

Iranian Journal of Oil & Gas Science and Technology, Vol. 13 (2024), No. 2, pp. 01–30
<http://ijogst.put.ac.ir>

Modeling the Cementation Exponent Using Pore Structure Descriptions in Heterogeneous Carbonate Formations Through Robust Intelligent Modeling

Alireza Rostami¹, Abbas Helalizadeh³, Mehdi Bahari Moghaddam^{3*}, and Aboozar Soleymanzadeh⁴

¹ PhD of Petroleum Engineering, Department of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran

² Associate Professor of Petroleum Engineering, Department of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran

³ Assistant Professor of Petroleum Engineering, Department of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran

⁴ Director of Research Affairs, Department of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran

Highlights

- The most advanced and hybrid approaches, namely support vector regression (SVR), eXtreme gradient boosting (XGBoost), probabilistic neural networks (PNN), decision trees (DT), multilayer perceptron neural networks trained with the Levenberg–Marquardt algorithm (MLPNN-LM), and random forest (RF), are employed for model development.
- A comprehensive literature database of cementation exponent data for tight and deep carbonate reservoirs is compiled and analyzed.
- Multiple statistical performance metrics and graphical analyses are utilized to evaluate model performance.
- The MLPNN-LM and RF models demonstrate the strongest agreement with the measured cementation exponent data compared with existing literature correlations and other machine learning techniques.
- The developed MLPNN-LM and RF models represent reliable candidates for simulating tight and deep carbonate reservoirs.

Received: February 23, 2024; revised: May 03, 2025; accepted: May 13, 2025

Abstract

In this study, several robust and computationally efficient machine learning methods are developed to estimate the cementation exponent in carbonate formations characterized by heterogeneous pore systems, thereby addressing a critical need for accurate petrophysical property prediction in complex geological settings. Unlike conventional empirical correlations, support vector regression (SVR), multilayer perceptron neural network (MLPNN), probabilistic neural network (PNN), random forest (RF), decision tree (DT), and eXtreme gradient boosting (XGBoost) are implemented as predictive algorithms to simulate the cementation exponent based on detailed pore structure descriptions and total porosity. To optimize the parameters of the MLPNN model, the Levenberg–Marquardt (LM) algorithm is integrated with MLPNN, resulting in a hybrid framework referred to as MLPNN-LM. This hybrid approach enhances parameter optimization, leading to improved predictive accuracy and faster convergence.

The database required for model development, validation, and testing is compiled from Ragland's dataset and is systematically divided into training, validation, and testing subsets. The results indicate high predictive

* Corresponding author:
Email: bahari@put.ac.ir

performance for the hybrid MLPNN-LM and RF models. Key statistical indicators include Average Absolute Percentage Relative Deviation AAPRD values of 4.3781% for MLPNN-LM and 4.8690% for RF, along with coefficients of determination R^2 of 0.8654 and 0.8731, respectively. These findings demonstrate that MLPNN-LM and RF provide more accurate cementation exponent estimates than widely used literature correlations and other machine learning approaches.

Sensitivity analysis reveals that interparticle, moldic, and connected vuggy pore types exert the most significant influence on model predictions. The reliability of the dataset and the robustness of the MLPNN-LM and RF models are further evaluated using Williams plots. Approximately 93.75% and 91.96% of the data points fall within the applicability domain of the MLPNN-LM and RF models, respectively, confirming their generalizability. Trend analysis further demonstrates strong agreement between predicted and measured cementation exponent values, particularly for deep formations with depths greater than 1200 m and tight carbonate reservoirs with total porosity below 10%, where traditional correlations often fail due to the complex behavior of the cementation exponent.

In conclusion, reliable and accurate computational frameworks are introduced for estimating the cementation exponent as a function of pore structure characteristics. These models are well suited for fluid flow simulation and reservoir characterization in carbonate and heterogeneous formations, especially in deep and tight carbonate reservoirs.

Keywords: Cementation exponent; Deep and tight carbonate reservoirs; Random forest (RF); Multilayer perceptron neural network (MLPNN); Levenberg–Marquardt (LM); Pore structure description

How to cite this article

Rostami, A. et al., Modeling the Cementation Exponent Using Pore Structure Descriptions in Heterogeneous Carbonate Formations Through Robust Intelligent Modeling, *Iran J. Oil Gas Sci. Technol.*, Vol. 13, No. 2, p. 1–30, 2024. DOI: <https://doi.org/10.22050/ijogst.2025.445375.1705>

1. Introduction

Evaluating hydrocarbon producibility in carbonate reservoirs using well log interpretation is generally more complex than in clastic formations. Comprehensive analog studies and local to regional geological models typically provide the initial framework for reservoir characterization. Over geological time, diagenetic processes exert a dominant influence on the evolution of carbonate pore systems and fluid mobility within the porous medium (Herrick and Kennedy, 1994; Herrick and Kennedy, 1995; Asquith, 1997). Diagenetic alterations, including cementation, dissolution, and recrystallization, substantially modify the porosity–permeability framework of carbonate rocks and, consequently, control hydrocarbon storage capacity and flow behavior.

As the understanding of these processes advances, accurate determination of water saturation (S_w) in hydrocarbon-bearing carbonate reservoirs increasingly depends on a comprehensive characterization of the rock fabric. The intrinsic properties of carbonate rocks, particularly porosity, permeability, and pore structure, strongly govern fluid distribution and transport within the reservoir. Detailed geological and petrophysical characterization therefore enables more reliable modeling of these properties and, in turn, more accurate estimation of S_w .

Among the various parameters influencing carbonate reservoir behavior, the cementation exponent m has been identified as a key factor in calculating the formation resistivity factor (FRF) and, subsequently, in estimating water saturation (Najafi-Silab et al., 2023; Soleymanzadeh et al., 2018; Kolah-kaj et al., 2022). However, previous attempts to derive generalized correlations for predicting the cementation exponent have encountered substantial limitations. Many of these approaches require prior identification of pore and rock types, which are essential for accurate estimation of the reservoir quality index (RQI). The RQI integrates porosity and permeability measurements to quantify the ease

of fluid flow through the rock matrix. Consequently, uncertainties in pore classification and rock typing directly propagate into errors in cementation exponent estimation and, ultimately, in reservoir quality assessment (Ragland, 2002; Herrick and Kennedy, 1995; Herrick and Kennedy, 1994; Asquith, 1997; Focke and Munn, 1987; Borai, 1987; Watfa and Nurmi, 1987). In addition, many of the proposed formulations are strongly dependent on specific petrophysical log datasets and are calibrated for individual wells or fields, thereby limiting their broader applicability.

In practice, the results derived from existing cementation exponent correlations often lack general applicability across different carbonate formations and geographical regions. In recent years, the need to develop accurate and broadly applicable methods for predicting the cementation exponent, particularly in deep carbonate and tight formations in Western Asia, has become increasingly evident (Soleymanzadeh et al., 2021a; Soleymanzadeh et al., 2021c). Machine learning (ML) has emerged as a powerful modeling tool in petroleum engineering and geosciences. Numerous studies have demonstrated the effectiveness of ML algorithms in predicting various parameters, including permeability (Rostami et al., 2022), aqueous nanofluid viscosity (Mahdavi-Ara et al., 2022), viscoelastic surfactant viscosity (Mahdaviara, Rostami, and Shahbazi, 2021), and carbon dioxide (CO₂) solubility in brine (Sayahi et al., 2021). Among the wide range of available algorithms, multilayer perceptron neural network (MLPNN), decision tree (DT), random forest (RF), probabilistic neural network (PNN), support vector regression (SVR), and eXtreme gradient boosting (XGBoost) have demonstrated strong capabilities in modeling complex nonlinear phenomena.

In this study, we apply MLPNN, DT, RF, PNN, SVR, and XGBoost to develop robust predictive models for estimating the cementation exponent. A comprehensive experimental database was compiled from open literature sources and divided into training (70% of the total dataset), validation (15%), and testing (15%) subsets for model development and performance evaluation (Ragland, 2002). The proposed models are grounded in a detailed understanding of porous media characteristics in potential reservoir formations. Because the dataset was obtained from core samples collected across diverse geographical regions and is closely associated with specific pore structure descriptions, the resulting models are expected to be broadly applicable to carbonate reservoirs in various regions.

The remainder of this paper is organized as follows. Section 2 describes the database used for modeling. Section 3 provides a detailed explanation of the implemented algorithms. Section 4 presents and discusses the modeling results using various post-processing tools. Finally, the key findings of the study are summarized in the Conclusion section.

2. Data presentation

In the present investigation, a comprehensive dataset of petrophysical properties relevant to carbonate reservoirs was compiled from open-source literature (Ragland, 2002). The experimental database includes the cementation exponent (m) as the target variable, together with multiple input variables representing different pore types and associated characteristics of carbonate reservoirs. Moldic (MO) pores result from the dissolution of pre-existing grains or fossils and typically form well-defined voids. Interparticle (BW) pores occur between sediment grains and contribute to overall rock porosity. Fracture (FR) pores are associated with cracks or fractures within the rock matrix and can significantly affect fluid flow. Intracrystalline (WI) pores are located within individual crystals, whereas intercrystalline (IX) pores occur between adjacent crystals. Fenestral (FN) pores are large, irregular cavities commonly observed in tidal and shallow marine carbonate environments. Non-fabric-selective dissolution pores may be either connected (NFS/CN), meaning they are integrated into the interconnected pore network, or isolated (ISVUG), indicating that they are disconnected from the primary pore system. Total porosity (PHIT) represents the total void volume within the rock.

Collectively, these variables influence the cementation exponent and are critical for accurately modeling fluid distribution and mobility in carbonate reservoirs.

The database comprises 112 experimental data points obtained from Ragland (2002), including 98 samples collected from ten wells in the United States and two international locations. The remaining 14 data points were obtained from Reservoirs Inc.'s carbonate rock catalog (Ragland, 2002). Prior to model development, the data were normalized to a range between 0 and 1 and randomly divided into training, validation, and test subsets to ensure a homogeneous distribution. Specifically, 70% of the dataset was allocated to the training subset, 15% to the validation subset, and the remaining 15% to the test subset.

The compiled databank includes two major carbonate lithologies, ranging from limestone to dolomite, each characterized by diverse rock textures. These data provide a robust foundation for developing generalized predictive models. A visual representation of variations in the cementation exponent, together with selected input parameters, is presented in two-dimensional heatmap diagrams in Figure 1. The results indicate that higher cementation exponent values are associated with increased moldic porosity, higher total porosity, lower connected vuggy (non-fabric-selective dissolution) porosity, and reduced interparticle porosity.

3. Developing the intelligent model

3.1. Multilayer perceptron neural network

A multilayer perceptron neural network is a type of artificial neural network that employs a feedforward architecture with multiple layers of neurons. These layers include an input layer, one or more hidden layers, and an output layer. By introducing nonlinearity into the model through activation functions, such as the sigmoid or rectified linear unit (ReLU), each neuron in a hidden layer enables the network to learn complex, nonlinear relationships in the data (Stebakov et al., 2022; Li et al., 2022; Chang, 2022).

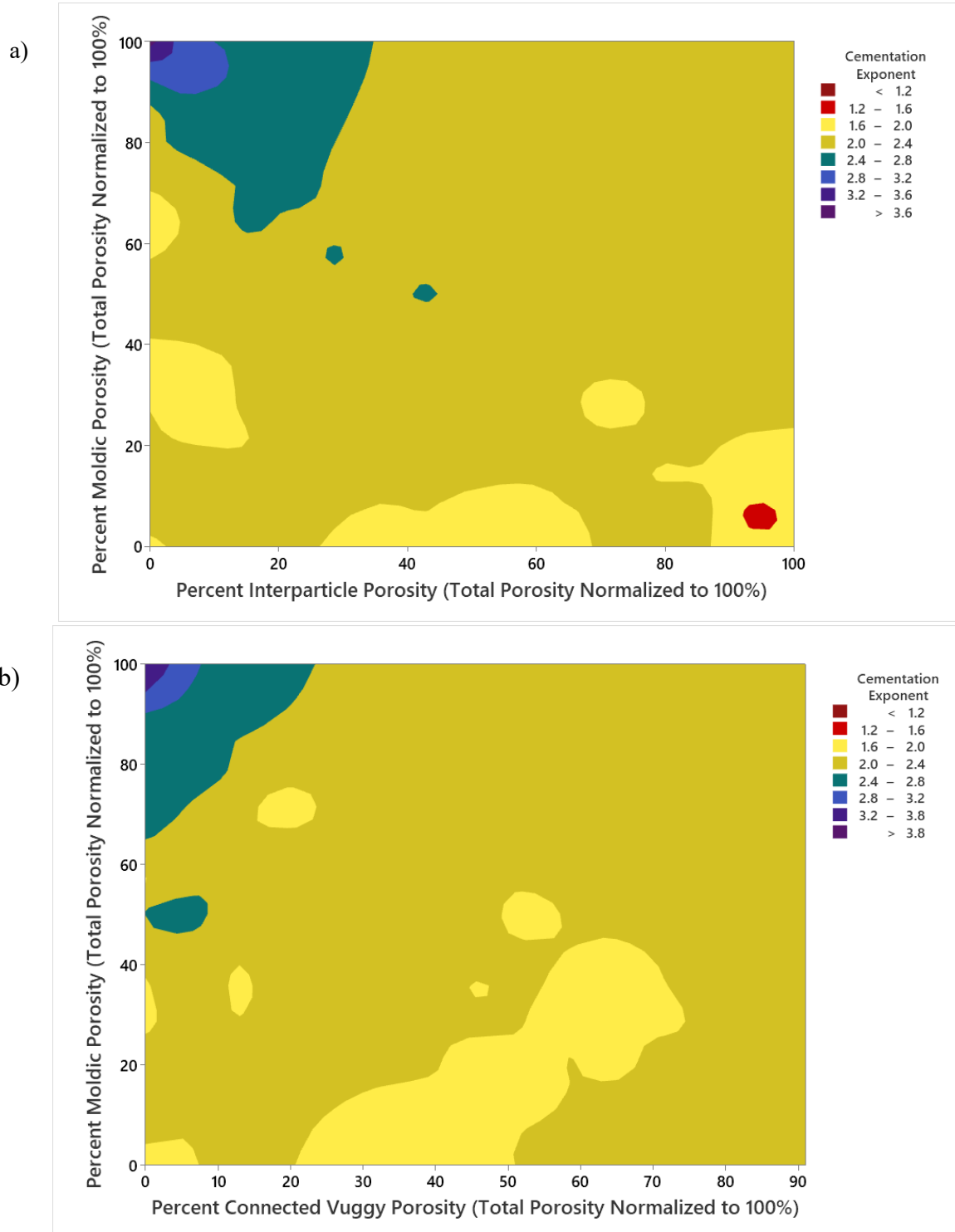
The Levenberg–Marquardt (LM) algorithm is an advanced optimization technique commonly used to train neural networks by minimizing the error function. The LM algorithm combines the strengths of the Gauss–Newton method and gradient descent to achieve faster and more accurate convergence. The training of an MLPNN aims to minimize the sum of squared differences between the network's predicted outputs and the actual outputs. During training, the LM algorithm iteratively updates the network's weights and biases to reduce this error (Fischer et al., 2024; Lv and Yu, 2024; Huang et al., 2023).

