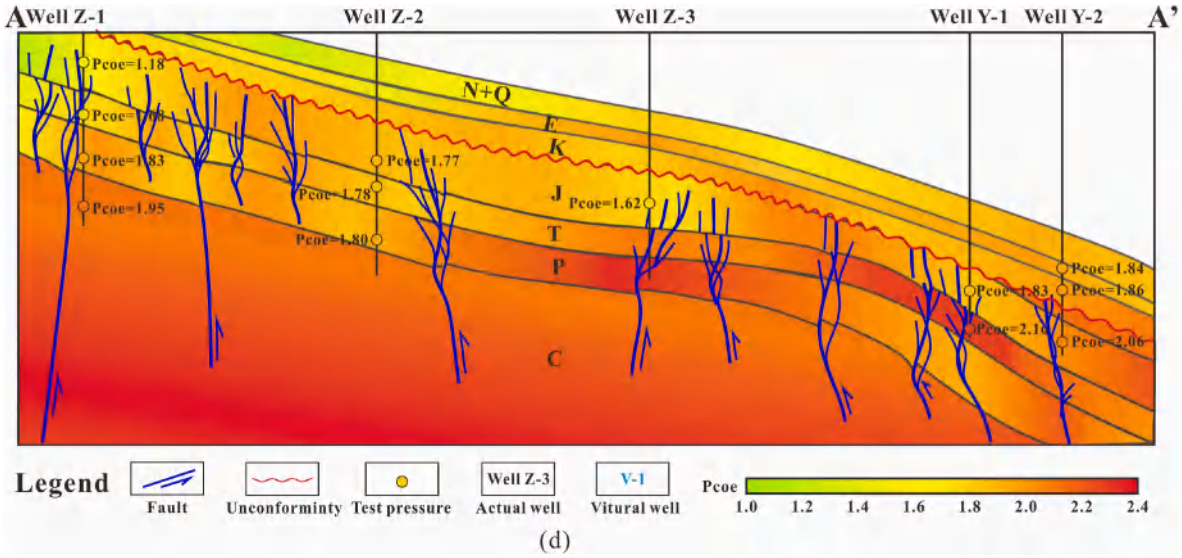


Fig. 22. (*continued*).

with regional training data by the same workflow.

5.4. Limitations and improvement of the model

Although the model can predict the pore pressure of deep strata with acceptable accuracy, it also has some limitations and its utilization can be improved in many aspects. Comparing to other GAN models, we utilized some guidance for enhancing the efficiency of the model for generating the results we want. While such guidance is also the constraining conditions, which contains too much specific messages for a certain geological setting and then lead to an overfitting of model. The generalization test also indicates an overfitting tendency of this model. Besides, the high efficiency and lower cost of this model relies on the seismic data. If we do not have the seismic data, or we cannot obtain the high-quality seismic data in some special conditions (such as the deep strata with extreme high temperature and pressure that will largely influence the seismic signal), the GANSim may be ineffective. Considering the factors that significantly influence the performance of the predicted models, there are several ways to improve the virtual well construction and pore pressure prediction models.

(1) The original dataset can be further optimized. The quality of the dataset is one of the most important factors for the performance of machine learning models. For pore pressure prediction, the original seismic and well log data can be optimized by detailed classification due to the different qualities of the original data in different depth ranges, well log or seismic data types, and even the equipment utilized during exploration. For well log curve generation, the data augmentation methods, including geometry transformation, noise injection and mixup augmentation is effective for enriching the training datasets, which may lead to the GAN generate more useful well log curve datasets for dealing with multiple geophysical problems.

(2) The virtual well construction model introduced in this research is only effective for pore pressure prediction due to the calibration of tested pressure data, which does not mean that the well log curve generated by G can completely replace the real curves because there is still a fundamental difference between the pixel to pixel correspondence and the depth range correspondence, which is applied in this research. However, if we can collect multiple tested data that are affected by well log data (such as the tested density, porosity, permeability and Poisson's ratio) and find a suitable corresponding way between seismic and well log data in pixel, the generated well log curve can be better calibrated and become more similar to the actual data.

(3) The semisupervised Pix2pix GAN is still the proper GAN model for virtual well construction because it provides possible ways to directly transfer seismic images to corresponding well log curves (Rajabi et al., 2023; Pu et al., 2024). In the future, with a large dataset and sufficient actual drilling wells for deep strata, the coupled seismic and well log data can be applied to the Pix2pix GAN model to generate an effective virtual well log dataset. Furthermore, the advanced GAN models (such as Style-GAN and SAGAN) are worth trying to generate high-quality well log curves with much more details. While the computation and time cost need to be considered.

(4) Since the current model has shown the overfitting feature, the overfitting calibration can be applied for further balancing the generalization and applicability for specific area. Generally, we hope to create a model with excellent generalization, while sometimes we do need a higher accuracy of the predicting result in some special situation, such as the precise prediction for the extremely high pore pressure, which is closely related to safety. In this case, a sequence-related feedback mechanism will be helpful. Based on a pre-trained model with strong generalization, the new real data for specific study area can be added and generate a sub-model by online learning or incremental learning. With more and more real data, the sub-model can be iterated and continuously optimized.

(5) The current matching between the 2D seismic image and the 1D well log curve data may lead to the difficulty in virtual data generation by GANs, which is already reflected on the generated image of well log. In this case, transferring the 2D seismic image (adjacent to drilling well) to 1D vector by encoding is an effective way to match the two types of data in dimension. Furthermore, the 1D-to-1D mapping would drastically reduce model complexity of GANs, ensure a more physically meaningful depth-to-depth alignment, and enable the model to predict other petrophysical properties of geological bodies more accurately.

