Prediction of centrifuge capillary pressure using machine learning techniques

Brandon Jeremy Bursey[1], *Erfan* Mohagheghian[1], *Edison* Sripal[1] and *Lesley Anne* James[1,*]

[1]Department of Process Engineering, Memorial University of Newfoundland, St. John's, NL, Canada

Abstract. In current literature in the petroleum industry, machine learning has been used to predict capillary pressure only on the centrifugal data points and not the complete capillary pressure curves generated from existing correlations after analysis. This paper will present novel information that will benefit the petroleum industry as it shows machine learning techniques can be used to obtain the complete capillary pressure curve which is the end goal in undertaking an SCAL centrifuge experiment.

This research involves testing core samples using a centrifuge set up to produce capillary pressure data points. Then, using a commercial SCAL interpretation software, the collected data is utilized to generate complete capillary pressure curves based on developed literature correlations. RCAL data for the core samples is also obtained to be used with the machine learning techniques. The machine learning models are then applied to the collected data to predict the capillary pressure curves. Optimization of the different machine learning techniques is done to improve the predictions.

The results show the machine learning techniques perform very well on the validation set after being trained on the training set. The machine learning models also provide reasonable prediction of the complete capillary pressure curves on the testing data set. Changing of the machine learning technique parameters also shows the effect on the overall precision and the improvements that can be made. Further research can be done to see the effectiveness of using machine learning techniques to predict other SCAL properties such as relative permeability. This can then greatly reduce the time needed to obtain these extremely important properties for reservoir characterization.

1 Introduction

Researchers continue to find new ways to reduce the time and improve the accuracy of reservoir characterization [1]. Specialized core analysis (SCAL) examines important reservoir characteristics such as capillary pressure and relative permeability to estimate incremental oil recovery during secondary water or gas flooding and tertiary enhanced oil recovery. Arguably, SCAL is more time consuming and has more uncertainty compared to routine core analysis and fluid characterisation. Capillary pressure (P_c) is the pressure between two immiscible phases in a capillary tube or, in this case, within the rock's pores [2]. Capillary pressure is measured in a lab through spontaneous imbibition, gravity drainage, and forced imbibition tests. P_c is used directly in reservoir simulators to account for unrecoverable oil and it is further used to estimate relative permeability or the relative conductivity of the rock to more than one fluid phase simultaneously. The three main test methods used to determine capillary pressure curves are centrifuge, porous plate, and mercury injection [2]. The advantages and disadvantages of these methods are summarized in Table 1. It is worth pointing out that mercury injection capillary pressure (MICP) does not respond to wettability components in the pores since Hg is an ideal non-wetting fluid. Hence MICP reflects only the pore geometry and therefore differs slightly from other tests that use oil/brine/gas.

The oil and gas industry has always been highly competitive and somewhat unpredictable. In recent years, companies have begun using machine learning techniques to combat several challenges and issues in data processing and handling in various oil and gas activities such as reducing risk factors and cost of maintenance [3]. In current literature, machine learning has been used to predict the experimental capillary pressure data points obtained from centrifugal and mercury injection tests with varying results [4-6]. For the case that did use centrifugal data, correlations were not applied to the experimental results to obtain a complete capillary pressure curve. This research aims to obtain complete capillary pressure curves generated from existing correlations after analysis of centrifugal data and then test the novelty in using machine learning techniques to predict the experimental capillary pressure curve and water saturation. Three machine learning techniques, including the artificial neural network, support vector regression and random forest regression were used to predict the capillary pressure from centrifuge test samples with routine core analysis data (porosity, permeability, grain density, and irreducible water saturation) being used as inputs. ML-based approaches generate a multi-point capillary pressure curve as a function of fluid saturation using the provided input parameters. The ML prediction is not just a single value just as the experimentally measured and calculated capillary pressure values are recorded at discrete fluid saturation points. Although experimental techniques are broadly used to generate the capillary pressure curve of core samples, they are expensive and require core samples to be in reservoir conditions that is very difficult to achieve. As a result, researchers ponder whether data-driven

* Corresponding author: ljames@mun.ca

methods can estimate SCAL data. Can one develop accurate machine learning methods to predict SCAL data such as capillary pressure of core samples from routine core analysis data without a need for time consuming and finicky laboratory experiments?

Table 1. Pc Methods – Advantages and Disadvantages [2]

Method	Advantages	Disadvantages
Centrifuge	- Applicable to most rock types - Simple and quick - Non-destructive test - Uses synthetic reservoir fluids	- Most expensive method - Raw production data needs to be corrected by the test lab
Porous Plate	- Best method to achieve a uniform saturation profile - Uses synthetic reservoir fluids - Non-destructive test	- Time consuming to achieve complete Pc curve - Time allowed to achieve equilibrium can completely govern resultant curve
Mercury Injection	- Low-cost technique - Rapid tests	- Not representative of reservoir fluids - Test is destructive (mercury remains in sample)

1.1 Machine Learning Technique Types

Machine learning is the broad study of programming algorithms that can learn through experience from the use of a variety of data [7]. Machine learning techniques are categorized into either supervised or unsupervised learning with supervised learning being further broken into classification and regression as shown in Fig. 1.
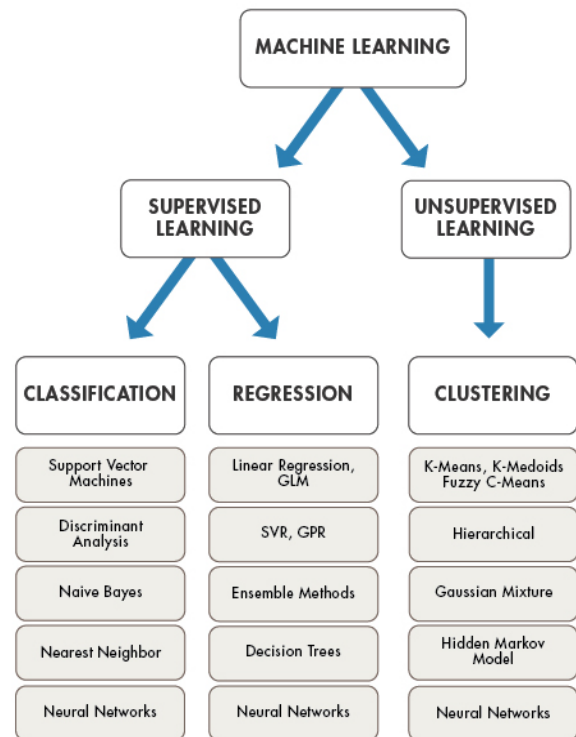
Fig. 1. Machine Learning Algorithm Breakdown [8]

For the purposes of this research, supervised learning regression techniques that were used are briefly explained as follows.

