geosciences

MDPI

Article

Integration Between Well Logs and CT Information to Estimate Petrophysical Properties Through a Neural Network Model

Edwar Hernando Herrera Otero [1,*], Josep Oriol Oms Llobet [2] and Eduard Remacha Grau [2]

1 School of Petroleum Engineering, Faculty of Physicochemical Engineering, Industrial University of Santander, Bucaramanga 680002, Colombia
2 Department of Geology, Autonomous University of Barcelona, 08193 Barcelona, Spain; joseporiol.oms@uab.cat (J.O.O.L.); eduard.remacha@uab.cat (E.R.G.)
* Correspondence: edherote@uis.edu.co

Abstract

Reservoir petrophysical characterization is traditionally performed through the interpretation of well logs validated with routine core analysis (RCAL), often excluding the integration of other tools such as computed tomography (CT), which provides interpretation of higher resolution. In this study, artificial neural network (ANN) models were applied to estimate porosity and permeability by integrating conventional logs with CT-derived data (RHOB and PEF), thereby validating the petrophysical model of Ciénaga de Oro Formation. Neural networks were trained in MATLAB® using a feed-forward regression network based on a multilayer perceptron (MLP) architecture, with RCAL measurements serving as a reference. Model performance was assessed by comparing predictions with laboratory data from two wells, yielding high accuracy (R^2 = 0.98 for permeability and R^2 = 0.90 for porosity) with mean absolute errors below 5%. Additional validation was performed using well logs and CT data from complete 3 ft sections, with the trained models successfully reproducing core heterogeneities at millimetric resolution. These results confirm the potential of integrating well logs and CT data with ANN to enhance petrophysical characterization and extend property estimation to wells lacking core or laboratory measurements. Furthermore, an interactive MATLAB® tool was developed, enabling users to load well logs and CT files as flat inputs, generate high-resolution predictions, validate results, and export the estimated values.

Keywords: artificial neural networks (ANN); Computed tomography (CT); machine learning; petrophysical characterization; well logs integration; routine core analysis (RCAL)

Academic Editors: Beatriz Menéndez and Agust Gudmundsson

Received: 9 October 2025
Revised: 26 December 2025
Accepted: 29 December 2025
Published: 31 December 2025

1. Introduction

Petrophysical reservoir characterization involves evaluating the spatial distribution of rock properties such as porosity, permeability, and saturation [1]. This process requires extensive interpretation and analysis of measurements obtained from multiple sources, including well logs, computed tomography (CT), cores, seismic data, and well tests. Conventionally, this information is derived from conventional well logs and, in selected intervals, from routine core analysis (RCAL), which provides the most reliable porosity and permeability measurements. However, extensive coring is costly and typically limited to a few wells and specific intervals, restricting the continuity and resolution of petrophysical models [2]. Consequently, models have been developed to estimate petrophysical properties from well logs in the absence of direct core data.

In addition, X-ray computed tomography has been applied in the industry as a non-destructive technique for reservoir rock description and characterization. Its advantage lies in acquiring high-resolution cross-sectional images without damaging the integrity of the sample, based on the differential attenuation of X-rays, which depends on the electron density of the material [3]. From CT images, physical properties such as bulk density (RHOB, g/cm^3) and photoelectric factor (PEF, barns/e$^-$) can be extracted as individual logs. These properties are closely related to mineralogy, porosity, permeability, and rock type. As a result, CT provides petrophysical information at millimetric resolution, one order of magnitude higher than conventional well logs.

Although advances have been made in using logs and CT for petrophysical characterization, there remains a need for methods that integrate both data sources to improve resolution and accuracy throughout the reservoir volume. This need arises from the practical and economic limitations of core acquisition and the growing demand for high-resolution petrophysical models for hydrocarbon production optimization.

To address these constraints, artificial neural networks (ANN) have emerged as effective tools for predicting petrophysical properties from multivariate datasets [4]. Their ability to capture nonlinear relationships between geophysical measurements and rock properties makes them particularly suitable for integrating logs and CT data. Algorithms such as linear regression, support vector regression (SVR), random forest (RF), ensemble methods, and deep learning have been shown to enhance accuracy over traditional frameworks, primarily in geologically complex environments.

Recent studies demonstrate progress in this field; numerous studies have applied artificial neural networks and other machine learning architectures to improve petrophysical characterization, such as ANN models predicting porosity from CT derived logs (RHOB and PEF) at 0.625 mm resolution in Colombian Andean rocks [4], multimodal deep learning approaches for direct inversion of geophysical logs [5], physics-informed neural networks (PINN) for porosity estimation from seismic data incorporating rock and wave physics [6], and physics aware CNN for predicting permeability directly from 3D CT rock images [7]. Similarly, a pervious study proposed a hybrid CNN–LSTM–PSO framework to capture both spatial and temporal dependencies in well logs [8]. Other works combined ANN and support vector machines (SVM) for permeability prediction in carbonate rocks, achieving robust results even with limited core data [9], or they employed ensemble or deep learning methods with micro-CT to predict porosity and permeability at core scale [10,11], while hybrid ANN–Random Forest (RF) models have been used to estimate fracture intensity from petrophysical logs [12] and further enhance predictive accuracy by integrating multi-level ensemble learning features [13].

These advances confirm the relevance of combining conventional logs and CT through neural network models to improve petrophysical estimation. However, limitations remain, including focusing on a single property, reliance on small datasets or mono-well applications, potential overfitting in complex models, lack of cross-well and cross-field validation, and challenges in extrapolating results in highly heterogeneous reservoirs.

Consequently, the current research is focused on developing hybrid and probabilistic frameworks, and it proposes a methodology for estimating porosity and permeability at millimetric resolution by simultaneously integrating conventional well logs with CT derived data using a multilayer perceptron (MLP) model. Unlike approaches focused on a single property, this method validates both porosity and permeability predictions against RCAL measurements. Validation demonstrates the model's ability to operate across dual scales: logs and CT, capturing fine scale heterogeneities that impact fluid flow and reservoir simulation. Moreover, the study shows the applicability of the methodology in wells lacking

https://doi.org/10.3390/geosciences16010021

core information or in non-cored intervals, providing a more comprehensive reservoir characterization and complementing traditional petrophysical evaluation methods.

2. Theoretical Framework: Artificial Neural Networks

Artificial Neural Networks (ANN) are computational models capable of inductively identifying patterns through learning algorithms based on training datasets, without requiring explicit modeling by the user [14]. Their application in geosciences has consolidated over the past decade, as they enable the capture of nonlinear relationships in multivariate data. Moreover, their use as a tool for well-log evaluation and for predicting petrophysical properties has proven equally, or even more, effective than conventional statistical methods [15].

There are different ANN architectures (Figure 1). However, the learning model required for training a neural network depends on the type of information to be processed. Machine learning models can handle linear data (numerical values), and their architecture is based on Multilayer Perceptrons (MLP).

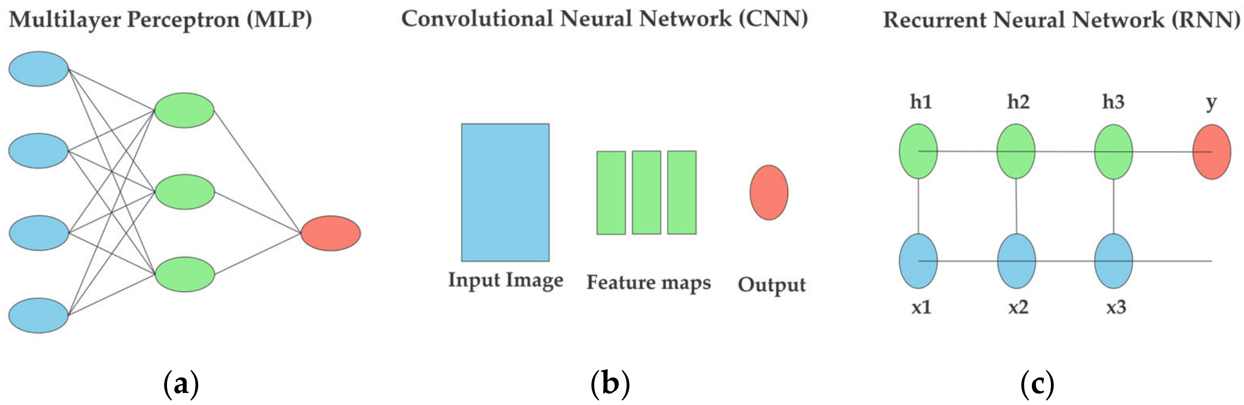


Figure 1. Schematic comparison of three neural network architectures commonly applied in geosciences. (**a**) Multilayer Perceptron (MLP): designed for tabular numerical data such as well logs. (**b**) Convolutional Neural Network (CNN): applied to spatial data such as images, (**c**) Recurrent Neural Network (RNN): suitable for sequential data such as well-log series. In this study, the MLP architecture was selected given the nature of the dataset (well logs and CT-derived curves), which does not require spatial or sequential processing.

Although there are more complex deep learning architectures that are designed to perform more sophisticated processes—such as Convolutional Neural Networks (CNN), mainly used for processing spatial data (e.g., images), and Recurrent Neural Networks (RNN), designed to process sequential data (e.g., text)—for this study, MLPs were selected. This choice is justified because they provide an adequate balance between computational simplicity and predictive capacity, in addition to being the most suitable architecture for integrating conventional well logs with numerical data derived from CT.

The majority of ANN applications in geosciences rely on MLP architectures [15], as they are widely used for numerical prediction tasks involving tabular datasets [16]. MLP networks are organized in a feed-forward structure composed of three types of layers: input, hidden, and output. Each layer consists of a set of neurons, and every neuron is fully connected to the neurons in the subsequent layer. Each connection is weighted, and the weights are iteratively adjusted during the training process. A typical representation of this network architecture is shown in Figure 2.

The optimization of network weights is commonly performed using backpropagation algorithms (Backpropagation Neural Network, BPNN) within a supervised learning scheme, in which a set of input–output vectors is employed to train the network. Among these, the Levenberg–Marquardt algorithm is widely used due to its efficiency in solving nonlinear regression problems [16].

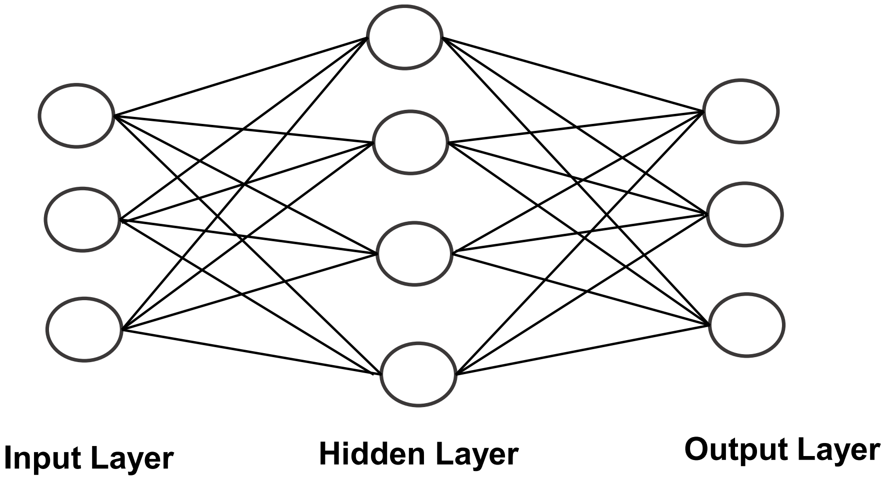


Figure 2. Simple architecture MLP (Multilayer perceptron).

Previous studies have demonstrated the effectiveness of MLP networks in predicting petrophysical properties from conventional well logs, including gamma ray (GR), deep resistivity (LLD), neutron porosity (NPHI), and bulk density (RHOB) [1,15]. The outputs of these BPNN models include lithology classification, porosity, and permeability estimation. More recently, successful applications combining ANN with CT-derived data have been reported, achieving porosity and permeability predictions at resolutions higher than those obtained with traditional well logs [4,11].

The main challenge in training ANN models lies in avoiding overfitting, in which the network memorizes the training examples and loses its ability to generalize to unseen data [1]. Overfitting typically occurs when the number of training iterations exceeds the optimal range. To mitigate this issue, the dataset is divided into three subsets: the training set, used to compute gradients and update weights; the validation set, which defines the stopping point of training since its error increases when overfitting begins; and the test set, which evaluates the generalization performance of the final model. This procedure allows continuous monitoring of model performance and ensures that training stops once the validation error increases [1].

Recent advances in petrophysical-property estimation can be grouped into three principal research directions: (i) empirical and regression-based methods relying on conventional well logs, (ii) high-resolution imaging approaches based on computed tomography (CT) and micro-CT that enable millimetric to micron-scale characterization, and (iii) machine-learning frameworks, particularly artificial neural networks (ANN) and hybrid deep-learning models, that capture complex nonlinear relationships among logs, images, and laboratory measurements.

Table 1 summarizes representative studies, indicating data sources, applied methods and applications, predicted properties, operative resolution/scale, and principal advantages and limitations. This structured overview highlights recurrent gaps in the literature—single-property focus, limited multi-scale integration, and difficulties in scaling core-derived results to well and field scales and positions the present work as addressing these gaps by integrating conventional well logs and CT-derived features within an ANN (MLP) framework validated against RCAL in two wells and extended across millimetric resolutions.

Table 1. Summary of representative previous studies on petrophysical property estimation.

Author	Data Source	Method & Application	Predicted Property	Resolution/Scale	Main Advantage/Limitation
[4]	RHOB-CT & PEF-CT (core CT scans, Colombian Andes)	ANN (MLP) to predict porosity from CT-derived logs (RHOB, PEF)	Porosity	~0.625 mm (CT-based)	Advantage: high-resolution porosity estimation from CT logs. Limitation: single-property focus; regional dataset.
[17]	Conventional well logs (multi-well datasets)	Deep RNN (bidirectional LSTM cascaded with FC) for missing-log prediction	Well-log curves (reconstruction)	well scale (depth-series)	Advantage: effective reconstruction of missing logs /leverages sequential dependence. Limitation: focus on log synthesis (not direct petrophysical inversion).
[8]	Conventional well-logging data	Hybrid CNN–LSTM–PSO model for well-log prediction (temporal + spatial features)	Well-log curves (e.g., PE)	well scale	Advantage: captures spatial + temporal patterns; optimized by PSO. Limitation: not using CT/limited to log interpolation/prediction.
[18]	Whole-core CT images (2D slices)	CNN for lithology classification from core CT scans	Lithofacies (classification)	core/mm-scale	Advantage: proven CNN workflow for CT images and lithology mapping. Limitation: classification task (not continuous φ/k prediction).
[10]	Micro-CT slices (carbonate plugs)	Stacked ensemble ML (multiple learners + meta-learner) for φ and absolute k prediction	Porosity, permeability	core/µm–mm (micro-CT)	Advantage: ensemble increases generalizability for core-scale φ & k. Limitation: core-scale results—scaling to wells/field is non-trivial.
[7]	3D rock images (micro-CT)	3D CNN/physics-aware CNN for permeability prediction from image geometry	Permeability	core/3D image scale	Advantage: fast end-to-end permeability estimation from 3D images. Limitation: requires 3D imaging and computational resources.
[19]	Well logs + core	ANN (backpropagation) for permeability prediction using well logs (numerical regression)	Permeability	well/core-linked	Advantage: straightforward ANN regression for k from logs. Limitation: depends on quality and quantity of core calibration points.
[20]	3D micro-CT images	Physics-aware/simulation-informed methods (CFD + ML approximations) for permeability estimation	Permeability	core/3D image scale	Advantage: integrates physical modelling with ML for robust k estimation. Limitation: high computational cost; core-scale.

3. Methodology and Case Study

The methodology was designed to integrate conventional well-log data with computed tomography (CT)-derived curves to estimate porosity and permeability using artificial neural networks (ANN). The selected architecture was a multilayer perceptron (MLP), implemented as a feed-forward regression network, which is particularly suitable for predictive tasks involving numerical input–output relationships.

The feed-forward regression models were developed in MATLAB® 2017a, which offers the advantage of enabling rapid model evaluation through regression parameters and performance plots. For each model, the objective was to identify the architecture and topology that best fit the training dataset while ensuring accurate prediction of petrophysical properties during validation.

The implemented workflow followed four sequential phases: (i) data preprocessing and normalization to ensure consistency between conventional and CT-derived curves; (ii) dataset construction by dividing the information into training, validation, and testing subsets; (iii) network design and training, where weights were adjusted using backpropagation to minimize error; and finally, (iv) prediction and validation, where the trained network was applied to integrate log and CT data, with results benchmarked against routine core analysis (RCAL) measurements.

This process enabled the development of ANN-based models capable of reproducing petrophysical properties with high fidelity across both well logs and CT-derived datasets, as schematically summarized in Figure 3.

