DOI: 10.30632/SPWLA-2022-0032

MACHINE LEARNING ASSISTED PREDICTION OF PERMEABILITY OF TIGHT SANDSTONES FROM MERCURY INJECTION CAPILLARY PRESSURE TESTS

Jassem Abbasi, University of Stavanger | Jiuyu Zhao, China University of Petroleum | Sameer Ahmed, University of Stavanger | Jianchao Cai, China University of Petroleum | Pål Østebø Andersen, University of Stavanger | Liang Jiao, China University of Geosciences

This paper was prepared for presentation at the SPWLA 63rd Annual Logging Symposium held in Stavanger, Norway, June 10-15, 2022.

ABSTRACT

Capillary pressure is an important parameter in both petrophysical and geological studies. It is a function of different porous media properties, in special, the pore structure of the rock. Mercury Injection Capillary Pressure (MICP) analysis is a consistent methodology for determining different petrophysical properties including porosity, and pore throat distribution. The matrix permeability is dependent on the pore size distribution but is not directly measured from MICP tests. In this work, we consider distinct parameters derived from MICP tests for the prediction of permeability by following a machine learning based approach.

Firstly, a vast range of MICP test results (246 samples) related to tight sandstones is gathered with a permeability range of 0.001 to 70 millidarcy. After quality checking of dataset, different theoretical permeability models are tested on the dataset and the results are analyzed. Also, different features related to the pore throat characteristics of rock is analyzed and the best characteristics are selected for the input variables to the machine learning model. The Support Vector Regression (SVR) approach is proposed with the Radial Basis Function (RBF) kernel for the prediction of rock permeability from MICP tests. A Particle Swarm Optimization (PSO) is applied for optimization of the model meta parameters in the validation process to avoid over or underfitting. The training of the model is carried out with random selection of 80% of samples while other points are applied for testing of the model.

The data analysis on the correlation between rock permeability and parameters of capillary pressure is studied and showed that using pore throat radius corresponding to saturation range of 0.4-0.8 and the median capillary pressure values obtained from the capillary pressure curves is suitable to be used as input features of the SVM model. Also, the porosity and Winland equation was considered as input features due to their acceptable correlation with the rock permeability. The results showed that the implemented SVM-PSO model can acceptably predict the experimentally measured permeability values with R^2 rate of over 0.88 for training and testing datasets.

This work represents an analysis of the relationship of capillary pressure curve specifications with permeability on a large MICP dataset, especially focused on tight sandstone rocks. The analysis provided new statistical and physics-based features with the highest correlations with the rock permeability that helped in significant improvement of the SVM-PSO prediction results.

INTRODUCTION

Multiphase flow in porous media happens in many man-made and natural processes such as during nonaqueous phase liquids (NAPLs) transport, CO_2 storage, enhanced oil recovery, etc. (Ahmad et al., 2016; Blunt, 2017). The relative distribution and movement of wetting and non-wetting phases in porous media are highly influenced by the capillary behavior of phases and controlled by two critical parameters. In classic definitions, these are the capillary pressure and the relative permeability of phases, both considered to be functions of the saturation of the wetting phase (Lin et al., 2018). Also, the efficiency of operations in oil recovery or geological storage is highly dependent on the permeability of the rock and the correct prediction of its distribution in areal and lateral directions affects decisions and technical solutions. The rock permeability is directly dependent on the geometrical attributes of rocks such as porosity, pore-size distribution, and pore network coordination number (Menke et al., 2021).

Capillary pressure is defined as the pressure difference across the interface of two immiscible phases. The wetting condition of porous media, interfacial properties of phases and the pore geometry of rock are the determining parameters in the capillary behavior of porous media. The porous plate method, centrifuge method, and mercury injection capillary pressure (MICP) method are the three most routine approaches of capillary pressure function measurement in geological

rocks (Abbasi and Andersen, 2021). In MICP tests, the mercury as a non-wetting phase is injected into the core samples up to high pressures (to 4000 bars). The porosity is calculated from the total volume of mercury injected at the maximum pressure (McPhee et al., 2015). Also, the saturation at each pressure stage is determined as the volume fraction of mercury entered into the rock. The capillary pressure at each stage is equal to the mercury injection pressure. However, it leads to permanent loss of the core samples due to retaining mercury in the pores after withdrawal. This test also gives insightful information related to the pore structure of the rock (Jiao et al., 2020). The measured capillary pressure at each saturation stage is related to effective pore throat size by the Young-Laplace equation:

$$P_c = \frac{2\sigma cos\theta}{r} \quad (1)$$

Where σ is mercury/air interfacial tension (IFT), θ is contact angle and r is the pore throat radius.

The permeability of cores is routinely measured in single-phase liquid or gas injection tests. Also, due to the high measurement time of tight samples, the pressure decay method may be used (Jones, 1997). However, for long decades, researchers are tried to provide methodologies to extract the permeability from other measurements such as capillary pressure tests, especially MICP tests. Purcell (1949) provided an analytical-based relation for the prediction of permeability from porosity and MICP curve properties such as the fraction of volume occupied by mercury and capillary pressure. Swanson (Swanson, 1981) defined the Swanson parameter that is the maximum value of the curve when the (S_{Hg}/P_c) is plotted versus the S_{Hg}, for the prediction of permeability from the capillary pressure data for clean sandstones and carbonates, separately. In MICP tests, S_b is the mercury saturation. However, Xiao et al. (2014) concluded that the Swanson equation is not successful in tight sandstones due to the ambiguous Swanson parameter values. Xiao et al. (2017) showed that in homogeneous sandstone rocks, the average pore throat radius can be calculated to find accurate permeability predictions. By analysis of tight gas sands, Rezaee et al. (2012) found that the dominant pore throat radius is in the mercury saturation of 10%, and this point is capable to be correlated with permeability.

In recent years the application of machine learning in geosciences is growing rapidly due to the large volume of available data that needs processing and analysis (Karpatne et al., 2019). Several applications of machine learning have been found in the rock-fluid properties (Hébert et al., 2020), and upscaling of porosity-permeability calculations from pore-scale images (Menke et al., 2021). Due to the importance of permeability, many of the studies are focused on the prediction of permeability from different data sources including special core analysis (SCAL) tests (Erofeev et al., 2019), wireline logs (Zhang et al., 2021), and Nuclear Magnetic Resonance (NMR) tests (Zhang et al., 2017). Feng et al. (2020) applied a support vector machine algorithm on 22 sets of MICP data and found that it is superior to currently available permeability models. However, there are still open opportunities for the development of models with more generalization capabilities.

This work at first intended to investigate the validity of current permeability models (i.e., Swanson, Purcell, Parachor, Fractal and Winland models) over a large set of MICP tests related to tight sandstones which are gathered from all over the world. Then, it is tried to find the pore-scale characteristics of the capillary pressure curves that are more relevant the most relevant to the rock permeability. Then, by mixing the most correlated theoretical and statistical features, it is focused on the deployment of machine aided models for the estimation of tight sandstone permeabilities from MICP tests. In the following, the theories behind the deployed models are introduced and the details of the developed machine learning model are provided. Then, the results obtained in the process of the model deployment is provided. Finally, the work ended with a conclusion.

METHODOLOGY

The methodology used in this work is based on machine learning algorithms coupled with optimization tools. As it is shown in Fig. 1, the work started with the data-gathering stage from industrial and also literature data. Afterwards, the data cleaning and anomaly analysis are carried out. Then, by statistical analysis of possible relevant features, the input parameters best suitable for the prediction of rock permeabilities are investigated and features with the best results are selected. Finally, the machine learning model is trained, and the results are validated against the testing dataset. More discussions about the methodology are provided in the next sections.

DATA GATHERING

In this work, a dataset with overall 248 samples related to tight sandstone rocks was gathered. Each sample included the measurements of porosity, permeability, and MICP capillary pressure curves. A part of the data was gathered from the Ordos Basin, located at the junction of the eastern tectonic domain and western tectonic domain, which is the second-largest

petroliferous basin in China (Wang and Wang, 2013). The study area is in the Yishaan Slope of the Ordos Basin, whose area is also the main area of tight (sandstone) oil exploration and development. These tight cores are taken from different layers of more than 20 wells in Yanchang Formation, in a total of 172 cores were tested by Poremaster PM-33-13 for MICP data, and CMS-300 was used to obtain porosity and permeability of these cores (Fan et al., 2019). Also, 76 MICP samples are gathered from different sources in the literature (Arabjamaloei et al., 2019; Eslami et al., 2013; Feng et al., 2020; Jiang et al., 2018; Lai and Wang, 2015; K. Liu et al., 2020; Liu et al., 2018; X. Liu et al., 2020; Y. Liu et al., 2020; Rezaee et al., 2012; Tran et al., 2018; Wang et al., 2018, 2019; Xiao et al., 2017, 2014). Figure 2 shows the capillary pressure curves for all cases in both cartesian and logarithmic scales. Also, the statistical distribution of porosity and permeability values of the tests are provided in Fig. 2.