1.1.1 Support Vector Regression

Support vector regression (SVR) is a popular technique where support vector machines are adapted for regression with a quantitative response [9]. The technique provides flexibility in defining the acceptable error for the model and determines an appropriate line or hyperplane fit to the data [10]. For SVR, the objective function is to minimize the L2-norm of the coefficient vector. One strength of SVR is a variety of kernel functions being available to select or modify depending on the requirements for your own predictions [11]. The kernel function transforms data that are not linearly separable to linearly separable in a higher dimensional space. As well, given its dependence on support vectors, not all training data is needed for predictions. The support vectors determine the decision boundaries between the data from different trends that are used to do regression. SVR is a supervised technique in which data are fed into the model in (input, output) pairs.

1.1.2 Artificial Neural Network

Artificial neural network (ANN) is another regression technique that learns from processing (input, output) data pairs supplied during the training-testing split of input data. The more examples and variety of inputs the model is given, the more accurate the model typically will predict outputs [9]. Fig. 2 below shows the architecture of an ANN model, where the X nodes are the input layer, the Z nodes are the hidden layer, and the Y nodes are the output layer. For this study, X consists of porosity, permeability, grain density, and irreducible water saturation and Y consists of capillary pressure and water saturation. The number of hidden layers

and epochs (number of cycles through the training data) can be varied to see their effect on the predictions. The input to every hidden layer is the linear combination of the outputs of the previous layer passed through a non-linear function, also called the activation function. Activation function can be one of hyperbolic tangent (tanh), sigmoid, Rectified Linear Unit (ReLU), linear, etc.
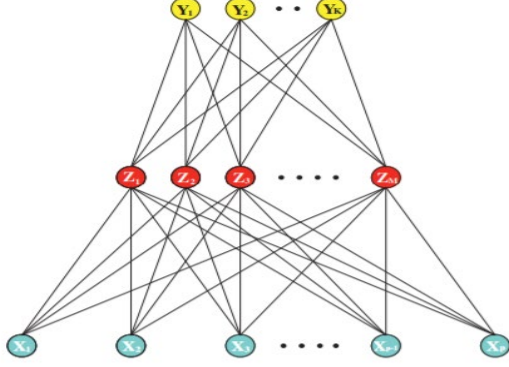

Fig. 2. Single Hidden Layer ANN [8]

1.1.3 Random Forest Regression

Random forest regression (RFR) is a supervised learning technique that utilizes ensemble learning for regression analysis. The method uses predictions from multiple learning algorithms to obtain a more accurate prediction [12]. RFR is also one of the most powerful regression techniques and is useful for non-linear relationships. A typical setup of a RFR model is where multiple decision trees are constructed on training data and the mean of the trees is outputted as the prediction. Each tree processes the input parameters, including the porosity, permeability, grain density, and irreducible water saturation to predict capillary pressure and water saturation. Therefore, capillary pressure and water saturation are the output of each tree. The final output will be the average of all trees outputs in the forest.

However, it should be pointed out that capillary pressure curves generated by machine learning methods and experimental methods are often not accurate and seldom match the distribution of wetting and non-wetting phases sensed by imaging techniques. An ML-based method can predict water saturation and capillary pressure curve much faster than laboratory methods and potentially with no more uncertainty. Also, ML-based methods can be generalized to different fields if enough samples are used to train the models. In this manner, the need for core samples from new fields in reservoir conditions is mitigated.

2 Literature Review

2.1 Centrifuge Capillary Pressure Correlations

The centrifuge capillary pressure problem involves solving the challenge of determining local saturations along the length of the core based on the averaged saturations measured during the centrifuge experiment [13]. To solve the centrifuge capillary pressure problem numerous correlations have been developed that take on the general form of Eq. 1.

$$S(P_{c1}) = a_1 \hat{S}(P_{c1}) + a_2 P_{c1} \frac{d\hat{S}(P_{c1})}{dP_{c1}} + E \tag{1}$$

where a_1, a_2, and E will vary depending on the correlation used [14]. The main assumptions are hydrostatic equilibrium being reached at each phase and the boundary condition of the capillary pressure being zero at the outflow face [13]. $S(P_{c1})$ is the local saturation and $\hat{S}(P_{c1})$ is the average saturation measured during the centrifuge experiment. P_{c1} is determined from Eq. 2.

$$P_{c1} = \frac{1}{2}\Delta\rho\omega^2(r_2^2 - r_1^2) \tag{2}$$

Where $\Delta\rho$ is the difference in the fluid densities and r_1 and r_2 are the radii at the inner and outer faces of the sample in the centrifuge core holder and ω is the rotation speed.

2.1.1 Hassler Brunner Method

The simplest solution for the centrifuge capillary problem was developed by Hassler and Brunner and it is also the poorest solution as it neglects both radial and gravity effects [15]. These assumptions can be satisfied for very short and narrow samples spun far from the rotation axis. In the case of the general centrifuge capillary problem, the Hassler-Brunner equation reduces to the following:

$$S(P_{c1}) = \hat{S}(P_{c1}) + P_{c1} \frac{d\hat{S}(P_{c1})}{dP_{c1}} \tag{3}$$

where $a_1 = 1$, $a_2 = 1$ and $E = 0$.

Forbes first solution demonstrated that this solution is always significantly lower, in terms of saturation, than the true $S(P_c)$ solution [14].

2.1.2 Forbes First Solution

Forbes built upon previous solutions to the centrifuge capillary problem and came up with his own solution that still neglects both radial and gravity effects but incorporates the difference in the centrifuge core radii [13]. Forbes solution took the form shown in Eq. 4.s

$$S(P_{c1}) = \hat{S}(P_{c1}) + \frac{1}{1+\alpha} P_{c1} \frac{d\hat{S}(P_{c1})}{dP_{c1}} \tag{4}$$

where $a_1 = 1$, $a_2 = \frac{1}{1+\alpha}$ and $E = 0$.

$$\alpha = \frac{1 - \sqrt{1-B}}{1 - 2\sqrt{1-B}} \tag{5}$$

$$B = 1 - \left(\frac{r_1}{r_2}\right)^2 \tag{6}$$

2.1.3 Spline Functions

Nordtvedt and Kolltvelt were first to approximate the wetting phase saturation as a function of the capillary pressure whereby the capillary pressure curve is approximated piecewise by a polynomial [16]. This parameter estimation technique is usually related to functions which are too simplistic to describe the shape of capillary pressure curves [15]. A linear system of equations is then obtained and the piecewise approximation polynomial takes on the form shown in Eq. 7.

$$S(P_{c1}) = a_{i,o} + a_{i,1} P_{c1} + a_{i,2} P_c^2 \quad i = 1,2,\dots n_I \tag{7}$$

Where i indicates the interval number and n_I refers to the total number of intervals.