One of the main advantages of using the LM algorithm for MLPNN training is its computational efficiency and speed, particularly for small- to medium-sized networks. It is especially effective for models where the error surface is smooth and continuous. The algorithm addresses such surfaces by linearizing the error function at each iteration and solving a linear system to update the parameters, resulting in more stable and faster convergence to a local or global minimum compared with standard backpropagation methods (Fischer et al., 2024; Lv and Yu, 2024; Huang et al., 2023).

The combination of MLPNN and the LM algorithm provides a powerful modeling framework applicable to diverse fields, including pattern recognition, time series prediction, and control systems. Its ability to capture complex data patterns while maintaining computational efficiency makes this approach a valuable tool in machine learning and artificial intelligence (Fischer et al., 2024; Lv and Yu, 2024; Huang et al., 2023).

3.2. Decision tree algorithm

Decision tree regression is a widely used and intuitive supervised learning algorithm for predicting continuous numerical values. The algorithm constructs a tree-like structure in which each internal node represents a decision rule, and each leaf node corresponds to a predicted value. This structure is created by recursively partitioning the data into subsets based on the values of input features. The root node, representing the entire dataset, serves as the starting point of the tree. The splitting process aims to maximize the homogeneity of resulting subsets, typically by minimizing criteria such as mean squared error (Costa and Pedreira, 2023; Blockeel et al., 2023; Loh, 2011).



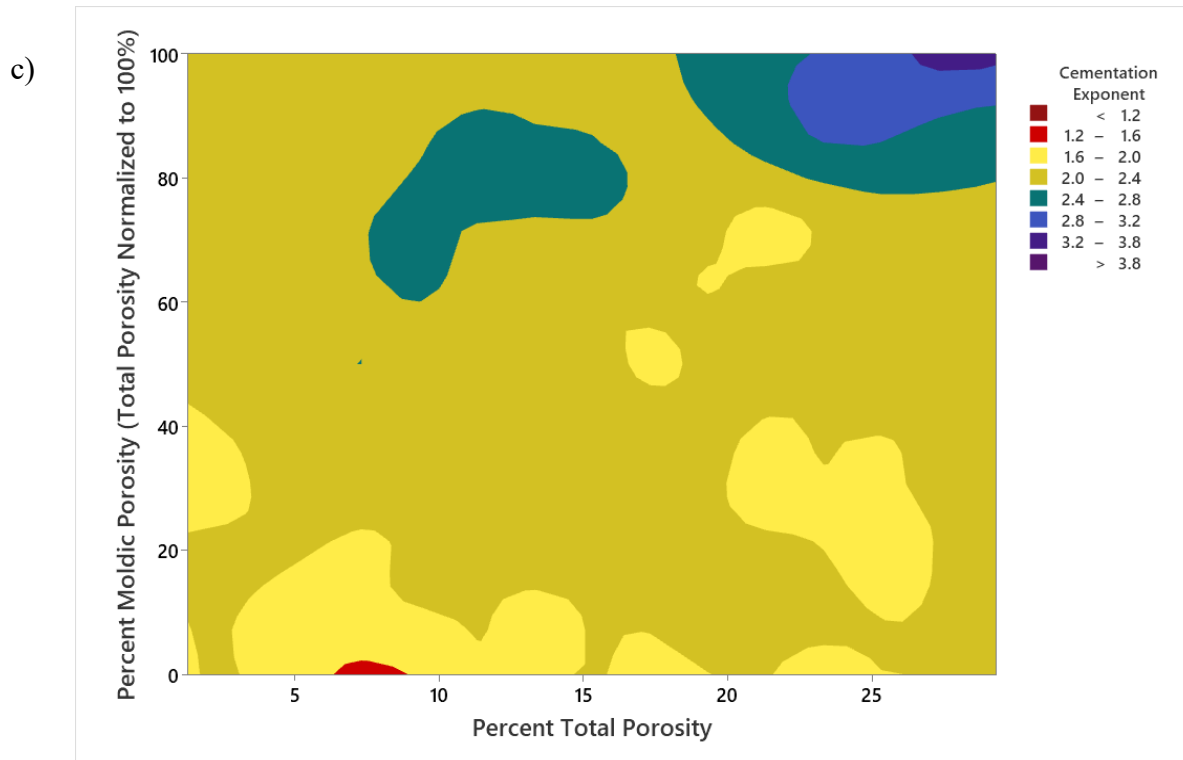


Figure 1

Two-dimensional heatmap diagrams illustrating variations in the cementation exponent with respect to key input variables: (a) cementation exponent versus interparticle (BW) and moldic (MO) pore types, (b) cementation exponent versus connected vuggy (NFS/CN) and moldic pore types, and (c) cementation exponent versus total porosity (PHIT) and moldic pore type (Ragland, 2002).

Decision trees are advantageous because they are easy to understand and interpret, making them suitable for explaining predictions to non-technical stakeholders. They can capture complex, nonlinear relationships between input features and the target variable and can handle both numerical and categorical data. However, decision trees have certain limitations. They are prone to overfitting, particularly when the tree is deep, which can result in poor generalization on unseen data. Additionally, decision trees can be unstable, as small variations in the dataset may produce significantly different tree structures (Costa and Pedreira, 2023; Blockeel et al., 2023; Loh, 2011). Despite these challenges, decision tree regression remains a popular choice due to its simplicity and effectiveness. Techniques such as pruning, ensemble methods (e.g., random forests), and boosting can be employed to reduce overfitting and improve model stability and performance (Costa and Pedreira, 2023; Blockeel et al., 2023; Loh, 2011).

3.3. Random forest algorithm

Random forest constructs multiple decision trees using a technique known as bootstrap aggregating, or bagging. For each tree, a random subset of the data is selected, and at each split, a random subset of features is considered (Breiman, 2001; Biau, 2012; Iranzad and Liu, 2024). This randomness promotes diversity among the trees, and combining their predictions reduces model variance and enhances predictive accuracy.

One of the primary advantages of random forest is its ability to handle large datasets with numerous features while minimizing the risk of overfitting. It can capture complex, nonlinear relationships between input features and the target variable, making it suitable for a wide range of applications

(Breiman, 2001; Biau, 2012; Iranzad and Liu, 2024). Furthermore, random forest provides feature importance scores, offering insights into the relative contribution of each feature to the predictions. However, random forest models can be computationally demanding, particularly for very large datasets, and they are generally less interpretable than individual decision trees due to the ensemble's complexity (Breiman, 2001; Biau, 2012; Iranzad and Liu, 2024).

3.4. Probabilistic neural network algorithm

Probabilistic neural networks are a class of neural networks designed to model uncertainty in predictions, making them particularly suitable for regression tasks. Unlike standard neural networks, which produce deterministic outputs, PNNs generate probability distributions for the target variable, allowing the derivation of both expected values and confidence intervals. This capability enables PNNs to quantify aleatoric uncertainty, which represents the inherent unpredictability in the input–output relationships of a system.

PNNs employ a kernel-based approach to estimate the probability density function of the target variable. The network architecture consists of four layers: an input layer, a pattern layer, a summation layer, and an output layer (Pourkamali-Anaraki et al., 2024; Chang, 2021). The input layer receives the features, the pattern layer calculates distances between the input and training samples, the summation layer aggregates the contributions of the pattern layer neurons, and the output layer produces the probability distribution of the target variable.

A key advantage of PNNs is their ability to provide accurate predictions with quantified uncertainty, which is critical in fields such as materials science, healthcare, and environmental engineering. PNNs are relatively straightforward to implement and can be trained more rapidly than many other neural network models. However, they can be memory-intensive and computationally demanding, particularly when handling large datasets. Additionally, the performance of PNNs is highly dependent on the choice of kernel function and its parameters, requiring careful tuning.

PNNs have been successfully applied in various scientific and engineering domains, including material science and structural engineering, where modeling complex input–output relationships is essential. Their ability to deliver reliable uncertainty estimates makes them especially valuable in applications where safety and reliability are critical (Pourkamali-Anaraki et al., 2024; Chang, 2021).

3.5. Support vector regression algorithm

Support vector regression is a supervised learning method used to predict continuous numerical values. It is an extension of the support vector machine (SVM) algorithm, which is primarily applied to classification tasks. SVR constructs a regression function that approximates the data within a specified tolerance by mapping the input features into a high-dimensional feature space using a kernel function. The algorithm employs an ϵ -insensitive loss function, meaning that errors within a defined margin (ϵ) are ignored, while deviations beyond this threshold are penalized. This approach enhances the model's robustness to noise and outliers. SVR seeks a function that is as smooth as possible while maintaining a maximum ϵ deviation from the true target values for all training data points (Smola and Schölkopf, 2004; Awad and Khanna, 2015; Yan et al., 2018).

A major advantage of SVR is its ability to capture both linear and nonlinear relationships between input features and the target variable through the use of kernel functions. Commonly used kernels include polynomial, linear, sigmoid, and radial basis function (RBF) kernels. SVR also exhibits strong generalization capability, performing well on unseen data. Furthermore, it provides sparse solutions, as only a subset of the training data (the support vectors) defines the regression function, improving

computational efficiency. However, SVR has limitations. It can be computationally intensive, especially with complex kernels and large datasets, and its performance is sensitive to kernel choice and hyperparameters such as the regularization parameter (C) and kernel parameter (γ). Additionally, SVR models are generally less interpretable than simpler regression approaches (Smola and Schölkopf, 2004; Awad and Khanna, 2015; Yan et al., 2018).

3.6. eXtreme gradient boosting algorithm

eXtreme gradient boosting is an advanced implementation of gradient boosting decision trees designed for efficiency and scalability. Developed by Chen and Guestrin (2016), XGBoost is widely recognized for its ability to handle large, complex datasets effectively. The algorithm constructs an ensemble of decision trees sequentially, with each successive tree correcting the errors of its predecessors (Brownlee, 2021; Chen, 2016). This iterative process continues until a specified stopping criterion is met, resulting in a highly accurate predictive model. XGBoost incorporates various regularization techniques to prevent overfitting and supports parallel processing, enhancing computational efficiency and scalability (Brownlee, 2021; Chen, 2016).

4. Results and discussion

In this section, we present and highlight the principal findings of the study.

4.1. Establishment of the proposed models

The primary objective of this investigation is to analyze the cementation exponent in carbonate pore systems, considering its dependence on various rock properties. Previous studies have extensively examined factors influencing the cementation exponent in carbonate reservoirs (Ragland, 2002; Najafi-Silab et al., 2023; Soleymanzadeh et al., 2021c; Kolah-kaj et al., 2022; Soleymanzadeh et al., 2021a; Soleymanzadeh et al., 2018; Herrick and Kennedy, 1994; Borai, 1987; Watfa and Nurmi, 1987; Asquith, 1997; Focke and Munn, 1987). A consistent trend observed in these studies highlights a strong correlation between the cementation exponent and parameters such as total porosity and specific pore types, including moldic, interparticle, and non-fabric-selective dissolution pores (Ragland, 2002; Najafi-Silab et al., 2023; Soleymanzadeh et al., 2021c; Kolah-kaj et al., 2022; Soleymanzadeh et al., 2021a; Soleymanzadeh et al., 2018; Herrick and Kennedy, 1994; Borai, 1987; Watfa and Nurmi, 1987; Asquith, 1997; Focke and Munn, 1987). To explore these relationships further, we compiled a comprehensive dataset from established open scientific literature (Chilingarian et al., 1990).

To quantify the relationship between the cementation exponent and input variables in the compiled database, we conducted a sensitivity analysis using Pearson correlation coefficients (Chok, 2010). Figure 2 presents the results of this analysis for the hybrid multilayer perceptron neural network with Levenberg–Marquardt optimization (MLPNN-LM) and RF models. The findings indicate that different input parameters contribute variably to the estimation of the cementation exponent, consistent with experimental observations. Among these variables, interparticle, moldic, and connected vuggy pore types are the most influential. In contrast, intracrystalline pores exhibit the lowest contribution. Nevertheless, due to the stochastic nature of carbonate rocks, some data points deviate from established trends, which adds complexity to accurate cementation exponent estimation.

