

# Physics-integrated neural networks for improved mineral volumes and porosity estimation from geophysical well logs



Prasad Pothana <sup>\*</sup>, Kegang Ling

Department of Energy and Petroleum Engineering, College of Engineering and Mines, University of North Dakota, Grand Forks, 58202, ND, USA

## ARTICLE INFO

### Article history:

Received 26 November 2024

Received in revised form

25 January 2025

Accepted 3 April 2025

Handling editor: Bo Ren

### Keywords:

Physics integrated neural networks

Petrophysics

Well logs

Oil and gas

Reservoir characterization

Mineralogy

Machine learning

## ABSTRACT

Accurate estimation of mineralogy from geophysical well logs is crucial for characterizing geological formations, particularly in hydrocarbon exploration, CO<sub>2</sub> sequestration, and geothermal energy development. Current techniques, such as multimineral petrophysical analysis, offer details into mineralogical distribution. However, it is inherently time-intensive and demands substantial geological expertise for accurate model evaluation. Furthermore, traditional machine learning techniques often struggle to predict mineralogy accurately and sometimes produce estimations that violate fundamental physical principles. To address this, we present a new approach using Physics-Integrated Neural Networks (PINNs), that combines data-driven learning with domain-specific physical constraints, embedding petrophysical relationships directly into the neural network architecture. This approach enforces that predictions adhere to physical laws. The methodology is applied to the Broom Creek Deep Saline aquifer, a CO<sub>2</sub> sequestration site in the Williston Basin, to predict the volumes of key mineral constituents—quartz, dolomite, feldspar, anhydrite, illite—along with porosity. Compared to traditional artificial neural networks (ANN), the PINN approach demonstrates higher accuracy and better generalizability, significantly enhancing predictive performance on unseen well datasets. The average mean error across the three blind wells is 0.123 for ANN and 0.042 for PINN, highlighting the superior accuracy of the PINN approach. This method reduces uncertainties in reservoir characterization by improving the reliability of mineralogy and porosity predictions, providing a more robust tool for decision-making in various subsurface geoscience applications.

© 2025 Petroleum Exploration and Production Research Institute Corporation, SINOPEC. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The accurate estimation of petrophysical properties, such as mineralogy, porosity, permeability, and fluid saturation, plays an important role in minimizing uncertainty during reservoir characterization (Guillon et al., 2022; Lima et al., 2022). Mineralogical distribution in the formation helps accurately characterize the lithological variations in the formation that impact the storage and flow capacity (Pothana et al., 2023, 2024; Burke et al., 2022). Additionally, having detailed mineralogical information helps in rock physics modeling, gaining insights into diagenesis and depositional environments (Burke et al., 2022; Sayers and Dasgupta, 2024). During geothermal energy development, hot fluids can

transform the existing minerals in the host rock due to chemical and physical reactions (D'amore and Panichi, 1985). For example, the hydrothermal alteration process produces silica residue and creates minerals like kaolinite, sulfur, sulfides, and aluminous sulfates (Rodgers et al., 2002). During permanent CO<sub>2</sub> storage, the presence of highly reactive minerals, for example, basalt formations, is favorable for in-situ mineralization, also known as mineral trapping (Kim et al., 2023). Hence, the evaluation of the mineral composition of subsurface targets holds significant importance.

The assessment of petrophysical properties through the utilization of geophysical well logs is a standard procedure. This practice is widely applied across various geoscience fields, including both conventional and unconventional petroleum system characterization, as well as the evaluation of reservoirs and seals for CO<sub>2</sub> sequestration, and the assessment of formations for geothermal energy development (Jacobi et al., 2008; Metwally et al., 2023; Sanchez and Ofori, 2013; Bhattacharya et al., 2023). Mineral volumes are estimated by an inverting system of equations that relates

\* Corresponding author.

E-mail address: [prasad.pothana@und.edu](mailto:prasad.pothana@und.edu) (P. Pothana).

Peer review under the responsibility of Editorial Board of Energy Geoscience.

geophysical log measurements and the formation of mineral and fluid volume fractions and their responses. Besides water saturation equations, the majority of the equations assume linear relationships between the measurements and model parameters (Doveton, 2014). Specifically, the optimization process involves logging tool measurements  $\mathbf{X}$  and inverting a response matrix ( $\mathbf{R}$ ) to estimate mineral and fluid volume fractions  $\mathbf{Y}$ . This can be mathematically represented as follows (see, e.g., Doveton (2014)):

$$\mathbf{Y} = \mathbf{R}^{-1} \cdot \mathbf{X} + \epsilon \quad (1)$$

In this equation,  $\epsilon$  indicates the noise parameter matrix. It is to be noted that the straightforward inversion of the response matrix ( $\mathbf{R}$ ) is impractical because of the matrix's complexity and potential ill-conditioning. In practice, the parameters, also known as endpoints, are iteratively fine-tuned, and the well-logs are computed through a forward modeling approach. This iterative procedure is continued until the error between the predicted well-logs and measured well-logs is less than the user-defined tolerance.

During the above process, typically, two critical issues are encountered: (i) the adjustment of the endpoints in the response matrix to achieve satisfactory predictions is very time-consuming, and (ii) the optimal mathematical solution may not represent the underlying geological reality due to the assumption used in the system of equations (see Equation (1)). Therefore, expertise in regional subsurface geology is essential to correctly assessing the validity of the optimum model. The presence of these two challenges motivates us to develop an automated process that will not only accelerate the process but also enable us to predict mineralogical distributions that closely resemble true geological mineralogy.

Machine learning (ML) techniques have recently become popular for pattern recognition and prediction in many fields, including geoscience. ML techniques are particularly well suited for capturing complex, non-linear relations between input data—for example, well logs—and target variables such as permeability and rock classification, among others. A recent, though nonexhaustive, overview of publications that utilized ML in the petrophysics domain is presented in Table 1. These show the broad range of applications of ML in petrophysics, including synthetic log generation, permeability prediction, rock type classification, and real-time decision-making during logging while drilling (LWD). A comprehensive review of machine learning applications in petrophysics can be found in Xu et al. (2019, 2022) and Syed et al. (2022).

To the best of our knowledge, only a limited number of

researchers have attempted to predict mineral volumes using machine learning (ML) from conventional geophysical logs. Laalam et al. (2022) evaluated various ML algorithms (LR, RFR, XGB, SVR, kNN and ETR) to predict mineral volumes from well logs. However, their study presents major limitations. Notably, the best-performing model, RFR, occasionally predicted negative mineral volumes, which violates fundamental physical principles. Bettir et al. (2024) attempted to solve the underdetermined problem, where the number of minerals and fluids is more than the number of well-logs, using RF, SVM, AdaBoost, and ANN algorithms. However, their study lacks robustness, as the model was not validated on a fully independent dataset but rather on a partitioned subset of the original training data, leaving questions about its generalizability. Consequently, ML models often function as "black boxes", offering limited interpretability and sometimes yielding results that are not physically consistent. These limitations highlight the need for approaches that not only use the predictive power of ML but also ensure physical consistency and interpretability.

Integrating physical constraints ensures predictions align with established physical principles while benefiting from automation and efficiency of ML models. By combining data-driven learning with prior knowledge, Physics-Integrated Neural Networks (PINNs) offer the potential for more accurate, reliable, and physically consistent estimates of mineralogy and porosity. Recently, physics-guided machine learning techniques have gained popularity in geoscience and petroleum engineering fields (Schuster et al., 2024; Sun et al., 2022). In the field of petrophysics, for instance, Shao et al. (2024) introduced a Petrophysics-Informed Machine Learning (PIML) method that integrates petrophysical data into neural networks to improve reservoir evaluation. Similarly, Lv et al. (2021) predicted permeability using well log data, demonstrating a broader range of accurate permeability predictions through physics-informed methods. Additionally, Zhao et al. (2024) investigated a physics-guided ML approach to predict shear sonic data based on well logs. However, their utilization of PINN in petrophysics remains relatively unexplored, particularly in the context of predicting mineral volumes from geophysical well logs.

This research investigates the effectiveness of Physics-Integrated Neural Networks (PINNs) in predicting mineralogy and porosity using conventional geophysical well logs while addressing gaps in previous methods. PINNs integrate domain-specific knowledge, such as petrophysical relationships and geological constraints, directly into the neural network architecture. The study compares the performance of PINNs with traditional Artificial

**Table 1**  
Recent literature on applications of machine learning in petrophysics.

Reference	Objective	ML Models Used
Mukherjee et al. (2024)	Petrographic Classification From Geophysical Logs	kNN,SVM, DT, RF, XG Boost and ANN
Iraji et al. (2023)	Rock type prediction and petrophysical characterization using multi-scale data	ResNet and 1D CNN
Burak et al. (2024)	Real-time prediction of formation lithology and tops using drilling parameters.	kNN,SVM, RF and ANN
Radwan et al. (2022)	Pore pressure prediction using well logs	RF, SVM, ANN, ADA and DT
Silva et al. (2020)	Classification of Petrofacies	DT, RF, Gradient Boosting, kNN, naive Bayes
Pan et al. (2023)	Multi-well petrophysical evaluation using well logs	DANN
Mohammadian et al. (2022)	Rock typing and permeability prediction	Extreme Gradient Boosting
Ali et al. (2021)	Shear sonic log prediction	DNN
Kim (2022)	Shear sonic log prediction	SVR-PSO, DNN-PSO, and LSTM-PSO
Chaikine and Gates (2020)	Shear sonic log prediction	Hybrid cRNN
Jiang et al. (2022)	Shear sonic log prediction	XGBoost-DFNN
Hussain et al. (2022)	Rock type classification, Lithofacies analysis	SOM
Houshamand et al. (2022)	Rock type classification, Lithofacies analysis	SVM, DT, kNN, NB, LR, XGBoost, Res-Net and cRNN
Jiang et al. (2021)	Rock type classification, Lithofacies analysis	RNN, BI-GRU and 1D CNN
Ifrene et al. (2023)	Fracture characterization	ANN and SVM
Matinkia et al. (2023)	Permeability Prediction	SSD-MLP
Wood (2020)	Porosity, permeability and water saturation Prediction	TOB
Feng et al. (2024)	Permeability Prediction	LGB-WOA

Neural Networks (ANNs), emphasizing the advantages of incorporating physical constraints into the machine learning framework. The findings contribute to advancements in reservoir characterization by reducing manual effort and uncertainty, thereby demonstrating the potential of PINNs for improving decision-making in subsurface geoscience applications, particularly hydrocarbon exploration, CO<sub>2</sub> storage and geothermal energy development.

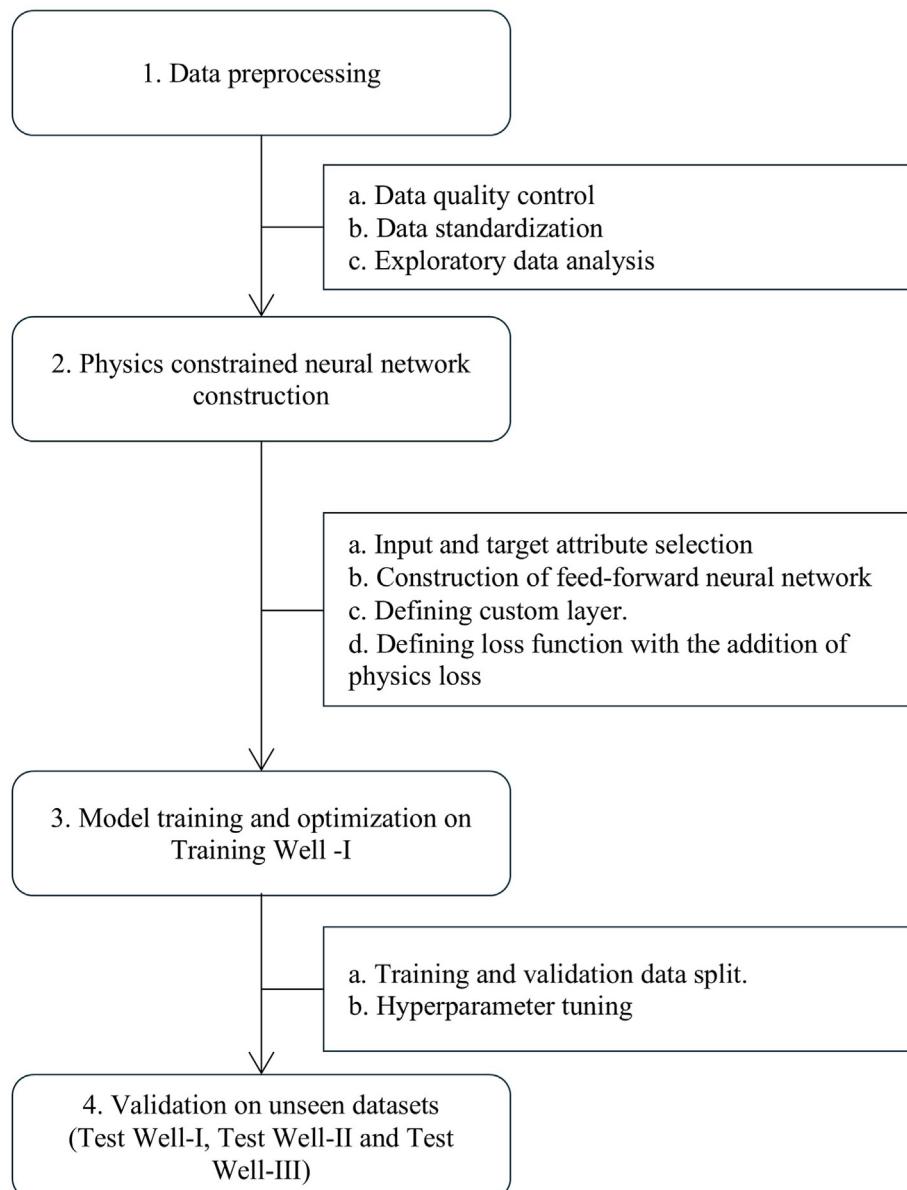
## 2. Methodology

### 2.1. Workflow

The workflow involves a sequence of stages (Fig. 1), beginning with comprehensive data preprocessing, which ensures data quality and prepares it for model training. Following this, input and target attributes are carefully selected based on domain-specific knowledge and exploratory data analysis, ensuring that both are

well-aligned with the physical processes being modeled. Feed-forward neural networks are employed with a detailed selection of hyper-parameters, including the number of layers, units per layer, activation functions, and regularization strategies. A critical component of this workflow is the definition and integration of physics-based constraints into the neural network architecture.