2.2 Current Machine Learning Applications

Busaleh et al. used several types of machine learning techniques to predict capillary pressure in carbonate oil reservoirs from data comprised of mercury injection drainage capillary pressure data [4]. Their data consisted of 202 carbonate core samples that contained at minimum 70 capillary pressure versus saturation points in addition to the corresponding porosity, permeability, and grain density for the core. For their machine learning techniques, the inputs used were porosity, permeability, grain density, and water saturation with their output being capillary pressure. They used a training-testing split of 70-30 of the core sample data. With the variety of machine learning techniques came with mixed results. ANN-trainbr, ANN-trainlm, ANN-trainscg, ANN-trainoss refer to neural networks using Bayesian regularization backpropagation, Liebenberg-Marquardt optimization, scaled conjugate gradient, and one-step secant methods, respectively.
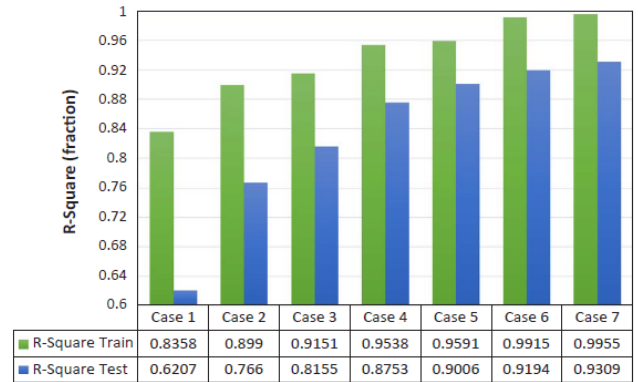
The decision tree algorithm performed the best on the training data however this is likely due to overfitting as it performed considerably worse on the testing set. The ANN-trainbr method performed the best overall; however, there is still significant room for improvement [4].

Jamshidian et al. obtained centrifuge capillary pressure data for 15 core samples each of which having 27 capillary pressure data points totalling 405 data points. They completed seven cases to see the effects of additional input parameters on the final predictions [5]. An ANN model with a Cuckoo optimization algorithm was used to predict capillary pressure as the only output which was then compared against the experimental centrifuge values. The training set consisted of 8 cores with the remaining 7 used for validation.

**Table 2.** ANN Models with Input Parameters [5]

ANN Modelling Case	Contributing Input Parameter(s)
Case 1	S_w
Case 2	S_w, k_{air}
Case 3	S_w, k_{air}, φ
Case 4	S_w, k_{air}, φ, $S_{w,irr}$
Case 5	S_w, k_{air}, φ, $S_{w,irr}$, k_w
Case 6	S_w, k_{air}, φ, $S_{w,irr}$, k_w, k_o @ $S_{w,irr}$
Case 7	S_w, k_{air}, φ, $S_{w,irr}$, k_w, k_o @ $S_{w,irr}$, ρ_{grain}

Table 2 above shows the different cases modelled with the addition of input parameters. Water saturation, air permeability, porosity, irreducible water saturation, water permeability, oil permeability at irreducible water saturation, and grain density were the contributing input parameters.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
R-Square Train	0.8358	0.899	0.9151	0.9538	0.9591	0.9915	0.9955
R-Square Test	0.6207	0.766	0.8155	0.8753	0.9006	0.9194	0.9309

Fig. 3. Correlation Coefficients for the Different Cases [5]

From Fig. 3 the increase of input parameters correlates with an increase in the correlation coefficient for both the training and testing data sets. Case 1 through 7 correlates to the ANN modelling case and contributing input parameters shown in Table 2. Also, the greatest prediction comes with all the input parameters being used. Fig. 4 below highlights the accuracy of the predictions for core number 9 under the case 7 scenario.
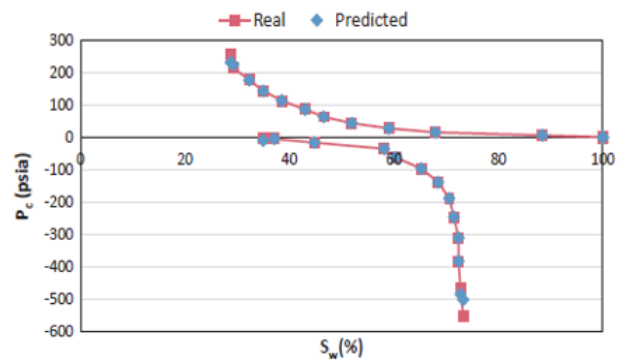
Fig. 4. Core Number 9 Experimental versus Predicted Values [5]

Kasha et al. developed a new method for capillary pressure estimation based on the Leverett J-function as shown in Eq. 8. This is a dimensionless function of water saturation describing capillary pressure also considering pore size and interfacial tension [6].

$$J(S_w) = \frac{P_c(S_w)\sqrt{\frac{k}{\phi}}}{\gamma \cos \theta} \tag{8}$$

$$r = 900.33 \sqrt{\frac{k}{\phi}} \tag{9}$$

Eq. 9 above is the Carman-Kozeny equation which was used to determine the average pore size. The data consisted of 214 samples of mercury injection data with permeability and porosity data and the J versus water saturation for each curve was calculated. The workflow consisted of the formulation of grouping features from porosity, permeability, and pore throat-radius. The formed grouping features can be seen in Eq. 10 and 11.

$$F_1 = 0.01(\ln k \ln r^{0.5} - 3.23) \tag{10}$$

$$F_2 = 0.39\left(\frac{\phi^{0.5}\ln k}{\ln r^{0.5}} - 0.11\right) \tag{11}$$

This approach was compared against the flow zone indicator (FZI) method, which is a rock typing method grouping capillary pressure data into different rock types [6].

$$F_{1(FZI)} = 0.37 \left(\log \frac{\phi}{1-\phi} + 2.71 \right) \quad (12)$$

$$F_{2(FZI)} = 0.28 \left(\log \sqrt{\frac{k}{\phi}} + 1.71 \right) \quad (13)$$

Next an unsupervised machine learning technique (clustering) was used to group each sample based on the grouping features. A technique called the within-cluster sum of squares (WCSS) was used to determine the optimum number of clusters.
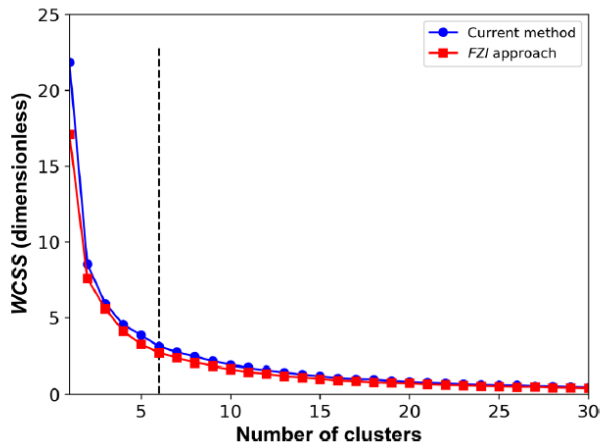
Fig. 5. WCSS Optimum Clusters [6]

The WCSS is the sum of squared deviations between each cluster's data points and the cluster's centroid. In this case, as shown in Fig. 5, four scenarios were considered based on the maximum curvature of 6 and three higher cluster scenarios of 10, 15 and 20 clusters.