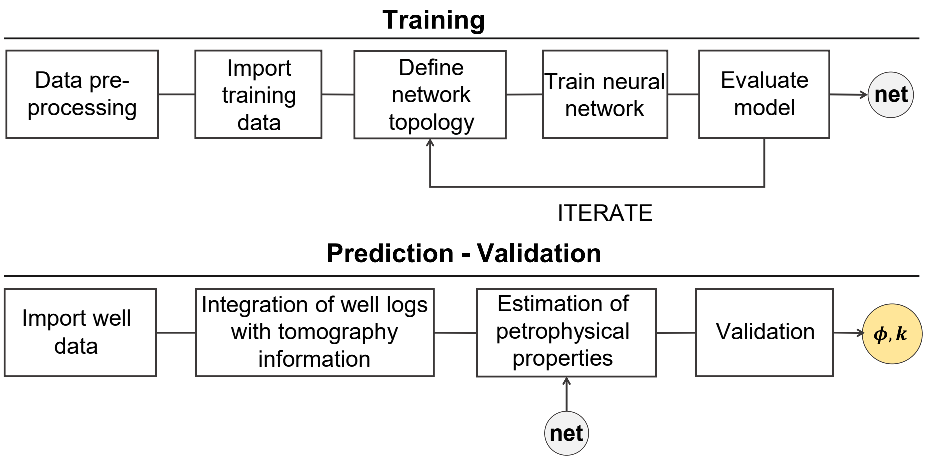


Figure 3. Schematic workflow of the artificial neural network (ANN) implementation for petrophysical property estimation. The process begins with preprocessing and normalization of conventional well logs (GR, SP, NPHI, RHOB) and CT-derived logs (RHOB-CT, PEF-CT), followed by dataset construction and division into training, validation, and testing subsets. The selected multilayer perceptron (MLP) feed-forward network is then trained using backpropagation, optimizing the weights until convergence. The trained model (net) is subsequently applied to integrate well-log and CT data for porosity (φ) and permeability (k) prediction, validated against routine core analysis (RCAL) measurements.

The methodology was applied to two study wells. Each dataset integrated well logs, tomographic data, and core-based petrophysical measurements. The analyzed intervals in both wells belong to the Ciénaga de Oro Formation (Sinú-San Jacinto Basin, Colombia) and are characterized by a predominantly siliciclastic lithology. The reservoir units are mainly composed of fine to medium grained sandstones with variable clay content, interbedded with thin shale layers. Core descriptions and petrophysical responses indicate comparable depositional facies and mineralogical composition across both wells, resulting in similar porosity and permeability ranges, and rock fluid behavior. This lithological consistency provides a controlled geological framework for evaluating the proposed ANN-based workflow and explains the coherent predictive performance observed between the two datasets.

Although an extensive amount of well-log and high-resolution CT data is available, the effective size of the supervised learning dataset is constrained by the number of laboratory-measured RCAL porosity and permeability samples. After neural network training, porosity and permeability were estimated across the entire well intervals, with high-resolution log-derived predictions validated against laboratory core measurements.

3.1. Data Pre-Processing and Normalization

Conventional well logs (GR, SP, NPHI, RHOB), CT-derived logs (RHOB-CT, PEF-CT), and petrophysical properties from routine core analysis (RCAL) were extracted as flat files. Normalization is recommended for the MLP when the training data covers a wide range of numerical values. For porosity modeling, normalization proved unnecessary. However, given the wide variability of permeability, a logarithmic transformation was applied to both input porosity values and output values for permeability model, which improved the stability of the training process.

Because of the resolution mismatch between conventional logs and CT data within a given interval, prior data integration was required before applying the trained neural network model. This step ensures that any new data set used for future predictions maintains the same number of samples across all data types. Integration was performed by upscaling the well-log data (~0.25 ft) to the resolution of the CT scans (~0.6 mm) through interpolation. The interpolation used depth values and measured log curves as known points to calculate new values aligned with the higher-resolution CT depths. This procedure guaranteed that both data sources fed into the model at the same sampling scale, achieving consistent integration while preserving data reliability. The interpolation results for the porosity model are shown in Figure 4, while the interpolation results for permeability model are shown in Figure 5.

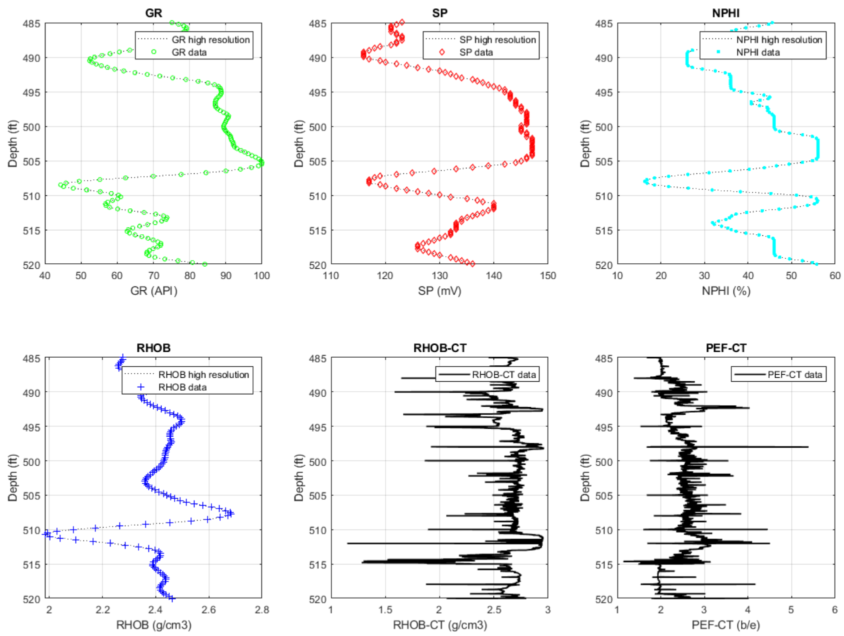


Figure 4. Results of the interpolation of well logs at tomography scale for the porosity model.

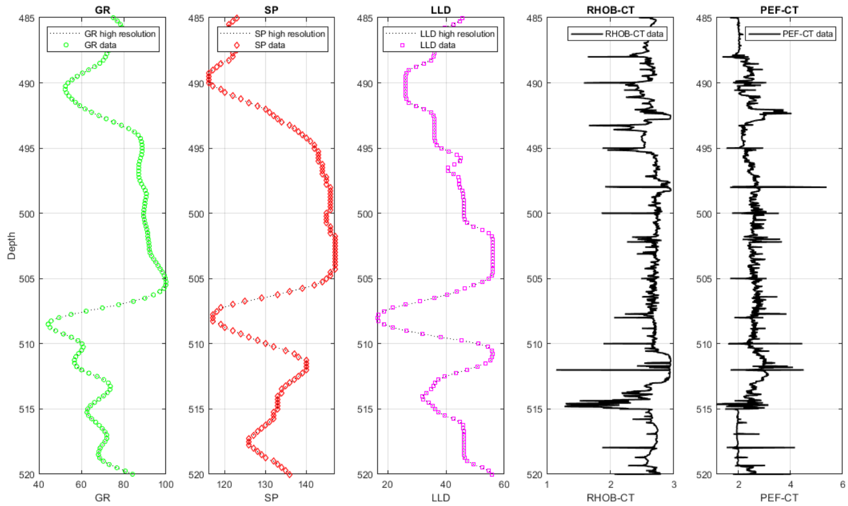


Figure 5. Results of the interpolation of well logs at tomography scale for the permeability model.

3.2. Training Dataset Construction

The training datasets were built for two wells and included conventional well logs, CT-derived parameters as input, and laboratory routine core analysis (RCAL) measurements as the output objective. For Well 1, the dataset comprised 61 porosity samples and 57 permeability samples within the interval 0–2180 ft, while Well 2 included 125 porosity samples and 119 permeability samples from 0–2270 ft. A summary of this dataset distribution is presented in Table 2.

Table 2. Summary of well-log, CT-derived, and RCAL data available for the study wells.

Data Type	Well 1: ANH-SSJ-La Estrella-1X	Well 2: ANH-SSJ-Nueva Esperanza-1X
Well logs	Lithological, petrophysical, and resistivity logs (0–2180 ft); vertical resolution 0.25 ft; 8721 data per log	Lithological, petrophysical, and resistivity logs (0–2270 ft); vertical resolution 0.25 ft; 9081 data per log
CT scans	303 cores, 899 sections (3 ft each) from 6–2190 ft; resolution ~0.6 mm; 1524 values per section; 1,090,936 total values	309 cores, 920 sections (3 ft each) from 9–2266 ft; resolution ~0.6 mm; 1524 values per section; 1,133,123 total values
RCAL	Porosity: 61 measurements and permeability: 57 measurements (199.8–2180.75 ft)	Porosity: 125 measurements and permeability: 119 measurements (89.4–2258.5 ft)

The tested configurations demonstrated that porosity model performed best when combining lithological (GR, SP) and petrophysical logs (RHOB, PEF), in addition to CT-derived logs (RHOB-CT, PEF-CT), as inputs, while permeability models achieved higher accuracy when using two lithological logs (GR, SP) and one resistivity log (LLD), consistent with previous findings [1]. For the permeability model, porosity (ϕ data) was included in the input dataset. The complete structure of the input–output datasets is presented in Table 3.

Table 3. Input and target datasets used in porosity and permeability prediction models.

Prediction Model	Input Dataset	Target Output Dataset
Porosity	Well logs: GR (API), SP (mV), NPHI (v/v), RHOB (g/cm^3); CT-derived curves: RHOB-CT (g/cm^3), PEF-CT ($barn/e^-$).	Porosity data
Permeability	Well logs: GR (API), SP (mV), LLD ($\Omega{\cdot}m$); CT-derived curves: RHOB-CT (g/cm^3), PEF-CT ($barn/e^-$), Porosity data.	Permeability data

3.2.1. Well Logs

Well logs consisted of lithological GR (API), SP (mV), petrophysical NPHI (v/v), RHOB (g/cm^3), and resistivity curves LLD ($\Omega\cdot$m) acquired at a vertical resolution of 0.25 ft, corresponding to 8721 data points per log in Well 1 and 9081 in Well 2. Well logs were subjected to a quality control workflow including depth consistency checks, detection and correction of outliers, and unit/format standardization [21,22]. To estimate porosity from petrophysical logs, the following equations are used:

- Shale volume (Vshale): estimated from gamma ray (GR) logs using the Larionov model [23] equation for Tertiary rocks and validated in selected intervals with core mineralogy and spectral gamma analysis.

$$\text{Vshale} = 0.083 * \left(2^{3.7\text{GR}} - 1\right), \tag{1}$$

- Porosity (∅): Total porosity computed from density (RHOB) is more accurate than neutron-based methods due to environmental corrections; the following equation was used to calculate the porosity from density log (RHOB) [24].

$$\varnothing\text{D} = \frac{\text{RHOBma} - \text{RHOBl}}{\text{RHOBma} - \text{RHOBf}} \tag{2}$$

where:

∅D = Total porosity calculated from the density log
RHOBma = Formation matrix density (2.65 & 2.71 g/cc)
RHOBf = Fluid density (1 g/cc)
RHOBl = Log value (g/cc)
Then, we proceed to calculate the effective porosity with the following equation:

$$\text{PHIE} = \varnothing\text{D} * (1 - \text{Vshale}), \tag{3}$$

where PHIE = Effective porosity.

3.2.2. Tomographic Data

This study utilized RHOB and PEF values extracted from cross-sectional images of over 300 core samples, spanning a vertical interval exceeding 2000 ft. Core samples were scanned with a medical tomography scanner at vertical resolution of approximately 0.6 mm, enabling high-resolution petrophysical characterization. CT-derived information was obtained from individual core scans, which resulted in 1,090,936 data points for Well 1 and 1,133,123 data points for Well 2.

3.2.3. Routine Core Analysis (RCAL)

Routine Core Analysis (RCAL) is traditionally the tool that provides the most reliable porosity and permeability measurements in the petrophysical field. This data served as baseline values (output target) for training the artificial neural network (ANN) models used in predicting petrophysical properties.

The RCAL data were obtained through laboratory measurements of absolute porosity and gas permeability in core samples from the two wells under study: for the ANH-SSJ-La Estrella-1X well (Well 1), 61 samples were used for the porosity model, while for the permeability model, 57 samples were used, with 4 samples being discarded because, although they had porosity measurements, they did not have permeability measurements because they were fractured; for the ANH-SSJ-Nueva Esperanza-1X well (Well 2), 125 samples were used for the porosity model, while for the permeability model, 119 samples were used, with 6 samples being discarded because they were fractured. These data cover specific depth intervals (e.g., 199.8–2180.75 ft for Well 1 and 89.4–2258.5 ft for Well 2).

The samples were prepared and analyzed in accordance with API RP 42 standards [25]. Absolute permeability was determined by using a gas permeameter, applying a controlled pressure gradient and measuring the steady-state gas flow through the core. The results were computed according to Darcy's law corrected for gas compressibility, considering the mean pressure and gas viscosity. A Klinkenberg correction was then applied to extrapolate to the non-slip condition and obtain the equivalent liquid permeability.

Laboratory measurements of porosity and permeability were quality-controlled using porosity-permeability cross-plots. Non-representative values were filtered, and depth-specific correlations were established between core data, well logs, and tomographic information. These RCAL data served as the reference dataset for validating the neural network models.

3.3. Neural Network Topology and Training

With the training dataset defined for prediction models, the optimal neural network topology was determined to ensure the model's accuracy in fitting the training data and its applicability to the target well conditions. With the training dataset defined, the next step was to determine the most appropriate network topology to obtain the model that best fits the training data while remaining suitable for the well conditions to be evaluated.

The MATLAB® Machine Learning Toolbox was used to configure the following topology-defining elements prior to training. These elements are: (i) Network structure: Determines the number of layers and neurons per layer. The hidden layer architecture is tuned by well and property to achieve the best model performance, prioritizing tuning of the validation set. (ii) Training algorithm: Backpropagation of Levenberg–Marquardt, adjusts network parameters based on input-output training pairs, efficient in solving non-linear regression problems and processing speed. (iii) Activation function: hyperbolic sigmoid tangent transfer function, governs signal propagation between layers, used in the neurons of each hidden layer, it handles an output range between −1 and 1 and (iv) Data partitioning: Splits data into training.

Training pseudocode for each model is presented in Appendix A. Appendix A.1: porosity and Appendix A.2: permeability. This study implemented identical topologies for both porosity and permeability prediction models, except for the number of neurons in the hidden layers, which varied between cases for each well. The general architecture used is described for porosity model in Table 4 and for permeability model in Table 5 below:

Table 4. Neural Network Topology for Porosity Prediction Model.

Topology	Description
Architecture	6 hidden layers, well 1:10 neurons per layer, well 2: 20 neurons in the 3 first layers and 12 neurons in the last 3 layers.
Training algorithm	Backpropagation of Levenberg–Marquardt
Activation function	Hyperbolic sigmoid tangent transfer function
Dataset distribution	Training: 70% Validation: 15% Test: 15%

The training process was executed in MATLAB® using the parameters defined in the selected network topology. The division of the dataset into training, validation, and testing subsets was performed internally by the MATLAB® Neural Network Toolbox. The software automatically applies a random split using the default partition of 70% for training, 15% for validation, and 15% for testing. This ensures that the validation and test subsets remain independent from the training process, preventing data leakage and allowing an unbiased evaluation of model generalization.

Table 5. Neural Network Topology for Permeability Prediction Model.

Topology	Description
Architecture	6 hidden layers, well 1: 20 neurons in the 3 first layers and 10 neurons in the last 3 layers, well 2: 12 neurons per layer.
Training algorithm	Backpropagation of Levenberg–Marquardt
Activation function	Hyperbolic sigmoid tangent transfer function
Dataset distribution	Training: 70% Validation: 15% Test: 15%

Randomization is executed at each training initialization, which minimizes potential bias due to ordering effects in the input data. Once completed, model performance was evaluated by analyzing regression plots for the training, validation, and testing subsets. Multiple iterations were performed with either fixed configurations or modified network characteristics until the best-performing prediction model was identified.

Model selection ultimately relied on the user's evaluation; therefore, in addition to regression parameters and performance plots, it was recommended to validate predictions within a defined interval by comparing estimated values with laboratory-measured data using the trained network.

The training phase produced the final ANN model, which could then be applied to future predictions and further evaluated through validation plots. Training was carried out iteratively to optimize the balance between training and validation errors, thereby avoiding overfitting. The final model selection was not based solely on minimizing error in the training or total datasets but prioritized the performance of the validation set, since this measure best reflects the ability of the model to generalize new data.

In the validation plots and in the performance metrics reported by MATLAB®, the value of R corresponds to the linear correlation coefficient between the network-estimated values and the laboratory-measured (RCAL) values. MATLAB computes this parameter directly within the Neural Network Toolbox as:

$$\mathrm{R} = \frac{cov\ (T, Y)}{\sigma T\ \sigma Y} \tag{4}$$

where

T represents the target (RCAL) values.

Y denotes the network predictions.

σT is the standard deviation of the target RCAL values.