Before examining the detailed methodology, it is worth mentioning that we deliberately used the phrase “Physics Integrated Neural Network (PINN)” to distinguish our approach from the more specific “Physics Informed Neural Network”. While physics-informed neural networks typically incorporate known physical laws, generally expressed as partial differential equations (PDEs) or ordinary differential equations (ODEs), directly into the training process (Raissi et al., 2019), our approach is more general and hybrid. We utilize both strict physical bounds implemented as custom layers within the network architecture and an additional physics-informed loss function in the form of non-linear equations. Physical principles are embedded during training by incorporating



**Fig. 1.** Workflow used in the present study.

geological constraints and governing equations directly into the model, ensuring that the network respects the underlying physical laws of the system. This physics-integrated approach is maintained throughout training, and the model is subsequently tested on unseen datasets to evaluate its generalization performance and adherence to both patterns based on data and physical consistency. The following sections present a detailed step-by-step procedure for each of these stages.

## 2.2. Data preprocessing

### 2.2.1. Data quality control

Raw geophysical well log data undergoes preprocessing to ensure its suitability for modeling. Geophysical logs represent physical measurements taken at regular depth intervals along a drilled wellbore. Therefore, it is necessary to ensure that all the measurements are depth synchronized. The first step in this regard is depth shifting, which addresses discrepancies that arise when data is collected over multiple runs using different logging tools. These discrepancies manifest as misalignment in depth between datasets, making it necessary to perform depth correction. This correction ensures that all recorded measurements correspond to the correct depth, thereby improving data consistency.

Well logs, particularly those generated by tools with shallow depths of investigation, are often influenced by borehole rugosity and washouts. In these cases, the log responses may not accurately reflect the true geological features. A careful examination of the logging measurements in conjunction with the borehole caliper log is necessary to identify and correct any distortions caused by irregularities in the borehole. Additional preprocessing steps include data cleaning and filtering to remove noise, managing missing values, and normalizing or standardizing input features to enhance model stability during training. Readers are encouraged to consult McDonald (2021), which provides an elegant summary of various geophysical well-log data quality issues, their impact on machine learning algorithms, and strategies for mitigating these effects.

### 2.2.2. Data standardization

The input features  $X$ , consisting of well-log data, are standardized using the *StandardScaler* from the *sklearn* library (Pedregosa et al. (2011)). The standardization process transforms the input data to have zero mean and unit variance using the following equation:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma} \quad (2)$$

where  $X$  is the original input feature;  $\mu$  is the mean of the feature; and  $\sigma$  is the standard deviation of the feature. The scaling is performed to ensure that each feature contributes equally to the training process, thereby avoiding biases due to the different scales of input features. This is particularly relevant when dealing with well-log data, where the range of values can differ significantly across different datasets. For example, porosity values typically range between 0 and 1 (v/v), gamma ray measurements range from 0 to several hundred GAPI, and resistivity can vary between 0.01 and several thousand ohm·m.

### 2.2.3. Exploratory data analysis

In this section, the correlations between input well-log data and target attributes are analyzed to identify significant features for prediction. By examining these relationships, we can determine which variables hold the most predictive power. Additionally, the distribution of both input and target attributes is evaluated to assess any skewness, as this can influence model performance.

## 2.3. Physics integrated neural network construction

The neural network model is designed to predict the quantities of the mineral volume fractions and porosity. The input to the model is the set of normalized well-log measurements  $X_{\text{scaled}}$ , and the output is the predicted mineral quantities  $Y$ . The model architecture consists of a set of fully connected layers (Dense layers), where the number of layers and the number of units in each layer are hyperparameters to be tuned (Fig. 2). The following operation can describe each dense layer:

$$\mathbf{z} = \mathbf{W}\mathbf{x} + \mathbf{b} \quad (3)$$

where  $\mathbf{W}$  is the weight matrix;  $\mathbf{x}$  is the input to the layer; and  $\mathbf{b}$  is the bias vector. The output of the layer is passed through a non-linear activation function, typically the ReLU (Rectified Linear Unit):

$$\text{ReLU}(\mathbf{x}) = \max(0, \mathbf{x}) \quad (4)$$

After applying the ReLU activation function, the training process continues by propagating the transformed output through the remaining layers of the neural network. ReLU introduces non-linearity to the model, allowing it to capture complex relationships between input well-log measurements and the target mineral quantities. Following the ReLU transformation, the updated values (activations) are passed to subsequent dense layers, where similar operations of linear transformation and non-linear activation are applied. Only the pertinent details are provided here. The construction of neural networks is a well-established area of research; for a more rigorous mathematical treatment, readers are encouraged to consult Gurney's foundational work (Gurney, 1997). Practical guidance on implementing neural networks using TensorFlow can be found in TensorFlow (2024) and Abadi et al. (2015).

In addition to the standard fully connected layers, the model incorporates physical constraints as a custom layer (Fig. 2). This custom layer encodes domain-specific knowledge, ensuring that the predicted mineral quantities and porosity adhere to known physical laws or geological constraints. During the training process, the network's parameters—the weight matrices  $\mathbf{W}$  and bias vectors  $\mathbf{b}$  are updated using backpropagation. However, unlike traditional training approaches, the loss function in this model has been modified to account for the physical constraints imposed by the custom layer and additional physical loss function to typical mean squared error (MSE) loss (Fig. 2); these are elaborated in detail in the following sections.

## 2.4. Physical constraints

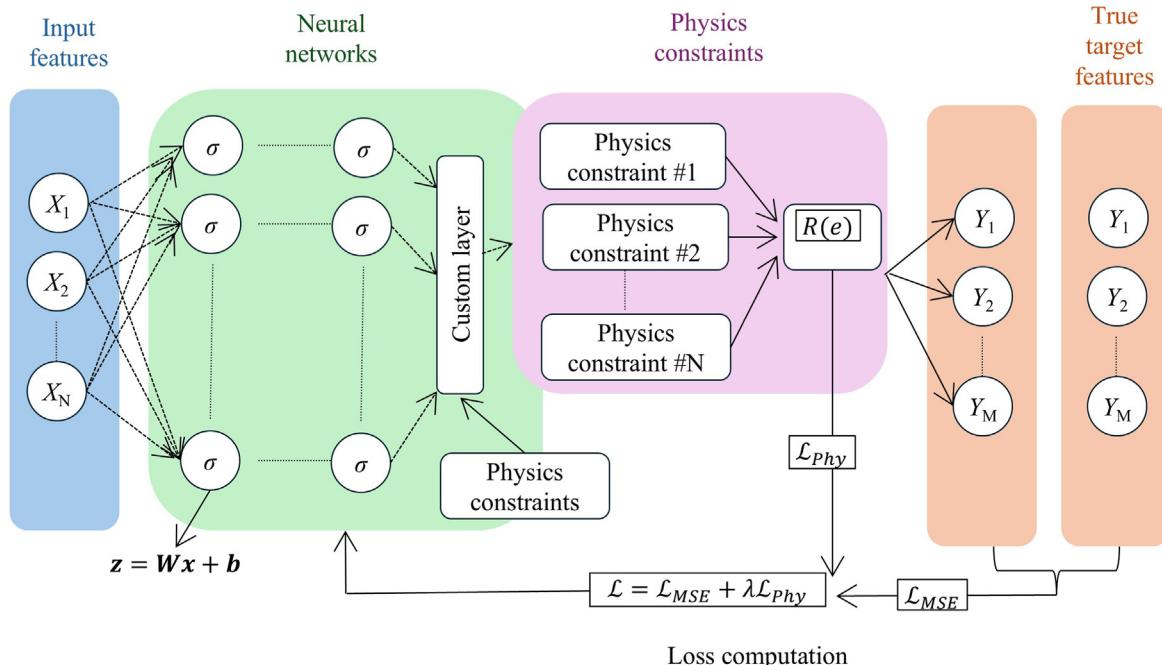
By integrating physical constraints directly into both the architecture and the training process, we aim to enhance the model's performance and adherence to physical laws. The following subsections describe various geological and petrophysical constraints that can be incorporated into the network.

### 2.4.1. Volume fraction boundaries

The volume fraction of each mineral within a rock formation is physically bound between 0 and 1. This ensures that the neural network does not predict unphysical mineral fractions that are either negative or exceed the total volume of the rock. Mathematically, this is represented as (see e.g. (Cheng et al., 2022),):

$$0 \leq V_i \leq 1 \quad \forall i \in M \quad (5)$$

where  $V_i$  is the volume fraction of mineral  $i$  and  $M$  is the set of all minerals present in the formation.



**Fig. 2.** Schematic of a Physics-Integrated Neural Network (PINN). It combines data-driven learning with physics-based constraints by minimizing a total loss comprising prediction error and physics residuals.

#### 2.4.2. Total mineral volume and porosity

The sum of all mineral volume fractions must be less than the total solid volume fraction of the rock, which is equal to the total volume minus the porosity ( $\phi$ ). This constraint reflects the principle that porosity represents the fraction of the rock that is pore space, and the rest is composed of various minerals. This is expressed as:

$$\sum_{i=1}^n V_i = 1 - \phi \quad (6)$$

where  $n$  is the number of mineral components;  $V_i$  is the volume fraction of mineral  $i$ ; and  $\phi$  is the porosity of the formation. This equation ensures that the network adheres to mass balance principles, making predictions of total rock volume realistic in terms of mineral content.

#### 2.4.3. Maximum porosity limit

Porosity in a rock formation is naturally constrained by geological factors such as compaction and cementation. Therefore, the network must ensure that the predicted porosity ( $\phi$ ) is always less than the maximum possible porosity ( $\phi_{\max}$ ), which is determined from regional geological knowledge and core sample analysis:

$$\phi < \phi_{\max} \quad (7)$$

where  $\phi_{\max}$  is the maximum porosity achievable under the given geological conditions.

#### 2.4.4. Complete saturation of pores

Reservoir rocks are typically saturated with various fluid constituents such as water, oil, and gas. It is essential to impose the physical constraint that the pore volume is entirely occupied by these fluids. This ensures that the sum of the volume fractions of all fluids equals the total porosity of the rock. Mathematically, this constraint can be expressed as:

$$V_W + V_O + V_G = \phi \quad (8)$$

where  $V_W$ ,  $V_O$ , and  $V_G$  denote the volume fractions of water, oil, and gas within the rock's pore space, respectively, and  $\phi$  represents the total porosity of the reservoir rock. Based on prior reservoir information, if any of the fluid constituents are absent, the corresponding terms can be omitted from the equation.

#### 2.4.5. Water saturation and resistivity model

In many formations, porosity and water saturation are related to resistivity through models like Archie's law, which describes the relationship between formation resistivity, water saturation, and porosity. In formations with low clay content, Archie's equation is commonly used (Archie, 1942; Doveton, 2014):

$$S_w^n - \frac{a}{\phi^m} \cdot \frac{R_w}{R_t} = 0 \quad (9)$$

where  $S_w$  is the water saturation;  $n$  is the saturation exponent;  $a$  is the tortuosity factor;  $m$  is the cementation factor;  $R_w$  is the formation water resistivity; and  $R_t$  is the true formation resistivity measured by logging tools. In clay-rich formations, Archie's equation becomes insufficient because clay conductivity cannot be ignored. In such cases, more complex models like the Indonesian Equation (Poupon and Leveaux, 1971), the Waxman-Smits Model (Waxman and Smits, 1968), or any other shaly-sand model can be used (Clavier et al., 1984; Gupta et al., 2023; Simandoux, 1963). These equations account for the excess conductivity contributed by the presence of clay minerals in addition to the formation of water. For instance, the Indonesian equation can be written as (Poupon and Leveaux, 1971):

$$\frac{1}{\sqrt{R_t}} = \left[ \frac{V_{sh}^{(1-V_{sh}/2)}}{\sqrt{R_{sh}}} + \frac{\phi^{m/2}}{\sqrt{aR_w}} \right] S_w^{n/2} \quad (10)$$

where  $V_{sh}$  is the shale volume fraction;  $R_{sh}$  is the resistivity of

the shale; and  $\phi$  is the porosity. These equations enforce realistic relationships between resistivity, porosity, and water saturation in the network, improving the physical relevance of predictions.