The results of the cluster scenarios based on the grouping features can be seen in Fig. 6.
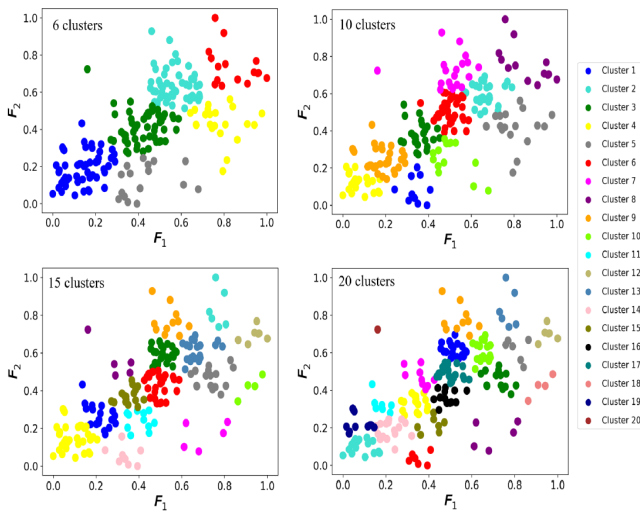
Fig. 6. Clustering Scenarios Based on Grouping Features [6]

From these generated clusters in each of the scenarios, an average J curve was calculated based on the J curves for each

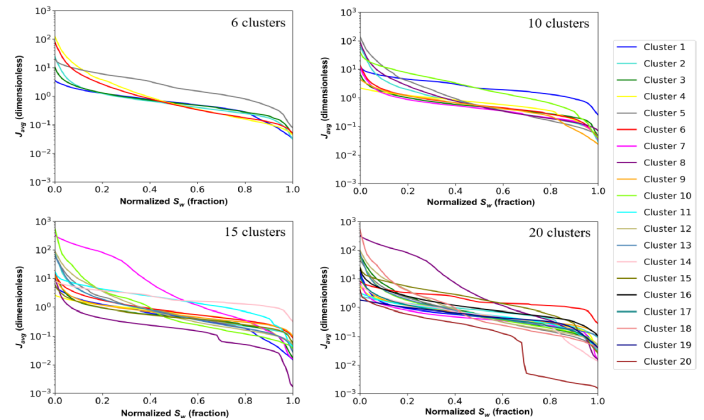sample that were identified in the specific cluster as seen in Fig. 7.

Fig. 7. Average J Curves Based on Clustering Scenarios [6]

An example of the average J curves calculated against the J curves for each sample identified in a cluster is shown in Fig. 8. This is showing the curves for cluster 1 and 2 for the 6 clusters scenario.
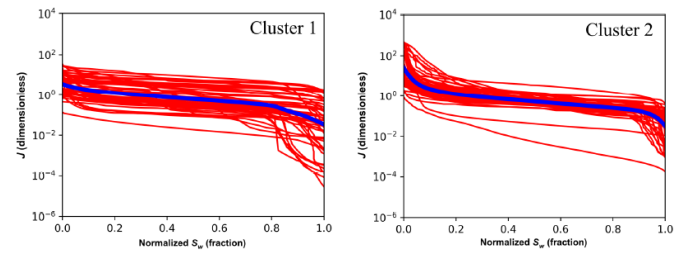

Fig. 8. Cluster 1&2 J Curves for 6 Cluster Scenario [6]

Next, classification algorithms (supervised machine learning) were used to develop maps for the prediction of corresponding clusters for new samples. The algorithms used were K-nearest neighbours (KNN), kernel support vector machine (KSVM), decision tree (DT), and random forest (RF) which were validated by using six carbonate rocks and comparing with the respective experimental data. The best predictions were for the 6 and 10 cluster scenarios with the accuracy for each algorithm found below in Table 3.

Table 3. Classification Algorithm Results [6]

Classification Algorithm	Current Method	
	6 Clusters (%)	10 Clusters (%)
KSVM	97.09	91.24
KNN	95.95	90.62
RF	95.29	87.75
DT	94.12	81.90

As shown, KSVM provided the highest accuracy in the predictions. This algorithm was then subsequently used to classify the validation set (6 carbonate rocks). The rock properties and the grouping features for each can be seen in Table 4 along with the identified cluster for each rock.

Table 4. Summary of Rock Properties and Validation Results [6]

Rock Information				Grouping Feature		Identified Cluster	
Rock Type	k (md)	Φ (%)	r (μm)	F1	F2	6 Clusters (KSVM)	10 Clusters (KSVM)
1	124.350	1.21	91271	0.91	0.10	4	5
2	26.710	11.87	13506	0.69	0.44	4	5
3	5.580	10.05	6709	0.57	0.39	3	6
4	0.586	8.64	2345	0.43	0.35	3	3
5	0.041	6.12	737	0.28	0.27	1	9
6	0.003	4.51	232	0.16	0.20	1	9

Busaleh et al. presented work completed on mercury injection data from carbonate rocks. Although they used a variety of machine learning techniques, results were poor, and more emphasis could have been placed on obtaining better results from a few techniques.

The work from Jamshidian et al. used centrifuge experimental data and an ANN with a Cuckoo optimization algorithm being used to predict capillary pressure with varying inputs. It is important to note here the predictions were completed on capillary pressure experimental centrifuge points with water saturation being used as an input. 7 different cases were conducted with each case involving an additional input parameter which in turn increased the accuracy of the predictions. The limitation here however was the use of the raw centrifuge experimental points. The use of an interpretation software was not used to apply literature correlations to obtain a complete capillary pressure curve.

Kasha et al. took a different approach in using an unsupervised algorithm, clustering, first followed by classification algorithms. The limitation here being the need for an abundance of mercury injection data for many samples to be clustered together and an accurate average J curve for those types of rocks to then be developed. Other approaches could have been taken as well with regards to the pore size distribution such as the Winland r_{35} values.

This research aims to show the novelty of using machine learning techniques to predict complete capillary pressure curves developed from conducting centrifuge tests followed by using SCAL interpretation software to apply literature correlations. This is turn can reduce the time to obtain a complete capillary pressure curve and help improve reservoir characterization which is the end goal in undertaking a SCAL centrifuge experiment. Although several researchers have used ML methods to predict SCAL data, they have focused on predicting a single value of the quantity of interest (relative permeability, capillary pressure, water saturation, etc.). This study assesses the capability of three machine learning algorithms in generating the capillary pressure curves that consists of both capillary pressure and water saturation. Therefore, the novelty of this paper is to simultaneously predict capillary pressure and water saturation values to generate the capillary pressure curve.

3 Methodology

3.1 Centrifuge Capillary Pressure Tests

For this research, multiple centrifuge capillary pressure tests were conducted. The setup consisted of a VINCI RC 4500 – capillary pressure refrigerated centrifuge along with a computer with the appropriate recording and interpretation software. The interpretation software used was CYDAR which offers a powerful solution for the interpretation of conventional and specialized core analysis experiments. CYDAR is also used to record the water saturation and speed versus time during the experiment. A separate VINCI software is used to set the rotational speeds and block times for the centrifuge as well as setting up the camera interfaces for the holding cups.