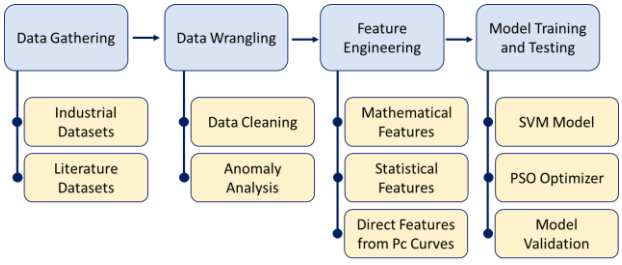

Fig. 1—The flowchart introducing the workflow of the permeability model development.

PERMEABILITY CORRELATIONS

Many researchers have developed models for the prediction of rock permeability by interpretation of MICP tests. These models may partially be based on theoretical hypotheses or based on empirical data sets. In this section, the most prominent correlations that were applied in this work are introduced.

SWANSON PERMEABILITY

To provide a method for prediction of rock brine permeability from capillary measurements, Swanson (Swanson, 1981) provided a correlation by introducing the Swanson parameter that is calculating the maximum point of the curve when the (S_{Hg}/P_c) is plotted versus the S_{Hg}. This point is closely related to the condition in which the non-wetting phase partially fills the effective pore volumes, has a determining role in controlling fluid flow in the rock pore system. For sandstone rocks, rock permeability (m^2) is calculated by:

$$K = 0.015\left(\frac{S_{Hg}}{P_c}\right)^{2.109}_{max} \qquad (2)$$

Where in this equation P_c is in Pa, and S_b is in fraction.

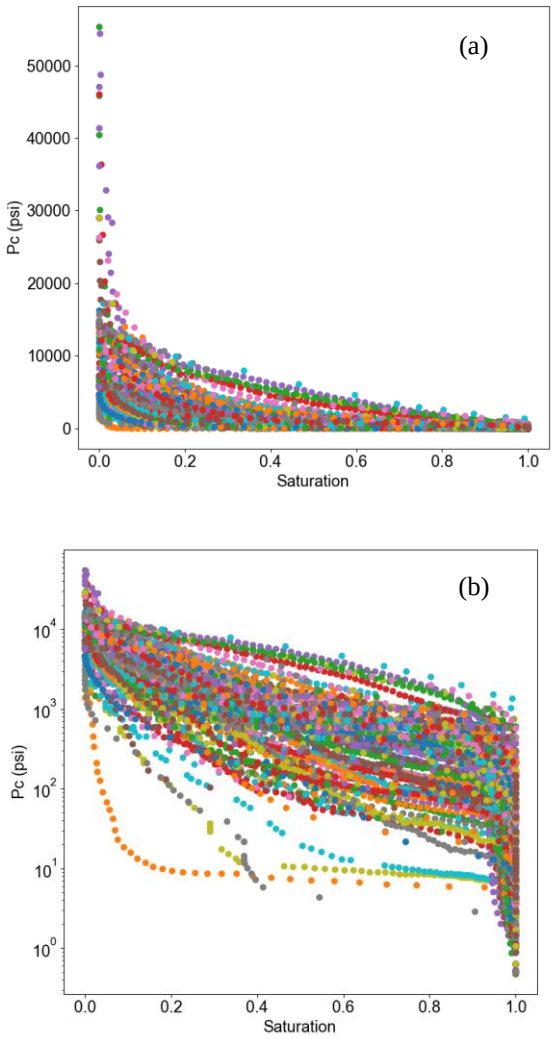

Fig. 1—The mercury injection capillary pressure curves for all the MICP samples in (a) cartesian scale, (b) Semi-log scale

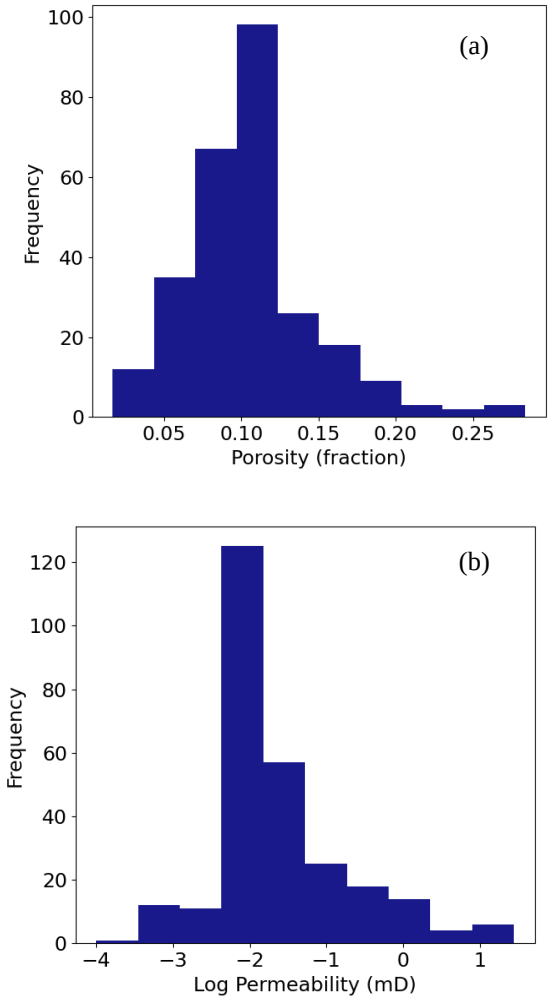

Fig. 2—Histogram distribution of (a) porosity and (b) permeability values

PURCELL EQUATION

Purcell by utilizing a simplified hypothesis or Poiseuille's equation, derived a semi-analytical equation to find the relationship of the permeability of a porous media to its porosity and capillary pressure curve (Purcell, 1949).

$$K = (6.47e-04)\phi \int_{S_{Hg}=0}^{S_{Hg}=100} \frac{dS_{Hg}}{P_c^2} \qquad (3)$$

Where P_c is in Pa, S_b is in fraction, and K is in m^2. The equation was verified over 26 MICP tests from sandstone rocks and showed that the permeability can be well predicted using the porosity and capillary pressure curves.

PARACHOR EQUATION

Guo et al. (2004), by extending the Swanson model, introduced a correlation that best fitted their MICP data by deriving capillary Parachor parameter $(S_b/P_c^2)_{max}$ that is defined as the maximum value of the (S_{Hg}/P_c^2) parameter versus the mercury saturation. So, then, the Parachor permeability model is defined as:

$$K = (5.29e-07)\left(\frac{S_{Hg}}{P_c^{\,2}}\right)_{max} \qquad (4)$$

here, P_c is in Pa, S_b is in fraction and K is in m^2.

WINLAND EQUATION

Winland presented an empirical correlation that relates the average pore radius and porosity to the rock air permeability (Kolodzie, 1980). After regression with different parameters, he found that the radius corresponding to the mercury saturation of 35% has the best correlation with the permeability of cores. However, the optimum corresponding saturation may be different in various cases (Wang et al., 2018). In this work, the reference correlation is used for the analysis of the permeability measurements:

$$\log r_{35} = 7.82 + 0.588 \log K_{air} - 0.864 \log \phi \qquad (5)$$

Where K is in m^2, ϕ is in fraction and r_{35} is in μm. In this equation, the pore radius corresponding to saturation of S_m is calculated by:

$$r_{Sm} = \frac{2\sigma cos\theta}{P_c(S_m)} \qquad (6)$$

Where σ is air-mercury interfacial tension (IFT) and θ is rock-air-mercury contact angle.