#### 2.4.6. Gamma ray and radio active minerals

Gamma-ray (GR) logs measure the natural radioactivity of the formation and measure Thorium, Potassium, and Uranium, which predominantly arise from clay and other radioactive minerals such as K-feldspar minerals (Katahara, 1995). Therefore, a constraint can be imposed based on the gamma-ray log to relate it to the clay and radioactive mineral volume fractions. This can be expressed as:

$$V_{\text{Clay}} + V_{\text{RA}} - k \cdot \frac{GR - GR_{\min}}{GR_{\max} - GR_{\min}} = 0 \quad (11)$$

where  $V_{\text{Clay}}$  and  $V_{\text{RA}}$  are the volume fractions of clay and radioactive minerals, such as feldspar, respectively;  $GR_{\min}$  and  $GR_{\max}$  are the minimum and maximum gamma ray values for the formation; and  $k$  is a scaling constant.

#### 2.4.7. Porosity relationships

In traditional petrophysical analysis, the total porosity of a reservoir rock is commonly estimated using porosity logs such as bulk density logs, neutron porosity logs, and sonic logs. These logs provide measurements that can be utilized to calculate porosity through specific equations. The porosity derived from the density log ( $\phi_D$ ) is calculated using (see e.g. (Asquith and Krygowski, 2006),):

$$\phi_D = \frac{\rho_{\text{ma}} - \rho_b}{\rho_{\text{ma}} - \rho_f} \quad (12)$$

where  $\rho_{\text{ma}}$  denotes matrix density ( $\text{g}/\text{cm}^3$ ), the density of the rock grains;  $\rho_b$  is bulk density ( $\text{g}/\text{cm}^3$ ) measured by the density log; and  $\rho_f$  represents fluid density ( $\text{g}/\text{cm}^3$ ), the density of the fluids within the pore spaces.

An integrated porosity using neutron and density logs can be estimated using the equation (see e.g. (Asquith and Krygowski, 2006),):

$$\phi_{\text{ND}} = \sqrt{\phi_D^2 + \phi_N^2} \quad (13)$$

where  $\phi_{\text{ND}}$  is porosity from combined density-neutron logs;  $\phi_D$  represents density-derived porosity; and  $\phi_N$  denotes neutron-derived porosity.

The sonic log measures the interval transit time ( $\Delta t$ ) of acoustic waves traveling through the formation. The porosity derived from the sonic log ( $\phi_S$ ) can be calculated using the Wyllie Time-Average Equation (Wyllie et al., 1956) (see e.g. (Asquith and Krygowski, 2006),):

$$\phi_S = \frac{\Delta t_{\log} - \Delta t_{\text{ma}}}{\Delta t_{\text{fl}} - \Delta t_{\text{ma}}} \quad (14)$$

where,  $\phi_S$  is sonic-derived porosity;  $\Delta t_{\log}$  denotes interval transit time ( $\mu\text{s}/\text{ft}$ ) recorded by the sonic log;  $\Delta t_{\text{ma}}$  is matrix transit time ( $\mu\text{s}/\text{ft}$ ), characteristic of the rock matrix; and  $\Delta t_{\text{fl}}$  denotes fluid transit time ( $\mu\text{s}/\text{ft}$ ), characteristic of the pore fluids.

However, a significant challenge in applying these equations is that the choice of matrix properties ( $\rho_{\text{ma}}$  and  $\Delta t_{\text{ma}}$ ) must accurately match the actual lithology of the formation to obtain reliable results. Therefore, the selection and application of these equations should be guided by petrophysical expertise and tailored to the specific geological context of the study area.

By embedding these physical constraints into the neural network architecture, the model is forced to generate predictions

that not only fit the data but also respect fundamental physical principles. This leads to more reliable predictions in the context of mineral volume fraction and porosity estimation. However, the choice of model (e.g., Archie's vs. Indonesian) must consider the geological context, and the sensitivity to parameters must be carefully managed to avoid errors.

#### 2.4.8. Incorporating uncertainty in physical constraints

In the above subsections, equations (5)–(8) are strictly enforced by the model, whereas equations (9)–(14) inherently hold some degree of uncertainty (Moore et al., 2011). This uncertainty arises from several factors, including errors or limitations in geophysical log measurements, which may stem from tool calibration, environmental conditions, or borehole irregularities. Furthermore, the input parameters used in these equations, such as Archie's parameters ( $a$ ,  $m$ , and  $n$ ), are often derived empirically and can vary significantly depending on lithology, pore geometry, and fluid properties (Archie, 1942; Doveton, 2014). Additionally, the assumptions used in the derivation of petrophysical equations introduce uncertainty. For instance, the assumption that only fluid contributes to the electrical conductivity of rocks may not hold true in cases where clays or other conductive minerals are present. Similarly, the empirical nature of many models adds another layer of imprecision. An example is the Wyllie Time-Average Equation (Wyllie et al., 1956), which employs a simplified, linearized relationship between the volume fractions of fluids and the rock matrix on acoustic travel time, ignoring more complex non-linear effects and rock fabric influences.

In this work, we incorporate this uncertainty explicitly into the computation of the physics loss using a variance term ( $\sigma^2$ ) to reflect our uncertainty about the accuracy of the underlying physical laws. By doing so, we adaptively model the physics residual as being distributed with some uncertainty, which is encoded through a normal distribution assumption.

For a given physical residual  $R(\mathbf{x}, \mathbf{y})$ , which depends on the input variables  $\mathbf{x}$  and the predicted outputs  $\mathbf{y}$ , we define the physics loss term under the assumption that the residuals are normally distributed. In this case, the residual is given by:

$$R(\mathbf{x}, \mathbf{y}) = \hat{y}(\mathbf{x}) - f(\mathbf{x}, \mathbf{y}) \quad (15)$$

where  $\hat{y}(\mathbf{x})$  represents the predicted outcome, and  $f(\mathbf{x}, \mathbf{y})$  represents the governing physical equation. The introduction of a variance parameter,  $\sigma^2$ , to account for uncertainty in this relationship modifies the loss function to the following form:

$$\mathcal{L}_{\text{physics}} = \frac{1}{2\sigma^2} E[R(\mathbf{x}, \mathbf{y})^2] \quad (16)$$

$\mathcal{L}_{\text{physics}}$  denotes physical loss function.  $E$  stands for the expectation or expected value of the random variable, which in this case is the squared residual  $R(\mathbf{x}, \mathbf{y})^2$ . The physics residuals, in this case, are normalized by the uncertainty term  $\sigma^2$ , which allows the model to account for imperfect knowledge of the underlying physical process. The expectation term is approximated via the empirical mean over the batch of data points. The parameter  $\sigma$  is adjusted based on the level of uncertainty inherent in the physical problem being modeled.

#### 2.4.9. Hyperparameter tuning

The model's hyperparameters, such as the number of layers, number of units in each layer, learning rate, and the weight  $\lambda_{\text{physics}}$  for the physics-informed loss, are tuned using the Keras Tuner library (O'Malley et al., 2019). The RandomSearch method is employed to explore a range of possible hyperparameter configurations. The objective of the tuning process is to minimize the

validation final loss function ( $\mathcal{L}$ ). The final loss function for the model is:

$$\mathcal{L} = \mathcal{L}_{MSE} + \lambda_{\text{physics}} \cdot \mathcal{L}_{\text{physics}} \quad (17)$$

where  $\mathcal{L}_{MSE}$  denotes mean squared error loss function.  $\mathcal{L}_{\text{physics}}$  represents physical loss function.

The mean absolute error (MAE) metric, used for evaluating the model, is calculated as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_{\text{true},i} - y_{\text{pred},i}| \quad (18)$$

**Custom Output Layer:** The output layer of the neural network is designed to handle the prediction of porosity separately from the other mineral components. This is achieved by introducing a *CustomOutputLayer*, which ensures that the porosity is limited to the range  $[0, \phi_{\max}]$ , while the sum of the other mineral predictions is constrained to 1 minus the predicted porosity.

Let  $y_{\text{mnrl}}$  represent the vector of predicted mineral quantities and  $y_{\text{por}}$  represent the predicted porosity. The output layer performs the following operations:

$$y_{\text{por}} = \phi_{\max} \sigma(\mathbf{W}_{\text{por}} \mathbf{x}) \quad (19)$$

where  $\sigma(x)$  is the sigmoid function (see e.g., (Chen et al., 2024)):

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (20)$$

This ensures that the predicted porosity is within the desired range. The mineral predictions are computed using a softmax function:

$$y_{\text{mnrl}} = \text{softmax}(\mathbf{W}_{\text{mnrl}} \mathbf{x}) \quad (21)$$

where the softmax function is defined as (see e.g., (Ren and Wang, 2023)):

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (22)$$

Finally, the total output of the layer is:

$$y = (1 - y_{\text{por}}) \cdot y_{\text{mnrl}} + y_{\text{por}} \quad (23)$$

This ensures that the sum of all mineral quantities plus porosity is equal to 1.

### 3. Application to field data

#### 3.1. Study area and geological background

The proposed methodology is applied to the Broom Creek Formation within the Williston Basin, located primarily in North Dakota. The Broom Creek Formation is of significant interest due to its status as a key site for the commercial-scale deployment of a dedicated CO<sub>2</sub> storage project (Peck et al., 2019). This project has long-term goals of developing a fully permitted geologic storage facility, contributing to carbon sequestration efforts (Steadman et al., 2006; Peck et al., 2019). The formation is part of the Permian-age Minnelusa Group and exhibits excellent reservoir properties conducive to geological CO<sub>2</sub> storage, including high porosity and permeability (Sorensen et al., 2009b). Moreover, the formation is capped by effective seals both above (Opeche formation) and below (Amsden formation), ensuring containment of

injected CO<sub>2</sub> and reducing the risk of leakage (Sorensen et al., 2009a).

The Williston Basin itself is geologically extensive, spanning across North Dakota, South Dakota, Montana, and parts of Canada. This basin holds significant economic and environmental importance, as it is rich in oil and gas resources, substantial lignite coal deposits, and valuable potash resources (Gerhard et al., 1982). As such, the Broom Creek Formation within the Williston Basin represents a promising site for advancing geological carbon storage technologies, particularly in regions associated with high levels of hydrocarbon production and the presence of coal power plants, where CO<sub>2</sub> emissions are a pressing concern (Peck et al., 2019). Monitoring mineralogical changes resulting from reactions with CO<sub>2</sub> during long-term CO<sub>2</sub> storage is essential.

The Broom Creek Formation predominantly consists of clean sandstone (with a reddish-brown to pink quartzarenite composition) interspersed with layers of dolomite, anhydrite, and fine-grained silt or clay (Pothana et al. (2024)). As depth increases, dolomite and anhydrite become more prevalent, while the upper sections are characterized by cleaner sandstone. Geological assessments indicate that the sandstone is generally well-sorted and subrounded, with occasional inclusions of anhydrite and dolomitic cement. In various intervals, dolomite and dolomitic sandstone, often containing anhydrite and clay, occur adjacent to sandstone layers. Although the clean sandstone possesses high porosity and permeability, the introduction of dolomite and anhydrite leads to a noticeable reduction in these reservoir qualities.

#### 3.2. Data

The data consists of well-log data and mineralogical distribution results from four wells. One well is used as a training well (Training Well-I), and the remaining three wells are used as blind wells, Test Well-I, Test Well-II, and Test Well-III, to validate the results. The input attributes and target attributes are tabulated in Table 2. It is to be noted that all the output features are simultaneously predicted during the training process. As detailed in the previous Section 2.2.1, depth correction has been applied to all individual logs to ensure proper alignment at the appropriate depths. Additionally, a thorough analysis of the borehole caliper was conducted for the observed section. The analysis revealed no significant borehole rugosity or washouts; therefore, no manual corrections were deemed necessary for the logs. The Training Well-I dataset consists of 519 data points. A statistical summary of both input and target attributes is presented in Table 3, along with visual representations in Fig. 3. It is evident that quartz is the dominant mineral, followed by dolomite, K-feldspar, anhydrite, and illite. The mean porosity is 0.18 (v/v), with a maximum value of 0.33 (v/v).

The distributions of illite, anhydrite, and K-feldspar are highly skewed, with the majority of values clustered at very low values close to zero, whereas porosity exhibits a more or less uniform distribution across the range of 0–0.33 (v/v). As previously discussed, the input logs display a wide range of values; for instance, neutron porosity is below 1, while gamma ray and resistivity

**Table 2**  
Input features and target features used in the present study.

Input features	Units	Target features	Units
Natural gamma-ray log (GR)	GAPI	Quartz	v/v
Formation resistivity (RT)	ohm·m	Dolomite	v/v
Formation bulk density (RHOZ)	g/cc	K-Feldspar	v/v
Neutron porosity (NPHI)	v/v	Anhydrite	v/v
Volumetric Photo-electric factor (U_PEFZ)	b/cc	Illite	v/v
Compressional slowness (DTCO)	μs/ft	Porosity	v/v

**Table 3**

Range, mean, and median values of input and target features, along with statistical indicators for data recorded in the Broom Creek Formation of Training Well-I. The sample size is 519. All the values are rounded to 2 decimal places.

Variable	Minimum	Mean	Median	Maximum
GR (GAPI)	15.05	50.28	38.73	206.59
RT (ohm·m)	0.25	5.94	1.06	170.19
RHOZ (g/cc)	2.15	2.44	2.39	2.97
NPHI (v/v)	0.00	0.19	0.20	0.35
DTCO (μs/ft)	48.34	75.90	76.03	99.61
U_PEFZ (b/cc)	4.11	6.75	6.36	13.75
Illite (v/v)	0.00	0.03	0.00	0.32
Anhydrite (v/v)	0.00	0.05	0.02	0.73
Dolomite (v/v)	0.00	0.28	0.22	0.87
K-Feldspar (v/v)	0.00	0.08	0.06	0.50
Quartz (v/v)	0.00	0.38	0.46	0.77
Porosity (v/v)	0.01	0.18	0.18	0.33

measurements vary by several hundred. Given this variability, it is recommended to standardize these values before using them in neural network training to prevent bias.