In this study, six robust modeling approaches were applied: support vector regression, a hybrid multilayer perceptron neural network with Levenberg–Marquardt optimization, probabilistic neural network, random forest, decision tree, and eXtreme gradient boosting. The aim was to comprehensively analyze the cementation exponent in carbonate reservoirs. For model development, the dataset was randomly divided into training (70%), validation (15%), and test (15%) subsets. The input parameters

included total porosity and various pore descriptors: interparticle, moldic, fenestral, intercrystalline, fracture, intracrystalline, isolated vuggy, and connected vuggy. These parameters demonstrated significant contributions to cementation exponent prediction, as discussed in the following sections.

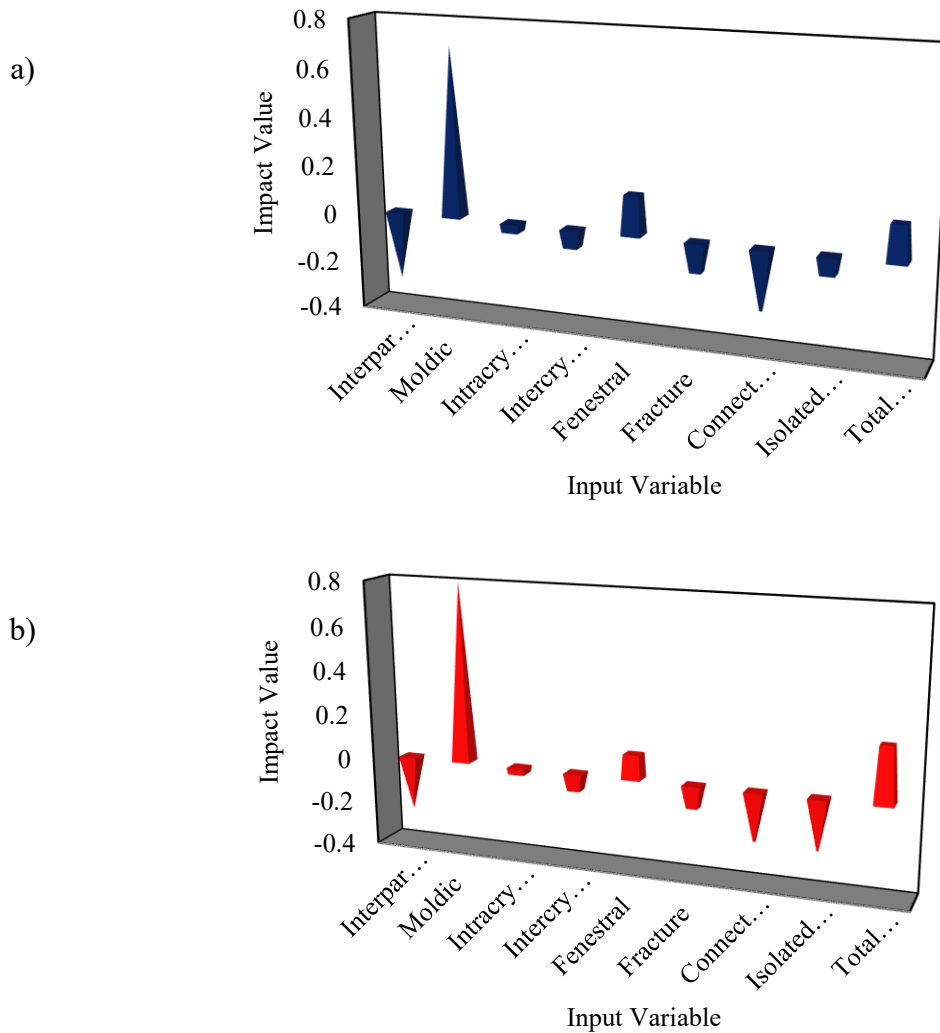


Figure 2

Relative correlation of the influencing parameters with the measured cementation exponent values for the implemented models: (a) MLPNN-LM and (b) RF.

$$m = f(PHIT, MO, BW, WI, IX, FN, FR, NFS/CN, ISVUG) \quad (1)$$

In Equation 1, the cementation exponent is represented by the symbol m . The sensitivity analysis results shown in Figure 2 indicate that certain variables, including intercrystalline, intracrystalline, and isolated vuggy pore types, exert negligible influence on the output estimation in both MLPNN-LM and RF models.

In this study, several robust and hybrid modeling approaches were employed to develop predictive models for the cementation exponent. These include the multilayer perceptron neural network with Levenberg–Marquardt optimization, probabilistic neural network, random forest, support vector regression, eXtreme gradient boosting, and decision tree. Prior to modeling, all input data were normalized to a range between 0 and 1 using the following definition:

$$X_N = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

In this expression, X represents the input and/or output parameters, while the subscripts N , max, and min denote the normalized, maximum, and minimum values of the parameter considered for normalization, respectively.

Furthermore, to quantify the sensitivity of Archie's parameters—including the cementation exponent, tortuosity factor, brine resistivity, true formation resistivity, and porosity—on water saturation calculations, Chen and Fang (1986) proposed a set of mathematical formulations. According to their study, the variance in water saturation can be decomposed into six components, with the fractional contribution of each component expressed as follows:

$$E_a = \left[\frac{\sigma_a}{a} \right]^2 \quad (3)$$

$$E_{R_w} = \left[\frac{\sigma_{R_w}}{R_w} \right]^2 \quad (4)$$

$$E_{R_t} = \left[\frac{\sigma_{R_t}}{R_t} \right]^2 \quad (5)$$

$$E_{PHIT} = \left[m \times \frac{\sigma_{PHIT}}{PHIT} \right]^2 \quad (6)$$

$$E_m = [\sigma_m \times \ln(PHIT)]^2 \quad (7)$$

$$E_n = [\sigma_n \times \ln(S_w)]^2 \quad (8)$$

In this formulation, the fractional error contribution, variance, brine resistivity, true formation resistivity, tortuosity factor, cementation exponent, saturation exponent, total porosity, and water saturation are denoted by E , σ , R_w , R_t , a , m , n , $PHIT$, and S_w , respectively. Assuming identical measurement uncertainties for all parameters in Archie's saturation equation, the cementation exponent (m) is found to have the largest impact on the variance of water saturation, excluding the effect of the saturation exponent (n). In contrast, the parameters R_w , R_t , and a contribute equally to the error in water saturation estimation. The influence of the saturation exponent n depends on the actual value of S_w^n and may represent either the least or the most critical factor. For porosity values below 7.5%, the cementation exponent exhibits the highest fractional error contribution to water saturation. However, for porosity values exceeding 7.5%, the saturation exponent becomes the dominant factor affecting water saturation estimation. For a detailed discussion, readers are referred to the comprehensive sensitivity analysis conducted by Chen and Fang (1986).

4.2. Validity analysis of the established methods

4.2.1. Statistical quality measures

The statistical validity of the proposed MLPNN-LM model was assessed using the following conventional statistical metrics:

$$MAD = \frac{\sum_{i=1}^N |m_i^{exp} - m_i^{pred}|}{N} \quad (9)$$

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(\frac{m_i^{exp} - m_i^{pred}}{m_i^{exp}} \right)^2} \quad (10)$$

$$APRD\% = \frac{100}{N} \sum_{i=1}^N \left(\frac{m_i^{exp} - m_i^{pred}}{m_i^{exp}} \right) \quad (11)$$

$$AAPRD\% = \frac{100}{N} \sum_{i=1}^N \left| \frac{m_i^{exp} - m_i^{pred}}{m_i^{exp}} \right| \quad (12)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (m_i^{exp} - m_i^{pred})^2 \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (m_i^{exp} - m_i^{pred})^2} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (m_i^{pred} - m_i^{exp})^2}{\sum_{i=1}^N (m_i^{pred} - \overline{m_i^{exp}})^2} \quad (15)$$

Using Equations 9–15, the statistical performance of the proposed MLPNN-LM model was evaluated. Specifically, the metrics employed include the average absolute percentage relative deviation (AAPRD), root mean square error (RMSE), average percentage relative deviation (APRD), mean absolute deviation (MAD), standard deviation (SD), and coefficient of determination (R^2). Experimentally measured and predicted values of the cementation exponent are denoted by the superscripts *exp* and *pred*, respectively.

The analysis of the parametric statistical values presented in Table 1 indicates that the MLPNN-LM model predicts the cementation exponent with the highest accuracy, despite the inherent challenges associated with estimating this parameter in carbonate formations. Among the alternative approaches, the random forest model demonstrates the next highest precision. Both MLPNN-LM and RF models provide highly accurate and consistent results, with the MLPNN-LM achieving $R^2 = 0.8654$, AAPRD = 4.3781%, SD = 0.0745, and RMSE = 0.1413, while the RF method attains $R^2 = 0.8731$, AAPRD = 4.8690%, SD = 0.0803, and RMSE = 0.1502.

Table 1 further shows that the PNN model, with an AAPRD of 91.13%, and the SVR model, with an AAPRD of 10.63%, exhibit substantial errors and are therefore unsuitable for further analysis. Similarly, the decision tree and XGBoost models display unreliable outcomes due to large discrepancies between the training, validation, and test subsets, indicative of model overfitting. Consequently, only the MLPNN-LM and RF approaches are considered for subsequent analysis.

4.2.2. Visual analysis

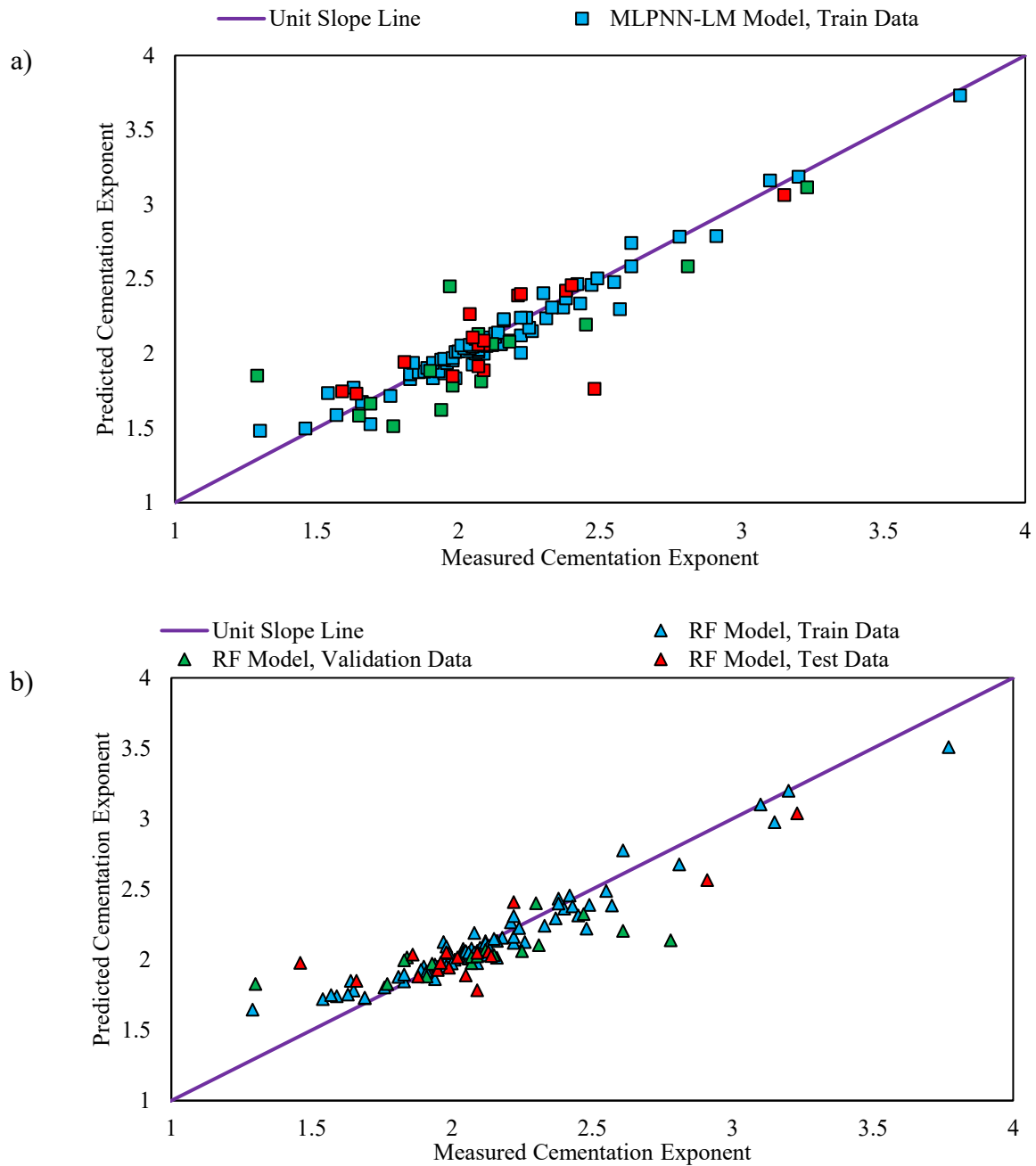
In this study, the parametric analysis is complemented by a visual evaluation of the proposed methodology's performance. Several visualization tools were employed for this purpose, including cross plots, index plots for comparing measured and predicted values, and relative error distributions. Specifically, Figure 3 presents the cross plots for the MLPNN-LM and RF models across the training, validation, and test subsets.