6. Conclusion

(1) This study introduces a novel method for pore pressure prediction in deep-strata petroleum resource exploration via GANs and BPNNs. To address the problem of scarce drilling wells and therefore the need for well log data for precise pore pressure

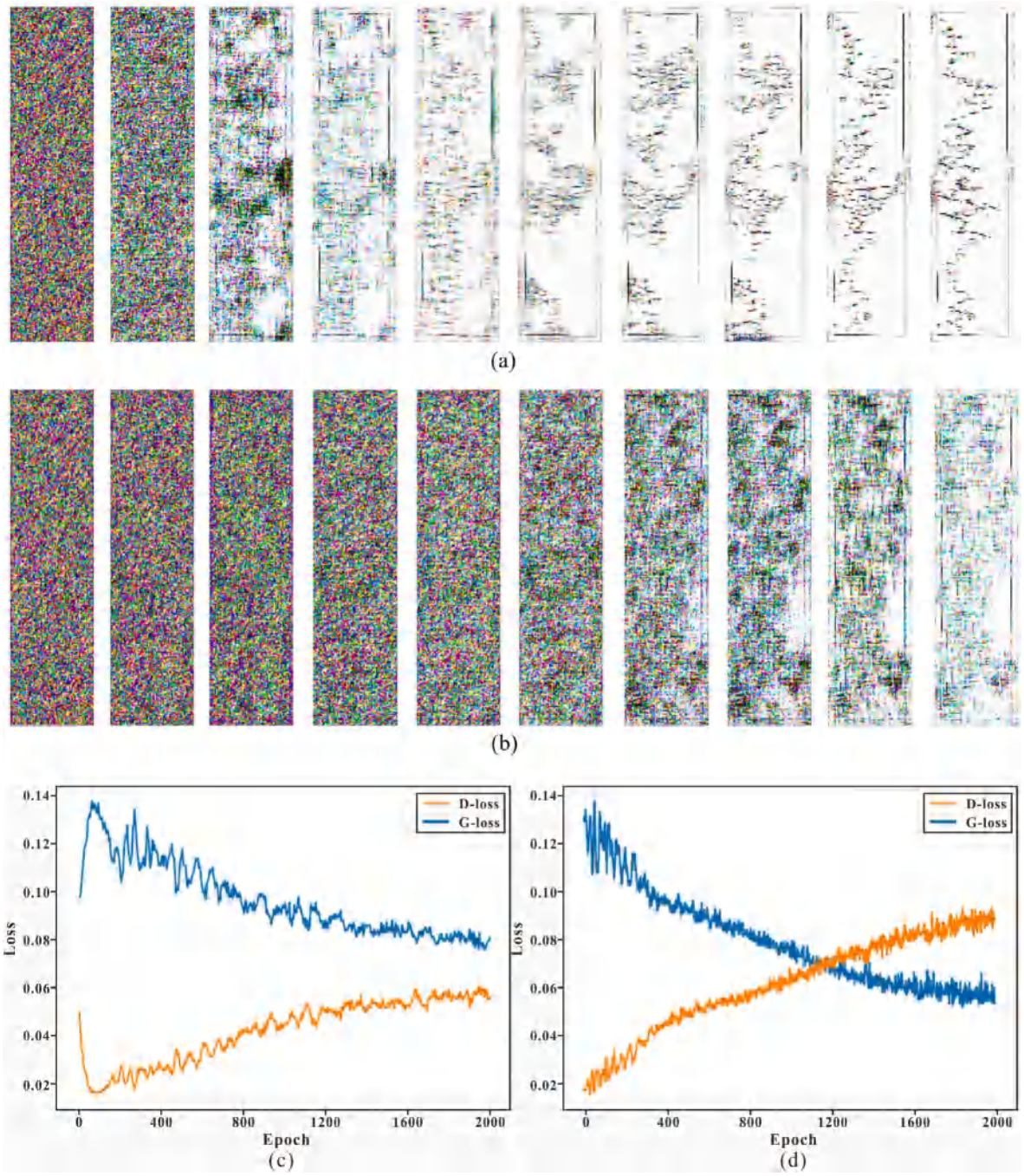


Fig. 23. Virtual well log curve constructed by the DCGAN with the GANSim framework. (b) Virtual well log curve constructed by the single DCGAN; (c) loss of D and G by the DCGAN with the GANSim framework; and (d) loss of D and G by the single DCGAN model.

prediction, we first apply the DCGAN algorithm and GANSim framework to construct a virtual well dataset to provide enough well log data with the guidance of seismic data in deep strata. Then, a pore pressure prediction model is established based on the BPNN with the calibration of the tested pressure.

(2) The combined methods for virtual well construction and deep-strata pore pressure prediction exhibit good performance. The optimized DCGAN with GANSim framework finally generates 2D well log curves that cannot be distinguished by the D guided by the actual well log curve images. The pore pressure prediction model of the BPNN also shows good performance and high prediction accuracy with the ensemble methods and calibration of the test data with the final Ccoe between the predicted and test pore pressures reaches 0.988. The uncertainty analysis shows that approximate 84.6 % predicted results locate in the confidence interval of 95 %, which indicates high reliability of the model. The generalization test with the data in another district shows a correlation coefficient of 0.90, which can fulfil the pore pressure predicting requirement.