Spearman's correlation coefficients between the input and target attributes were computed and presented in Fig. 4. Unlike Pearson's correlation coefficient, which captures only linear relationships, Spearman's coefficient assesses the rank-based (non-linear) correlations between variables. Given that well logs frequently exhibit non-linear relationships with petrophysical properties, Spearman's correlation was deemed more appropriate for this analysis. As shown in the figure, porosity is highly correlated with the input logs, while the mineral components exhibit moderate correlations. Notably, quartz and dolomite demonstrate a strong negative correlation, which is likely indicative of depositional changes. Similarly, quartz and porosity are strongly positively correlated, whereas the presence of dolomite and anhydrite

negatively impacts porosity, as reflected in the observed negative correlations. These inherent correlations among the target attributes underscore the rationale for simultaneously predicting all target variables using a Physics Integrated Neural Network (PINN), leveraging appropriate relationships and constraints to enhance the accuracy of the predictions.

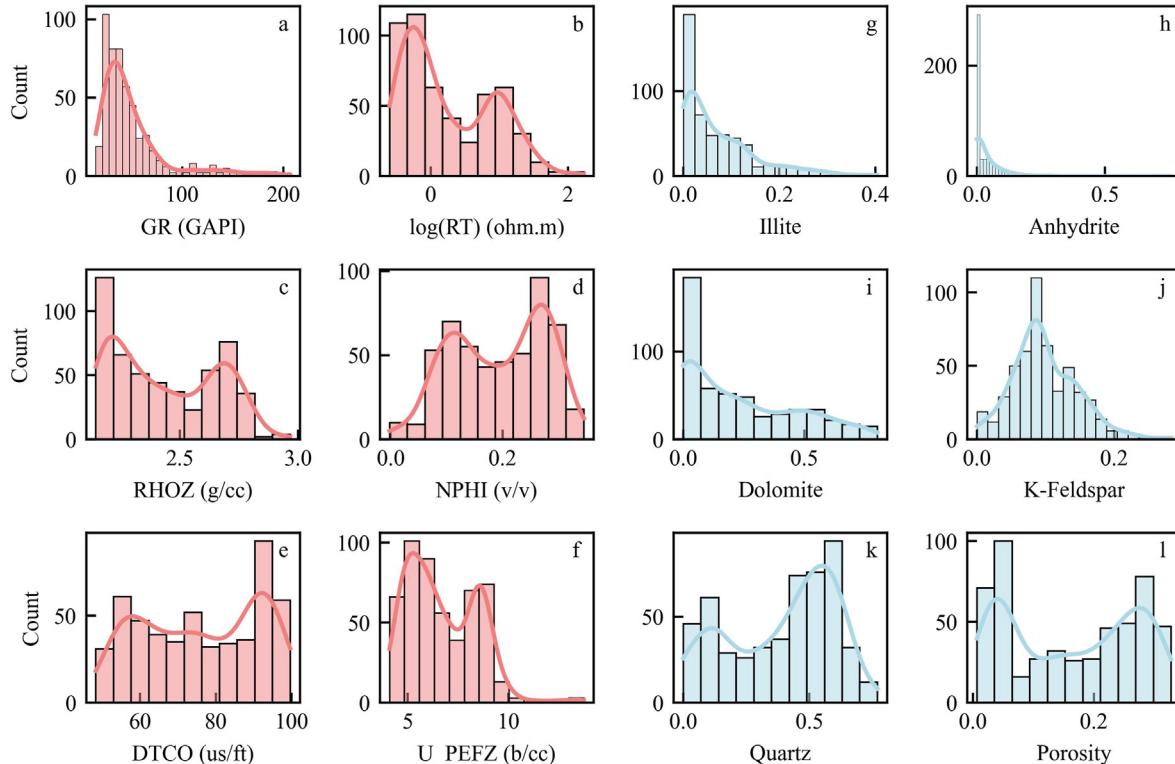
### 3.2.1. Multimineral petrophysical evaluation

The ground truth values for mineral volumes and porosity in this study were derived from a multimineral petrophysical analysis. This analysis involves selecting input logs, identifying dominant minerals, formulating an inverse matrix, and iteratively refining the parameters to achieve a close match between measured and predicted well logs. This process is carried out while ensuring consistency with geological, core, and drilling data. For this study, we followed the procedure outlined in Pothana et al. (2024). Furthermore, the estimated mineral volumes were validated against XRD results and core description data, while porosity estimations were cross-verified with routine core analysis results.

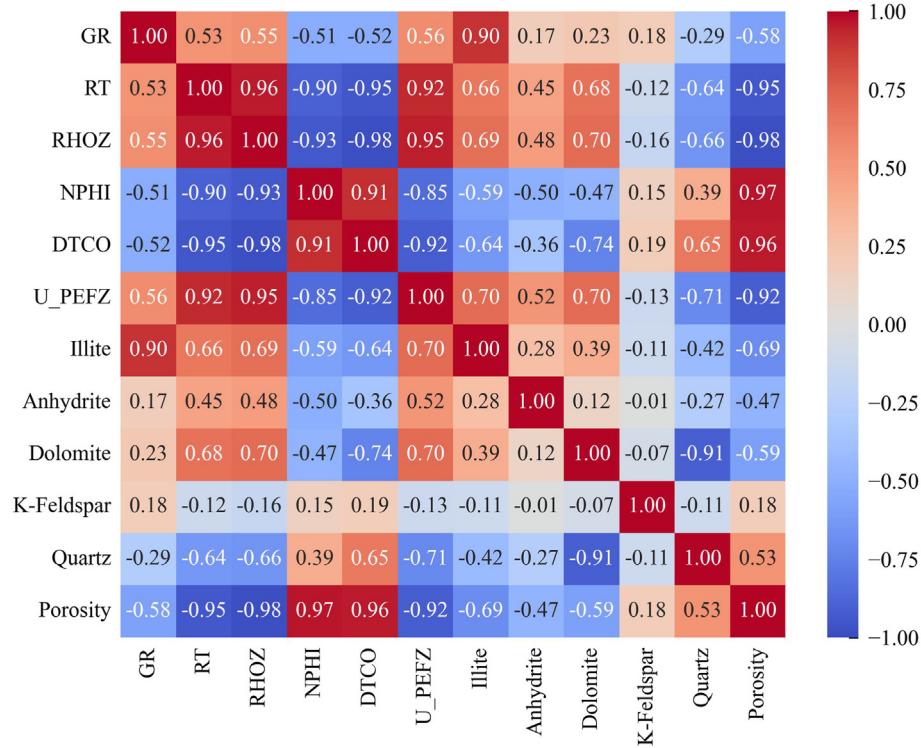
### 3.3. Physics-integrated neural network training process

Prior to training, the input features were standardized according to Equation (2). A custom output layer was developed to integrate the physical relationships defined by Equations (5)–(7). These physical constraints were incorporated using the methodology outlined in Section 2.4.9. It is important to note that the Broom Creek Formation is a deep saline aquifer devoid of hydrocarbons, thus rendering the inclusion of Equation (8), which accounts for hydrocarbon saturation, unnecessary.

Archie's equation (Equation (9)) was employed to capture the resistivity porosity relationship. The rationale behind this choice stems from the minimal clay content in the Broom Creek Formation



**Fig. 3.** Statistical distribution of input features (a–f) and target attributes (g–l) of the training well. The continuous curve represents the KDE distribution of the data.



**Fig. 4.** Spearman's correlation matrix between the input and target features. Each cell represents the Spearman's correlation coefficient, indicating the strength and direction of the monotonic relationship between the corresponding features.

(approximately 3 % on average), which implies a negligible contribution of clay-bound water conductivity. As the formation is fully saturated with brine, we set  $S_w = 1$ , and the resistivity of the brine was taken as 0.05 ohm·m at 150 °F. Furthermore, standard Archie's parameters were applied, with  $a = 1$  and  $n = 2$ , while the cementation exponent ( $m$ ) was treated as a variable bounded between [1.5, 2.5], allowing the model to account for lithological variability across the formation. It is important to note that, in the present case where  $S_w = 1$ , the values of  $n$  do not influence the results. However, in a more generalized scenario where hydrocarbons are present ( $S_w \neq 1$ ), the choice of  $n$  becomes critical.

Additionally, the Gamma-ray (GR) equation for radioactive minerals (Equation (11)) was incorporated, with K-feldspar serving as a proxy for the radioactive mineral content. The constant  $k$  in the equation was treated as a variable, with bounds set between [0.5, 1.5], allowing the model to estimate  $k$  through the training process based on data. We deliberately chose not to incorporate the porosity relationships discussed in Section 2.4.7, given the strong correlation between porosity and the various geophysical logs, as evidenced by the Spearman correlation heatmap (Fig. 4). These high correlation values indicate that the input logs have sufficient predictive power for porosity, making the inclusion of additional porosity-predictive equations unnecessary in this context. However, it is important to emphasize that the selection of physical laws for neural network training should be assessed on a case-by-case basis, and generalizations should be avoided. Our approach here is tailored specifically to the characteristics of the Broom Creek Formation, and different geological settings may require the integration of different equations to optimize predictive accuracy.

Following the initial setup, we fine-tuned the PINN's hyperparameters using a comprehensive search process. The custom output layer was designed to handle the porosity and mineral outputs separately. The porosity output was bounded between

0 and 0.4 using a sigmoid activation function (see Section 2.4.9). In contrast, a softmax activation constrained the mineral components to ensure that their combined contribution remained consistent with physical constraints (see Section 2.4.9).

The fine-tuning of the PINN involved multiple aspects, starting with the number of hidden layers, which varied between one and three layers. Each layer's number of units was adjusted in the range of [16, 64], with increments of 8, allowing for sufficient flexibility in model complexity. The learning rate was fine-tuned within a range of  $1 \times 10^{-6}$  to  $1 \times 10^{-1}$  using logarithmic scaling to explore both lower and higher values. This ensured that the model's convergence rate was optimized for both stability and efficiency. To further incorporate the physical relationships, two distinct physics losses were defined. The first physics loss focused on the relationship between porosity and RT. The second physics loss modeled the relationship between GR, illite, and K-feldspar quantities. A tunable factor scaled both losses,  $\lambda_{\text{physics}}$ , which was varied between 0.01 and 0.5 to balance the contribution of physical constraints to the overall loss function (see Section 2.4.9).

The search for optimal hyperparameters was conducted using a random search approach, exploring 1000 trials with two executions per trial using RandomSearch Tuner (Keras, 2024). The model was trained for 100 epochs per trial with a batch size of 32, and a validation split of 30 % was applied to ensure reliable performance evaluation. The optimum parameters found during the fine-tuning process are presented in Table 4. Fig. 5 depicts the model convergence during the training process.

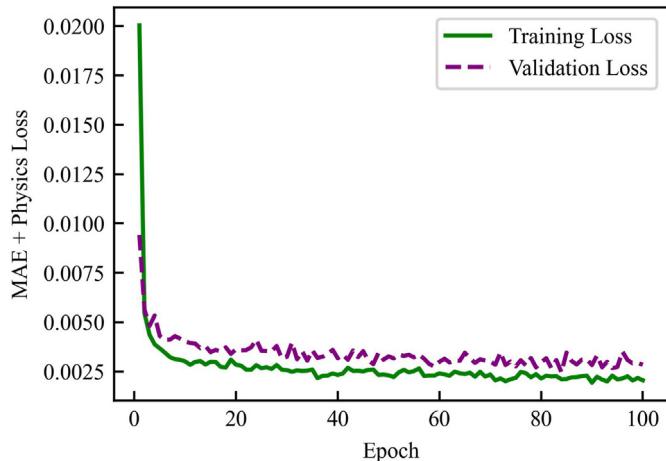
### 3.4. Error metric

As previously discussed, geophysical logs represent physical measurements taken at regular depth intervals along the borehole. To perform a meaningful error analysis, it is essential to estimate an

**Table 4**

Hyperparameters used for training the physics-integrated neural network model.

Parameter	Value
Activation function	ReLU
Number of layers	3
Units in layer-1	32
Units in layer-2	48
Units in layer-3	48
Learning rate	0.0041
Lamda Physics	0.41
$\sigma$	1.0



**Fig. 5.** Evolution of training loss and validation loss throughout the PINN training process.

error curve at each depth point, allowing for the examination of errors at specific depths. This depth-specific error analysis provides additional information beyond the cumulative error calculated over the entire log. The error log, a combination of Weighted RMSE and Sum Constraint Error at each depth point, is computed using the following formula:

$$\text{Total Error} = \sqrt{\sum_i \left( (Y_{\text{true},i} - Y_{\text{pred},i})^2 \cdot \frac{Y_{\text{true},i}}{\sum_i Y_{\text{true},i}} \right)} + \left| 1 - \sum_i Y_{\text{pred},i} \right| \quad (23)$$

In this equation,  $Y_{\text{true},i}$  represents the ground true value of the  $i$ -th mineral constituent, while  $Y_{\text{pred},i}$  is the predicted value for the  $i$ -th mineral constituent. This function combines two essential aspects: Weighted RMSE (first term) and Sum Constraint Error (second term). The expression  $\frac{Y_{\text{true},i}}{\sum_i Y_{\text{true},i}}$  is the normalized true value for

the  $i$ -th mineral, ensuring that the error is weighted by the importance of each constituent, with dominant minerals having a larger impact on the error. In other words, More dominant minerals should contribute more to the error if they are predicted incorrectly, reflecting their significance in the composition. The term

$\left| 1 - \sum_i Y_{\text{pred},i} \right|$  represents the absolute difference between 1 and the sum of predicted values, which enforces the constraint that the total predicted mineral composition must sum to exactly 1. This

sum constraint error ensures that the prediction respects the physical limitation of the system.