FRACTAL ANALYSIS OF PORE STRUCTURE

The fractal theory is widely used for the analysis of pore structures in sandstone rocks, firstly proposed by Burn and Mandelbrot (1984). They found that the size distribution of the pores in sponges follows the power law. Actually, the cumulative distribution of pores with size greater than or equal to λ has been confirmed to

follow:

$$N(L \geq \lambda) = \left(\frac{\lambda_{max}}{\lambda}\right)^{D_f} \quad (7)$$

Where D_f is defined as the pore-size fractal dimension and λ_{max} is the maximum pore size. The fractal dimension represents the fractal specifications of pores and especially, it gives insightful information about the complexity of the pore network of the rocks and can be calculated using different approaches such as scanning electron microscopy (SEM), thin section analysis, X-ray computed tomography scan and also MICP tests. There are several methods for calculation of fractal dimension of rocks from MICP data including 2D capillary tube models, 3D sphere models, thermodynamic models and 3D capillary tube models (Ge et al., 2016). Wang et al. (2018) showed that the 3D capillary tube model, is the most appropriate model for the prediction of rock properties and pore structures from MICP tests. In one of these models, Li, 2010 proposed a relation with the mercury saturation capillary pressure as:

$$S_{Hg} \propto P_c^{-(2-D_f)} \quad (8)$$

Considering this relationship, the fractal dimension of the 3D capillary tube model is calculated by plotting the log-log scale of S_{Hg} and P_c curve. Assuming the slope of $log(S_{Hg}) - log(P_c)$ curve is m, the fractal dimension is defined as (Li, 2010):

$$D_f = m + 2 \quad (9)$$

Afterwards, the specific surface area of rock is calculated by:

$$S_P = \frac{3}{r_g}\frac{1-\phi}{\phi}; \quad (10)$$

Where r_g is the average grain radius. The permeability (m^2) is then calculated by:

$$K = 1.6\left(\frac{1-\phi}{S_p}\right)\left(\frac{0.952\phi^2}{1-\phi}\right)^{2/(D_f-1)} \quad (11)$$

ANOMALY ANALYSIS

Outliers (anomalies) are defined as a group of the original samples in the dataset that show abnormal behavior in comparison to the majority of the population. These abnormalities can be due to erroneous measurements/calculations or even the presence of rare cases in the population. There are several methods for the detection of outliers in the literature. In this work, a Random Sample Consensus (RANSAC) method is used (Choi et al., 2009). This method follows an automatic non-deterministic algorithm and fits a model on random subsets of inliers from the complete data set. The advantage is that it can find the outliers in multidimensional data sets with a high level of accuracy even when a substantial fraction of anomalies is present in the dataset. The results of anomaly detection of the cores are shown in Fig. 3. In this process, the porosity-permeability data are applied for anomaly detection.

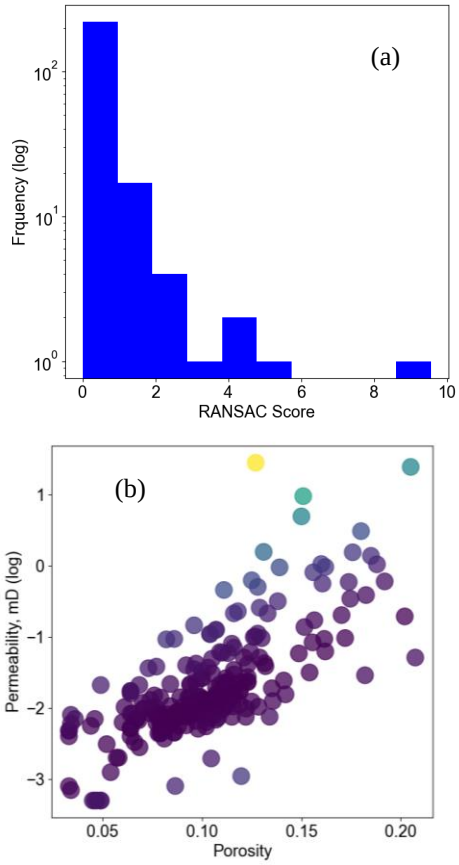

Fig. 3—The results of anomaly analysis applied to the dataset. (a) The histogram of RANSAC outlier analysis scores, (b) Anomaly analysis where the colors show the outlier score of each point. Points with the lowest outlier score have darker colors. While the points with high RANSAC scores are colored in red.

Figure 3a shows the histogram distribution of anomaly scores found by the RANSAC algorithm. The points with higher scores are more probable to be considered

anomalies. In Fig. 3b, the color of points is set based on the anomaly score provided by the RANSAC algorithm. It shows that there may be an abnormality in Porosity-Permeability scatter trend in cases with medium porosity but very high permeability. These cases can be due to the presence of non-reported fissures/fractures in the cores, or also can be due to measurement errors. However, it cannot be confidently considered as a measurement error.

Since this work is intended to hold the generality of the model, the suspected anomaly points are not removed from the training model. This will help to ensure that the developed model will be adequately general to be used in the future predictions.

KENDAL'S TAU CORRELATION COEFFICIENT

There are different models to quantify the dependency of two quantities like Pearson, Spearman and Kendal's Tau correlations. Different works showed the relative advantage of Kendal's Tau model in normally distributed datasets (Croux and Dehon, 2010; Puth et al., 2015). In this work, Kendal's Tau correlation coefficient is introduced that is defined as the similarity of the ordering of the data when separately ranked based on each parameter (Kendall, 1976). This parameter is explicitly defined as below for any pair of (x_i,x_j) and (y_i,y_j) :

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} sgn(x_i - x_j)\, sgn(y_i - y_j) \qquad (12)$$

Here *sgn* is the sign function. Two parameters are completely irrelevant if the τ converges to zero and are directly or inversely correlated if τ equals *1* or *-1*, respectively.

INPUT FEATURES

The main part of the study is designing proper input parameters for the training of the SVM-PSO model. The most routine feature for the prediction of permeability is rock porosity. Figure 3b shows the correlation between the porosity versus permeability values. There is a linear increasing trend in permeability (log) versus porosity scatter plot. Kendal's tau coefficient is 0.54 showing that the porosity generally is a determining factor for the permeability prediction. So, the porosity is used as a relevant data source.

Also, to integrate the machine learning model with the physics-based theories, the calculated permeability values with the different permeability models including Swanson, Purcell, Winland, Parachor, and Fractal models are considered as the possible features. The theories behind these equations are provided in previous sections. The more successful model will be selected as the input features.

Moreover, plus the previous models, two statistically calculated parameters are extracted from the capillary pressure curves:

- **Average P_c:** The average value of capillary pressure for each point in capillary pressure curves. This value is related to the average pore throat size values in rock.

- **Median P_c:** The capillary pressure value lied on the midpoint of the capillary pressure curve.

Also, in another part of the feature engineering process, it has been decided to directly import some parts of the capillary pressure curve as the pore-throat radius. In the calculations, capillary pressure points related to the different saturation ranges are extracted from the capillary pressure curves. The extracted capillary pressure values are then converted to the pore-throat radius using Eq.(6 and then their average values are used as features. More information about the validity of these values and their correlation with the rock permeability is provided in the next sections.

SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) algorithm was first suggested by Cortes and Vapnik (1995) and is a subset of supervised learning methods, and primarily introduced for classification (pattern recognition) purposes in projects like hand-writing and face recognition. However, it was also successful in regression problems. In its simplest form, SVM uses a linear fitting hyperplane to regress on the dataset with a minimal error margin. Given the training dataset $\{(x_i, y_i)\}_{i=1}^{n}$ (*n* is the number of training samples), where x_i is the matrix of input variables (with dimension 1 by *I*, where *I* is the number of inlet variables), and y_i is the outlet variable, that in this work is the logarithm of permeability ($logK_i$), $i=1,\ldots,N$, the hyperplane equation in two-dimensional (x,y) space is defined as a subspace of dimension *n - 1*:

$$f(x) = \vec{w} \cdot \vec{x} + b \qquad (13)$$

Here, scaler b is defined as the offset of regression line and the vector w (1 by I) is called weight vector and defines a direction perpendicular to the hyperplane. The prediction function above requires a small w. The regression parameters of the hyperplane function are calculated by minimizing the objective function, subjected to the constraints:

$$\text{Minimize:} \quad \frac{1}{2}\omega^T \cdot \omega$$

$$\text{Subject to:} \quad \begin{cases} \vec{w} \cdot \vec{x} + b - \vec{y} \le \varepsilon \\ \vec{y} - \vec{w} \cdot \vec{x} - b \le \varepsilon \end{cases} \qquad (14)$$

Where ε is the epsilon-distance within which no penalty is associated in the training loss function with points predicted within a distance epsilon from the actual value (see Fig. 4). In this equation, the error values less than ε should be ignored. To increase the generalization capability of the model, the relaxation variable (ζ) is introduced to include the errors associated to the points where the target error values exceed ε (see Fig. 4):

$$\text{Minimize: } \frac{1}{2}\omega^T \cdot \omega + C\sum_{i=1}^{n}\zeta_i$$

$$\text{Subject to:} \quad \begin{cases} \vec{w} \cdot \vec{x} + b - \vec{y} \le \varepsilon + \zeta \\ \vec{y} - \vec{w} \cdot \vec{x} - b \le \varepsilon - \zeta \end{cases} \qquad (15)$$

The constant $C > 0$ determines the trade-off between the flatness of the model (model complexity) and the amount to which prediction errors larger than ε are tolerated.