σY is the standard deviation of the ANN predictions.

This coefficient quantifies the degree of alignment between both datasets. Standard deviations normalize the covariance term, ensuring that the correlation coefficient R always ranges between -1 and 1. Although it is not strictly equivalent to a statistical R^2, MATLAB interprets R as an indicator of the regression fit between predicted and true values. Values close to 1 indicate robust model performance and confirm the ANN's capability to capture the nonlinear relationships between input features and petrophysical properties. For additional details, see the MATLAB® Neural Network Toolbox documentation.

3.4. Petrophysical Property Estimation and Results Validation

The prediction–validation phase was implemented through a compiled tool in MATLAB®, which allowed loading input data as flat files, integrating them by interpolation, and estimating the target petrophysical property using an embedded neural network. The tool also permitted the user to import an alternative trained network if superior prediction performance was achieved. Once the estimation was completed, the results could be evaluated through validation plots and exported at the resolution of the CT-derived data.

The compiled tool integrated conventional well logs with CT-derived information to feed the trained ANN, thereby generating porosity or permeability predictions for the interval of interest. Additionally, the tool offered the option to upload laboratory measurements, enabling direct validation of estimated values against RCAL data. The validation process included joint evaluation of well logs, CT curves, and the predicted values, which were compared with measured data when available.

Model performance was assessed using regression plots and statistical metrics, including the coefficient of determination (R), root mean squared error (RMSE), and mean absolute error (MAE). External validation covered intervals exceeding 2000 ft in both wells, where predicted values were benchmarked against RCAL laboratory measurements to ensure accuracy and robustness of the trained ANN models.

4. Results

The ANN models trained for porosity and permeability showed a high fit with the laboratory data (RCAL) in both study wells. The coefficients of determination (R^2) exceeded 0.89 in all cases, reaching a maximum of 0.98 for permeability. Prediction errors remained low, with RMSE values < 0.05 for porosity and < 0.1 on a logarithmic scale for permeability (Table 6).

Table 6. Performance of ANN models in training and validation.

Well	Property	R^2 (Train)	R^2 (Val)	RMSE
1	Porosity	0.90	0.88	0.045
1	Permeability	0.98	0.95	0.090 *
2	Porosity	0.89	0.87	0.048
2	Permeability	0.92	0.90	0.110 *

* Permeability values calculated on a logarithmic scale.

4.1. Evaluation of Petrophysical Property Prediction Models

In all cases, good results were obtained during neural network training in the first three and four iterations. Regression graphs and coefficients of determination are used to evaluate prediction models, comparing the estimated values with the direct laboratory measurements used to train the neural networks.

Next, the results of the porosity and permeability prediction models are compared for the two case studies. The regression graphs confirm the good performance of the models, showing that the estimated values closely align with the laboratory measurements.

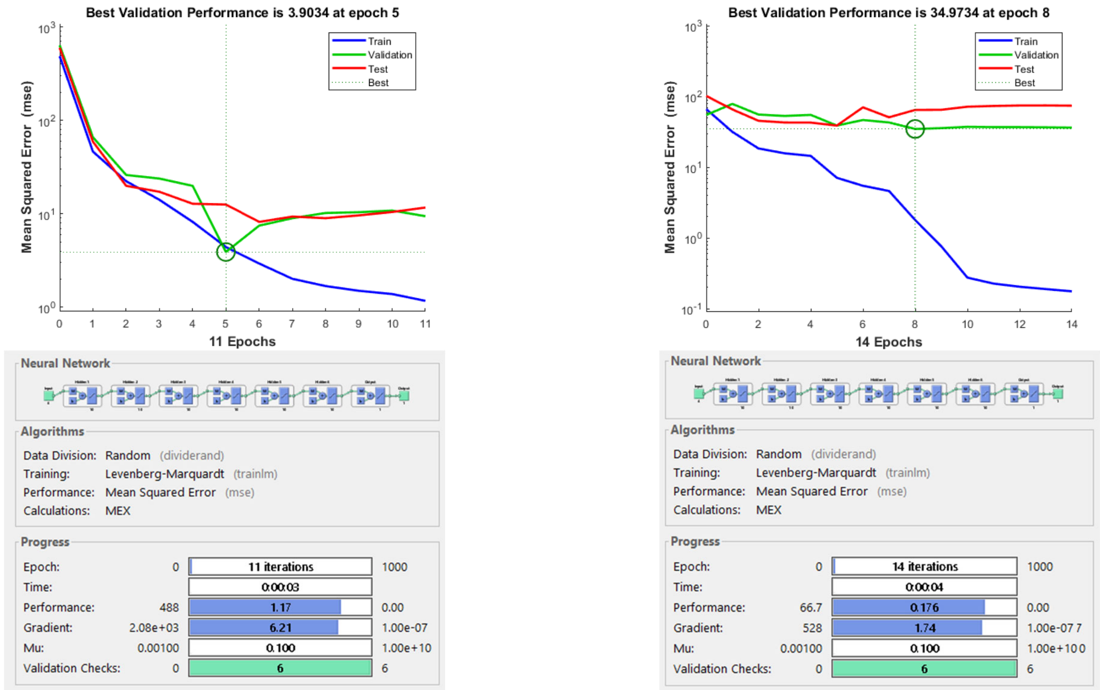
4.1.1. Porosity Prediction Model

The results obtained iteratively for the porosity model are shown in Table 7. The trained model that showed the best fit of the validation set for well 1 is the one obtained in iteration #2. Meanwhile, the trained model that showed the best fit of the validation set for well 2 is the one obtained in iteration #4.

Figure 6 summarizes the internal training process for the porosity-well 1 MLP network using the MATLAB® nntraintool interface, including the topology, training algorithm (trainlm), random data division, and early stopping behavior. The stopping criterion was reached based on validation checks (consecutive increases in validation MSE). The figure displays the network architecture, training algorithm, performance metric (MSE), gradient, damping factor (Mu), and the evolution of the error during training.

Table 7. Regression parameters for training the porosity prediction full input dataset model based on tomography records in different iterations. Source: Own elaboration.

Well	Run	Training R^2	Validation R^2	Test R^2	Total R^2
1	1	0.646	−0.111	0.51	0.482
	2	0.918	0.976	0.617	0.901
	3	0.997	0.73	0.807	0.907
2	1	0.955	0.961	0.863	0.87
	2	0.998	0.726	0.858	0.938
	3	0.997	0.869	0.957	0.947
	4	0.992	0.968	0.867	0.981
	5	0.969	0.507	0.76	0.835

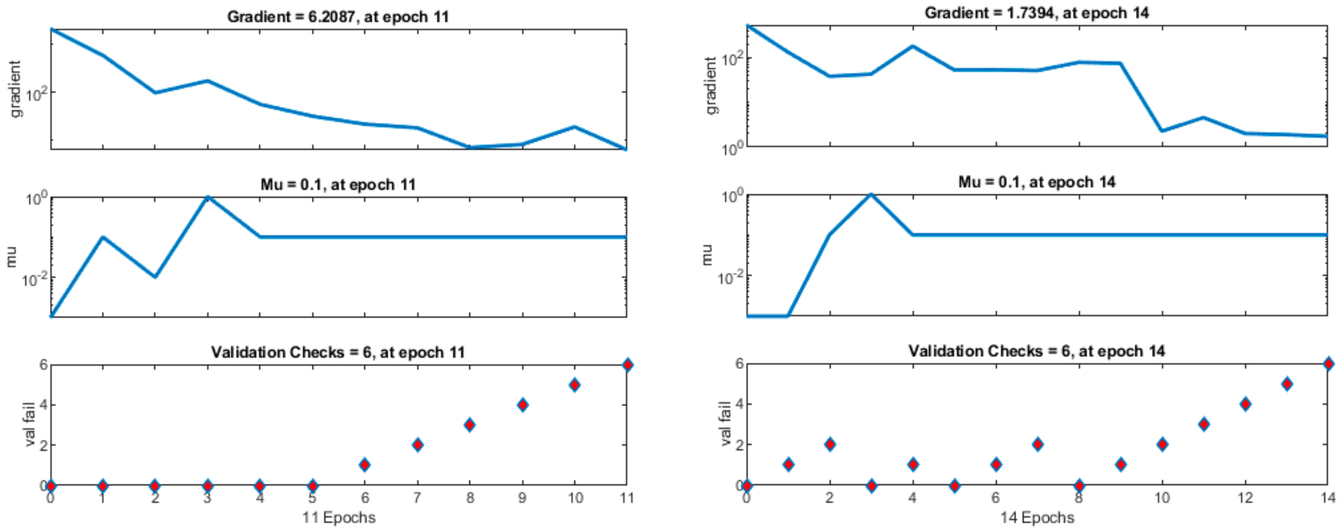


(**a**) (**b**)

Figure 6. Training outs of the MATLAB® Neural Network Training Tool (nntraintool) for the Porosity−well 1 MLP model. (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.

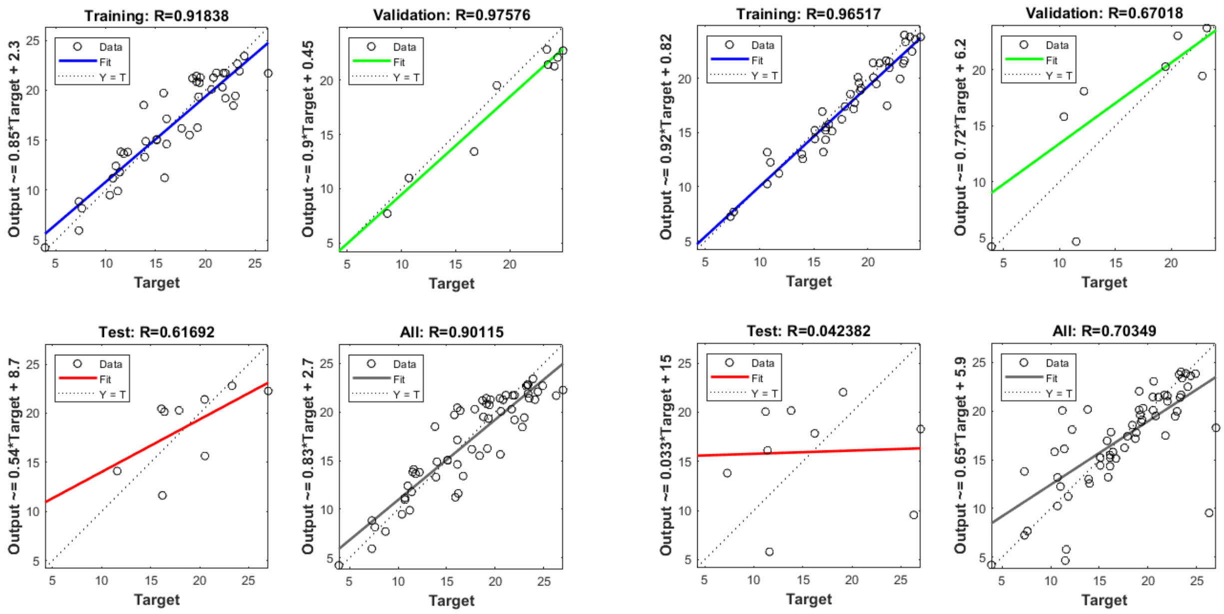
The evolution of the gradient, Mu, and validation checks during training for porosity-well 1 model (Figure 7) illustrates the stability of the optimization process and the point at which MATLAB® triggered early stopping after six consecutive validation failures, preventing overfitting. The gradient decreases consistently over the training epochs, indicating stable convergence of the optimization process. The Mu parameter transitions toward 0.1, reflecting the adaptive behavior of the Levenberg–Marquardt update rule. Validation failures increase to six consecutive checks at epoch 11, triggering MATLAB®'s early stopping criterion and preventing overfitting.

The regression graphs for the porosity model shown in Figure 8 show that models fit in the training, validation and test sets for well 1. Each subplot compares laboratory-measured porosity (x-axis) against the ANN-estimated porosity (y-axis), allowing visual evaluation of the regression performance. The dotted line represents the ideal fit (Y = T), while the solid lines correspond to the linear regression obtained for each subset. These results confirm consistent full input dataset model performance, demonstrating strong correlation values (0.90), indicating robust generalization and absence of overfitting.



(**a**) (**b**)

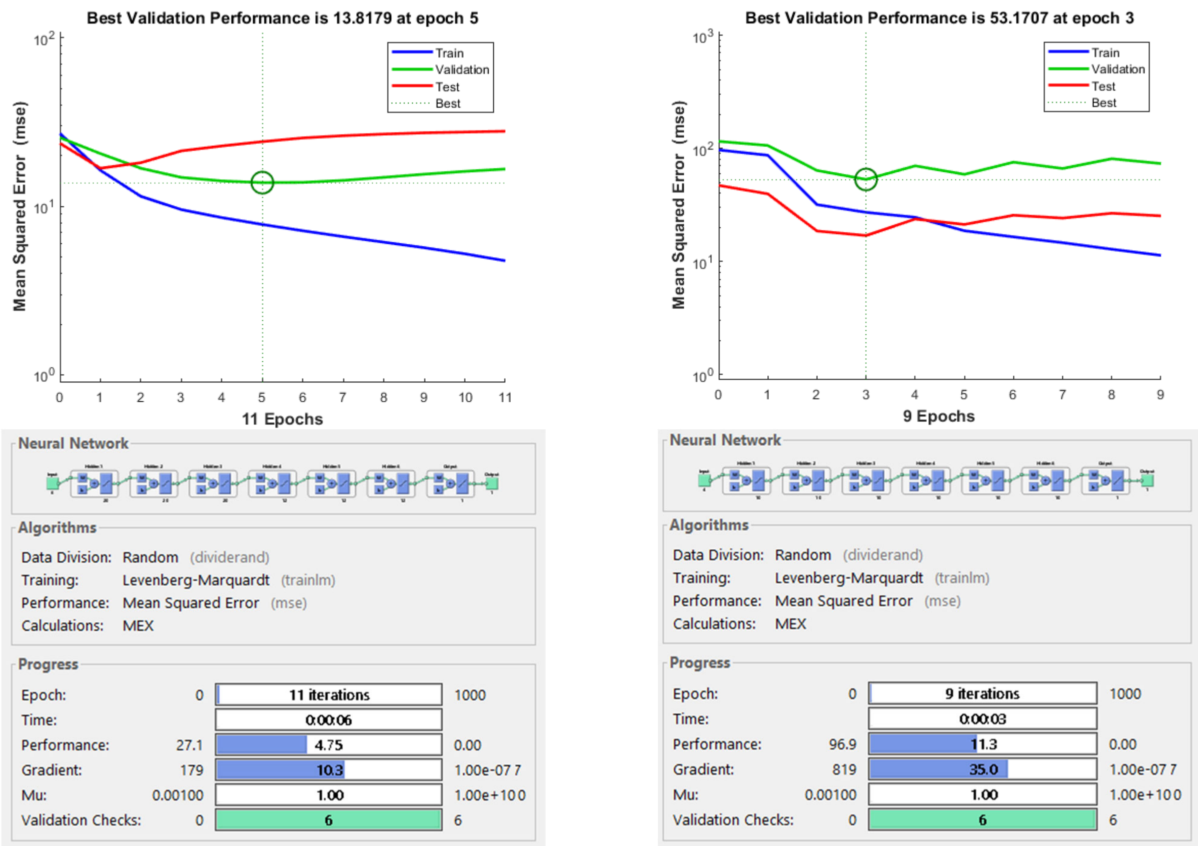
Figure 7. Gradient, Mu, and validation-failure evolution during ANN training using the Levenberg–Marquardt algorithm (trainlm) for Porosity well 1 model. (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.



(**a**) (**b**)

Figure 8. Regression graphs from the training dataset of the porosity prediction model. Well 1. The estimated values show a coefficient of determination of Regression graphs $R^2 \approx 0.90$ compared to laboratory measurements (RCAL). (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.

Figure 9 summarizes the internal training process for the porosity-well 2 MLP network using the MATLAB® nntraintool interface. The training diagnostics indicate that the model converged efficiently and with stable learning dynamics. The best validation performance at epoch 5 indicates that the early stopping criterion effectively prevented overfitting. The proximity of the validation and test curves suggests good generalization capability. Overall, the model achieved a reliable fit to the nonlinear relationship between the input logs and porosity.