Core samples were first cleaned followed by drying out in an oven for approximately 1-2 days. Following drying the core samples, their length and diameter were recorded using a Vernier calliper. Next, the dry weight of the cores were recorded. The petrophysical properties of the cores used in the machine learning can be seen in Table 5 below.

Table 5. Core Petrophysical Properties

| Core #	Porosity	Permeability (mD)	Grain Density (g/mL)	Swirr
Core 1	0.222	300	2.011	0.045
Core 2	0.217	180	2.089	0.098
Core 3	0.317	2300	1.761	0.033
Core 4	0.316	2300	1.736	0.064
Core 5	0.308	2300	2.034	0.068
Core 6	0.308	2173	1.811	0.2
Core 7	0.285	2216	1.627	0.23
Core 8	0.265	2557	1.844	0.162
Core 9	0.261	1631	1.864	0.066
Core 10	0.296	2868	2.032	0.306
Core 11	0.2828	3591	1.792	0.058
Core 12	0.267	2900	2.033	0.185
Core 13	0.27	2200	1.686	0.23

Next, a synthetic brine was prepared for the simulation of formation water, the composition is shown in Table 6 below.

Table 6. Synthetic Brine Formulation

Synthetic Brine Formulation		
NaCl	50.397 g/kg	96.96 g/L
CaCl$_2$*2H$_2$O	13.202 g/kg	14.16 g/L
MgCl$_2$*6H$_2$O	2.760 g/kg	2.96 g/L
KCl	0.439 g/kg	0.46 g/L
Na$_2$SO$_4$	0.345 g/kg	0.37 g/L
Subtotal	107.132 g/kg	114.91 g/L
Add Water	892.868 g/kg	957.69 g/L
Total	1000 g/kg	1072.60 g/L

The cores were then loaded into a vessel where the approximately 1 L of brine was pumped in to saturate the cores for approximately one day. Due to the high permeability of the cores one day proved sufficient in saturating the cores. After saturation had occurred, the wet weight of the cores was measured. This data was then used to calculate the porosity and grain density of each sample.

The cores were then loaded into centrifuge core holders where the top, bottom pipe, and sleeve were determined from the core length as shown in Fig. 9a.
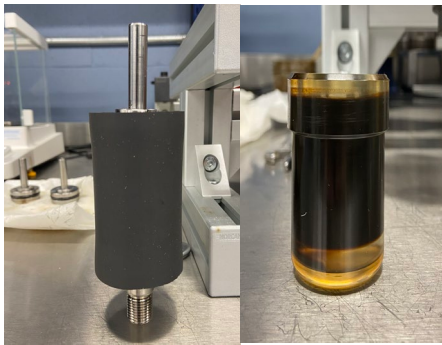
Fig. 9a) Assembled Core Sample (left) **9b)** Holding Cup (right)

The core holders were pressurized to 2000 psi using a hand pump and silicone oil. Small amounts of deionized water and oil were put into the holding cups to create an interface as seen in Fig. 17b. The core holders were also weighed to ensure symmetrical weight before being loaded in the centrifuge.

Once loaded in the centrifuge, the recorded data was inputted in CYDAR, and the recording software was loaded. Multi-step speeds (between 500 to either 3500 or 4500 rpm in increasing 500 increments) as well as step durations were specified, and the camera was set to record the oil-water interface for the respective cups as shown in Fig. 10. This figure shows an oil-water system with the right arrow indicating the water side and the left arrow indicating the oil side. The arrow pointing between the two indicates the oil-water interface. Throughout the drainage experiment as the oil is forced into the core and the water out, the interface will gradually move towards the left. The camera was also checked periodically to ensure it was still properly recording the interface.
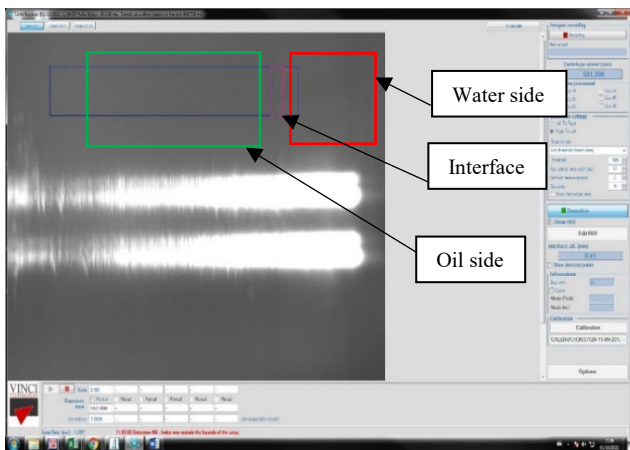
Fig. 10. Camera Interface

Upon completion of the centrifuge test, the next step was to process the recorded data. This required the use of CYDAR interpretation software. The water saturation versus time data was cleaned to exclude any outlier data that the camera may have recorded, and the block times and durations were recorded and entered for CYDAR to calculate the capillary pressure based on the step speeds. For each core, an analytical fit was calculated for the experimental data and Forbes first solution correlation [13] was applied to the experimental data. An example of a complete capillary pressure curve for a

sample after processing can be seen in Fig. 11. A bi-exponential fit was the analytical method used in this case. As shown, both analytical and Forbes methods lack rigor to fit the experimental data.
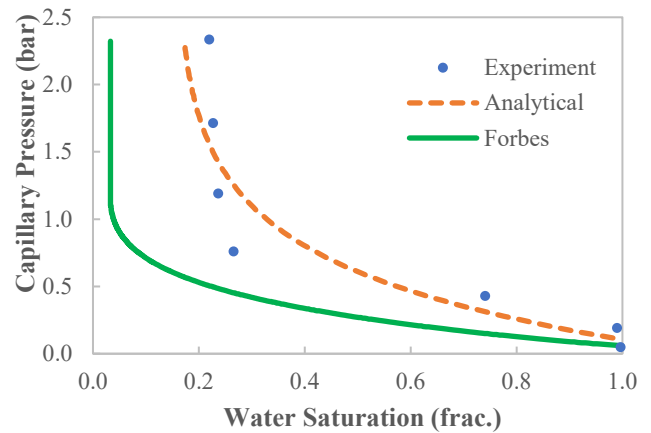

Fig. 11. Core 3 Processed Data (drainage capillary pressure)

For each of the 13 cores the Forbes first solution was applied to the experimental data. Using CYDAR, data from the Forbes curves was discretized to be used for the machine learning techniques in order to predict the Forbes curves.