However, in complex classification or regression problems, SVM maps nonlinear regression problems from low dimensional feature spaces into linear regression problems with higher dimensional feature spaces. The mapping of parameters in space is achieved by using nonlinear transforming functions which are called kernel functions. There are a variety of kernel functions including Radial Basis Function (RBF), polynomial functions, and gaussian functions. The RBF kernel function (1 by I) in SVM regression problems is defined as:

$$k(x_i, x_j) = \exp\left(-\gamma\left(x_i - x_j\right)^2\right) \qquad (16)$$

Where γ represents the distribution width in the kernel function that tunes the prediction accuracy. Also, x and x' are two different observations in the dataset. To solve this nonlinear problem, the main methodology is to construct a Lagrange function from the objective function and the corresponding constraints that will be called the primal objective function, by introducing a set of variables (c). By applying the suitable kernel function in the SVM model, it is tried to maximize the below function (Smola and Schölkopf, 2004):

$$\text{Maximize:} \quad -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) - \varepsilon\sum_{i1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i,j=1}^{n} y_i(\alpha_i - \alpha_i^*) \qquad (17)$$

$$\text{Subject to:} \quad \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \quad \text{and } \alpha_i, \alpha_i^* \in [0, C]$$

In this operation, the Lagrangian multipliers (α_i and α_i^*) are optimized for each sample to minimize the error between the measured and predicted permeabilities. Likewise, the expansion of original hyperplane equation may be written as:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x_i, x) + b \qquad (18)$$

After calculation of Lagrange multipliers α_i and α_i^*, by considering that $K(x_i, x_j) = \Phi^T(x_i) \cdot \Phi(x_i)$, we can find an optimal weight vector of the hyperplane as

$$\omega = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\Phi(x_i) \qquad (19)$$

where, in large C values, the SVM tends to be overfitting while in small C cases, the SVM model leans towards underfitting. The support vector regression operation equation above has three metaparameters (C, γ, and ε). Since it is accepted that the accuracy of an SVM model relies on a correct setting of these metaparameters, these values needed to be optimized in the training stage. The values of these metaparameters control the learning speed and generalization ability of the model.

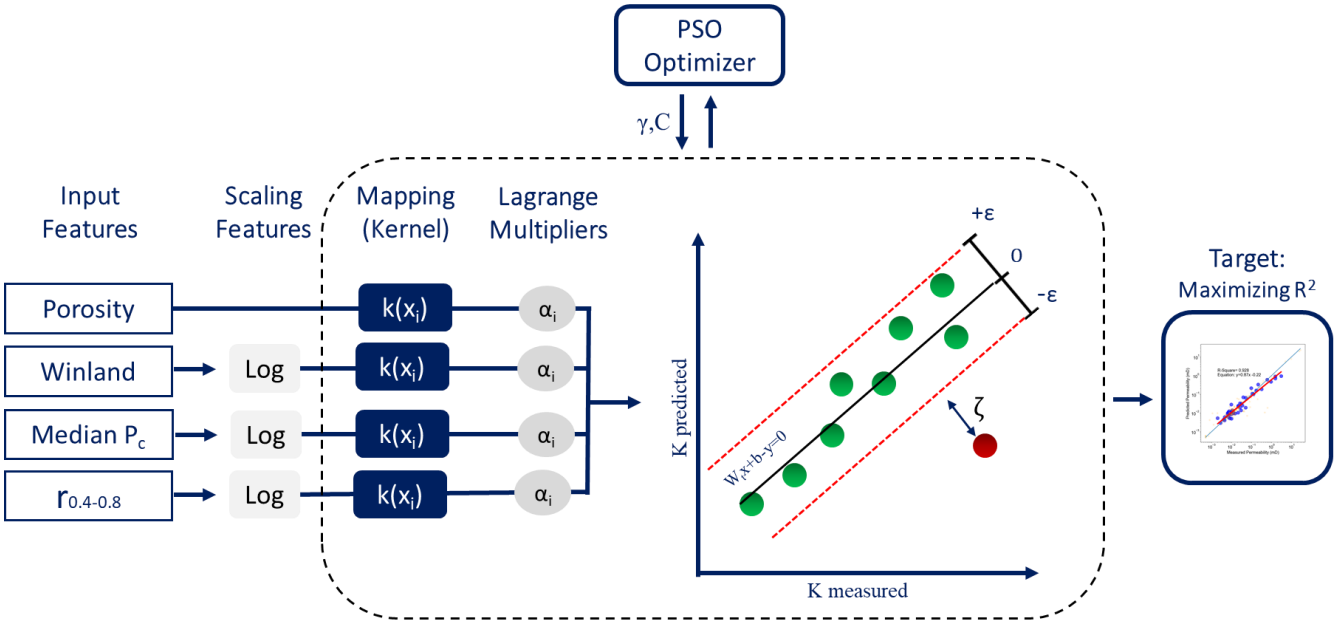

Fig. 4—A simple flowchart introducing the main specifications of the implemented SVM model coupled with the PSO algorithm.

PARTICLE SWARM OPTIMIZATION

The tuning parameters of different machine learning algorithms, which may have large impacts on the efficiency of the training process, are needed to be optimized during the training process. In this work, Particle Swarm Optimization (PSO) is applied as the optimization algorithm. This nature-inspired algorithm is only one of the evolutionary optimization methods that minimize the objective function by iterative improvement of the solutions. The method works by considering a population (called swarms) of the possible solutions (particles). The population of swarms move through the search space until the optimum solution is found. The advantage of the PSO algorithm is its tolerance in non-homogeneous conditions. This algorithm is originally introduced by Kennedy and Eberhart (1995), but a large number of variants are introduced after that. The position of a particle from x_k^i (it can be C or γ in the SVM model) will be evolved to x_{k+1}^i as:

$$x_{k+1}^i = x_k^i + v_{k+1}^i \tag{20}$$

Where the v_{k+1}^i is the movement velocity:

$$v_{k+1}^i = v_k^i + c_1 r_1 (p_k^i - x_k^i) + c_2 r_2 (p_k^g - x_k^i) \tag{21}$$

The subscript k indicates the increment of a time. p_k^i is the optimum position of the swarm i at time k so far, while p_k^g represents the global optimum position for all swarms at time k. r_1 and r_2 are random values between 0 and 1. Also, c_1 and c_2 are the cognitive and social scaling parameters, respectively, which are selected such that $c_1 = c_2 = 2$ to give a mean equal 1 when they are multiplied by r_1 and r_2. More information on the theoretical aspects related to this algorithm is provided in Kameyama (2009).

STATISTICAL ASSESSMENT

To analyze the fitting quality of the models and predictions, different statistical parameters were used that is introduced in the following. The coefficient of determination or R^2 score is the amount of the changes in the dependent variable that is predictable from the independent variable:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (22)$$

In this equation, y_i is the measured variable, x_i is the predicted variable, and $\overline{y_i}$ is the overall mean of the measured vector. Also, the Mean Squared Error (*MSE*) measures the mean squared difference between the predicted values and the measured values:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n} \quad (23)$$

The Mean Absolute Error (*MAE*) as is clear from its name, is the average value of all absolute errors:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (24)$$

MODEL IMPLEMENTATION

In this work, the support vector machine regressor algorithm is applied for the prediction of rock permeability from MICP test results. The introduction of the SVM model is provided in previous sections. Due to the complexity of the investigating phenomena, it is decided to use the RBF kernel because of its scaling capabilities. As it was shown, the SVM model and the coupled RBF kernel have 3 metaparameters that controls the accuracy and generality of the model. Since both SVM and RBF models have metaparameters that needed to be optimized, here the PSO algorithm is coupled with the SVM model. Figure 3 shows the simplified flowchart of the used model. The PSO optimizer is coupled with the SVM model to find the optimum value meta parameters and the coefficient of determination (R^2 score) is selected as the objective function to be maximized. The regularization coefficient (C) and RBF width parameter (γ) are imported as the tuning variables to fit the model. The epsilon-distance is selected to be a low value of 0.05 to improve the accuracy of the model. So, the number of dimensions in PSO model is 2 meaning that particles are crawling in a 2D space where axis are C and γ. Also, 10 particles (swarms) are added to the PSO optimizer to look up the optimized value of meta parameters. More information about the implemented model is provided in Table 1. The iteration number of 600 is selected for the active swarms to search for the meta parameters.