#### 4. Results and discussion

The Physics-Informed Neural Network (PINN) training was conducted using data from one well, designated as Training Well-I, while the validation was performed on three unseen (blind) wells: Test Well-I, Test Well-II, and Test Well-III. The cumulative mineralogical distributions, along with the input geophysical logs for Training Well-I, are displayed in Fig. 6. Similarly, the results for the testing wells are presented in Figs. 7–9.

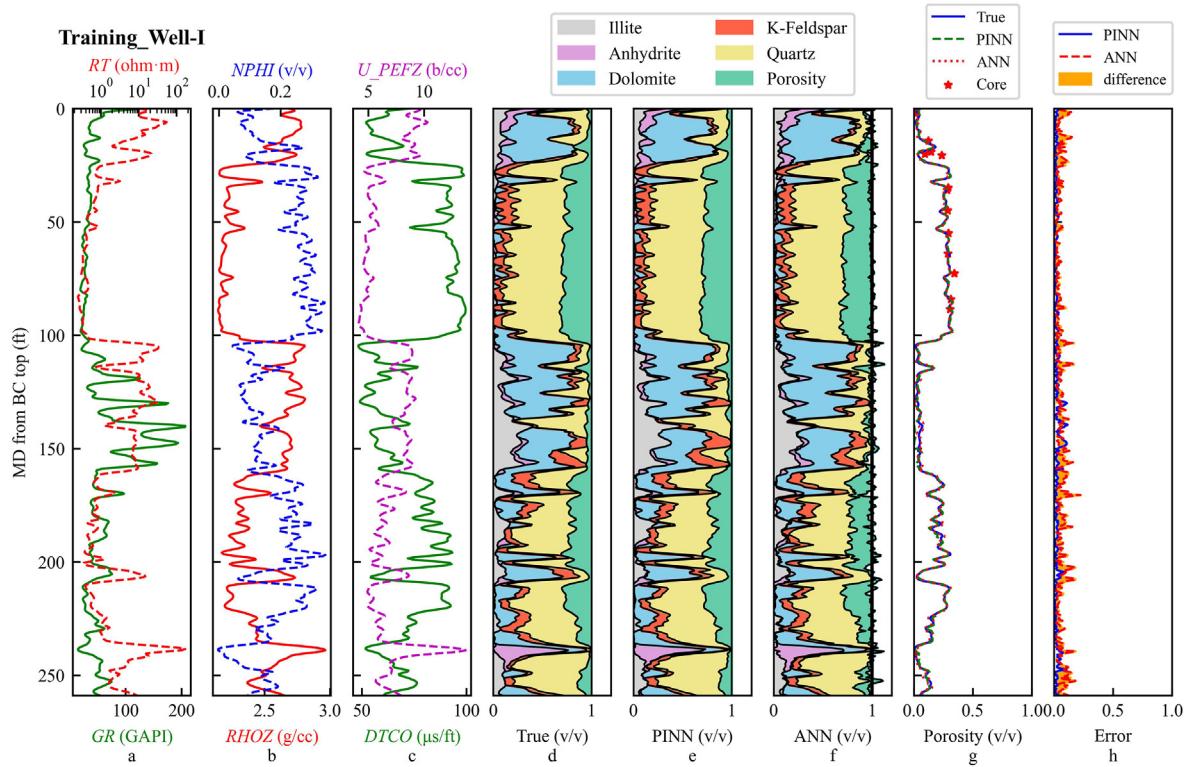
In Figs. 6–9, the ground truth values are depicted in track-d, while the PINN predictions are shown in track-e, and the ANN results are illustrated in track-f. Each mineral is distinctly represented using unique color coding for clarity. The results clearly demonstrate that the mineralogical and porosity predictions generated using PINN closely align with the true values obtained through manual petrophysical analysis. The proposed methodology effectively predicts mineralogical variations within the formation with high accuracy. Notably, key reservoir minerals such as Quartz and Feldspar are predicted with high precision, along with the accurate identification of impermeable layers, including Dolomite and Anhydrite.

Notably, the predicted mineral volume fractions remain within the theoretical bounds, which is a significant indicator of PINN's ability to incorporate physical laws into its predictions. This accuracy contrasts with the traditional Artificial Neural Network (ANN), which was used as a baseline comparison in our study. Unlike PINN, the ANN does not incorporate additional physical constraints or loss functions. The results show that PINN significantly reduces error rates compared to ANN, providing predictions that are more consistent with geological and physical principles. It can be clearly seen from the track-f of Figs. 6–9 that the ANN model often predicts cumulative mineral sums exceeding 1, violating mass conservation laws, which further underscores its limitations.

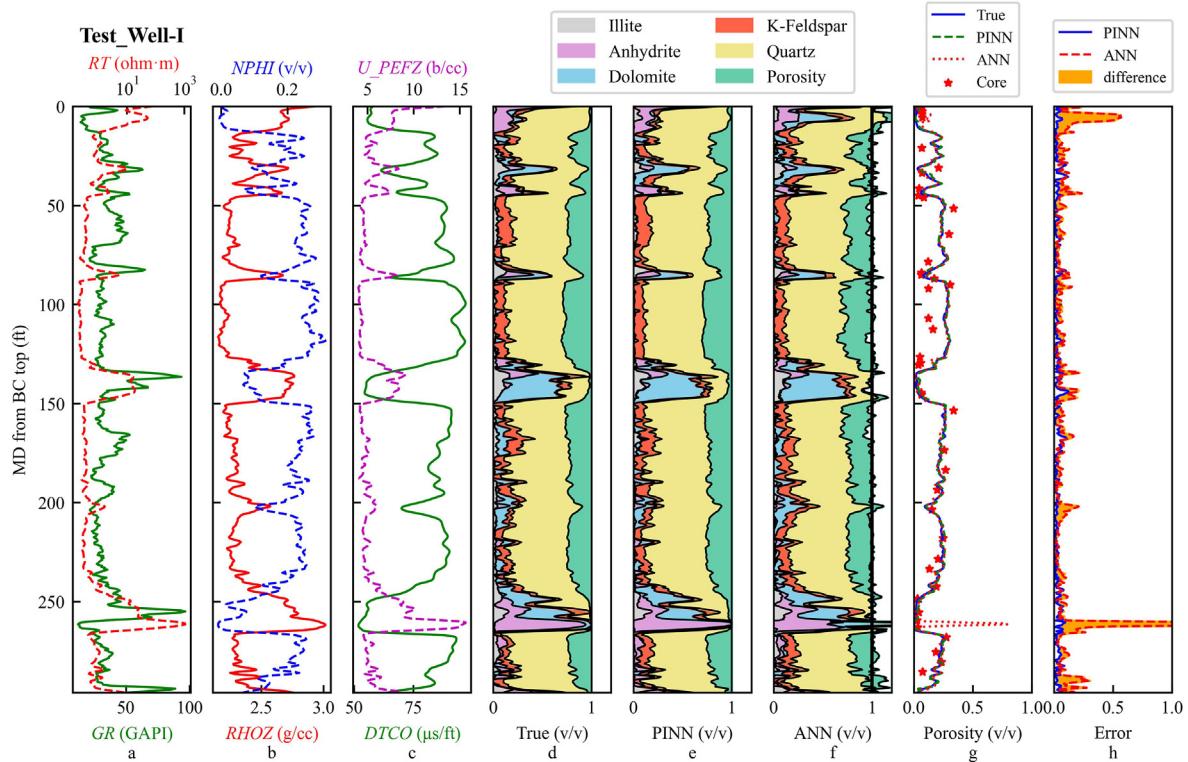
The PINN-predicted porosity is superimposed on the true porosity (estimated from manual petrophysical analysis) and the core-derived porosity, as shown in track-g in Figs. 6–9. It is evident that the predicted values align exceptionally well with the core-derived data, further highlighting the predictive capability of the PINN model.

An interesting observation can be made in the Training Well-I results (Fig. 6), where the interval from 100 to 150 ft, representing dolomitic sections, shows a slightly lower error for ANN compared to PINN. However, a deeper analysis of the predictions for the test wells (Figs. 7–9) reveals that the ANN error consistently surpasses that of the PINN model, particularly in Test Well-II (Fig. 8). This outcome highlights a classic limitation of ANN models: their propensity for overfitting training data, leading to reduced generalization when applied to unseen datasets. In contrast, PINN's integration of physical constraints and adherence to petrophysical laws allows it to avoid overfitting, resulting in more reliable and generalized predictions.

While the overall performance of PINN surpasses that of ANN, some areas still require improvement. In particular, the prediction accuracy for anhydrite peaks warrants further enhancement. Several factors could contribute to this issue. First, the data-driven nature of the model may cause bias due to the relatively limited sampling of these thin anhydrite layers compared to the more abundant data for other minerals. This sampling imbalance likely skews the training process towards the more dominant minerals. To address this, acquiring additional data or employing data augmentation techniques could help improve the representation of less frequent lithologies. Second, the inherent vertical resolution



**Fig. 6.** Input logs and predictions for Training Well-I: Tracks a through f present various input logs, predictions, and error analyses. Track a displays RT and GR logs, while Track b shows NPHI and RHOZ logs. Track c includes U\_PEFZ and DTCO logs. Track d illustrates the cumulative distribution of true mineralogy derived from manual petrophysical analysis. Track e represents the cumulative distribution of mineralogy and porosity predicted using PINN, and Track f presents the same predictions using ANN. True porosity, PINN-predicted porosity, and core-derived porosity are presented in track g. The final track h displays the error logs for both PINN and ANN, computed using equation (23).



**Fig. 7.** Input logs and Predictions for Test Well-I.

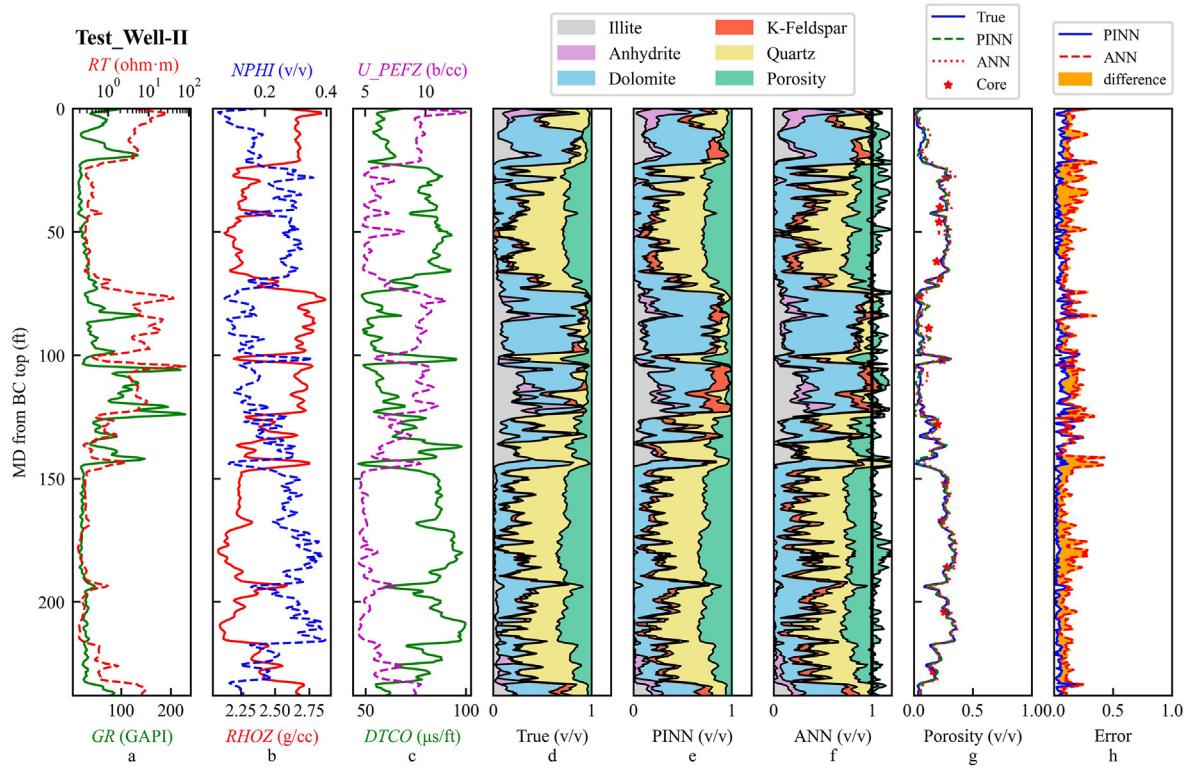


Fig. 8. Input logs and Predictions for Test Well-II.