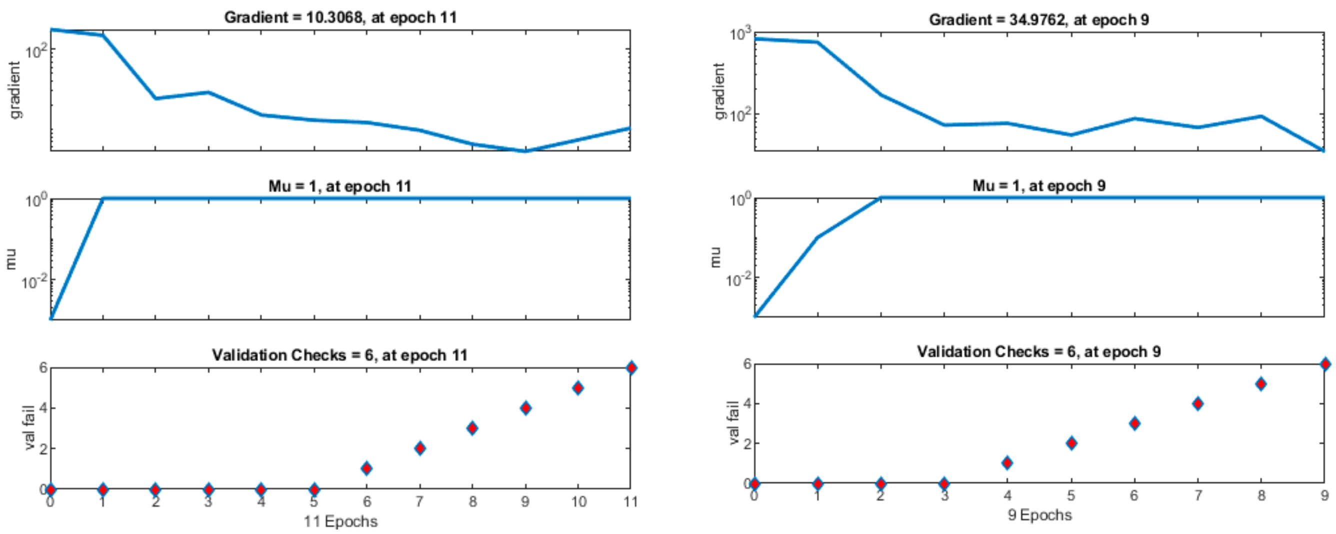
(**a**) (**b**)

Figure 9. Training outs of the MATLAB® Neural Network Training Tool (nntraintool) for the Porosity−well 2 MLP model, indicating that the model converged efficiently and with stable learning dynamics. (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.

Figure 10 shows the evolution of the neural network training of over 11 epochs for porosity-well 2 model. The gradient exhibits an overall decreasing trend, indicating a progressive reduction in the error, although slight fluctuations appear toward the end. The mu parameter shows initial variations associated with the Levenberg–Marquardt adjustment process, stabilizing around 0.1 during the final epochs.

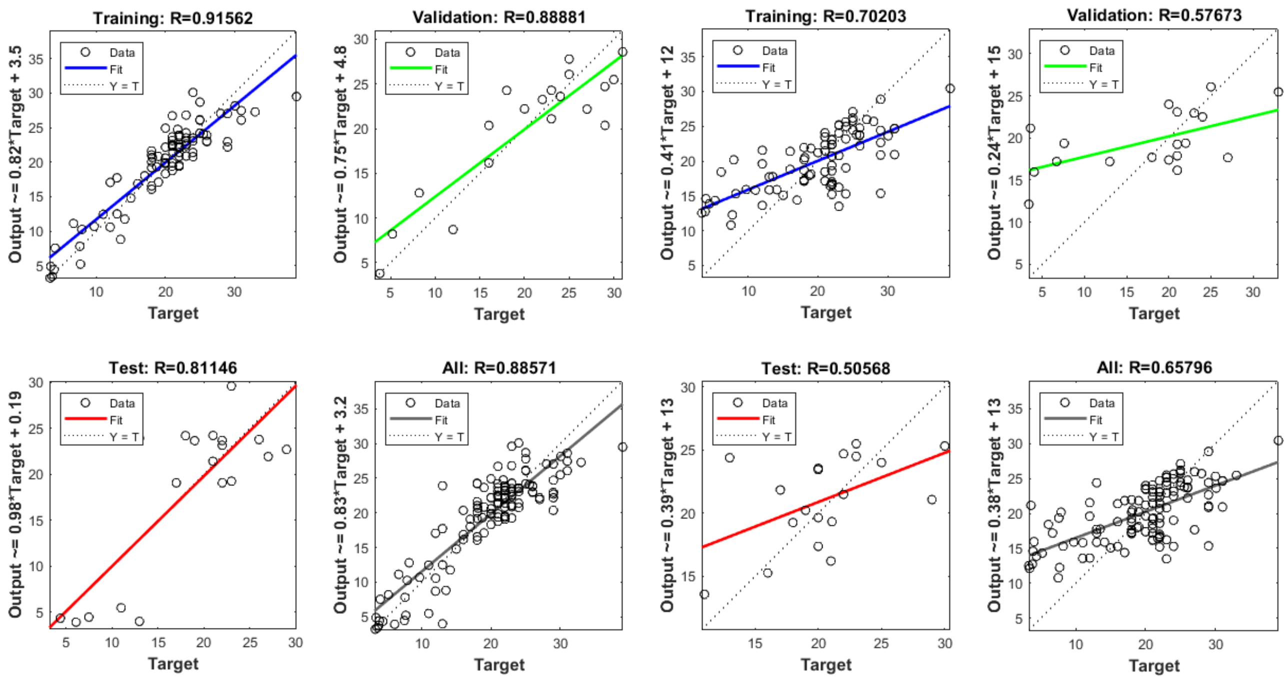
The regression graphs for the porosity model shown in Figure 11 show a good fit of the model in the training and validation sets for well 2. The regression graph that least fits the model is the test graph shown with a red line, which obtains an R of 0.81, which is generally acceptable for the model. The total fits of the training dataset for the porosity prediction models were up to 91.8% in well 1 and 91.5% in well 2.

The results show clear differences between the porosity model trained with integrated well logs plus CT-derived attributes (RHOB-CT and PEF-CT) and the model trained using only conventional well logs. When CT information is included, the predictive performance consistently improves across all data subsets, reaching a correlation coefficient of R = 0.92 (all) with stable fits in the training, validation, and test sets. In contrast, the model without CT features exhibits significantly lower performance (R = 0.66 all), with flatter regression slopes, greater scatter, and reduced ability to track laboratory measurements in the validation and test subsets. These results confirm that CT-based attributes provide additional information on rock heterogeneity, enriching the feature space and enhancing the predictive capability of the ANN model.



(**a**) (**b**)

Figure 10. Gradient, Mu, and validation failure evolution during ANN training using the Levenberg–Marquardt algorithm (trainlm) for Porosity−well 2 model, indicating a progressive reduction of the error. (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.



(**a**) (**b**)

Figure 11. Regression graphs of the training dataset for the porosity prediction model for Well-2. The estimated values maintain a robust fit with $R^2 \approx 0.91$, confirming the model's generalization ability across different datasets. (**a**) Full input dataset CT-Integrated model. (**b**) No-CT as a input for the model.

4.1.2. Permeability Prediction Model

The results obtained iteratively for the permeability model are shown in Table 8. The trained model that showed the best fit of the validation set for well 1 is the one obtained in iteration #4. Meanwhile, the trained model that showed the best fit of the validation set for well 2 is the one obtained in iteration #3.

Table 8. Regression parameters for training the permeability prediction model from tomography records in four different iterations. Source: Own elaboration.

Well	Run	Training R^2	Validation R^2	Test R^2	Total R^2
1	1	0.877	0.865	0.802	0.855
	2	0.914	0.829	0.865	0.882
	3	0.914	0.633	0.772	0.839
	4	0.915	0.888	0.811	0.886
	5	0.826	0.799	0.748	0.813
2	1	0.971	0.891	0.716	0.937
	2	0.837	0.924	0.766	0.834
	3	0.952	0.944	0.85	0.923
	4	0.929	0.702	0.88	0.9

Figure 12 shows the training process of the permeability MLP model for well 1 using MATLAB®'s nntraintool. A consistent decrease in error is observed until the best validation performance is reached at epoch 5. Beyond this point, the validation error begins to increase slightly, indicating the onset of overfitting while the training error continues to decrease.

The Levenberg–Marquardt algorithm parameters, such as the gradient and mu, remain within appropriate ranges, reflecting a stable optimization process. Additionally, six failed validation checks are recorded, triggering early stopping. Overall, the figure demonstrates an efficient modeling process with good fit and adequate generalization during the early epochs.

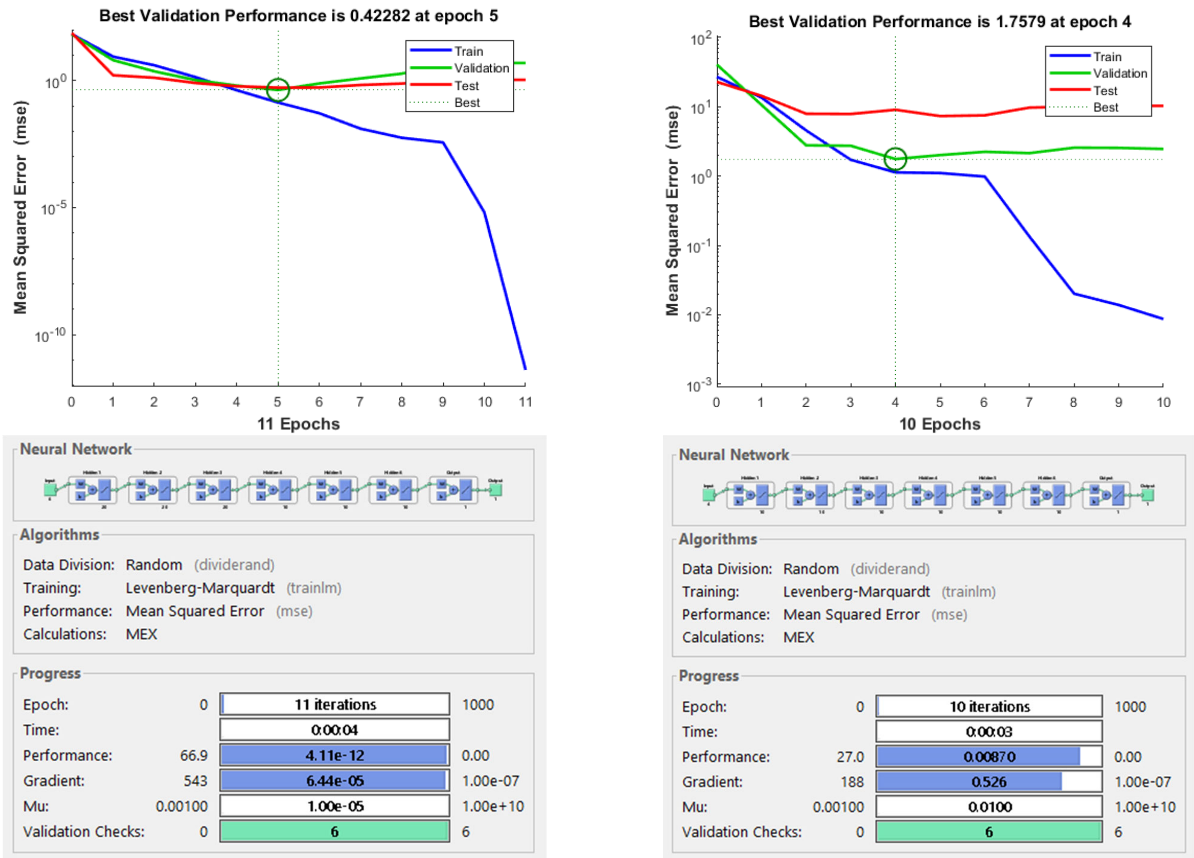
Figure 13 shows the evolution of the gradient, the mu parameter, and the validation failures during the 11 training epochs of the permeability model for well 1. The gradient decreases progressively, indicating a stable adjustment of the network and reaching a very low final value (6.4358×10^{-5}). The mu parameter exhibits initial variations associated with the Levenberg–Marquardt algorithm but tends to decrease toward the final epochs, reflecting increased stability in the optimization process. At the same time, the number of failed validation checks increases steadily until reaching six, which triggers early stopping to prevent overfitting.

The regression graphs compare the permeability measured in the laboratory (x-axis) with the permeability estimated by ANN (y-axis) on a logarithmic scale, for the permeability model. Results shown in Figure 14 indicate three good fits in the training and validation sets for well 1. The test regression graph shown with a red line has an R correlation of 0.86, which is acceptable for the model.

Figure 15 illustrates the relationship between the permeability values predicted by the ANN model and the experimental RCAL measurements after applying the logarithmic transformation during training for Well 1. The solid line represents the regression fit, while the dashed line denotes the ideal 1:1 relationship.

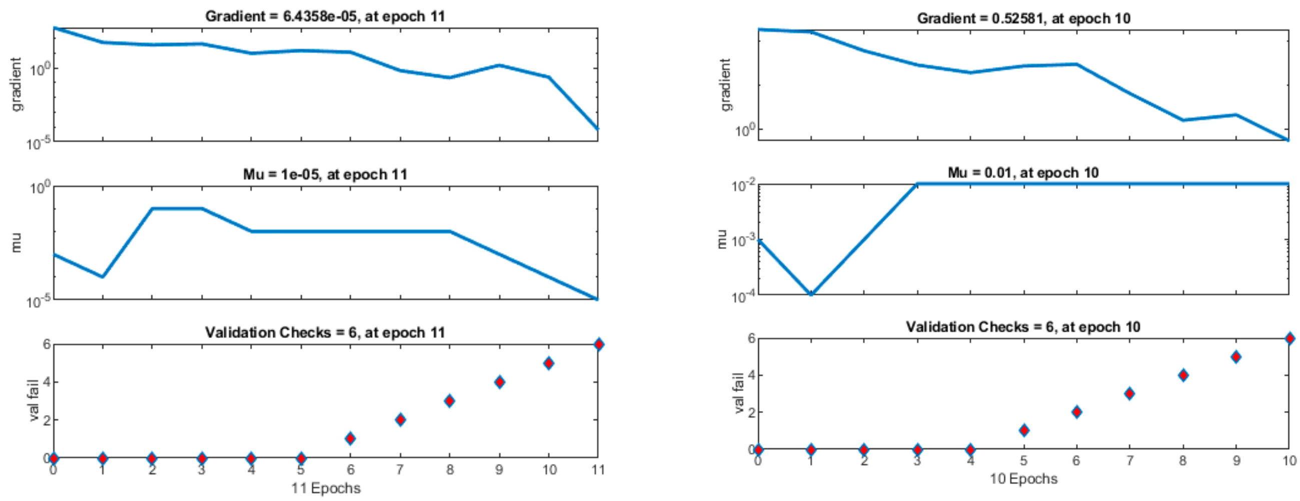
The data points show a generally coherent alignment with the fitted regression line, indicating that the model reliably captures permeability variations across the analyzed range. Although some deviations appear at higher permeability values expected due to the inherently high dispersion of this petrophysical property, the overall correlation is strong (R = 0.94590), demonstrating robust model performance.

The logarithmic transformation effectively stabilizes variance and enhances linearity between predictions and true values, enabling more consistent fitting even for samples with highly contrasting permeability levels.



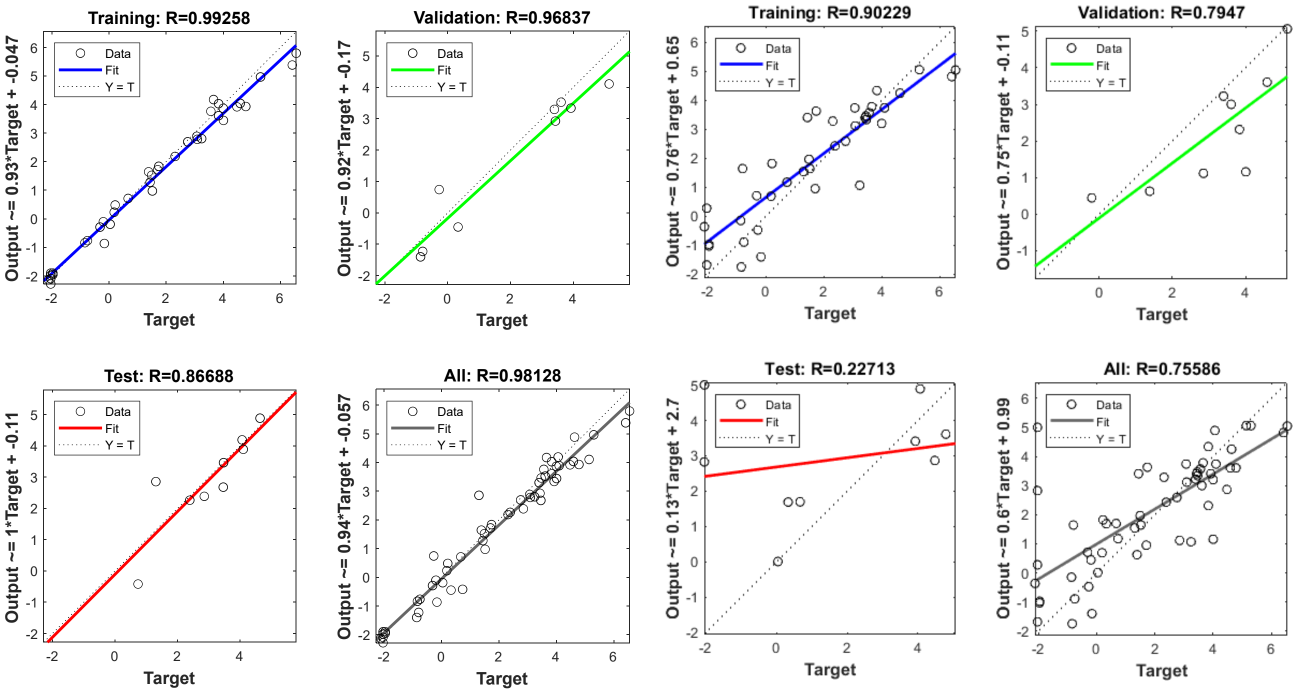
(**a**) (**b**)

Figure 12. Training outs of the MATLAB® Neural Network Training Tool (nntraintool) for the Permeability−well 1 MLP model, reflecting a stable optimization process. (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.