Table 1

Statistical parameters of the proposed models for the training, validation, and test subsets.

Model	Parameter	Train	Validation	Test	All
MLPNN-LM	MAD	0.0554	0.1817	0.1448	0.0881
	SD	0.0399	0.1432	0.0937	0.0745
	APRD%	0.5944	1.0936	0.0174	0.5826
	AAPRD%	2.7076	9.6718	6.7490	4.3781
	MSE	0.0061	0.0576	0.0459	0.0199
	RMSE	0.0783	0.2399	0.2141	0.1413
	R^2	0.9584	0.7058	0.6407	0.8654
RF	MAD	0.0699	0.1808	0.1455	0.0985
	SD	0.0535	0.1304	0.1102	0.0803
	APRD%	-0.5429	0.7569	-0.8297	-0.3877
	AAPRD%	3.4454	8.8685	7.3178	4.8690
	MSE	0.0099	0.0625	0.0397	0.0225
	RMSE	0.0997	0.2500	0.1992	0.1502
	R^2	0.9469	0.5723	0.7692	0.8731
DT	MAD	0.0000	0.2782	0.1388	0.0627
	SD	0.0000	0.1923	0.0888	0.0825
	APRD%	0.0000	1.7740	-1.3416	0.0826
	AAPRD%	0.0000	13.7619	7.1263	3.1392
	MSE	0.0000	0.1438	0.0279	0.0260
	RMSE	0.0000	0.3792	0.1669	0.1613
	R^2	1.0000	0.0828	0.8383	0.8256
XGBOOST	MAD	0.0008	0.1583	0.1993	0.0553
	SD	0.0005	0.1308	0.1179	0.0689
	APRD%	-0.0025	0.6028	2.0966	0.4117
	AAPRD%	0.0368	7.8154	9.5356	2.6829
	MSE	0.0000	0.0637	0.0675	0.0201
	RMSE	0.0011	0.2524	0.2599	0.1418
	R^2	0.9999	0.4927	0.6136	0.8675
PNN	MAD	1.9276	1.9346	2.0109	1.9413
	SD	0.8944	0.9387	0.9683	0.9128
	APRD%	89.3566	93.6492	96.7544	91.1311
	AAPRD%	89.3566	93.6492	96.7544	91.1311
	MSE	3.8616	3.8841	4.1499	3.9088
	RMSE	1.9651	1.9708	2.0372	1.9771
	R^2	0.0473	0.2296	0.3475	0.0369
SVR	MAD	0.2025	0.2276	0.3024	0.2216
	SD	0.1403	0.1562	0.2019	0.1537

Model	Parameter	Train	Validation	Test	All
	APRD%	-1.2380	0.1721	3.6450	-0.2742
	AAPRD%	9.5326	11.7068	14.5106	10.6280
	MSE	0.0869	0.0739	0.1759	0.0986
	RMSE	0.2948	0.2719	0.4194	0.3139
	R^2	0.4355	0.3729	0.2159	0.4099

**Figure 3**

Cross plots of the predicted versus measured cementation exponent for the implemented models: (a) MLPNN-LM and (b) RF.

Notably, the data points in these subsets closely align with the $Y = X$ line (45-degree line), indicating a strong agreement between the experimentally measured cementation exponent values and the predictions produced by the MLPNN-LM and RF models. Furthermore, Figure 4 presents the relative deviations of the predicted values from the corresponding experimental targets, providing additional evidence of the accuracy and reliability of the established MLPNN-LM and RF approaches.

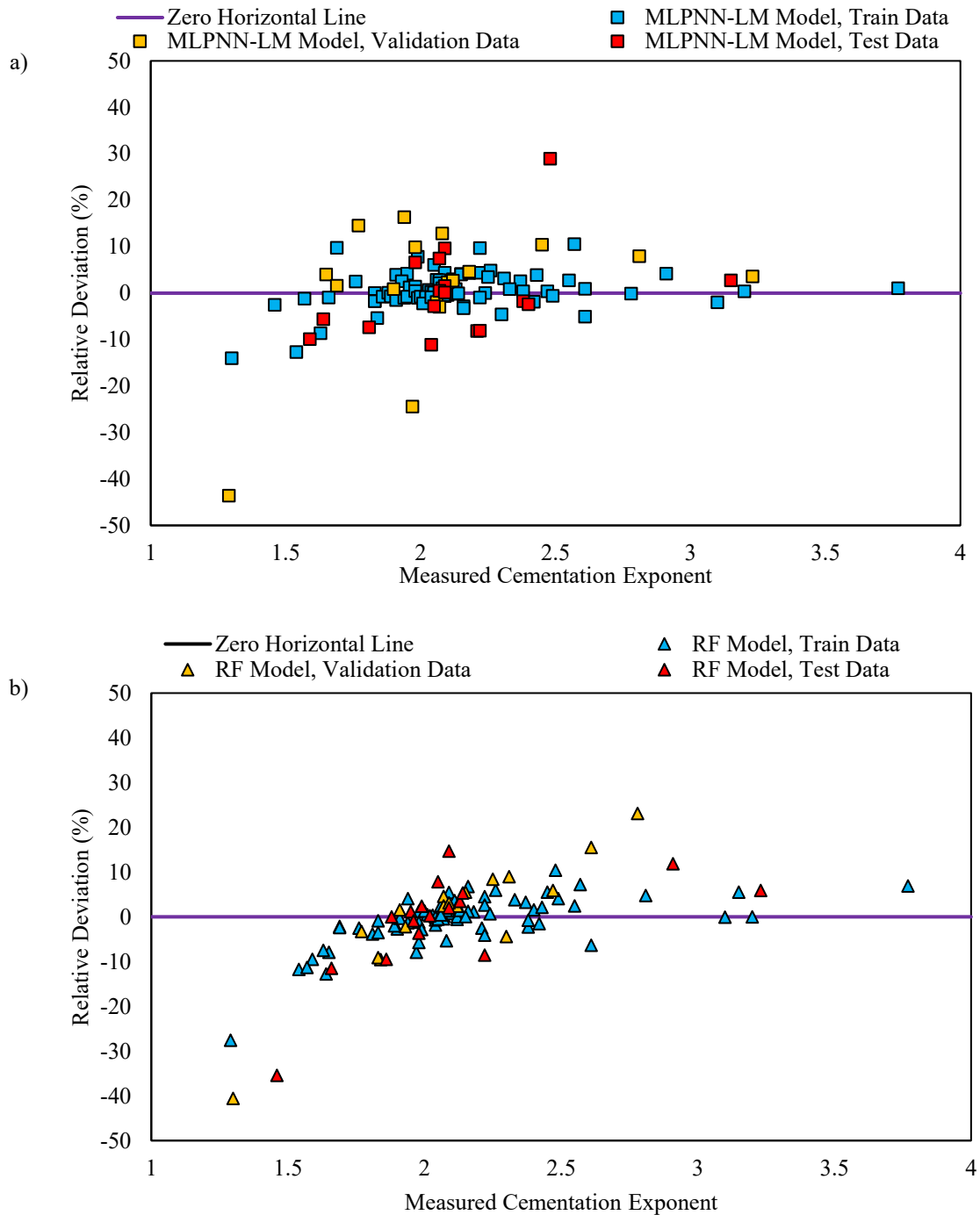


Figure 4

Distribution of the relative error between predicted and measured cementation exponent values for the training, validation, and test subsets: (a) MLPNN-LM and (b) RF.

Notably, the clustering of data points around the zero-error line indicates a strong correspondence between the predicted and actual cementation exponent values. Overall, these results demonstrate the robustness of the proposed approaches for estimating the cementation exponent in heterogeneous carbonate pore systems. Both MLPNN-LM and RF models provide highly accurate predictions, as evidenced by the close alignment of the estimated values with the $Y = X$ line in Figure 3 and the zero-error line in Figure 4. The associated relative errors for MLPNN-LM and RF predominantly fall within approximately -14% to $+16\%$ and -11.5% to $+15.5\%$, respectively. Figure 5 offers an additional graphical evaluation of model precision, showing that the training, validation, and test subsets all exhibit strong agreement with the experimental dataset. These results confirm the capability of the proposed techniques to reliably predict the cementation exponent in heterogeneous carbonate pore systems.

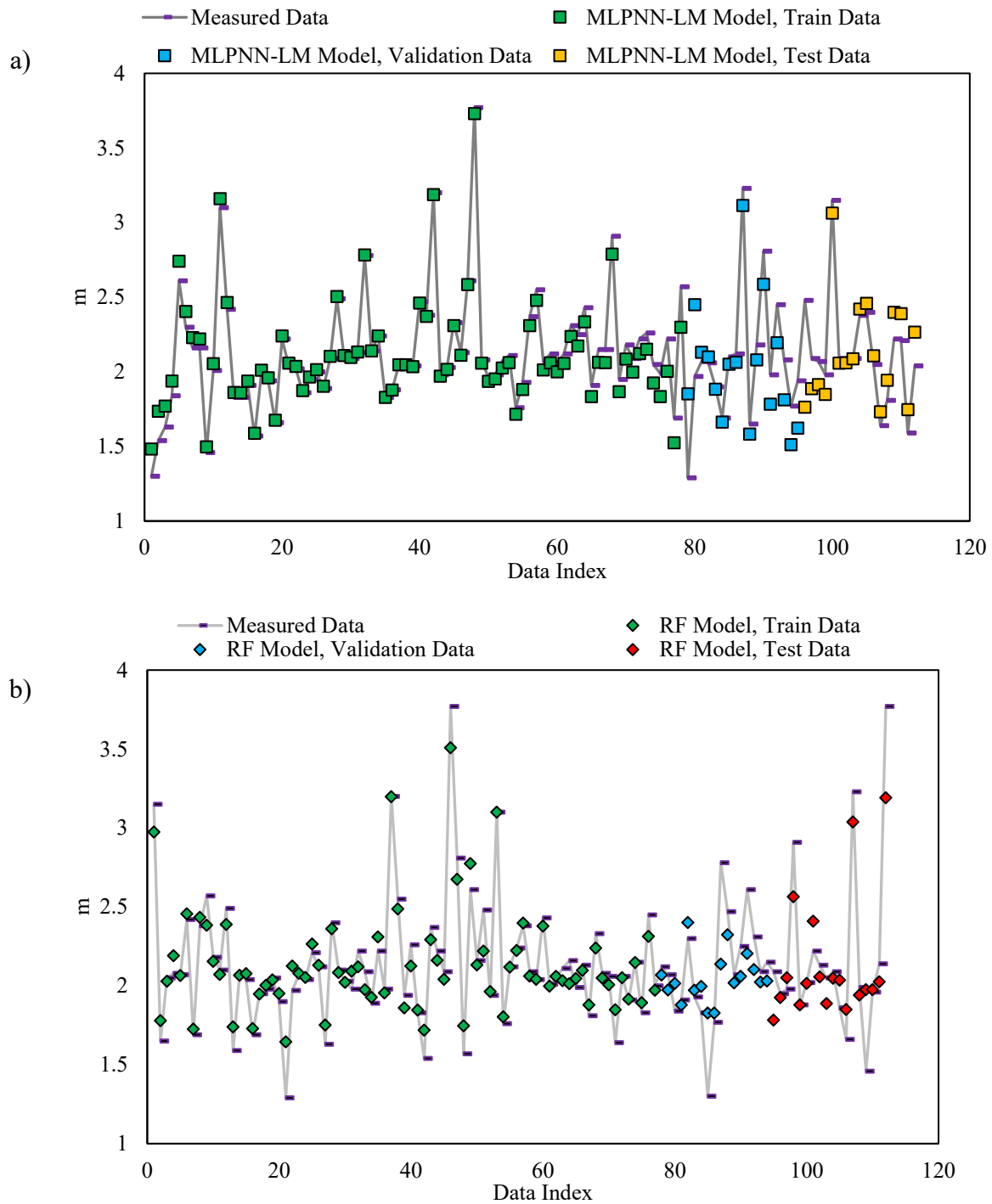
The advantages of the MLPNN-LM and RF algorithms extend beyond their high accuracy and resistance to overfitting. These methods are also highly adaptable to diverse geographical regions and can be readily updated as new data become available. Unlike many traditional literature correlations and models that rely solely on total porosity as an input, this study incorporates detailed pore-type descriptions to improve predictive performance. Moreover, the analysis indicates that total porosity has a relatively minor influence on cementation exponent estimation compared to moldic, interparticle, and connected vuggy porosities.