(3) In the actual situation of limited drilling well and test data, the virtual well construction based on the GANSim and DCGAN frameworks shows better performance than the normal GAN and Cycle-GAN, which are both characterized by relatively high loss of D and G over 0.5 and still do not show a stable status, with an obvious tendency to increase or decrease after 2000 training cycles. The results also indicate the suitability of the DCGAN for

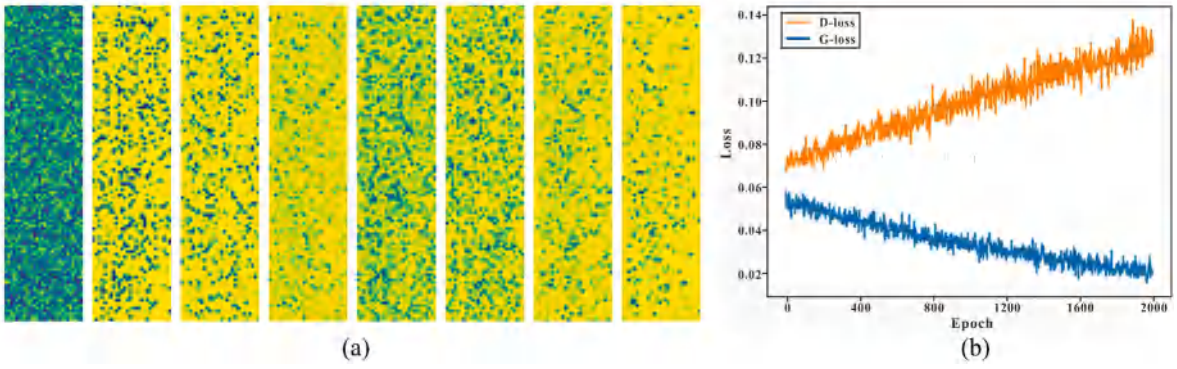


Fig. 24. (a) Performance of normal GANs in virtual well log curve generation; and (b) loss record of D and G by normal GANs.

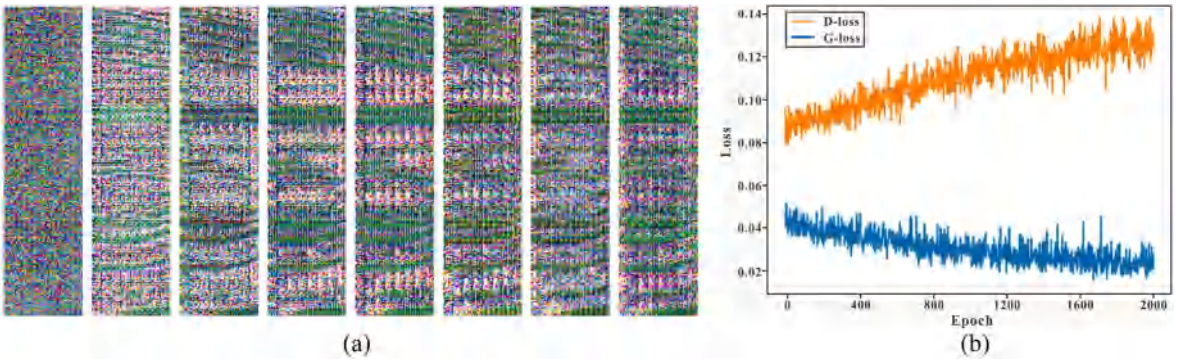


Fig. 25. (a) Performance of Cycle-GAN in virtual well log curve generation; and (b) loss records of D and G obtained by Cycle-GAN.

Table 6
Pore pressure prediction results by multiple methods in Junngar basin (JB) and Sichuan Basin (SB).

Type of method	Name of method	Basin	Coefficient corelation (R^2)	MAE	MSE	RMSE	References
ML	**BPNN-GAN**	JB	0.98	2.76	10.03	3.17	/
		SB	0.90	2.28	7.33	2.71	
	LSTM-BP	JB	0.95	4.67	25.62	5.06	Song et al. (2022)
		SB	0.94	1.77	4.52	2.12	
	CNN-LSTM	JB	0.93	5.64	39.60	6.29	Fang (2022)
		SB	0.82	3.14	13.36	3.66	
	CNN-Eaton	JB	0.89	5.76	59.10	241.86	/
		SB	0.82	3.22	12.85	3.59	
Physical informed	**Eaton**	JB	0.55	13.22	241.86	15.55	Eaton (1975)
		SB	0.63	4.45	26.84	5.18	
	Bowers	JB	0.60	12.76	211.58	14.55	Bowers (2002)
		SB	0.65	4.26	25.81	5.08	
	Fillipone	JB	0.61	12.84	208.09	14.43	Fillippone. (1982)
		SB	0.70	4.20	21.53	4.64	

image generation and for enhancing the efficiency of the GANSim framework.

(4). There are several ways to enhance the models, including optimizing the original dataset, adding the constraining test data, applying advanced GAN models and use the sequential data to construct a feedback mechanism of the model. The use of the GAN model also provides a reference for the use of artificial intelligence in deep-strata petroleum exploration or other problems with limited original data.

CRediT authorship contribution statement

Qiaochu Wang: Writing – original draft, Funding acquisition, Conceptualization. **Dongxia Chen:** Project administration, Conceptualization. **Fuwei Wang:** Data curation. **Zijie Yang:** Investigation. **Zaiquan Yang:** Software. **Sha Li:** Investigation. **Yuqi Wang:** Writing – review & editing.