3.2 Machine Learning Techniques

A spreadsheet was developed to collect all the test data for the machine learning techniques before pre-processing of the data from 13 core samples. The machine learning workflow followed in this case can be seen in Fig. 12.
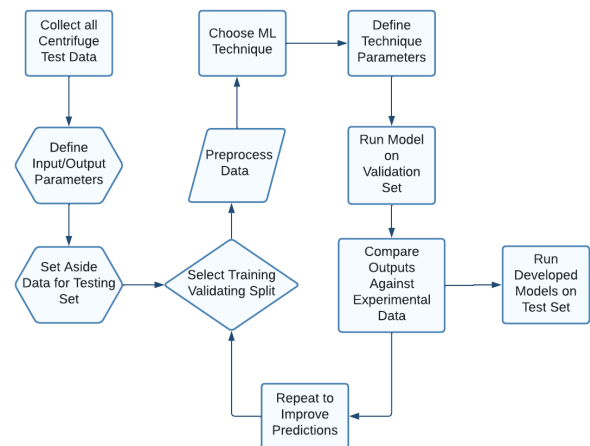
Fig. 12. Machine Learning Workflow

The inputs for the machine learning techniques as mentioned previously were the routine core analysis data of porosity, permeability, grain density, and irreducible water saturation. The outputs were water saturation and capillary pressure; subsequently, multi-output regression models were developed. Data for 2 core samples were put into a separate file to be used as a test set after model development. Note that for these 2 samples, 1995 (input, output) data points were recorded. For the remaining 11 cores, a training-testing split of 70-30 was used to train the models. Using these 11 cores,

10,955 (input, output) data points were recorded and used to train and validate the ML models. Depending on the technique used, the parameters were defined for each; for ANN the epochs, activation function, hidden layers, and neurons, for RFR the number of trees, and for support vector regression the tolerance and max iterations. Standardization and normalization feature scaling were also implemented to see how they would affect overall performance. Eqs. 14 and 15 show the methods for standardization and normalization, respectively.

$$X' = \frac{X - \mu}{\sigma}, \tag{14}$$

where μ is the mean of the feature values and σ is the standard deviation of the feature values.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}, \tag{15}$$

where X_{min} and X_{max} are the minimum and maximum of the feature, respectively.

Once the models were run on the training and testing data, their performance was measured using 2 measurements, including the mean squared error (MSE) and correlation coefficient (R^2) where n corresponds to the total number of data points.

$$MSE = \frac{1}{n} . \Sigma(actual - predicted)^2 \tag{16}$$

Finally, once all models were developed on the validation data set, they were then tested and evaluated on the unseen data set. All machine learning methods were implemented using Python 3.7 programming language along with Keras library with Tensorflow 2.3 backend.

4 Results and Discussion

The results from all ML algorithms on the validation data is shown in Table 7.

Table 7. ML Models Results – Test Data

ML Algorithm	MSE	R² (Validation Data)
Random Forest	2.09×10⁻⁵	0.9997
Neural Network	0.00222	0.9937
Support Vector	0.00436	0.9543

All models performed well with random forest being the best with a correlation coefficient of 0.9997. Figs. 13 and 14 show the prediction versus experimental data for ANN for capillary pressure and water saturation, respectively.
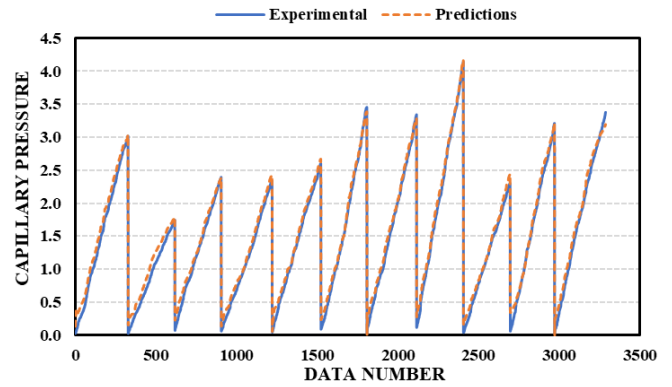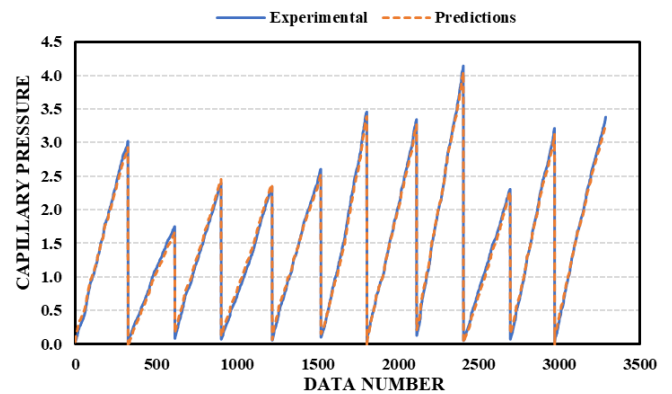

Fig. 13. ANN Validation Data Results – Capillary Pressure
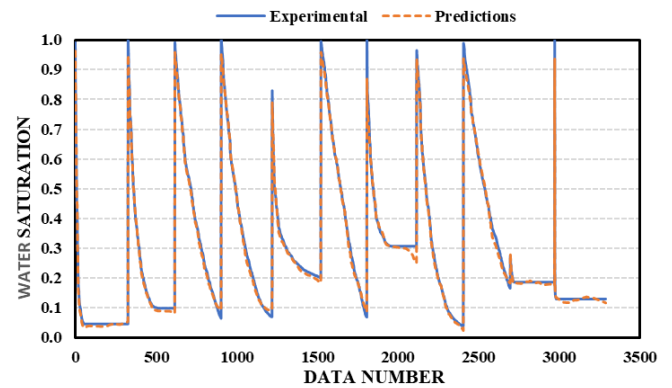
Fig. 14. ANN Validation Data Results – Water Saturation

Similarly, Figs. 15 and 16 highlight the results for the SVR model and Fig. 17 and 18 are the results for the RFR model on the testing data.

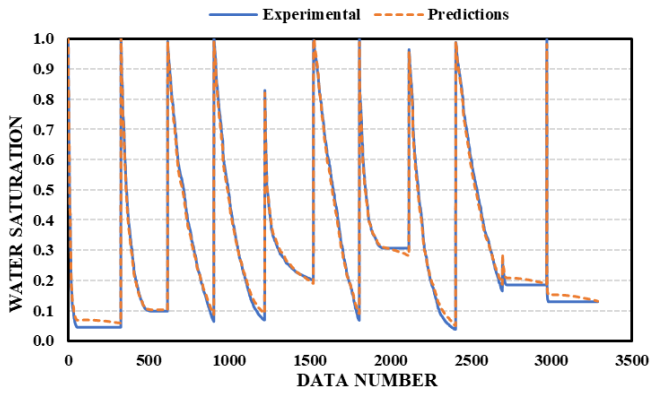**Fig. 15.** SVR Validation Data Results – Capillary Pressure