After randomly dividing the dataset, 80% of the samples were used for the training of the model and others are selected as the testing dataset (no validation dataset is needed in SVM models). It should be mentioned that the SVM models are not scale-invariant, meaning that the respective scale of variable influences the accuracy of the model. Because of that, it is important to reduce the scale of all inputs to the same scale of magnitude. So, the input parameters related to the permeability and pore-throat radius are transformed to the logarithmic scale.

To avoid overfitting of the model, the fitting of predicted values to the measured permeabilities are cross validated after finishing the whole optimization stage. It is checked that the fitting score of test data and training data tend to an almost similar value. Also, it is tried to remove the irrelevant input features to improve the generality of the model.

Table 1—The hyper-parameters related to the PSO algorithm and SVM model.

Metaparameters	Value
Number of Dimensions	2
Number of particles	10
C1	0.7
C2	0.7
Iterations	600
Kernel	RBF
Tolerance	1E-3
Objective Function	R^2 Score

RESULTS

After introducing the main specifications of the workflow, in this section, the results of feature selection, and model training is provided. The trained model also is tested, and the results compared to the other relationships of calculating permeability.

FEATURE ENGINEERING

In this section, the correlation of different information that could be obtained from the MICP tests and the rock permeability is investigated that will be used as the features (inputs) in the machine learning modeling section. At first, the capability of empirical/semi-analytical correlations in the prediction of rock permeability is investigated. In Fig. 5a, the porosity-

permeability relationship is shown. Also, the calculated permeability values for 5 different models including Swanson, Purcell, Winland, Parachor, and Fractal models are compared with the actual values in Fig. 5 (b-f). The mathematical details of the correlations are provided in the previous sections. As it is shown in Fig. 5, the predictions follow the overall trend of permeability. However, there are some noises in the measured values that is a good indication that the calculated values may be highly influenced by the shape of the capillary curve, meaning that any measurement errors in these curves may lead to large prediction errors. Figure 6 shows the prediction errors for different permeability models. It is shown that the predictions related to the Winland, and Swanson model and simple linear regression of Porosity-Permeability scatter plot has the minimum errors. The fractal model resulted in the highest error values. It is while Fig. 5 shows that the fractal model could predict the permeability for a large fraction of the dataset and only a smart portion of the data had large errors. Purcell and Parachor models almost shown similar results and prediction errors. On the other hand, the calculated Kendall's tau in Table 2 shows that the Winland r_{35} and Porosity-Permeability relationship have the highest correlation with the permeability values (τ>0.5). Other correlations have τ values of less than 0.4. In machine learning models the importance of the correlation coefficient is undeniable. So, in this work, considering the MSE and correlation coefficients, the Porosity itself and the Winland model is selected as the most suitable features for the prediction of rock permeability.

Table 2—The statistical evaluation of the empirical permeability models and their correlation with the measured permeabilities

Models	MAE (Log mD)	MSE (Log mD)	Correlation Coefficient
Porosity-Permeability	0.34	0.77	0.545
Winland	0.92	0.98	0.645
Swanson	0.77	1.34	0.387
Purcell	2.48	6.85	0.283
Parachor	1.81	4.74	0.294
Fractal	2.40	10.32	0.151

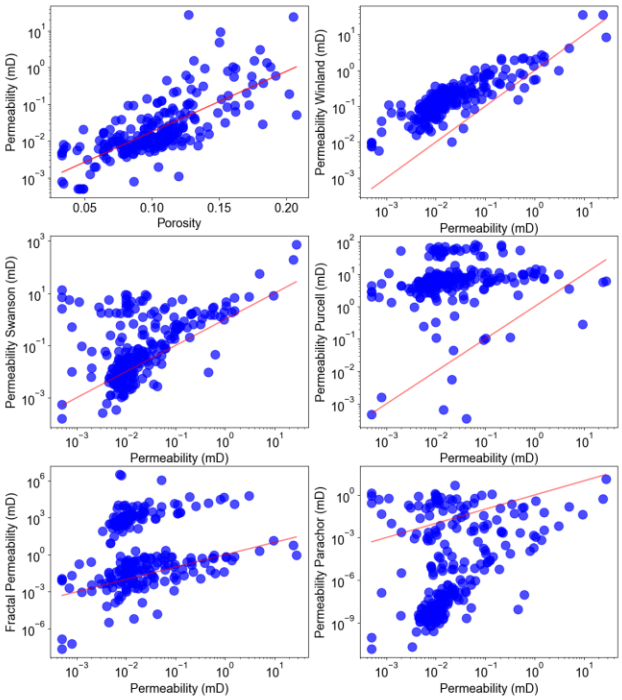

Fig. 5—The calculated values of permeability for different correlations. (a) Porosity-Permeability plot, (b) Winland method, (c) Swanson method, (d) Purcell method, (e) Fractal model, (f) Parachor model. The red lines show the 1:1 permeability value.

Table 3: The statistical features extracted from the capillary pressure curves and their correlation with the measured permeability values.

Feature	Correlation Coefficient
Average Pc	-0.380
Median Pc	-0.599

Figure 7 shows the scatter plot related to the average equivalent pore-throat radius values for different saturation ranges of capillary pressure curves. It is clear from the figure that overall, the permeability is correlated with the capillary pressure shape and trends. However, here it is tried to find the ranges with the highest correlation values. As Table 4 shows Kendal's correlation coefficient values for these ranges, the most optimum sampling range is in the range of 0.4-0.8 with where the calculated pore radii have the most correlations with the rock permeability. This is in line with the results of previous studies such as Kolodzie (1980). Actually, the pores filled in this saturation range has the highest population and highest contribution to rock permeability. On the other hand, the saturation

ranges of 0-0.2 have the worst results that are acceptable considering that the low saturation values are related to the tightest pores. These pores cannot be sufficiently effective in the fluid transmissibility of rock.

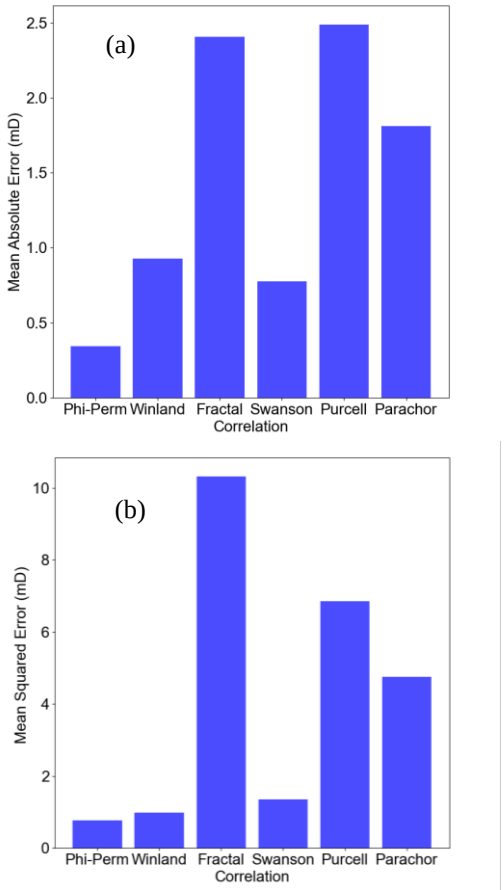

Fig. 6—Comparison of the results of different correlations, (a) Mean absolute error, (b) Mean squared errors

Table 4—The statistical parameters related to the sampled pore throat radiuses for different saturation ranges.