Declaration of competing interest

No conflict of interest exits in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

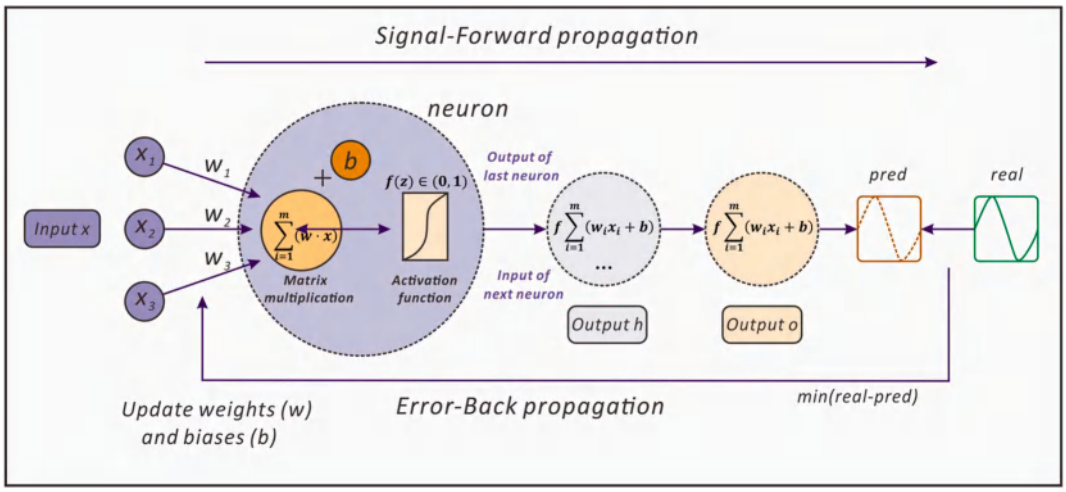
Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant No. 42302141) and Science Foundation of China University of Petroleum Beijing (Grant No. 2462025YJRC005). We

gratefully acknowledge the Shengli Oilfield Branch of the China Petroleum & Chemical Corporation for providing field test data and reservoir data.

Appendix

(1) The Framework of the backpropagation neural network is shown in the following figure:



Data availability

Data will be made available on request.

References

Abdellatif, A., Elsheikh, A.H., Graham, G., et al., 2022. Generating unrepresented proportions of geological facies using generative adversarial networks. Comput. Geosci. 162, 105085. https://doi.org/10.1016/J.CAGEO.2022.105085.

Ahmad, A.B., Tsuji, T., 2021. Machine learning for automatic slump identification from 3D seismic data at convergent plate margins. Mar. Petrol. Geol. 133, 105290. https://doi.org/10.1016/J.MARPETGEO.2021.105290.

Aldhubri, A., Lu, J., Fu, G., 2024. SAGAN: skip attention generative adversarial networks for few-shot image generation. Digit. Signal Process. 149, 104466.

Alharbi, O.Q., Alarifi, S.A., 2023. Productivity index prediction for single-lateral and multilateral oil horizontal Wells using machine learning techniques. ACS Omega 8 (7), 7201–7210.

Alzubaidi, F., Mostaghimi, P., Swietojanski, P., et al., 2021. Automated lithology classification from drill core images using convolutional neural networks. J. Petrol. Sci. Eng. 197, 107933.

Amaya, J.E., Camargo, E., Aguilar, J., et al., 2021. A proposal for a cooperative cross-entropy method to tackle the unit commitment problem. Comput. Ind. Eng. 162, 107764. https://doi.org/10.1016/J.CIE.2021.107764.

Bai, M., Zhang, Z., Cao, K., et al., 2023. Prediction of compressive strength of fly ash-slag based geopolymer paste based on multi-optimized artificial neural network. Materials 16 (3), 1090. https://doi.org/10.3390/MA16031090.

Bashkova, S.E., Kozhevnikova, E.E., Yakovlev, Y.A., et al., 2022. The deep structure of the proterozoic complex of the sedimentary cover the east European platform: oil and gas potential. Geotectonics 57 (4), 428–445.

Bessmeltsev, M., Solomon, J., 2019. Vectorization of line drawings via polyvector fields. ACM Trans. Graph. 38 (1), 1–12. https://doi.org/10.1145/3202661.

Bhatti, B.A., Broadwater, R., 2020. Distributed nash equilibrium seeking for a dynamic micro-grid energy trading game with non-quadratic payoffs. Energy 202, 117709. https://doi.org/10.1016/J.ENERGY.2020.117709.

Bowers, G.L., 2002. Detecting high overpressure. Lead. Edge 21 (2), 174–177.

Bowers, G.L., 1995. Pore pressure estimation from velocity data; accounting for overpressure mechanisms besides undercompaction. SPE Drilling and Completions, June, 1995, 89–95.

Byeon, H., Shabaz, M., Shrivastava, K., et al., 2024. Deep learning model to detect deceptive generative adversarial network generated images using multimedia forensic. Comput. Electr. Eng. 113, 109024. https://doi.org/10.1016/J.COMPELECENG.2023.109024.

Chaki, S., Routray, A., Mohanty, W.K., 2022. A probabilistic neural network (PNN) based framework for lithology classification using seismic attributes. J. Appl. Geophys. 199, 104578.

Chen, Y., Wu, J., Zhang, Y., et al., 2022. Research on hyperparameter optimization of concrete slump prediction model based on response surface method. Materials 15 (13), 4721. https://doi.org/10.3390/MA15134721.

Chen, C., Liu, W., Tan, X., et al., 2023. Semi-supervised Cycle-GAN for face photo-sketch translation in the wild. Comput. Vis. Image Understand. 235, 103775. https://doi.org/10.1016/J.CVIU.2023.103775.

Cheng, B.J., Tian, J.X., Luo, S.Y., et al., 2022. Method and practice of deep favourable shale reservoirs prediction based on machine learning. Petrol. Explor. Dev. 49 (5), 1056–1068.