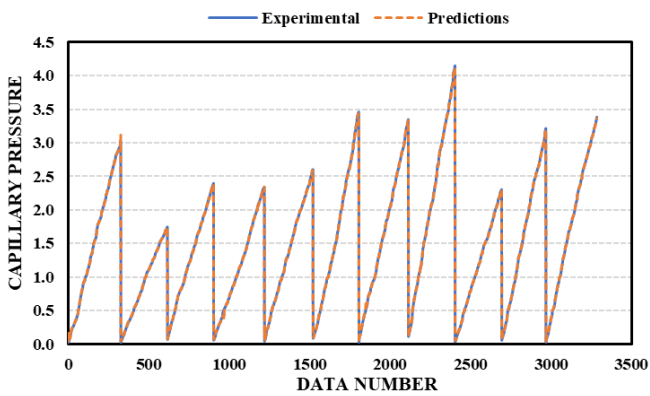**Fig. 16.** SVR Validation Data Results – Water Saturation

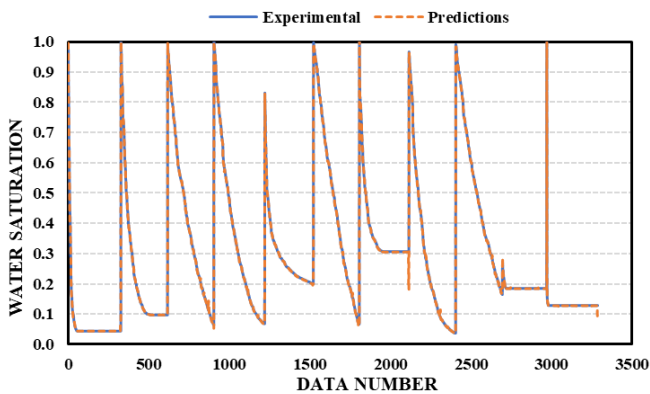**Fig. 17.** RFR Validation Data Results – Capillary Pressure

Fig. 18. RFR Validation Data Results – Water Saturation

A sensitivity analysis was also completed on the ANN parameters of hidden layers, neurons, and activation functions. The base case for the following graphs is a model with 32 neurons, 2 hidden layers and a rectified linear unit (ReLU) activation function (i.e. when seeing the effects of changing neurons, 2 hidden layers and ReLU activation function is used).
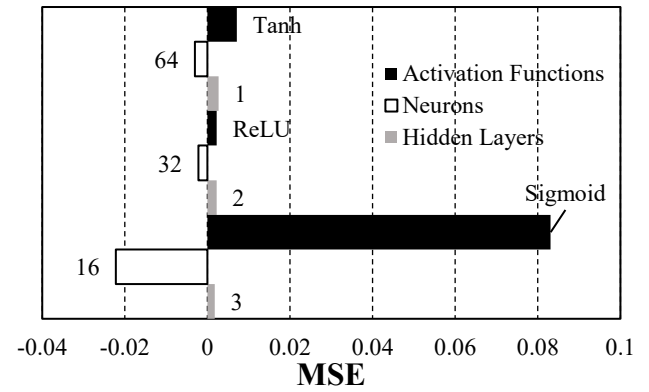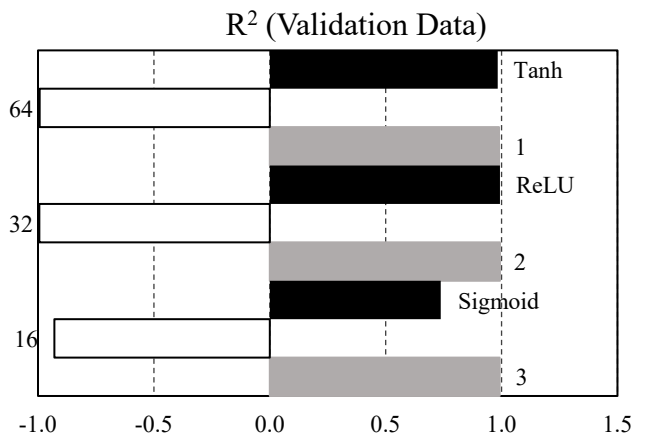

Fig. 19. Varying ANN MSE Results

Fig. 20. Varying ANN R^2 Results

As can be seen from Figs. 19&20 the biggest change in prediction accuracy comes with the change in using Tanh as the activation function. Minimal changes occurred in the prediction accuracy with varying the hidden layers with 2 hidden layers performing the best. Finally, for the varying of neurons, the greatest error occurred when using 16, while the greatest prediction happened using 32.

With the ML models developed, the test set of the 2 cores that was set aside was then given to the models for predictions. The following graphs highlight the accuracy of each model showing the test data results with the results on the test set when using normalized and standardized feature scaling.
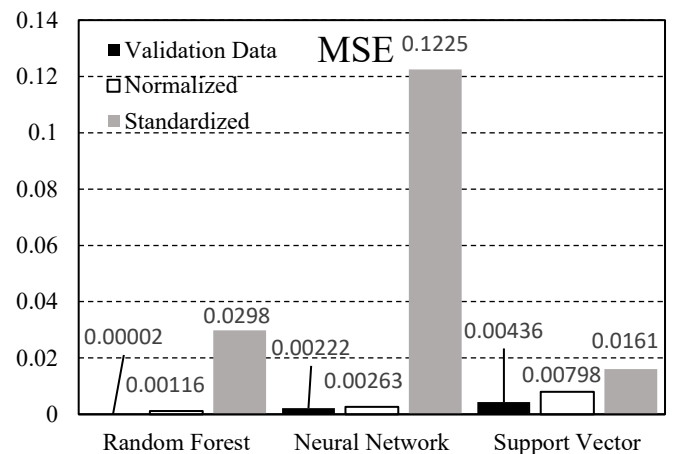
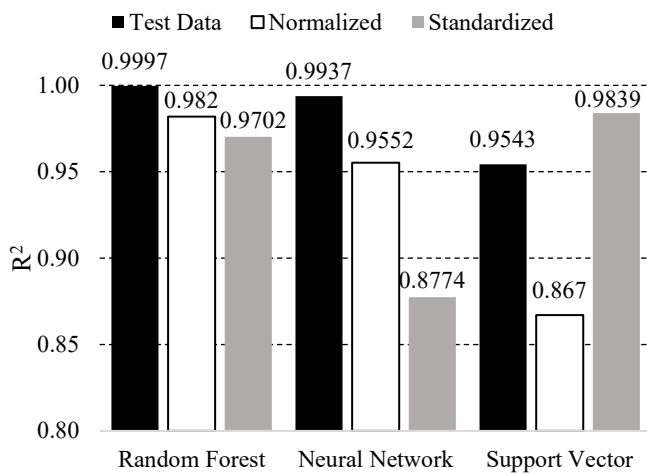**Fig. 21.** MSE Results of Models for Test and Validation Data

Fig. 22. R^2 Results of Models for Validation Data

Following the evaluation of the validation data, unseen data consisting of 2 cores was used to test the accuracy of the models. The results for each of the models can be seen in Table 8.