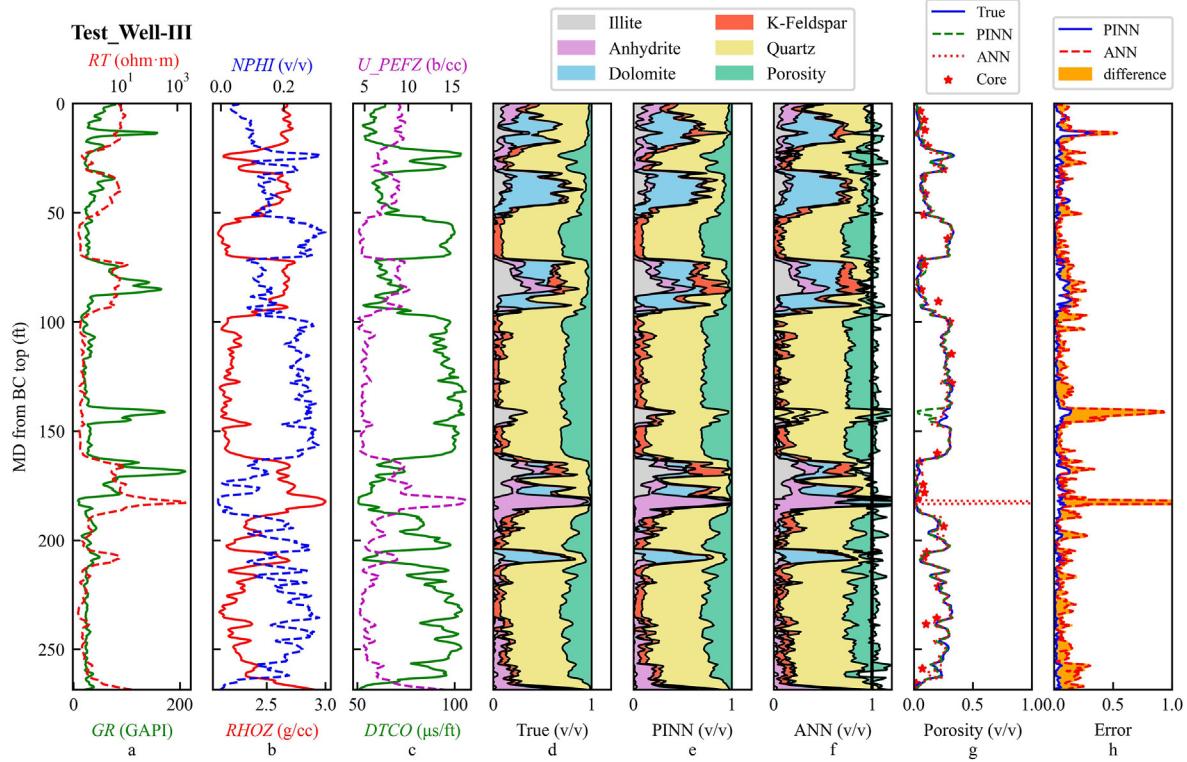
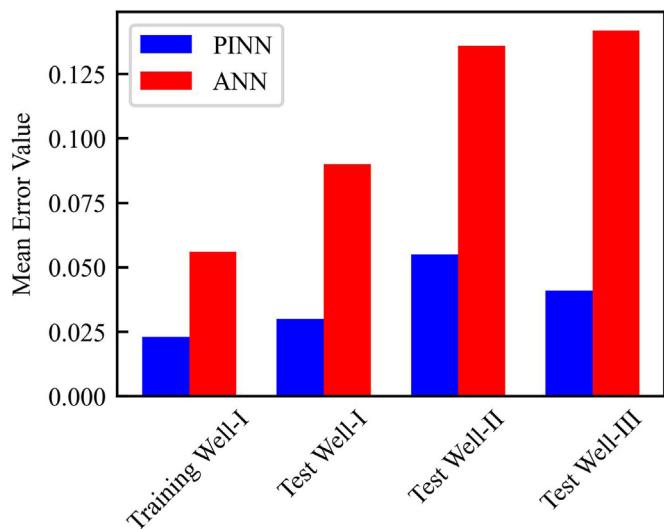


Fig. 9. Input logs and Predictions for Test Well-III.

limitations of well logs may also play a role. Sharp lithological changes in the formation may appear differently across varying well log window lengths, which affects the model's ability to detect such variations precisely. Addressing these discrepancies by ensuring comparable vertical resolutions across different logs could lead to more accurate predictions in these challenging intervals. Additionally, incorporating multi-scale modeling approaches or enhancing the resolution of the training data could further refine the prediction accuracy in such cases.

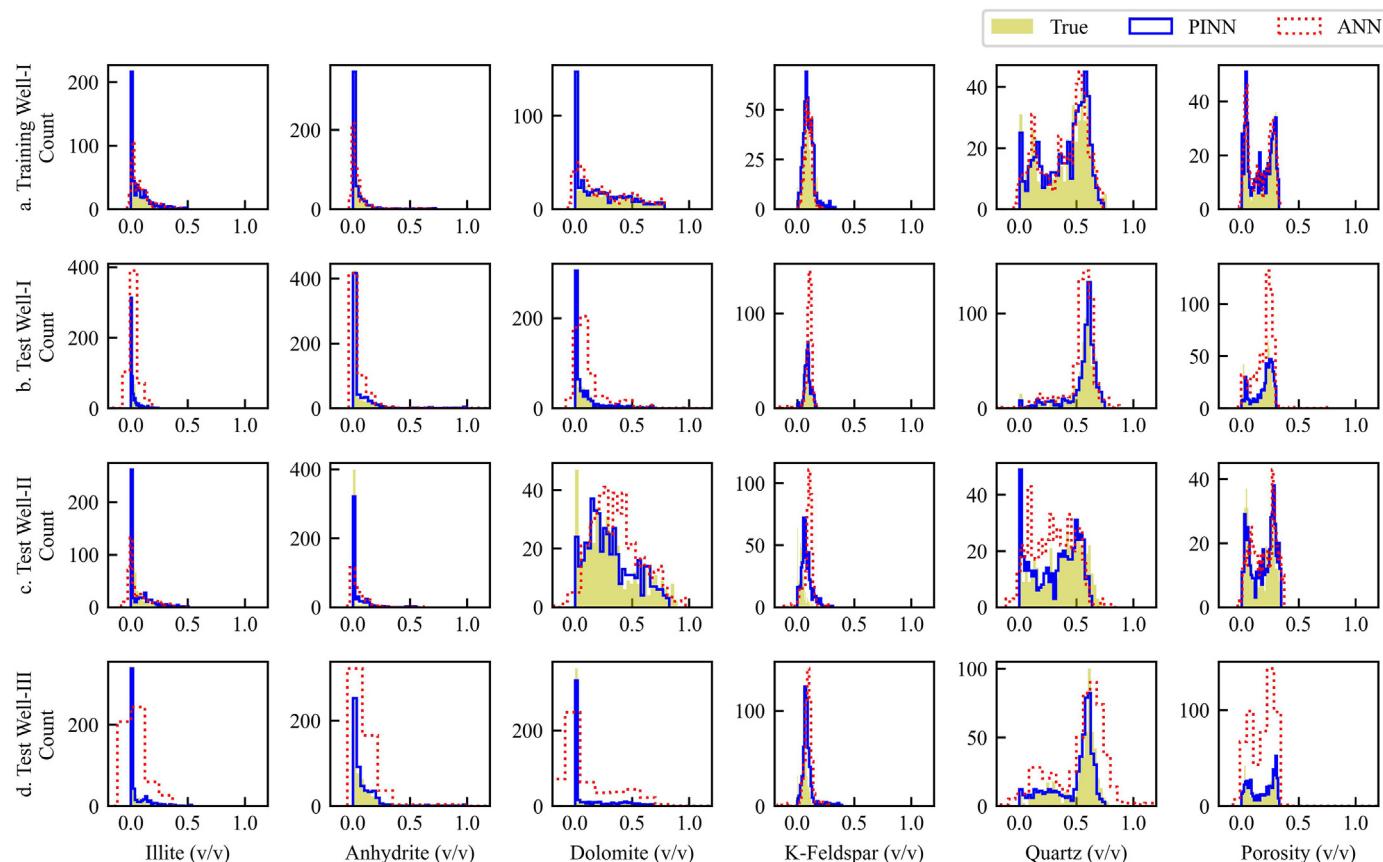
The distributions of all minerals and porosity for the true values, PINN-predicted values, and ANN-predicted values are presented as histograms in Fig. 10. This plot includes data from both the training well and the blind wells. It is evident that the predictions made by the PINN closely align with the actual values (shaded region). In contrast, the predictions from the traditional ANN exhibit significant discrepancies, including inconsistencies such as negative values and cumulative sums exceeding 1, which violate physical constraints. These results highlight the superior performance and reliability of the PINN model in adhering to geological and physical laws.

The error distributions, computed using equation (23), for PINN and ANN are presented in Fig. 11. The error statistics for PINN and ANN predictions are shown in Table 5, and mean value comparisons are presented in Fig. 12. The PINN error distribution is close to zero. The results show that the PINN consistently exhibits lower error values compared to the ANN across all wells. In particular, for Test Well-I and Test Well-III, the ANN error exceeds a value of 1 in certain intervals, rendering the corresponding predictions



**Fig. 11.** Error distributions predicted by PINN and ANN for a. Training Well-I, b. Test Well-I, c. Test Well-II, and d. Test Well-III. The plots consistently demonstrate that the errors from the ANN model are higher than those from the PINN model across all wells.

physically invalid due to significant overestimation. This further highlights the superior performance of the PINN in maintaining accuracy and consistency.



**Fig. 10.** Distribution of minerals and porosity. Each row corresponds to data from a specific well: a. Training Well-I, b. Test Well-I, c. Test Well-II, and d. Test Well-III. In each plot, the shaded regions represent the true values derived from routine petrophysical analysis. The unfilled histograms represent the predictions, with the blue histograms indicating predictions from the PINN model and the red histograms indicating predictions from the ANN model.

**Table 5**

Combined error statistics of PINN and ANN predictions for the training well and test wells. All the values are rounded to 3 decimal places.

Well	Statistics	PINN	ANN
Training Well-I	Min	0.001	0.005
	Mean	0.023	0.056
	Median	0.019	0.047
	Max	0.115	0.224
Test Well-I	Min	0.004	0.006
	Mean	0.030	0.090
	Median	0.023	0.057
	Max	0.123	3.039
Test Well-II	Min	0.002	0.008
	Mean	0.055	0.136
	Median	0.044	0.119
	Max	0.28	0.427
Test Well-III	Min	0.001	0.005
	Mean	0.041	0.142
	Median	0.03	0.087
	Max	0.335	7.421

#### 4.1. Practical implications of mineralogical estimation in CO<sub>2</sub> reservoir characterization

The accurate estimation of mineral volumes in geological formations is a cornerstone for multiple geoengineering applications. This is particularly vital in the characterization of CO<sub>2</sub> storage reservoirs, where understanding the mineralogical composition directly influences assessments of storage capacity, geomechanical

stability, and long-term integrity. In the current study, mineralogical information serves as a key input for evaluating the suitability of the Broom Creek Formation for CO<sub>2</sub> sequestration.

The dominant pore volume in this formation, critical for CO<sub>2</sub> storage, is associated with sandstone lithology, which primarily comprises quartz and feldspar minerals. These minerals not only offer higher porosity and permeability but also present relatively stable long-term storage conditions. Conversely, sections dominated by dolomite exhibit minimal porosity and are impermeable. The differentiation between these lithologies is, therefore, important for optimizing reservoir assessment and management strategies.

Thin-section studies of the reservoir reveal that the primary grain constituents, quartz and feldspar, are often cemented by dolomite, anhydrite, and quartz overgrowths. Within the reservoir sandstone sections, this cementation minimally reduces the porosity and permeability while increasing the stiffness. Fig. 13 illustrates the spatial distribution of grains and cementation materials from a photomicrograph within the porous sandstone section in Test Well-I, reinforcing the critical need for accurate mineralogical mapping.

Long-term CO<sub>2</sub> storage necessitates a comprehensive understanding of the interactions between injected CO<sub>2</sub>, brine, and reactive minerals within the reservoir. For example, CO<sub>2</sub> dissolution in water forms carbonic acid, which reacts with dolomite, leading to its dissolution (Wang et al., 2013). This process can simultaneously enhance porosity but may also result in secondary mineral precipitation, potentially clogging pore spaces and altering flow

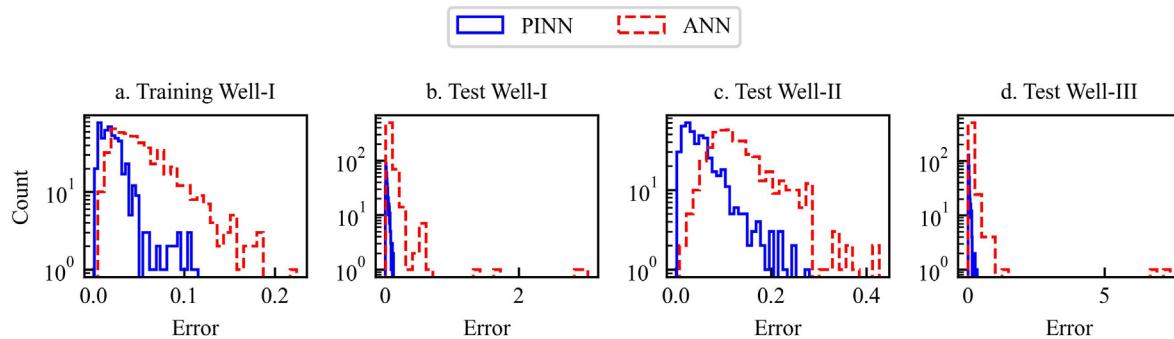


Fig. 12. Comparison of the mean error values for PINN and ANN computed across the depth for each well.