Chitkeshwar, A., 2024. The role of machine learning in earthquake seismology: a review. Arch. Comput. Methods Eng. 28, 1–13.

Colombera, L., Mountney, N.P., 2020. Accommodation and sediment - supply controls on clastic parasequences: a Meta - analysis. Sedimentology 67, 1667–1709.

Creswell, A., White, T., Dumoulin, V., et al., 2018. Generative adversarial networks: an overview. IEEE Signal Process. Mag. 35 (1), 53–65.

Cui, Y., Liu, H., Wang, Q., et al., 2022. Investigation on the ignition delay prediction model of multi-component surrogates based on back propagation (BP) neural network. Combust. Flame 237, 111852. https://doi.org/10.1016/J.COMBUSTFLAME.2021.111852.

De Souza, V.L., Marques, B.A., Batagelo, H.C., et al., 2023. A review on generative adversarial networks for image generation. Comput. Graph. 114, 13–25.

Dorotic, M., Stagno, E., Warlop, L., 2024. AI on the street: Context-dependent responses to artificial intelligence. Int. J. Res. Market. 41 (1), 113–137.

Eaton, B.A., 1975. The Equation for Geopressure Prediction from Well Logs. Society of Petroleum Engineers of AIME. Paper SPE 5544.

Ebraheem, M.O., Ibrahim, H.A., Ewida, H.F., et al., 2022. Identification of hydrocarbon-bearing zones within the early cretaceous reservoir rocks using well logging and seismic reflection data in Al-Baraka field, West Komombo, Egypt. J. Petrol. Sci. Eng. 218, 111037. https://doi.org/10.1016/J.PETROL.2022.111037.

Eshun, R.B., Islam, A.K., Bikdash, M., 2024. A deep convolutional neural network for the classification of imbalanced breast cancer dataset. Healthcare Analytics, 100330.

Fan, M., Zuo, K., Wang, J., et al., 2023. A lightweight multiscale convolutional neural network for garbage sorting. Systems and Soft Computing 5, 200059. https://doi.org/10.1016/J.SASC.2023.200059.

Fang, H., 2022. Enrichment mechanism and prospects of deep oil and gas. Acta Geol. Sin. 96 (3), 742–756.

Farhadinia, B., Ahangari, M.R., Heydari, A., et al., 2024. A generalized optimization-based generative adversarial network. Expert Syst. Appl. 248, 123413. https://doi.org/10.1016/J.ESWA.2024.123413.

Fillippone, W.R., 1982. Estimation of formation parameters and the prediction of overpressures from seismic data. SEG Technical Program Expanded Abstracts 1, 502-503.

Fukuda, M., Kotaki, S., Nozawa, M., et al., 2024. A cycle generative adversarial network for generating synthetic contrast-enhanced computed tomographic images from non-contrast images in the internal jugular lymph node-bearing area. Odontology 41 (2), 1–10.