(**a**) (**b**)

Figure 13. Gradient, Mu, and validation-failure evolution during ANN training using the Levenberg–Marquardt algorithm (trainlm) for Permeability−well 1 model, demonstrating an efficient training process with good convergence and appropriate control of validation performance. (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.



(**a**) (**b**)

Figure 14. Regression graphs of the training dataset for the permeability prediction model for Well−1. The model achieves $R^2 \approx 0.98$ on a logarithmic scale, reproducing the RCAL values with high fidelity. (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.

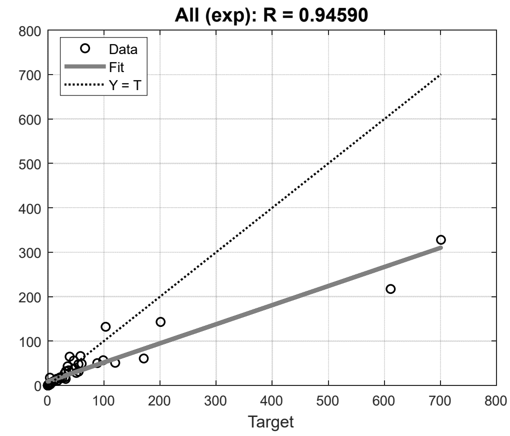
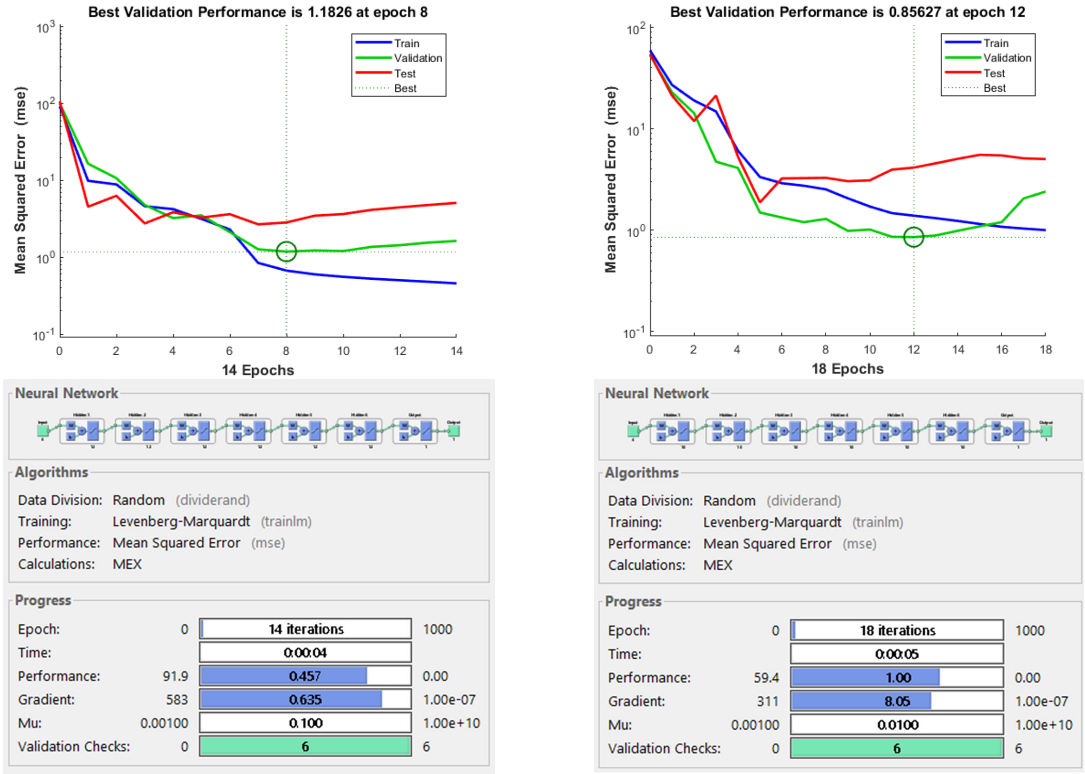


Figure 15. Crossplot between predicted and laboratory-measured permeability values (RCAL) after applying the logarithmic transformation during model training for Well 1. The model exhibits a strong correlation (R = 0.94590), indicating effective capture of permeability trends despite natural dispersion at high values.

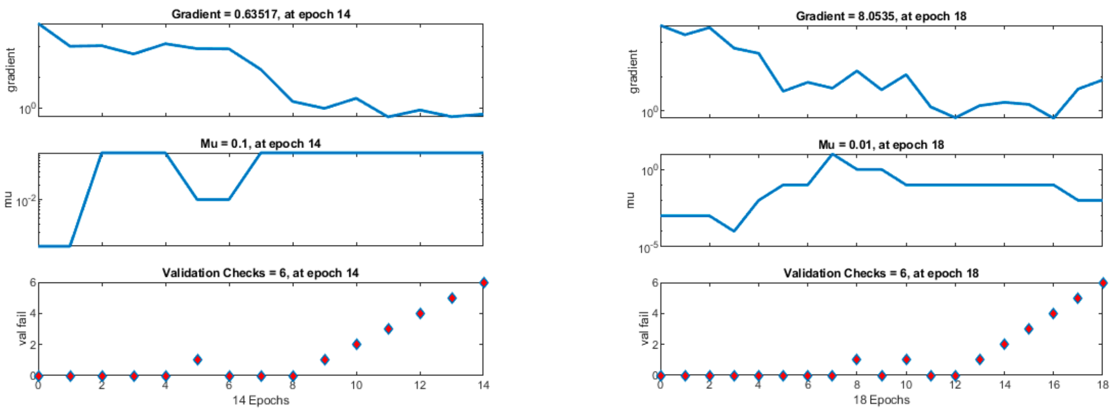
Figure 16 presents the training results of the permeability MLP model for well 2 using MATLAB®'s nntraintool. The error decreases consistently and reaches its best validation performance at epoch 8, where the model exhibits its highest generalization capability. After this point, the validation curve shows a slight increase, indicating the onset of overfitting while the training error continues to improve. Process indicators such as the gradient, mu, and validation failures confirm proper convergence and a correctly triggered early stopping criterion.



(**a**) (**b**)

Figure 16. Training outs of the MATLAB® Neural Network Training Tool (nntraintool) for the Permeability—well 2 MLP model, reflecting an efficient training process and a model with strong performance during the initial epochs. (**a**) Full input dataset CT—Integrated model. (**b**) No—CT model.

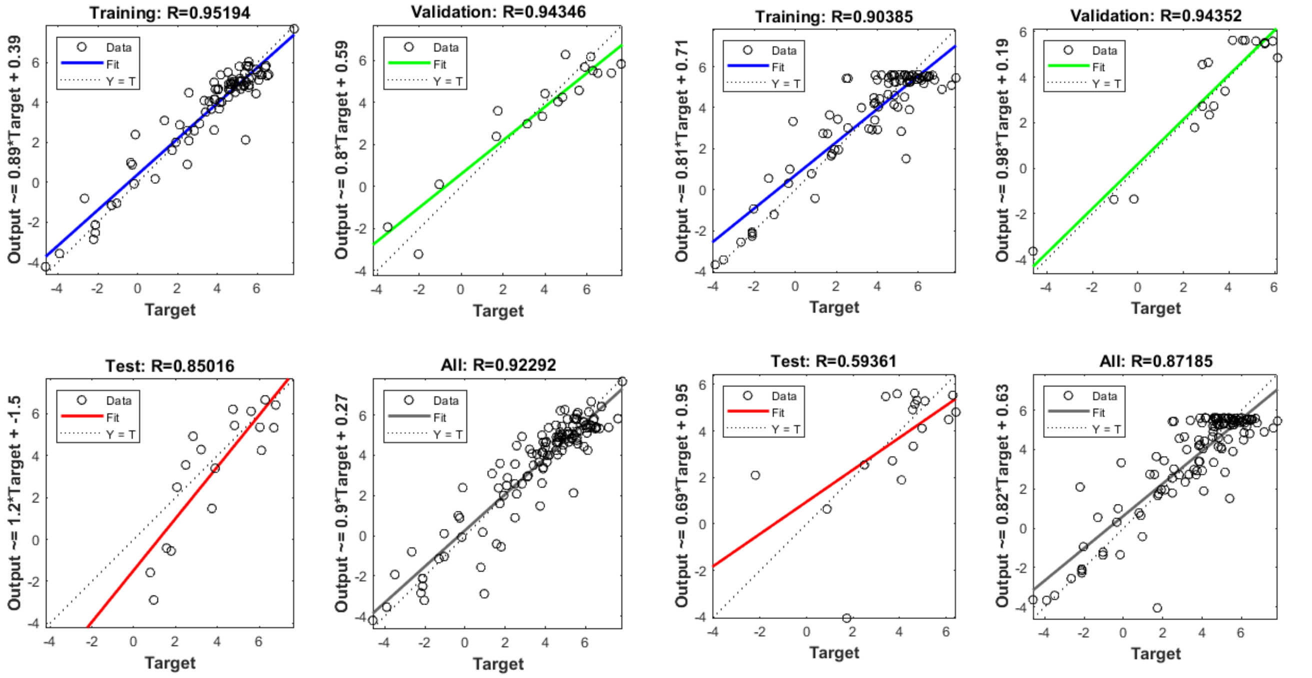
Figure 17 shows the evolution of the gradient, the mu parameter, and the validation failures during the training of MLP permeability—well 2 model. The gradient decreases consistently, indicating stable convergence, while mu exhibits moderate variations that reflect an appropriate adjustment of the learning step. Validation failures remain low throughout most of the process, demonstrating good generalization capability.



(**a**) (**b**)

Figure 17. Gradient, Mu, and validation-failure evolution during ANN training using the Levenberg–Marquardt algorithm (trainlm) for Permeability—well 2 model, the three indicators reveal an efficient training process, with controlled convergence and no significant signs of overfitting. (**a**) Full input dataset CT—Integrated model. (**b**) No—CT model.

The regression graphs for the permeability model shown in Figure 18 indicate three good fits in the training and validation sets for well 2. The test regression graph shown with a red line has an R correlation of 0.85, which is acceptable for the model. Quite good results were obtained in both wells. The total fits of the training data set for the permeability prediction models were up to 99% in well 1 and 95% in well 2.



(**a**) (**b**)

Figure 18. Regression graphs from the training dataset of the permeability prediction model for Well−2. A satisfactory fit is observed with $R^2 \approx 0.89$ on a logarithmic scale, validating the applicability of the model. (**a**) Full input dataset CT−Integrated model. (**b**) No−CT model.

Figure 19 displays the relationship between the permeability values predicted by the model and the experimental RCAL measurements without applying the logarithmic transformation for well 2. The plot shows a moderate correlation (R = 0.75693), with noticeable dispersion at higher permeability levels, an expected behavior due to the natural variability and scale sensitivity of this property. The fitted regression line exhibits a slope lower than the 1:1 reference, indicating a tendency of the model to underestimate higher values. Overall, the figure demonstrates acceptable performance, though with reduced stability compared to the log-transformed case.

The comparison clearly shows that including CT−derived features (RHOB−CT, PEF−CT) substantially improves permeability model performance. The CT model achieves consistently higher correlation values across all subsets (R ≈ 0.92–0.95), whereas the model without CT exhibits a marked reduction, particularly in the test set (R ≈ 0.59). This reduction is expected when fewer input features are available, limiting the model's ability to capture nonlinear relationships. Overall, the results demonstrate that CT features provide essential information that significantly strengthens predictive accuracy.

In this section, a good fit was observed between the data estimated by the trained neural networks and the laboratory petrophysical data (RCAL). To validate this, the results are shown at two different scales below: first, the validation logs generated by the compiled model at 90 ft intervals for each prediction model (Section 4.2), and second, the results of both models at a higher resolution scale in a 3 ft interval compared to tomography images (Section 4.3).

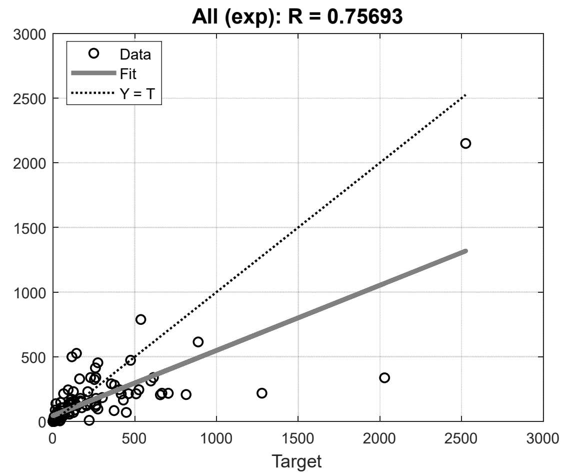


Figure 19. Crossplot between predicted and laboratory-measured permeability values (RCAL) after applying the logarithmic transformation during model training for Well 2. The model shows moderate correlation (R = 0.75693) and increased dispersion at higher permeability values, reflecting the inherent variability of this property and the reduced stability of predictions when raw permeability data are used.

4.2. Validation of Results of Porosity and Permeability Models at Well Logs Scale

The trained neural network was exported to feed the integration model. To validate the previously trained prediction models, well log data and tomography information (individual RHOB and PEF logs) from the two study wells were used, but this time evaluated over an extension covering an interval between 65.25 and 2167.75 ft, with an average resolution of 0.6 mm. This was the range that contained the complete information from both logs and tomography.

4.2.1. Results of Porosity Model at Well Logs Scale

Estimates were obtained from the porosity prediction model, extrapolated over the entire interval of the well. Figures 20 and 21 show the results obtained over a specific interval for each study well. Note that the graph on the right shows the porosity predicted by the network (solid line) and the known laboratory porosity (red dots); an excellent fit between these data can be seen.

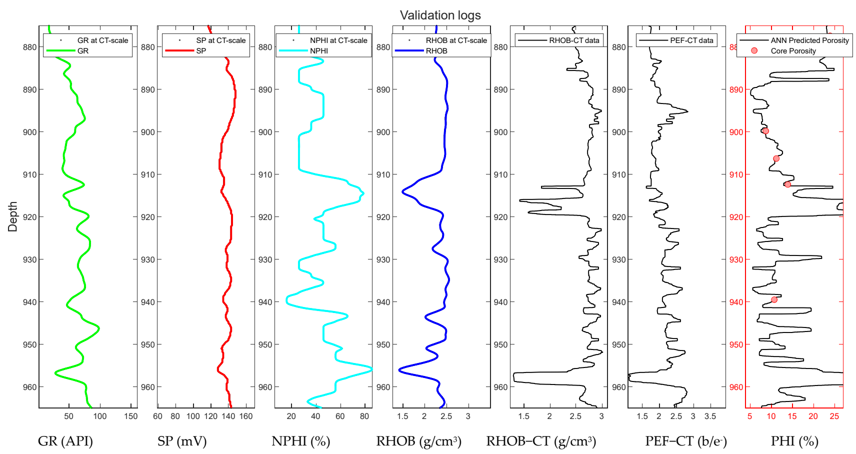


Figure 20. Results of the porosity prediction model for Well−1. (875–975 ft). The results closely align with RCAL measurements, demonstrating the model's ability to capture petrophysical trends at the log scale (~0.5 ft).

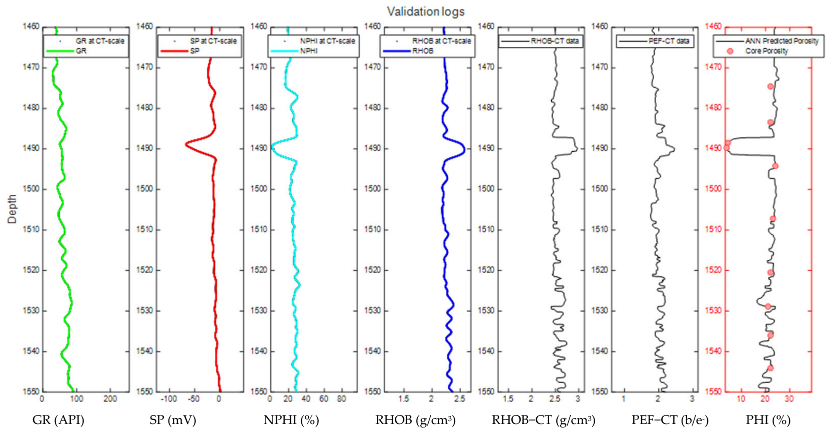


Figure 21. Results of the porosity prediction model for Well−2. (875–975 ft). Consistency between estimated values, RCAL, and conventional records is confirmed.