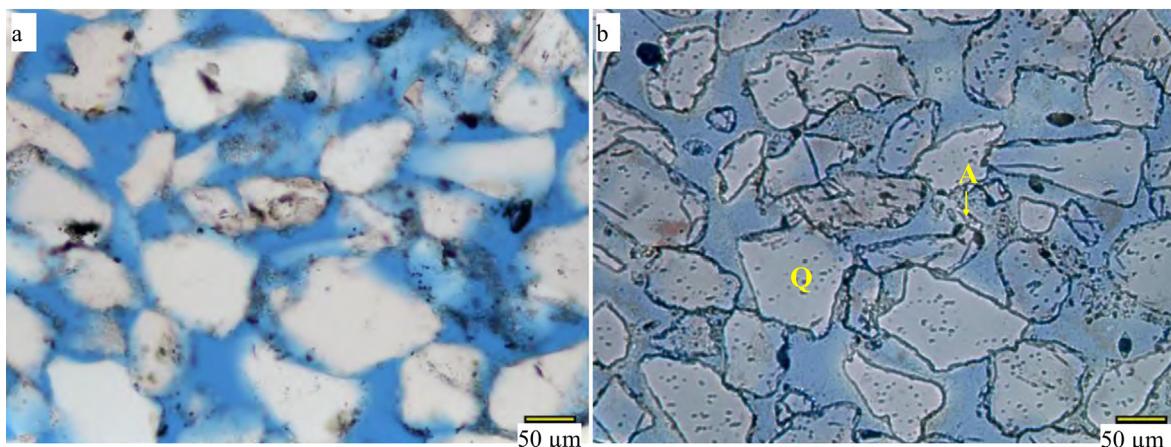


Fig. 13. Photomicrographs from Test Well-I, Broom Creek Formation (Depth: 228.5 m from top): a. Plane-Polarized (20 × PP) light and b. Reflected Light (20 × RL) views of a sandstone sample predominantly composed of quartz (80 %), feldspar (5 %), and anhydrite cement (10 %). Anhydrite crystallization between quartz grains locally obstructs intercrystalline porosity and separates the grains.

dynamics (Famouri et al., 2021; Wang et al., 2013). Additionally, the elastic properties of sandstone stiffened by dolomite and anhydrite cement indicate potential geomechanical implications. Dissolution at grain contacts could weaken the structural integrity of the reservoir, making time-lapse geomechanical monitoring a critical component of long-term storage assessments.

The methodology presented in this study significantly enhances the ability to quantify mineral volumes rapidly and robustly. Unlike traditional methods, which are often time-intensive, this approach integrates physics-informed constraints with data-driven models, ensuring both accuracy and physical consistency. By reducing the uncertainty in mineralogical estimations, this method enables better predictions of storage capacity and risks associated with mineral dissolution and precipitation processes.

## 5. Conclusions

Traditional ML models often work as “black boxes” and offer less intuition of interpretability. Sometimes, the predictions violate fundamental physical laws. This research successfully implements physics-integrated neural networks (PINN) to predict mineral volumes and porosity using geophysical well logs. We presented various physical laws that can be incorporated and the elaborated procedure on how these additional constraints can be integrated into the neural network architecture. Integrating physics into neural networks significantly increases the accuracy of mineralogy and porosity predictions from geophysical well logs. PINNs make sure that the predictions will follow physical laws; thus, their outputs are more in tune with geological realities compared to traditional artificial neural networks. Additionally, PINN performed better than traditional ANNs by considerably reducing over-fitting and enhancing predictive performance on unseen data. The average mean error across the three blind wells is 0.123 for ANN and 0.042 for PINN, highlighting the superior accuracy of the PINN approach. Finally, PINN improves the fidelity of mineralogical and porosity predictions by coupling data-driven learning with physical laws, thus helping in better resource management and development strategies.

## CRediT authorship contribution statement

**Prasad Pothana:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kegang Ling:** Validation, Supervision, Resources, Funding acquisition.

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 4o in order to improve the english language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Declaration of competing interest

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

## Acknowledgments

Authors would like to thank the North Dakota Industrial Commission (NDIC) for their financial support. This work used advanced

cyberinfrastructure resources provided by the University of North Dakota Computational Research Center.

## Abbreviations

<i>GR</i>	Gamma Ray log (GAPI)
<i>RT</i>	Formation Resistivity ( $\text{ohm}\cdot\text{m}$ )
<i>RHOB</i>	Bulk Density log ( $\text{g}/\text{cc}$ )
<i>NPHI</i>	Neutron Porosity ( $\text{v}/\text{v}$ )
<i>DTCO</i>	Compressional Slowness ( $\mu\text{s}/\text{ft}$ )
<i>U_PEFZ</i>	Volumetric Photo-electric Factor ( $\text{b}/\text{cc}$ )
<i>V<sub>Quartz</sub></i>	Volume Fraction of Quartz ( $\text{v}/\text{v}$ )
<i>V<sub>Dolomite</sub></i>	Volume Fraction of Dolomite ( $\text{v}/\text{v}$ )
<i>V<sub>K-Feldspar</sub></i>	Volume Fraction of K-Feldspar ( $\text{v}/\text{v}$ )
<i>V<sub>Anhydrite</sub></i>	Volume Fraction of Anhydrite ( $\text{v}/\text{v}$ )
<i>V<sub>Illite</sub></i>	Volume Fraction of Illite ( $\text{v}/\text{v}$ )
$\phi$	Porosity ( $\text{v}/\text{v}$ )
$a$	Tortuosity Factor
$m$	Cementation Factor
$n$	Saturation Exponent
$S_w$	Water Saturation ( $\text{v}/\text{v}$ )
$R_w$	Formation Water Resistivity ( $\text{ohm}\cdot\text{m}$ )
$R_t$	True Formation Resistivity ( $\text{ohm}\cdot\text{m}$ )
$V_{sh}$	Volume Fraction of Shale ( $\text{v}/\text{v}$ )
$R_{sh}$	Shale Resistivity ( $\text{ohm}\cdot\text{m}$ )
$k$	Scaling Constant (Gamma-Ray relation)
$\rho_m$	Matrix Density ( $\text{g}/\text{cc}$ )
$\rho_b$	Bulk Density ( $\text{g}/\text{cc}$ )
$\rho_f$	Fluid Density ( $\text{g}/\text{cc}$ )
$\Delta t_{log}$	Interval Transit Time (Sonic Log) ( $\mu\text{s}/\text{ft}$ )
$\Delta t_{ma}$	Matrix Transit Time ( $\mu\text{s}/\text{ft}$ )
$\Delta t_{fl}$	Fluid Transit Time ( $\mu\text{s}/\text{ft}$ )
$\mathcal{L}_{\text{physics}}$	Physics Loss Function
$\mathcal{L}_{\text{MSE}}$	Mean Squared Error Loss Function
$\mathbf{Y}$	Predicted Output Matrix
$\mathbf{X}$	Input Matrix
$\mathbf{R}$	Response Matrix (Petrophysical)
$\mathbf{W}$	Weight Matrix in Neural Network
$\mathbf{z}$	Layer Output in Neural Network
$\mathbf{b}$	Bias Term in Neural Network
$\sigma(x)$	Sigmoid Activation Function
$\text{ReLU}(x)$	Rectified Linear Unit Function
$\text{softmax}(z)$	Softmax Activation Function
$\epsilon$	Noise Term
$\lambda_{\text{physics}}$	Weight for Physics Loss Function
$\phi_{\text{max}}$	Maximum Porosity ( $\text{v}/\text{v}$ )
$V_{Clay}$	Volume Fraction of Clay ( $\text{v}/\text{v}$ )
$V_{RA}$	Volume Fraction of Radioactive Minerals ( $\text{v}/\text{v}$ )
$E$	Expected Value/Mean

## References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: large-scale machine learning on heterogeneous systems. URL: <https://www.tensorflow.org/softwareavailablefromtensorflow.org>.
- Ali, M., Jiang, R., Ma, H., Pan, H., Abbas, K., Ashraf, U., Ullah, J., 2021. Machine learning-a novel approach of well logs similarity based on synchronization measures to predict shear sonic logs. *J. Petrol. Sci. Eng.* 203, 108602. <https://doi.org/10.1016/j.petrol.2021.108602>.
- Archie, G., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics. *Transactions of the AIME* 146, 54–62. <https://doi.org/10.2118/942054-G>.
- Asquith, G., Krygowski, D., 2006. Porosity logs. In: *Basic Well Log Analysis*, second ed. AAPG, Tulsa, OK. volume 16 of *AAPG Methods in Exploration Series*, pp. ix,