Gafni, R., Aviv, I., Kantsepolsky, B., et al., 2024. Objectivity by design: the impact of AI-driven approach on employees' soft skills evaluation. Inf. Software Technol. 170, 107430.
Gao, Z.Y., Cui, J.G., Fan, X.R., et al., 2023. Action mechanisms of abnormal fluid pressure on physical properties of deep reservoirs: a case study on Jurassic Toutunhe Formation in the southern margin of Junggar Basin, NW China. Petrol. Explor. Dev. 50 (6), 1398–1410.
Gene, J., Park, S., Shin, H.C., et al., 2024. Hybrid optical convolutional neural network with convolution kernels trained in the spatial domain. Neurocomputing, 127251.
Goodfellow, I., McDaniel, P., Papernot, N., 2018. Making machine learning robust against adversarial inputs. Commun. ACM 61 (7), 56–66. https://doi.org/10.1145/3134599.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al., 2020. Generative adversarial networks. Commun. ACM 63 (11), 139–144. https://doi.org/10.1145/3422622.
Hastie, T., Tibshirani, R., Friedman, J., 2008. The Elements of Statistical Learning, second ed. Springer. ISBN 0-387-95284-5.
He, Y., Kang, S.H., Morel, J.-M., 2023. Topology- and perception-aware image vectorization. J. Math. Imag. Vis. 65, 874–893.
Hottman, C.E., Johnson, R.K., 1965. Estimation of formation pressures from log-derived shale properties. Journal of Petroleum Technology 17 (6), 717–722.
Hou, X., Zhou, X., Liu, L., 2024. Reconstruction of ship propeller wake field based on self-adaptive loss balanced physics-informed neural networks. Ocean. Eng. 309, 118341.
Huang, S.Y., An, W.J., Zhang, D.S., et al., 2023. Image classification and adversarial robustness analysis based on hybrid quantum–classical convolutional neural network. Opt. Commun. 533, 129287. https://doi.org/10.1016/J.OPTCOM.2023.129287.
Huang, S.K., Chao, W.T., Lin, Y.X., 2024. Conditional generation of artificial earthquake waveforms based on adversarial networks. Soil Dynam. Earthq. Eng. 180, 108622.
Imamverdiyev, Y., Sukhostat, L., 2019. Lithological facies classification using deep convolutional neural network. J. Petrol. Sci. Eng. 174, 216–228.
Izumi, T., Amagasaki, M., Ishida, K., et al., 2022. Super-resolution of sea surface temperature with convolutional neural network- and generative adversarial network-based methods. J. Water Clim. Change 13 (4), 1673–1683. https://doi.org/10.2166/WCC.2022.291.
Jia, C.Z., Pang, X.Q., Song, Y., 2023. Whole petroleum system and ordered distribution pattern of conventional and unconventional oil and gas reservoirs. Pet. Sci. 20 (1), 1–19.
Kim, S., Hong, Y.J., Lim, J.T., et al., 2023. Improved prediction of shale gas productivity in the Marcellus shale using geostatistically generated well-log data and ensemble machine learning. Comput. Geosci. 181, 105452.
Kreimeyer, K., Dang, O., Spiker, J., Munoz, A.M., Rosner, G., Ball, R., Botsis, T., 2021. Feature engineering and machine learning for causality assessment in pharmacovigilance: lessons learned from application to the FDA adverse event reporting system. Comput. Biol. Med. 135, 104517.
Kuntalp, M., Düzyel, O., 2024. A new method for GAN-based data augmentation for classes with distinct clusters. Expert Syst. Appl. 235, 121199. https://doi.org/10.1016/J.ESWA.2023.121199.
Lataifeh, Carrasco, X.A., Elnagar, A.M., et al., 2024. Human–machine co-creation: a complementary cognitive approach to creative character design process using GANs. J. Supercomput. 1–37.
Law, B.E., Dickinson, W.W., 1985. Conceptual-model for origin of abnormally pressured gas accumulation in low-permeability reservoirs. AAPG (Am. Assoc. Pet. Geol.) Bull. 69 (8), 1295–1304.
Lim, B., Yu, H., Yoon, D., Nam, M.J., 2021. Machine learning derived AVO analysis on marine 3D seismic data over gas reservoirs near South Korea. J. Petrol. Sci. Eng. 197, 108105. https://doi.org/10.1016/J.PETROL.2020.108105.
Liu, Y., Wang, X., Wu, K.Y., et al., 2019. Late Carboniferous seismic and volcanic record in the northwestern margin of the Junggar Basin: implication for the tectonic setting of the West Junggar. Gondwana Res. 71, 49–75.
Liu, Y.Y., Ma, X.H., Zhang, X.W., Guo, W., Kang, L.X., Yu, R.Z., et al., 2021a. A deep-learning-based prediction method of the estimated ultimate recovery (EUR) of shale gas wells. Pet. Sci. 18 (5), 1450–1464.
Liu, L., Chalaturnyk, R., Deisman, N., et al., 2021b. Anisotropic borehole response from pressuremeter testing in deep clay shale formations. Can. Geotech. J. 58 (8), 1159–1179.
Lu, X.S., Zhao, M.J., Zhang, F.Q., et al., 2022. Characteristics, origin and controlling effects on hydrocarbon accumulation of overpressure in foreland thrust belt of southern margin of Junggar Basin, NW China. Petrol. Explor. Dev. 49 (5), 991–1003.
Lyra, L.O., Fabris, A.E., Florindo, J.B., 2024. A multilevel pooling scheme in convolutional neural networks for texture image recognition. Appl. Soft Comput. 152, 111282. https://doi.org/10.1016/J.ASOC.2024.111282.
Ma, H.L., Jiang, L., Wang, Z., et al., 2024. Prediction and effectiveness study of fracture distribution based on an artificial neural network technology and production performance data detection method: a case study of Ordovician carbonate formation in the Tahe oilfield, Tarim Basin. Petrol. Sci. Technol. 11, 1–6.
Malakouti, S.M., Menhaj, M.B., Suratgar, A.A., 2023. The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. Clean. Eng. Technol. 15, 100064. https://doi.org/10.1016/J.CLET.2023.100664.
Merriam, E.R., Petty, J.T., Maloney, K.O., et al., 2018. Brook trout distributional response to unconventional oil and gas development: landscape context matters. Sci. Total Environ. 628, 338–349.
Mohammadjafari, S., Ozyegen, O., Cevik, M., et al., 2021. Designing mm-wave electromagnetic engineered surfaces using generative adversarial networks. Neural Comput. Appl. 33, 11309–11323. https://doi.org/10.1007/S00521-020-05656-2.
Mousavi, S.M., Gregory, C.B., 2023. Machine learning in earthquake seismology. Annu. Rev. Earth Planet Sci. 51, 105–129.
Osborne, M.J., Swarbrick, R.E., 1997. Mechanisms for generating overpressure in sedimentary basins: a reevaluation. AAPG (Am. Assoc. Pet. Geol.) Bull. 81, 1023–1041.
Pfeifer, A., Feijoo, F., Duić, N., 2023. Fast energy transition as a best strategy for all? The nash equilibrium of long-term energy planning strategies in coupled power markets. Energy 284, 129109. https://doi.org/10.1016/J.ENERGY.2023.129109.
Pradhan, B., 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput. Geosci. 51, 350–365.
Pu, Z., Koutti, L., Masmoudi, L., de Oliveira, J.V., 2024. A super resolution method based on generative adversarial networks with quantum feature enhancement: application to aerial agricultural images. Neurocomputing 577, 127346.
Puhachov, I., Neveu, W., Chien, E., et al., 2021. Keypoint-driven line drawing vectorization via PolyVector flow. ACM Trans. Graph. 40 (6), 1–17.
Puzyrev, V., Salles, T., Surma, G., et al., 2022. Geophysical model generation with generative adversarial networks. Geoscience Letters 9 (1), 32.
Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. Comp. Sci. https://doi.org/10.48550/arXiv.1511.06434.
Rafiei, A., Rad, M.G., Sikora, A., Kamaleswaran, R., 2024. Improving mixed-integer temporal modeling by generating synthetic data using conditional generative adversarial networks: a case study of fluid overload prediction in the intensive care unit. Comput. Biol. Med. 168, 107749. https://doi.org/10.1016/J.COMPBIOMED.2023.107749.
Rahmanifard, H., Alimohamadi, H., Gates, I., 2020. Well Performance Prediction in Montney Formation Using Machine Learning Approaches. SPE/AAPG/SEG Unconventional Resources Technology Conference.
2021 International energy outlook, 2021. US Energy information Administration, Washington.
2022 International energy outlook, 2022. US Energy information Administration, Washington.
Raissi, M, Perdikaris, P, Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378 (1), 686–707.
Rajabi, M.M., Komeilian, P., Wan, X., et al., 2023. Leak detection and localization in water distribution networks using conditional deep convolutional generative adversarial networks. Water Res. 238, 21–26. https://doi.org/10.1016/J.WATRES.2023.120012.
Reda, M., Fathy, M., Mosaad, M., Alshehri, F., Ahmed, M.S., 2024. 3D static modeling and sequence stratigraphy using well logs and seismic data: an example of Abu Roash G member in Bahga oilfield. Energy Geoscience, 100303.
Roy, J., Saha, S., 2022. Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K-fold cross validation approach. Artif. Intell. Geosci 3, 28–45. https://doi.org/10.1016/J.AIIG.2022.07.001.
Saengrung, A., Abtahi, A., Zilouchian, A., 2007. Neural network model for a commercial PEM fuel cell system. J. Power Sources 172 (2), 749–759. https://doi.org/10.1016/J.JPOWSOUR.2007.05.039.
Sangari, A., Sethares, W., 2016. Convergence analysis of two loss functions in soft-max regression. IEEE Trans. Signal Process. 64 (5), 1280–1288. https://doi.org/10.1109/TSP.2015.2504348.
Sapijaszko, G., Mikhael, W.B., 2018. An overview of recent convolutional neural network algorithms for image recognition. https://doi.org/10.1109/MWSCAS.2018.8623911.
Saud, S., Jamil, B., Upadhyay, Y., et al., 2020. Performance improvement of empirical models for estimation of global solar radiation in India: a k-fold cross-validation approach. Sustain. Energy Technol. Assessments 40, 100768. https://doi.org/10.1016/J.SETA.2020.100768.
Saxena, D., Cao, J.N., 2021. Generative adversarial networks (GANs) challenges, solutions, and future directions. ACM Comput. Surv. 54 (3), 1–42.
Semenov, A., Boginski, V., Pasiliao, E.L., 2019. Neural networks with multidimensional cross-entropy loss functions, 57-62. https://doi.org/10.1007/978-3-030-34980-6_5.
Shafee, A., Awaad, T.A., 2021. Privacy attacks against deep learning models and their countermeasures. J. Syst. Architect. 114, 101940. https://doi.org/10.1016/J.SYSARC.2020.101940.
Shehata, A.A., Osman, O.A., Nabawy, B.S., 2021. Neural network application to petrophysical and lithofacies analysis based on multi-scale data: an integrated study using conventional well log, core and borehole image data. J. Nat. Gas Sci. Eng. 93, 104015.
Shi, Y., Song, X.Z., Song, G.F., 2021. Productivity prediction of a multilateral-well geothermal system based on a long short-term memory and multi-layer perceptron combinational neural network. Appl. Energy 282, 116046.
Sim, E.-A., Lee, S., Oh, J., Lee, J., 2021. GANs and DCGANs for generation of topology optimization validation curve through clustering analysis. Adv. Eng. Software 152, 102957. https://doi.org/10.1016/J.ADVENGSOFT.2020.102957.
Skvortsov, V.A., 2020. Assessment of the oil and gas potential of the basement of the southern part of the siberian platform and deep-seated oil exploration. Dokl. Earth Sci. 492 (1), 302–305.
Snee, J.E.L., Zoback, M.D., 2022. State of stress in areas of active unconventional oil and gas development in North America. AAPG (Am. Assoc. Pet. Geol.) Bull. 106 (2), 106.
Song, Y., Li, Z., Jiang, Z.X., Luo, Q., Liu, D.D., Gao, Z.Y., 2017. Progress and development trend of unconventional oil and gas geological research. Petrol. Explor. Dev. 44 (4), 638–648.