4.2.2. Results of Permeability Model at Well Logs Scale

Estimates were obtained from the permeability prediction model, extrapolated over the entire interval of the well. Figures 22 and 23 show the results obtained for a specific interval for each study well. Note that the graph on the right shows the permeability predicted by the network (solid line) and the permeability known from laboratory tests (red dots); an excellent fit between these data can be seen.

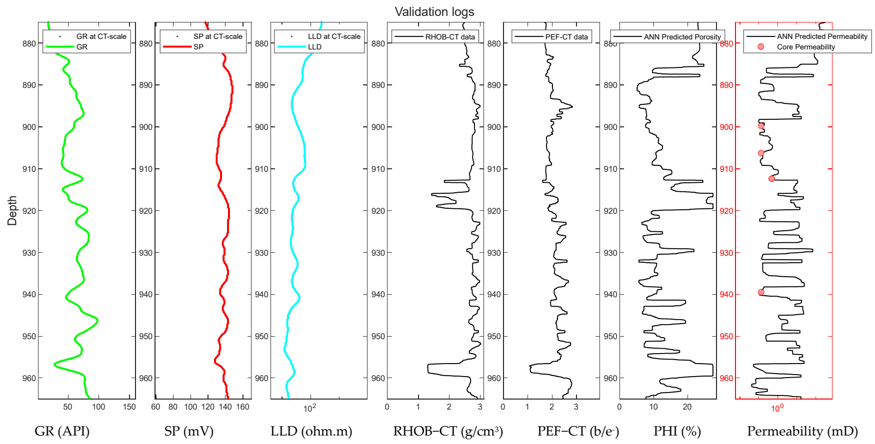


Figure 22. Results of the permeability prediction model for well-1 (interval 875–975 ft). The model accurately reproduces the variations recorded in the laboratory and in conventional logs.

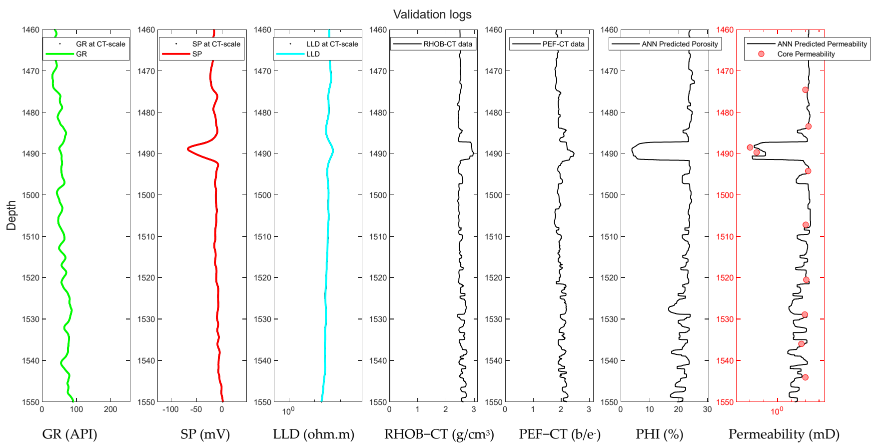


Figure 23. Results of the permeability prediction model for well−2 (interval 875–975 ft). The model maintains a robust fit consistent with RCAL values.

As can be seen, the porosity prediction models (Figures 20 and 21) and the permeability prediction models (Figures 22 and 23) fit the laboratory measurements quite well. Furthermore, they correlate closely with the tool measurements at that scale of logs.

4.3. Validation of Results of Porosity and Permeability Models CT−Scale

The main advantage of the proposed models over conventional approaches despite the fact that traditional log scale techniques (0.25–0.5 ft) already provide reasonable porosity estimations is their ability to also predict permeability with comparable accuracy, without requiring prior flow unit modeling or rock-package classification. In addition, the method yields reliable results at significantly higher vertical resolution, allowing the characterization of heterogeneities at finer scales and supporting.

The porosity estimates show a good agreement with RCAL measurements, which reinforces the validity of the ANN based predictions and confirms the consistency of the results with respect to experimental core derived data. To verify this capability, model outputs for both case studies were generated at 3 ft intervals and compared against corresponding CT-scan core images. In both cases, the predicted porosity and permeability distributions exhibited strong visual and spatial coherence with tomographic observations, demonstrating that the ANN based model reproduces the internal textural patterns observed in core data (Figures 24–29).

Taken together, the results demonstrate that the proposed approach not only accurately predicts petrophysical properties at the log scale, but also extends the characterization capability to finer scales, consistently integrating laboratory information, conventional logs, and CT.

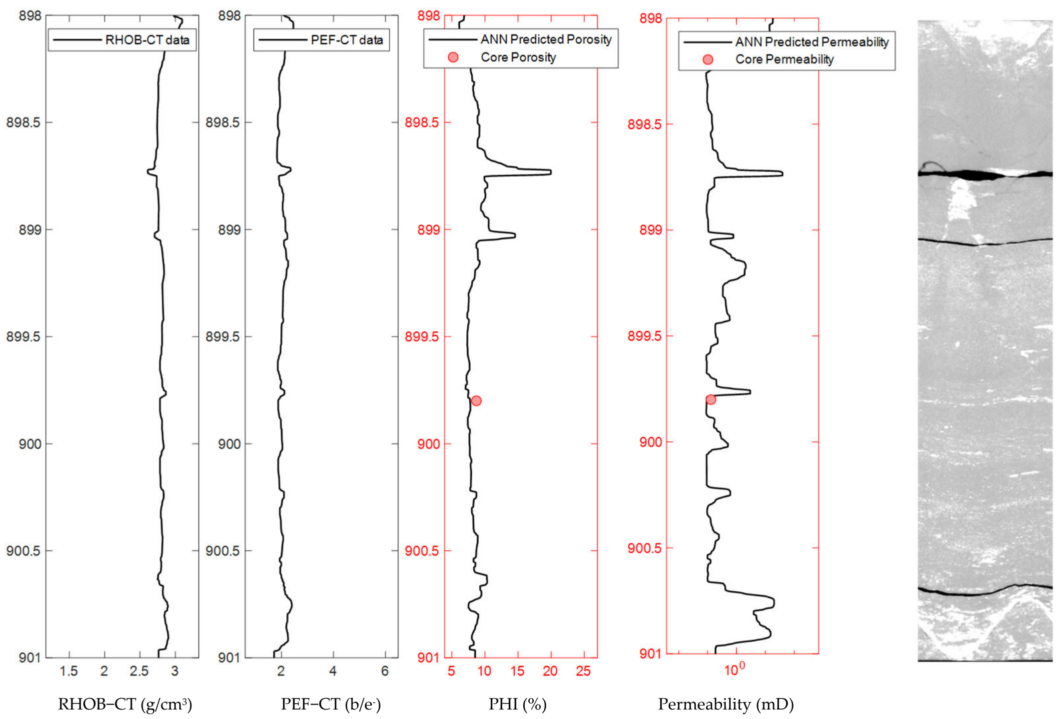


Figure 24. Comparison of porosity and permeability predictions using ANN results versus CT images in Well−1 (interval 898–901 ft). There is evidence of correspondence between internal textures patterns observed trough tomographic images and variations in estimated petrophysical properties by ANN. The results closely align with RCAL measurements, demonstrating the model's ability to capture petrophysical trends at the CT scale (3 ft).

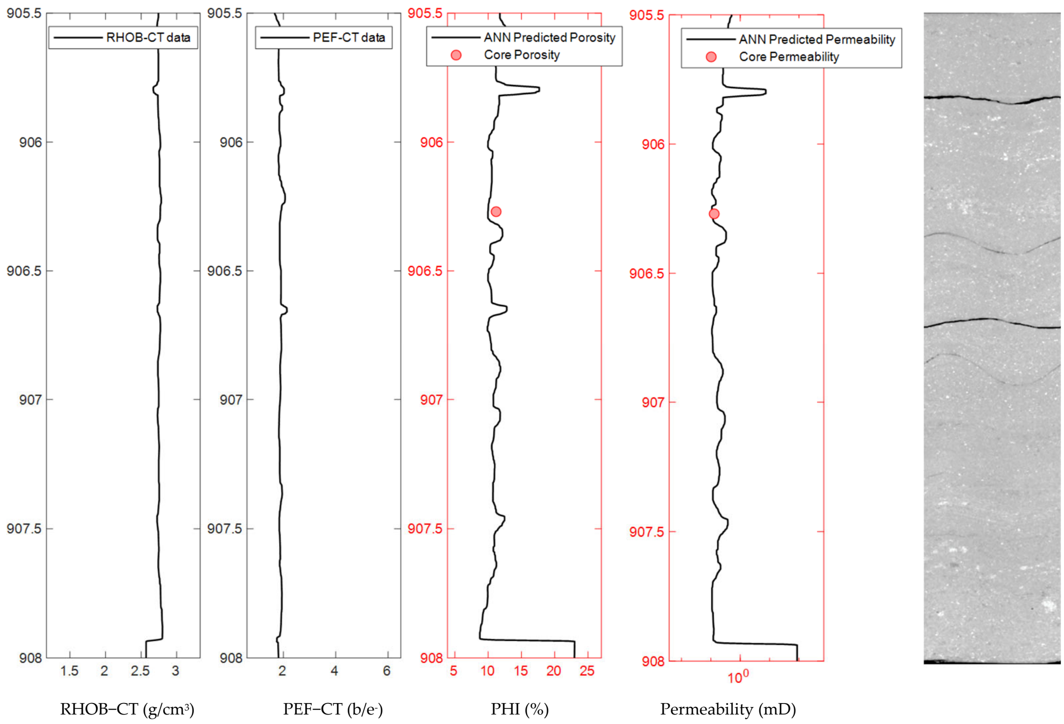


Figure 25. Comparison of ANN predictions and CT images in Well−1 (interval 905.5–908 ft). The model captures transitions in porosity and permeability observed in the texture of the scanned core. The results closely align with RCAL measurements, demonstrating the model's ability to capture petrophysical trends at the CT scale (3 ft).

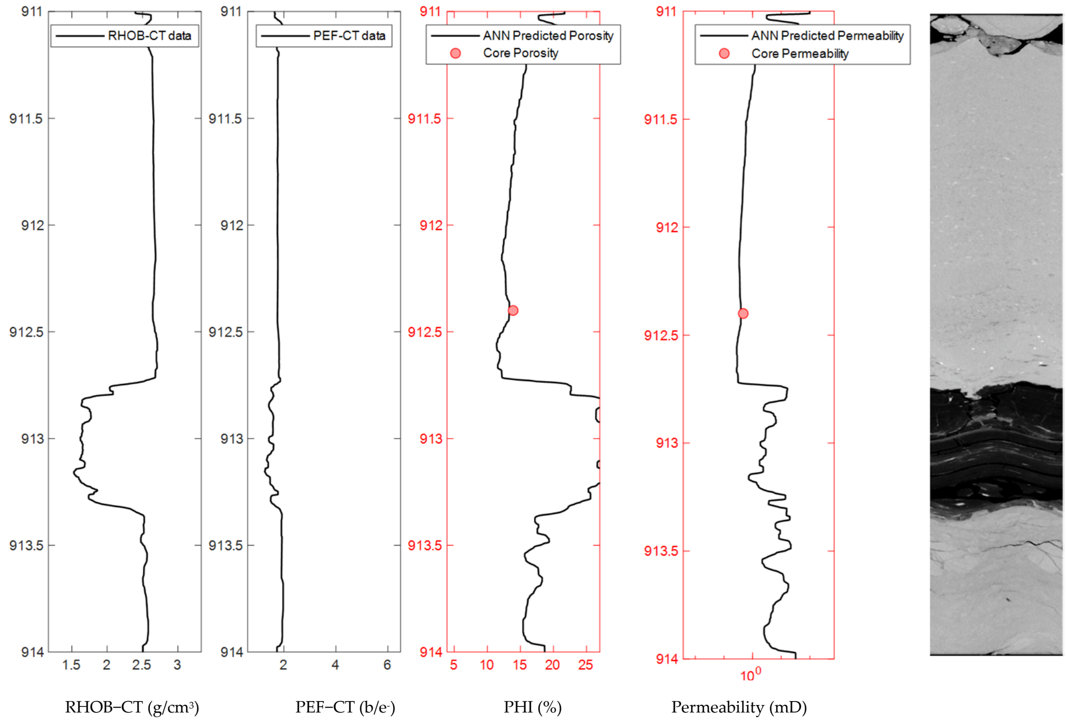


Figure 26. ANN results and CT images in Well−1 (interval 911–914 ft). Fine adjustment of petrophysical properties estimation is observed with local variations identified in the tomography.

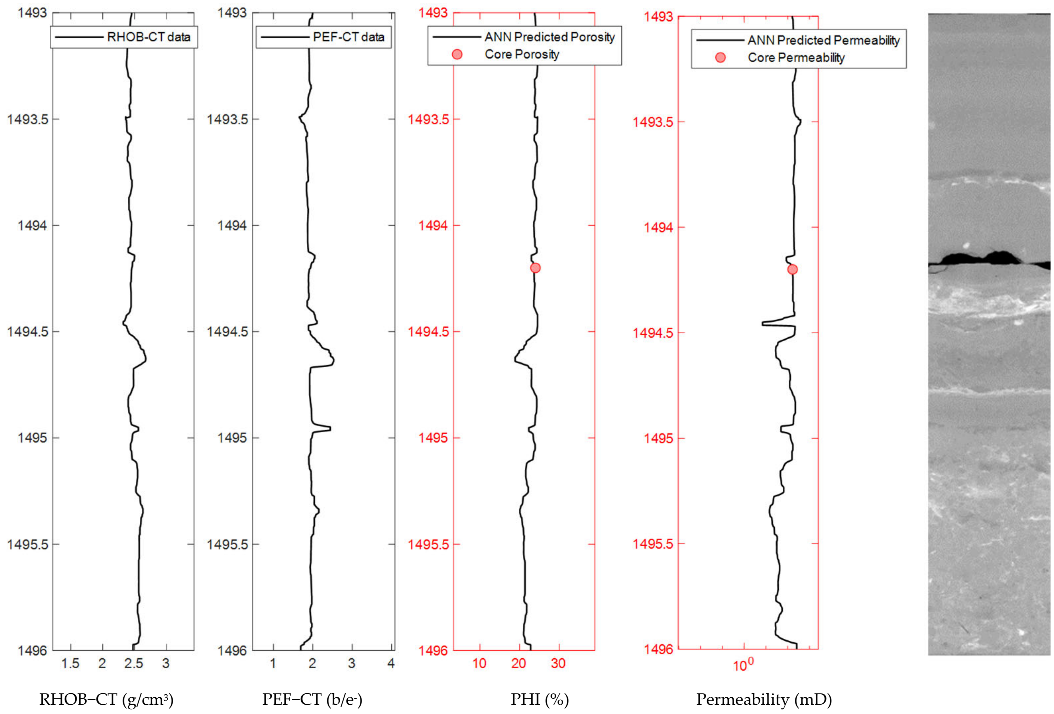


Figure 27. ANN results comparison with CT images in Well−2 (interval 1493–1496 ft). The model reproduces local variations in porosity and permeability associated with textural heterogeneities in the rock.

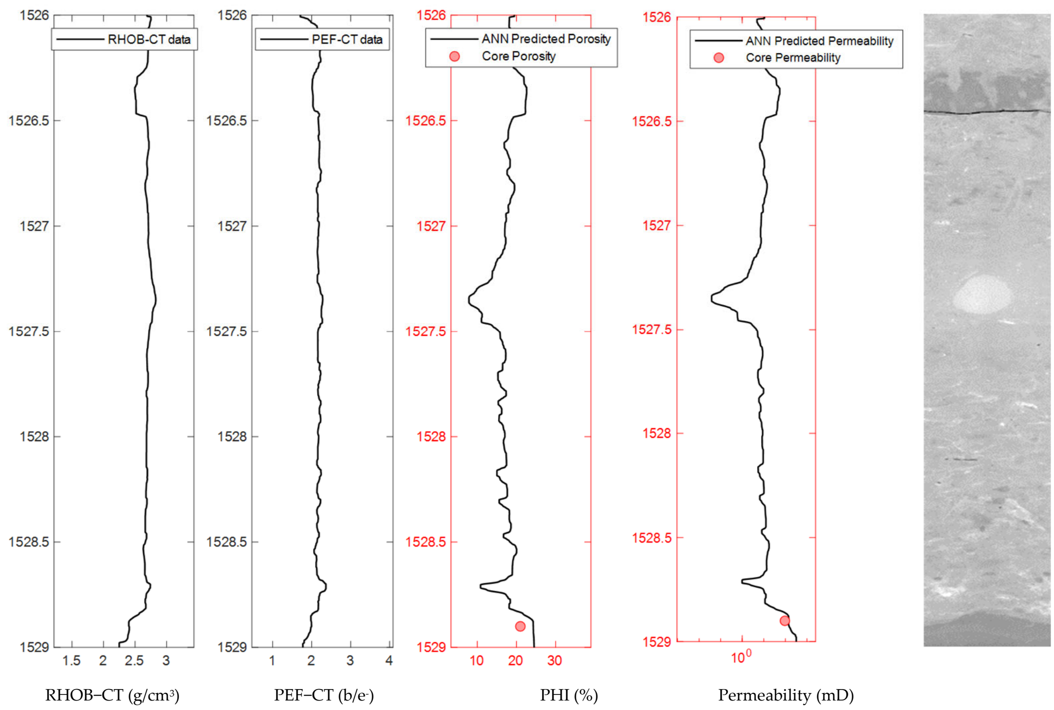


Figure 28. ANN results and CT images in Well−2 (interval 1526–1529 ft). There is evidence of consistency between the tomography records and the ANN predictions.