Saturation Range	Correlation Coefficient
$r_{0\text{-}0.2}$	0.273
$r_{0.2\text{-}0.4}$	0.450
$r_{0.4\text{-}0.6}$	0.563
$r_{0.6\text{-}0.8}$	0.587
$r_{0.8\text{-}1}$	0.476
$r_{0.4\text{-}0.8}$	0.580

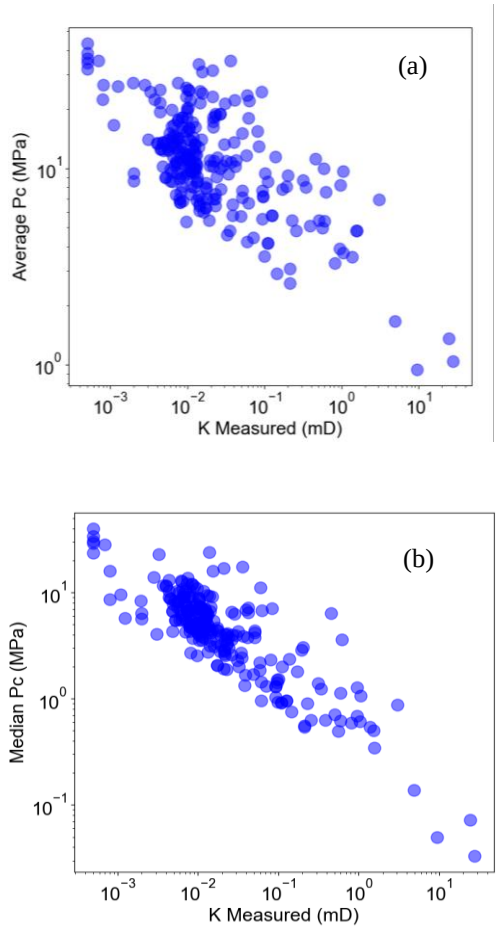

Fig. 7—Correlation of the capillary pressure curve derived features with the measured permeability. More details are provided in Table 3. (a) Average capillary pressure, (b) Median capillary pressure

Finally, the optimum combination of features is selected based on Kendal's correlation coefficient and also the MSE of predictions. Since the training of the machine learning model is based on the relevance of the features, the features with a correlation coefficient higher than 0.5 were prioritized (Fig. 9). So, the below list is selected as the inputs of the SVM-PSO model:

- Porosity
- Winland permeability model (log mD)
- The median value of capillary pressure (log bar)
- The average pore-throat radius calculated from capillary pressure curves at the wetting phase saturation range of 0.4 to 0.8 (log μm)

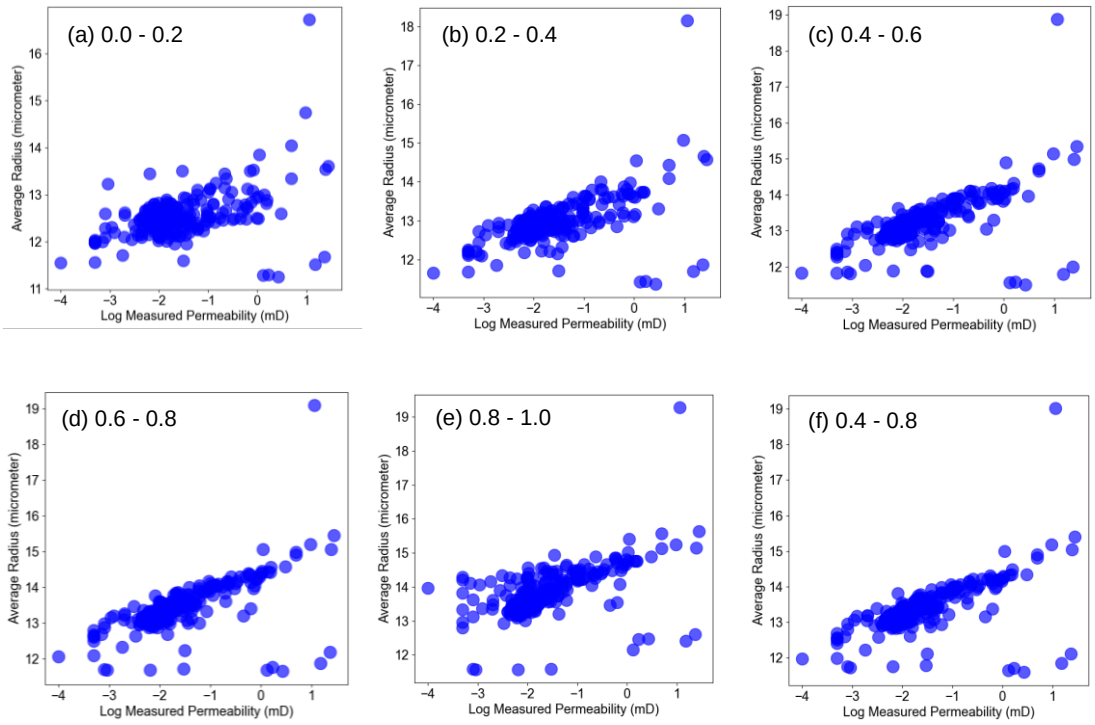

Fig. 8—Correlation of the pore radii (capillary pressure curves) versus the measured permeability for different saturation ranges.

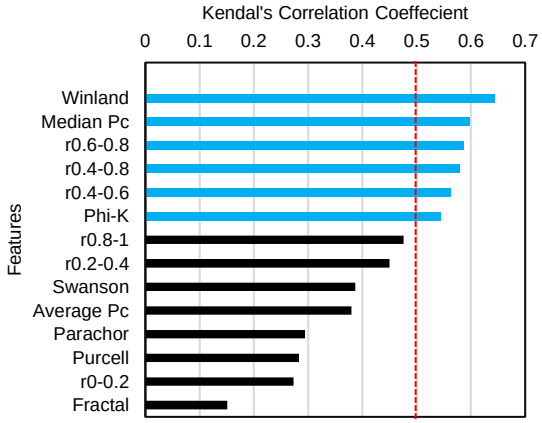

Fig. 9—The tornado chart related to the value of Kendal's correlation coefficient (absolute value) ordered from high to low. The features with the τ score higher than 0.5 are selected for the model development process.

MODEL TRAINING AND VALIDATION

After adding the training dataset (80% of the dataset) to the implemented SVM-PSO model, the model is trained, and the results are shown in Fig. 10. The R^2 score of the training process was 0.89 that is significant in comparison to the previously introduced models. Also, the other 20% of the samples are separated in the testing process. Figure 10c shows the predicted values of permeability versus measured ones in the testing stage. As Fig. 10d shows, the accuracy of predictions is significantly higher than correlations like the Winland equation. The detailed statistics on the comparison of the SVM-PSO model and Winland equation is shown in Table 5 that depicts that the correlation coefficient score is improved. In previous features, the maximum correlation coefficient was not higher than 0.645, while in the SVM-PSO model, it increased to the level of 0.74 that is an improvement. Also, the *MSE* value for the Winland equation was around 0.95, while this value is reduced to the level of 0.074 in SVM-PSO model testing results. It shows a significant improvement in the level of permeability predictions for the SVM-PSO model.