Table 8. ML Models Results – Unseen Data

ML Algorithm	MSE	R^2 (Unseen Data)
Random Forest	0.0248	0.9734
Neural Network	0.1034	0.8886
Support Vector	0.0438	0.9527

As shown in Fig. 23, although RFR performed the best overall on the validation data and unseen data with regards to the performance metrics, it is very noisy when predicting unseen data. SVR performed the greatest overall with a smoother curve, however the entry capillary pressure was predicted higher.
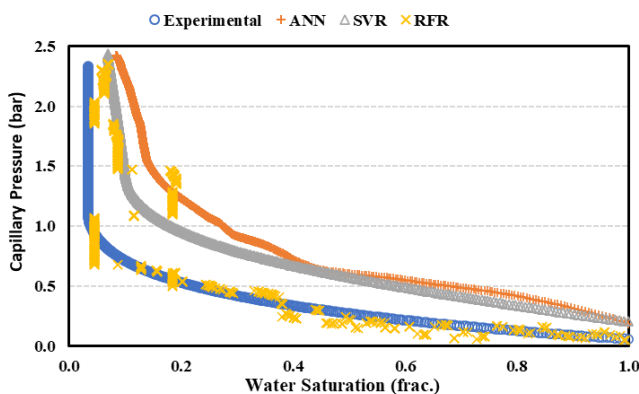

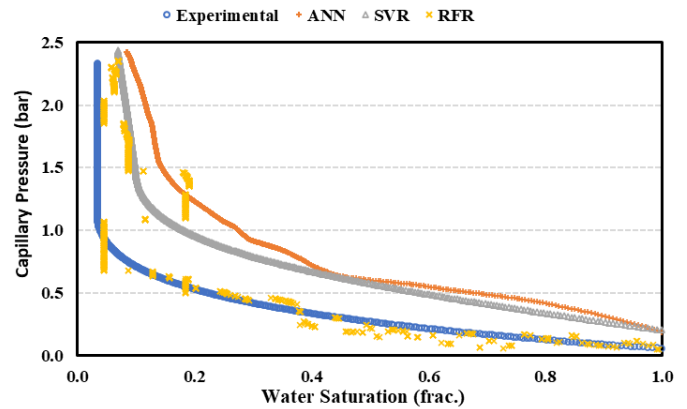**Fig. 23.** Unseen Data Model Results – First Core

Fig. 24. Unseen Data Model Results – Second Core

Similar to the first core, RFR was noisy again in predictions for the second core. ANN and SVR curves were smoother but showed differences in the entry capillary pressure.

5 Conclusions

This work investigated various machine learning techniques in predicting complete drainage capillary pressure curves. Centrifuge capillary pressure tests were conducted with SCAL interpretation software being used to apply the Forbes First solution correlation. Data for 13 cores was obtained along with the respective RCAL data of porosity, permeability, grain density, and irreducible water saturation for each core. Once all the data was collected, three machine learning models were developed using training and validation data sets with some of the collected data set aside to be used as a test data set. Some of the key findings from this work were that the three techniques used all performed very well on the validation data set as evident from the performance metrics used. Another takeaway is the effectiveness of conducting a sensitivity analysis as evident in the results from the ANN model when changing the neurons, hidden layers, and activation function used.

Some recommendations for future work are to conduct centrifuge experiments on more cores of different rock types to see how the models will perform when using a variety of rock types. As well as, incorporating additional input data such as pore size distribution to see the effects. Additional cores could also be added to the unseen data for the models to increase the accuracy of the predictions. With the success of applying machine learning techniques to capillary pressure, future work can also be to conduct other SCAL tests such as relative permeability to see the effectiveness of machine learning techniques in predicting other reservoir characteristics.

Acknowledgements

References

1. S.B. Suslick, D. Schiozer, M.R. Rodriguez. Uncertainty and Risk Analysis in Petroleum Exploration and Production. Terrae **6**(1): 30-21 (2009). https://www.ige.unicamp.br/terrae/V6/PDF-N6/T-a3i.pdf

2. C. McPhee, J. Reed, I. Zubizarreta. Core Analysis – A Best Practice Guide. https://app.knovel.com/hotlink/toc/id:kpCAABPG03/core-analysis-best-practice/core-analysis-best-practice

3. A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, H. Oza. Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Research **6**, 379-391, (2021).

4. Y.R. Busaleh, A, Abdulraheem, T. Okasha. Prediction of Capillary Pressure for Oil Carbonate Reservoirs by Artificial Intelligence Technique. SPE Asia Pacific Oil and Gas Conference Exhibition. (2016). https://doi.org/10.2118/182173-MS

5. M. Jamshidian, M. Mansouri Zadehm M. Hadian, R. Moghadasi, O. Mohammadzadeh. A Novel Estimation Method for Capillary Pressure Curves Based on Routine Core Analysis Data Using Artificial Neural Networks Optimized by Cuckoo Algorithm – A Case Study. (2018). Fuel, **220**, 363–378. https://doi.org/10.1016/j.fuel.2018.01.099

6. A.A. Kasha, A. Sakhaee-Pour, I.A. Hussein. 'Machine Learning for Capillary Pressure Estimation.' SPE Res Eval & Eng **25**. (2022): 1-20. doi: https://doi-org.qe2a-proxy.mun.ca/10.2118/208579-PA

7. T. Mitchell. Machine Learning. New York: McGraw Hill. (1997). OCLC 36417892.

8. Mathworks Help Center. Machine Learning in MATLAB https://www.mathworks.com/help/stats/machine-learning-in-matlab.html

9. T. Hastie, R. Tibshirani, J. Friedman. The Eléments of Statistical Learning. Springer New York. https://doi.org/10.1007/978-0-387-84858-7_12

10. A. Shmilovici. Support Vector Machines. In : O. Maimon, L. Rokach. (eds) Data Mining and Knowledge Discovery Handbook. (2005). Springer, Boston, MA. https://doi-org.qe2a-proxy.mun.ca/10.1007/0-387-25465-X_12

11. Rodríguez-Pérez, R., Bajorath, J. Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery. J Comput Aided Mol Des (2022). https://doi-org.qe2a-proxy.mun.ca/10.1007/s10822-022-00442-9

12. K. Fawagreh, M.M. Gaber, E. Elyan. Random forests: from early developments to recent advancements. System Science and Control Engineering. Volume 2. (2014). https://doi-org.qe2a-proxy.mun.ca/10.1080/21642583.2014.956265

13. P. Forbes. Simple and Accurate Methods for Converting Centrifuge Data into Drainage and Imbibition Capillary Pressure Curves. SPWLA-1994-V35n4a3, 35(04), 13. (1994).

14. Z.A. Chen, D.W. Ruth. Measurement and Interpretation of Centrifuge Capillary Pressure Curves-The SCA Survey Data. The Log Analyst, 36(05), 13. (1995).

15. P. Forbes. Centrifuge Data Analysis Techniques : An SCA Survey on the Calculation of Drainage Capillary Pressure Curves from Centrifuge Measurements. 20. (1997).

16. J.E. Nordtvedt, K. Kolltvelt. Capillary Pressure Curves From Centrifuge Data by Use of Spline Functions. SPE-19019-PA, 6(04), 497–501. https://doi.org/10.2118/19019-PA (1991).