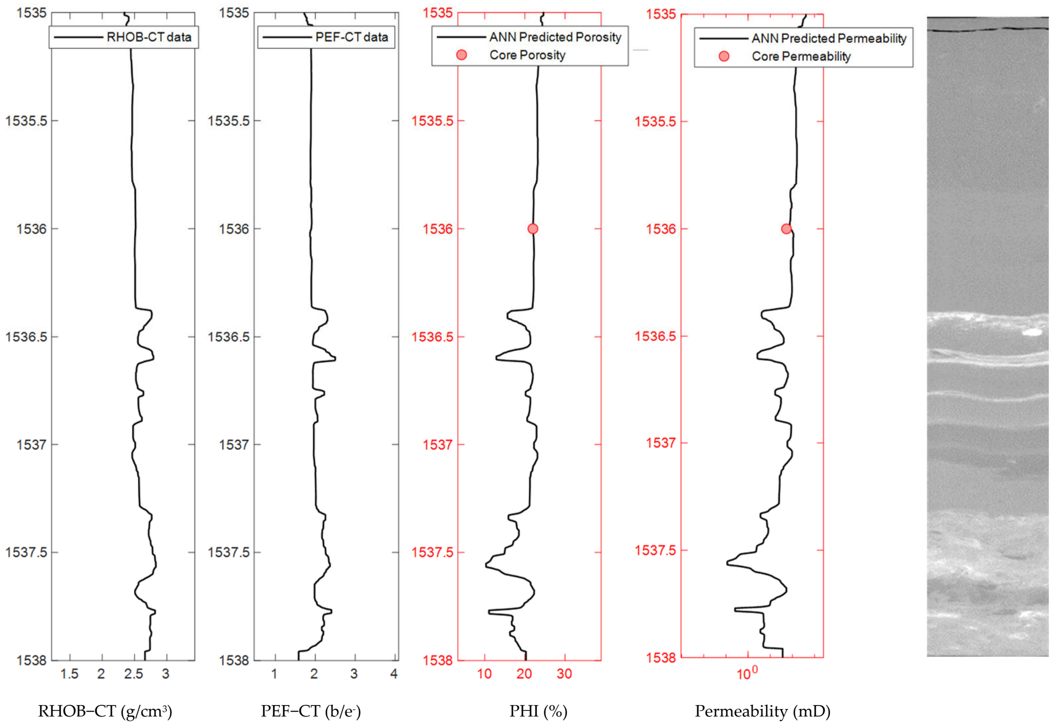


Figure 29. ANN–CT validation in Well−2 (interval 1535–1538 ft). The model maintains consistency in the estimation of petrophysical properties at millimeter resolution, capturing small range heterogeneity.

5. Discussion

The artificial neural network (ANN) models developed in this study exhibit strong agreement between predicted and laboratory measured petrophysical properties (RCAL). This section analyzes the key factors influencing model performance, compares the proposed approach with alternative machine-learning methods, and discusses its applicability across different spatial scales.

5.1. Performance of ANN Models

The results obtained show that the integration of conventional logs and parameters derived from computed tomography using ANN allows porosity and permeability to be estimated with high accuracy and millimeter resolution. R^2 values greater than 0.89 in all cases, together with low prediction errors, confirm the robustness of the approach in two different wells. Regression graphs and log and CT scale validations show that the model captures both general trends and local heterogeneities.

Overall, the comparison between the two tested models, one trained using conventional well logs combined with CT derived attributes (RHOB−CT and PEF−CT) and another trained solely with conventional logs, highlights the direct impact of CT information on predictive performance. The comparison demonstrates that incorporating CT derived features significantly enhances predictive accuracy and model stability, reinforcing the benefit of combining conventional logs and CT data to improve nonlinear regression performance in petrophysical estimation.

5.2. Comparison with Previous Studies

This paper's findings are consistent with previous studies reporting the applicability of ANN in estimating petrophysical properties from conventional logs [1,15] and with work that has explored the integration of CT and machine learning, demonstrating the feasibility of predicting porosity from CT logs (RHOB and PEF) [4].

Similarly, more recent studies used learning ensembles on micro−CT images to estimate porosity and permeability, showing the usefulness of tomographic information for capturing small-scale heterogeneity [6]. The present work extends these approaches by simultaneously integrating conventional and CT logs and by validating porosity and permeability models against laboratory data in two wells, reproducing millimeter-scale heterogeneity.

To contextualize the performance of the proposed MLP models, we examined the recent literature comparing commonly used regression algorithms for petrophysical prediction (porosity and permeability) on well-log and core derived datasets (Table 9). A consistent pattern emerges: gradient-boosted tree ensembles (e.g., XGBoost) and Random Forest (RF) are frequently the best-performing baselines on tabular well-log datasets, providing robust R^2/MAE improvements after careful hyperparameter tuning. Several applied studies report XGBoost slightly outperforming RF for porosity/permeability prediction when extensive feature engineering is applied [26–29].

Table 9. Literature-based comparison of ML methods for petrophysical property prediction.

Methods Compared	Main Finding
XGBoost vs. Random Forest	Both algorithms performed well; XGBoost slightly superior after tuning, highlighting gradient boosting's robustness for well-log regression [29].
ANN vs. XGBoost	ANN and XGBoost both achieved good predictive accuracy; XGBoost showed higher computational efficiency [26].
ANN, RF, SVR, GPR	Ensemble methods (RF/XGBoost) robust for tabular log data; ANN excels with image-based or CT-enhanced inputs [30].
Optimized XGBoost vs. Baseline ML models	Hybrid optimization significantly improved XGBoost prediction stability and accuracy; $R^2 > 0.95$ for carbonate cores [31].
RF, XGBoost, SVR	RF and XGBoost provided the highest accuracy; feature-selection improved model interpretability and reduced overfitting [27].
XGBoost vs. Traditional Regression	XGBoost achieved lower RMSE (0.087 mD) and higher R^2 (0.96) than linear and polynomial regressions [28].
ANN, RF, SVR	ANN and RF achieved comparable accuracy; ANN more adaptable for continuous real-time prediction [32].
CNN vs. Ensemble ML	CNNs outperformed ensembles on image-based datasets; ensembles better for tabular log summaries [10].

Conversely, artificial neural networks (ANN/MLP) remain competitive, often matching or exceeding tree methods, when (i) larger and more diverse training datasets are available, (ii) CT or image-derived features are included, or (iii) the problem requires learning complex nonlinear interactions between features. Studies on image based permeability estimation show deep learning and CNN based approaches excel on image inputs, while ensemble methods dominate by tabular summaries [10,30–32].

Support for uncertainty quantification is another differentiator: Gaussian Process Regression (GPR) provides principled uncertainty estimates but scales poorly with dataset size, and SVR performs reasonably for small datasets but is sensitive to kernel selection and scaling [27]. Overall, the literature suggests that no single method universally dominates; rather, method choice depends on data modality, dataset size, and the availability of engineered features. These findings motivate our MLP approach when integrating conventional logs with CT derived features and provide the rationale for the observed performance in this study.

5.3. Advantages of the Proposed Approach

One of the Strengths of this approach lies in its ability to predict porosity and permeability without requiring the explicit definition of flow units or prior rock type classification, as the artificial neural network inherently integrates lithological and petrophysical variability within its learning structure.

Additionally, the method enables porosity (Figure 30) and permeability (Figure 31) estimation with enhanced vertical resolution, capable of capturing finer scale heterogeneities than those calculated through conventional techniques [33]. Furthermore, the ANN-based prediction yielded higher correlation values (R= 0.90), whereas the conventional approach reported correlation coefficients of 0.84 for porosity model.

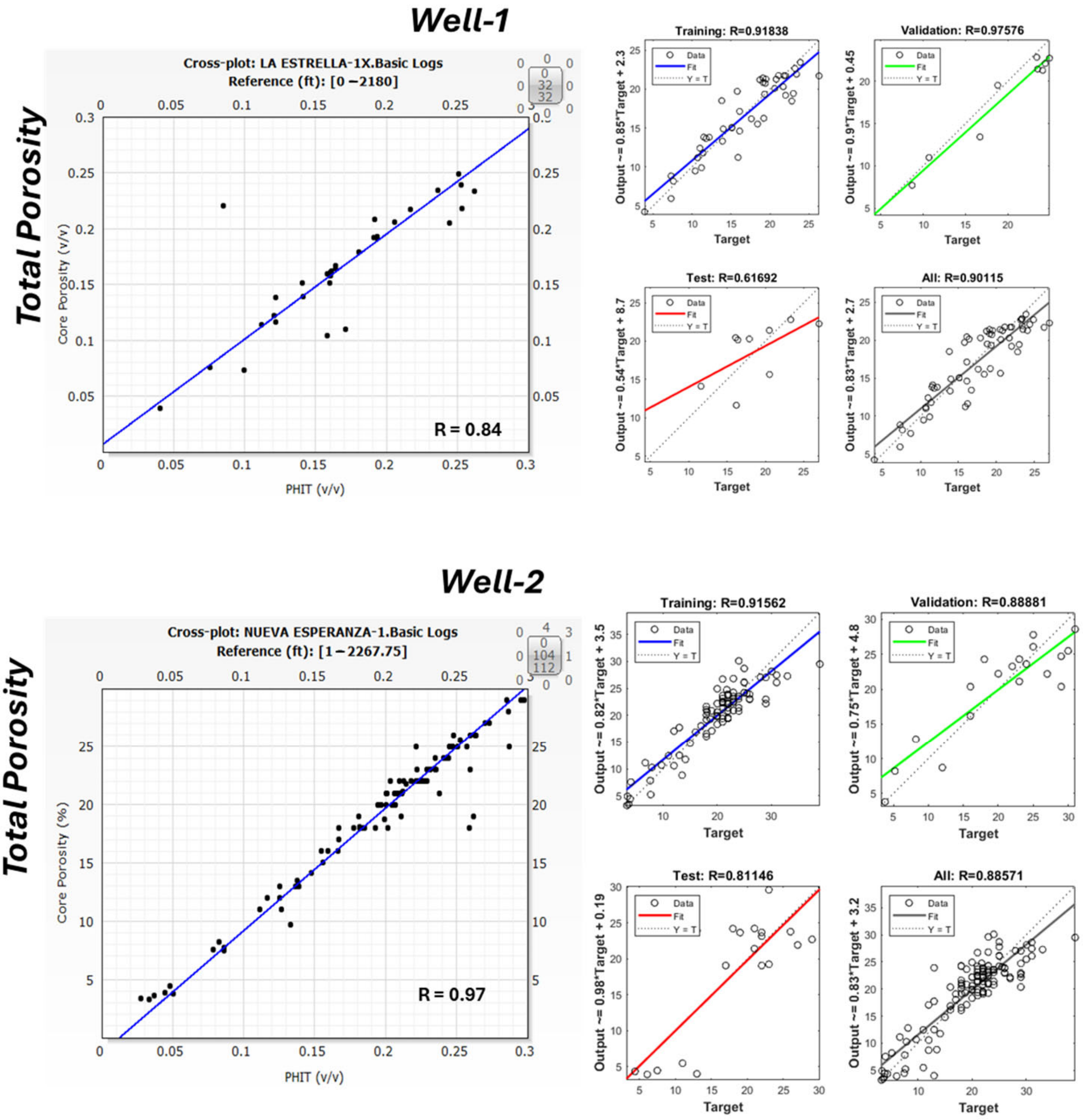


Figure 30. Cross-plot comparison of porosity estimated using conventional methods (left) [33] versus the estimation obtained through ANN−MLP predictive model (right). The ANN based model exhibits a stronger correlation against laboratory RCAL measurements, highlighting its improved capacity to capture nonlinear relationships between log responses and porosity.

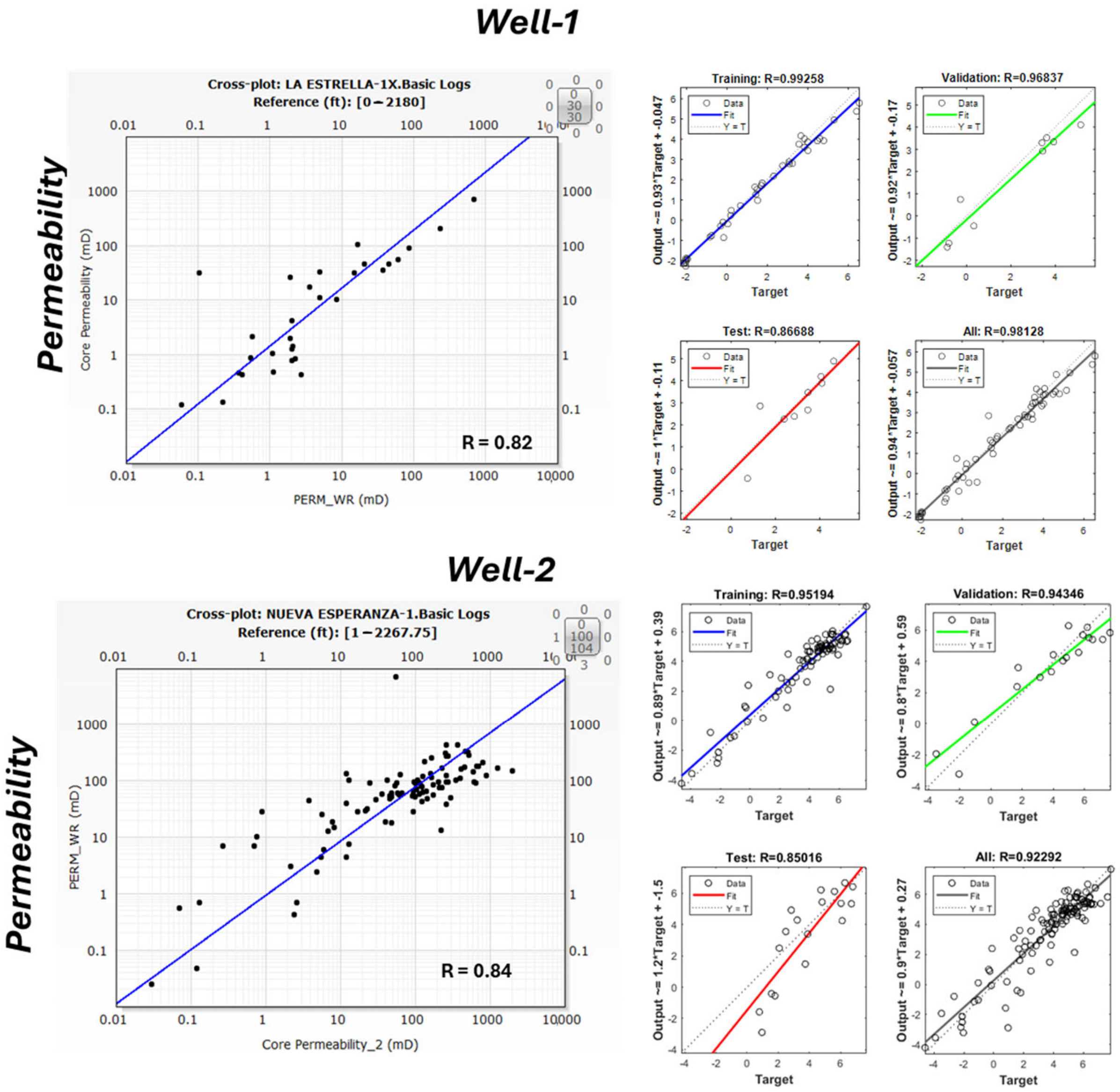


Figure 31. Cross-plot comparison of permeability obtained using conventional methods (left) [33] versus the permeability estimated by the ANN−MLP predictive model (right). The ANN based estimation demonstrates superior agreement with laboratory RCAL permeability, indicating better predictive performance and enhanced suitability for high-resolution petrophysical characterization.

This demonstrates that this proposed approach improved resolution and workflow simplicity; the data driven ANN framework calibrated with RCAL offers an alternative for petrophysical prediction.