4.2.3. Outlier detection

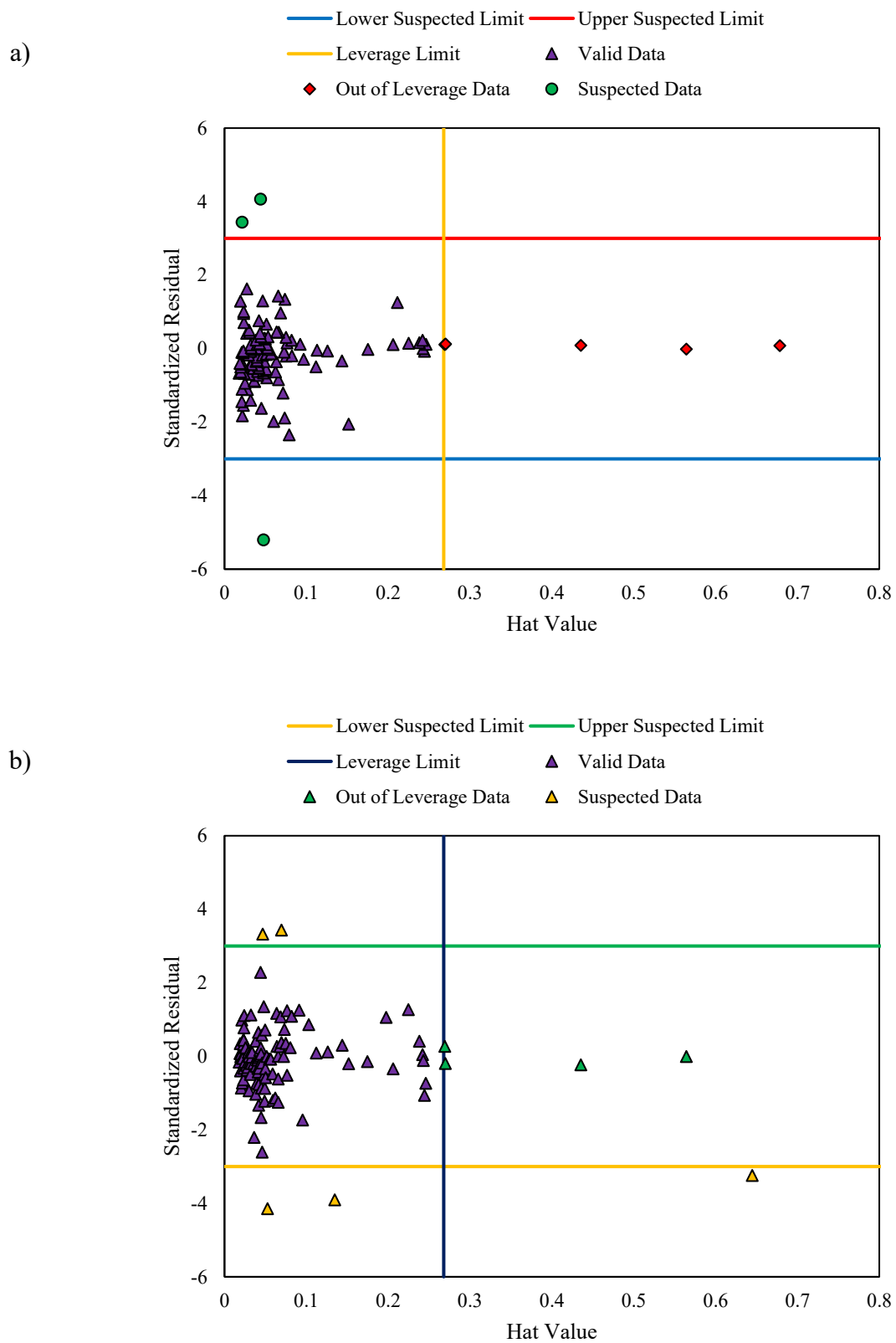
In scientific research, evaluating data points that deviate significantly from the general trend—commonly referred to as outliers—is critical for several technical reasons. First, outliers may indicate errors or anomalies in data collection, measurement, or recording. Identifying these points allows researchers to determine whether they represent genuine observations or erroneous data requiring correction. Second, outliers capture the natural variability within the dataset. Recognizing and accounting for this variability contributes to the development of more robust models that better reflect real-world conditions. Moreover, outliers can substantially influence statistical analyses and predictive model performance. Ignoring such data points may lead to biased results, incorrect conclusions, or inaccurate predictions. Consequently, proper assessment of outliers is essential for ensuring reliable research outcomes. Additionally, outliers may reveal unexpected phenomena or patterns, providing opportunities for new discoveries and advancements. Finally, evaluating outliers aids in validating the performance and generalizability of predictive models by illustrating how models respond to atypical observations.

These anomalous points may arise from various sources, including measurement or experimental errors. Detecting and addressing these outliers is crucial to ensure the reliability and accuracy of the proposed methodologies (Rousseeuw and Leroy, 1987; Gramatica, 2007; Hemmati-Sarapardeh et al., 2016; Goodall, 1993). One effective technique for outlier detection is the leverage method. For a detailed discussion of outlier analysis principles, readers are referred to the cited literature (Rousseeuw and Leroy, 1987; Gramatica, 2007; Hemmati-Sarapardeh et al., 2016; Goodall, 1993).

In this study, outliers were identified using Williams' plot in combination with the leverage approach. Figure 6 presents the Williams' plot for the MLPNN-LM model, where the standardized residuals are plotted against the hat values (leverage indices) to facilitate diagnostic interpretation. As shown in Figure 6, approximately 93.75% of the dataset for the MLPNN-LM model and 91.96% for the RF model fall within the rectangular box defining the applicability domain. This observation further confirms the consistency and reliability of the developed MLPNN-LM and RF models across the employed dataset.

**Figure 5**

Assessment of the agreement between predicted and measured cementation exponent values for the developed models: (a) MLPNN-LM and (b) RF.

**Figure 6**

Williams' plots for the implemented models: (a) MLPNN-LM and (b) RF.

4.2.4. Comparison with existing methods

In this study, the performance of the proposed models is evaluated through a comprehensive comparison with established methods reported in the scientific literature. A selection of widely used empirical and semi-empirical models is presented in prior studies (Rostami et al., 2024; Ragland, 2002). For further details regarding the specific correlations analyzed in this work, readers are referred to the original sources (Rostami et al., 2024; Ragland, 2002).

Beyond traditional empirical models, Rostami et al. (2024) recently explored advanced machine learning approaches, including gene expression programming (GEP) and a radial basis function neural network coupled with ant colony optimization (RBFNN-ACO), for estimating the cementation exponent. The outcomes of these literature models are summarized in Table 2 to facilitate comparative analysis.

The validity of the literature models presented in Table 2 was evaluated using multiple statistical error metrics, including RMSE, AAPRD%, APRD%, SD, and MAD. Based on these metrics, the models can be ranked in terms of predictive accuracy as follows:

MLPNN-LM > RF > RBFNN-ACO (Rostami et al., 2024) > GEP-I (Rostami et al., 2024) > GEP-II (Rostami et al., 2024) > Borai equation (Borai, 1987) > Nugent equation (Ragland, 2002) > Shell equation (Neustaedter, 1968; Watfa and Nurmi, 1987) > Asquith equation (Asquith, 1997) > Focke and Munn equation (Focke and Munn, 1987).

Table 2

Statistical parameters of the newly proposed models in comparison with conventional correlations and existing machine learning models reported in the literature.

Model	AAPRD%	SD	APRD%	MAD	RMSE
MLPNN-LM	4.3781	0.0749	0.5826	0.0881	0.1413
RF	4.8690	0.0803	-0.3877	0.0985	0.0803
RBFNN-ACO (Rostami et al., 2024)	6.2800	0.0900	-0.2300	0.1300	0.1700
GEP-I (Rostami et al., 2024)	6.3900	0.0900	-0.5900	0.1300	0.1700
GEP-II (Rostami et al., 2024)	7.4500	0.1100	-1.6000	0.1500	0.2100
Borai Eqn. (Borai, 1987)	12.3022	0.1702	3.6035	0.2759	0.4061
Nugent Eqn. (Ragland, 2002)	16.2853	0.1990	10.6681	0.3604	0.4629
Shell Eqn. (Neustaedter, 1968; Watfa and Nurmi, 1987)	20.2914	0.2446	18.2422	0.4634	0.5940

Model	AAPRD%	SD	APRD%	MAD	RMSE
Asquith Eqn. (Asquith, 1997)	28.4128	0.3855	23.6334	0.6044	0.7924
Focke and Munn Eqn. (Focke and Munn, 1987)	31.3222	0.3908	-22.9935	0.6278	0.7497

In heterogeneous carbonate pore systems, the MLPNN-LM model exhibits the highest precision in estimating the cementation exponent, outperforming all other literature correlations and machine learning approaches. Figure 7 presents a cross-plot analysis illustrating the deviations of the cementation exponent values predicted by the various methods from the experimentally measured data.

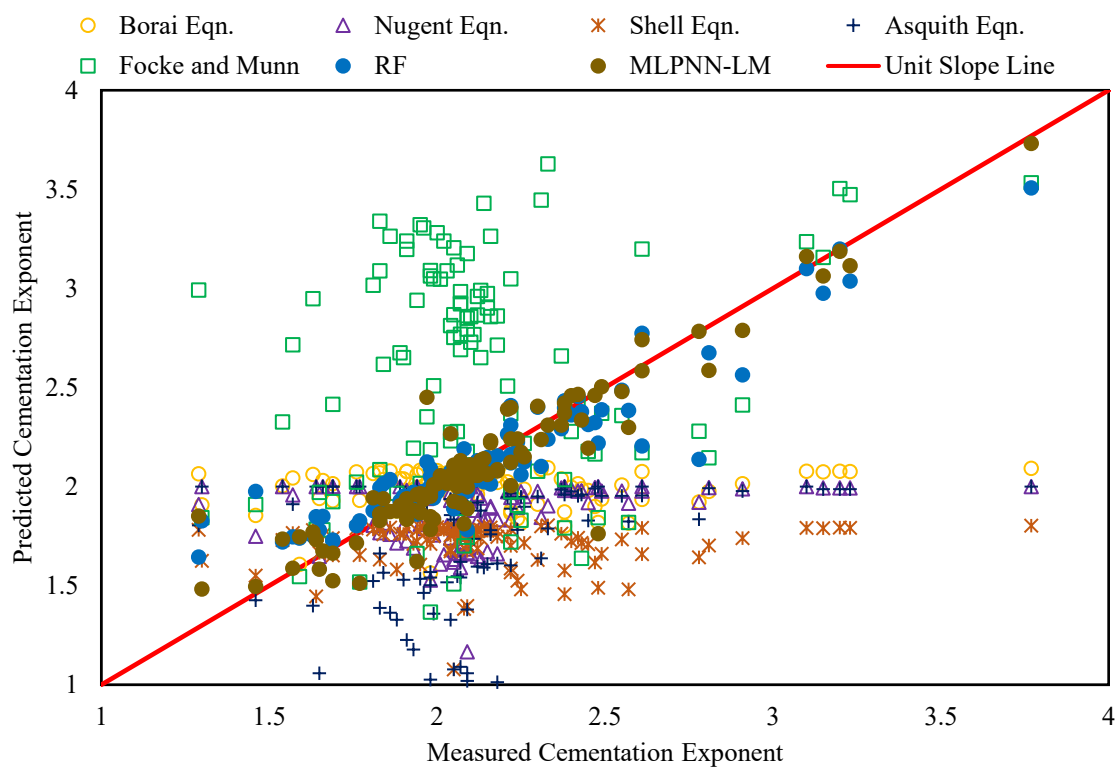
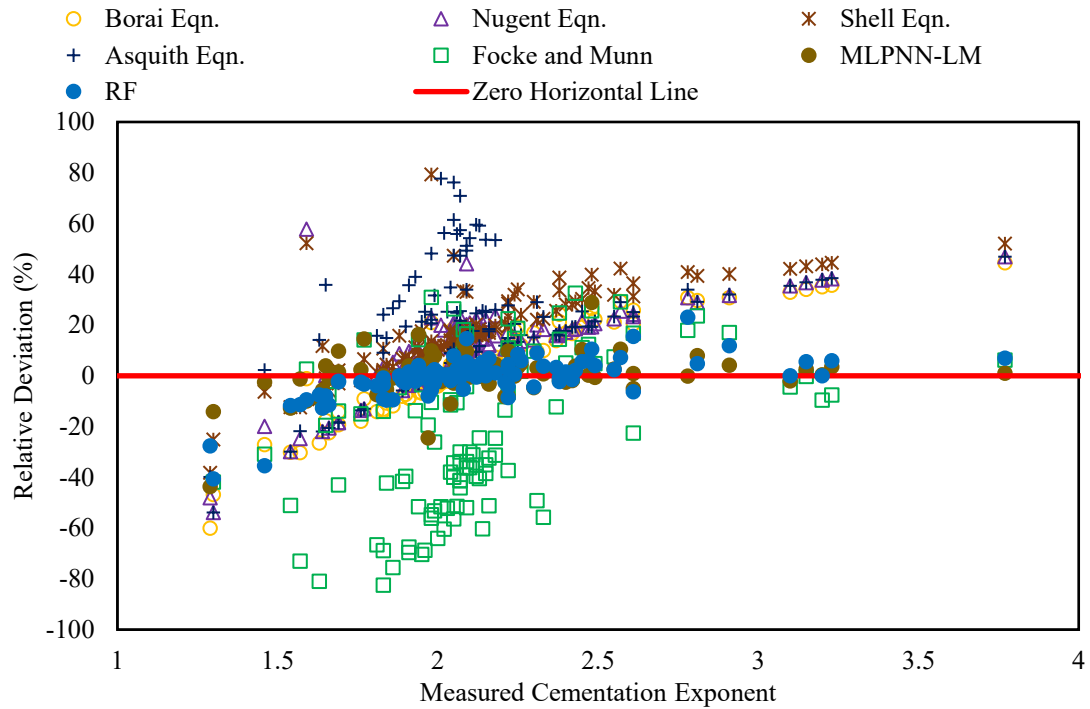