Song, S., Mukerji, T., Hou, J., 2021. GANSim: conditional facies simulation using an improved progressive growing of generative adversarial networks (GANs). Math. Geosci. 53, 1413–1444. https://doi.org/10.1007/S11004-021-09934-0.

Song, S., Mukerji, T., Hou, J., et al., 2022a. GANSim-3D for conditional geomodeling: theory and field application. Water Resour. Res. 58 (7), 1–34. https://doi.org/10.1029/2021WR031865.

Song, S., Mukerji, T., Hou, J., 2022b. Bridging the gap between geophysics and geology with generative adversarial networks. IEEE Geosci. Remote Sens. Lett 99, 1–11. https://doi.org/10.1109/TGRS.2021.3066975.

Song, S., Zhang, D., Mukerji, T., Wang, N., 2023. GANSim-surrogate: an integrated framework for stochastic conditional geomodelling. J. Hydrol. 620, 129493. https://doi.org/10.1016/J.JHYDROL.2023.129493.

Sun, C., Li, C., Liu, Y., et al., 2019. Prediction method of concentricity and perpendicularity of aero engine multistage rotors based on PSO-BP neural network. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2941118.

Swarbrick, R.E., Osborne, M.J., 1998. Mechanisms that generate abnormal pressures: an overview. AAPG Memoir 70, 13–34.

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. Pet. Sci. 18 (5), 1281–1293.

Timm, N., Botha, J., Jordaan, S., 2023. Max-SAT-based synthesis of optimal and Nash equilibrium strategies for multi-agent systems. Sci. Comput. Program. 228, 102946. https://doi.org/10.1016/J.SCICO.2023.102946.