- 244.**
- Bettir, N., Mellal, I., Dehdouh, A., Rabiei, M., Malki, M.L., Allam, L., 2024. Application of Machine Learning Algorithms for Minerals Volume Prediction in Unconventional Reservoirs Using Conventional Well Logs. <https://doi.org/10.56952/ARMA-2024-1029>. D0325041R012.
- Bhattacharya, S., Bakhshian, S., Hovorka, S., Uroza, C., Hosseini, S., Bump, A., Trevino, R., Olariu, I., Haagsma, A., 2023. Integrated petrophysical studies for subsurface carbon sequestration. In: SPWLA Annual Logging Symposium. SPWLA. D031S001R003.
- Burak, T., Sharma, A., Hoel, E., Kristiansen, T.G., Welmer, M., Nygaard, R., 2024. Real-time lithology prediction at the bit using machine learning. Geosciences 14. <https://doi.org/10.3390/geosciences14100250>.
- Burke, L.A., Birdwell, J.E., Paxton, S.T., 2022. Multimineral petrophysics of thermally immature eagle ford group and cretaceous mudstones, u.s. geological survey gulf coast 1 research wellbore in central Texas. Interpretation 10, T151–T165. <https://doi.org/10.1190/INT-2021-0094.1>.
- Chaiikine, I.A., Gates, I.D., 2020. A New Machine Learning Procedure to Generate Highly Accurate Synthetic Shear Sonic Logs in Unconventional Reservoirs. <https://doi.org/10.2118/201453-MS>. D031S027R007.
- Chen, Y., Li, L., Li, W., Guo, Q., Du, Z., Xu, Z., 2024. Chapter 2 - fundamentals of neural networks. In: Chen, Y., Li, L., Li, W., Guo, Q., Du, Z., Xu, Z. (Eds.), AI Computing Systems. Morgan Kaufmann, pp. 17–51. <https://doi.org/10.1016/B978-0-323-395399-3.00008-1>. URL: <https://www.sciencedirect.com/science/article/pii/B9780323953993000081>.
- Cheng, L., Jin, G., Michelena, R., Tura, A., 2022. Practical bayesian inversions for rock composition and petrophysical endpoints in multimineral analysis. SPE Reservoir Eval. Eng. 25, 849–865. <https://doi.org/10.2118/210576-PA>.
- Clavier, C., Coates, G., Dumanoir, J., 1984. Theoretical and experimental bases for the dual-water model for interpretation of shaly sands. Soc. Petrol. Eng. J. 24, 153–168. <https://doi.org/10.2118/6859-PA>.
- D'amore, F., Panichi, C., 1985. Geochemistry in geothermal exploration. Int. J. Energy Res. 9, 277–298. <https://doi.org/10.1002/er.4440090307>.
- Doveton, J.H., 2014. Compositional analysis of mineralogy. In: Principles of Mathematical Petrophysics. Oxford University Press. <https://doi.org/10.1093/oso/9780199978045.003.0009>.
- Famouri, F., Azin, R., Osfouri, S., Arya, N., 2021. Experimental investigation of the geochemical and mineralogical interaction between CO<sub>2</sub> and carbonate: evaluation of CO<sub>2</sub> sequestration in dolomite-calcite formations. Energy and Climate Change 2, 100029. <https://doi.org/10.1016/j.egycc.2021.100029>. URL: <https://www.sciencedirect.com/science/article/pii/S2666278721000064>.
- Feng, P., Wang, R., Sun, J., Yan, W., Chi, P., Luo, X., 2024. An interpretable ensemble machine-learning workflow for permeability predictions in tight sandstone reservoirs using logging data. GEOPHYSICS 89, MR265–MR280. <https://doi.org/10.1190/geo2023-0657.1>.
- Gerhard, L.C., Anderson, S.B., Lefever, J.A., Carlson, C.G., 1982. Geological development, origin, and energy mineral resources of williston basin, North Dakota. AAPG (Am. Assoc. Pet. Geol.) Bull. 66, 989–1020.
- Guillon, V., Delamaide, E., Rousseau, D., Morvan, M., 2022. On the Importance of Reservoir Rock Mineralogy on Design and Performance of Surfactant-Based EOR Processes. <https://doi.org/10.2118/200247-MS>. D0225055R001.
- Gupta, A.R., et al., 2023. A theoretical approach for water saturation estimation in shaly sandstones. Geoenergy Science and Engineering 228, 212001. <https://doi.org/10.1016/j.geoen.2023.212001>.
- Gurney, K., 1997. In: An Introduction to Neural Networks, first ed. CRC Press. <https://doi.org/10.1201/9781315273570>. eBook Published 30 January 2017.
- Houshmand, N., GoodFellow, S., Esmaeili, K., Ordoñez Calderon, J.C., 2022. Rock type classification based on petrophysical, geochemical, and core imaging data using machine and deep learning techniques. Applied Computing and Geosciences 16, 100104. <https://doi.org/10.1016/j.acags.2022.100104>.
- Hussain, M., Liu, S., Ashraf, U., Ali, M., Hussain, W., Ali, N., Anees, A., 2022. Application of machine learning for lithofacies prediction and cluster analysis approach to identify rock type. Energies 15. <https://doi.org/10.3390/en15124501>.
- Ifrene, G., Irofti, D., Ni, R., Egenhoff, S., Pothana, P., 2023. New insights into fracture porosity estimations using machine learning and advanced logging tools. Fuels 4, 333–353. <https://doi.org/10.3390/fuels4030021>. URL: <https://www.mdpi.com/2673-3994/4/3/21>.
- Iraji, S., Soltanmohammadi, R., Matheus, G.F., Bassi, M., Vidal, A.C., 2023. Application of unsupervised learning and deep learning for rock type prediction and petrophysical characterization using multi-scale data. Geoenergy Science and Engineering 230, 212241. <https://doi.org/10.1016/j.geoen.2023.212241>.
- Jacobi, D., Gladkikh, M., LeCompte, B., Hursan, G., Mendez, F., Longo, J., Ong, S., Bratovich, M., Patton, G., Shoemaker, P., 2008. Integrated petrophysical evaluation of shale gas reservoirs. In: SPE Unconventional Resources Conference/Gas Technology Symposium, SPE, p. 114925. <https://doi.org/10.2118/114925-MS>.
- Jiang, C., Zhang, D., Chen, S., 2021. Lithology identification from welllog curves via neural networks with additional geologic constraint. GEOPHYSICS 86, IM85–IM100. <https://doi.org/10.1190/geo2020-0676.1>.
- Jiang, R., Ji, Z., Mo, W., Wang, S., Zhang, M., Yin, W., Wang, Z., Lin, Y., Wang, X., Ashraf, U., 2022. A novel method of deep learning for shear velocity prediction in a tight sandstone reservoir. Energies 15. <https://doi.org/10.3390/en15197016>.
- Katahara, K.W., 1995. Gamma ray log response in shaly sands. Log. Anal. 36. URL: <https://api.semanticscholar.org/CorpusID:132546530>.
- Keras, 2024. Randomsearch Tuner. (Accessed 30 September 2024).
- Kim, J., 2022. Synthetic shear sonic log generation utilizing hybrid machine learning techniques. Artificial Intelligence in Geosciences 3, 53–70. <https://doi.org/10.1016/j.aiig.2022.09.001>.
- Kim, K., Kim, D., Na, Y., Song, Y., Wang, J., 2023. A review of carbon mineralization mechanism during geological co<sub>2</sub> storage. Heliyon 9, e23135. <https://doi.org/10.1016/j.heliyon.2023.e23135>. URL: <https://www.sciencedirect.com/science/article/pii/S2405844023103434>.
- Laalam, A., Boualam, A., Ouadi, H., Djezzar, S., Tomomewo, O., Mellal, I., Bakelli, O., Merzoug, A., Chemmakh, A., Latreche, A., Berrehal, B.E., 2022. Application of machine learning for mineralogy prediction from well logs in the bakken petroleum system. D012S063R002. <https://doi.org/10.2118/210336-MS>.
- Lima, M.C., Pontedeiro, E.M., Ramirez, M.G., Favoretto, J., dos Santos, H.N., van Genuchten, M.T., Borghi, L., Couto, P., Raoof, A., 2022. Impacts of mineralogy on petrophysical properties. Transport Porous Media 145, 103–125. <https://doi.org/10.1007/s11242-022-01829-w>.
- Lv, A., Cheng, L., Aghighi, M.A., Massoumi, H., Roshan, H., 2021. A novel workflow based on physics-informed machine learning to determine the permeability profile of fractured coal seams using downhole geophysical logs. Mar. Petrol. Geol. 131, 105171. <https://doi.org/10.1016/j.marpetgeo.2021.105171>.
- Matinkia, M., Hashami, R., Mehrad, M., Hajsaeedi, M.R., Velayati, A., 2023. Prediction of permeability from well logs using a new hybrid machine learning algorithm. Petroleum 9, 108–123. <https://doi.org/10.1016/j.petlm.2022.03.003>.
- McDonald, A., 2021. Data quality considerations for petrophysical machine-learning models. Petrophysics – The SPWLA Journal of Formation Evaluation and Reservoir Description 62, 585–613. <https://doi.org/10.30632/PV62N6-2021a1>.
- Metwally, A.M., Mabrouk, W.M., Mahmoud, A.I., Eid, A.M., Amer, M., Noureldin, A.M., 2023. Formation evaluation of abu madi reservoir in baltim gas field, nile delta, using well logs, core analysis and pressure data. Sci. Rep. 13, 19139.
- Mohammadian, E., Kheirollahi, M., Liu, B., Ostadhassan, M., Sabet, M., 2022. A case study of petrophysical rock typing and permeability prediction using machine learning in a heterogenous carbonate reservoir in Iran. Sci. Rep. 12, 4505. <https://doi.org/10.1038/s41598-022-08575-5>.
- Moore, W.R., Ma, Y.Z., Urdea, J., Bratton, T., 2011. Uncertainty analysis in well-log and petrophysical interpretations. In: Uncertainty Analysis and Reservoir Modeling: Developing and Managing Assets in an Uncertain World. American Association of Petroleum Geologists. <https://doi.org/10.1306/13301405M963478>.
- Mukherjee, B., Kar, S., Sain, K., 2024. Machine learning assisted state-of-the-art-of petrographic classification from geophysical logs. Pure Appl. Geophys. 1–33doi: <https://doi.org/10.1007/s00024-024-03563-4>.
- O'Malley, T., Burszttein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al., 2019. Kerastuner. <https://github.com/keras-team/keras-tuner>.
- Pan, W., Torres-Verdin, C., Duncan, I.J., Pyrcz, M.J., 2023. Improving multiwell petrophysical interpretation from well logs via machine learning and statistical models. Geophysics 88, D159–D175. <https://doi.org/10.1190/geo2022-0151.1>.
- Peck, W.D., Ayashi, S.C., Klapperich, R.J., Gorecki, C.D., 2019. The North Dakota integrated carbon storage complex feasibility study. Int. J. Greenh. Gas Control 84, 47–53. <https://doi.org/10.1016/j.ijggc.2019.03.001>. URL: <https://www.sciencedirect.com/science/article/pii/S1750583618307837>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
- Pothana, P., Ifrene, G., Ling, K., 2023. Stress-dependent petrophysical properties of the bakken unconventional petroleum system: insights from elastic wave velocities and permeability measurements. Fuels 4, 397–416. <https://doi.org/10.3390/fuels4040025>. URL: <https://www.mdpi.com/2673-3994/4/4/25>.
- Pothana, P., Ifrene, G., Ling, K., 2024. Integrated petrophysical evaluation and rock physics modeling of broom creek deep saline aquifer for geological CO<sub>2</sub> storage. Fuels 5, 53–74. <https://doi.org/10.3390/fuels5010004>. URL: <https://www.mdpi.com/2673-3994/5/1/4>.
- Poupon, A., Leveaux, J., 1971. Evaluation of Water Saturation in Shaly Formations. SPWLA-1971-0.
- Radwan, A.E., Wood, D.A., Radwan, A.A., 2022. Machine learning and datadriven prediction of pore pressure from geophysical logs: a case study for the Mangahewa gas field, New Zealand. J. Rock Mech. Geotech. Eng. 14, 1799–1809. <https://doi.org/10.1016/j.jrmge.2022.01.012>.
- Raiissi, M., Perdikaris, P., Karniadakis, G., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving non-linear partial differential equations. J. Comput. Phys. 378, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>.
- Ren, J., Wang, H., 2023. Chapter 3 - calculus and optimization. In: Ren, J., Wang, H. (Eds.), Mathematical Methods in Data Science. Elsevier, pp. 51–89. <https://doi.org/10.1016/B978-0-44-318679-0-00009-0>. URL: <https://www.sciencedirect.com/science/article/pii/B9780443186790000090>.
- Rodgers, K.A., Cook, K.L., Browne, P.R.L., Campbell, K.A., 2002. The mineralogy, texture and significance of silica derived from alteration by steam condensate in three New Zealand geothermal fields. Clay Miner. 37, 299–322. <https://doi.org/10.1100/0009855023720035>.
- Sanchez, J., Ofori, Y., 2013. Petrophysical evaluation for a geothermal project in tenerife. In: Second EAGE Sustainable Earth Sciences (SES) Conference and Exhibition. European Association of Geoscientists & Engineers cp-361.
- Sayers, C.M., Dasgupta, S., 2024. Elastic properties of unconsolidated sandstones of interest for carbon storage. Geophys. Prospect. 72, 617–632. <https://doi.org/10.1111/1365-2478.13438>.
- Schuster, G.T., Chen, Y., Feng, S., 2024. Review of physics-informed machine learning

- inversion of geophysical data. *GEOPHYSICS* 0, 1–91. <https://doi.org/10.1190/geo2023-0615.1>.
- Shao, R., Wang, H., Xiao, L., 2024. Reservoir evaluation using petrophysics informed machine learning: a case study. *Artificial Intelligence in Geosciences* 5, 100070. <https://doi.org/10.1016/j.aiig.2024.100070>. URL: <https://www.sciencedirect.com/science/article/pii/S266654412400011X>.
- Silva, A.A., Tavares, M.W., Carrasquilla, A., Missagia, R., Ceia, M., 2020. Petrofacies classification using machine learning algorithms. *Geophysics* 85, WA101–WA113. <https://doi.org/10.1190/geo2019-0439.1>.
- Simandoux, P., 1963. Dielectric measurements in porous media and application to shaly formation. *Revue de L'Institut Français du Pétrole* 18, 193–215.
- Sorensen, J.A., Bailey, T.P., Dobroskok, A.A., Gorecki, C.D., Smith, S.A., Fisher, D.W., Peck, W.D., Steadman, E.N., Harju, J.A., 2009a. Characterization and modeling of the broom creek formation for potential storage of co<sub>2</sub> from coal-fired power plants in North Dakota. *Search Discov* 1, 1–14.
- Sorensen, J.A., Smith, S.A., Dobroskok, A.A., Peck, W.D., Belobraydic, M.L., Kringsstad, J.J., Zeng, Z.W., 2009b. Carbon dioxide storage potential of the broom Creek Formation in NorthNorth Dakota: a case study in site characterization for large-scale sequestration. In: Carbon Dioxide Sequestration in Geological Media—State of the Science. American Association of Petroleum Geologists. <https://doi.org/10.1306/13171244St593378>.
- Steadman, E.N., Daly, D.J., de Silva, L.L., Harju, J.A., Jensen, M.D., O'Leary, E.M., Peck, W.D., Smith, S.A., Sorensen, J.A., 2006. Plains co<sub>2</sub> reduction (pcor) partnership (phase 1) final report/july–september 2005 quarterly report. Energy and Environ. Res. Cent., Univ. Of ND, Grand Forks.
- Sun, A.Y., Yoon, H., Shih, C.Y., Zhong, Z., 2022. Applications of physicsinformed scientific machine learning in subsurface science: a survey. In: *Knowledge Guided Machine Learning*. Chapman and Hall/CRC, pp. 111–132.
- Syed, F.I., AlShamsi, A., Dahaghi, A.K., Neghabhan, S., 2022. Application of ml & ai to model petrophysical and geomechanical properties of shale reservoirs – a systematic literature review. *Petroleum* 8, 158–166. <https://doi.org/10.1016/j.petim.2020.12.001>.
- TensorFlow, 2024. Tutorials—tensorflow Core. <https://www.tensorflow.org/tutorials>. (Accessed 29 September 2024).
- Wang, X., Alvarado, V., Swoboda-Colberg, N., Kaszuba, J.P., 2013. Reactivity of dolomite in water-saturated supercritical carbon dioxide: significance for carbon capture and storage and for enhanced oil and gas recovery. *Energy Convers. Manag.* 65, 564–573. <https://doi.org/10.1016/j.enconman.2012.07.024>. URL: <https://www.sciencedirect.com/science/article/pii/S019689041200310X>. global Conference on Renewable energy and Energy Efficiency for Desert Regions 2011 “GCREEDER” 2011.
- Waxman, M., Smits, L., 1968. Electrical conductivities in oil-bearing shaly sands. *Soc. Petrol. Eng. J.* 8, 107–122. <https://doi.org/10.2118/1863-A>.
- Wood, D.A., 2020. Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data. *J. Petrol. Sci. Eng.* 184, 106587. <https://doi.org/10.1016/j.petrol.2019.106587>.
- Wyllie, M.R.J., Gregory, A.R., Gardner, L.W., 1956. Elastic wave velocities in heterogeneous and porous media. *GEOPHYSICS* 21, 41–70. <https://doi.org/10.1190/1.1438217>.
- Xu, C., Fu, L., Lin, T., Li, W., Ma, S., 2022. Machine learning in petrophysics: advantages and limitations. *Artificial Intelligence in Geosciences* 3, 157–161. <https://doi.org/10.1016/j.aiig.2022.11.004>.
- Xu, C., Misra, S., Srinivasan, P., Ma, S., 2019. When Petrophysics Meets Big Data: what Can Machine Do? <https://doi.org/10.2118/195068-MS> D041S038R002.
- Zhao, L., Liu, J., Xu, M., Zhu, Z., Chen, Y., Geng, J., 2024. Rock-physicsguided machine learning for shear sonic log prediction. *GEOPHYSICS* 89, D75–D87. <https://doi.org/10.1190/geo2023-0152.1>.