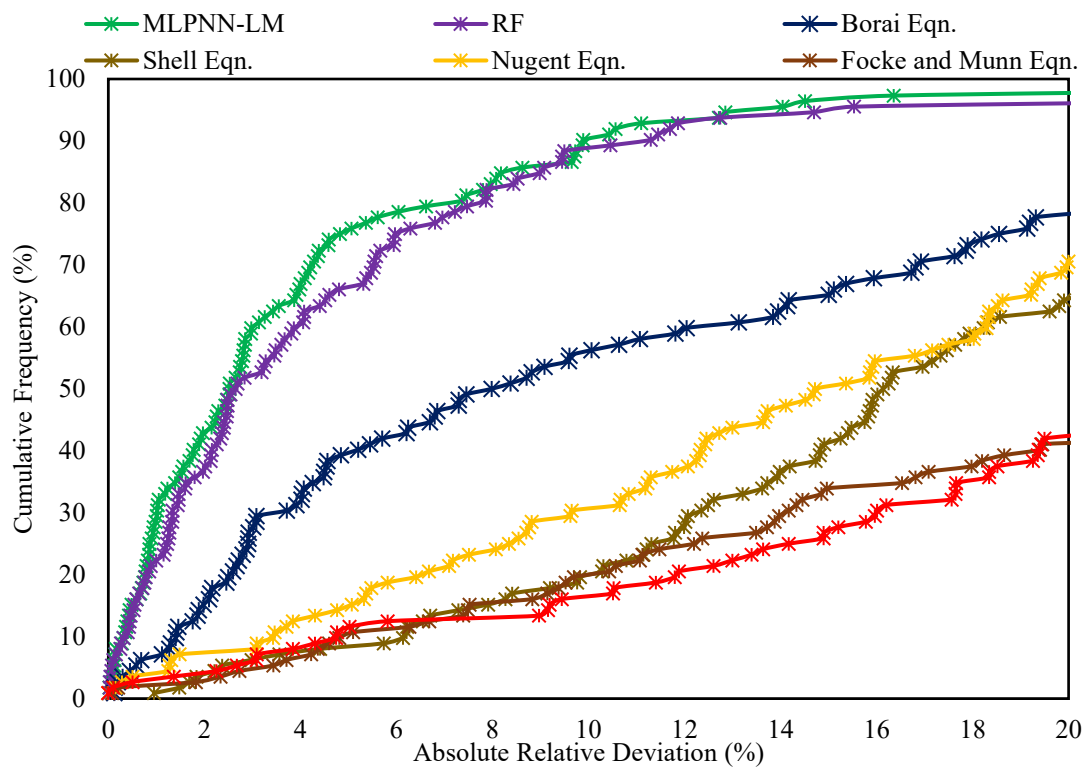
Figure 7

Cross-plot comparison of the predicted cementation exponent values from the proposed MLPNN-LM and RF models against existing correlations reported in the literature.

Notably, the data points corresponding to the MLPNN-LM and RF models cluster closely around the 45° line ($Y = X$), highlighting their superior performance compared to existing literature correlations. In contrast, the literature models exhibit systematic deviations, indicating inconsistencies in their cementation exponent predictions. Figure 8 illustrates the distribution of relative errors for each model with respect to the measured data. The strong performance of the proposed MLPNN-LM and RF models is evident, as their data points closely follow the zero-error line, confirming their accuracy and reliability.

**Figure 8**

Distribution of relative errors for the proposed MLPNN-LM and RF models compared to existing correlations reported in the literature.

**Figure 9**

Cumulative frequency of absolute percentage relative deviations for the proposed MLPNN-LM and RF models compared to existing published correlations.

The cumulative frequency diagram presented in Figure 9 provides an additional graphical approach for evaluating the performance of the proposed models relative to commonly used literature correlations across the entire dataset. Considering an absolute percentage relative deviation of 20%, the proportions of predictions within this threshold are as follows: MLPNN-LM: 98.2%; RF: 96.4%; Borai equation (Borai, 1987): 78.6%; Nugent equation (Ragland, 2002): 70.5%; Shell equation (Neustaedter, 1968; Watfa and Nurmi, 1987): 64.3%; Asquith equation (Asquith, 1997): 42.9%; and Focke and Munn equation (Focke and Munn, 1987): 41.9%. These results indicate that the proposed methods outperform existing literature correlations, as higher cumulative frequencies reflect greater robustness for a given absolute percentage relative deviation.

Among the models analyzed, MLPNN-LM demonstrates the highest predictive accuracy, particularly for tight carbonate samples with total porosity below 10% in deep formations. Figure 10 illustrates the relationship between cementation exponent and total porosity, showing that the MLPNN-LM model closely aligns with the experimentally measured values, confirming its reliability in such reservoir conditions.

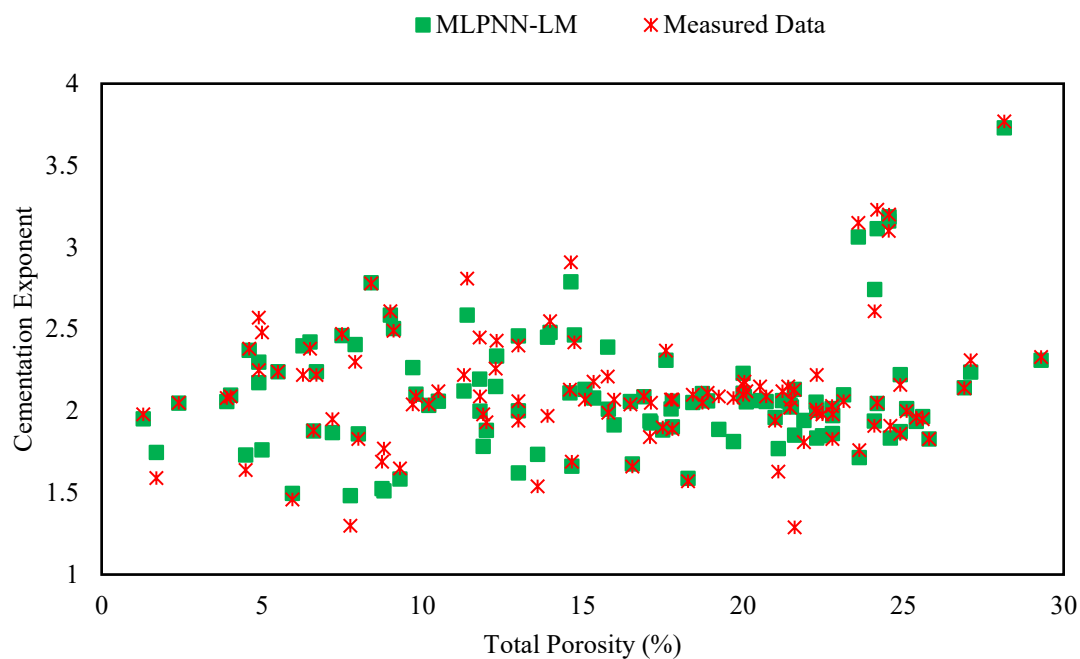


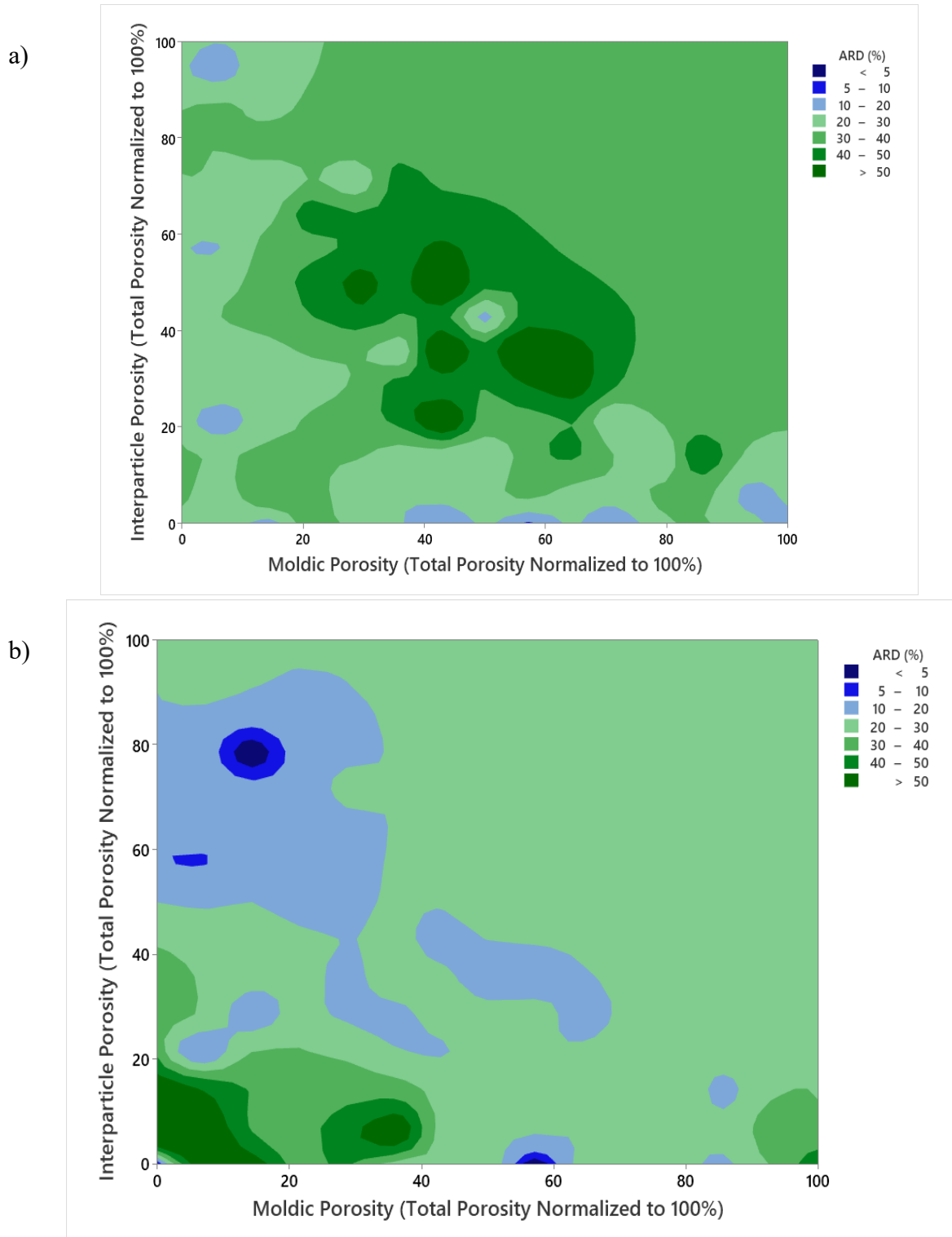
Figure 10

Comparison of measured cementation exponent values with predictions from the MLPNN-LM model, highlighting performance in deep formations and tight carbonate reservoirs.

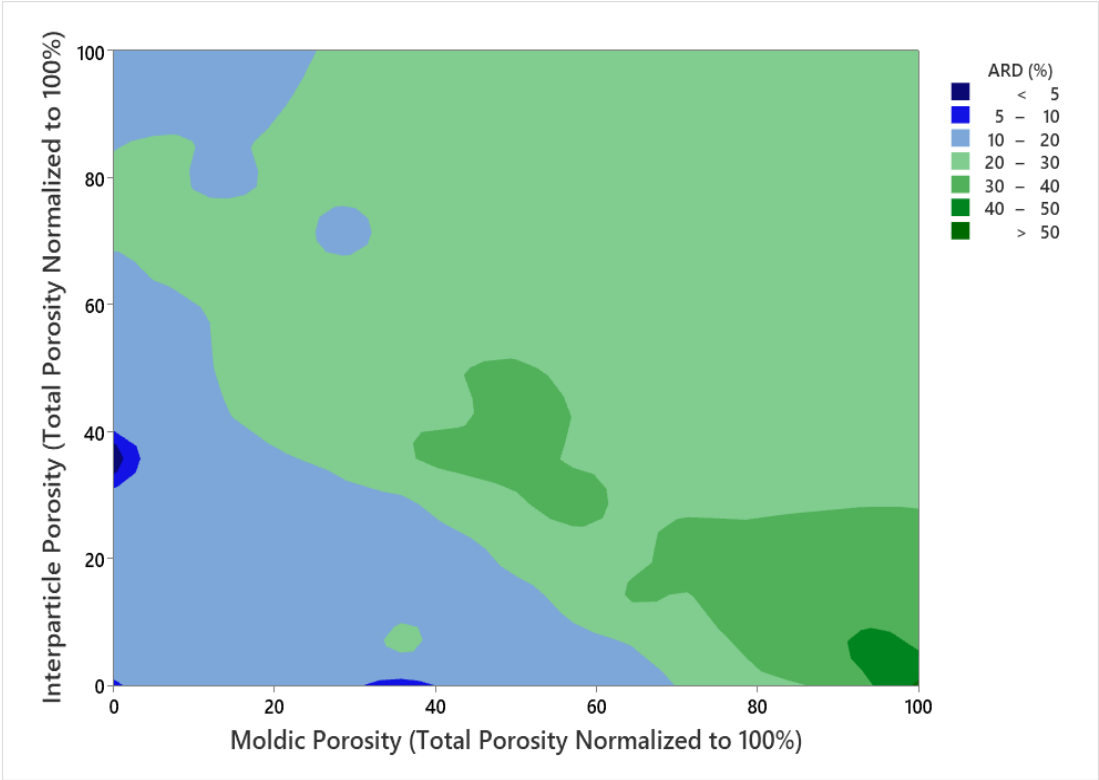
Evidently, the proposed MLPNN-LM model performs exceptionally well in estimating the cementation exponent within deep formations and tight carbonate reservoirs, underscoring a key advantage over conventional literature correlations. Existing empirical models often struggle to predict the cementation exponent accurately in such conditions, a limitation effectively addressed by the MLPNN-LM approach.