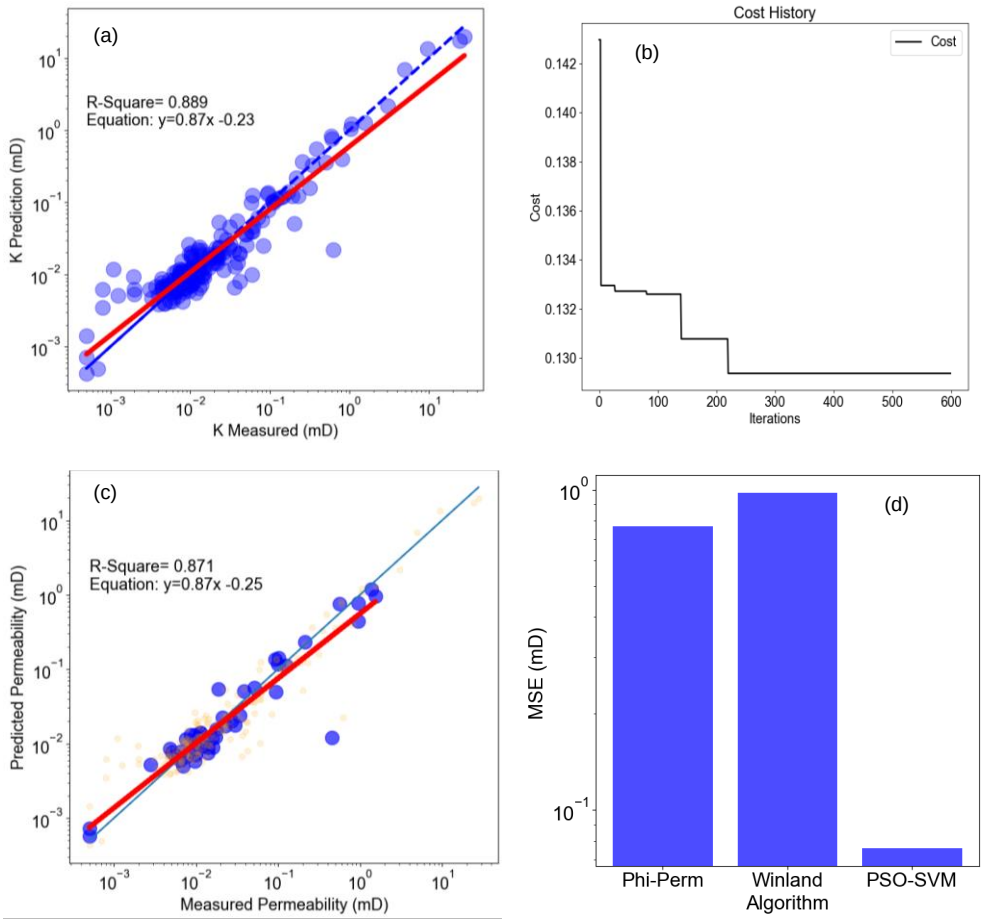

Fig. 10—The results of training and testing of the Support Vector Machine model using the PSO optimization algorithm. (a) comparing the predicted values and actual measured values. The red line shows the fitted curves on the predicted values and the blue line shows the y=x trend. (b) The convergence trend of PSO model for iterations, (c), The test results for prediction of core permeability from MICP data using the combination of SVM and PSO algorithm. Orange points in the background (low transparency) shows the training results (d) The mean absolute percentage error related to different machine learning algorithms and the selected correlations.

Table 5—The training and testing results of SVM-PSO permeability prediction model and their comparison to the Winland correlation.

Algorithm	R^2	MSE (log mD)	Correlation Coefficient
SVM-PSO Training	0.89	0.067	0.72
SVM-PSO Testing	0.87	0.074	0.74
SVM-PSO Total	0.88	0.071	0.72

CONCLUSION

The permeability is a determining parameter in fluid flow analysis in geosciences. Due to the role of pore structures on the flow capacity of rocks, several models are previously provided to predict the rock permeability from the mercury injection capillary pressure data. In this work, a large dataset of MICP tests related to tight sandstones from all over the world is gathered. Comparison to predictions obtained from the models shows that they could not acceptably predict the rock permeabilities. The most accurate model was Winland r_{35} equation. To improve the predictions a study is carried out to find the most relevant features to be imported as the input variables to the machine learning

model. The median capillary pressure value calculated from the capillary pressure curve was acceptably correlated with the rock permeability. Also, the average pore throat radius of rock related to the pore related to the saturation range of 0.4-0.8 best fits with the rock permeability measurements, comparing to the other saturation ranges.

The SVM-PSO machine learning model fitted with the actual permeability values with R^2 of 0.89 significantly higher than the R^2 values of theoretical based permeability models. This shows that merging machine-learning algorithms with current empirical or physics-based models can significantly improve permeability modeling practices.

ACKNOWLEDGEMENTS

Andersen acknowledges the Research Council of Norway and the industry partners, ConocoPhillips Skandinavia AS, Aker BP ASA, Vår Energi AS, Equinor ASA, Neptune Energy Norge AS, Lundin Norway AS, Halliburton AS, Schlumberger Norge AS, and Wintershall DEA, of The National IOR Centre of Norway for support.

REFERENCES

Abbasi, J., Andersen, P.Ø., 2021. Theoretical Comparison of Two Setups for Capillary Pressure Measurement by Centrifuge 2021, 1–18. doi:https://doi.org/10.3997/2214-4609.202133154

Ahmad, N., Wörman, A., Sanchez-Vila, X., Jarsjö, J., Bottacin-Busolin, A., Hellevang, H., 2016. Injection of CO2-saturated brine in geological reservoir: A way to enhanced storage safety. Int. J. Greenh. Gas Control 54, 129–144. doi:10.1016/j.ijggc.2016.08.028

Arabjamaloei, R., Daniels, D., Ebeltoft, E., Petersen, E., Pitman, R.J., Ruth, D., 2019. Validation of permeability and relative permeability data using mercury injection capillary pressure data. E3S Web Conf. 89. doi:10.1051/e3sconf/20198901001

Blunt, M.J., 2017. Multiphase Flow in Permeable Media. Cambridge University Press. doi:10.1017/9781316145098

Burn, R.P., Mandelbrot, B.B., 1984. The Fractal Geometry of Nature, The Mathematical Gazette. WH freeman New York. doi:10.2307/3615422

Changtao, Y., Shuyuan, L., Hailong, W., Fei, Y., 2018. Pore structure characteristics and methane adsorption and desorption properties of marine shale in Sichuan Province, China. RSC Adv. 8, 6436–6443.

Choi, S., Kim, T., Yu, W., 2009. Performance evaluation of RANSAC family. Br. Mach. Vis. Conf. BMVC 2009 - Proc. 1–12. doi:10.5244/C.23.81

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Croux, C., Dehon, C., 2010. Influence functions of the Spearman and Kendall correlation measures. Stat. Methods Appl. 19, 497–515. doi:10.1007/s10260-010-0142-z

Erofeev, A., Orlov, D., Ryzhov, A., Koroteev, D., 2019. Prediction of Porosity and Permeability Alteration Based on Machine Learning Algorithms. Transp. Porous Media 128, 677–700. doi:10.1007/s11242-019-01265-3

Eslami, M., Kadkhodaie-Ilkhchi, A., Sharghi, Y., Golsanami, N., 2013. Construction of synthetic capillary pressure curves from the joint use of NMR log data and conventional well logs. J. Pet. Sci. Eng. 111, 50–58. doi:10.1016/j.petrol.2013.10.010

Fan, C., Zhong, C., Zhang, Y., Qin, Q., Shun, H.E., 2019. Geological Factors Controlling the Accumulation and High Yield of Marine-Facies Shale Gas: Case Study of the Wufeng-Longmaxi Formation in the Dingshan Area of Southeast Sichuan, China. Acta Geol. Sin. 93, 536–560. doi:10.1111/1755-6724.13857

Feng, F., Wang, P., Wei, Z., Jiang, G., Xu, D., Zhang, Jiqiang, Zhang, Jing, 2020. A New Method for Predicting the Permeability of Sandstone in Deep Reservoirs. Geofluids 2020. doi:10.1155/2020/8844464

Ge, X., Fan, Y., Deng, S., Han, Y., Liu, J., 2016. An improvement of the fractal theory and its application in pore structure evaluation and permeability estimation. J. Geophys. Res. Solid Earth 121, 6333–6345.

Guo, B., Ghalambor, A., Duan, S., 2004. Correlation between sandstone permeability and capillary pressure curves. J. Pet. Sci. Eng. 43, 239–246. doi:10.1016/j.petrol.2004.02.016

Hébert, V., Porcher, T., Planes, V., Léger, M., Alperovich, A., Goldluecke, B., Rodriguez, O., Youssef, S., 2020. Digital core repository coupled with machine learning as a tool to classify and assess petrophysical rock properties. E3S Web Conf. 146. doi:10.1051/e3sconf/202014601003

Jiang, Z., Mao, Z., Shi, Y., Wang, D., 2018. Multifractal characteristics and classification of tight sandstone reservoirs: A case study from the Triassic Yanchang Formation, Ordos Basin, China. Energies 11. doi:10.3390/en11092242

Jiao, L., Andersen, P.Ø., Zhou, J., Cai, J., 2020. Applications of mercury intrusion capillary pressure for pore structures: A review. Capillarity 3, 62–74. doi:10.46690/capi.2020.04.02

Jones, S.C., 1997. A Technique for Faster Pulse-Decay Permeability Measurements in Tight Rocks. SPE Form. Eval. 12, 19–26. doi:10.2118/28450-PA

Kameyama, K., 2009. Particle swarm optimization-a survey. IEICE Trans. Inf. Syst. 92, 1354–1361.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V., 2019. Machine Learning for the Geosciences: Challenges and Opportunities. IEEE Trans. Knowl. Data Eng. 31, 1544–1554. doi:10.1109/TKDE.2018.2861006

Kendall, M.G., 1976. Rank Correlation Methods 4th edn (London: Griffin). High Wycombe, Bucks.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: Proceedings of ICNN'95-International Conference on Neural Networks. IEEE, pp. 1942–1948.