5.4. Limitations and Future Tests

Despite the favorable results, certain limitations must be considered. First, the relatively small size of the datasets used (57 and 125 samples per well) restricts the model's generalization capacity. This aspect is critical, as neural networks with limited datasets can overfit and lose predictive robustness [6].

Secondly, although the method shows potential for extrapolation to wells without CT information, validation of this scenario could be complemented with additional external testing. This validation can include: (i) applying the model to nearby wells without coring, but with complete conventional logs (GR, SP, NPHI, RHOB), to evaluate its predictive capacity using external well logs alone; (ii) performing inter-well cross-validations, training on one well and predicting on another, in order to test the generalization of the model; and (iii) comparing predictions against independent data from pressure tests or wireline formation tests, which can provide indirect verification of permeability. Recent studies have demonstrated the usefulness of similar approaches when integrating logs and seismic data [12], but their extension to non-CT scenarios still requires empirical validation.

Finally, the millimeter resolution achieved depends directly on the quality of the tomographic images; previous research has shown that the accuracy of porosity estimation using CT is strongly conditioned by the resolution and quality of the scan, which implies additional acquisition costs [11].

For future work, it is recommended to implement additional predictive approaches recently reported in the literature, such as Gradient Boosting techniques (e.g., XGBoost), Random Forest regression, Support Vector Regression, and Gaussian Process models. Incorporating these methods under equivalent training–validation–testing conditions would enable a systematic benchmarking of predictive performance against the MLP-based framework presented in this study.

This comparative evaluation would allow quantifying differences in accuracy, computational efficiency, robustness to limited datasets, and sensitivity to input feature selection (with and without CT-derived variables). Such extensions would strengthen the generalization of the methodology and provide deeper insight into the suitability of AI-based models for petrophysical property estimation in diverse subsurface settings.

In summary, the objective of this work is not to claim regional-scale generalization, but rather to evaluate the feasibility and the added value of integrating CT derived features with conventional well logs within a controlled geological setting through the performance of ANN–MLP models.

5.5. Practical Implications

The proposed approach is particularly useful in fields where core acquisition is limited or economically unfeasible. By integrating log and CT data using ANN, it is possible to extend laboratory measurements to uncored wells, improve the quality of static reservoir models, and characterize in greater detail the heterogeneities that control fluid flow.

This suggests potential for application in geological–petrophysical modeling and dynamic simulation projects at more realistic scales. Despite its limitations, the results demonstrate that integrating conventional logs and CT with ANN offers a promising path for improving high-resolution petrophysical characterization.

6. Conclusions

The integration of conventional well logs with parameters derived from computed tomography (CT) to feed petrophysical property prediction models using artificial neural networks (ANN) allows porosity and permeability to be estimated with high accuracy, achieving a satisfactory fit between the estimated data and laboratory data, with coefficients of determination (R^2) greater than 0.89 in both wells analyzed. As a result, reliable values are obtained at a higher level of resolution compared to that achieved conventionally.

The prediction models implemented show robust prediction capabilities at both the log scale (0.25–0.5 ft) and the CT scale (~0.6 mm), which facilitates the multiscale characterization of petrophysical heterogeneities.

Unlike traditional approaches, the proposed methodology does not require the explicit definition of flow units and rock package classification to estimate permeability. However, the model generates good results since it is fed with logs, which are used to implicitly model rock types and thus relate porosity values to permeability data.

The main limitation of the study lies in the small size of the datasets and the dependence on the quality of the CT scans, which restricts the generalization of the model to other geological environments. In this sense, results are constrained to formations and wells with comparable lithological and petrophysical characteristics, and extension to more heterogeneous settings would require additional tests, wells and broader datasets.

High-resolution petrophysical data can be useful for extending or extrapolating direct laboratory measurements. Alternatively, the prediction model can be extrapolated to wells that do not have such measurements but do have the logs that feed the model. In other words, it is possible to apply the previously trained neural networks to nearby wells without coring, which, despite not having tomography information, probably have apparent density (RHOB) and photoelectric factor (PEF) logs. In this case, the proposed workflow has the advantage that high-resolution data were used during the training of the neural networks, improving the quality of the training and, possibly, the quality of the prediction data.

In this sense, the approach is a promising tool for extrapolating laboratory measurements to uncored wells, optimizing static reservoir models, and capturing heterogeneities that impact fluid flow in production and storage processes.

Author Contributions: Conceptualization, E.H.H.O., J.O.O.L. and E.R.G.; methodology, E.H.H.O.; software, E.H.H.O.; validation, E.H.H.O., J.O.O.L. and E.R.G.; formal analysis, E.H.H.O.; investigation, E.H.H.O.; resources, E.H.H.O.; data curation, E.H.H.O.; writing—original draft preparation, E.H.H.O.; writing—review and editing, E.H.H.O., J.O.O.L. and E.R.G.; visualization, E.H.H.O.; supervision, J.O.O.L. and E.R.G.; project administration, E.H.H.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Hydrocarbons Agency (ANH), Minciencias and the scientific committee of the project "Physical and Quality Stratigraphy of the Drill Cores Recovered by ANH in the Basins of the Lower Magdalena Valley, Cesar, and Ranchería" (Contract No. FP44842—454—2017, UIS—MINCIENCIAS). The APC was funded by Contract No. FP44842—454—2017, UIS—MINCIENCIAS.

Data Availability Statement: The datasets presented in this article are not immediately available because they are owned by ANH. Requests for access to the datasets should be directed to the corresponding entity: ANH, SGC, or the Petroleum Information Bank of Colombia.

Acknowledgments: Gratitude is expressed to the Industrial University of Santander for granting the study commission and providing the necessary resources and logistical support for the development of this research. Acknowledgment is also extended to the Research Group in Computed Tomography for Reservoir Characterization (GIT—UIS) and its director, Nicolás Santos, for the execution of laboratory tests and the accomplishment of the proposed objectives.

Conflicts of 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.

Abbreviations

The following abbreviations are used in this manuscript:

UIS	Universidad Industrial de Santander
GIT	Group in Tomography for Reservoir Characterization
UAB	Universidad Autónoma de Barcelona
Minciencias	Ministry of Science
ANH	National Hydrocarbons Agency
SGC	Colombian Geological Survey
ANN	Artificial Neural Network
MLP	Multilayer Perceptron

CT	Computed Tomography
RHOB	Density Log in Well Logging
PEF	Photoelectric Effect
GR	Gamma Ray
SP	Spontaneous Potential log
RCAL	Routine core analysis
PEF-CT	Photoelectric factor from Computed Tomography
RHOB-CT	Bulk density from Computed Tomography

Appendix A

Appendix A.1

Training Pseudocode for Porosity Model:

```
% ======= START =======
% ------- Load training data -------
% Read integrated well-log and CT data
logsctdata = read_matrix("integration_for_porosity_file")
% Transpose the matrix so that columns correspond to samples
logsctdata = transpose(logsctdata)
% Read matrix with porosity values
porosityTrain = read_matrix("porosityTrain_file")
% Transpose the porosity matrix
porosityTrain = transpose(porosityTrain)
% ------- Neural network configuration -------
% Create a feedforward neural network (multilayer perceptron)
% Topology: 6 hidden layers, each with 10 neurons. The number of neurons is adjusted according %    to the architecture of the neural network.
% Training algorithm: Levenberg–Marquardt backpropagation
net = create_network([10, 10, 10, 10, 10, 10], training_algorithm = "trainlm")
% Define dataset split ratios
net.training_ratio = 70%
net.validation_ratio = 15%
net.testing_ratio = 15%
% ------- Network training -------
% Train the neural network using the input and target datasets
(trained_net, training_data) = train(net, logsctdata, porosityTrain)
% ======= END OF PROCESS =======
```

Appendix A.2

Training Pseudocode for Permeability Model:

```
% ======= START =======
% ------- Load training data -------
% Read integrated well-log and CT data
logsctdata = read_matrix("integration_for_permeability_file")
% Transpose the matrix so that columns correspond to samples
logsctdata = transpose(logsctdata)
% Apply natural logarithm to porosity
logsctdata(6, :) = log(logsctdata(6, :))
% Read matrix containing permeability values
permeabilityTrain = read_matrix("permeabilityTrain_file")
% Transpose the permeability matrix
```

permeabilityTrain = transpose(permeabilityTrain)
% Transform permeability by applying the natural logarithm
permeabilityTrain = log(permeabilityTrain)
% ------- Neural network configuration -------
% Create a feedforward neural network (multilayer perceptron)
% Topology: 6 hidden layers, each with 10 neurons. The number of neurons is adjusted according % to the architecture of the neural network.
% Training algorithm: Levenberg–Marquardt backpropagation (trainlm)
% Note: The default activation function in hidden layers is the hyperbolic tangent sigmoid (tansig)
net = create_network([10, 10, 10, 20, 20, 20], training_algorithm = "trainlm")
% Define dataset split ratios
net.training_ratio = 70%
net.validation_ratio = 15%
net.testing_ratio = 15%
% ------- Network training -------
% Train the neural network using the input and target datasets
(trained_net, training_data) = train(net, logsctdata, permeabilityTrain)
% ======= END OF PROCESS =======

References

1. Kohli, A.; Arora, P. Application of Artificial Neural Networks for Well Logs. In Proceedings of the International Petroleum Technology Conference, Doha, Qatar, 19–22 January 2014; University of Petroleum and Energy Studies: Dehradun, India, 2014.
2. Anifowose, F.; Adeniye, S.; Abdulraheem, A.; Al-Shuhail, A. Integrating seismic and log data for improved petroleum reservoir properties estimation using non-linear feature-selection based hybrid computational intelligence models. *J. Pet. Sci. Eng.* **2016**, *145*, 230–237. [CrossRef]
3. Siddiqui, S.; Khamees, A. Dual-Energy CT-Scanning Applications in Rock Characterization. In Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, TX, USA, 26–29 September 2004; SPE-90520-MS. pp. 1–9.
4. Ortiz Meneses, A.; Herrera, E.; Santos Santos, N. Porosity prediction from X-ray computed tomography logs (RHOB and PEF) using Artificial Neural Networks (ANN). *Boletín Geol.* **2020**, *42*, 141–149. [CrossRef]
5. Alyaev, S.; Elsheikh, A.H. Direct multi-modal inversion of geophysical logs using deep learning. *Earth Space Sci.* **2022**, *9*, e2021EA002186. [CrossRef]
6. Vashisth, D.; Mukerji, T. Direct Estimation of Porosity from Seismic Data using Rock and Wave Physics Informed Neural Networks (RW-PINN). *Lead. Edge* **2022**, *41*, 840–846. [CrossRef]
7. Tang, P.; Zhang, D.; Li, H. Predicting permeability from 3D rock images based on CNN with physical information. *J. Hydrol.* **2022**, *606*, 127473. [CrossRef]
8. Wu, L.; Dong, Z.; Li, W.; Jing, C.; Qu, B. Well-Logging Prediction Based on Hybrid Neural Network Model. *Energies* **2021**, *14*, 8583. [CrossRef]
9. Yalamanchi, P.; Datta Gupta, S. Estimation of pore structure and permeability in tight carbonate reservoir based on machine learning (ML) algorithm using SEM images of Jaisalmer sub-basin, India. *Sci. Rep.* **2024**, *14*, 930. [CrossRef]
10. Kalule, R.; Abderrahmane, H.; Alameri, W.; Sassi, M. Stacked ensemble machine learning for porosity and absolute permeability prediction of carbonate rock plugs. *Sci. Rep.* **2023**, *13*, 9855. [CrossRef]
11. Chawshin, K.; Berg, C.F.; Varagnolo, D.; Lopez, O. Automated porosity estimation using CT-scans of extracted core data. *Comput. Geosci.* **2022**, *26*, 595–612. [CrossRef]
12. Zaiery, M.; Kadkhodaie, A.; Arian, M.; Maleki, Z. Application of artificial neural network models and random forest algorithm for estimation of fracture intensity from petrophysical data. *J. Petrol. Explor. Prod. Technol.* **2023**, *13*, 1877–1887. [CrossRef]
13. Azadivash, A.; Soleymani, H.; Seifirad, A.; Sandani, A.; Yahyaee, F.; Kadkhodaie, A. Robust fracture intensity estimation from petrophysical logs and mud loss data: A multi-level ensemble modeling approach. *J. Pet. Explor. Prod. Technol.* **2024**, *14*, 1859–1878. [CrossRef]
14. Salas, R. *Redes Neuronales Artificiales*; Departamento de Computación, Universidad de Valparaíso: Valparaíso, Chile, 2015.
15. Fung, C.C.; Wong, K.W.; Eren, H.; Charlebois, R.; Crocker, H. Modular artificial neural network for prediction of petrophysical properties from well log data. *IEEE Trans. Instrum. Meas.* **1997**, *46*, 1295–1299. [CrossRef]
16. Isasi Viñuela, P.; Galván León, I.M. *Redes Neuronales Artificiales. Un Enfoque Práctico*; Pearson Educación: Hong Kong, China, 2004.

17. Pham, N.; Zabihi-Naeini, E. Missing Well Log Prediction Using Deep Recurrent Neural Networks. In Proceedings of the 81st EAGE Conference and Exhibition, London, UK, 3–6 June 2019; European Association of Geoscientists & Engineers: Utrecht, Netherlands, 2019; pp. 1–5.
18. Chawshin, K.; Berg, C.F.; Varagnolo, D.; López, O. Lithology classification of whole core CT scans using convolutional neural networks. *SN Appl. Sci.* **2021**, *3*, 668. [CrossRef]
19. Abdel Azim, R.; Aljehani, A. Neural Network Model for Permeability Prediction from Reservoir Well Logs. *Processes* **2022**, *10*, 2587. [CrossRef]
20. Gärttner, S.; Alpak, F.O.; Meier, A.; Ray, N.; Frank, F. Estimating permeability of 3D micro-CT images by physics-informed CNNs based on DNS. *Comput. Geosci.* **2023**, *27*, 245–262. [CrossRef]
21. Asquith, G.; Krygowski, D.; Henderson, S.; Hurley, N. *Basic Well Log Analysis*; American Association of Petroleum Geologists: Tulsa, OK, USA, 2004; Volume 16.
22. Ellis, D.V.; Singer, J.M. *Well Logging for Earth Scientists*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2007.
23. Larionov, V.V. *Borehole Radiometry*; Nedra: Moscow, Russian, 1969.
24. Wyllie, M.R.J.; Gregory, A.R.; Gardner, L.W. Elastic Wave Velocities in Heterogeneous and Porous Media. *Geophysics* **1956**, *21*, 41–70. [CrossRef]
25. American Petroleum Institute (API). *Recommended Practice for Core Analysis Procedures (API RP 42)*, 2nd ed.; API Publishing Services: Washington, DC, USA, 1998.
26. Simanjuntak, R.; Irawan, D. Applying Artificial Neural Network and XGBoost to Improve Data Analytics in Oil and Gas Industry. *Indones. J. Energy* **2021**, *4*, 26–35. [CrossRef]
27. Riyadi, Z.A.; Olutoki, J.O.; Hermana, M.; Latif, A.H.A.; Suananda Yogi, I.B.S.; Kadir, S.J.A. Machine learning prediction of permeability distribution in the X field Malay Basin using elastic properties. *Results Eng.* **2024**, *24*, 103421. [CrossRef]
28. Zhao, J.; Wang, Q.; Rong, W.; Zeng, J.; Ren, Y.; Chen, H. Permeability Prediction of Carbonate Reservoir Based on Nuclear Magnetic Resonance (NMR) Logging and Machine Learning. *Energies* **2024**, *17*, 1458. [CrossRef]
29. Nugroho, I.; Trisna, M.; Saroji, S. An Implementation of XGBoost and Random Forest Algorithm to Estimate Effective Porosity and Permeability on Well log data at Fajar Field, South Sumatera Basin, Indonesia. *Indones. J. Appl. Phys.* **2024**, *14*, 271–280. [CrossRef]
30. Xu, C.; Fu, L.; Lin, T.; Li, W.; Ma, S. Machine learning in petrophysics: Advantages and limitations. *Artif. Intell. Geosci.* **2022**, *3*, 157–161. [CrossRef]
31. Huang, C.; Zhu, X.; Lu, M.; Zhang, Y.; Yang, S. XGBoost algorithm optimized by simulated annealing genetic algrithm for permeability prediction modeling of carbonate reservoirs. *Sci. Rep.* **2025**, *15*, 14882. [CrossRef] [PubMed]
32. Hassan, S.; Mohamed, A.; Ibrahim, A.F.; Elkatatny, S. Real-Time Prediction of Petrophysical Properties Using Machine Learning Based on Drilling Parameters. *ACS Omega* **2024**, *9*, 17066–17075. [CrossRef] [PubMed]
33. Edwar, H.; Oms, O.; Eduard, R. Integration of Qualitative and Quantitative Approaches for 3D Geostatistical Modeling of the Ciénaga De Oro Formation, Southern Sinú-San Jacinto Basin, Colombia. *Appl. Sci.* **2025**, *15*, 12374. [CrossRef]