To further assess model performance, the variation of absolute relative deviation (ARD) percentages was evaluated across a broad range of connected vuggy and moldic porosities for all considered methods. The results are illustrated using 2D heatmap diagrams in Figure 11. In these heatmaps, dark green represents ARD levels greater than 50%, while dark blue corresponds to ARD values below 5%.

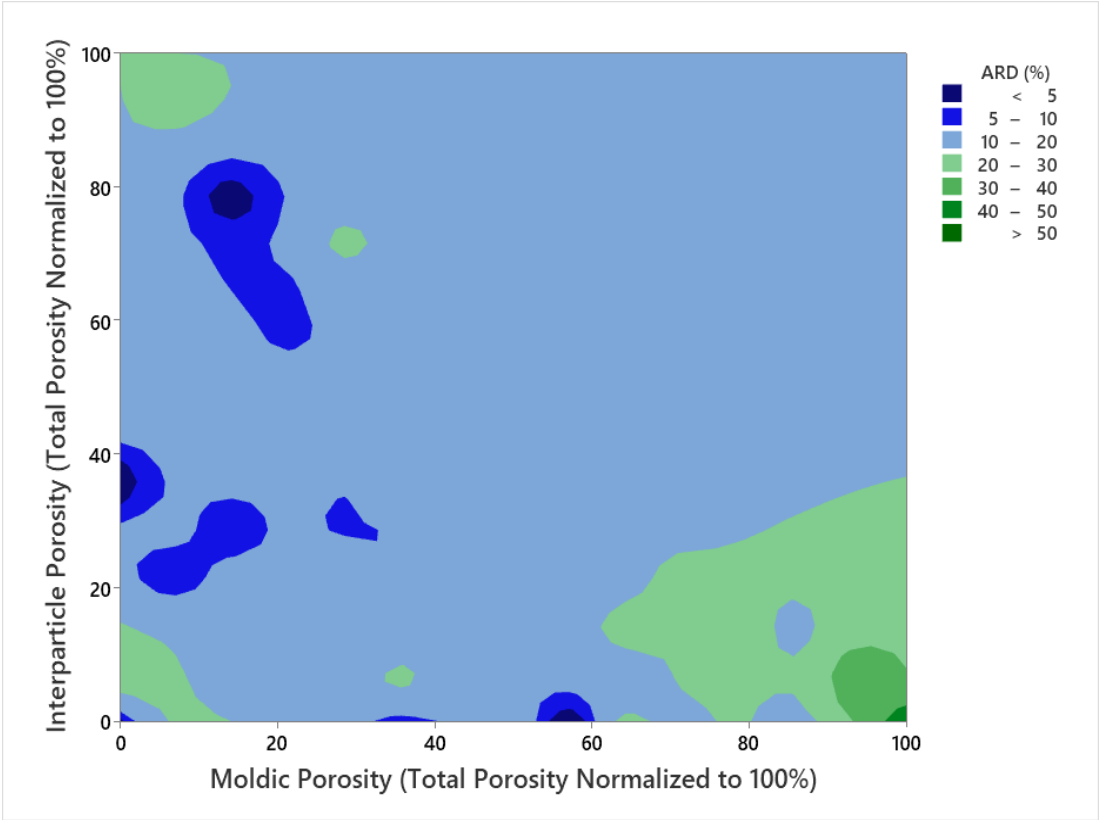
Intermediate colors indicate ARD values between 5% and 50%. Notably, the MLPNN-LM heatmap (Fig. 11g) exhibits predominantly dark and light blue shading, reflecting consistently low errors, whereas Focke and Munn's correlation (Focke and Munn, 1987) shows mainly light and dark green areas, indicating higher errors and poorer performance.



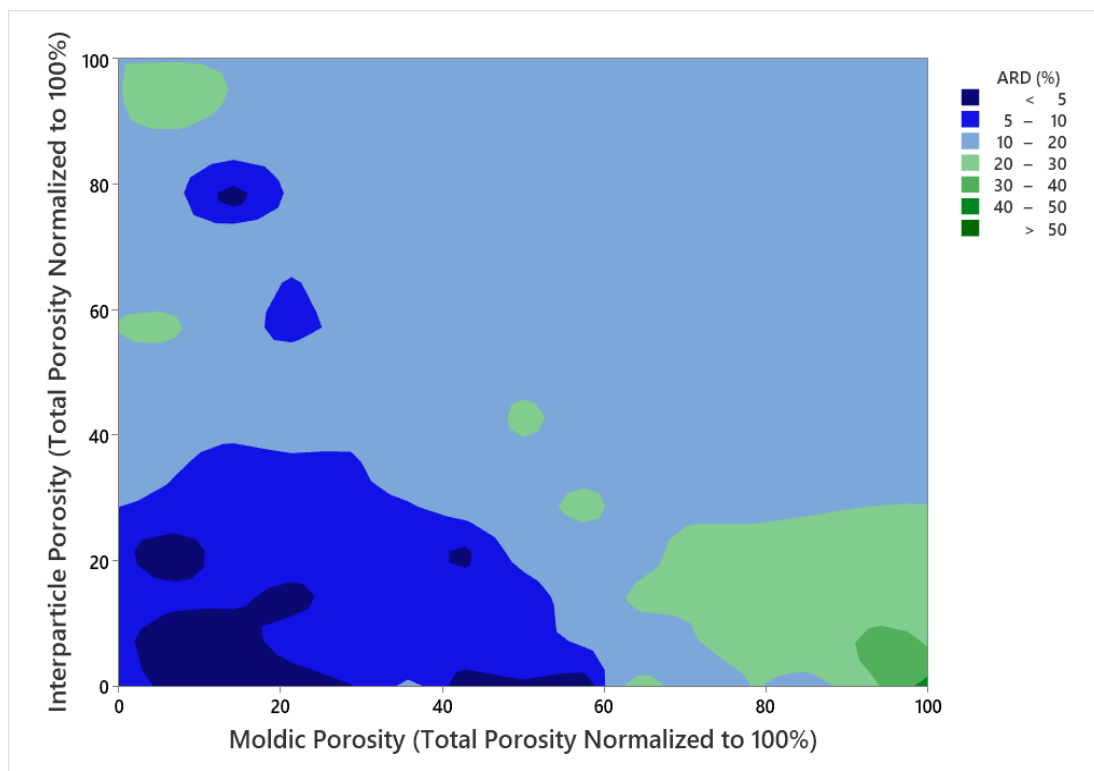
c)



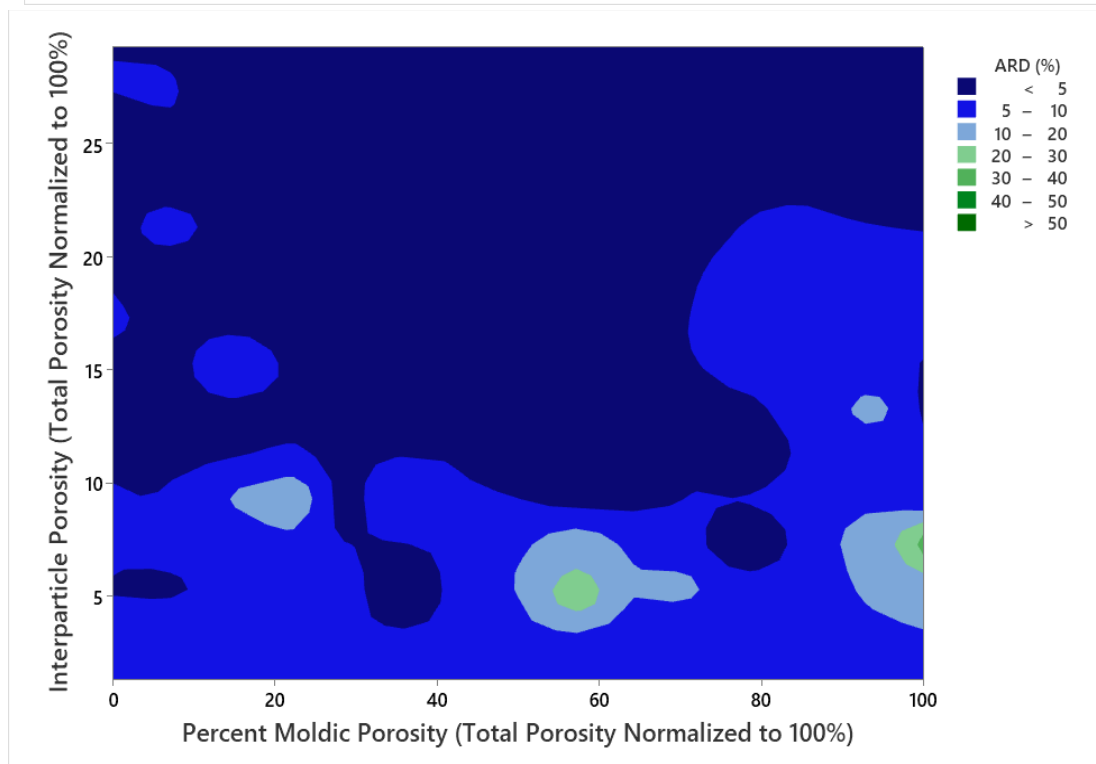
d)



e)



f)



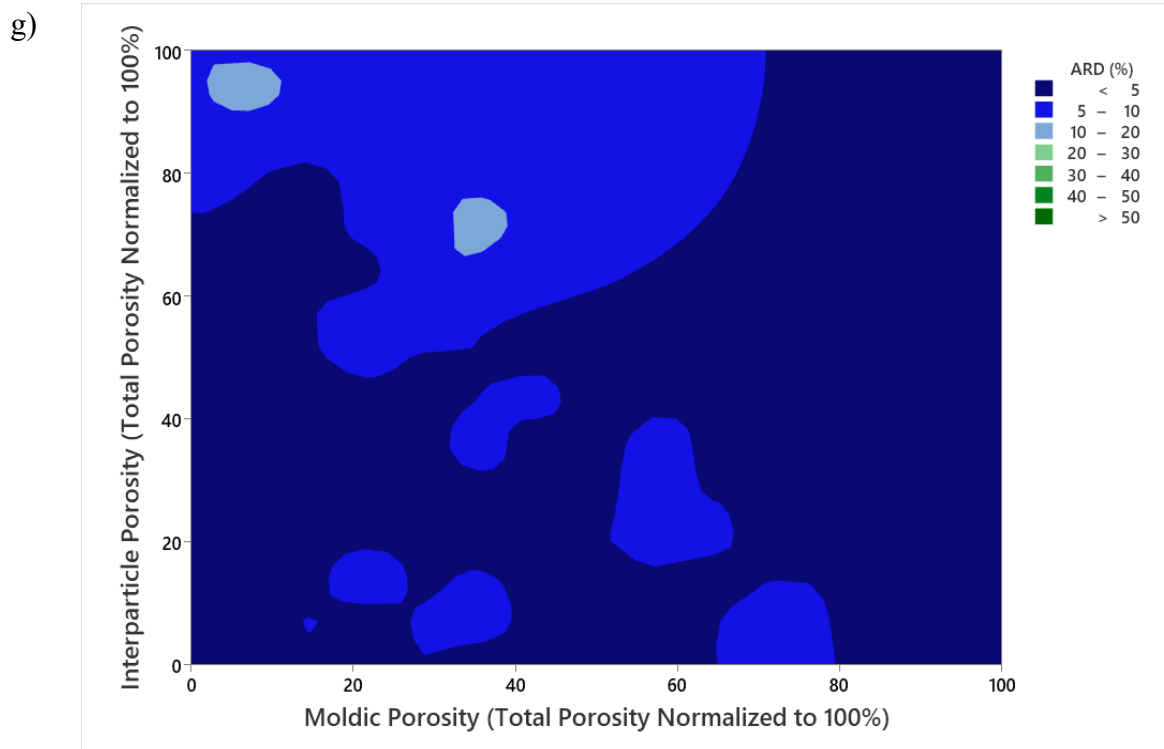


Figure 11

Contour plots of absolute relative deviation (ARD) across various ranges of moldic and interparticle pore types for the different models: (a) Focke and Munn equation (Focke and Munn, 1987), (b) Asquith equation (Asquith, 1997), (c) Shell equation (Neustaedter, 1968; Watfa and Nurmi, 1987), (d) Nugent equation (Ragland, 2002), (e) Borai equation (Borai, 1987), (f) RF, and (g) MLPNN-LM.

The models developed in this study leverage a comprehensive databank to generate globally applicable cementation exponent predictions. Consequently, the proposed MLPNN-LM and RF models provide reliable estimates, particularly for tight and deep carbonate formations, even when extrapolated beyond the original dataset. These models are suitable for both industrial and academic applications, including reservoir simulation and characterization. By incorporating pore description analysis, prediction accuracy is significantly improved compared to existing correlations that rely solely on total porosity.

Furthermore, the proposed methods can be seamlessly integrated into commercial software and modeling modules, offering valuable tools for researchers and practitioners working with porous media. By applying full-set petrophysical logs, including image logs, reservoir descriptions can be quantitatively established at each depth, determining the contributions of different pore types. Using the approaches developed here, MLPNN and RF modules can be implemented in well-known petrophysical software such as Geolog or Techlog. By integrating these modules with established pore descriptions from key wells, a global relationship for estimating the cementation exponent across the reservoir can be constructed.