Tingay, M.R.P., Hillis, R.R., Swarbrick, R.E., Morley, C.K., Damit, A.R., 2009. Origin of overpressure and pore-pressure prediction in the Baram province, Brunei. AAPG Bulletin 93 (1), 51–74.

Ulfers, A., Zeeden, C., Wagner, B., et al., 2022. Borehole logging and seismic data from Lake Ohrid (North Macedonia/Albania) as a basis for age-depth modelling over the last one million years. Quat. Sci. Rev. 276, 107295. https://doi.org/10.1016/J.QUASCIREV.2021.107295.

Van Balen, R.T., Verweij, J.M., Van Wees, J.D., et al., 2022. Deep subsurface temperatures in the roer valley graben and the peelblock, the Netherlands - new results. Neth. J. Geosci. 81 (1), 19–26.

Verma, A., Meenpal, T., Acharya, B., 2024. Action-guided CycleGAN for Bi-directional video prediction. IETE Tech. Rev. https://doi.org/10.1080/02564602.2024.2327566.

Vrettos, K., Koltsakis, E., Zibis, A.H., et al., 2024. Generative adversarial networks for spine imaging: a critical review of current applications. Eur. J. Radiol. 171, 111313. https://doi.org/10.1016/J.EJRAD.2024.111313.

Wang, Q., Chen, D., Li, M., et al., 2023. A novel method for petroleum and natural gas resource potential evaluation and prediction by support vector machines (SVM). Appl. Energy 351, 121836.

Wang, J., Heng, J., Xiao, L., et al., 2017. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. Energy 125, 591–613.

Wei, Q., Lin, Y.T., Gao, G., et al., 2023. Research on Formation pressure prediction method for ultra-deep tight sandstone based on collocated cokriging. Processes 11 (7), 2010.

Xu, S.D., Chen, S.P., Zhao, H.B., et al., 2024. Basin-range coupling relationship and prediction of favorable targets for ultra-deep subsalt exploration in the Tarim Basin, Northwest China. J. Asian Earth Sci. 260, 105955.

Xue, P., Wang, H., Luo, T., et al., 2022. Clear sky color modeling based on BP neural network. Build. Environ. 226, 109715. https://doi.org/10.1016/J.BUILDENV.2022.109715.

Yang, F., Li, J.Z., Lu, S., et al., 2023. Carboniferous to early Permian tectono-sedimentary evolution in the western Junggar Basin, NW China: implication for the evolution of Junggar Ocean. Front. Earth Sci. 11, 1237367.

Ye, L., Peng, Y., Li, Y., Li, Z., 2024. A novel informer-time-series generative adversarial networks for day-ahead scenario generation of wind power. Appl. Energy 364, 123182.

Yilmaz, B., Korn, R., 2024. A comprehensive guide to generative adversarial networks (GANs) and application to individual electricity demand. Expert Syst. Appl., 123851

Zamo, M., Mestre, O., Arbogast, P., et al., 2014. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: deterministic forecast of hourly production. Energy 105, 792–803.

Zhang, H., Goodfellow, I., Metaxas, D., et al., 2019. Self-attention generative adversarial networks. In: International Conference on Machine Learning. PMLR, pp. 7354–7363.

Zhang, Z.Y., Zhu, G.Y., Han, J.F., et al., 2021. Genesis and preservation of the giant ultradeep Hadexun petroleum accumulation in the Tarim basin, China. J. Petrol. Sci. Eng. 208, 109249.

Zhang, F.Q., Gingras, M., Shan, C.D., et al., 2023a. The effect of coupling of tectonic compression and overpressure on porosity of deep reservoirs: a case study of southern margin of Junggar Basin, northwest China. Front. Earth Sci. 11, 1220105.

Zhang, K., Wang, W., Lv, Z., et al., 2023b. LKDPNet: Large-Kernel Depthwise-Pointwise convolution neural network in estimating coal ash content via data augmentation. Appl. Soft Comput., 110471 https://doi.org/10.1016/J.ASOC.2023.110471.

Zhang, Y., Ma, B.B., Jiang, S., et al., 2024a. Formation mechanisms of anomalously high reservoir quality in deep-buried volcaniclastic sandstones, central Junggar basin, northwestern China. Mar. Petrol. Geol. 27, 106772.

Zhang, W., Wu, C.D., Liu, S.B., et al., 2024b. Impact of disequilibrium compaction and unloading on overpressure in the southern Junggar Foreland Basin, NW China. Mar. Petrol. Geol., 106819

Zhi, D.M., Liu, M.X., Chen, X.W., et al., 2022. Stratigraphic characteristics and sediment-filling process of the early Permian fengcheng formation in the Northwestern Margin of the Junggar Basin, Northwest China. Front. Earth Sci. 8 (10), 945563.

Zhu, D.Y., Liu, Q.Y., He, Z.L., et al., 2020. Early development and late preservation of porosity linked to presence of hydrocarbons in Precambrian microbialite gas reservoirs within the Sichuan Basin, southern China. Precambr. Res. 342, 105694.

Zou, C.N., Qiu, Z., Zhang, J.Q., Li, Z.Y., Wei, H.Y., et al., 2022. Unconventional petroleum sedimentology: a key to understanding unconventional hydrocarbon accumulation. Engineering 18 (11), 62–78.

Zrelli, A., Amiri, A., Sebei, K., et al., 2024. Sequence stratigraphy of the paleozoic successions in Ghadames Basin, southern Tunisia: insights from integrated interpretation of well logs and seismic data. J. Afr. Earth Sci. 213, 105218.