Kolodzie, S., 1980. Analysis of Pore Throat Size and Use of the Waxman-Smits Equation To Determine Ooip in Spindle Field, Colorado. Soc. Pet. Eng. AIME, SPE. doi:10.2523/9382-ms

Lai, J., Wang, G., 2015. Fractal analysis of tight gas sandstones using high-pressure mercury intrusion techniques. J. Nat. Gas Sci. Eng. 24, 185–196. doi:10.1016/j.jngse.2015.03.027

Li, K., 2010. Analytical derivation of Brooks–Corey type capillary pressure models using fractal geometry and evaluation of rock heterogeneity. J. Pet. Sci. Eng. 73, 20–26.

Lin, Q., Bijeljic, B., Pini, R., Blunt, M.J., Krevor, S., 2018. Imaging and Measurement of Pore-Scale Interfacial Curvature to Determine Capillary Pressure Simultaneously With Relative Permeability. Water Resour. Res. 54, 7046–7060. doi:10.1029/2018WR023214

Liu, K., Mirzaei-Paiaman, A., Liu, B., Ostadhassan, M., 2020. A new model to estimate permeability using mercury injection capillary pressure data: Application to carbonate and shale samples. J. Nat. Gas Sci. Eng. 84, 103691. doi:10.1016/j.jngse.2020.103691

Liu, M., Xie, R., Li, C., Li, X., Jin, G., Guo, J., 2018. Determining the segmentation point for calculating the fractal dimension from mercury injection capillary pressure curves in tight sandstone. J. Geophys. Eng. 15, 1350–1362. doi:10.1088/1742-2140/aab1d8

Liu, X., Xie, Y., Shu, H., Fan, X., Wang, G., Luo, Y., Ma, X., Qin, Z., 2020. Effect of pore structure on oil-bearing property in the third member of Paleogene Funing Formation in Subei Basin, East China. Energy Sci. Eng. 8, 2187–2202. doi:10.1002/ese3.657

Liu, Y., Xian, C., Li, Z., Wang, J., Ren, F., 2020. A new classification system of lithic-rich tight sandstone and its application to diagnosis high-quality reservoirs. Adv. Geo-Energy Res. 4, 286–295. doi:10.46690/ager.2020.03.06

McPhee, C., Reed, J., Zubizarreta, I., 2015. Core analysis: a best practice guide. Elsevier.

Menke, H.P., Maes, J., Geiger, S., 2021. Upscaling the porosity–permeability relationship of a microporous carbonate for Darcy-scale flow with machine learning. Sci. Rep. 11, 1–10. doi:10.1038/s41598-021-82029-2

Purcell, W.R., 1949. Capillary Pressures - Their Measurement Using Mercury and the Calculation of Permeability Therefrom. J. Pet. Technol. 1, 39–48. doi:10.2118/949039-g

Puth, M.-T., Neuhäuser, M., Ruxton, G.D., 2015. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. Anim. Behav. 102, 77–84.

Rezaee, R., Saeedi, A., Clennell, B., 2012. Tight gas sands permeability estimation from mercury injection capillary pressure and nuclear magnetic resonance data. J. Pet. Sci. Eng. 88, 92–99.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222.

Swanson, B.F., 1981. Simple Correlation Between Permeabilities and Mercury Capillary Pressures. JPT, J. Pet. Technol. 33, 2498–2504. doi:10.2118/8234-PA

Tran, H., Sakhaee-Pour, A., Bryant, S.L., 2018. A Simple Relation for Estimating Shale Permeability. Transp. Porous Media 124, 883–901. doi:10.1007/s11242-018-1102-6

Wang, F., Jiao, L., Liu, Z., Tan, X., Wang, C., Gao, J., 2018. Fractal analysis of pore structures in low permeability sandstones using mercury intrusion porosimetry. J. Porous Media 21, 1097–1119. doi:10.1615/JPorMedia.2018021393

Wang, F., Yang, K., You, J., Lei, X., 2019. Analysis of pore size distribution and fractal dimension in tight sandstone with mercury intrusion porosimetry. Results Phys. 13, 102283. doi:10.1016/j.rinp.2019.102283

Wang, Jianmin, Wang, Jiayuan, 2013. Low-amplitude structures and oil-gas enrichment on the Yishaan Slope, Ordos Basin. Pet. Explor. Dev. 40, 52–60. doi:10.1016/S1876-3804(13)60005-1

Xiao, L., Liu, D., Wang, H., Li, J., Lu, J., Zou, C., 2017. The applicability analysis of models for permeability prediction using mercury injection capillary pressure (MICP) data. J. Pet. Sci. Eng. 156, 589–593. doi:10.1016/j.petrol.2017.06.042

Xiao, L., Liu, X.P., Zou, C.C., Hu, X.X., Mao, Z.Q., Shi, Y.J., Guo, H.P., Li, G.R., 2014. Comparative study of models for predicting permeability from nuclear magnetic resonance (NMR) logs in two Chinese tight sandstone reservoirs. Acta Geophys. 62, 116–141. doi:10.2478/s11600-013-0165-6

Zhang, Chong, Cheng, Y., Zhang, Chaomo, 2017. An improved method for predicting permeability by combining electrical measurements and mercury injection capillary pressure data. J. Geophys. Eng. 14, 132–142. doi:10.1088/1742-2140/14/1/132

Zhang, G., Wang, Z., Mohaghegh, S., Lin, C., Sun, Y., Pei, S., 2021. Pattern visualization and understanding of machine learning models for permeability prediction in tight sandstone reservoirs. J. Pet. Sci. Eng. 200, 108142. doi:10.1016/j.petrol.2020.108142

ABOUT THE AUTHORS

Jassem Abbasi is PhD candidate in Petroleum Technology at University of Stavanger. His focus is on the application of Deep Learning methods in improving the prediction of physical processes, especially flow in porous media. Before starting his PhD, for almost 5 years, he worked with different institutions/ companies as both researcher and engineer in different technical disciplines such as reservoir engineering, reservoir simulation, enhanced oil recovery (EOR), reservoir fluid properties, and software development.

Jiuyu Zhao is a Geological Resources and Geological Engineering PhD candidate in China University of Petroleum (Beijing), tutored by Professor Jianchao Cai now, and received his Post-Graduate degree of Oil and Gas Development Engineering in China University of Petroleum (Beijing). He focuses on digital core technology and deep learning algorithms about rock image processing, reconstruction and flow simulation

Sameer Ahmed is a Reservoir Engineer, currently working as Support Manager at Resoptima AS. His area of interest is focused on Ensemble based reservoir modelling. He received his master degree in Petroleum Engineering from University of Stavanger Norway.

Jianchao Cai is a professor of Geological Resources and Geological Engineering at the China University of Petroleum (Beijing). For more than 15 years, his work has been focused on the petrophysical characterization and micro-transport phenomena in porous media, and fractal theory as well as its applications. He has served as a visiting scholar at University of Tennessee-Knoxville (USA.), and at King Abdullah University of Science and Technology (Saudi Arabia). Additionally, he is the Founder and Editor-in-Chief of Advances in Geo-Energy Research and serves as Associate Editor or a member of the Editorial Board for several international journals. He received an award from the National Science Foundation of China for Outstanding Youth Foundation in 2017 and the First Prize of Natural Science of Hubei Province in 2021. He has managed and completed more than 30 projects to date on oil and gas reservoir evaluation and published more than 170 peer-reviewed journal articles, 6 books, and numerous book chapters.

Pål Østebø Andersen is Associate Professor at University of Stavanger and holds a PhD in Petroleum Technology. He has authored over 80 technical papers and received the 2021 SPE Cedric K Ferguson Medal. His main research interests include simulation, special core analysis, design and interpretation of experiments, machine learning and the energy transition.

Liang Jiao is an oil and natural gas engineering PhD in Professor Jianchao Cai research group at China University of Geosciences in Wuhan. He focuses on pore structure characterization and seepage law of tight oil/shale oil reservoirs. He received Post-Graduate degree in China University of Petroleum (Beijing).