5. Conclusions

In this study, we aim to characterize the cementation exponent in carbonate and heterogeneous formations using several advanced heuristic and self-organizing methods, including support vector regression, multilayer perceptron neural network optimized with Levenberg-Marquardt, probabilistic neural network, random forest, decision tree, and eXtreme gradient boosting. To accomplish this, a comprehensive dataset was compiled from the literature, comprising the cementation exponent as the

output variable and pore-type descriptors—moldic, interparticle, fracture, intracrystalline, intercrystalline, fenestral, connected vugs, isolated vugs—and total porosity as input variables. The dataset was then randomly partitioned into training, validation, and test subsets. Using these non-symbolic approaches, the MLPNN-LM, RF, SVR, PNN, XGBoost, and DT models were developed, and their performance was rigorously evaluated through both statistical metrics and visual assessment techniques.

1. The hybrid MLPNN-LM approach, followed by the RF model, exhibits the lowest errors in estimating the cementation exponent. Specifically, MLPNN-LM achieves $R^2 = 0.8654$ and AAPRD% = 4.3781%, while RF attains $R^2 = 0.8731$ and AAPRD% = 4.8690%, across both carbonate and heterogeneous pore systems.
2. The DT and XGBoost models are unreliable due to significant discrepancies among the training, validation, and test subsets, resulting in overfitting. Additionally, the PNN and SVR approaches are unsuitable for predicting the cementation exponent because of their high estimation errors.
3. The proposed MLPNN-LM and RF methods demonstrate substantially higher accuracy compared to commonly used literature correlations and existing machine learning models. For context, AAPRD% values for current machine learning algorithms and literature correlations predicting the cementation exponent range between 6.28%–7.45% and 12.30%–31.32%, respectively. This marked improvement highlights the superiority of the MLPNN-LM and RF approaches. Moreover, their performance remains robust in tight and deep carbonate pore systems, further confirming their advantage over conventional methods and existing machine learning tools.
4. Sensitivity analysis indicates that moldic porosity is the most influential input parameter for predicting the cementation exponent. Contrary to conventional literature correlations that often emphasize total porosity, our results reveal that detailed pore descriptions—particularly moldic, interparticle, and connected vuggy porosities—play the dominant roles in controlling the variability of the cementation exponent.
5. Outlier analysis using Williams' plot demonstrates that approximately 93.75% of the MLPNN-LM estimates and about 91.96% of the RF estimates fall within the valid applicability domain, confirming the robustness and reliability of both models.
6. Trend analysis indicates that the MLPNN-LM model exhibits a high degree of agreement with measured data across a broad range of total porosities. Notably, for total porosities below 10%, MLPNN-LM accurately captures the complex variations of the cementation exponent, closely matching the observed values. This highlights the suitability of the proposed approach for deep formations and tight carbonate reservoirs.
7. In summary, the robust methodologies developed in this study are highly valuable for simulating porous media flow and conducting petrophysical characterizations in petroleum engineering. They are particularly effective for deep formations (depths greater than 1200 m) and tight reservoirs (total porosity below 10%), providing reliable and accurate estimations of the cementation exponent in complex carbonate systems.

Conflict of interest

The authors declare that they have no relevant financial or non-financial competing interests to disclose.

Nomenclature

AAPRD	Average absolute percentage relative deviation
APRD	Average percentage relative deviation
BW	Interparticle pores
DT	Decision tree
FN	Fenestral pores
FR	Fracture pores
FRF	Formation resistivity factor
IX	Intercrystalline pores
LM	Levenberg–Marquardt
MAD	Mean absolute deviation
ML	Machine learning
MLPNN	Multilayer perceptron neural network
MLPNN-LM	Multilayer perceptron neural network with Levenberg–Marquardt optimization
MO	Moldic pores
PHIT	Total porosity
PNN	Probabilistic neural network
R^2	Coefficient of determination
RBF	Radial basis function
ReLU	Rectified linear unit
RF	Random forest
RMSE	Root mean square error
RQI	Reservoir quality index
SD	Standard deviation
SVM	Support vector machine
SVR	Support vector regression
WI	Intracrystalline pores
XGBoost	eXtreme gradient boosting

References

- Asquith, G., The importance of determining pore type from petrophysical logs in the evaluation of a Permian Wolfcamp reentry northern Midland Basin, *The Log Analyst*, Vol. 38, No. 03, p. 37-46, 1997.
- Awad, M., and Khanna, R., *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, 268 p. Springer Nature, 2015.
- Biau, G., Analysis of a random forests model, *The Journal of Machine Learning Research*, Vol. 13, No. 1, p. 1063–1095, 2012.
- Blockeel, H., Devos, L., Frénay, B., Nanfack, G., and Nijssen, S., Decision trees: from efficient prediction to responsible AI, *Frontiers in Artificial Intelligence*, Vol. 6, Article ID 1215873, 2023.

- Borai, A., A new correlation for the cementation factor in low-porosity carbonates, *SPE Formation Evaluation*, Vol. 2, No. 04, p. 495–499, 1987.
- Breiman, L., Random forests, *Machine Learning*, Vol. 45, No. 1, p. 5–32, 2001.
- Brownlee, J., A gentle introduction to XGBoost loss functions, *Machine Learning Mastery*, 2021. Available: <https://machinelearningmastery.com/xgboost-loss-functions>.
- Chang, D. T., Probabilistic deep learning with probabilistic neural networks and deep probabilistic models, arXiv preprint, arXiv:2106.00120, 2021.
- Chang, P., Multi-layer perceptron neural network for improving detection performance of malicious phishing URLs without affecting other attack types classification, arXiv preprint, arXiv:2203.00774, 2022.
- Chen, H. C., and Fang, J. H., Sensitivity analysis of the parameters in Archie's water saturation equation, *The Log Analyst*, Vol. 27, No. 05, 1986.
- Chen, T., and Guestrin, C., XGBoost: a scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, San Francisco, California, USA, p. 785–794, 2016.
- Chen, T., Story and lessons behind the evolution of XG-Boost, 7 p. 2016.
- Chilingarian, G. V., Chang, J., and Bagrintseva, K. J., Empirical expression of permeability in terms of porosity, specific surface area, and residual water saturation of carbonate rocks, *Journal of Petroleum Science and Engineering*, Vol. 4, No. 4, p. 317–322, 1990.
- Chok, N. S., ms, Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data, Ph.D. thesis, University of Pittsburgh, Pittsburgh, 2010.
- Costa, V. G., and Pedreira, C. E., Recent advances in decision trees: an updated survey, *Artificial Intelligence Review*, Vol. 56, No. 5, p. 4765–4800, 2023.
- Fischer, A., Izmailov, A. F., and Solodov, M. V., The Levenberg–Marquardt method: an overview of modern convergence theories and more, *Computational Optimization and Applications*, Vol. 89, No. 1, p. 33–67, 2024.
- Focke, J., and Munn, D., Cementation exponents in Middle Eastern carbonate reservoirs, *SPE Formation Evaluation*, Vol. 2, No. 02, p. 155–167, 1987.
- Goodall, C. R., 13 Computation using the QR decomposition, *Handbook of Statistics*, Elsevier, Vol. 9, p. 467–508, 1993.
- Gramatica, P., Principles of QSAR models validation: internal and external, *QSAR & Combinatorial Science*, Vol. 26, No. 5, p. 694–701, 2007.
- Hemmati-Sarapardeh, A., Ameli, F., Dabir, B., Ahmadi, M., and Mohammadi, A. H., On the evaluation of asphaltene precipitation titration data: modeling and data assessment, *Fluid Phase Equilibria*, Vol. 415, p. 88–100, 2016.
- Herrick, D. C., and Kennedy, W. D., Electrical efficiency; a pore geometric theory for interpreting the electrical properties of reservoir rocks, *Geophysics*, Vol. 59, No. 6, p. 918–927, 1994.
- Herrick, D., and Kennedy, W. D., Formation resistivity factor and permeability relationships in rocks characterized by secondary solution porosity, *SPWLA 36th Annual Logging Symposium*, OnePetro, 1995.

- Huang, X., Cao, H., and Jia, B., Optimization of Levenberg Marquardt algorithm applied to nonlinear systems, *Processes*, Vol. 11, No. 6, p. 1794, 2023.
- Iranzad, R., and Liu, X., A review of random forest-based feature selection methods for data science education and applications, *International Journal of Data Science and Analytics*, Vol. 20, p. 1–15, 2024.
- Kolah-Kaj, P., Kord, S., and Soleymanzadeh, A., Application of electrical rock typing for quantification of pore network geometry and cementation factor assessment, *Journal of Petroleum Science and Engineering*, Vol. 208, Article ID 109426, 2022.
- Li, F., Zhang, X., Lu, A., Xu, L., Ren, D., and You, T., Estimation of metal elements content in soil using x-ray fluorescence based on multilayer perceptron, *Environmental Monitoring and Assessment*, Vol. 194, No. 2, p. 95, 2022.
- Loh, W. Y., Classification and regression trees, *WIREs Data Mining and Knowledge Discovery*, Vol. 1, No. 1, p. 14–23, 2011.
- Lv, X., and Yu, Z., A modified Levenberg–Marquardt algorithm for low order-value optimization problem, *Journal of Applied Mathematics and Computing*, Vol. 70, No. 6, p. 5629–5644, 2024.
- Mahdaviara, M., Rostami, A., and Shahbazi, K., Smart learning strategy for predicting viscoelastic surfactant (VES) viscosity in oil well matrix acidizing process using a rigorous mathematical approach, *SN Applied Sciences*, Vol. 3, No. 10, p. 815, 2021.
- Mahdavi-Ara, M., Rostami, A., Shahbazi, K., Shokrollahi, A., and Ghazanfari, M. H., Estimating aqueous nanofluids viscosity via GEP modeling: correlation development and data assessment, *Iranian Journal of Chemistry and Chemical Engineering*, Vol. 41, No. 1, p. 266–283, 2022.
- Najafi-Silab, R., Soleymanzadeh, A., Kolah-Kaj, P., and Kord, S., Electrical rock typing using Gaussian mixture model to determine cementation factor, *Journal of Petroleum Exploration and Production Technology*, Vol. 13, No. 5, p. 1329–1344, 2023.
- Neustaedter, R. H., Log evaluation of deep Ellenburger gas zones, *SPE Deep Drilling and Development Symposium*, Monahans, Texas, SPE-2071-MS, 1968.
- Pourkamali-Anaraki, F., Husseini, J. F., and Stapleton, S. E., Probabilistic neural networks (PNNs) for modeling aleatoric uncertainty in scientific machine learning, *IEEE Access*, Vol. 12, p. 178816–178831, 2024.
- Ragland, D. A., Trends in cementation exponents (m) for carbonate pore systems, *Petrophysics*, Vol. 43, No. 05, 2002.
- Rostami, A., Helalizadeh, A., Moghaddam, M. B., and Soleymanzadeh, A., New insights into estimating the cementation exponent of the tight and deep carbonate pore systems via rigorous numerical strategies, *Journal of Petroleum Exploration and Production Technology*, Vol. 14, No. 6, p. 1605–1629, 2024.
- Rostami, A., Kordavani, A., Parchekhari, S., Hemmati-Sarapardeh, A., and Helalizadeh, A., New insights into permeability determination by coupling Stoneley wave propagation and conventional petrophysical logs in carbonate oil reservoirs, *Scientific Reports*, Vol. 12, Article ID 11618, 2022.
- Rousseeuw, P. J., and Leroy, A. M., *Robust regression and outlier detection*, John Wiley & Sons, 2003.

- Sayahi, T., Tatar, A., Rostami, A., Anbaz, M. A., and Shahbazi, K., Determining solubility of CO₂ in aqueous brine systems via hybrid smart strategies, *International Journal of Computer Applications in Technology*, Vol. 65, No. 1, p. 1–13, 2021.
- Smola, A. J., and Schölkopf, B., A tutorial on support vector regression, *Statistics and Computing*, Vol. 14, No. 3, p. 199–222, 2004.
- Soleymanzadeh, A., Helalizadeh, A., Jamialahmadi, M., and Soulgani, B. S., Development of a new model for prediction of cementation factor in tight gas sandstones based on electrical rock typing, *Journal of Natural Gas Science and Engineering*, Vol. 94, Article ID 104128, 2021.
- Soleymanzadeh, A., Helalizadeh, A., Jamialahmadi, M., and Soulgani, B. S., Investigation of analogy between thermal and electrical properties of some reservoir rocks, *Bulletin of Engineering Geology and the Environment*, Vol. 80, No. 1, p. 507–517, 2021.
- Soleymanzadeh, A., Jamialahmadi, M., Helalizadeh, A., and Soulgani, B. S., A new technique for electrical rock typing and estimation of cementation factor in carbonate rocks, *Journal of Petroleum Science and Engineering*, Vol. 166, p. 381–388, 2018.
- Soleymanzadeh, A., Kolah kaj, P., Kord, S., and Monjezi, M., A new technique for determining water saturation based on conventional logs using dynamic electrical rock typing, *Journal of Petroleum Science and Engineering*, Vol. 196, Article ID 107803, 2021.
- Stebakov, I., Kornaev, A., Popov, S., and Savin, L., Fault diagnosis systems for rotating machines operating with fluid-film bearings, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, Vol. 236, No. 10, p. 1917–1931, 2022.
- Watfa, M., and Nurmi, R., Calculation of saturation, secondary porosity and producibility in complex Middle East carbonate reservoirs, *SPWLA Annual Logging Symposium*, Paper No. SPWLA-1987, 1987.
- Yan, C., Shen, X., and Guo, F., An improved support vector regression using least squares method, *Structural and Multidisciplinary Optimization*, Vol. 57, No. 6, p. 2431–2445, 2018.

**COPYRIGHTS**

©2023 by the authors. Licensee Iranian Journal of Oil & Gas Science and Technology. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>)